

Gesundheitsdaten als sozialpolitisches Gut

Warum wir eine Debatte über ihre Nutzung für die Forschung brauchen

Jan Paul Heisig, Jianghong Li und Jutta Allmendinger

Summary: Effective design of social and health policy requires a solid evidential basis. In Germany, the potentials of “routine” data – data that are routine byproducts of the everyday workings of the health care or social insurance systems – remain relatively untapped. We need a broad societal debate about the potentials and the pitfalls of using routine data for purposes that serve the public interest. This debate should be informed by examples from other countries, including the West Australian Data Linkage System described in this article.

Kurz gefasst: Eine effektive und zielgenaue Gestaltung von Gesundheits- und Sozialpolitik bedarf einer soliden empirischen Grundlage. Die Potenziale sogenannter „Routinedaten“, die bei den alltäglichen Routinen des Gesundheitssystems oder der Sozialversicherung quasi nebenbei anfallen, werden in Deutschland bisher nur unzureichend genutzt. Es braucht eine breite gesellschaftliche Diskussion über Chancen und Risiken einer gemeinwohlorientierten Nutzung von Routinedaten. Konkrete Beispiele aus anderen Ländern – wie das hier vorgestellte West Australian Data Linkage System – können dabei wertvolle Anregungen liefern.

Armut kann krank machen, und Krankheit kann zu Armut führen. Sozialpolitik ist deswegen auch Gesundheitspolitik, Gesundheitspolitik ist auch Sozialpolitik – und eine moderne, effektive und zielgenaue Gestaltung von Gesundheits- und Sozialpolitik bedarf einer soliden empirischen Grundlage. Nur mithilfe umfangreicher, detaillierter und weitgehend unverzerrter Daten können soziale und gesundheitliche Ungleichheiten beschrieben, Handlungsbedarf identifiziert und die Wirkungen politischer Interventionen quantifiziert werden.

Die Datengrundlage ist in Deutschland grundsätzlich gut, aber sie könnte – so eine zentrale These unseres Beitrags – noch besser sein. Die Datengrundlage ist gut, weil wir über eine Reihe Befragungen verfügen, die regelmäßig Informationen zur sozialen und gesundheitlichen Lage der Bevölkerung erheben – in vielen Fällen sogar als Längsschnittstudien, die dieselben Personen wiederholt befragen und damit individuelle Verlaufsmuster abbilden können. Wichtige Beispiele hierfür sind die verschiedenen Studien des Robert Koch-Instituts, die NAKO Gesundheitsstudie, das Sozio-oekonomische Panel oder das Nationale Bildungspanel. Dieses breite Portfolio an Befragungsdaten ließe sich durch größere Fallzahlen oder inhaltliche Ergänzungen noch weiter ausbauen. Wenn wir sagen, dass die Datengrundlage besser sein könnte, meinen wir jedoch vor allem, dass die Potenziale sogenannter „Routinedaten“ in Deutschland bisher nur unzureichend genutzt werden. Hierbei handelt es sich um Daten, die bei den alltäglichen Routinen des Gesundheitssystems, der Sozialversicherung oder der Arbeitsverwaltung quasi nebenbei entstehen – oft werden sie deshalb auch als „prozessproduziert“ bezeichnet.

Wichtige Beispiele für prozessproduzierte Gesundheitsdaten sind die Diagnose-, Behandlungs- und Abrechnungsdaten, die bei niedergelassenen Ärztinnen und Ärzten und in Krankenhäusern anfallen und zum Teil an die Krankenkassen gemeldet werden. Routinedaten, die Aufschluss über Erwerbs- und Einkommensverläufe geben können, entstehen zum Beispiel im Kontext von Arbeitslosen- und Rentenversicherung. Aufgrund der fortschreitenden Digitalisierung von Verwaltungsprozessen liegen solche Routinedaten zunehmend in digitaler Form vor. Das eröffnet neue Chancen, weil verschiedene Datenquellen ohne großen Aufwand verknüpft werden können und sich der Kreis der Datennutzer*innen fast ohne zusätzliche Kosten erweitern lässt. Es entstehen aber auch neue Risiken. Gerade durch das Zusammenfügen von Datenquellen kann ein „Buch des Lebens“ entstehen, das zwar für Forschungszwecke sehr nützlich ist, aber eben auch höchst sensible Informationen von vor der Geburt bis hin zum Tod enthält. Wird dieses Buch unzureichend abgesichert, können verheerende Formen des Datenmissbrauchs oder -diebstahls die Folge sein.

Warum sind Routinedaten eine potenziell so wichtige Ergänzung der vorhandenen Befragungsdaten? Ein entscheidender Vorteil ist die enorme Größe prozessproduzierter Datensätze. Krankenversicherungs- und ärztliche Abrechnungsdaten beispielsweise liegen für fast jede in Deutschland lebende Person vor, wenn auch bei verschiedenen Trägern. Die Generierung ähnlich großer Datensätze mithilfe von Befragungen wäre mit enormem Aufwand und Kosten verbunden. Die Größe von Routinedatensätzen erlaubt zum Beispiel eine deutlich feingliedrigere räumliche Unterteilung als klassische Umfragedaten. So lassen sich neue Fragen stellen, die mit herkömmlichen Daten kaum zu beantworten

wären, etwa zum Zusammenhang zwischen Wohnumgebung und Gesundheit. Auch seltene, aber oftmals verheerende Erkrankungen wie zum Beispiel bestimmte Krebsarten oder schwere psychische Erkrankungen wie posttraumatische Belastungsstörungen lassen sich mit Routinedatensätzen untersuchen, während sie in herkömmlichen Befragungsdatensätzen kaum oder sogar überhaupt nicht vorkommen. Oft sind große Fallzahlen zudem eine entscheidende Voraussetzung für die Untersuchung von Kausalzusammenhängen mit modernen quasiexperimentellen Verfahren. Viele dieser Verfahren sind auf Vergleiche zwischen sehr spezifischen Subgruppen angewiesen, um daraus Schlussfolgerungen über kausale Effekte von gesundheits- und sozialpolitischen Interventionen oder anderen Faktoren zu ziehen. In regulären Befragungen sind die relevanten Subgruppen oftmals nicht groß genug, um empirische Befunde statistisch hinreichend abzusichern.

Ein weiterer wichtiger Vorteil verknüpfter digitaler Routinedaten ist ihre schnelle Verfügbarkeit. So können sie einen wichtigen Beitrag zum sogenannten *nowcasting*, zur Gesellschaftsbeobachtung in Echtzeit, leisten. In Verbindung mit anderen Datenquellen, etwa zum Suchverhalten im Internet, können Routinedaten so zum Beispiel effizientere Frühwarnsysteme für die Ausbreitung von ansteckenden Krankheiten wie Covid-19 oder für unentdeckte Nebenwirkungen von Medikamenten ermöglichen. Mögliche Bedrohungen für die öffentliche Gesundheit können schneller und zuverlässiger erkannt werden.

Die gemeinwohlorientierte Nutzung und Verknüpfung von Gesundheitsdaten ist in vielen anderen Ländern bereits weit fortgeschritten. Fälle von schwerem Missbrauch oder Datendiebstahl sind bisher kaum bekannt, was aber nicht bedeutet, dass solche Ereignisse unvorstellbar sind – und natürlich können wir auch nicht ausschließen, dass es bereits zu derartigen Vorfällen gekommen ist, ohne dass eine breitere Öffentlichkeit davon erfahren hätte. Auch in Deutschland ist bereits eine begrenzte Nutzung von Routinedaten möglich. Bestimmte Daten der Arbeitslosen- und Rentenversicherung etwa können für Forschungszwecke genutzt werden. Auch die Nutzung von Daten der gesetzlichen Krankenkassen ist möglich; der Zugang erfolgt jedoch über die einzelnen gesetzlichen Krankenkassen. Das ist nicht nur sehr aufwendig, sondern es schmälert den Nutzen der Daten auch deshalb, weil viele Projekte sich auf die Versicherten einer einzelnen Krankenkasse beschränken müssen, was sich nachteilig auf die Repräsentativität der Untersuchungspopulation auswirkt.

Die gesellschaftliche Diskussion über das Für und Wider der Datennutzung und über angemessene Regeln für den Umgang mit Gesundheitsdaten in Deutschland hinkt allerdings weit hinterher. Das zeigte sich zuletzt in der zum Teil durchaus berechtigten Kritik an Teilen des Digitale-Versorgung-Gesetzes aus dem Jahr 2019, insbesondere an der Schaffung eines Forschungsdatenzentrums, das die Abrechnungsdaten der verschiedenen gesetzlichen Krankenkassen in pseudonymisierter Form zusammenführen und für die Forschung verfügbar machen soll.

Deutschland braucht in dieser Frage den öffentlichen Diskurs. Wird er nicht geführt, bleiben die Potenziale verknüpfter Gesundheits- und Sozialdaten weiter weitgehend ungenutzt, zum Nachteil von Forschung, Gesundheits- und Sozialsystem und Patient*innen. Trotz der Chancen sollten aber auch die erheblichen Risiken nüchtern und vorbehaltlos anerkannt werden. Ohne Frage können die unsachgemäße Nutzung oder gar Veröffentlichung sensibler Gesundheitsdaten gravierende Folgen haben, etwa in Form von sozialer Stigmatisierung oder Diskriminierung auf dem Arbeits- und Versicherungsmarkt. Datensicherheit und Datenschutz müssen daher höchste Priorität haben.

Wichtig ist, dass der nötige gesellschaftliche Diskurs nicht bei null anfangen muss. Die Erfahrungen anderer Länder und Expert*innenempfehlungen, wie sie in den letzten Jahren etwa im Rahmen der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) erarbeitet wurden, können wichtige Impulse für die Diskussion in Deutschland liefern. So listet eine offizielle Empfehlung des OECD-Rates zur Regulierung von Gesundheitsdaten aus dem Jahr 2016 zwölf Kernprinzipien für die nationalen Regulierungsrahmen auf.



Jan Paul Heisig ist Leiter der Forschungsgruppe Gesundheit und soziale Ungleichheit. Seine Arbeitsschwerpunkte sind soziale Ungleichheiten in den Bereichen Gesundheit, Bildung und Arbeitsmarkt sowie statistische Methoden. (Foto: David Ausserhofer)

jan.heisig@wzb.eu

Jianghong Li ist wissenschaftliche Mitarbeiterin in der Forschungsgruppe der Präsidentin. Sie erforscht soziale, wirtschaftliche und kulturelle Determinanten der Gesundheit und Entwicklung von Kindern (siehe auch ihren Artikel im Heft).

jianghong.li@wzb.eu

Jutta Allmendinger ist Präsidentin des WZB und Professorin für Bildungssoziologie und Arbeitsmarktforschung an der Humboldt-Universität zu Berlin.

jutta.allmendinger@wzb.eu

Literatur

Dunn, Halbert L.: „Record Linkage“. In: *American Journal of Public Health*, 1946, Jg. 36, H. 12, S. 1412–1416.

Hansen, Michele/Kurinczuk, Jennifer J./Bower, Carol/Webb, Sandra: „The Risk of Major Birth Defects after Intracytoplasmic Sperm Injection and In Vitro Fertilization“. In: *New England Journal of Medicine*, 2002, Jg. 346, H. 10, S. 725–730.

Hofmann, Sarah/Mühlenweg, Andrea: „Eine Perspektive für Analysen des Zusammenhangs von Bildung und Gesundheit: Potential und Limitationen von Routinedaten der Gesetzlichen Krankenversicherung (GKV)“. In: *Sozialer Fortschritt*, 2017, Jg. 66, H. 1, S. 49–75.

Holman, C. D'Arcy J./Bass, A. John/Rosman, Diana L./Smith, Merran B./Semmens, James B. et al.: „A Decade of Data Linkage in Western Australia: Strategic Design, Applications and Benefits of the WA Data Linkage System“. In: *Australian Health Review*, 2008, Jg. 32, H. 4, S. 766–777.

OECD: *Recommendation of the Council on Health Data Governance*, OECD/LEGAL/0433. Paris: Organisation for Economic Co-operation and Development 2019.

Jenseits solcher allgemeinen Prinzipien lässt sich auch von positiven Beispielen viel lernen. Der australische Bundesstaat Westaustralien hat im Laufe der letzten Jahrzehnte eine beispielhafte vernetzte Dateninfrastruktur geschaffen. Das Western Australia Data Linkage System (WADLS) verknüpft Gesundheits- und Sozialdaten aus 39 Quellen über einen Zeitraum von bis zu 50 Jahren. Neben Gesundheitsdaten beinhaltet das WADLS eine Vielzahl weiterer Routinedaten etwa aus Bildung, Justiz- und Strafvollzug, Kinderschutz und Wohnen. Über die verschiedenen Routinedatensätze hinaus wurde eine Reihe von weiteren Datensätzen aus Bevölkerungsumfragen oder medizinischen Fall-Kontroll-Studien in das WADLS integriert. Diese Verknüpfung kann die Potenziale solcher sozialwissenschaftlichen und medizinischen Datensätze deutlich erhöhen, weil Querschnittbefragungen mit längsschnittlichen Lebensverlaufsdaten aus den Routinedaten angereichert werden können.

Das WADLS wird intensiv genutzt und genießt hohe gesellschaftliche Akzeptanz. Dafür gibt es eine Reihe von Erklärungen. Ein entscheidender Faktor ist konsequente Einbeziehung verschiedenster gesellschaftlicher Gruppen. Die Zusammenarbeit von Forschung, Gesundheitswesen, Sozialversicherung, Verwaltung und Patient*innenvertretungen war entscheidend für die Entwicklung einer effizienten technischen Infrastruktur und für die Etablierung von Sicherheits- und Nutzungsprotokollen mit breiter gesellschaftlicher Akzeptanz. Ein wichtiger Baustein der dauerhaften Patient*innenbeteiligung ist das Amt einer beziehungsweise eines vollzeitbeschäftigten Patient*innenbeauftragten. Die Sichtbarkeit von Forschungsergebnissen auf Grundlage des WADLS, unter anderem durch eine wöchentliche Beilage in der wichtigsten Tageszeitung des Bundesstaats, unterstreicht den Kollektivgutcharakter des Systems.

Eine entscheidende Rolle für die Akzeptanz des WADLS spielt sicher auch das anspruchsvolle Datensicherheitskonzept. Verknüpfung, Speicherung und Auswertung der WADLS-Daten unterliegen strengen Regeln. Forschungsprojekte werden in der Regel mehrfach geprüft und auch die Datennutzer*innen müssen bestimmte Anforderungen erfüllen, um Zugriff auf vertrauliche Daten zu erhalten (zum Beispiel Überprüfung möglicher Vorstrafen). Ein zentrales Instrument zum Schutz der Privatsphäre ist das sogenannte „Trennungsprinzip“ („separation principle“): Die einzelnen prinzipiell verknüpfbaren Datensätze werden dezentral und getrennt voneinander gespeichert und verwaltet. Im WADLS existiert somit kein vollständiger integrierter Datensatz, der sämtliche Informationen aus allen prinzipiell verfügbaren Datensätzen enthielte. So könnte das detaillierte und sensible „Buch des Lebens“ zwar prinzipiell durch Zusammenführen aller im System vorhandenen Daten „geschrieben“ werden; in der Praxis geschieht dies aber nicht. Zudem werden alle Informationen, die eine unmittelbare Identifikation von Einzelpersonen ermöglichen (zum Beispiel Adresse, Geburtsdatum, Sozialversicherungsnummer), grundsätzlich getrennt von den individuellen Gesundheits- und Sozialdaten gespeichert. In der Langfassung unseres Beitrags (siehe Literaturhinweis) werden die Bausteine des WADLS-Datenschutzkonzepts ausführlich beschrieben.

Das Beispiel zeigt, wie eine umfassende nutzer*innenfreundliche und datenschutzkonforme Infrastruktur für verknüpfte Gesundheits- und Sozialdaten aussehen könnte. Es zeigt zudem, wie sich eine breite gesellschaftliche Akzeptanz einer solchen Infrastruktur erreichen lässt: durch die konsequente Aufklärung über den Nutzen einer gemeinwohlorientierten Verwendung von Gesundheitsdaten und durch die Einbindung verschiedener Interessengruppen in die Entwicklung und Überwachung des Systems. Deutschland muss den westaustralischen Ansatz nicht eins zu eins kopieren, aber es wäre fahrlässig, das Thema einer gemeinwohlorientierten Nutzung von prozessproduzierten Gesundheitsdaten nicht entschiedener anzugehen. Die Corona-Pandemie hat uns zuletzt deutlich vor Augen geführt, wie wichtig eine solide Datenlage und darauf basierende wissenschaftliche Erkenntnisse auch für weitreichende politische Entscheidungen sein können.

Dieser Beitrag ist die überarbeitete und gekürzte Fassung eines Kapitels, das unter dem Titel „Gesundheitsdaten als öffentliches Gut“ in dem von Philip van der Eijk, Detlev Ganten und Roman Marek herausgegebenen Band „Was ist Gesundheit? Interdisziplinäre Perspektiven aus Medizin, Geschichte, Philosophie und Religion“ bei De Gruyter erscheinen wird.