Dr. rer. pol. Dominik von Stillfried

Big Data und Digitalisierung: Handlungsbedarf für Versorgungsforschung

Zusammenfassung

Datengrundlagen und Analysemöglichkeiten für die Versorgungsforschung nehmen rapide zu. Soll Deutschland als Forschungsstandort internationale Bedeutung behalten, reicht allein eine finanzielle Förderung der Versorgungsforschung nicht aus. Vielmehr bedarf es einer politischen Initiative zur Schaffung einer nationalen Forschungsinfrastruktur. Diese sollte auf dem Prinzip des data linkage von administrativen Daten, klinischen Daten und unstrukturierten Alltagsdaten beruhen. Notwendig ist dafür eine Rechtsgrundlage, die den Aufbau von Datenbanken über längere Zeitreihen und die Zusammenführung von Daten unterschiedlicher Quellen auf pseudonymisierte Einzelpersonen gestattet. Deutschland kann sich hierbei auf internationale Erfahrungen wie z.B. aus dem australischen Population Health Research Network stützen.

Schlüsselwörter

Digitalisierung, Big Data, Versorgungsforschung, Forschungsinfrastruktur, Forschungsstandort Deutschland

Die Gesundheitssysteme der Industrieländer sind extrem komplex. Wer sie verstehen, steuern oder verbessern will, braucht dazu zwingend eine funktionsfähige Versorgungsforschung. Wer außerdem Innovationen im Kontext von Alltagssituationen bewerten will, kommt an einer leistungsfähigen Versorgungsforschung nicht vorbei. Darüber gibt es keinen Dissens in Deutschland. Förderprogramme des BMBF in den Aufbau von Forschungsstrukturen und für die Zusammenführung von Daten sowie die Verankerung des Innovationsfonds im Sozialgesetzbuch zeugen auch von der Bereitschaft, in die Entwicklung der Versorgungsforschung in Deutschland zu investieren. Allein, es fehlt an den notwendigen Rechtsgrundlagen, aufgrund derer Finanzinvestitionen in die Versorgungsforschung zur Entwicklung einer nachhaltigen und leistungsfähigen Forschungsinfrastruktur heranreifen können. Ohne Daten gibt es keine Forschung, und ohne geeignete Rechtsgrundlage fehlt der deutschen Versorgungsforschung der notwendige Datenzugang. Der formale Anlass, um dieses „heiße Eisen" anzupacken ist gegeben: Die anstehende Umsetzung der europäischen Datenschutzverordnung in deutsches Recht. Der Zeitrahmen dafür ist begrenzt, nämlich bis Ende 2018. Allzu lange sollte die (alte und die neue) Bundesregierung auch aus anderen Gründen nicht warten. Andere Parlamente, etwa im Vereinigten Königreich, haben den internationalen Wettbewerbsdruck längst erkannt und bereits gehandelt. In den nächsten Jahren wird sich entscheiden, ob Deutschland bei ‚Big Data' in der Versorgungsforschung führend oder langfristig ein internationales Nischendasein in der Versorgungsforschung führen wird. An dieser Frage wird sich auch entscheiden, ob Deutschland perspektivisch weiter ein Exportland von Gesundheitstechnologie sein oder zu einem Importland werden wird. Wer den Forschungsstandort Deutschland auch im Bereich der Versorgungsforschung an der Weltspitze halten will, muss jetzt die Ressorts Forschung, Gesundheit, Wirtschaft und Inneres dazu verpflichten, ein gemeinsames Aktionsprogramm Versorgungsforschung aufzulegen, welches baldmöglichst vom Bundestag beschlossen werden müsste.

>> Um die Voraussetzungen für die Versorgungsforschung in Deutschland langfristig günstiger zu gestalten, muss der Gesetzgeber jetzt mit Hochdruck an zwei Stellen ansetzen.

Erstens: Die Versorgungsforschung braucht für die meisten Fragestellungen Daten von Einzelpersonen und diese in möglichst langen Zeitreihen. Der Aufbau einer entsprechenden Forschungsdatenbank ist nach geltenden Rechtsmaßstäben im Bereich der Sozialdaten nur dann zulässig, wenn die Zielsetzung des Forschungsvorhabens eindeutig speziell diese Datensätze in Zeitreihe erfordert; nach Abschluss der Analyse muss die Forschungsdatenbank gelöscht werden (§ 75 SGB X). Das erhöht nicht nur massiv den Personal-, Zeit-, und Finanzaufwand, bis eine Auswertung überhaupt starten kann. Eine Forschungsdatenbank kann somit gewissermaßen auch nur im Rückspiegel aufgebaut werden, denn welche Forschungsfragen in fünf oder zehn Jahren relevant sein werden, vermag heute niemand präzise zu sagen. Heute bereits Datengrundlagen für künftige Forschungsfragen aufzubauen, fällt unter das Unwort „Vorratsdatenhaltung" und ist verboten.

Zweitens: Die Versorgungsforschung muss, um aussagefähige und praktisch nutzbare Ergebnisse zu erzielen, Daten über Einzelpersonen aus einer zunehmenden Anzahl von Quellen zusammenführen. Dazu gehören z.B. Abrechnungsdaten aus der gesetzlichen Krankenversicherung, Leistungsdaten aus anderen Bereichen der Sozialversicherung, Daten aus dem Melderegister sowie aus speziellen Registern (z.B. Krebsregister), Todeszeitpunkte und -ursachen, medizinische Befunddaten, Daten aus Biobanken, Daten zum Einkaufs- und Fahrverhalten, Daten aus stationären und mobilen Geräten über Abläufe des alltäglichen Lebens, sowie aus (fortlaufenden) Befragungen. Mit zwei Worten: Big Data. Da die – auf jeweils gleiche Personenidentitäten – notwendige Zusammenführung der Daten aus so unterschiedlichen Quellen, und womöglich über eine längere Zeitreihe, für die Zwecke der Versorgungsforschung nicht durch eine einzige Stelle geleistet werden kann, ist der Aufbau einer Forschungsinfrastruktur notwendig. Und die braucht eine Rechtsgrundlage. Datenhalter müssen verpflichtet werden, ihre Daten auf Anfrage in einer bestimmten Weise in einen „Pool" zu geben. Die Koordinatoren dieses Prozesses benötigen Berechtigungen und müssen bestimmten Auflagen an den Datenschutz und die Datensicherheit genügen.

Wenn es gelingt, auf diesen beiden Baustellen in den kommenden zwei bis drei Jahren substantielle Fortschritte zur Errichtung einer Forschungsinfrastruktur in der Versorgungsforschung zu erzielen, dann bleibt der Forschungsstandort Deutschland weiterhin weltweit in der Spitzengruppe. Es ist völlig klar, dass hiermit eine sehr große Herausforderung verbunden ist. Aber: sie ist grundsätzlich leistbar, wie uns andere Länder beweisen.

Wieso eigentlich Versorgungsforschung?

Bevor wir uns internationalen Rollenmodellen für einen möglichen Forschungs-Entwicklungsplan in Deutschland widmen, noch einmal eine Reflexion. Warum erhebt die Versorgungsforschung im Grundsatz einen so breiten Anspruch auf Teilhabe am weltweit rapide wachsenden Datenschatz?

Versorgungsforschung überprüft, wie die medizinische Versorgung unter Alltagsbedingungen funktioniert und wirkt. Sie ergänzt damit unverzichtbar die klinische Forschung. Deren Ziel ist es, den Nachweis über Wirksamkeit und Sicherheit von Behandlungsmethoden zu führen. Dies aber ist nur unter kontrollierten Laborbedingungen möglich. Wäh-

rend die klinische Forschung zwei oder mehrere Behandlungsverfahren vergleicht, betrachtet die Versorgungsforschung Populationen, z.B. alle Versicherten mit Wohnort in bestimmten Regionen.

Versorgungsforschung untersucht die Anwendung und die im Alltag erreichten Ergebnisse von Behandlungsmethoden. Sie stellt die Frage „Was kommt tatsächlich bei den Patienten an?“ und ermittelt Einflussfaktoren, die unterschiedliche Inanspruchnahmemuster, Behandlungsverläufe oder Ergebnisse (Outcomes) erklären. Damit liefert sie wichtige Informationen zur Verbesserung der Organisation, der Prozesse und der Ergebnisse der medizinischen Versorgung. Sie ergänzt somit die „Laborergebnisse“ der klinischen Forschung.

Uneinheitliche Versorgung trotz einheitlicher Rahmenbedingungen

Es wäre grundfalsch anzunehmen, dass eine neue wissenschaftliche Erkenntnis im Bereich der klinischen Forschung oder ein neues Gesetz zur Regulierung des Gesundheitswesens einfach aufgrund ihres jeweiligen Allgemeingültigkeitsanspruchs einheitliche Ergebnisse in der realen Welt hervorrufen.

Versorgungsforschung deckt Lücken zwischen Anspruch und Wirklichkeit auf. So basieren nahezu alle Gesundheitssysteme der Industrieländer auf dem Leitbild, dass jeder Bürger, zumindest aber jeder Sozialversicherte, Anspruch auf eine medizinische Versorgung nach dem aktuellen Stand der medizinischen Erkenntnisse hat. Nahezu alle Versorgungs- und Versicherungssysteme zielen darauf, allen Bürgern in dieser Hinsicht verlässliche Rahmenbedingungen zu bieten.

Die Versorgungsforschung zeigt, dass trotzdem in allen Industrieländern erhebliche regionale und kleinräumige Versorgungsunterschiede bestehen. Die Gründe liegen teils in der Bevölkerungs- oder Versichertenstruktur und in allgemeinen, administrativ- ökonomischen Rahmenbedingungen, teils in der Organisation der Versorgungssysteme. Auch Abläufe und Kommunikation zwischen den an der Versorgung beteiligten Einrichtungen und das Verhalten Einzelner spielen eine Rolle. In großen Datenmengen lassen sich bestimmte Muster erkennen, die auf Ursachen regionaler Variationen schließen lassen, wie etwa verschiedene Gesundheits- und Versorgungsatlanten weltweit demonstrieren.

Abseits der standardisierten Laborbedingungen der klinischen Forschung ist es entscheidend, Relevanz und Wirkungsrichtung der verschiedenen möglichen Einflüsse zu kennen, um letztlich wirksame Maßnahmen zur Verbesserung der Ergebnisqualität in der medizinischen Versorgung bestimmen zu können. Dies gilt umso mehr, je komplexer die medizinische Versorgung wird. Welchen Effekt haben etwa die komplexen Therapieregime in der Onkologie, die in Abhängigkeit gendiagnostischer Befunde oft im Verlauf einer Krankheitsepisode mehrfach gewechselt werden müssen? Welche Bedeutung haben etwa die Versorgungsstrukturen bei Apotheken, Ärzten und Krankenhäusern, Bildungsstand, Wohnverhältnisse und andere Lebensumstände sowie die Compliance der Patienten im Verhältnis dazu? Inwieweit sind erwünschte und unerwünschte Ergebnisse abhängig von lokalen Rahmenbedingungen oder vom individuellen Verhalten der Behandler und ihrer Patienten?

Ohne Daten keine Forschung

Während klinische Forschung für jede Studie eigene Daten erhebt, benötigt Versorgungsforschung zwingend Zugang zu bestehenden „Alltagsdaten“; es ist viel zu aufwändig, für jede Fragestellung erneut flächendeckende Daten zur Versorgungslage von Bevölkerungsgruppen zu erheben.

Administrative Daten:

Üblicher Weise nutzt Versorgungsforschung vor allem sogenannte Routine- oder Sekundärdaten, also Daten, die primär für administrative Zwecke erhoben werden. Hierzu gehören z.B. Abrechnungsdaten von Ärzten, Krankenhäusern und Apotheken aus der gesetzlichen Krankenversicherung, aber auch Angaben zum Versicherten und zum Versicherungsverhältnis, Leistungsdaten aus anderen Sozialversicherungszweigen, Daten von Einwohnermeldeämtern. Der Vorteil dieser Daten ist, dass sie den jeweiligen Ausschnitt der Bevölkerung nahezu vollzählig erfassen und in der Regel populationsbezogen vollständig vorliegen. Nachteile bestehen darin, dass

- die Daten bei vielen Institutionen verstreut verwaltet werden und für Forschungszwecke nicht zugänglich sind. Zudem können sie nicht oder nur mit unverhältnismäßig hohem Aufwand mit anderen Daten zusammengeführt werden.
- die erhobenen Datenmerkmale nur den administrativen Zwecken genügen müssen. Unter Gesichtspunkten der Forschung bleiben hierdurch viele Fragen offen, die nur durch zusätzliche Daten beantwortet werden können.

Klinische Daten:

Bei jeder ambulanten oder stationären Behandlung entsteht eine individuelle Dokumentation der Anamnese, der diagnostischen Befunde, der Diagnosen und Behandlungsmaßnahmen, der Patientenpräferenzen sowie der Kontrollbefunde. Anhand dieser Daten ließen sich viele Fragen spezifischer als etwa auf Basis von Abrechnungsdaten beantworten. Diese Daten sind mangels strukturierter elektronischer Patientenakten bislang zumeist nicht elektronisch auswertbar. Auch sind die Inhalte der Dokumentation nicht abgestimmt oder vereinheitlicht. Zudem muss in jedem Einzelfall eine Einwilligung der Beteiligten zur Nutzung der Daten vorliegen. Wie u.a. im Rahmen der Nationalen Kohorte für eine vergleichsweise begrenzte Zahl von Probanden gezeigt, wäre auf dieser Basis eine Zusammenführung mit administrativen Routinedaten möglich. Sie scheitert außerhalb besonders geförderter Projekte aber meist am Aufwand.

Primärdaten:

In besonderen Fällen wird versucht, diese Lücken durch die zusätzliche Erhebung von Primärdaten zu schließen, etwa durch Befragungen von Versicherten, Ärzten oder Einrichtungen. Dies gelingt in der Regel nur im Rahmen von geförderten Studien. Danach stehen die Einzeldaten für weitere Analysen meist nicht mehr zur Verfügung und können auch nicht mit anderen Datenbeständen zusammengeführt werden. Einen Mittelweg zwischen klinischen Daten und Primärdaten eröffnen Register, die bereits erhobene klinische Daten für Zwecke der Versorgungsforschung in einem eigenen Datenbestand zusammenfassen. Typischerweise konzentrieren sich Register auf eine bestimmte Erkrankung (z.B. Krebserkrankungen, Herzinfarkt) oder Maßnahme (z.B. die Implantation einer Gelenkendoprothese).

Surveillance- bzw. Melderegisterdaten:

Diese Daten werden vorzugsweise auf gesetzlicher Grundlage von staatlichen oder staatlich beauftragten Institutionen erhoben und dienen der Überwachung bestimmter gesundheitlicher Gefahren für die Bevölkerung, z.B. durch Infektionskrankheiten. Formal besitzen sie Charakteristika sowohl von Primär- als auch Sekundärdaten. Teils stehen die erhobenen Daten im Rahmen der Gesundheitsberichterstattung in aggregierter Form und eingeschränktem Umfang zur Verfügung. Eine Verknüpfung mit anderen Datenquellen ist nicht möglich.

Unstrukturierte Alltagsdaten:

In rasender Geschwindigkeit entstehen parallel zu den bisher genannten Datenquellen Datenbestände, deren Namensgebung noch im Fluss ist. Die Nutzung des Internets, von sozialen Netzwerken, von Apps sowie von mobilen Endgeräten, die zunehmend mit Sensorik ausgestattet sind, schafft eine Fülle von Daten über Interessen, Meinungen und Verhaltensweisen von Individuen. Perspektivisch wird dieser Datenbestand durch das „Internet der Dinge" gigantisch gesteigert werden. Diese Informationen gehören derzeit kommerziellen Anbietern und werden für kommerzielle Zwecke gehandelt und genutzt. Insbesondere durch die Zusammenführung der Informationen aus verschiedenen Quellen, lassen sich mit diesen Daten Profile erstellen, die viel über Lebensstil, Risikoneigung, gesundheitsfördernde oder –schädliche Verhaltensweisen sowie Entstehung und Umgang mit Krankheiten aussagen können. Dennoch stehen diese Daten der Wissenschaft, insbesondere der Versorgungsforschung, bisher nicht zur Verfügung.

Bei „Big Data" muss Deutschland vom Zuschauer zum Akteur werden

Als „Big Data" werden große personenbezogene oder -beziehbare, stark differenzierte Datenmengen bezeichnet, die mit Softwareunterstützung schnell auswertbar gemacht werden. Für manche scheint dies gleichbedeutend mit der Annahme, dass allein durch die Fülle an Informationen handlungsrelevante Muster erkennbar werden, manche haben schon das Ende theoriegeleiteter Analysen vorhergesagt. Diese Erwartung wird sicherlich enttäuscht werden. Vielfach ist auch die Qualität der Daten schlicht nicht ausreichend, um weitreichende Schlussfolgerungen zu ziehen. In diese Kategorie fallen viele der über mobile Geräte oder das Internet generierten Daten. So musste etwa Google seine Erwartung korrigieren, aufgrund von Häufigkeit und geografischer Entwicklung der Nutzung bestimmter Suchbegriffe im Internet z.B. Epidemien vorhersagen zu können. Gleichwohl müssen angesichts des rapide steigenden Datenvolumens rechtzeitig Weichen gestellt werden, um der Wissenschaft den Zugang zu potenziell relevanten Daten zu sichern. Ob und inwieweit Daten relevant sind, kann aber oftmals erst entschieden werden, wenn mit ihnen an konkreten Forschungsfragen gearbeitet wird. Derzeit scheint die Entwicklung von „Big Data" an Deutschland weitestgehend vorbei zu gehen. Da die meisten Anbieter derjenigen Hard- und Software, die die entsprechenden Daten ihrer Nutzer generieren, aus dem Ausland kommen, fließen auch die unstrukturierten Alltagsdaten dorthin zurück. Zudem gelten viele der im Inland auch von deutschen Unternehmen erhobenen Daten als „proprietär", also als Geschäftsgeheimnis.

Warum aber sollten die etwa in einem modernen Fahrzeug der Oberklasse bei jeder Fahrt generierten Daten allein dem Hersteller des Fahrzeugs gehören? Wem sollen künftig etwa die Daten der intelligenten Stromzähler in Häusern und Wohnungen gehören, mit denen das Alltagsverhalten der Bewohner abgebildet werden kann? Die Kenntnis, was und wie gut mit den Daten ausgesagt werden kann, darf aus Sicht einer effizienten Versorgungsforschung keinesfalls allein der internen Bewertung des Geräteanbieters oder des Stromkonzerns überlassen bleiben, denn dann bleiben der Öffentlichkeit auch die Gefahren verborgen, die potenziell mit einem solchen privaten Informationsvorsprung verbunden sind.

Ein zukunftsfähiges Forschungsdatennetz in Deutschland kann dem entgegenwirken. Eine leistungsfähige Versorgungsforschung benötigt ein Forschungsdatennetz, in dem Daten aus unterschiedlichen Quellen nach dem Prinzip von „Big Data" permanent zusammengeführt und für bestimmte wissenschaftliche Fragestellungen verantwortungsvoll ausgewertet werden können. Je umfassender der Datenbestand, desto größer der Nutzen für die Versorgungsforschung. Mit Umfang und Nutzbarkeit der Daten steigt auch das Interesse potenzieller Kooperationspartner. Ein entsprechender „Datenkern" wird, wie Beispiele aus Kalifornien zeigen, weitere „Datenspenden" aus Universitäten, von Unternehmen und von Einzelpersonen erzeugen, wenn sich hierdurch weitere Erkenntnisse ergeben.

Grundzüge eines Forschungsdatennetzes für die Versorgungsforschung in Deutschland

Von internationalen Beispielen lernen

In den skandinavischen Ländern existieren seit Jahrzehnten amtliche Register für Routine- und Registerdaten; alle Einträge können jeweils eindeutig einem Bürger und z.B. dem behandelnden Arzt zugeordnet werden. Auf dieser Basis können Analysen über langfristige Behandlungsverläufe durchgeführt werden. Rückschlüsse darauf, welche Behandlungsmaßnahmen Qualität und Wirtschaftlichkeit fördern, sind dadurch möglich.

Ein anderes Beispiel liefert das australische Population Health Research Network (PHRN). Dies ist ein Zusammenschluss aller staatlichen Stellen auf regionaler und föderaler Ebene, die im australischen Gesundheitssystem über Routinedaten verfügen. Dank pseudonymisierter Datenzusammenführung (data linkage) können Wissenschaftler diese Daten nachträglich zur Erforschung von Einflüssen auf die gesundheitliche Entwicklung nutzen. So gewinnen sie zahlreiche Hinweise auf Verbesserungsmöglichkeiten.

Das Beispiel Australiens scheint auf die Ausgangslage in Deutschland weitaus einfacher anwendbar als skandinavische Vorbilder. Eine Lehre aus diversen Projekten der Zusammenführung von Daten in Deutschland ist, dass das dezentral verteilte Wissen bei den Datenherren benötigt wird, um die Qualität der Daten und die Aussagefähigkeit zusammengeführter Daten zu sichern. Auch dies spricht eher für das australische Modell, in dem dieses Wissen seitens der Koordinatoren des PHRN und durch die Struktur des PHRN explizit eingebunden wird. Um das Wissen zu erhalten, müssen die Teildatenbestände auch weiterhin von den Quellen genutzt und gepflegt werden. Die Wissenschaft erhält somit gewissermaßen eine Kopie verfügbarer Daten.

Datentrias der Versorgungsforschung

Etwa 95% aller Deutschen gehen mindestens einmal jährlich zum Arzt. Routinedaten aus der Sozialversicherung erfassen die wichtigsten Ereignisse in der alltäglichen medizinischen Versorgung nahezu vollständig. Angaben aus Melderegistern könnten zudem weitere Informationslücken schließen. Eine populationsvollständige Routinedatengrundlage bildet das Rückgrat der Versorgungsforschung.

Die Routinedaten habe zwei Lücken: Zum einen sind sie oftmals nicht spezifisch genug. Zum anderen fehlen Informationen über das Geschehen zwischen Ereignissen, z.B. zwischen dem Besuch einer Arztpraxis und einem Krankenhausaufenthalt, die administrativ erfasst werden. Diese Lücken können durch eine geeignete Forschungsdateninfrastruktur schrittweise geschlossen werden.

Klinische Daten sind sehr gut dokumentiert. Sie liefern wesentlich spezifischere Informationen und folgen meist einer durch das Krankheits- und Behandlungsgeschehen bestimmten Struktur.

Die Daten aus dem Internet, aus Apps und von Sensoren der Endgeräte liefern schließlich wichtige Informationen über verlaufsbestimmende Ereignisse, Meinungen, Lebensumstände, Präferenzen und Verhaltensweisen der Patienten. Sie können ggf. durch Primärdaten aus Befragungen und Register ergänzt werden.



Abb. 1: Datentrias der Versorgungsforschung

Versorgungsforschung braucht eine ständige, leistungsfähige Infrastruktur

Menge und Fülle der Daten schließen aus, die Datengrundlage für jede Fragestellung neu zu legen. Insofern müsste zunächst die langjährige Datenvorhaltung bei den Quellen in einer Weise gesichert werden, die auch später noch eine personenbezogene Zusammenführung von einzelnen Datensätzen mit denen anderer Quelldatenbanken erlaubt.

Auf Basis einer Rechtsgrundlage, die den dauerhaften Aufbau von Datenbanken zur Versorgungsforschung ermöglicht, sowie mittels einer Koordinationsstelle, die einen funktionsfähigen Datenzugang für die wissenschaftliche Nutzung gewährleisten kann, wäre zu erwarten, dass dann an mehr als einer Stelle zusammengeführte Datenbanken entstehen, deren Bestände von anderen Forschern mitgenutzt werden können. Anstelle einer einzigen Datentransparenzstelle etwa beim DIMDI würden bereits voraggregierte und gepflegte Datenbanken (bezogen auf den heutigen Stand in Deutschland z.B. die AOK Versichertenstichprobe Hessen und die Daten von vier Krankenkassen an der Universität Bremen) für eine Fragestellung gemeinsam analysiert und ggf. mit weiteren ergänzenden Daten (etwa aus Krebsregistern, dem Gesundheitssurvey DEGS und Versichertenbefragungen) kombiniert werden können.

Es ist davon auszugehen, dass hierdurch bestimmte Datenbanken als „Bausteine" für Studien der Versorgungsforschung entstehen, die zunächst im Zuge einer Studie entwickelt werden und dann von Dritten wiederum als eine von mehreren Quellen weitergenutzt werden können.

Dieses Modell hätte den Charme, dass extreme Engpasssituationen vermieden werden wie sie derzeit beim DIMDI oder in Forschungsdatenzentren des Statistischen Bundesamts mit monate- und jahrelangen Verzögerungen von Auswertungsergebnissen bestehen.

Zuerst Routinedaten zusammenführen

Grundlage einer Forschungsinfrastruktur für die Versorgungsforschung in Deutschland sollte daher ein Data Linkage bestehender Routinedaten sein. Deren Kern wiederum sollten die Abrechnungsdaten in der gesetzlichen Krankenversicherung sowie der privaten Krankenversicherung sein. Diese sind heute auf viele datenhaltende Stellen verteilt und somit faktisch nicht in zusammengeführter Form auswertbar. Die Datentransparenzverordnung (DaTraV) hat auf Basis der Daten des Risikostrukturausgleichs der gesetzlichen Krankenkassen einen Anfang gemacht. Die Daten können bezogen auf pseudonymisierte Krankenversicherungsnummern ausgewertet werden. Der Datenbestand ist jedoch unvollständig, zum Zeitpunkt der Auswertung veraltet und das Verfahren der Auswertung ungeeignet, um von einer Vielzahl von Wissenschaftlern mit kurzen Bearbeitungszeiten genutzt werden zu können.

Zudem sind weitere Routinedaten notwendig. Daten aus anderen Zweigen der Sozialversicherung, wie z.B. das Vorliegen von Pflegebedürftigkeit, die Inanspruchnahme von Rehabilitationsmaßnahmen, ein Renteneintritt, das Vorliegen von Arbeitslosigkeit etc. sind wichtige Informationen, die in den Daten der gesetzlichen Krankenversicherung fehlen. Dort finden sich auch wichtige epidemiologische Rahmendaten, etwa zur Geburt, zum Wohnort, zum Bildungsstand sowie zu Todeszeitpunkt und -ursache, die nicht extra erhoben werden müssten.

Hierfür kann Deutschland sich am Erfolgsmodell des australischen PHRN orientieren. Dies geht noch über die im aktuellen Aktionsplan Versorgungsforschung des BMBF bereits vorgesehenen forschungsbezogenen Datenregister hinaus. Notwendig ist eine gesetzliche Grundlage, mit welcher der Gesetzgeber einen in der Forschung verwurzelten Akteur oder ein Konsortium mit der Koordinationsaufgabe des Data Linkage bestehender Datenquellen beleiht.

Die Zukunft von Anfang an mitdenken

Auch ein deutsches Forschungsdatennetz in der Versorgungsforschung aus methodischen Gründen müsste bei den Routinedaten der Sozialversicherung und staatlicher Stellen seinen Ursprung nehmen. Der Gesetzgeber dürfte hier aber nicht Halt machen. Die technologische Entwicklung schafft einen die Routinedaten schnell übersteigenden Bestand an weitgehend unstrukturierten Alltagsdaten von zunehmender epidemiologischer Relevanz. Hier sollte nach dem gleichen Grundsatz verfahren werden wie bei den Routinedaten: Daten, die für administrative oder kommerzielle Gründe „sowieso" erhoben werden, werden der Versorgungsforschung zugänglich gemacht. Dies würde bedeuten, dass Deutschland von Unternehmen, die im Internet, per App oder Sensor Daten erheben, eine Kopie der Daten zur wissenschaftlichen Nutzung einfordert. Um für Fragestellungen der Versorgungsforschung genutzt werden zu können, sollten auch diese Daten, soweit technisch möglich, dem Nutzer in pseudonymisierter Form zugeordnet werden.

Literatur

1. Garbe E, Pigeot I. (2015) Der Nutzen großer Gesundheitsdatenbanken für die Arzneimittelrisikoforschung. Bundesgesundheitsbl. 58: 829. doi:10.1007/s00103-015-2185-7
2. Jacobs S. Stallmann C, Pigeot I. (2015) Verknüpfung großer Sekundär- und Registerdatenquellen mit Daten aus Kohortenstudien Bundesgesundheitsbl. 58: 822. doi:10.1007/s00103-015-2184-8
3. Swart, E., Stallmann, C., Powietzka, J. et al. (2014) Datenlinkage von Primär- und Sekundärdaten Ein Zugewinn auch für die kleinräumige Versorgungsforschung in Deutschland? Bundesgesundheitsbl. 57: 180. doi:10.1007/s00103-013-1892-1
4. Swart, E., Thomas, D., March, S. et al. (2011) Die Eignung von Sekundärdaten zur Evaluation eines Interventionsprojekts. Praev Gesundheitsf 6: 305. doi:10.1007/s11553-011-0309-y
5. Aspinall EJ, Hutchinson SJ, Janjua NZ, et al. (2015) Trends in mortality after diagnosis of hepatitis C virus infection: an international comparison and implications for monitoring the population impact of treatment. J Hepatol.62(2):269-77. PubMed PMID: 25200903. Epub 2014/09/10. eng.
6. Falster MO, Jorm LR, Douglas KA, Blyth FM, Elliott RF, Leyland AH. (2015) Sociodemographic and health characteristics, rather than primary care supply, are major drivers of geographic variation in preventable hospitalizations in Australia. Med Care. 53(5):436-45. PubMed PMID: 25793270. Pubmed Central PMCID: Pmc4396734. Epub 2015/03/21. eng.
7. Holman CD, Bass AJ, Rouse IL, Hobbs MS. (1999) Population-based linkage of health records in Western Australia: development of a health services research linked database. ANZJPH 23 (5): 453–459
8. http://www.phrn.org.au/

Datenspende ermöglichen

Schafft der Gesetzgeber die Voraussetzungen für eine Infrastruktur, die eine regelhafte Auswertung zusammengeführter administrativer Routinedaten und unstrukturierter Alltagsdaten ermöglicht, sind wesentliche Elemente eines zukunftsfähigen Forschungsdatennetzes vorhanden, welches durch Datenspenden aus dem Bereich der medizinischen Befunddaten ergänzt werden könnte.

Weiterer wichtiger Bestandteil eines zukunftsfähigen Forschungsdatennetzes müssen deshalb flexible Regeln zur Ermöglichung von Datenspenden sein. Schon heute ist es Einzelpersonen möglich, der Nutzung ihrer klinischen Daten für Forschungszwecke zuzustimmen. Aufgrund der besonderen Sensibilität der klinischen Daten dürfte dieses Einwilligungsmodell auch weiterhin Grundlage für die Einbeziehung klinischer Daten in das Forschungsdatennetz bleiben.

Heute existieren jedoch vielfache rechtliche Hürden, um klinische Daten z.B. mit administrativen Routinedaten der Sozialversicherung zusammenzuführen, selbst wenn die Einwilligung der beteiligten Patienten und Ärzte vorliegt.

In Zeiten einer zunehmenden Subspezialisierung der Medizin, der Entwicklung sogenannter personalisierter Therapien sowie einer rasanten medizinischen und medizintechnischen Entwicklung, die ihrerseits wie etwa in der Humangenetik enorme Datenmengen generiert, bilden klinische Daten den dritten unverzichtbaren Pfeiler des deutschen Forschungsdatennetzes.

Aufgrund der besonderen Präzision und des steigenden Informationsgehalts klinischer Daten müssen diese auf der Grundlage des Prinzips der Datenspende, nämlich durch Einwilligung der Patienten, Bestandteil des Datenkörpers werden können. Im Rahmen der Einwilligung können hierbei Art und Umfang der Auswertungsmöglichkeiten und besondere Anforderungen an den Datenschutz im Bedarfsfall individuell geregelt werden.

Datenschutzregeln an veränderte Bedingungen adaptieren

Internationale Erfahrungen belegen, dass der Umgang mit solchen komplexen Datenbanken mit wirksamen Datenschutzanforderungen verbunden werden kann. Für jeden wissenschaftlichen Nutzer einer Datenbank im Netz könnte gelten, dass der Zugang nur über einen jeweils zentralen Datenzugang erfolgt. Teile des Datenkörpers werden nicht an einzelne Forscher abgegeben. Somit können sämtliche Datenzugriffe überprüft und etwa nach dem dänischen Modell sanktioniert werden. Dort erfolgt der Datenzugang für einzelne Forscher immer über eine Institution, z.B. die Universität. Verletzt ein Forscher den Datenschutz, wird der Datenzugang für die gesamte Institution zunächst befristet, im Wiederholungsfall vollständig gesperrt. Zusammen mit technischen Maßnahmen des Datenschutzes, wie etwa geeigneter Pseudonymisierung der Namensbezüge, können entsprechende Sanktionsmaßnahmen die Einhaltung der Datenschutzanforderungen wirksam unterstützen.

Autorenerklärung

Dieser Beitrag entstand im Zusammenhang der Arbeit von Dr. Dominik Graf von Stillfried als Geschäftsführer des Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland (Zi). Der Autor dankt Frau Prof. Iris Pigeot, Bremen, sowie Mitarbeitern des Zentralinstituts für wertvolle Kommentare.

Making big data work for health sciences in Germany – What needs to be done now

In spite of uniform regulation health care is subject to geographic variation within each health system. In order to know what is actually happening to the patients, and why, we need health services research. The answers can be drawn from various sources of data, administrative, clinical, and personal. Relevant data is created by an ever increasing degree of digitalization of day-to-day-life. Germany, however, has not yet made proper use of its data sources for health services research. In order to stay abreast with other countries Germany now needs to embark on a program to create a lasting data infrastructure for scientific research. There needs to be a legal basis to allow for long time series of personally identifiable data as well as data linkage across the various data bases. In doing so, Germany can refer to prior experiences in other countries such as in the Australian Population Health Research Network.

Keywords

digitalization of all aspect of life, big data, health services research, life sciences, research infrastructure, international competition

Jetzt beginnen und den Anschluss nicht weiter verlieren

Die Errichtung eines deutschen Forschungsdatennetzes für die Versorgungsforschung ist ein maßgebliches Infrastrukturprojekt. Auch unter Nutzung internationaler Vorerfahrungen hat ein solches Projekt eine Entwicklungszeit von etwa zehn Jahren – von der Beratung im Bundestag bis zur praktischen Nutzung für Wissenschaftler. Wer will, dass die Versorgungsforschung im Jahr 2030 maßgebliche Beiträge zur Verbesserung des deutschen Gesundheitswesens und der medizinischen Versorgung in anderen Industrieländern leistet, muss heute dafür die Voraussetzungen schaffen. Zum Einstieg sollte bereits das Zeitfenster zur Implementierung der europäischen Datenschutzrichtlinie genutzt werden. Zur Vorbereitung eines beratungsfähigen Gesetzentwurfs wäre deshalb ein Schritt noch in dieser Legislaturperiode erforderlich, etwa die Einrichtung einer gemeinsamen Kommission von Experten, die auf Basis internationaler Erfahrungen bis Anfang 2018 einen technischen und organisatorischen Verfahrensvorschlag inklusive ggf. erforderlicher Rechtsänderungen entwickelt. Wer könnte diesen Schritt wirksam vornehmen? Hier wäre vermutlich das Kanzleramt am Zuge. <<

Dr. rer. pol. Dominik Graf von Stillfried
ist Geschäftsführer des Zentralinstituts für die kassenärztliche Versorgung in der Bundesrepublik Deutschland (Zi) in Berlin. Von 2006 bis 2008 war er Leiter der Stabsstelle Versorgungsforschung bei der Kassenärztlichen Bundesvereinigung, von 1998 bis 2006 Leiter des Dezernates Grundsatzfragen bei der Kassenärztlichen Bundesvereinigung und von 1996 bis 1998 leitete er das Sachgebiet Managed Care bei der BARMER Ersatzkasse.
Kontakt: DStillfried@zi.de

