

Beiträge und Analysen

Gesundheitswesen

aktuell 2022

herausgegeben von Uwe Repschläger,
Claudia Schulte und Nicole Osterkamp

Jonas Schreyögg

Warum wir einen verbesserten Zugang zu Forschungsdaten im Gesundheitswesen benötigen,
Seite 14–31

doi: 10.30433/GWA2022-14

Jonas Schreyögg

Warum wir einen verbesserten Zugang zu

Forschungsdaten im Gesundheitswesen benötigen

Beim öffentlichen Diskurs zu den Themen Digitalisierung und Big Data im Gesundheitswesen in Deutschland tritt eine starke Ambivalenz zutage. Einerseits existieren hohe Erwartungen an Verfahren der Künstlichen Intelligenz (KI) unter Nutzung dieser Daten, andererseits sehen viele Menschen große Risiken im Aufbau und der Nutzung großer Datenbestände. Über die Risiken wird deutlich mehr geschrieben als über den Nutzen. Der Beitrag thematisiert die aktuelle Situation in Deutschland beim Zugang zu Forschungsdaten im Gesundheitswesen, zeigt den Mehrwert einer verbesserten Nutzung von Forschungsdaten auf, geht auf die Ausgestaltung des Zugangs zu Forschungsdaten in ausgewählten anderen Ländern (insbesondere Finnland und Frankreich) als Best-Practice-Beispiele ein und spricht Empfehlungen aus.

Ausgangslage: Probleme bei Zugang und Nutzung relevanter Datensätze des deutschen Gesundheitswesens

Es existieren große Datenmengen im deutschen Gesundheitssystem aus generierten Systemdaten. Diese sind jedoch entweder nur in Ausschnitten zugänglich (beispielsweise Krankenkassendaten), durch restriktive Nutzungsbedingungen kaum für Forschungsprojekte nutzbar (beispielsweise stationäre Abrechnungsdaten gemäß § 21 KHEntgG beim Statistischen Bundesamt) oder sie sind gar nicht für Forschungszwecke zugänglich (beispielsweise Daten der Kalkulationskrankenhäuser beim Institut für das Entgeltssystem im Krankenhaus – InEK). Administrativ besonders aufwendig ist eine Verlinkung verschiedener Datensätze, die aber immer wichtiger wird – nicht zuletzt zur Anwendung neuer Methoden wie Machine Learning.

Insgesamt hinkt Deutschland anderen europäischen Staaten in Bezug auf den Aufbau und den Zugang zu Forschungsdaten deutlich hinterher (OECD 2015; Deutschland ist in diesem Bericht nicht inkludiert. Dennoch wird deutlich, dass andere Staaten trotz ähnlicher Datenschutzgesetze eine liberale Auslegung praktizieren). Gerade im Kontext von Forschungsprojekten des EU-Forschungsrahmenprogramms (Horizon Europe) wird besonders deutlich, dass andere Länder, beispielsweise England und Finnland,

wesentlich bessere Datenzugänge bereitstellen. Deutschland wird diesbezüglich zunehmend auch von osteuropäischen Ländern überholt. Ein guter Forschungsdatenzugang im Gesundheitswesen ist jedoch essenziell für die Innovationsfähigkeit des Wissenschaftsstandorts Deutschland, aber auch für die evidenzbasierte Weiterentwicklung des Gesundheitssystems.

Im deutschen Gesundheitswesen existieren große Datenmengen. Dies betrifft vor allem Struktur- und Leistungsdaten aus allen Sektoren, die bei verschiedenen Trägern und Dienstleistern in unterschiedlichen Konfigurationen vorliegen. Die Qualität der Struktur- und Leistungsdaten, beispielsweise die Kodierqualität im stationären Sektor, ist im Vergleich mit anderen Industriestaaten als sehr gut einzustufen, was regelmäßig bei komparativen EU-Projekten auffällt. Kaum ein Land weist eine so hervorragende Datenbasis auf. Allerdings liegen diese Struktur- und Leistungsdaten bisher nicht so zusammengeführt mit allen relevanten Informationen vor, dass sie für die überwiegende Zahl an Forschungsfragen nutzbar wären. Dies ist auch auf die Struktur des deutschen Gesundheitswesens zurückzuführen. Während beispielsweise in England die Daten aller Bürger im National Health Service (NHS) gebündelt und über NHS Digital für Forschungszwecke bereitgestellt werden, muss sich die Forschung in Deutschland oftmals auf die Daten einzelner Krankenkassen und damit auf einen Ausschnitt der Versorgung stützen. Außerdem bestehen in Deutschland für die Nutzung dieser Daten für wissenschaftliche Zwecke große administrative Hürden, die bei genauerer Betrachtung nicht auf Datenschutzgesetze selbst, sondern deren Interpretation und Umsetzung zurückzuführen sind und eine adäquate wissenschaftliche Nutzung de facto verhindern (Schreyögg 2017 und 2020).

In den letzten Jahren wurden die stationären Abrechnungsdaten gemäß § 21 KHEntgG sowie die Krankenkassendaten prinzipiell als Individualdatensatz pseudonymisiert zur Verfügung gestellt. Die stationären Abrechnungsdaten gemäß § 21 KHEntgG stehen seit einigen Jahren als DRG-Statistik im Forschungsdatenzentrum (FDZ) Bund zur Verfügung. Diese Daten werden im Rahmen der DRG-Kalkulation durch das InEK generiert. Die Kostendaten der Krankenhäuser, die auch im Rahmen der Kalkulation generiert werden, sind für wissenschaftliche Zwecke bisher nicht geöffnet und werden auch

nicht an das Statistische Bundesamt übermittelt. Ambulante und stationäre Abrechnungsdaten (gemäß 295 SGB V, § 300 SGB V i.V.m. § 31 SGB V und § 301 SGB V) sowie Daten aus anderen Sektoren standen auch bisher über § 303a-e SGB V beim Deutschen Institut für medizinische Dokumentation und Information (DIMDI) für Forschungszwecke zur Verfügung. Diese Daten wurden durch das Bundesversicherungsamt (BVA) für den morbiditätsorientierten Risikostrukturausgleich (Morbi-RSA) erfasst und dann ans DIMDI übermittelt. Bei beiden Datensätzen stellen sich grundsätzliche Probleme im Umfang der bereitgestellten Informationen sowie dem erlaubten Prozedere der operativen Datennutzung.

Im Rahmen der DRG-Statistik werden im Prinzip alle Informationen zur Abrechnung der Krankenhäuser an das Statistische Bundesamt übermittelt. So werden jährlich etwa 19 Millionen Krankenhausfälle anhand von Leistungs- und Strukturdaten erfasst. Leistungsbezogene Daten enthalten Informationen zu jedem Behandlungsfall, unter anderem die Haupt- und Nebendiagnosen nach ICD-10, durchgeführte Operationen und Prozeduren nach dem Operationen- und Prozedurenschlüssel (OPS), Informationen zum Verlauf des Aufenthalts, demografische Daten sowie Art und Höhe der abgerechneten Entgelte. Auch krankenhausbezogene Informationen sind in dem Datensatz enthalten. Die Daten können entweder per kontrollierter Ferndatenverarbeitung, also durch die Einsendung von Analyseskripten ohne direkten Zugriff auf den Volldatensatz, oder an Gastwissenschaftlerarbeitsplätzen an einem der 18 Standorte der FDZ der statistischen Ämter ausgewertet werden.

Allerdings sind bereits bei der Datenaufbereitung bestimmte Grundvoraussetzungen nicht gegeben, um die Daten für einen Großteil der Fragestellungen sinnvoll nutzen zu können. Erstens darf der öffentlich verfügbare und offiziell vom InEK zertifizierte Grouper, der von den Krankenhäusern zur Eingabe von Fällen und zur Ermittlung der Fallpauschalen genutzt wird, nicht zum Datensatz hinzugespielt werden, was dazu führt, dass die Daten nicht über die Jahre vergleichbar sind. Dadurch scheiden viele Fragestellungen zu Vergütung und Vergütungsanreizen aus, die Impulse für die Weiterentwicklung der stationären Vergütung in Deutschland liefern könnten. Eine weitere große Hürde ist die Begrenzung auf wenige Variablen aus externen Quellen, beispielsweise aus den gesetzlichen Qualitätsberichten der Krankenhäuser, die hinzugefügt

werden können. Zudem müssen die Variablen aus öffentlichen Informationen stammen. Befragungen von Krankenhäusern zu institutionellen Gegebenheiten können nur unter sehr restriktiven Bedingungen zum Datensatz hinzugespielt werden.

Zuletzt scheidet eine Nutzung für viele Forschungsprojekte allein deshalb aus, weil eine Einsendung von Analyseskripten ohne direkten Zugriff auf den Volldatensatz den heutigen Anforderungen an komplexe statistische Analysen nicht gerecht wird. Der Zeitaufwand allein bei der Datenaufbereitung, die mit unzähligen Einsendungen von Analyseskripten verbunden ist, wird damit für die meisten Forschungsprojekte wesentlich zu hoch. Diese Praxis stammt noch aus den 1980er beziehungsweise 1990er Jahren, als Analysen weniger komplex waren. Für heutige multivariate Analysen mit möglichst kausalem Anspruch und algorithmische Verfahren, beispielsweise Machine-Learning-Verfahren, ist dies nicht praktikabel (Schreyögg 2017; Milstein und Schreyögg 2020; SVR 2021).

Im Rahmen der Krankenkassendaten nach § 303a ff. SGB V beim DIMDI fehlen bisher bestimmte Informationen, die nicht von den Krankenkassen an das BVA gemeldet werden. Dies sind zum einen Kennziffern für die stationären und ambulanten Leistungserbringer. Das bedeutet, man kann keine leistungserbringerbezogenen Fragestellungen mit diesen Daten bearbeiten. Beispielsweise können so grundlegende Fragen nach der Versorgungsqualität, wie Unterschiede in der poststationären Mortalität oder Wiedereinweisung von Patienten zwischen Krankenhausträgern oder Größenklassen, nicht beantwortet werden. Weiterhin fehlten Angaben zu OPS-Kodes, also zu durchgeführten Operationen und Prozeduren im stationären Bereich. Diese und weitere Defizite wurden durch das Digitale-Versorgungs-Gesetz (DVG) behoben, indem nun die Daten direkt von den Kassen abgefragt und zusammengeführt werden.

Die Nutzung der kontrollierten Ferndatenverarbeitung der nach § 303a ff. SGB V Daten soll durch das DVG und der damit verbundenen Rechtsverordnung überwunden werden. Das neu geschaffene Forschungsdatenzentrum Gesundheit beim Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) soll künftig den Nutzern die Arbeit auf dem Volldatensatz ermöglichen. Das neu geschaffene Forschungsdatenzentrum könnte damit, sofern es die erforderliche personelle Ausstattung erhält, wie in anderen Ländern

üblich, die Beratung und Bereitstellung der Datensätze übernehmen. Das Forschungszentrum Gesundheit könnte auch der zentrale Ort für künftige notwendige Erweiterungen beispielsweise um weitere Datensätze und Verlinkung von Datensätzen sein. Auch die Möglichkeit eines Zugriffs auf den Volldatensatz über eine gesicherte Verbindung vom eigenen Arbeitsplatz aus, wie er beispielsweise bei der BARMER bereits seit vielen Jahren für Forscher ermöglicht wird, wird durch das DVG vorgesehen. Insofern schaffen diese Regelungen – vorbehaltlich der Umsetzung – deutliche Verbesserungen in der Forschungsdatennutzung.

Anträge zur Nutzung dieses neuen Datensatzes sollen in der zweiten Jahreshälfte des Jahres 2022 gestellt werden können. Der Volldatensatz wird nur ab dem Datenjahr 2021 zur Verfügung stehen, was anfangs die Analysemöglichkeit auf Querschnittsanalysen beschränkt. Für eine kausale Analyse von zum Beispiel Politikinterventionen wird der Datensatz noch um viele Datenjahre anwachsen müssen. Gleichzeitig heißt dies, dass Forscherinnen und Forscher bis auf Weiteres für die Analyse von Interventionen auf andere Datensätze, etwa von einzelnen Krankenkassen, angewiesen sind.

Aktuell bleiben jedoch auch zahlreiche weitergehende Fragen ungelöst. Unklar bleibt unter anderem, ob die Krankenkassendaten zukünftig als Panel genutzt werden können, beispielsweise zum Vergleich der poststationären Mortalität von Krankenhausträgern über Jahre hinweg. Eine wesentliche Frage ist zudem, ob in dem neu zu schaffenden Forschungszentrum auch Registerdaten oder Daten aus der elektronischen Patientenakte (ePA) mit Krankenkassendaten verknüpft werden können. Bisher ist die Verknüpfung von Krankenkassendaten mit anderen Informationen zwar mit Einschränkungen auch schon möglich. Die Beantragung und Durchführung einer solchen Verknüpfung ist jedoch administrativ so aufwendig und langwierig, dass sie für viele Forschende eine abschreckende Wirkung hat (Schreyögg 2020).

Nutzung weiterer Datenquellen

Bei Nutzung von Registern und der Daten aus der ePA jeweils als einzelne Datenquelle oder in Kombination mit Struktur- und Leistungsdaten kommen weitere Probleme hinzu. Medizinische Register dienen als wichtige Datenquelle für eine patientenorientierte Versorgungsforschung unter Alltagsbedingungen. Das Deutsche Netzwerk

Versorgungsforschung (DNVF) definiert Register als „eine möglichst aktive, prospektive, standardisierte Dokumentation von Beobachtungseinheiten zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen, für die ein präziser Bezug zur Zielpopulation darstellbar ist“.

Register unterstützen die Beobachtung von Patientinnen und Patienten in definierten Leistungsbereichen über einen längeren Zeitraum. Zwar existieren in Deutschland verschiedenste krankheits-, produkt- oder prozedurenbezogene Register. Diese umfassen allerdings nicht selten nur einen kleinen Ausschnitt der relevanten Population (beispielsweise eine Region), und der Zugang ist oftmals unklar und intransparent. Nationale Register mit einer hohen Abdeckung von Patientenpopulationen, die für zahlreiche Fragestellungen wünschenswert wären, wie in England oder den nordischen Staaten, existieren hierzulande kaum (Schreyögg 2020). Bisher ist auch unklar, wie die neu einzurichtende ePA für Forschungszwecke genutzt werden kann. Die im Koalitionsvertrag vorgesehene Opt-Out-Lösung ist ein großer Fortschritt gegenüber der bisher geltenden Opt-In-Regelung, da so mit einer höheren Wahrscheinlichkeit der Vollständigkeit des ePA-Datensatzes gerechnet werden kann. Dennoch lässt der Koalitionsvertrag offen, ob die ePA-Daten künftig auch im Forschungsdatenzentrum Gesundheit bereitgestellt werden, das bedeutet, ob das Opt-Out-Prinzip auch für eine Nutzung der ePA-Daten für Forschungszwecke gelten soll.

Zu diesen bisher genannten Datenbeständen kommen die Datenbestände der Medizininformatikinitiative und des Netzwerks Universitätsmedizin (NUM), die für bestimmte Fragestellungen, auch in Kombination mit den oben genannten Datenbeständen, wichtige Informationen liefern. Erwähnenswert ist ebenfalls die Initiative zur Förderung der Nationalen Forschungsdateninfrastruktur. Das durch Bund und Länder gemeinsam geförderte Programm soll die Forschungsdatenbestände deutscher Wissenschaftseinrichtungen systematisch erschließen, verknüpfen, langfristig nutzbar und zugänglich sowie international anschlussfähig machen. Mit Blick auf Gesundheitsdaten wird insbesondere das Konsortium NFDI4Health gefördert, das die Sichtbarkeit und Zugänglichkeit von Forschungsdaten im Gesundheitswesen erhöhen möchte. Darüber hinaus existieren zahlreiche weitere Datenquellen im deutschen Gesundheitswesen, beispielsweise aus der gesetzlichen Qualitätssicherung, die für die Forschung relevant sind.

Problem: Die Verknüpfung unterschiedlicher Datenquellen zu Analysezwecken

Letztlich stellt sich bei allen Datenquellen die Frage, wie diese mit anderen Datenquellen künftig verknüpft werden können. Viele Datenquellen werden erst dann für die Forschung wertvoll und nutzbar, wenn sie mit anderen Datenquellen verknüpft werden. Ein gutes Beispiel hierfür ist die fehlende Verknüpfbarkeit der Daten aus der Impfsurveillance des Robert Koch-Instituts (RKI) und der Pharmakovigilanz des Paul-Ehrlich-Instituts (PEI) zu den COVID-19-Impfstoffen mit den Krankenkassenroutinedaten. Eine Verknüpfung dieser Datenquellen scheiterte an dem Fehlen eines Unique Identifiers (UI) und hätte viel differenziertere Aussagen zu Wirkungen und möglichen Nebenwirkungen von COVID-19-Impfstoffen in Deutschland ermöglicht (Schröder et al. 2021). Durch diese zusätzlichen Informationen hätte man potenziell auch unsichere Bevölkerungsgruppen von der Sicherheit der Impfstoffe überzeugen können.

Insgesamt ist der Status quo des Zugangs und der Nutzungsmöglichkeiten von Forschungsdaten im Gesundheitswesen – trotz deutlicher Verbesserungen im Rahmen des DVG – unbefriedigend. Es existiert ein deutlicher Handlungsbedarf, um zu anderen Staaten, unter anderem Großbritannien und den nordischen Staaten, aufzuschließen. Der Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (SVR), die Deutsche Forschungsgemeinschaft (DFG) sowie viele andere haben kürzlich auf diese erhebliche Problematik für den Forschungs- und Innovationsstandort Deutschland sowie die gesellschaftlichen Folgen hingewiesen (SVR 2021; DFG 2021).

Gesellschaftlicher Mehrwert eines verbesserten Zugangs zu Forschungsdaten

Auch wenn sich der Zugang in Deutschland zu Forschungsdaten in den letzten Jahren, auch durch Gesetzesinitiativen wie das DVG, verbessert hat und das Thema im Rahmen der Forschungsförderung, beispielsweise durch die Nationale Forschungsdateninfrastruktur (NFDI), mehr Aufmerksamkeit erhält, sind die Möglichkeiten der Datennutzung im Vergleich zu anderen europäischen Staaten nach wie vor stark eingeschränkt. Die Verbesserung des Forschungsdatenzugangs erhält jedoch immer noch nicht die Aufmerksamkeit in Politik und Medien, die notwendig wäre, um zu anderen Ländern in

Europa aufzuschließen. Als ein wichtiger Grund ist hier anzuführen, dass der Mehrwert eines verbesserten Zugangs zu Forschungsdaten im Gesundheitswesen vor allem auch in der Kombination von Krankenkassendaten, Daten der ePA und von Registerdaten liegt. An dieser Stelle stehen neben Politikern nicht zuletzt die Wissenschaftlerinnen und Wissenschaftler selbst in der Pflicht, diesen Mehrwert stärker zu veranschaulichen.

Es existieren zahlreiche Anwendungsfelder, die von einer Verbesserung des Zugangs zu Forschungsdaten profitieren würden. Einige dieser Anwendungsfelder sind in Tabelle 1 mit Umsetzungsbeispielen dargestellt. Oftmals ist der Nutzen relativ klar operationalisierbar, allerdings für den nicht involvierten Bürger beziehungsweise die Bürgerin nicht einfach zu erfassen. Ein gutes Beispiel für einen gesellschaftlichen Nutzen ist die Möglichkeit zur Evaluation von Politikinterventionen. In Deutschland ist es nach wie vor nicht üblich, die Wirkung von Gesetzesänderungen wissenschaftlich zu evaluieren. Zwar enthalten Gesetze manchmal sogar die Verpflichtung zur Evaluation von Reformen, unter anderem bei der DRG-Einführung. Die eigentliche Evaluation leidet aber dann regelmäßig darunter, dass die notwendigen Zugänge zu den Datenquellen nicht ermöglicht werden (können), sodass oftmals die relevanten Forschungsfragen gar nicht beantwortet werden können. Bei der DRG-Einführung ist beispielsweise bis heute die empirische Evidenz rar und zu manchen Fragestellungen, beispielsweise der Wirkung der DRG-Einführung auf die Versorgungsqualität, existieren bis heute keine Arbeiten (Milstein und Schreyögg 2020).

Tabelle 1: Anwendungsfelder für einen verbesserten Forschungsdatenzugang und Beispiele für eine Umsetzung

Anwendungsfeld	Zweck	Beispiele	potenzielle Datenquellen
Weiterentwicklung von Diagnostik, Therapie und Versorgung	Verbesserung der individuellen Behandlung	Evaluation von neuen Versorgungsformen, beispielsweise im Rahmen von Innovationsfondsprojekten	klinische Studiendaten, Registerdaten, Behandlungsdaten
Evaluation von Politikinterventionen	Weiterentwicklung der Gesetzgebung, Steuerung des Gesundheitssystems	Evaluation von hausarztzentrierter Versorgung oder Disease-Management-Programmen (Gerlach und Szecsenyi 2018; Weber et al. 2019), Gutachten zur Mengenentwicklung im stationären Sektor (Schreyögg et al. 2014)	Abrechnungsdaten, Behandlungsdaten, Registerdaten

Anwendungsfeld	Zweck	Beispiele	potenzielle Datenquellen
gesundheits- ökonomische Evaluation	effiziente Ressourcen- allokation im Solidarsystem	beispielsweise Modellierung von langfristigen Effekten von Impfungen, beispielsweise Rotavirus-Impfung (Aidelsburger et al. 2013)	Abrechnungsdaten, Behandlungsdaten, Registerdaten, klinische Studiendaten
Gesundheits- berichterstattung und Transparenz des Behandlungs- geschehens	Informationen für Entscheidungs- träger, Bürger- information	Gesundheitsberichterstattung des Bundes (www.gbe-bund.de), Patienteninformationsportal „Weiße Liste“ (<a href="http://www.weisse-
liste.de">www.weisse- liste.de)	Abrechnungsdaten, Behandlungsdaten, Register
Überprüfung von Arzneimittel- und Produktsicherheit nach Markt- einführung	Patienten- sicherheit	Verordnungsanalysen von neuen oralen Antikoagulanzen im Hin- blick auf korrekte Verordnung (Rottenkolber et al. 2020) und Blutungsrisiken (Souverain et al. 2020)	Abrechnungsdaten, Registerdaten
ergänzende Studien zur Zulassung von Arzneimitteln	Medikamenten- zulassung	mathematische Modellierungen und Simulationen zur Arznei- mittelsicherheit, beispielsweise bei der Zulassung von Kinder- arzneimitteln (Läer 2019)	Studiendaten, Behandlungsdaten
Überprüfung der Sicherheit von Behandlungen	Patientensicher- heit, Verbesserung der individuellen Behandlung	Überprüfung der Sicherheit von Implantaten, Gesetzliche Qualitätssicherung gemäß § 136 SGB V	Abrechnungsdaten, Behandlungsdaten, Register
Nutzenbewertung von Arzneimitteln in der Regel- versorgung	effiziente Ressourcen- allokation im Solidarsystem	beispielsweise frühe Nutzen- bewertung im Rahmen des AMNOG-Verfahrens nach § 35a SGB V	klinische Studiendaten, Registerdaten, Abrechnungsdaten
Weiterentwicklung der Messung von Qualität	Verbesserung der individuellen Be- handlung, Patien- tensicherheit, Bürgerinformation	Qualitätssicherung mit Routinedaten im Bereich der stationären Versorgung (WidO 2019)	Abrechnungsdaten, Behandlungsdaten, Register
Versorgungs- und Bedarfsplanung	Abbau von Über-, Unter- und Fehlversorgung	beispielsweise Gutachten zur Krankenhauslandschaft in NRW (Topf et al. 2019)	Register, Abrechnungsdaten
Produkt- und Verfahrens- innovationen	Entwicklung von Anwendungen aus der klinischen Praxis	(Weiter-)Entwicklung von digitalen Gesundheitsanwen- dungen oder Implantaten	Register, Abrechnungsdaten, Behandlungsdaten

Quelle: leicht modifiziert nach SVR (2021)

Es zeigt sich, dass es bereits bei der Konzeption von Reformen wichtig ist, deren Wirkung anhand von realen Daten abschätzen zu können. Oftmals zeigt sich in der Politikberatung, dass Politikerinnen und Politiker vor Konzeption eines neuen Gesetzes nach wissenschaftlicher Evidenz suchen, dann allerdings oftmals die Antwort erhalten,

dass die entsprechende Fragestellung in Deutschland aufgrund des fehlenden Datenzugangs nicht bearbeitet werden kann. Auch nach erfolgten Politikinterventionen (beispielsweise Terminservicestellen oder Pflegepersonaluntergrenzen) wird häufig von Politik und Verbänden Evidenz zur Wirkung dieser Reformen erwartet, um möglicherweise zu einem späteren Zeitpunkt nachsteuern zu können.

Andere europäische Länder, unter anderem England oder die Niederlande, zeigen, dass Reformen häufig rückgängig gemacht werden, wenn sich in Evaluationen nicht die gewünschte Wirkung zeigt. Auch die Erprobung von Politikinterventionen durch randomisierte Experimente ist generell wünschenswert. Falls Letzteres aber nicht möglich oder gewünscht ist, sollte zumindest die Möglichkeit bestehen, auf die bereits genannten Datenquellen zuzugreifen, anhand derer in der Regel ebenfalls eine Evaluation möglich ist. Ein guter Forschungsdatenzugang kann die Evidenz für Politikinterventionen deutlich verbessern und so Entscheidungsgrundlagen schaffen, Politikinterventionen effektiver auszurichten. Beispielweise hätte die Zusammenführung von Krankenkassendaten und Daten aus der ePA ein hohes Potenzial zur Evaluation der hausarztzentrierten Versorgung.

Es sollte bei der Diskussion um einen verbesserten Zugang zu Forschungsdaten nicht unerwähnt bleiben, dass dieser ein wichtiger Parameter bei der Standortwahl von Spitzenwissenschaftlern in Klinik, Versorgungsforschung und darüber hinaus darstellt. Nicht selten haben ausländische Wissenschaftlerinnen und Wissenschaftler Vorbehalte gegen den Standort Deutschland, da aus EU-Projekten oder anderen internationalen Projekten bekannt ist, wie restriktiv der Datenzugang hierzulande gestaltet ist. Spitzenwissenschaftlerinnen und -wissenschaftler im Gesundheitswesen geben wichtige Impulse für einen Wissenschafts- und Innovationsstandort Deutschland, da sie durch innovative Studien oftmals den Weg zu vielversprechenden Versorgungsinterventionen ebnen, internationale Forschungsgelder attrahieren oder auch über Spin-Offs wiederum Start-Ups gründen. Ihre Wirkung geht weit über die Publikation von Studien in wissenschaftlichen Fachzeitschriften mit wissenschaftlichem Impact hinaus. Daher sollte der Zugang zu Forschungsdaten als wichtiger Parameter für den Innovations- und Forschungsstandort Deutschland gesehen werden (Schreyögg 2020; SVR 2021).

Best-Practice-Beispiele aus anderen Ländern

Bei der möglichen Implementierung eines angemessenen und datenschutzkonformen Zugangs zu Gesundheitsdaten werden für Deutschland häufig sehr komplexe Szenarien diskutiert. Die Auslegung der Datenschutz-Grundverordnung (DSGVO) durch Datenschutzbehörden und andere Behörden erfolgt in Deutschland sehr restriktiv und verhindert einen angemessenen Zugang zu Gesundheitsdaten. Gleichzeitig wird eine heterogene Auslegung in der föderalen Struktur der Länder oftmals zur Hürde für Forscherinnen und Forscher. Als Beispiel seien hier Multi-Center-Trials genannt, die in unterschiedlichen Bundesländern durchgeführt werden. Dabei existieren in zahlreichen EU-Ländern bereits DSGVO-konforme Modelle für einen angemessenen Zugang zu Gesundheitsdaten. Das Grundprinzip aller Modelle ist eine zentrale Anlaufstelle für Wissenschaftlerinnen und Wissenschaftler und andere Interessenten, bei der sie Anträge zur Nutzung von gesundheitsbezogenen Daten, falls notwendig auch verknüpft mit anderen Datenquellen, stellen können. Eine zentrale Anlaufstelle heißt nicht, dass die zur Verfügung gestellten Daten auch zwingend an diesem zentralen Ort gespeichert werden. Vielmehr können sie projektbezogen aus dezentralen Orten durch diese zentrale Stelle zusammengezogen werden.

In der Regel prüft ein sogenanntes Use-and-Access-Committee die Anträge und entscheidet, ob Daten bereitgestellt werden sollen und, wenn ja, auf welchem Aggregationsniveau. Wissenschaftlerinnen und Wissenschaftler benötigen häufig Daten auf Individualebene, die gegebenenfalls auch mit anderen Datenquellen, beispielsweise mit Geodaten, verlinkt werden. Diese Datenbereitstellung unterliegt dann einem besonderen Schutz und geht in der Regel mit einer besonderen Prüfung der antragstellenden Organisation einher. Andere mögliche Antragsteller, beispielsweise Start-Ups, benötigen jedoch häufig nur Aggregatdaten, an die deutlich geringere Anforderungen in Bezug auf den Datenschutz gestellt werden. Ein Start-Up für eine Diabetes-App könnte beispielsweise daran interessiert sein, wie viele Menschen in Deutschland von der Diagnose „Diabetisches Fußsyndrom“ betroffen sind. Selbstverständlich kann analog auch forschenden Unternehmen aus dem Bereich Pharmazie und Medizintechnik unter bestimmten Voraussetzungen ein Zugang zu Gesundheitsdaten ermöglicht werden.

Viele EU-Länder haben mittlerweile einen Prozess für die Bereitstellung von Gesundheitsdaten für Forschungszwecke definiert und implementiert. Besonders führend sind dabei die skandinavischen Länder und England als ehemaliges EU-Land. Im Folgenden seien in Kürze Finnland und Frankreich als Beispiele für einen gelungenen Prozess zur Datenbereitstellung angeführt. Frankreich wird deshalb ausgewählt, weil es Deutschland strukturell ähnlicher ist als die skandinavischen Länder und England.

In Frankreich wurde im Jahr 2019 der sogenannte Health Data Hub (HDH) geschaffen, um den Datenzugang für Forschung und Entwicklung zu stärken. Der HDH soll zur zentralen Stelle für Antragstellende aus Wissenschaft und Forschung beziehungsweise Entwicklung werden. Dies soll über eine Plattform erfolgen, die eine bereits bestehende große Datenbank aus Abrechnungs-, Diagnose- und Sterberegisterdaten mit weiteren Datenbeständen vernetzt. Dazu sollen Behandlungsdaten der Krankenhäuser, Daten aus Patientenakten des ambulanten Sektors sowie weitere Datenquellen verlinkt werden. Es sollen sukzessive weitere dezentral gelagerte Datenbestände hinzukommen, die dann projektbezogen verlinkt werden können. Alle Forschungsprojekte im öffentlichen Interesse können beantragt werden. Ein Projekt kann im öffentlichen Interesse liegen, wenn es bestimmte Zwecke und Voraussetzungen erfüllt. Zweck kann beispielsweise eine Verbesserung der Versorgung oder eine Verbesserung des Gesundheitssystems sein.

Auch private Unternehmen können eine Nutzung von Daten für Projekte beantragen. Über die Anträge entscheidet eine Kommission, die sich aus gemeinsam von den Ministerien für Gesundheit sowie für Bildung und Forschung berufenen Experten zusammensetzt (Ministère des Solidarités et de la Santé 2020). Der HDH führt eine Antragsberatung durch und stellt nach erfolgter Bewilligung die Daten bereit. Individualdaten werden pseudonymisiert über eine gesicherte VPN-Verbindung bereitgestellt. Das heißt, Nutzerinnen und Nutzer können von ihrem eigenen Arbeitsplatz aus den Volldatensatz per Fernzugang sehen und diesen mit Statistikprogrammen bearbeiten. Es ist somit sichergestellt, dass auch eine notwendige Aufbereitung des Datensatzes für statistische Zwecke vorgenommen werden kann (SVR 2021).

Findata ist im finnischen Ministerium für Gesundheit und Soziales angesiedelt. Ziel von Findata ist ebenfalls die Gewährleistung eines zentralen angemessenen und datenschutzsicheren Zugangs und Prozesses zur Nutzung von Gesundheits- und Sozialdaten und gegebenenfalls weiteren Datenquellen, unter anderem zur Inanspruchnahme von Arzneimitteln, Daten aus Registern, der Todesursachenstatistik sowie der elektronischen Patientenakte. Findata gewährleistet dabei die Antragsbearbeitung, die Verknüpfung und die Bereitstellung der beantragten Daten in einem für Statistikprogramme nutzbaren und einheitlichen Format. Auch Findata ermöglicht die Nutzung von Gesundheitsdaten für Unternehmen. Anders als in Frankreich wird hier jedoch ex-ante festgelegt, dass Unternehmen (Entwicklungs- und Innovationsprojekte) keinen Zugang zu Individualdaten haben dürfen, sondern nur Aggregatdaten erhalten. Die berechneten Gebühren hängen von zahlreichen Variablen ab, unter anderem dem Aufwand der Datenbereitstellung und dem Aggregationsniveau. Die Gebührentabelle ist sehr transparent und unter folgendem Link zu finden: <https://findata.fi/en/pricing/>.

Wie in Frankreich werden die Daten nach Antragsgenehmigung aufbereitet und pseudonymisiert in geschützten virtuellen Räumen für Forschende per Fernzugriff zur Verfügung gestellt. Auch hier können Nutzerinnen und Nutzer direkt den pseudonymisierten Volldatensatz sehen und diesen anhand eines Statistikprogramms aufbereiten und anschließend statistische Auswertungen selbst durchführen (ohne hierfür wie beispielsweise bei der Krankenhausstatistik in Deutschland eine Schleife über die Mitarbeiterinnen und Mitarbeiter des Statistischen Bundesamtes nehmen zu müssen). Findatas Aktivitäten beruhen auf einer eigens geschaffenen gesetzlichen Grundlage, dem Gesetz über die Sekundärnutzung von Gesundheits- und Sozialdaten, welches die Datenverarbeitung, die Kombination mit Registerdaten und Datenschutzaspekte regelt. Ein Antrag auf Datennutzung besteht unter anderem aus einem Forschungsbeziehungsweise Projektplan inklusive Angabe des Nutzungszwecks, der benötigten Datensätze und methodischen Anforderungen an die Datenbereitstellung, Angaben zu den beteiligten Personen und zur Finanzierung sowie zu den getroffenen Sicherheitsvorkehrungen. Ein Ethikvotum ist nur dann erforderlich, wenn eigens erhobene Daten beispielsweise aus Survey- mit Sekundärdaten auf Individualebene verlinkt werden sollen (<https://findata.fi/en/>; SVR 2021).

Empfehlungen zur Verbesserung des Datenzugangs in Deutschland

Die Defizite des Zugangs zu Gesundheitsdaten für Forschungs- und Entwicklungszwecke in Deutschland sind seit langer Zeit in Fachkreisen bekannt. Während der Coronapandemie wurde dann auch für die allgemeine Öffentlichkeit deutlich, dass die wesentlichen Entscheidungen zur Impfung oder zu Eindämmungsmaßnahmen überwiegend auf Basis von Studien anderer Länder getroffen wurden. Dies ist nicht etwa auf die fehlende Qualifikation des Wissenschaftssystems in Deutschland zurückzuführen, sondern auf die Probleme des Datenzugangs. Insbesondere sind dabei der bisher fehlende Zugang auf den GKV-Volldatensatz, das Fehlen eines Unique Identifiers für eine Verlinkung von Datenquellen, das Fehlen von Registern mit nationaler Abdeckung sowie der Zeitverzug bei der Bereitstellung von Daten des ambulanten Systems zu nennen. Dabei wurde auch deutlich, dass die Problematik des Datenzugangs nicht primär ein Problem des Datenschutzes, sondern in erster Linie ein Datenschutzauslegungs- und Prozessproblem ist.

Andere EU-Länder zeigen, dass ein angemessener Zugang zu Gesundheitsdaten für Forschungs- und Entwicklungszwecke DSGVO-konform ermöglicht werden kann. Dabei zeigt sich, dass die definierten Prozesse über die Länder hinweg relativ ähnlich sind. Es wird üblicherweise eine zentrale Stelle eingerichtet, die Anträge entgegennimmt und dann die Daten, die oftmals dezentral gelagert sind, projektbezogen verknüpft und diese über eine VPN-Verbindung den Nutzerinnen und Nutzern bereitstellt.

In Deutschland würde es sich anbieten, zunächst analog zum finnischen Weg eine gesetzliche Grundlage im Rahmen eines Forschungsdatengesetzes zu schaffen, das dann die umfassende Bereitstellung und Verlinkung verfügbarer Datenquellen gewährleisten kann. Ein solches Forschungsdatengesetz ist im aktuellen Koalitionsvertrag vorgesehen und würde nicht nur Rechtssicherheit für einen DSGVO-konformen Prozess schaffen, es würde auch die derzeit starke Heterogenität der Datenschutzauslegung der Bundesländer beseitigen. In diesem Forschungsdatengesetz sollte auch unbedingt die Möglichkeit eines Unique Identifiers geschaffen werden, die für eine Verknüpfung von Datensätzen eine wichtige Voraussetzung ist.

Als zentrale Anlaufstelle könnte dann das Forschungsdatenzentrum Gesundheit beim BfArM fungieren. Dort würden Anträge eingehen und von einem unabhängigen, fachlich kompetenten und interdisziplinär besetzten Gremium (das sogenannte Use and Access Committee) in einem Kurzverfahren bewertet werden (SVR 2021). Die dort dann angebotenen Datensätze könnten, wie im Koalitionsvertrag vorgesehen, weiterhin dezentral gelagert werden. Sie würden bei entsprechenden Anträgen projektbezogen aus den dezentralen Orten zusammengezogen. Es wäre besonders wichtig, dass Forschende dort die Nutzung von Daten beantragen können, die mehrere Datenbestände verlinkt, zum Beispiel Abrechnungsdaten und elektronische Patientenakte oder auch Registerdaten. Diese neu zu schaffende Forschungsdaten-Infrastruktur könnte auch die Datenbestände der Medizininformatik-Initiative und andere Datenbestände aus öffentlich geförderten Projekten integrieren. Die Verknüpfung von verschiedenen Datensätzen ist auch deshalb so wichtig, weil man so einen breiten Datensatz mit vielen Variablen generiert, der oft auch als Big Data bezeichnet wird. Insbesondere für die Anwendung von KI-Methoden ist eine große Zahl von Variablen besonders wichtig. Mit einer innovativen Verknüpfungsmöglichkeit von Datensätzen könnte Deutschland perspektivisch im Datenzugang international sogar eine Spitzenposition einnehmen (Schreyögg 2020).

Neben der Schaffung einer zentralen Stelle und eines administrativ schlanken Prozesses sollte auch geklärt werden, wie und in welchem Umfang die Daten aus der ePA genutzt werden können. Das im Koalitionsvertrag vorgesehene Opt-Out-Verfahren ermöglicht potenziell eine hohe Vollständigkeit. Um den höchstmöglichen Nutzen zu erzielen, sollten die Daten aus der ePA grundsätzlich für Forschungszwecke zugänglich sein, analog zu Routinedaten, da dies auch aus Gründen der Solidarität, gerade mit Blick auf hochpreisige Therapeutika geboten ist. In diesem Kontext stellt sich eine ähnliche Frage wie bei Impfquoten, das bedeutet, erst durch eine hohe Teilnahme entsteht ein substanzieller Nutzen für die Gesellschaft (Schreyögg 2020).

Neben der Nutzbarkeit von ePA-Daten ist der Aufbau von qualitativ hochwertigen Registern mit nationaler Abdeckung eine zentrale Herausforderung für die nächsten Jahre. Hierfür sollte ein unabhängiges Gremium zur Erfassung und Bewertung bestehender Register sowie zur Identifikation von noch fehlenden, aber potenziell nutzen-

stiftenden Registern geschaffen werden (SVR 2021). Förderungswürdige Register sollten staatlich finanziert und deutschlandweit ausgebaut werden. Als Voraussetzung für eine Förderung sollten Mindestqualitätsstandards definiert werden, die diese Register zu erfüllen haben (Windeler 2021). Zuletzt sollte ein „Register für Register“ geschaffen werden, um Forschenden einen transparenten Überblick zu Informationen über Zielsetzung, Datenqualität, Umfang, Zugangsmöglichkeiten und -voraussetzungen zu geben.

Literatur

- Aidelsburger, P., Dietl, M., Böhm, K., Grabein, K., Wasem, J., Koch, J. et al. (2013). Gesundheitsökonomische Evaluation der Rotavirus-Impfung in Deutschland. Endbericht. Robert Koch-Institut. Berlin. Online unter https://www.rki.de/DE/Content/Infekt/Impfen/Forschungsprojekte/abgeschlossene_Projekte/Rotavirus-Impfung/Rota_inhalt.html (Download am 31. August 2022).
- DFG – Deutsche Forschungsgemeinschaft (2021). Stellungnahme der Kommission für Pandemieforschung. Online unter https://www.dfg.de/download/pdf/foerderung/corona_infos/stellungnahme_daten_gesundheitsforschung.pdf (Download am 6. Juli 2022).
- Gerlach, F. M. und J. Szecsenyi (2018). Evaluation der Hausarztzentrierten Versorgung (HZV) in Baden-Württemberg. Zusammenfassung der Ergebnisse – Ausgabe 2018. Online unter https://www.aok.de/gp/fileadmin/user_upload/Arzt_Praxis/Aerzte_Psychotherapeuten/Vertraege_Vereinbarungen/Hausarztzentrierte_Versorgung/Baden-Wuerttemberg/HZV-Evaluation_Broschuere_2020.pdf (Download am 31. August 2022).
- Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG). Online unter https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBI&jumpTo=bgbl119s2562.pdf#_bgbl_%2F%2F*%5B%40attr_id%3D%27bgbl119s2562.pdf%27%5D__1578918584390 (Download am 6. Juli 2022).
- Läer, S. (2019). Pharmakokinetische und pharmakodynamische Modellierung bei Kinderarzneimitteln: Schwerpunkt Physiologie-basiertes pharmakokinetisches Simulieren. In: Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 141–142. S. 66–73. doi: 10.1016/j.zefq.2019.03.003.

- Milstein, R. und J. Schreyögg, (2020). Empirische Evidenz zu den Wirkungen der Einführung des G-DRG Systems. In: Klauber, J., Geraedts, M., Friedrich, J. et al. (Hrsg.). Krankenhausreport 2020. Berlin, Heidelberg.
- Ministère des Solidarités et de la Santé (2020). Arrêté du 26 mai 2020 portant nomination des membres du Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé n° 0140 du 9 juin 2020, 2020 (texte n° 34). Online unter www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORF-TEXT000041976239&dateTexte=&categorieLien=id (Download am 6. Juli 2022).
- OECD (2015). Health Data Governance, Health Policy Studies. OECD, Paris.
- Rottenkolber, M., Schmiedl, S., Ibáñez, L., Sabaté, M., Ballarín, E., Vidal, X. et al. (2020). Prescribers' compliance with summary of product characteristics of dabigatran, rivaroxaban and apixaban – A European comparative drug utilization study. In: *Basic Clin Pharmacol Toxicol*, 128 (3). S. 440–454. doi: 10.1111/bcpt.13517.
- Schreyögg, J., Bäuml, M., Krämer, J., Dette, T., Busse, R. und A. Geissler (2014). Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG. Endbericht. Hamburg Center for Health Economics.
- Schreyögg, J. (2017). Big Data: Datenbestände für Wissenschaft und Patienteninformation effektiver nutzen. Bertelsmann Stiftung (Hrsg.). (Der digitale Patient). Online unter <https://blog.der-digitale-patient.de/big-data-datenbestaende-effektiver-nutzen/> (Download am 6. Juli 2022).
- Schreyögg, J. (2020). Datenschatz für die Versorgung von morgen. In: *Gesundheit und Gesellschaft G+G-Wissenschaft* 23 (2). S. 33–37.
- Schröder, H., Repschläger, U. und J. Walker (2021). Daten bündeln gegen Corona. In: *Gesundheit und Gesellschaft (G+G)*. Online unter <https://gg-digital.de/2021/05/thema-des-monats/daten-buendeln-gegen-corona/index.html> (Download am 6. Juli 2022).
- Sovereign, P. C., van den Ham, H. A., Huerta, C., Merino, E. M., Montero, D., León-Muñoz, L. M. et al. (2020). Comparing risk of major bleeding between users of different oral anticoagulants in patients with nonvalvular atrial fibrillation. In: *British Journal of Clinical Pharmacology*, 87 (3). S. 988–1000. doi: 10.1111/bcp.14450.
- SVR – Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (2018). Bedarfsgerechte Steuerung der Gesundheitsversorgung. Göttingen.

- SVR – Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (2021). Digitalisierung für Gesundheit – Ziele und Rahmenbedingungen eines dynamisch lernenden Gesundheitssystems. Göttingen.
- Topf, T., Ameln, W. von, Franken, M., Bösel, T., Fischer, V., Peukert, J. et al. (2019). Krankenhauslandschaft Nordrhein-Westfalen. Gutachten im Auftrag des Ministeriums für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen (MAGS). Online unter https://broschuerenservice.mags.nrw/mags/shop/Gutachten_Krankenhauslandschaft_Nordrhein-Westfalen/1 (Download am 31. August 2022).
- Weber, W., Berendes, A. und F. Potthoff (2019). Bericht der strukturierten Behandlungsprogramme der gesetzlichen Krankenkassen – Indikation Asthma bronchiale zum 30.06.2020. Berichtszeitraum 01.01.2006 bis 31.12.2018. Online unter https://www.g-ba.de/downloads/17-98-4941/2019-12-19_DMP-Evaluationsbericht_Asthma-bronchiale_infas_MNC.pdf (Download am 31. August 2022).
- WIdO – Wissenschaftliches Institut der AOK (2019). QSR-Verfahren. Revisionsbericht. Berlin.
- Windeler, J. (2021). Sammeln mit System. In: Gesundheit und Gesellschaft (G+G). Online unter <https://www.gg-digital.de/2021/05/sammeln-mit-system/index.html> (Download am 6. Juli 2022).