

Zi-Gutachten „Internationale Datengrundlagen für die Versorgungsforschung“

Wo man schon heute Daten aus diversen Quellen nutzt

Obwohl schon im November 2019 fertiggestellt, wurde das im Auftrag des Zentralinstituts für die kassenärztliche Versorgung in Deutschland (Zi) erstellte Gutachten „Internationale Datengrundlagen für die Versorgungsforschung“ aufgrund der Covid-19-Pandemie erst Ende September dieses Jahres veröffentlicht. Das Thema ist nach und besonders in Hinblick auf Covid-19 brisant, denn in dieser internationalen Vergleichsstudie haben der Gesundheitsökonom Prof. Dr. Reinhard Busse (TU Berlin) und sein Team untersucht, wie die Datengrundlage in Deutschland verbessert werden könnte: Es ist noch viel Luft nach oben!

>> Nach einer sehr ausführlichen Darstellung der in zwölf sorgfältig ausgesuchten Ländern vorhandenen Datenverfügbarkeit und -nutzung postuliert Busse, dass eine gute Datengrundlage für Forschungszwecke vor allem dann vorliege, „wenn alle Informationen zur Interaktion von einzelnen Patienten mit dem Gesundheitssystem an einer zentralen Stelle gehalten werden“. So könnten leistungserbringer- und ggf. sektorenübergreifende Versorgungsprozesse abgebildet werden, im optimalen Fall auch im Zeitverlauf. Diese Bedingung ist nach Aussage des im Auftrag des Zi vorgelegten Gutachtens grundsätzlich in denjenigen Gesundheitssystemen erfüllt, die ein „Gatekeeping“-System in der Primärversorgung verankert haben; ein typisches Beispiel hierfür seien die Hausärzte in Großbritannien. Das typischste Beispiel für eine Datenbank, die von Hausärzten erfassten Daten zusammenführt, ist laut Zi-Gutachten der Clinical Practice Research Datalink (CPRD), der Daten von Praxen zusammenführt, die sich in das Programm eingeschrieben haben („Opt-in“). Patienten dieser Praxen können sich wiederum von der Datenübermittlung ausschließen lassen („Opt-out“), im Einklang mit der nationalen Opt-out-Regelung des NHS. Die von CPRD gehaltenen Daten können für Beobachtungsstudien beantragt oder zur Durchführung von experimentellen Studien unterstützend herangezogen werden. Sie können darüber hinaus mit anderen Datensätzen von NHS Digital verknüpft werden, um reichere, sektorenübergreifende Einsichten zu ermöglichen. Bereits mehr als 2.000 wissenschaftliche Studien haben CPRD-Daten genutzt, erläutert das Gutachten.

Als weiteres positives Beispiel wird Schottland angeführt, in dem 2016 der Scottish Primary Care Information Resource (SPIRE) gegründet worden sei, um Daten aus schottischen Praxen zusammenzubringen, um so den dort tätigen Hausärzten die Möglichkeit zu geben, sich über ihr Patientenkollektiv im Vergleich zu informieren, die Versorgung besser zu planen, sowie den schottischen National Health Service (NHS) dabei zu unterstützen, die Qualität der Versorgung zu monitoren und zu verbessern.

Die Initiative betont, dass es bei SPIRE nicht um eine zentrale Datenbank aller Informationen zu schottischen Patienten geht, sondern um die zielspezifische Zusammenführung von Parametern aus Hausarzt-Daten für konkrete, bewilligte Zwecke. Daten würden daher auch nicht länger gespeichert, als es für die Bearbeitung der jeweiligen Anträge notwendig sei, die Informationen würden danach vernichtet.

Auch die Datenbank zur Primärversorgung des niederländischen Forschungsinstituts

Daten in der Primärversorgung					
Land	Initiative	Abdeckung	Opt-in/Opt-out	Zugriff für Forschung	Beispiele
UK	CPRD	Ca. 650 Praxen in 2016	Patienten können aus Datasharing aussteigen; Hausärzte nehmen freiwillig an der Übermittlung an CPRD teil. Ebenso ist die Zustimmung der Praxis für die Verlinkung der Daten notwendig	Zugriff ist für Wissenschaftler, Gesetzgeber/NHS-Beschäftigte und andere (inkl. Industrie) möglich	>2.300 Artikel, https://www.cprd.com/bibliography <ul style="list-style-type: none"> • Änderungen in den Todesursachen von COPD-Patienten 2005-2015 (Gayle et al. 2019) • Vergleichende Nutzenbewertung in der COPD-Arzneimitteltherapie (Suisa et al. 2019) • Internationaler Vergleich von „high-cost, high-need“-Patienten (Tanke et al. 2019)
UK Schottland	SPIRE	Rollout in Umsetzung	Opt-in für Praxen; Patienten können Nutzung ihrer Daten ausschließen	Ja, auf Antragsbasis (geplant)	Noch keine
Niederlande	NIVEL PC Database	Repräsentative Stichprobe von ca. 10% der Bevölkerung	Opt-in für Praxen; Opt-out für Patienten	Ja, auf Antragsbasis; NIVEL eigene Forschung	<ul style="list-style-type: none"> • Nutzung von außerplanmäßiger Primärversorgung während der Pflegereform nach Einkommen (Jansen et al. 2018) • Nutzung von Email-Beratung in der Primärversorgung (Huygens et al. 2018)

Tab. 1: Daten in der Primärversorgung, Quelle: Zi-Gutachten 2020.

NIVEL setzt Daten aus der Primärversorgung ein, allerdings kommen hier die Daten aus den elektronischen Praxisverwaltungssystemen auch von Leistungserbringern jenseits der Hausärzte zusammen. In den Niederlanden werde dieses Modell auch von anderen Institutionen bzw. Initiativen genutzt.

Allen diesen Initiativen, so ein Zwischenfazit, sei gemein, dass neben den Forschungsmöglichkeiten die Daten auch explizit für ein Feedback an die teilnehmenden Praxen genutzt werden, inklusive einer Möglichkeit, eigene Ergebnisse mit den von anderen Praxen zu vergleichen. Dies gäbe es auch in Ländern wie England, Kanada und den USA, wo für eine erfolgreiche Umsetzung eine „robuste technische Infrastruktur sowie

Daten aus verteilten Datensätzen					
Land	Fallstudie	Was wird verlinkt und wo liegen die Daten?	Wer führt das Linkage durch?	Zugriff für Forschung	Beispiele für veröffentlichte Analysen
Australien	Population Health Research Network (PHRN)	Es wird unterschieden zwischen Daten, die routinemäßig verknüpft werden und Daten, die zusätzlich verknüpft werden können. Dabei gibt es bei den zusätzlichen Daten Unterschiede nach verantwortlicher Region (Jurisdiction). Routinemäßig verknüpft werden Daten aus: Geburtenregister, Perinatale Daten (Schwangerschaft, Geburt etc.), Sterberegister, Krebsregister, Mental Health, öffentliche Notaufnahmen (Public Emergency Departments), private und öffentliche Krankenhausfälle. Die Daten sind bei den entsprechenden Institutionen angesiedelt.	In Abhängigkeit von den angefragten Datensätzen und der zuständigen Region (Jurisdiction) gibt es eine verantwortliche Data-Linkage Unit (DLU). Die DLUs sind alle Teil des PHRN.	Zugriff für die Forschung ist klar reguliert und läuft über ein Antragsverfahren. Zudem fallen Gebühren für die Verknüpfung bzw. Datennutzung an.	Überblick über einzelne Studien: https://www.phrn.org.au/for-the-community/case-studies/ Interessante Beispiele: <ul style="list-style-type: none"> • Suizidrisiko nach Inhaftierung in Queensland (Spittal et al. 2016) • Verbesserung der Schätzung von Mortalitätsratios in Krankenhäusern (Spilsbury et al. 2017)
Kanada	Social Data Linkage Environment (SDLE)	Zur Verfügung stehen Daten aus der Arbeitskräfteerhebung, nationalen Gesundheitsbefragungen, Krankenhausentlassungen, Zensus-Daten, umweltbezogene Makrodaten und eine Reihe von Provinzdaten über Kriminalität, Bildung, psychische Gesundheit usw. Die Daten sind in einer gesicherten Umgebung von Statistics Canada gespeichert; es gibt eine Trennung des DRD, der Source Data files und des Key Registry.	Statistics Canada	Ja, ein Zugang zu verknüpften Daten bzw. zur Datenverknüpfung kann beantragt werden	Übersicht positiv evaluierter Anträge: https://www.statcan.gc.ca/eng/record/2018
Kanada – Ontario	IS/ES linked data	Daten aus verschiedenen administrativen Quellen, Surveydaten sowie Registerdaten werden verknüpft. Es ist ausschließlich die Bevölkerung aus Ontario in den Daten abgebildet. Die Daten sind auf den Servern von IC/ES in einem geschlossenen Computing-System gehalten.	IC/ES	Der Zugang ist für Forscher außerhalb des IC/ES seit 2009 möglich und setzt einen erfolgreichen Antrag voraus.	Übersicht Publikationen: https://www.ices.on.ca/Publications/Journal-Articles Interessante Beispiele: <ul style="list-style-type: none"> • Zugang zu psychiatrischen Leistungen für Pflegepatienten (Perlman et al. 2019) • Luftverschmutzung und Risiko für Vorhofflimmern und Schlaganfall (Shin et al. 2019)
Kanada – British Columbia	PopData BC	Daten auf Individualebene, de-identifiziert und im Längsschnitt zur Bevölkerung von British Columbia. Daten kommen von verschiedenen „Partnern“ (u.a. Krebsregistern, Gesundheitsministerium, Statistics Canada). Es können eigene Datensätze der beantragenden Forscher auch verknüpft werden. Daten sind auf dem zentralen Server („Secure Research Environment“) gespeichert.	PopData BC	Ja, Zugang ist für die Forschung geregelt. Es gibt ein Antragsystem.	Liste Forschungsprojekte: https://www.popdata.bc.ca/ria/pro-jects Interessantes Beispiel: <ul style="list-style-type: none"> • Neugeborenen-Outcomes bei Vätern unter Medikation für Multiple Sklerose (Lu et al. 2014)
Kanada – Manitoba	Manitoba population research data repository	Administrative-, Survey-, Register- und weitere Daten werden für die Einwohner von Manitoba verlinkt. Daten kommen von verschiedenen Organisationen. Daten sind auf den Servern des Manitoba Center for Health Policy gespeichert.	Manitoba population data research data repository	Ja, Zugang ist für die Forschung geregelt. Es gibt ein Antragsystem.	Interessante Beispiele: <ul style="list-style-type: none"> • Kindergesundheit und Gerechtigkeit (Nickel et al. 2014) • Autismus-Risiko bei pränataler Antibiotikaexposition (Hamad et al. 2019) • Zusammenhang zwischen den sozialen Determinanten von Gesundheit und Versorgungsqualität (Katz et al. 2018)
Neuseeland	Integrated data infrastructure (IDI)	Beinhaltet Mikrodaten der Einwohner und Haushalte in Neuseeland. Die Daten sind im Längsschnitt und de-identifiziert. Die Daten kommen von verschiedenen Ministerien, Befragungen sowie von Nicht-Regierungs-Organisationen. Die Daten decken ein sehr breites Feld ab, u.a. den Gesundheitsbereich, Bildungsbereich und Polizeistatistiken. Daten sind auf den Servern der IDI gespeichert (>166 Milliarden Beobachtungen)	Stats NZ	Ja, Zugang ist für die Forschung geregelt. Es gibt ein Antragsystem.	Übersicht aller Forschungsprojekte: https://cdm20045.contentdm.o-clc.org/digital/collection/p20045coll17 Interessante Beispiele: <ul style="list-style-type: none"> • Zusammenhang zwischen Ruralität, Freilandexposition und Aufmerksamkeits-Defizit-Hyperaktivitäts-Störung (Donovan et al., fortl.) • Korrelation bestimmter Erkrankungen und das Vorliegen einer privaten (Zusatz)-Versicherung (Warren & Zhang 2018) • Finanzielle Belastung durch Verletzungen für ältere Arbeitnehmer (Davie & Lilley 2018)

Tab. 2: Modelle mit einer zentralen Stelle, die verteilte Datensätze (unterschiedlicher Datenhalter) verknüpft, Quelle: Zi-Gutachten 2020.

Kurzlink

zum Zi-Gutachten: <https://bit.ly/34uythz>

Zitationshinweis

Stegmaier, P.: „Wo man schon heute Daten aus diversen Quellen nutzt“, in: „Monitor Versorgungsforschung“ (06/20), S. 12-14; doi: <http://doi.org/10.24945/MVF.06.20.1866-0533.2257>

starke wissenschaftliche und politische Unterstützung unabdingbar“ gewesen seien.

Im deutschen Gesundheitssystem gibt es durchaus etwas vergleichbares: das Zi-Praxispanel. Ziel dieses Panel sei die Beobachtung von Qualität und Kosten, aber auch die Nutzung der Daten für die Versorgungsforschung. Der Aufbau der fachärztlichen Versorgung in Deutschland bedinge allerdings, dass im Gegensatz zu den oben beschriebenen Beispielen auch Fachärzte in diesem Praxispanel teilnehmen.

Daten aus unabhängig entstandenen Beständen

Ebenfalls wurden im Laufe der Recherche zum vorgelegten Gutachten verschiedene Datenquellen mit unterschiedlichem Inhalt identifiziert, die über eine oder mehrere Verlinkungsvariablen verknüpft werden können. Dies seien keine Daten, die für den primären Nutzungszweck integriert erhoben worden sind oder zu einem neuen Datenbestand zusammengeführt werden, sondern erst für den sekundären Nutzungszweck (in diesem Fall für die Forschung) verknüpft werden müssen. Dies könne zum einen notwendig sein, da die Daten für den primären Zweck nur einen sehr abgegrenzten Bereich abdecken (für Deutschland führt das Gutachten die §21er KHEntG-Daten als Beispiel an); oder zum anderen, weil noch eine zusätzliche Erweiterung eines bereits umfangreichen Datensatzes notwendig sein kann, um bestimmte Forschungsfragestellungen entsprechend beantworten zu können. Hier sind beispielsweise Verknüpfungen von Daten der gesundheitlichen Versorgung mit Daten zur Berufstätigkeit/Arbeitslosigkeit oder von

klinischen Daten mit Abrechnungsdaten gemeint.

Das übergreifende Ziel dieser Daten-Initiativen ist es, Daten aus verschiedenen Quellen, insbesondere von verschiedenen öffentlichen Institutionen, miteinander zu verlinken. Damit könnten zum Beispiel Daten zur gesundheitlichen Versorgung mit Daten aus dem Bildungsbereich und Kriminalitätsstatistiken verknüpft werden.

Auf Basis zentraler Datenhaltungs-Stellen für die Zusammenführung unabhängig entstandener Datensätze eröffnen sich nach Meinung der Gutachten-Autoren Möglichkeiten für die Versorgungsforschung. Beispiele dafür seien Initiativen wie eDRIS in Schottland oder PHARMO in den Niederlanden sowie die finnische Aktion FinData.

Unterschiedlich und relevant im Hinblick auf die Überlegungen für eine potenzielle Ideengewinnung für den deutschen Raum sei das Ausmaß der Konsolidierung der unterschiedlichen Datenbestände. So knüpfen zum Beispiel das australische Population Health Research Network (PHRN) und das kanadische Social Data Linkage Environment (SDLE) die Daten projektspezifisch zusammen, während die Quelldatensätze grundsätzlich verteilt bleiben. Im Gegensatz dazu fügt die „integrierte“ neuseeländische Integrated Data Infrastructure (IDI) eingegangene Daten mithilfe eines „Spine“ zusammen. Interessant im Ansatz des PHRN sei auch, dass die regionalen Behörden auf Bundesstaatsebene eigene Verlinkungsstellen haben, die im Netzwerk des PHRN kooperieren, die aber auch eigene Verlinkungen durchführen. Dieser Gedanke, so die Autoren des Gutachtens, könnte auch für das deutsche Gesundheitssystem in Erwägung gezogen werden. Sie denken da beispielsweise an eine Kombination mit einer Ausgestaltung des bereits 2011 gegründeten German Record Linkage Centers (GRLC) der Universität Duisburg-Essen in Kooperation mit dem Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA).

Unterschiede im Verlinkungsprozess

Eine erfolgreiche Verlinkung setzt nach Ansicht der Gutachten-Autoren die Möglichkeit einer eindeutigen Zuordnung von Informationen aus separaten Datensätzen voraus. Dazu seien geeignete Schlüsselvariablen notwendig, die in allen zu verlinkenden Datensätzen vorhanden sein müssen. Für eine personenbezogene Verlinkung sei es aus datenschutzrechtlichen Gründen zudem häufig notwendig, statt eindeutiger personenidentifizierender Merkmale aus den Originaldaten (wie etwa eine Versicherungs- oder Ausweisnummer) Pseudonyme einzusetzen. Dabei müsse jedoch für alle zu verlinkenden Datenquellen dasselbe Pseudonymisierungsverfahren angewendet werden.

Die Zuständigkeiten für die Pseudonymisierung der Daten sowie der Prozess der De-Identifikation der Datenbestände könne, so

die Autoren weiter, erheblich zur Bewahrung der Anonymität und zum Schutz der Privatsphäre beitragen.

In Deutschland wäre es bisher datenschutztechnisch herausfordernd gewesen, einen breiten Einsatz der Verlinkung von gesundheitsbezogenen Daten zu etablieren. Nichtsdestotrotz würden jedoch erfolgreiche Fallbeispiele zeigen, dass die entsprechenden Möglichkeiten besser ausgenutzt werden können. Auch seien mit der „Guten Praxis Datenlinkage“ methodische Standards zur Verlinkung personenbezogener Daten im Rahmen von Forschungsvorhaben formuliert worden.

Zwar gehe nach Meinung der Autoren eine detaillierte technische Beschreibung möglicher De-Identifikations- und Verlinkungsoptionen über den Rahmen des vorgelegten Gutachtens hinaus, nichtsdestotrotz erscheint es ihnen sinnvoll, beispielhaft vereinzelte Ansätze zu beleuchten, um erste Ideen für zukünftige Diskussionen zu erlauben.

So gebe es in der Stichprobe der in das Gutachten eingegangenen Fallstudien ein Cluster, welches ein Verfahren basierend auf einer zugrundeliegenden, „verankernden“ Liste von Personen nutzt, um mittels identifizierender Merkmale diese Personen in den zu verlinkenden Datensätzen zu identifizieren und somit Daten zu verknüpfen. Beispielshaft genannt wird hier das im neuseeländischen IDI verwandte „IDI Spine“. Dieses Spine stelle, so die Erklärung im Gutachten, eine Liste von Individuen dar, die Informationen der Finanzämter mit Geburten- und Visum-Angaben kombiniert. Ziel des Spines sei es, alle möglichen Individuen zu identifizieren, die jemals in Neuseeland gewohnt haben (Zielpopulation), die aber nur ein einziges Mal zu erfassen.

Alle Mikrodaten der im IDI abgebildeten Sektoren werden dann mit dem Spine verknüpft und könnten somit auch miteinander verlinkt werden, jedoch ohne ein eindeutiges identifizierendes Merkmal über die unterschiedlichen Sektoren hinweg zu verwenden. Ähnlich funktioniert auch das „Population Spine“ (PS) in Schottland, welches für die Verlinkungen über eDRIS herangezogen wird. PS bilde alle Personen in Schottland ab, die jemals in Kontakt mit NHS Scotland waren, samt ihrer Identifikatoren.

Die hier genannten Verlinkungsinitiativen würden aufzeigen, dass auch ohne das Vorhandensein eines eindeutigen Identifikators eine Verknüpfung auf Basis anderer identifizierender Merkmale möglich ist und seit mehreren Jahren erfolgreich eingesetzt wird. <<

von: MVF-Chefredakteur Peter Stegmaier

Novartis Deutschland



Neue Wege in der Medizin

Bei Novartis gehen wir die größten medizinischen Herausforderungen unserer Gesellschaft mit wissenschaftlicher Innovation an. Unsere Forscherinnen und Forscher treiben die Wissenschaft voran, um das Verständnis von Krankheiten zu vertiefen und neue Produkte zu entwickeln, die unerfüllte gesundheitliche Bedürfnisse befriedigen. Unsere Leidenschaft gilt der Erforschung neuer Methoden, um das Leben zu verbessern und zu verlängern.

