

Dr. Enno Swart, Institut für Sozialmedizin und Gesundheitsökonomie
Otto-von-Guericke-Universität Magdeburg
email: enno.swart@med.ovgu.de

Statement zum Vortrag über das PHRN

Das *Population Health Research Network (PHRN)* zeigt eindrucksvoll, dass es Vorreiter und Pioniere braucht, um vorliegende Gesundheitsdaten vollumfänglich und zum Wohle der Bevölkerung nutzen zu können. Das *PHRN* wird dabei flankiert/gestützt durch gesetzliche Maßnahmen zu Forschungsinitiativen --> es braucht, neben engagierten Forschern, einen aufgeschlossenen und zukunftsorientierten Gesetzgeber, speziell einen offenen Dialog zwischen Wissenschaft und Datenschutz, der die gegenseitigen Interessen und Belange anerkennt. Zwar gibt es in Deutschland ein Memorandum des Berliner und des Hessischen Datenschutzbeauftragten aus dem Jahr 2002, das aber angesichts neuer technischer Möglichkeiten und (datenschutz-)rechtlicher Vorschriften sowie neuer Forschungsansätze (Datenlinkage) überholt ist.

Die Bayrische Staatsregierung bzw. das Bayrische Staatsministerium für Gesundheit und Pflege hat das Potenzial/ den Nutzen der Gesundheitsdaten für die öffentliche Gesundheitsversorgung erkannt. Als Vertreter der Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Epidemiologie (DGEpi) und der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und Koordinator des Kompetenznetzes Sekundär- und Registerdaten der NAKO Gesundheitsstudie freut es mich umso mehr, dass dieses Vorhaben nun angegangen wird. Das geplante „Bayrische Gesundheitsdatenzentrum“ ist damit Vorreiter für eine regionale Gesundheitsberichterstattung (mithilfe von Sekundärdaten) und trägt den aus der wissenschaftlichen Community immer lauter werdenden Stimmen Rechnung, ein Forschungsdatenzentrum für Gesundheitsdaten zu etablieren. Bislang gibt es in Deutschland zwar Forschungsdatenzentren einzelner Dateneigner wie der statistischen Ämter auf Bundes- und Landesebene oder von Sozialversicherungsträgern, doch (noch) keinen integrierten Ansatz

Die Verknüpfung mehrerer Datenquellen wird in Australien im *PHRN* bereits aktiv durchgeführt, was sich nicht zuletzt in einer großen Anzahl wissenschaftlicher Publikationen widerspiegelt. In Deutschland wird man meines Erachtens noch mehrere Jahre auf die Implementierung und Arbeitsfähigkeit einer umfassenden, dateneignerübergreifenden Forschungsinfrastruktur warten müssen. Ursache dafür sind verschiedene Hürden, die derzeit einer breiten wissenschaftlichen Nutzung sensibler Daten entgegenstehen. Das betrifft vor allem die für klinische, epidemiologische und Versorgungsforscher hochinteressanten Routinedaten der Sozialversicherungsträger und hier wiederum der Abrechnungsdaten der gesetzlichen Krankenversicherung (GKV).

Hier sind zum einen gesetzliche Vorgaben zu nennen, vor allem das Sozialgesetzbuch (SGB) und die Bundes- und Landesdatenschutzgesetze (BDSG, LDSG). Die beinhalten konkrete und umfassende Auflagen bzgl. informed consent, der Definition konkreter Forschungsfragen und der zeitlichen Begrenzung der Datennutzung als Voraussetzung für eine wissenschaftliche Nutzung sensibler Daten. Auf die Notwendigkeit eines Patienteneinverständnisses hat der Bayerische Landesdatenschutzbeauftragte in einer Stellungnahme zum geplanten bayerischen Gesundheitsdatenzent-

rum bereits explizit hingewiesen. Die verschiedenen Zuständigkeiten innerhalb der föderalen Strukturen Deutschlands auf Ebene der Bundesländer (mit je eigenen Aufsichtsbehörden und Datenschutzbeauftragten,) bedingen einen hohen personellen, zeitlichen und organisatorischen Aufwand bei der Erschließung von Sekundärdatenquellen.

Nicht zu unterschätzen ist die weit verbreitete Skepsis in der Bevölkerung gegen den Aufbau und Betrieb umfassender Datenbanken mit sensiblen Daten, auch wenn diese ausschließlich wissenschaftlichen Zwecken dienen und von öffentlichen Institutionen getragen werden (Stichwort ‚Big Data‘ und ‚NSA-Affäre‘). Das Recht auf informationelle Selbstbestimmung ist in Deutschland ein vom Verfassungsgericht bestätigtes Grundrecht, das beim Aufbau einer Datenbank zu berücksichtigen ist.

Zum anderen stellt die technische und logistische Umsetzung einer Forschungsdatenbank Forscher immer noch vor große Herausforderungen: Es existieren nur teilweise definierte und standardisierte Datensätze (Krebsregister, Zi). Selbst bei Vorliegen gesetzlicher Vorschriften zur Struktur und zu Inhalten von Gesundheitsdaten, wie etwa bei den Abrechnungsdaten der Gesetzlichen Krankenversicherung (GKV), ist ein hoher Aufwand zur Homogenisierung der Daten vor der Einspielung in eine Forschungsdatenbank notwendig. Hinzu kommt die eingesetzte Erfassungs- bzw. Verwaltungssoftware bei der prinzipiell jeder Dateneigner frei in der Auswahl ist. Entsprechend heterogen sind die vorliegenden (Roh-)Daten in Format und Inhalt. Spezielle „Standardisierungsstellen“ existieren nicht. Hierbei muss weiterhin beachtet werden, dass einige Datenquellen regional oder institutionell begrenzt (z.B. auf eine der aktuell 113 gesetzlichen Krankenkassen, Stand 1.1.2017) und deren Beschaffung von mehreren Dateneignern in vielen Fällen somit unumgänglich sind. Den einzigen kassenübergreifenden Datensatz für Gesundheitsdaten (im engeren Sinne) auf nationaler Ebene beherbergt aktuell nur das Zentralinstitut für die kassenärztlichen Versorgung (Zi), das wissenschaftliche Forschungsinstitut der niedergelassenen Ärzte Deutschlands. (Wenn man von der Tatsache absieht, dass auch das Zi keine Daten zu den rund 10-15% privat krankenversicherten Personen hat [Beamte, Freiberufler, Selbstständige]). Ein nationales Mortalitätsregister für Deutschland befindet sich gerade in einer sehr frühen Aufbauphase nach mehreren Jahren Grundsatzdiskussion und Gesetzgebungsverfahren.

Der gesetzliche Auftrag, die Interessen der Dateneigner und/oder deren rechtlich/institutionellen Gegebenheiten erschweren die Erschließung der Daten ebenfalls. Während der primäre Verwendungszweck der Daten zumeist die Abrechnung von erbrachten Versichertenleistungen ist, rückt die potenzielle Nutzung der Daten für die Forschung (aus Kassensicht) in den Hintergrund und wird dominiert von deren Kerngeschäft. Existierende gesetzliche Grundlagen legen die Hürden für Forschung zudem hoch (§ 75 SGB X) oder sind zu unkonkret, was einen großen Interpretationsspielraum zulässt. Dies betrifft Ablauf und Dauer von Genehmigungsverfahren und die von den jeweiligen Aufsichtsbehörden geltend gemachten Voraussetzungen für die Nutzung von Sozialdaten im Allgemeinen und Krankenkassendaten im Besonderen).

Erfreulicherweise hat eine Initiative zur Novellierung des § 75 SGB X im Zuge der Umsetzung der 2016 neu gefassten EU-Datenschutzgrundverordnung jüngst Früchte getragen, sodass ab 2018 eine (spürbare?) Erleichterung beim Zugang und der wis-

senschaftlichen Nutzung der Sozialdaten zu erwarten ist, ohne dass derzeit die konkreten Veränderungen absehbar sind ().

Die Struktur der gesetzlichen Krankenversicherung macht zudem in Deutschland aktuell einen Veränderungsprozess durch, der durch die Fusion gesetzlicher Krankenkassen geprägt ist. Daraus resultieren knappe Personalressourcen, Veränderungen in der Gewichtung der Unternehmensziele und des unternehmerischen Selbstverständnisses (verstärkter Fokus auf den Wettbewerb zwischen einzelnen Kassen und der damit verbundenen Sicherung der Unternehmensposition am Markt). All dies wirkt sich momentan hemmend auf den Aufbau einer nationalen Forschungsinfrastruktur aus. Es fehlt bei vielfach zu beobachtendem grundsätzlichem Interesse an Kapazität seitens der Dateneigner diesen Prozess zu unterstützen.

Wissenschaftliche Expertise ist dem gegenüber jedoch in Deutschland vorhanden und es gibt es einen starken Zuwachs an methodischem, organisatorischem, rechtlichem Erfahrungswissen in der umfassenden Nutzung und Verlinkung verschiedener Datenquellen. Nicht zuletzt gibt es eine aktive Community, die sich bspw. über die AGENS austauscht und wissenschaftliche Standards etabliert. Mit der TMF - Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. steht auf der IT-technischen Seite ein kompetenter Ansprechpartner zur Verfügung und das ‚German Record Linkage Center‘ stellt das für Datenverknüpfungen notwendige methodische Knowhow zur Verfügung.

Das PHRN zeigt, dass der Aufbau einer solchen Forschungsstruktur mit Geduld gelingen kann. D. h. es braucht vor allem Zeit, um alle Datenquellen zu erschließen (Dieser Prozess ist auch beim PHRN noch nicht beendet).

Die beschriebenen Herausforderungen, ein ähnliches Vorhaben in Deutschland umzusetzen, werden in ihrer Vielzahl und Komplexität aktuell in der NAKO Gesundheitsstudie (NAKO), der bisher größten epidemiologischen Kohortenstudie in Deutschland, deutlich. Für eine Kohorte von insgesamt 200.000 Teilnehmern sollen verschiedenste Datenquellen erschlossen werden, die im Rahmen eines anlassbezogenen Linkage (die Formulierung einer konkreten Forschungsfrage ist Voraussetzung) individuell miteinander verknüpft werden können. Den Forschern werden dafür passgenaue, pseudonymisierte Datensätze über eine Transferstelle (ähnlich dem PHRN Data Delivery System (DDS)) zur Verfügung gestellt. Die Linkage Keys (Pseudonyme) selbst werden zweigeteilt generiert und verwaltet. Studienintern erfolgt dies über das ‚Trust Center‘ der NAKO, für die externe Nutzung wird durch die Transferstelle ein externes Pseudonym generiert, welches nur einmalig verwendet wird. Die Forschungsdatenbank enthält neben primären Befragungs- und Untersuchungsdaten, Bioproben und Ergebnissen bildgebender Verfahren (Sonographie, MRT) auch Sekundärdaten. Die Studie ist zunächst auf 10 Jahre angelegt, eine Fortführung über das bisherige Ende 2023 ist von allen Wissenschaftlern intendiert. Die Datenbank soll ab 2018/2019 für Dritte zugänglich sein.

Eine Datenfernverarbeitung ist (aktuell?) nicht vorgesehen, die Forscher rufen die benötigten Daten über ein Portal ab. [Ggf. Grafik der NAKO Struktur zeigen] Daten, die neben den Primärdaten genutzt werden können, umfassen einen Großteil der in Deutschland verfügbaren Datenquellen für Gesundheits- und Sozialdaten: gesetzliche Krankenkassen, private Krankenversicherungen, Krebsregister, Daten des Zi, Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) zur individuellen

Erwerbsbiographie und der Deutschen Rentenversicherung, Sterberegister, Patientenakten.

Mit der Etablierung des Bayrischen Gesundheitsdatenzentrums verbinde ich die Hoffnung, dass mit einer gewissen Strahlkraft die Nützlichkeit eines solchen in anderen Bundesländern wahrgenommen wird. Das Zi stellt mit seinen Datensätzen zu ambulanter vertragsärztlicher Versorgung und den Arzneimittelverordnungen, einen für den Start optimalen Kooperationspartner dar. Dieser Prototyp kann andere Bundesländer dazu motivieren, den gleichen Weg zu beschreiten und damit die Nutzung von bereits vorhandenen Gesundheitsdaten in der epidemiologischen Forschung auf eine neue Stufe zu heben.

Die AGENS begrüßt daher das Vorhaben für ein Bayr. Gesundheitsdatenzentrum und die frühzeitige Kooperation mit dem *PHRN*. Die AGENS selbst steht beiden Projekten/Einrichtungen für einen regen fachlichen Austausch jeder Zeit gern zur Verfügung.