

RatSWD Working Paper Series

RatSWD.
Rat für Sozial- und
WirtschaftsDaten

284

Für einen besseren Datenzugang im Bereich Gesundheit

Stellungnahme aus dem
Verein für Socialpolitik

Martin Fischer, Hendrik Jürges,
Stefan Mangelsdorf, Simon Reif,
Hannes Ullrich, Amelie Wuppermann

Dezember 2023

www.ratswd.de

RatSWD Working Papers

des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers*-Reihe startete Ende 2007. In dieser Online-Publikationsreihe werden konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autorinnen bzw. Autoren dar und nicht die des RatSWD. Die Zuwendungsgeber des RatSWD haben die Publikationen nicht beeinflusst.

Herausgeberin oder Herausgeber der RatSWD Working Papers-Reihe ist die/der Vorsitzende des RatSWD:

seit 2020 Monika Jungbauer-Gans

2014–2020 Regina T. Riphahn

2009–2014 Gert G. Wagner

2007–2008 Heike Solga

Für einen besseren Datenzugang im Bereich Gesundheit

Stellungnahme aus dem Verein für Socialpolitik

Martin Fischer, Hendrik Jürges, Stefan Mangelsdorf, Simon Reif,
Hannes Ullrich und Amelie Wuppermann*

Version 1.2 (23.10.2023)

<https://doi.org/10.17620/02671.88>

Dieses Paper ist außerdem beim Verein für Socialpolitik erschienen:
https://www.socialpolitik.de/de/vfs_stellungnahmen_daten_gesundheit

* Wir danken Stefanie March für zahlreiche hilfreiche Kommentare und Verbesserungsvorschläge.

Inhalt

1. Hintergrund	4
2. Bestandsaufnahme: der Zugang zu Gesundheitsdaten in Deutschland und seine Bewertung durch die Wissenschaftsgemeinschaft	5
2.1 Gesundheitsdaten in Deutschland	5
2.2. Bewertung des Zugangs zu Gesundheitsdaten durch die Wissenschaftsgemeinschaft.....	8
3. Relevanz verknüpfbarer Gesundheitsdaten in den Wirtschaftswissenschaften.....	9
4. Deutschland verliert gegenüber Nachbarländern an Boden	12
4.1 Beispiele für Forschung mit verknüpften Mikrodaten	13
4.2. Einfluss auf die Politikgestaltung.....	14
5. Maßnahmen zur Verbesserung des Zugangs zu Gesundheitsdaten in Deutschland	16
6. Fazit	18
Anhang.....	19

1. Hintergrund

Die Attraktivität des Forschungsstandortes Deutschland hängt von den Voraussetzungen ab, die Forscherinnen und Forscher hier für ihre Arbeit vorfinden. Viele Faktoren tragen dazu bei: Arbeitsbedingungen, Gehälter, oder lokale Netzwerke exzellenter Forscherinnen und Forscher. Für empirisch arbeitende Forscherinnen und Forscher ist der Zugang zu Daten, die international konkurrenzfähige Forschung erlauben, buchstäblich lebensnotwendig. Dies betrifft die empirische Wirtschaftsforschung insgesamt im gleichen Maße wie die empirische Gesundheitsforschung in Form von epidemiologischer und sozialmedizinischer Forschung, Versorgungsforschung und interdisziplinärer Forschung im Kontext des Gesundheitssystems zwischen Medizin und Ökonomie, beispielsweise die Gesundheitsökonomik.

Da Daten zur Beantwortung vieler zentraler Forschungsfragen nicht eigenständig im notwendigen Umfang und in notwendiger Qualität erhoben werden können, müssen Beobachtungsdaten aus dem Feld genutzt werden. Sind solche Daten nicht verfügbar bzw. werden diese nicht für die Forschung zur Verfügung gestellt, dann wandern exzellente Forscherinnen und Forscher ab. Entweder arbeiten sie mit ausländischen Daten und unterstützen die wissenschaftliche Politikberatung anderer Länder, indem etwa Reformen im Gesundheitswesen in diesen Ländern auf ihre Auswirkungen hin untersucht werden. Oder sie verlassen Deutschland möglicherweise ganz bzw. kehren nach Aufenthalt im Ausland nicht nach Deutschland zurück. Umgekehrt wirkt ein guter Datenzugang als Pull-Faktor für exzellente Forscherinnen und Forscher aus dem Ausland.

Aus wirtschaftswissenschaftlicher Perspektive zeigt sich gerade im Gesundheitsbereich eine Diskrepanz zwischen Forschungsbedarfen und den am Forschungsstandort Deutschland vorgefundenen Bedingungen. Eine Befragung unter Mitgliedern des Vereins für Socialpolitik aus jüngster Zeit belegt dies eindrucksvoll.¹ In den Gesundheitswissenschaften bestehen schon seit längerem Initiativen, die sich für die Verbesserung des Datenzugangs einsetzen, wie etwa die Arbeitsgruppe "Erhebung und Nutzung von Sekundärdaten" (AGENS) oder die AG "Validierung und Linkage von Sekundärdaten" beim Deutschen Netzwerk für Versorgungsforschung. Dies unterstreicht, dass der Zugang zu Gesundheitsdaten nicht nur unter Gesundheitsökonom*innen als kritisch angesehen wird.

Das deutsche Gesundheitswesen ist ein komplexes, historisch gewachsenes Zusammenspiel verschiedener Institutionen, und es unterscheidet sich in manchen Bereichen stark von dem anderer westlicher Länder. Beispielhaft seien das Nebeneinander von gesetzlicher und privater Krankenversicherung, die Art der Vergütung im ambulanten Bereich, die doppelte Vorhaltung von Fachärzten, sowohl ambulant als auch stationär, oder auch die Art der Organisation des Rettungsdienstes genannt. Zwar lassen sich aus Forschungsergebnissen, die mit Daten aus anderen Ländern – und entsprechend anderen institutionellen Kontexten – gewonnen wurden, wichtige Impulse für die Weiterentwicklung des Gesundheitssystems in Deutschland ableiten.

¹ Einzelergebnisse aus dieser Befragung berichten wir in Abschnitt 2.2.

Für die Identifikation von Problemen und auch die Evaluation von Reformen im deutschen Gesundheitswesen benötigt die gesundheitsökonomische Forschung allerdings Gesundheitsdaten aus Deutschland. Überspitzt formuliert stellt sich die Frage, ob es der deutschen Gesundheitspolitik gleichgültig sein kann, dass exzellente deutsche Wissenschaftlerinnen und Wissenschaftler lieber über das amerikanische, englische oder dänische als das deutsche Gesundheitssystem forschen. Wäre stattdessen nicht das Umgekehrte wünschenswert, nämlich dass (auch) internationale Wissenschaftlerinnen und Wissenschaftler über das deutsche Gesundheitssystem forschen?

2. Bestandsaufnahme: der Zugang zu Gesundheitsdaten in Deutschland und seine Bewertung durch die Wissenschaftsgemeinschaft

2.1 Gesundheitsdaten in Deutschland

In Deutschland sind durchaus hochwertige Daten für gesundheitsökonomische Forschung vorhanden. Die Datenlandschaft ist indes aufgrund der Vielzahl und Heterogenität an Datenhaltern sehr komplex. Nicht alle Daten, die erfasst werden, stehen für die Forschung zur Verfügung. Zudem liegen Daten aus dem Gesundheitsbereich meist nur fragmentiert vor, und der Zugang ist oft intransparent geregelt. Gemeint sind hier etwa Abrechnungsdaten der GKV und PKV, Daten der gesetzlichen Unfallversicherung und der Rentenversicherungsträger, Daten aus der elektronischen Patientenakte (ePA), Daten aus Krankheitsregistern (z.B. klinische Krebsregister), Daten aus Schuleingangsuntersuchungen, Daten der Beihilfestellen, Daten der Qualitätssicherung, der Apothekenrechenzentren sowie der Kalkulationskrankenhäuser. Aufgrund mangelnder Verknüpfbarkeit auch mit anderen Daten, z.B. Befragungsdaten oder Daten aus anderen Zweigen der Sozialversicherung, können die bestehenden Gesundheitsdaten oft nicht in dem Maße verwendet werden, das für viele gesundheitspolitisch relevante Forschungsprojekte erforderlich wäre. Im Folgenden geben wir einen beispielhaften Überblick über bestehende Gesundheitsdaten – ohne Anspruch auf Vollständigkeit – und spezifische Probleme im Zugang.

Die erste Gruppe sind Daten der amtlichen Statistik, wie Geburtsregisterdaten, die Krankenhausstatistik, die DRG-Statistik, die Todesursachenstatistik oder der Mikrozensus. Sie sind als Mikrodaten an Gastwissenschaftsarbetsplätzen und durch Remote Access bei den Forschungsdatenzentren von Bund und Ländern verfügbar. Abgesehen von teilweise langen Wartezeiten auf Termine vor Ort oder bei der Bearbeitung von Auswertungen schränken hier vor allem die mangelnde Verknüpfbarkeit von Beobachtungseinheiten i) innerhalb eines Datensatzes ii) zwischen Datensätzen und iii) mit externen Quellen den Umfang und die Aussagekraft von Forschungsprojekten ein. Dabei handelt es sich bei der mangelnden

Verknüpfbarkeit überwiegend um eine regulatorische Einschränkung, denn in den Rohdaten sind zumeist Merkmale enthalten, die eine Verknüpfung im Prinzip erlauben würden.²

Die zweite Gruppe umfasst Abrechnungsdaten. Diese entstehen bei der Abrechnung von Leistungen zwischen den Leistungserbringern (z.B. Arztpraxen, Krankenhäuser, Apotheken) und den gesetzlichen Krankenkassen, den kassenärztlichen Vereinigungen oder den privaten Krankenversicherungen. Für die gesetzlichen Versicherungen ergibt sich hier ein besseres – wenn auch unvollständiges – Bild der Inanspruchnahme von Gesundheitsleistungen als bei Privatversicherungen. Mit Ausnahme der Daten zu Krankenhausabrechnungen, die zur amtlichen Statistik gehören, ist der Zugang zu diesen Daten für die Wissenschaft nur intransparent geregelt.³ Durch das Forschungsdatenzentrum Gesundheit am BfArM soll in Zukunft ein einheitlicher Zugang zu den individuellen Abrechnungsdaten aus der GKV ermöglicht werden. Außerdem soll auch ein Weg geschaffen werden, mit dem diese über die Zeit und über Sektorengrenzen hinweg verknüpft werden können. Zu diesem Zweck ist die Schaffung einer „anlassbezogenen“ Forschungskennziffer angedacht. Der Start des FDZ Gesundheit wird allerdings seit der ersten Ankündigung wiederholt verschoben und ist aktuell nicht absehbar. Zusätzlich sind weitere Daten, die im Rahmen der Abrechnung anfallen, bis jetzt nicht für den Datenpool am FDZ Gesundheit vorgesehen, was teilweise dem Problem fehlender einheitlicher Identifikationsnummern geschuldet ist. So werden Daten über medizinische Leistungen fehlen, die durch andere Kostenträger (etwa Renten- oder Unfallversicherung) bezahlt werden. Ohne diese Leistungen ergibt sich nur ein unvollständiges Bild der betroffenen Personen. Auch nicht enthalten sein sollen Indikatoren der Medikamente ausgebenden Apotheken. Außerdem sind Daten zu Rabattverträgen zwar öffentlich, allerdings fehlen hier entscheidende Informationen für fundierte Analysen.

Des Weiteren gibt es Gesundheitsdaten, die auf gesetzlicher Grundlage für andere Zwecke erhoben werden. Hierzu zählen u.a. die Schuleingangsuntersuchungen, die potenziell Aufschluss über den Gesundheitszustand aller Kinder in Deutschland geben könnten. Allerdings ist der Wissenschaft der Zugang bislang nur vereinzelt gelungen und nicht allgemein geregelt. Zudem sind die Daten aufgrund der föderalen Struktur heterogen. Ein weiteres Beispiel sind die Krankenhauspläne der einzelnen Bundesländer, die vielfach nicht maschinenlesbar vorliegen. Auch gibt es eine große Anzahl an Krankheitsregistern, bspw. die klinischen und epidemiologischen Krebsregister, die verknüpft mit weiteren Informationen

² Aufgrund der Heterogenität der Datenquellen gibt es aber auch technische Hürden. Eine Übersicht zum aktuellen Stand der Möglichkeiten und Einschränkungen der Verknüpfung administrativer Gesundheitsdaten untereinander oder mit Surveydaten geben March, Stefanie et al. (2018): "Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme." *Gesundheitswesen* 80: e20–e31. Im August 2023 erschienen ist das "White Paper: Verbesserung des Record Linkage für die Gesundheitsforschung in Deutschland" der Nationalen Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten (nfdi4health) mit einer Vielzahl von Use Cases.

³ Prinzipiell ist über §75 SGB X die Übermittlung von Sozialdaten für die Forschung als „Kann“-Bestimmung normiert. Problematisch sind in der Anwendung die Heterogenität der Datenbestände und die unterschiedliche Auslegung der Kann-Bestimmung durch einzelne Datenhalter.

über die betroffenen Individuen für die gesundheitsökonomische Forschung wertvoll wären, zu denen der Zugang aber uneinheitlich geregelt ist.⁴

Im Zusammenhang mit der externen Qualitätssicherung können Daten über das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) ausgewertet werden. Allerdings gestaltet sich die Auswertung in verschiedener Hinsicht als sehr schwierig. Die Probleme reichen von einer mangelnden Verknüpfbarkeit einzelner Datenbestände (wie Geburtshilfe und Perinatalversorgung) über fehlerhafte und unvollständige Dummy-Datensätze bis hin zu einer lediglich einmaligen Übermittlung der Endresultate.

In Deutschland gibt es außerdem eine Reihe von Befragungs- und Untersuchungssurveys, die genuin für wissenschaftliche Analysen erhoben werden. Eine Verknüpfung dieser Befragungsdaten mit anderen administrativen Daten ist teilweise schon gelungen, z.B. des Survey of Health, Ageing and Retirement in Europe (SHARE) mit Daten der Rentenversicherung, und sollte weiter vorangetrieben werden. Der Datenzugang bei diesen Surveys ist über entsprechende Scientific Use Files oder Forschungsdatenzentren möglich. Bei den regelmäßig durchgeführten Gesundheitssurveys bspw. des RKI stoßen externe Wissenschaftlerinnen und Wissenschaftler jedoch auf das Problem, dass bestimmte Teile der Daten (z.B. Interviewmonat) nicht zur Verfügung gestellt werden und die Daten nicht über Regionalkennziffern mit wichtigen Kontextinformationen verknüpft werden können.⁵

Eine Vielzahl von weiteren Daten der Leistungserbringer hat großes Potenzial für die gesundheitsökonomische Forschung, z.B. im Bereich der Notfallversorgung. Daten zur Rettungsdiensteinsätzen sind aber je nach Bundesland eingeschränkt für die Forschung nutzbar, nicht für die Forschung freigegeben oder werden angeblich nicht erhoben. Auch Daten zu Behandlungsfällen in Krankenhausnotaufnahmen sind für die Forschung kaum verfügbar. Zwar sind Informationen zu Notaufnahmefällen und -auslastung in den einzelnen Krankenhäusern vorhanden, allerdings gibt es keine Möglichkeit, diese Daten strukturiert für eine Vielzahl von Krankenhäusern zu nutzen. Reine Notaufnahmefälle sind nicht in der DRG-Statistik enthalten, lassen sich aber auch in ambulanten Abrechnungsdaten der Krankenhäuser nur unvollständig identifizieren. Hieraus ergibt sich ein Bild, dass sowohl die Zersplitterung, die mangelnde Verknüpfbarkeit, sowie der fehlende Zugang zu Daten bisher kaum überwindbare Hürden für die empirische Forschung darstellen.

Wie sich in einigen derzeit laufenden Gesetzgebungsprozessen zeigt, ist das Problem des Datenzugangs schon in der Politik angekommen. Allerdings laufen die Initiativen zur Verbesserung der Nutzung von Gesundheitsdaten in der Forschung parallel und getrennt von Initiativen zur Nutzung aller anderen Datenquellen, so dass die genannten Probleme nur

⁴ Im Zuge der Errichtung des Forschungsdatenzentrums Gesundheit sollen die Krebsregister mit den GKV-Abrechnungsdaten verknüpfbar sein.

⁵ Manchmal werden die Daten auch nur unter der Voraussetzung einer Zubilligung von Koautorenschaften in wissenschaftlichen Veröffentlichungen geteilt. Das mag bei Umfragedaten, die Forschende eigeninitiativ in ihrer eigenen Forschungszeit und mit eigener Finanzierung erhoben haben, gerechtfertigt sein, nicht aber bei Umfragedaten, die Behörden im Rahmen ihres gesetzlichen Auftrags erheben.

unvollständig adressiert werden. So hat das Bundeskabinett kürzlich den Entwurf für ein Gesundheitsdatennutzungsgesetz des Bundesministeriums für Gesundheit beschlossen (demzufolge u.a. eine Datenzugangs- und Koordinierungsstelle beim BfArM eingerichtet und das Forschungsdatenzentrum Gesundheit beim BfArM weiterentwickelt werden soll).⁶ Parallel wird das Dateninstitut des Bundes federführend vom Bundesministerium für Wirtschaft und Klimaschutz und vom Bundesministerium des Innern und für Heimat entwickelt. Das Bundesministerium für Bildung und Forschung arbeitet seinerseits an einem Forschungsdatengesetz.

2.2. Bewertung des Zugangs zu Gesundheitsdaten durch die Wissenschaftsgemeinschaft

Die im vorangegangenen Abschnitt beschriebenen Defizite und Probleme beim Zugang zu Gesundheitsdaten spiegeln sich in einer negativen subjektiven Bewertung der wissenschaftlichen Community wider. Der Verein für Socialpolitik (VfS) hat zu Beginn des Jahres 2023 eine Online-Befragung seiner Mitglieder zur Zufriedenheit mit dem Zugang zu Forschungsdaten durchgeführt. Dass über 500 Mitglieder des Vereins für Socialpolitik an der Befragung teilgenommen haben, zeigt, dass das Thema den Ökonominnen und Ökonomen in Deutschland sehr wichtig ist.

Die Teilnehmerinnen und Teilnehmer konnten sich selbst, mit Möglichkeit der Mehrfachnennung, zentralen Arbeitsgebieten zuordnen. Im Folgenden berichten wir über die Antworten der 42 Befragten aus Deutschland, die „Gesundheitsökonomik“ als Forschungsschwerpunkt angaben. Jeweils etwa zwei Drittel dieser Befragten nannten als Datenquellen für ihre Forschung Umfragedaten (z.B. das Sozio-ökonomische Panel), administrative Mikrodaten und selbst erhobene Experimental- oder Befragungsdaten.

Zentrales Ergebnis der Befragung ist eine weit verbreitete Unzufriedenheit mit dem Datenzugang in Deutschland. 69% der Befragten gaben an, „eher unzufrieden“ oder „ganz unzufrieden“ zu sein, 31% waren „eher zufrieden“ und niemand war ganz zufrieden. Der Aussage „der Datenzugang in Deutschland ist sehr gut“ stimmten nur 12% eher oder voll zu. Die restlichen 88% stimmten eher nicht oder gar nicht zu.

Die drei meistgenannten Probleme und Mängel beim Datenzugang waren die unzulängliche Bereitstellung vorhandener Daten (76% Nennungen), die fehlende Verknüpfbarkeit verschiedener Datenquellen (67%) und restriktive Datenschutzvorgaben (62%). Eher selten wurden „finanzielle Hürden“ (7%), eine begrenzte Nutzungsdauer (17%) oder mangelnde Nutzerfreundlichkeit (24%) genannt.

⁶ Zum Gesundheitsdatennutzungsgesetz bzw. zum Forschungsdatenzentrum Gesundheit haben Vertreterinnen und Vertreter der Versorgungsforschung bereits ausführlich ihre Wünsche und Visionen formuliert. Vgl. etwa Schmitt, Jochen et al. (2023): „Das Gesundheitsdatennutzungsgesetz – Potenzial für eine bessere Forschung und Gesundheitsversorgung“, *Gesundheitswesen* 85: 215–222 oder March, Stefanie et al. (2023): „Forschungsdatenzentrum Gesundheit – Vision für eine Weiterentwicklung aus Sicht der Forschung.“ *Gesundheitswesen* 85 (Suppl. 2): S145–S153.

Als Folge der Datenzugangsprobleme sahen 75% der Befragten negative Auswirkungen auf ihren Forschungserfolg und 88% gaben an, dass sie Aufgrund der Probleme beim Datenzugang Projekte nicht durchführen konnten. 85% der befragten Gesundheitsökonominnen in Deutschland sahen als praktische Auswirkung des Datenzugangs Erschwernisse, relevante und aktuelle Politikberatung zu fundieren.

Derartige Standortnachteile führen dazu, dass der (positiven) Aussage „Der Datenzugang in Deutschland erlaubt es, im internationalen Forschungswettbewerb mitzuhalten“ nur 28% der Befragten eher oder voll zustimmten. Konsequenz ist, dass die Befragten auf ausländische Daten ausweichen. 70% der Befragten gaben an, schon einmal mit ausländischen Daten gearbeitet zu haben, davon taten dies 45%, weil geeignete deutsche Daten nicht zur Verfügung standen. 93% der Befragten stimmten der Aussage „Der Datenzugang in Deutschland ist schlechter als in vergleichbaren Ländern“ eher oder voll zu. In der Gesamtschau ergibt sich ein recht negatives Stimmungsbild, das sich im Übrigen nicht sehr von dem der Befragten mit anderen Forschungsschwerpunkten unterscheidet.

3. Relevanz verknüpfbarer Gesundheitsdaten in den Wirtschaftswissenschaften

Der Gesundheitssektor weist mit einem Anteil von über 13% des BIP und erheblichem staatlichen Lenkungsanteil ein großes Potential für Fehlanreize und Ressourcenverschwendung auf. Die Wirtschaftswissenschaften untersuchen die Wirkungsfaktoren, die zu beobachteten Ergebnissen von Marktprozessen führen, und evaluieren gesellschaftliche und politische Möglichkeiten, diese Ergebnisse durch Interventionen zu verbessern. Daraus folgt, dass bei der wirtschaftswissenschaftlichen Gesundheitsforschung nicht nur Informationen über den Gesundheitszustand der Bevölkerung, medizinische Behandlungen und deren Ergebnisse benötigt werden. Dies sind wichtige Ergebnisgrößen, aber empirisch arbeitende Ökonomen sind vor allem an der Funktion von Märkten und der Wirkung von Anreizen im Gesundheitssystem und darüber hinaus interessiert. Daher ist es essentiell, Informationen aus dem Gesundheitssystem mit Daten zu nicht-gesundheitlichen Merkmalen (Bildung, Erwerbsbeteiligung, Einkommen, Familienkontext, usw.) verknüpfen zu können.

Politikmaßnahmen haben typischerweise komplexe Wirkungskanäle, die anhand von aggregierten Daten in der Regel nicht nachvollziehbar sind. Da individuelle Personen, Firmen und Institutionen ihre Zielsetzungen unter heterogenen Rahmenbedingungen optimieren, verdecken aggregierte Daten individuelle Verhaltensreaktionen auf Politikmaßnahmen und sich verändernde Rahmenbedingungen. Auf Basis von Aggregatdaten können Wirkungskanäle somit nicht zuverlässig erfasst werden.⁷ Um Politikmaßnahmen und Regulierungen ex ante zu gestalten oder ex post zu evaluieren, sind daher Individualdaten, die auf niedrigstmöglicher

⁷ Das Phänomen ist als ökologischer Fehlschluss bzw. Simpson's Paradox bekannt.

Aggregationsebene (z. B. individuelle Personen und Produktionsstätten) verknüpft werden können, essentiell.

An Beispielen für wichtige, aktuelle Forschungsfragen, die für die Verbesserung des deutschen Gesundheitssystems bearbeitet werden sollten, aber aufgrund der aktuellen Datenlage nicht oder nur eingeschränkt durchführbar sind, mangelt es nicht (s.u.). In der gesundheitspolitischen Diskussion werden Datenlücken aber meist nur dann thematisiert, wenn die Datengrundlage für akute Entscheidungen fehlt. Viele Fragestellungen im Zusammenhang mit der Coronavirus-Pandemie sind hierfür beispielhaft. Aber auch die Diskussion zur aktuellen Krankenhausreform, deren Auswirkungen nur bedingt im Vornherein simuliert und somit abgeschätzt werden können oder die offene Frage, ob Versorgungsengpässe bei Arzneimitteln ein nationales oder regional begrenztes Phänomen sind, stehen exemplarisch für den Mangel an Gesundheitsdaten, die Wissenschaft und Politik zur Verfügung stehen.

In Deutschland wird bereits eine Vielzahl an Forschungsprojekten zur Verbesserung des Gesundheitssystems durchgeführt und mit großen Summen vom Staat gefördert (zum Beispiel auch durch den Innovationsfonds). Oft zeichnen sich diese Projekte im Bereich der Versorgungsforschung dadurch aus, dass eine (komplexe) Intervention konzipiert und getestet wird. Entsprechend können prospektiv Daten erhoben werden, zum Beispiel auch mit kooperierenden Krankenkassen. Für viele gesundheitsökonomische Fragestellungen, die sich mit der Funktion von Märkten befassen, ist ein solches, prospektives Vorgehen ungeeignet. Es wäre ethisch fragwürdig und politisch kaum durchsetzbar, in zufällig ausgewählten Regionen das Versicherungssystem zu ändern. Stattdessen nutzen moderne ökonometrische Verfahren in der Gesundheitsökonomie Quasiexperimente und retrospektive Daten, um kausale Rückschlüsse zu ermöglichen. Ohne geeignete Daten jedoch sind viele Analysen nicht möglich.⁸

Im Folgenden nennen wir beispielhaft einige Fragen, deren Beantwortung für die langfristige Entwicklung des Gesundheitssystems wichtig wäre, mit den aktuell in Deutschland verfügbaren Daten aber nicht oder nur unter unverhältnismäßig hohem Aufwand möglich ist.⁹

- Ist der Übergang von der gesetzlichen zur privaten Krankenversicherung (und zurück) durch Adverse Selektion bzw. Risikoselektion geprägt und in welchem Ausmaß? Idealerweise hätte man für die Analyse dieser Fragestellung eine gemeinsame Datenbasis von gesetzlicher und privater Krankenversicherung, die es erlaubte,

⁸ Ein Beispiel ist die Einführung (und Abschaffung) der Praxisgebühr. Die Auswirkung von Zuzahlungen auf die Nachfrage nach medizinischen Leistungen ist Kernthema der Gesundheitsökonomik und von höchstem wissenschaftlichen und gesundheitspolitischen Interesse. Die Evaluation der Praxisgebühr erfolgte aber eher zufällig auf Eigeninitiative einiger Forschergruppen und musste zu Beginn mit im Prinzip untauglichen Umfragedaten auskommen.

⁹ Selbstverständlich kann für einzelne Projekte über die Anbahnung persönlicher Kontakte zu Entscheidungsträgern bei (einzelnen) Krankenkassen oder Krankenhäusern Interesse an den eigenen Forschungsfragen geweckt und Datenquellen für (einzelne) Forschungsfragen erschlossen werden. Solche Insellösungen sind aber weder effizient noch transparent.

Übergänge zu verfolgen. Dazu bedürfte es eines einheitlichen Identifikationsmerkmals und einer Harmonisierung der Datenbestände von GKV und PKV.

- Wie würde sich die Versorgung von gesetzlich und privat versicherten Patienten verändern, wenn die Vergütung im ambulanten Sektor angeglichen würde? Würde sich die Gesundheit der gesetzlich Versicherten verbessern und die der privat Versicherten verschlechtern. Dazu müsste man mittels harmonisierter Datenbestände von GKV und PKV untersuchen, ob das Vergütungssystem Anreize zur anbieterinduzierter Nachfrage ausübt.
- Führen Selektivverträge in der GKV zu besserer Versorgung oder dienen diese der Risikoselektion? Idealerweise wüsste man, welche Krankenversicherung welche Leistungen selektiv anbietet und von welchen Versicherten dies genutzt wird.
- Führt Wettbewerb zwischen Apotheken zu niedrigeren Kosten für nicht-rezeptpflichtige Arzneimittel? Idealerweise wüsste man, welches nicht-rezeptpflichtige Arzneimittel in welcher Apotheke wie viel kostet.
- Werden Patientinnen und Patienten in Krankenhäusern mit der – gemessen an der medizinisch angezeigten Behandlung – adäquaten technischen und personellen Ausstattung behandelt? Idealerweise könnte man die Informationen zur Ausstattung von Krankenhäusern aus den Qualitätsberichten mit den Informationen zur Krankenhausbehandlung aus der DRG Statistik verknüpfen und den langfristigen Gesundheitszustand in ambulanten Daten nachverfolgen.
- Sind Richtlinien zur Risikoschwangerschaftsvorsorge und zur Versorgung von Neugeborenen effektiv? Welchen Einfluss hat die Kreißaalauslastung auf Geburtsindikatoren und Gesundheit von Neugeborenen und Müttern? Um diese Fragen zu untersuchen, müsste man detaillierte Informationen über die werdenden Mütter (z.B. aus dem Mutterpass) mit Informationen über die Neugeborenen bzw. Mütter nach der Geburt oder zur Auslastung der Kreißsäle mit Informationen zu den erfolgten Geburten verknüpfen können.
- Sollten sozioökonomische Merkmale – wie in anderen Ländern – im Morbi-RSA berücksichtigt werden? Um diese Frage zu untersuchen, müssten Daten mit individuellen sozioökonomischen Merkmalen auf individueller Ebene mit Abrechnungsdaten aus der GKV verknüpft werden.
- Wie und warum hängen sozioökonomische und gesundheitliche Ungleichheit zusammen und welche Rolle spielt dabei das Gesundheitssystem? Dazu müssten wiederum individuelle Informationen zum sozioökonomischen Status, zum Gesundheitszustand und zur Inanspruchnahme von Gesundheitsleistungen verknüpft werden.
- Fragen zum Einfluss des Schulsystems auf die Gesundheit von Kindern können derzeit in Deutschland nur mit lückenhaften Daten bearbeitet werden, da bspw. Abrechnungsdaten aus der GKV keine Informationen zur besuchten Klassenstufe oder Schulform von Kindern und Jugendlichen enthalten.

Die genannten Beispiele zeigen, dass die Verknüpfung von Daten aus der Gesundheitsversorgung essentiell ist, um Erkenntnisse über institutionelle, organisatorische und regulatorische Bestimmungsfaktoren der Qualität der Gesundheitsversorgung sowie deren gesellschaftliche Implikationen zu gewinnen. Bei den hierfür notwendigen Analysen stehen häufig die Akteure im Mittelpunkt, die unter gegebenen Rahmenbedingungen eine Vielzahl gesundheitsrelevanter Entscheidungen treffen. Dies können zum Beispiel Ärztinnen und Ärzte, Patientinnen und Patienten, Pflegepersonal, Führungspersonen in Krankenhäusern und bei Gesundheitsdienstleistern, aber auch Verwaltungseinheiten auf kommunaler oder regionaler Ebene sein.

Möchte man beispielsweise die Qualität der Versorgung im allgemeinärztlichen Bereich dokumentieren und verstehen, ist es essentiell, individuelle Merkmale jeder Patientin und jedes Patienten sowie deren Interaktionen mit den Leistungserbringern berücksichtigen zu können, und dies für eine möglichst große Stichprobe. Da Interaktionen in komplexen Netzwerken in der Gesamtgesellschaft stattfinden, ist für eine empirische Messung kausaler Zusammenhänge, wie zum Beispiel die Relevanz der Behandlungsqualität für die Gesundheit von Patientinnen und Patienten, oft sogar die Beobachtung der Grundgesamtheit nötig. Aufgrund der potenziell sehr langfristigen Auswirkungen der Gesundheitsversorgung auf die individuelle Gesundheit, müssen zusätzlich Individualdaten über viele Jahre verknüpft und analysiert werden können. Einige internationale Studien haben beispielsweise die Auswirkung von gesundheitlichen Risikofaktoren im Kindesalter auf den späteren Bildungserfolg oder den Erfolg im Arbeitsmarkt, aber auch die psychische Gesundheit untersucht. Damit solche, politisch äußerst relevanten Untersuchungen möglich sind, müssen individuelle Personen und andere Entscheidungseinheiten in pseudonymisierter Form langfristig, in sicheren und geschützten Daten, beobachtbar sein.

4. Deutschland verliert gegenüber Nachbarländern an Boden

Die Auswertung verknüpfbarer Administrativdaten hat international in den letzten zwei Jahrzehnten zahlreiche hochwertige Forschungsbeiträge hervorgebracht. Tatsächlich kommt die beste wissenschaftliche, politikrelevante Evidenz fast immer aus Ländern, die bei der Bereitstellung und Verknüpfung von Administrativdaten vorangehen. Hierzu gehören von Beginn an die skandinavischen Länder. In der Folge kann dort auch eine stärker evidenzbasierte Politik durchgeführt werden. Nagaraj und Tranchero zeigen empirisch für die USA, wie wichtig die Bereitstellung von verknüpften Daten für die Forschung und deren Rezeption in der Politik ist.¹⁰ So können verknüpfte Administrativdaten ein transparentes Bild des Zustands der Gesundheitsversorgung ermöglichen und erlauben der Wissenschaft, die Rolle von Anreizen und Restriktionen sowie wirksame politische Hebel zur Verbesserung der

¹⁰ Nagaraj, Abhishek und Matteo Tranchero (2023), "How does data access shape Science? Evidence from the impact of US Census's research data centers on economics research", National Bureau of Economic Research Working Paper No. w31372.

Gesundheitsversorgung zu identifizieren. Die Anzahl der Länder, in denen Forschenden der Zugang zu Administrativdatensätzen, sowie die Möglichkeit der Verknüpfung untereinander, ermöglicht wird, erhöht sich stetig. Unter diesen Ländern befinden sich in jüngerer Zeit auch Nachbarländer Deutschlands wie die Niederlande und Österreich. In dem Maße, in dem die Nachbarländer beim Datenzugang vorangehen, verschlechtert sich die relative Position von Forscherinnen und Forschern in Deutschland weiter.

Oben wurde gezeigt, dass die überwiegende Mehrheit der befragten Gesundheitsökonominnen in Deutschland glaubt, aufgrund der Probleme und Hindernisse beim Datenzugang im internationalen Wettbewerb Nachteile zu haben und daher teils mit ausländischen Daten forscht. Im Anhang listen wir die wesentlichen organisatorischen, technischen und rechtlichen Rahmenbedingungen der Praxis des Datenzugangs in den Niederlanden, Österreich und Dänemark auf, um zu zeigen: die Forderungen nach einem verbesserten Datenzugang einschließlich der Möglichkeit der Verknüpfung sind realistisch. Wenn sie in unseren Nachbarländern mit berechtigten Belangen des Datenschutzes vereinbar sind, dann können sie es bei uns auch sein.

4.1 Beispiele für Forschung mit verknüpften Mikrodaten

Verknüpfte Mikrodaten haben über die wirtschaftswissenschaftlichen Fachbereiche hinweg eine Vielzahl wegweisender Studien ermöglicht, zum Beispiel zur Wirkung von Politikmaßnahmen auf finanzielle Vorsorgeentscheidungen, zu den Ursachen für steigende Lohnungleichheit, zu möglichem Ausweichverhalten nach Einführung einer Vermögenssteuer, zur optimalen Ausgestaltung der Arbeitslosenversicherung oder zu der Rolle von Kindern bzgl. Geschlechterunterschieden auf dem Arbeitsmarkt. Auch in der Gesundheitsökonomik wurden bereits zahlreiche Studien anhand von verknüpften Individualdaten mit wichtigen Erkenntnissen veröffentlicht.

Zum Beispiel haben Handel et al. anhand von niederländischen Daten untersucht, welchen Einfluss sozioökonomische Faktoren auf die Wahl von Krankenversicherungsleistungen haben.¹¹ Diese und ähnliche Fragen sind essentiell für das Design von Krankenversicherungen. Hierfür sind Daten von vereinzelt Krankenversicherungen in der Regel nicht ausreichend, sondern es müssen die individuellen Entscheider in der Gesamtbevölkerung beobachtet werden können.

Viele Beispiele für wichtige Forschungsarbeiten mit verknüpften Mikrodaten stammen aus Dänemark. Dazu gehören kausale Studien

- über die Auswirkungen einer ADHS-Diagnose auf die Gesundheit und beobachtete mutmaßliche Kriminalität von Jugendlichen,¹²

¹¹ Handel, Benjamin R., Jonathan T. Kolstad, Thomas Minten und Johannes Spinnewijn (2020), "The social determinants of choice quality: evidence from health insurance in the Netherlands", National Bureau of Economic Research Working Paper No. w27785.

¹² Dalsgaard, Søren, Helena Skyt Nielsen und Marianne Simonsen (2014), "Consequences of ADHD medication use for children's outcomes", *Journal of Health Economics* 37:137-151.

- zu den langfristigen Auswirkungen von infektiösen Atemwegskrankheiten auf Haushaltsmitglieder,¹³
- zum Effekt der Verhütungspille auf die psychische Gesundheit,¹⁴
- über den Nutzen von frühzeitiger medizinischer Versorgung von Neugeborenen für sie selbst sowie ihre Familienmitglieder,¹⁵
- zu den Auswirkungen von schweren Krankheiten auf das Verhalten und die Gesundheit von Haushaltsmitgliedern,¹⁶
- zur Relevanz von ärztlicher Kontinuität in der Primärversorgung¹⁷ sowie
- zu den starken systematischen Unterschieden in der Behandlungsqualität in einem vermeintlich weitgehend harmonisierten und standardisierten Gesundheitssystem.¹⁸

Aber auch Erkenntnisse über das Potential von Daten für Anwendungen von künstlicher Intelligenz können durch die Verknüpfung von Administrativdaten mit medizinischen Daten gewonnen werden.¹⁹

4.2. Einfluss auf die Politikgestaltung

In deutlicher Erinnerung bleibt die SARS-Cov-2 Pandemie, in der Politik und Öffentlichkeit in Israel und den skandinavischen Ländern regelmäßige disaggregierte Berichte zu Infektionszahlen, zum Beispiel nach Berufsgruppen, Alter und sozioökonomischen Eigenschaften, zur Hand hatten. So konnten Maßnahmen zielgerichteter und zeitnaher umgesetzt werden, sowie Test- und Behandlungskapazitäten angepasst werden. Beispielsweise wurde die Kapazitätsauslastung der Labore in Deutschland zunächst gar nicht, später nur auf Bundeslandebene durch den Interessenverband der akkreditierten medizinischen Labore erhoben und veröffentlicht. Solche Informationen auf lokaler Ebene wären für die Ermittlung lokaler Messfehler der Covid-Inzidenz sowie zur Bewertung der Qualität und Schnelligkeit der möglichen Kontaktnachverfolgung und zur Ableitung politischer Handlungsempfehlungen zu den Testkapazitäten notwendig gewesen. Im Vergleich hierzu konnten in Dänemark die individuellen Covid-Tests (PCR-Tests, sowie Schnelltests aus Testzentren) auf individueller

¹³ Daysal, N. Meltem, Hui Ding, Maya Rossin-Slater und Hannes Schwandt (2021), "Germs in the family: The long-term consequences of intra-household endemic respiratory disease spread", National Bureau of Economic Research Working Paper No. w29524.

¹⁴ Costa-Ramon, Ana, Meltem Daysal und Ana Rodríguez-González (2023), "The Oral Contraceptive Pill and Adolescents' Mental Health", CEPR Discussion Paper No. 18269.

¹⁵ Daysal, N. Meltem, Marianne Simonsen, Mircea Trandafir und Sanni Breining (2022), "Spillover effects of early-life medical interventions." *Review of Economics and Statistics* 104, 1:1-16.

¹⁶ Fadlon, Itzik und Torben Heien Nielsen (2019), "Family health behaviors." *American Economic Review* 109(9): 3162-3191. Fadlon, Itzik und Torben Heien Nielsen (2021), "Family labor supply responses to severe health shocks: Evidence from Danish administrative records." *American Economic Journal: Applied Economics* 13(3): 1-30.

¹⁷ Simonsen, Marianne, Lars Skipper, Niels Skipper und Peter Rønø Thingholm (2021), "Discontinuity in care: Practice closures among primary care providers and patient health care utilization." *Journal of Health Economics* 80:102551.

¹⁸ Huang, Shan und Hannes Ullrich (2023), "Provider effects in antibiotic prescribing: Evidence from physician exits", BSE Discussion Paper No. 18.

¹⁹ Huang, Shan, Michael Allan Ribers und Hannes Ullrich (2022), "Assessing the value of data for prediction policies: The case of antibiotic prescribing", *Economics Letters* 213:110360.

Personenebene verknüpft werden, zum Beispiel mit den im selben Haushalt lebenden Personen. In Dänemark konnte man so bereits im Sommer 2020 die Übertragungsrisiken in Haushalten quantifizieren und die Rolle von Alter und individueller Virenlast für Übertragungsrisiken in Haushalten beschreiben. Im weiteren Verlauf der Pandemie konnte auf diese Erkenntnisse aufgebaut werden.²⁰

In Schweden konnten Fragen zur Effektivität von Schulschließungen durch eine Auswertung von Gesundheitsdaten mit Haushaltsregistern und Schulregister auf Individualebene nahezu in Echtzeit durchgeführt werden. So konnte bereits im Sommer 2020 quasi-experimentelle Evidenz dafür geliefert werden, dass Schulbesuch von Kindern kein stark erhöhtes Risiko in Bezug auf eine Covid Erkrankung für die Erwachsenen in ihrem Haushalt darstellen.²¹ Des Weiteren wurden die ergriffenen Maßnahmen parallel zur laufenden Pandemie in mehreren detaillierten Regierungsberichten (SOU 2020:80, SOU 2021:89, SOU 2022:10) begleitend kritisch evaluiert. Die meisten Aspekte in dieser Evaluierung wären in Deutschland gar nicht zu untersuchen gewesen, da entsprechende Daten überhaupt nicht, nur unzureichend oder mit großem zeitlichen Verzug erfasst worden sind.

Das Verständnis der Wichtigkeit verknüpfter Administrativdaten hat in Dänemark bereits früh zu attraktiven Schnittstellen zwischen Wissenschaft und Politik geführt. Im Jahr 2001 wurde zum Beispiel an der Universität Kopenhagen das Economic Policy Research Network ins Leben gerufen, um den Austausch zwischen Wissenschaft und Politik mit starker empirischer Fundierung zu beleben.²² In diesem Netzwerk finden unter anderem ganztägige Veranstaltungen statt, auf welchen sich Ministerien, die basierend auf verknüpften Administrativdaten auch eigene Analysen erstellen, mit Forscherinnen und Forschern der Universität Kopenhagen, die ebenso ihre empirischen Arbeiten vorstellen, austauschen. Das weit verbreitete empirische Arbeiten mit verknüpften Administrativdaten erlaubt schlussendlich eine stärkere Durchlässigkeit zwischen Wissenschaft und Politik. Diese spiegelt sich unter anderem in Weiterbildungsangeboten von Universitäten für Behörden, Ministerien und Praktiker, in Forschungs Kooperationen und auch in gemeinsamen Stellenbesetzungen sowie in fachlichen ad hoc Beratungen wider.

²⁰ Siehe z.B. Lyngse, Frederik Plesner et al. (2022), "Nationwide study on SARS-CoV-2 transmission within households from lockdown to reopening, Denmark, 27 February 2020 to 1 August 2020", *Eurosurveillance* 27(6): 2001800 und Lyngse, Frederik Plesner et al. (2022), "Household transmission of SARS-CoV-2 Omicron variant of concern subvariants BA. 1 and BA. 2 in Denmark", *Nature Communications* 13(1): 5760.

²¹ Vlachos, Jonas, Edvin Hertegård und Helena B. Svaleryd (2021), "The effects of school closures on SARS-CoV-2 among parents and teachers", *Proceedings of the National Academy of Sciences*, 118(9): e2020834118.

²² Siehe <https://www.economics.ku.dk/research/ResearchCentres/epru/eprn/about-eprn/>

5. Maßnahmen zur Verbesserung des Zugangs zu Gesundheitsdaten in Deutschland

Die vorangehenden Abschnitte haben die Defizite im Datenzugang bzgl. Gesundheitsdaten und deren negative Auswirkungen auf Forschung und Politikberatung beschrieben. Zur Verbesserung der Voraussetzungen für international konkurrenzfähige gesundheitsökonomische Forschung in Deutschland und für Deutschland schlagen wir daher folgende Schritte und Maßnahmen vor:

1. Alle administrativen Datenquellen, d.h. Daten aus der öffentlichen Verwaltung, aus der Sozialversicherung, Daten der amtlichen Statistik, aber auch Daten aus Forschungsprojekten, die aus öffentlichen Quellen finanziert wurden, sollten für die weitere Forschung zugänglich gemacht werden. Wichtig ist, dass dies nicht in aggregierter Form, sondern in Form von Individualdaten erfolgt.
2. Datenquellen sollten untereinander verknüpft werden können. Explizit ist hier auch die Verknüpfung von Gesundheitsdaten mit anderen Daten, z. B. Steuerdaten, gemeint. Dazu muss die Zusammenarbeit zwischen der geplanten Datenzugangs- und Koordinierungsstelle für Gesundheitsdaten beim BfArM mit der entsprechenden Koordinierungsstelle anderer Forschungsdaten, bspw. dem Dateninstitut des Bundes, gesetzlich geregelt werden. Für die Zusammenführung von Daten ist aber in einem ersten Schritt eine einheitliche Identifikationsnummer für Personen erforderlich. Mögliche Konflikte zwischen den schutzwürdigen Interessen des Einzelnen und den Interessen der Forschung bei der Zusammenführung von Daten sollten durch eine Widerspruchsregelung gelöst werden.
3. Um eine übersichtliche Dateninfrastruktur zu schaffen, ist eine öffentlich finanzierte Plattform mit Datenübersicht und Datenmittler sinnvoll (angesiedelt etwa beim Dateninstitut des Bundes). Die Verknüpfung von Daten aus verschiedenen Quellen kann der Datenmittler als Datentreuhänder ermöglichen, d.h. der Datenmittler verwaltet die Zuordnungsschlüssel, mit denen Nutzer die Verknüpfung unterschiedlicher pseudonymisierter Datensätze vornehmen können. In diesem Kontext ist auch die Nationale Forschungsdateninfrastruktur (NFDI) als Plattform denkbar, die jedoch unterschiedliche Konsortien für personenbezogene Gesundheitsdaten (nfdi4health) und für die Sozial- und Wirtschaftswissenschaften (KonsortSWD) unterhält.
4. Für den Zugang für Forschungszwecke zu administrativen und allen aus öffentlichen Mitteln finanzierten Daten sollen transparente und einheitliche Regelungen geschaffen werden. Wer unter welchen Voraussetzungen Zugang zu welchen Daten bekommt, sollte so offen wie möglich und vor allem diskriminierungsfrei entschieden werden. Eine ganz wesentliche Rolle kommt hier dem Antragsverfahren zu. Datenzugang sollte auf Grundlage formeller, nicht materieller Kriterien gewährt werden, d.h. nicht in Abhängigkeit von einer den Datenhaltern/-eigentümern genehmen Fragestellung,

sondern danach, ob Nutzerinnen und Nutzer die jeweiligen schutzwürdigen Interessen der Datensubjekte wahren. Auch sollte sichergestellt werden, dass ein Verwendungszweck die Replikation bestehender Analysen sein kann. Das bedeutet: ein Nutzungsantrag sollte nicht schon alleine deshalb abgelehnt werden dürfen, weil jemand anders einen sehr ähnlichen Antrag in der Vergangenheit bewilligt bekommen hat. Replizierbarkeit von Studien verbessert die Qualität von Forschungsergebnissen und trägt dadurch zum Erkenntnisgewinn und zur Steigerung der Glaubwürdigkeit der Wissenschaft bei.

5. Ein Datennutzungsplan soll für jede öffentliche Institution die erhobenen Daten auflisten und die Antrags- und Zugangsverfahren für die entsprechenden Daten beschreiben. Der Datennutzungsplan wird jährlich erneuert und zentral, z.B. vom Dateninstitut des Bundes veröffentlicht.
6. Ein wesentliches rechtliches Hindernis beim Datenzugang stellt die föderale Struktur des Datenschutzes dar: jedes Bundesland hat ein eigenes Datenschutzgesetz, so dass z. B. Krebsregister in verschiedenen Bundesländern Merkmale uneinheitlich bereitstellen. Der Datenschutz in Bezug auf Zugang zu administrativen Mikrodaten muss daher bundeseinheitlich geregelt werden. Es darf nicht darauf ankommen, in welchem Bundesland der Nutzer wohnt oder arbeitet. Insbesondere müssen die in der europäischen Datenschutz-Grundverordnung festgelegten Ausnahmen für die Datenverarbeitung im öffentlichen Interesse und der Wissenschaft in der Praxis des Antragsverfahrens und Datenzugangs verankert sein, so dass rechtliche Bedenken und Risiken in diesem Schritt kein Hinderungsgrund sein können.
7. Bei Gesetzen, die Datenerhebungen umfassen (z.B. im Rahmen von Evaluierungen der Auswirkungen des Gesetzes, zur Gewährung von Fördergeldern, etc.) sollte festgelegt werden, dass und wo diese Daten der Forschung zugänglich gemacht werden, wobei zusätzlich die Kompatibilität mit dem geplanten Europäischen Gesundheitsdatenraum berücksichtigt werden sollte.
8. Selbstverständlich darf Datenzugang etwas kosten. Es muss jedoch darauf geachtet werden, dass Gebühren oder Beiträge nicht so hoch sind, dass bestimmte Forschergruppen (Nachwuchs) systematisch ausgeschlossen werden. Vergünstigte Tarife für Nachwuchswissenschaftler (wie im aktuellen FDZ System) oder spezielle auf die Finanzierung des Datenzugang ausgerichtete Förderlinien der DFG sind als Lösungen denkbar. International üblich sind Gebühren oder Beiträge im Bereich von 5.000 bis 10.000 Euro für ein typisches Forschungsprojekt. Das bisherige Gebührenmodell der FDZ des Statistischen Bundesamtes, in dem ein fester Betrag pro Datensatz und -jahr gezahlt werden muss, sollte nicht übernommen werden, da es nur in einem sehr schwachen Verhältnis zum Ressourcenaufwand bei den statistischen Ämtern steht.

6. Fazit

Im internationalen Vergleich leidet die empirische gesundheitsökonomische Forschung in Deutschland, insbesondere bezüglich der Beantwortung drängender Forschungsfragen rund um das nationale Gesundheitssystem, unter schlechter Datenverfügbarkeit und -verknüpfbarkeit. Dies hat Auswirkungen auf die Qualität der möglichen Analysen und Folgen für die Themen- und Standortwahl angewandter empirischer Forscherinnen und Forscher mit dem Schwerpunkt Gesundheit. Durch Vergleiche mit den nächsten Nachbarn im Ausland wird das vergebene wissenschaftliche Potential für Grundlagenforschung und angewandte Politikberatung deutlich. Dabei gibt es offensichtlich gangbare Wege für eine bessere Verfügbarkeit und Verknüpfung von Forschungsdaten im Gesundheitsbereich, im Einklang mit den europäischen Datenschutzbestimmungen. Eine Umsetzung der hier abgeleiteten, konkreten Handlungsempfehlungen sollte aus unserer Sicht einen wichtigen Beitrag zur Verbesserung der Forschung im Bereich der empirischen Gesundheitsforschung leisten.

Anhang

Tabelle 1: Austrian Micro Data Center²³

Zugangserlaubnis	Abgestuft nach gewähltem technischen Zugang und Re-Identifizierbarkeit der Individuen oder Firmen. Grundsätzlich können alle wissenschaftlich arbeitenden Personen im In- und Ausland Zugang bekommen. Remote Access auf Basis einer entsprechenden vertraglichen Vereinbarung nur für Wissenschaftlerinnen und Wissenschaftler, die an akkreditierten wissenschaftlichen Einrichtungen tätig sind (Universitäten, außeruniversitäre Forschungseinrichtungen). Akkreditierung ist auch für ausländische Forschungseinrichtungen möglich. Forscher unterzeichnen eine Verpflichtungserklärung zum Datenschutz mit Hinweis auf straf- und zivilrechtliche Konsequenzen einer Verletzung
Antragstellung	Ausführlich beschriebenes Forschungsvorhaben, mit Angaben zu Methoden der Analyse, angestrebte, Ergebnissen, spezifizierten Datenkategorien, zu denen Zugang benötigt wird, in räumlicher und zeitlicher Hinsicht spezifizierte Art der Statistikdaten, die für das Forschungsvorhaben unbedingt benötigt werden (Grundsatz der Datenminimierung).
Externe Daten	Selbst eingebrachte Daten können gemäß rechtlichen Rahmenbedingungen über Pseudonyme bzw. verschlüsselte Unternehmenskennzahlen mit den Mikrodaten des AMDC verknüpft werden.
Technischer Zugang	Gesicherter Onlinezugang (Remote Access) zu potentiell indirekt identifizierbaren vertraulichen Daten für Wissenschaftlerinnen und Wissenschaftler in akkreditierten Einrichtungen. Daneben gibt es noch faktisch anonyme Daten an wissenschaftlich arbeitende Personen in Form von Scientific Use Files und Vor-Ort-Zugang zu Mikrodaten im Safe Center von Statistik Austria durch alle fachlich geeigneten Personen.
Kosten	Für Onlinezugang pauschal ca. 3.000 Euro Grundgebühr plus ca. 200 Euro pro virtuellem Arbeitsplatz. Nutzung eines virtuellen Arbeitsplatzes ca. 100 Euro pro Monat. Stundensatz von ca. 100 Euro für Service und Beratungsleistungen (z. B. Einbringen eigener Daten). In Summe schätzungsweise 5.000 bis 10.000 Euro pro Projekt.
Datenschutz	Technisch – Verpflichtung zur Einrichtung einer geschützten Zugangsstelle; zwei Faktor- Authentifizierung für den Zugang. Den Forschenden ist verboten, gleichzeitig das Internet zu nutzen, zu versuchen, statistische Einheiten zu re-identifizieren, Bildschirmkopien zu erstellen, usw. Automatisierte Kontrolle der Wahrung der Geheimhaltung seitens Statistik Austria durch geeignete Routinen.

²³ Siehe Statistik Austria: <https://www.statistik.at/services/tools/services/center-wissenschaft/austrian-micro-data-center-amdc>

Tabelle 2: Dänische²⁴ oder andere skandinavische Register²⁵

Zugangserlaubnis	Für Wissenschaftlerinnen und Wissenschaftler in autorisierten Forschungsumgebungen (z.B. Universitäten, Forschungseinrichtungen, Ministerien) in Dänemark, die eine Datennutzungsvereinbarung mit Statistics Denmark abgeschlossen haben. Koautorinnen und Koautoren aus dem Ausland können ebenso einen Zugang für gemeinsame Forschungsprojekte erhalten. In jedem Fall müssen alle Wissenschaftlerinnen und Wissenschaftler, die einen Zugang erhalten möchten, einen Vertrag mit Statistics Denmark unterschreiben, der auch die deutlichen rechtlichen Konsequenzen bei Fehlverhalten beschreibt.
Antragstellung	Projektbeschreibung mit Begründung des gesellschaftlichen Interesses, inklusive Liste der gewünschten Daten und Variablen, sowie die Variablen, anhand welcher Datensätze (zum Beispiel die Personenidentifikationsnummer) verknüpft werden sollen. Bei der Verwendung von individuellen Personendaten ist ebenso eine Erlaubnis bei der eigenen Forschungseinrichtung einzuholen. Für Gesundheitsdaten ist in der Regel darüber hinaus ein weiterer Antrag bei der Gesundheitsdatenbehörde ²⁶ nötig.
Externe Daten	Sollen externe Daten zugespielt werden, die auch in der Projektbeschreibung beschrieben und erklärt werden müssen, ist die Erlaubnis der externen Datenbesitzer nötig.
Technischer Zugang	Die Daten können per VPN vom eigenen Rechner aus bearbeitet werden.
Kosten	Die Bereitstellungskosten ergeben sich aus dem Aufwand für Statistics Denmark, der auf Stundenbasis vergütet werden muss. Für ein gängiges Projekt fallen ca. 5000-10000 Euro an Bereitstellungskosten an. Für rechenintensive Anwendungen können Forschungseinrichtungen und Wissenschaftlerinnen und Wissenschaftler darüber hinaus in dedizierte Hardware bei Statistics Denmark investieren.
Datenschutz	Autorisierte Wissenschaftlerinnen und Wissenschaftler verpflichten sich vertraglich, in den de-identifizierten aber nicht anonymisierten Daten keine Individuen zu identifizieren. Um Ergebnisse außerhalb des gesicherten Forschungsdatenzentrums verwenden zu können, müssen diese nach festgelegten Mindestkriterien anonymisiert werden und publikationsfertig exportiert werden. Neben einem automatisierten Filter, der verdächtige Dateien blockiert, werden exportierte Ergebnisse von Mitarbeiter*innen bei Statistics Denmark überprüft. Bei mutmaßlichen Verstößen werden innerhalb kurzer Zeit Konsequenzen umgesetzt. Grundsätzlich wird im Sinne der EU-DSGVO wird bei der Antragsstellung das Prinzip der Datenminimierung angewandt.

²⁴ Siehe Statistics Denmark: <https://www.dst.dk/en/TilSalg/Forskningsservice>

²⁵ Antragstellung und Zugangsvoraussetzungen in Schweden sind nahezu identisch. Offenkundig über schwedische wissenschaftliche Institutionen und mit etwas geringeren zu erwarteten Kosten (3000-6000 Euro). Siehe Statistics Sweden: <https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/>

²⁶ Siehe Danish Health Data Authority: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers

Tabelle 3: Niederlande²⁷

Zugangserlaubnis	Für Wissenschaftlerinnen und Wissenschaftler an autorisierten wissenschaftlichen Institutionen.
Antragstellung	Kurze Projektbeschreibung inklusive Liste der gewünschten Daten und Verknüpfungsvariablen. Die Bearbeitungszeit von Seiten Statistics Netherlands sind wenige Wochen. Das Hinzufügen zusätzlicher Daten oder Benutzer zu einem Projekt ist schnell und unbürokratisch möglich.
Externe Daten	Sollen externe Daten zugespielt werden, die auch in der Projektbeschreibung beschrieben und erklärt werden müssen, ist die Erlaubnis der externen Datenbesitzer nötig.
Technischer Zugang	Die Daten können per VPN vom eigenen Rechner aus bearbeitet werden, jedoch muss man sich hierfür in der EU oder in einem Land mit EU-Datenabkommen befinden.
Kosten	Die Kosten hängen von der Anzahl benutzter Datensätze und der Anzahl der Nutzer ab. Für ein typisches wirtschaftswissenschaftliches Forschungsprojekt fallen ca. 5000 Euro an Kosten an.
Datenschutz	Nutzer müssen versichern, keine Individuen zu identifizieren. Um Ergebnisse verwenden zu können, werden sie erst von CBS (Statistics Netherlands) geprüft und freigegeben, so dass aus den freigegebenen Daten keine Individualdaten identifiziert werden können. Nicht freigegebene Ergebnisse darf man in keiner Weise verwenden. Vor dem Einloggen gibt es jeweils Kontrollfragen zum Datenschutz. Innerhalb des Datenzugangs kann das CBS den Computergebrauch der Nutzer komplett überwachen, aber das geschieht im Hintergrund.

²⁷ Siehe Statistics Netherlands: <https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research>

Impressum

Kontakt:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Geschäftsstelle

Am Friedrichshain 22

10407 Berlin

office@ratswd.de

<https://www.ratswd.de>

Die Geschäftsstelle des RatSWD wird als Teil von KonsortSWD im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert - Projektnummer: 442494171.

Berlin, Dezember 2023



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

DOI: [10.17620/02671.88](https://doi.org/10.17620/02671.88)

Zitationsvorschlag:

Fischer, M., Jürges, H., Mangelsdorf, S., Reif, S., Ullrich, H. & Wuppermann, A. (2023). *Für einen besseren Datenzugang im Bereich Gesundheit. Stellungnahme aus dem Verein für Socialpolitik.* (RatSWD Working Paper 284/2023). Berlin. Rat für Sozial- und Wirtschaftsdaten (RatSWD).

<https://doi.org/10.17620/02671.88>