

Gutachten im Auftrag des Bundesministeriums für Gesundheit (BMG)

Record Linkage der Daten des FDZ Gesundheit mit extern gehaltenen Forschungsdaten anhand des Beispiels einer Verknüpfung mit Daten einer Kohorte des NAPKON

Version 2.1
vom 23. Mai 2023

- Zusammenfassung -

Gutachter

Prof. Dr. med. **Jochen Schmitt**, MPH ¹ (Koordination des Gutachtens)

Dr. med. **Peter Ihle** ²

Prof. Dr. med. **Wolfgang Hoffmann**, MPH ³

Prof. Dr. rer. nat. **Jens-Peter Reese**, MPH ⁴

Dr. rer. nat. **Olaf Schoffer**, Dipl.-Stat. ¹

¹ Professur für Sozialmedizin und Versorgungsforschung, Zentrum für Evidenzbasierte Gesundheitsversorgung (ZEGV), Medizinische Fakultät Carl Gustav Carus, TU Dresden sowie Universitätsklinikum Carl Gustav Carus Dresden

² PMV forschungsgruppe, Klinik und Poliklinik für Psychiatrie, Psychosomatik und Psychotherapie des Kindes- und Jugendalters an der Medizinischen Fakultät und Uniklinik Köln, Universität zu Köln

³ Abteilung Versorgungsepidemiologie und Community Health, Institut für Community Medicine, Universitätsmedizin Greifswald

⁴ Professur für Versorgungsforschung und Public Health, Institut für Klinische Epidemiologie und Biometrie, Julius-Maximilians-Universität Würzburg

Zusammenfassung

Versorgungsnahe Daten (VeDa) sind die wesentliche Ressource für die wissenschaftlich fundierte Beschreibung, Evaluation und Weiterentwicklung der Gesundheitsversorgung und des Gesundheitssystems in Deutschland. Eine ganz wesentliche Quelle von VeDa sind Abrechnungsdaten der Gesetzlichen Krankenversicherung (GKV-Routinedaten). Das Forschungsdatenzentrum Gesundheit (FDZ Gesundheit) am Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) hat für die Gesundheitsforschung mit VeDa eine überragende Bedeutung, weil es Längsschnittdaten von allen mehr als 70 Millionen GKV-Versicherten in Deutschland sektorübergreifend zur ambulanten und stationären medizinischen Versorgung, sowie zur Arzneimittelversorgung verfügbar machen wird. Trotz der hohen Repräsentativität und Vollständigkeit der dokumentierten Diagnosen, Prozeduren und Arzneimittelverordnungen fehlen jedoch in den GKV-Routinedaten häufig wichtige zusätzliche Informationen zur Beantwortung von wichtigen Forschungsfragestellungen, etwa zu Patienten-berichteten Ergebnissen, zu soziodemografischen Faktoren, zur Diagnostik und Bildgebung, zu Laborergebnissen und Biomarkern sowie zu Patientenpräferenzen. Für die qualitativ hochwertige wissenschaftliche Beantwortung von versorgungsrelevanten Fragestellungen müssen die GKV-Routinedaten daher häufig mit anderen VeDa (Registerdaten, Klinikdaten) oder/und mit Befragungsdaten und Daten aus Kohortenstudien individuell verknüpft werden (Data Linkage).

Das vorliegende Gutachten analysiert zunächst den derzeitigen Status quo der Umsetzung von Studien mit Linkage von GKV-Routinedaten mit Forschungsdaten basierend auf der informierten Einwilligung von Studienteilnehmenden. Dies wird exemplarisch am Beispiel der NAPKON-Kohortenstudie des Netzwerks Universitätsmedizin (NUM) betrachtet, welche zur Untersuchung von Verläufen einer Coronavirus-Erkrankung (COVID-19) und von Risikofaktoren für eine Long-COVID Erkrankung primäre Forschungsdaten mit GKV-Routinedaten verknüpfen möchte. Unter Einbeziehung der Perspektive von Patientenvertretenden, Krankenkassenvertretenden, Expertinnen und Experten für Ethik, Forschenden sowie von Klinikerinnen und Klinikern erarbeitet das Gutachten darauf aufbauend eine Modelllösung für die Verknüpfung der im FDZ Gesundheit zukünftig vorgehaltenen GKV-Routinedaten mit externen Forschungsdaten und geht insbesondere auf technisch-organisatorische Voraussetzungen ein. Ausgehend von dem Fallbeispiel NAPKON, eigenen Erfahrungen der Gutachter und den Ergebnissen der Expertenbefragungen zeigt das Gutachten weiteren gesetzlichen Regelungsbedarf zur Forschung mit GKV-Routinedaten auf und gibt entsprechende Empfehlungen für die Berücksichtigung in dem von der Bundesregierung im Koalitionsvertrag angekündigten Gesundheitsdatennutzungsgesetz.

Die gegenwärtigen Rahmenbedingungen für die wissenschaftliche Nutzung von Behandlungsdaten in Deutschland sind unbefriedigend. Das große Potenzial, die Qualität,

Sicherheit und Effizienz der medizinischen Versorgung durch die Versorgungsforschung zu verbessern sowie geeignete Public-Health-Maßnahmen zu planen und umzusetzen wird bisher bei Weitem nicht genutzt. Es bestehen momentan erhebliche datenschutzrechtliche, regulatorische und organisatorische Hürden und in der Konsequenz eine fehlende Verlässlichkeit und Planbarkeit bzgl. der Möglichkeit der Verknüpfung von GKV-Routinedaten mit Forschungsdaten. In der derzeitigen Praxis dominieren in Deutschland bei Fragen zur Nutzung von Gesundheitsdaten für die wissenschaftliche Forschung die theoretischen Risiken die Diskussionen und datenschutzrechtlichen Entscheidungen. Die Chancen und der zu erwartende gesellschaftliche Nutzen der Forschung mit GKV-Routinedaten fallen dagegen kaum ins Gewicht.

Um ein angemessenes Abwägen von Nutz- und Schadenspotenzialen bei der regulatorischen Bewertung von Forschungsvorhaben zu ermöglichen, sollten klar abgrenzbare Kriterien sowohl für Datenschutzrisiken, als auch für Nutzenchancen entwickelt werden, die dann transparent bei der Abwägung zur Bewilligung oder Ablehnung von Forschungsvorhaben heranzuziehen sind. Patientenvertreter befürworten die wissenschaftliche Nutzung von GKV-Routinedaten, Krankenkassen berichten, dass ihre Versicherten die Daten „in der Regel bereitwillig für die Forschung teilen“ und die klinischen Fachvertreter erwarten von der wissenschaftlichen Nutzung der Daten einen unmittelbaren Nutzen für die Patientenversorgung und die medizinische Qualität. Obwohl es in Deutschland keine bekannten Fälle von Missbrauch der Daten im Forschungskontext gibt, macht der Datenschutz viele wichtige Projekte schwierig umsetzbar, wenn nicht gar unmöglich. Bzgl. Datenschutz ist es wesentlich, zwischen theoretischer Separierbarkeit und praktischer Identifizierbarkeit zu unterscheiden. Wenn eine Variablenkombination selten vorkommt, verursacht dies nicht automatisch eine Identifizierbarkeit. Die Identifizierung einer bestimmten Person in einem pseudonymisierten Datensatz erfordert einen hohen Aufwand und in aller Regel kriminelle Energie. Dieser Sachverhalt muss unbedingt bei der Datenschutzbewertung berücksichtigt werden. Angemessene Regelungen zur Vermeidung und Ahndung von Scientific Misconduct existieren seit Jahren und sind konsequent anzuwenden. In der gesundheitswissenschaftlichen Forschungs-Community existiert eine hohe Datenschutzorientierung. Neben den klaren Sanktionsmöglichkeiten ist es vor allem diese ausgeprägte Kultur der Selbstkontrolle, die Missbrauchsrisiken minimiert hat.

Auch bei informierter Einwilligung von Studienteilnehmende in die Verknüpfung der über sie erhobenen Forschungsdaten mit ihren bei ihrer Krankenkasse gespeicherten GKV-Routinedaten ist es im Beispiel der NAPKON-Kohortenstudie bisher nicht gelungen, GKV-Daten tatsächlich für die wissenschaftliche Nutzung zu erhalten. Gründe sind redundante, nicht standardisierte, restriktive Prüfungen des Datenschutzes durch unterschiedliche Institutionen (Datenschützer der Einrichtungen, Ethikkommissionen, Aufsichtsbehörden der Krankenkassen) mit häufig widersprüchlichen Ergebnissen sowie eine fehlende Verpflichtung

der Krankenkassen, dem Willen ihrer Versicherten zur Verfügbarmachung der GKV-Routinedaten zu entsprechen.

Aktuell können Forschungsprojekte mit Linkage von sektorenübergreifenden GKV-Routinedaten nur mit aktiver Unterstützung durch Krankenkassen durchgeführt werden. Größere Krankenkassen erhalten sehr viele Anfragen zur Forschungsdatennutzung, denen sie häufig nicht entsprechen können oder nach eigener Priorisierung nicht entsprechen wollen. Es fehlen transparente Entscheidungsprozesse und kassenübergreifende Gremien zur Koordination der Forschungsanfragen. Die momentane Praxis birgt die Gefahr von Interessenkonflikten und resultierender selektiver Berichterstattung/Publication Bias. Krankenkassen und ihre wissenschaftlichen Institute haben vielfach eine hohe Expertise in der Versorgungs-, Qualitäts- und Patientensicherheitsforschung mit GKV-Routinedaten. Sie sind häufig nicht nur für die Datenbereitstellung zuständig, sondern auch für die Planung und Umsetzung von Forschungsprojekten. Dennoch und gerade deshalb sollte eine klarere Transparenz darüber hergestellt werden, wer was mit GKV-Routinedaten beforcht. Wirksame Maßnahmen wie prospektive Studienregistrierung sowie Berichts- und Qualitätsstandards sind aus der Klinischen Forschung bekannt und sollten für die Versorgungsforschung mit GKV-Routinedaten ebenfalls umgesetzt werden. Da auch nach erfolgreicher Implementierung des FDZ Gesundheit weiterhin mehrere Szenarien bestehen werden, bei denen die Krankenkassen für die Datenbereitstellung direkt beteiligt werden müssen (z. B. Interventionsstudien, Evaluation Neuer Versorgungsformen des Innovationsfonds, Versorgungsmanagement, Studien, die GKV-Daten benötigen, die momentan nicht für das FDZ Gesundheit vorgesehen sind), schlägt das Gutachten entsprechende Regelungen zur Erhöhung der Planbarkeit von Forschungsprojekten und der Transparenz von Entscheidungsprozessen für oder gegen die Unterstützung von Forschungsprojekten durch die Krankenkassen vor.

Bei Nutzung von GKV-Daten auf Basis einer informierten Einwilligung und Vorliegen eines zustimmenden Votums der zuständigen Ethikkommission sollte zukünftig kein Antrag nach § 75 SGB X notwendig sein. Hintergrund für diese Empfehlung ist, dass alle Voraussetzungen für die Sozialdatennutzung bereits durch die zuständige Ethikkommission geprüft werden und durch die Mehrfachprüfung kein Datenschutznutzen, aber Aufwände auf Forschungsseite produziert werden. Das FDZ Gesundheit hat das Potenzial, entscheidend zum Ausbau der sekundären Forschungsnutzung von Behandlungsdaten in Deutschland beizutragen. Dies ist ethisch und politisch geboten, um die Qualität, Sicherheit und Wirtschaftlichkeit der Gesundheitsversorgung evidenzgeleitet zu erhöhen, neue medizinische Anwendungen zu entwickeln und damit die Gesundheit und Lebensqualität der Bevölkerung zu verbessern. Um dieses Potenzial zu heben und den politischen Willen einer breiten und qualitativ hochwertigen wissenschaftlichen Nutzung von GKV-Routinedaten zu erfüllen, muss das FDZ Gesundheit für Wissenschaftlerinnen und Wissenschaftler praktikable Lösungen vorsehen. Dazu zählt neben der Verfügbarmachung möglichst aktueller Daten u. a. die Möglichkeit, per personalisiertem Remotezugang pseudonymisierte Einzeldatensätze vom FDZ Gesundheit

analysieren zu können. Um den wissenschaftlichen Anforderungen der Versorgungsforschung mit GKV-Routinedaten zu entsprechen, sollten Forschende nach Bewilligung von Nutzungsanträgen diesen Remotezugang zu einem möglichst breiten Datensatz erhalten. Dies ist notwendig, weil bei GKV-Routinedatenanalysen – anders als etwa bei Klinischen Studien – ohne genaue Kenntnis des Datenkörpers nicht alle Analyseschritte abschließend geplant werden können. Synthetische Daten oder „verrauschte“ Daten bzw. Anwendung von differential privacy oder verteiltes Rechnen sind für die Planung und Durchführung von GKV-Routinedatenanalysen nach aktuellem Kenntnisstand nicht geeignet.

Einige für die Gesundheitsversorgung äußerst relevante Fragestellungen können nur mit einem vollständigen, krankenkassenübergreifenden Datensatz aller Versicherter in der notwendigen wissenschaftlichen Qualität bearbeitet werden. Hierzu zählen beispielsweise Fragestellungen der Qualitätssicherung, der Patientensicherheitsforschung, der Arzneimittelsicherheit, des kleinräumigen Versorgungsmonitorings, zu seltenen Erkrankungen sowie zur Leitliniengerechtigkeit der Versorgung in spezifischen Patientengruppen. Die Einrichtung des FDZ Gesundheit ermöglicht die Forschung in diesen wichtigen Arbeitsfeldern. Bereits einzelne fehlende Daten können diese Ergebnisse verzerren oder eine Frage gänzlich unbeantwortbar machen. Aus diesem Grund sollte die Übermittlung der GKV-Daten an das FDZ Gesundheit ohne individuelle Widerspruchsmöglichkeit erfolgen.

Aktuell ist die Datenlieferung von den Krankenkassen an das FDZ im jährlichen Turnus festgeschrieben. Je nach Forschungsinhalt oder Dringlichkeit von datenbasierten Entscheidungen (bspw. Pandemische Gefährdungslage) sollten Lieferungen im kürzeren Zeitabständen möglich sein. Die COVID-19 Pandemie hat gezeigt, dass GKV-Routinedaten für gesamtgesellschaftlich hochrelevante Fragestellungen eine ganz wesentliche, wenn nicht die einzige Datenbasis für evidenzbasierte Entscheidungen darstellen, dass GKV-Routinedaten aber in Deutschland im internationalen Vergleich sehr viel langsamer und unvollständiger oder nicht direkt verlinkbar (z. B. Impfdatenprojekt RiCO) verfügbar waren. Die Verlinkbarkeit insbesondere bei neuen Datenerhebungen sollte aus gesetzgeberischer Sicht immer mitgedacht werden. Vor dem Hintergrund einer notwendigen besseren Vorbereitung auf zukünftige Pandemien und Krisen (Pandemic Preparedness) und für die Sicherstellung einer internationalen Wettbewerbsfähigkeit der Versorgungsforschung sollte eine sektorenbezogene Beschleunigung der Liefertermine an das FDZ Gesundheit erfolgen. Stationäre Aufenthalte können bereits wenige Tage nach Entlassdatum übermittelt und Verordnungsdaten innerhalb eines Monats verfügbar gemacht werden. Auch die quartalsmäßige Abrechnung der ambulanten ärztlichen Daten über die Kassenärztlichen Vereinigungen (KV) an die Krankenkassen sollten unter dem Gesichtspunkt der schnelleren Nutzarmachung auf den Prüfstand gestellt werden. So könnten in einer Art "fast lane" Verfahren kritische Ereignisse tagesgenau berichtet werden, beispielsweise der Anstieg von Atemwegserkrankungen in der ambulanten Versorgung, als Frühwarnsystem, noch bevor es

zu einer Häufung von stationären Aufnahmen kommt. Dieses Verfahren sollte flexibel auf die jeweilige (heterogene) Gefährdungslage reagieren können.

Momentan ist der Kreis der Forschenden mit ausgewiesener Expertise im Umgang mit GKV-Daten in Deutschland begrenzt. Um einerseits einen breiten wissenschaftlichen Zugang zu den Daten des FDZ Gesundheit zu ermöglichen und gleichzeitig eine hinreichend hohe Qualität der Forschungsarbeiten mit den Daten sicherzustellen, sollte ein Zertifikat entwickelt werden, das Voraussetzung zur Datennutzung wird. Eine mögliche Referenz bildet das für die Beteiligung an Arzneimittelstudien notwendige GCP (Good Clinical Practice)-Zertifikat. Zusätzlich zu einem Zertifikat „Gute Praxis Sekundärdatenanalyse und Linkage“ sollte zur Nachwuchsförderung ein Curriculum für die wissenschaftliche Nutzung und das Linkage von GKV-Routinedaten entwickelt werden.

Momentan können ausschließlich die Daten von GKV-Versicherten für Forschungszwecke nach § 303a ff. SGB V genutzt werden. Der Ausschluss der privat Versicherten stellt nach Überzeugung der Gutachter und der befragten Patientenvertretungen ein Gerechtigkeitsproblem dar. Das selektive Fehlen der Daten der PKV-Versicherten kann auch je nach Fragestellung zu einer Beeinträchtigung der Qualität der Forschungsergebnisse führen. Der Gesetzgeber sollte daher auch die perspektivische Einbeziehung von Versorgungsdaten der PKV in das FDZ Gesundheit vorsehen. Hierbei kann auf erste Projekte, die Versorgungsdaten innerhalb der PKV zu nutzen, aufgesetzt werden.

Das größte wissenschaftliche Potenzial zur Verbesserung der Gesundheitsversorgung und Gesundheit der Bevölkerung liegt im Linkage von GKV-Routinedaten mit anderen Daten. Dies ist heute schon Goldstandard etwa in der NAKO Gesundheitsstudie oder beim Broad Consent der Medizininformatik Initiative. Ein Linkage ist i. d. R. notwendig, um kausale Schlüsse zu ziehen. Wir sehen daher die Linkagemöglichkeit der GKV-Daten mit anderen Datenquellen als eine entscheidende Voraussetzung für den Erfolg des FDZ Gesundheit. Wie die NAKO Gesundheitsstudie zeigt, ist die Akzeptanz der Studienteilnehmenden zur Nutzung und Linkage von GKV-Routinedaten mit Primärforschungsdaten groß. Grundsätzlich sollte das Linkage wie im Fallbeispiel dieses Gutachtens auf Basis einer informierten Einwilligung (informed consent) erfolgen. In der Rechtsgrundlage des FDZ Gesundheit (§303 a bis f SGB V) sowie der zugehörigen Datentransparenzverordnung findet sich derzeit kein Hinweis auf ein konsentbasiertes Datenlinkage, unabhängig davon, ob das Linkage beim Forscher (nach Export der Daten durch das FDZ Gesundheit) oder im FDZ stattfindet (mit Übermittlung der klinischen Primärdaten aus der anfragenden Studie an das FDZ). Eine entsprechende Anpassung der Datentransparenzverordnung wird daher dringend empfohlen.

Im Rahmen größerer prospektiver Kohortenstudien und Register mit konsentbasiertem Datenlinkage ist für die Forschungsinfrastruktur der Studiengruppe eine Datenannahme- und Aufbereitungsstelle, eine Datentreuhänderstelle sowie ein multiperspektivisches Use &

Access Board vorzusehen. Nach Genehmigung des Forschungsvorhabens und Vorliegen eines entsprechenden Ethikvotums erfolgt die Lieferung der GKV-Routinedaten vom FDZ Gesundheit an die Forschungsinfrastruktur der Studiengruppe, die das Datenlinkage dann fragestellungsbezogen umsetzt. Diese Vorgehensweise ist bereits gängige Praxis bei der Kooperation von Forschenden mit Krankenkassen und sollte analog bei der zukünftigen Verfügbarmachung der GKV-Routinedaten durch das FDZ Gesundheit umgesetzt werden.

Für das konsentbasierte Linkage sind im FDZ Gesundheit technische, organisatorische und rechtliche Rahmenbedingungen zu schaffen. Die vorliegende Einwilligung inkl. der Zustimmung zum Datenlinkage muss als Grundlage für die Nutzung von Daten des FDZ Gesundheit durch entsprechende Regelungen juristisch belastbar anerkannt werden. Bei Nutzung der Daten von Krankenkassen und Vorliegen einer Einwilligung sollte auf einen Antrag nach §75 SGB X verzichtet oder das Antragsverfahren beschleunigt werden.

Die Vertrauensstelle DaTraV des FDZ Gesundheit muss organisatorisch in die Prozesse für ein konsentbasiertes Linkage einbezogen werden, damit die in Studienteilnehmerlisten bereitgestellte KVNR in das zugehörige PüP transferiert und an das FDZ Gesundheit projektbezogen übermittelt werden. Wenn zusätzlich Studiendaten übermittelt werden, müssen auch diese an das FDZ Gesundheit weitergeleitet und dort zusammen mit den verlinkten Routinedaten bereitgestellt werden. Hierfür sind im FDZ Gesundheit entsprechende logistische Voraussetzungen sowie die notwendige IT-Infrastruktur zu schaffen, um diese Daten nach der Verknüpfung in einer sicheren Umgebung bereitstellen zu können. Bei Anfrage eines Forschungsprojekts auf Übermittlung von GKV-Routinedaten sind entsprechende Prozesse für die Bereitstellung von Einzeldatensätzen der Studienteilnehmenden zu schaffen. Das Datenlinkage zwischen den Datenquellen erfolgt über temporäre Lieferpseudonyme, die zusammen mit der KVNR übermittelt wurden und die eindeutige Zuordnung der Individualdaten ermöglichen. Bei Datenexport sind diese Lieferpseudonyme in den übermittelten Datensatz zu integrieren.

Das Vorhalten eines Versichertenverzeichnisses in einer Vertrauensstelle DaTraV im Sinne eines Identitätsmanagements wird empfohlen. Hierdurch verbessert sich die Qualität und es eröffnen sich weitere Funktionen und Möglichkeiten. Aus den Daten des FDZ Gesundheit vorselektierte Versicherte könnten angeschrieben werden, um an spezifischen Studien teilzunehmen oder es könnten Daten gezielt bei den Versicherten nacherhoben werden. Dies eröffnet insbesondere für seltene Erkrankungen ein enormes Forschungspotential. Auf Basis des Versichertenverzeichnisses können auch Studien- oder Registerdaten ohne Vorliegen der KVNR verlinkt werden, indem die in beiden Datenquellen verfügbaren identifizierenden Angaben des Probanden/der Probandin verlinkt werden. Das wechselseitige Datenlinkage zweier Datenquellen (z. B. Erkrankungsregister mit GKV Gesundheit) ermöglicht die Bearbeitung neuer Fragestellungen, die allein auf Basis der jeweils verfügbaren Einzeldatenquellen nicht möglich wäre. Alternativ zum Aufbau eines bundesweiten

Versichertenverzeichnis in der Vertrauensstelle DaTraV beim RKI können auch bestehende Strukturen und Stellen, z. B. Datenstelle der Rentenversicherung (DSRV), diese Funktion übernehmen.

In besonderen Fällen, wenn Nutzenpotenziale beträchtlich sind, das öffentliche Interesse an der Forschung oder Planung erheblich ist und eine Einwilligung der betroffenen Personen nicht umsetzbar oder unzumutbar ist, sollte das Linkage komplementärer VeDa auch ohne Einwilligung ermöglicht werden. In diesem Fall sollte – unabhängig davon, ob die GKV-Daten vom FDZ Gesundheit oder von Krankenkassen direkt bezogen werden – ein Antrag nach §75 SGB X auf die Übermittlung von Sozialdaten für die Forschung gestellt werden. Ein Beispiel bildet hier das Linkage von GKV-Daten mit Daten der Klinischen Krebsregister (KKR). Das vom Innovationsfonds geförderte Projekt WiZen zur Wirksamkeit der Versorgung in onkologischen Zentren zeigte, dass das Linkage von GKV-Daten mit Daten ausgewählter KKR anhand der pseudonymisierten Krankenversicherungsnummer (KVNR) realisiert werden kann. Ein momentanes Hindernis für die grundsätzliche Ermöglichung eines Linkages bundesweiter KKR-Daten mit GKV-Daten ist die Uneinheitlichkeit der landesgesetzlichen Regelungen in den Krebsregistern. Hier ist dringend eine Harmonisierung notwendig, damit der große Datenschatz, der durch die Verbindung von GKV-Routinedaten und KKR-Daten entsteht, für die onkologische Spitzenforschung, beispielsweise für die Ableitung von Leitlinienempfehlungen genutzt werden kann und damit die medizinische Versorgung von krebserkrankten Menschen in Deutschland verbessert werden kann.