

Beiträge und Analysen

# Gesundheitswesen

## aktuell 2022

herausgegeben von Uwe Repschläger,  
Claudia Schulte und Nicole Osterkamp

**Peter Ihle, Katharina Schneider, Steffen Heß**

Das Forschungsdatenzentrum Gesundheit – Routinedaten der Gesetzlichen Krankenversicherung  
für die Gesundheits- und Versorgungsforschung,  
Seite 32–47

doi: 10.30433/GWA2022-32

Peter Ihle, Katharina Schneider, Steffen Heß

## **Das Forschungsdatenzentrum Gesundheit – Routinedaten der Gesetzlichen Krankenversicherung für die Gesundheits- und Versorgungsforschung**

Routinedaten, insbesondere die Daten der Gesetzlichen Krankenversicherung (GKV), sind inzwischen ein fester Bestandteil in der Gesundheits- und Versorgungsforschung geworden. Der Beitrag beschäftigt sich mit der „Erfolgsstory“ der Routinedaten und deren Bereitstellung im Forschungsdatenzentrum Gesundheit (FDZ). Dabei wird die historische Entwicklung bis zum Stapellauf des FDZ aufgezeigt und auf grundsätzliche Vorteile und Limitationen von GKV-Routinedaten eingegangen. Weiterhin wird die geplante Arbeitsstruktur des FDZ beschrieben und abschließend die Beteiligung des FDZ bei der Planung des europäischen Gesundheitsdatenraums umrissen.

### **Zielsetzung**

Erste Gehversuche zur Sekundärnutzung von medizinischen Abrechnungsdaten reichen bis in die 1980er Jahre zurück (Schwefel et al. 1987). Inzwischen sind die GKV-Routinedaten als Datenquelle für die medizinische Forschung fester Bestandteil geworden. So nutzen beispielsweise die Projekte des G-BA-Innovationsfonds zu den beiden Themengebieten Versorgungsforschung und Versorgungsformen, aber auch viele weitere Forschungsprojekte diese GKV-Routinedaten. Als weiterer konsequenter Schritt der wissenschaftlichen Nutzung von GKV-Routinedaten kann das geplante Forschungsdatenzentrum Gesundheit (FDZ) am Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) in Bonn angesehen werden. Das FDZ Gesundheit wird zukünftig die GKV-Routinedaten von etwa 73,4 Millionen Versicherten (Bundesgesundheitsministerium 2022), entsprechend knapp 88 Prozent der Gesamtbevölkerung Deutschlands für die Forschung bereitstellen und damit eine neue Ära in der Nutzung dieser Datenbasis einläuten.

Der Vollständigkeit halber sei an dieser Stelle darauf hingewiesen, dass neben den GKV-Routinedaten, die nach aktueller Rechtsgrundlage im FDZ Gesundheit verfügbar sein werden, auch Daten anderer Sozialleistungsbereiche existieren, beispielsweise der Pflege- oder Rentenversicherung, die ebenfalls für Forschungsprojekte genutzt, aber in diesem Beitrag nicht weiter behandelt werden.

## GKV-Routinedaten

GKV-Routinedaten sind Abrechnungsdaten und entstehen im Abrechnungskontext bei der medizinischen Behandlung von Patientinnen und Patienten. Leistungserbringer, beispielsweise Krankenhäuser oder Arztpraxen, müssen ihre Leistungen und abrechnungsbegründende Diagnosen an die jeweils versichernde Krankenkasse übermitteln, worauf entsprechende Zahlungen erfolgen. Dieser Prozess erfolgt unter Einschaltung weiterer Stellen, wie Kassenärztliche Vereinigungen oder Apothekenabrechnungszentren. Die Abrechnungsdaten werden bei den Krankenkassen versichertenbezogen gespeichert und unterliegen vorgegebenen Löschrufen.

Die Vor- und Nachteile von GKV-Routinedaten sowie ihre Eignung für Forschungszwecke sind anderweitig ausführlich aufgezeigt worden (Swart et al. 2014). In diesem Beitrag soll daher auf die drei wichtigsten Eigenschaften dieser Routinedaten eingegangen werden: Vollständigkeit, Selektionsfreiheit und Populationsbezug.

Die Abrechnungsdaten der gesetzlichen Krankenkassen bilden das medizinische Versorgungsgeschehen nahezu lückenlos ab, da sie aus allen Versorgungssektoren und von allen Leistungserbringern Daten bereitstellen und damit in idealer Weise auch die Schnittstellen der Versorgung für die Forschung zugänglich machen. Die aus Sicht der Versorgungsforschung wichtigen Profile sollen hier nur stichwortartig wiedergegeben werden: ambulante ärztliche Versorgung und stationäre Aufenthalte mit Leistungen und Diagnosen, Arzneimittel, Arbeitsunfähigkeiten, weitere Leistungsbereiche wie Heil- und Hilfsmittel, Hebammenleistungen oder häusliche Krankenpflege (weitergehende Informationen siehe Swart et al. 2014). Für die genannte Vollständigkeit entscheidend ist, dass jede dokumentierte Leistung mehr oder weniger an Finanzströme gekoppelt und der oder die Dokumentierende direkt davon profitiert und damit das Dokumentationsverhalten positiv verstärkt wird. Ob hierbei zuweilen auch falsche Anreize gesetzt werden, wurde im Rahmen von Upcoding im DRG-System in der Vergangenheit diskutiert (Bundesgesundheitsministerium 2017) und muss natürlich bei der Routinedatenanalyse und Interpretation der Ergebnisse als Möglichkeit grundsätzlich bedacht werden. Neben der vollständigen Dokumentation der Leistungen enthalten die Routinedaten im ambulanten Sektor sogenannte abrechnungsbegründende Diagnosen. Das Pendant im stationären Bereich sind die Hauptentlassungsdiagnosen

und weitere Diagnosen, die eng mit den Operationen und Prozeduren (dokumentiert nach OPS) und dem DRG-Entgeltsystem assoziiert sind. Daneben finden sich im stationären Sektor auch weitere Diagnosekategorien, wie Aufnahme- und Nebendiagnosen.

Die zweite wichtige Eigenschaft von Routinedaten ist das weitgehende Fehlen von Selektionseffekten. Die Abrechnungsdaten werden unabhängig von den Ausgangsvoraussetzungen der Versicherten dokumentiert. Damit gibt es beispielsweise keinen Erinnerungsbias oder sprachlich beziehungsweise sozioökonomisch begründete Verzerrungen, wenn die Daten direkt beim Versicherten erhoben würden. Anders als Befragungen oder Studien auf der Basis von Patienteneinwilligungen können Forschungsprojekte mit Routinedaten auf Basis des § 75 SGB X ohne Einwilligung der Versicherten durchgeführt werden, sofern die Einholung als unzumutbar dargelegt wird. Diese Vorgehensweise wird im genannten Paragraphen als Möglichkeit aufgeführt, muss gegenüber den Aufsichtsbehörden jeweils projektbezogen und hinreichend gut begründet werden und setzt einen positiven Bescheid der zuständigen Aufsichtsbehörde voraus. Eine systematische Verzerrung der Daten durch Selektion von Patientinnen und Patienten, die zur Teilnahme bereit sind, kann dadurch vermieden werden. Das ist insbesondere dann von Vorteil, wenn Forschungsfrage und Nichtbereitschaft für die Teilnahme inhaltlich assoziiert sind, beispielsweise die Erreichbarkeit des Gesundheitssystems durch vulnerable Gruppen oder Adhärenzfragen. Unerkannte Verzerrungen durch Nicht-Teilnehmende sind insbesondere bei der Interpretation der Forschungsergebnisse regelmäßig Herausforderungen beziehungsweise schränken die Generalisierbarkeit der Aussagen ein. Eine unverzerrte Versichertenbasis mit Routinedaten besitzt diesbezüglich einen methodischen Vorteil im Datenzugang oder kann gegebenenfalls auch für so genannte Responder-Analysen eingesetzt werden, um Selektionseffekte zu erkennen.

Ergänzend zu den beiden bereits genannten Vorteilen - Vollständigkeit und Selektionsfreiheit – ist die Kenntnis über die versicherte Grundgesamtheit der jeweiligen Krankenkasse (SGB V § 288 Versichertenverzeichnis, Sozialgesetzbuch 2011) zu nennen. Damit steht zum einen der für epidemiologische Kennzahlen wie Prävalenz oder Inzidenz benötigte Nenner zur Verfügung, zum anderen kann auch die Nichtinanspruchnahme von GKV-finanzierten Leistungen eindeutig abgebildet und anteilmäßig auf die (beispielsweise unter einem bestimmten Risiko stehende) Zielpopulation bezogen werden.

Jedoch existieren bestimmte Limitationen: Abgesehen vom Fehlen bestimmter Gruppen, beispielsweise der Privatversicherten, ist hier vor allem die Tatsache zu nennen, dass nur dokumentiert und an die Krankenkasse übermittelt wird, was einerseits abrechnungsrelevant und andererseits zusätzlich an Dokumentation erforderlich ist. Leistungen außerhalb des GKV-Systems, also beispielsweise sogenannte Individuelle Gesundheitsleistungen (IGEL) oder für die Selbstmedikation erworbene, nicht verschreibungspflichtige „Over the Counter“ (OTC) Präparate, sind in Routinedaten genauso wenig enthalten wie die Dokumentation von Untersuchungsergebnissen oder Laborwerten. Andere notwendige oder relevante Angaben, beispielsweise Schweregrade von Erkrankungen, müssen meist näherungsweise operational definiert und approximiert werden (Schubert et al. 2017). Ebenso wenig enthalten sind Angaben zum Lebensstil (beispielsweise Ernährungsgewohnheiten, sportliche Betätigung) oder im Zusammenhang mit Pflegefragen relevante Informationen wie die Lebensumstände (beispielsweise „alleinlebend“). Für andere relevante Einflussfaktoren wie Rauchen oder Adipositas existieren zwar entsprechende ICD-Angaben, diese werden jedoch vornehmlich morbiditätsbezogen dokumentiert und können damit nur eingeschränkt als Indikatoren verwendet werden.

Bilanziert betrachtet überwiegen jedoch die beschriebenen Vorteile von Routinedaten erheblich und haben sie zu einer wertvollen Datenquelle für die medizinische Forschung werden lassen.

## Vorteile eines Forschungsdaten zentrums

Bei der wissenschaftlichen Nutzung von GKV-Routinedaten ist zu beachten, dass Routinedaten primär zum Zwecke der Abrechnung und fallbezogen entstanden sind. Bei der Sekundäranalyse werden die Daten regelhaft aus unterschiedlichen Sektoren und über längere Zeiträume versichertenbezogen zusammengeführt. Durch diese „Gesamtschau“ im Nutzungskontext können zuweilen fehlerhafte, implausible oder widersprüchliche Angaben entstehen. Im Gegensatz zu klinischen Studien, bei denen die Daten auf Basis eines Studienprotokolls mit geeigneten qualitätssichernden Maßnahmen erhoben werden, müssen diese qualitätssichernden Schritte nachträglich erfolgen und umfassen umfangreiche Aufbereitungsschritte und Plausibilitätsprüfungen (Horenkamp-Sonntag

et al. 2014), insbesondere dann, wenn die Daten selektiv für ein spezifisches Forschungsvorhaben von einer oder mehreren Krankenkassen bereitgestellt werden. Hierfür – und das zeigt die Erfahrung aus bisher durchgeführten Projekten – ist ausreichend Zeit einzuplanen und es bedarf essenzieller Kenntnisse der spezifischen Daten (siehe hierzu Swart et al. 2015). Dieser Punkt macht deutlich, dass es effizienter sein kann, Daten an eine hiermit betraute Stelle zu übermitteln, zentral aufzubereiten und zu verwalten.

Ein Beispiel für eine bereits bestehende Gesundheitsdatenbank ist die Kölner CoReNet-Datenbank des Instituts für Medizinsoziologie, Versorgungsforschung und Rehabilitationswissenschaft (IMVR) der Humanwissenschaftlichen Fakultät und der Medizinischen Fakultät der Universität zu Köln (KöR), bei der eine Datenbereitstellung per personalisiertem VPN-Zugang für externe Antragsteller nach positivem Bescheid des Steering-Komitees und des GKV-Beirats der kooperierenden Krankenkassen möglich ist (Scholten et al. 2012). Die hierbei gewonnenen Erfahrungswerte bezüglich der Datenaufbereitung und Datenbereitstellung zeigen, dass die Daten im Rahmen von Plausibilitätsprüfungen und Analyseschritten sukzessive an Qualität zunehmen. Das Vorhalten von Daten und die kontinuierliche Nutzung einer Datenquelle in einem Forschungsdatenzentrum bringt daher unter Qualitätsaspekten große Vorteile mit sich. Sowohl durch eigene Qualitäts- und Plausibilitätsprüfungen als auch durch erwünschte Rückmeldungen der Nutzenden können hier wichtige Weiterentwicklungen erfolgen. Aus diesem Grund ist ein enger und regelmäßiger Austausch mit der Community der Nutzungsberechtigten unabdingbar (siehe hierzu auch Swart et al. 2021).

Eine Aufbereitung und Bereitstellung von Daten durch eine zentrale Stelle hat zusätzlich den Vorteil, dass die Krankenkassen insofern entlastet werden, als dass diese dann in einem planbaren, regelmäßigen Turnus die Daten übermitteln, statt diese jeweils auf Anfrage der Forschenden für die individuellen Projekte zur Verfügung zu stellen. Der durch die vorgegebene Turnuslänge entstehende zeitliche Verzug der Daten muss dabei wiederum bei der Planung der Analysen berücksichtigt werden, was bei der Notwendigkeit hochaktueller Daten ein zu bedenkender Faktor sein kann. Andererseits entfällt für die Forschenden selbst der Schritt, sich eigenständig mit ihrem Anliegen an die Krankenkassen wenden zu müssen. Der Antrag kann dann direkt bei der daten-

haltenden Stelle gestellt werden. Dies kann viel Zeit einsparen, sodass dringliche Forschungsvorhaben unter Umständen schneller umgesetzt werden können.

Ein weiterer sehr entscheidender Faktor ist, dass in einem Forschungsdatenzentrum die Gesundheitsdaten aller gesetzlichen Krankenversicherungen zusammengeführt und bereitgestellt werden können. So stehen viel größere Datenmengen für „Big Data“-Analysen zur Verfügung und ermöglichen den Nutzungsberechtigten die Anwendung moderner Analysemethoden. Zudem werden Verzerrungseffekte, die gegebenenfalls bei der Nutzung von Daten einzelner Krankenkassen (Hoffmann und Icks 2012) aufgrund der unterschiedlichen Versichertenstrukturen entstehen können, vermieden.

Auch im Sinne des Datenschutzes ist die Bereitstellung der sensiblen Gesundheitsdaten durch eine hierfür speziell eingerichtete, öffentliche Stelle ein wichtiger Schritt. Eine einheitliche und transparente Vorgehensweise fördert die öffentliche Wahrnehmung und kann das Vertrauen in die Nutzung der Gesundheitsdaten in der Bevölkerung wesentlich stärken.

### **Historie**

Die Anfänge der wissenschaftlichen Sekundärnutzung von Krankenkassendaten reichen bis in die 1980er Jahre zurück, siehe hierzu beispielsweise in Schwefel et al. (1987). Ausgehend von dem Konzept Versichertenstichprobe (Ferber 1997) wurde als Auftragsgutachten im Rahmen des Methodenforschungsprojekts des Statistischen Bundesamtes „Stichprobe aus der Gesetzlichen Krankenversicherung“ (Ihle et al. 1999) eine Stichprobe mit konstantem Auswahlatz mit knapp einer Million Versicherten vorgeschlagen, die mithilfe eines Hashwerts aus der Krankenversicherungsnummer zufällig über alle Versicherten aller Krankenkassen gezogen werden sollte. Eine Realisierung dieses Konzepts erfolgte jedoch lediglich auf regionaler Ebene und wird aktuell für jährliche Nachlieferungen angewendet (Ihle et al. 2005). Im Nachhinein stellt sich allein die Beauftragung dieses Gutachtens als visionär dar, benötigte es danach mehrerer Anläufe im Gesetzgebungsverfahren bis im Rahmen der Datentransparenzverordnung (Sozialgesetzbuch 2011) die

Daten im Jahre 2014 erstmalig aus dem so genannten morbiditätsorientierten Risikostrukturausgleich als GKV-weite Datenbasis im Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) für Analysen nutzbar wurden. Diese Forschungsdaten konnten nach Antragstellung mithilfe selbst geschriebener SQL-Skripte – geprüft und ausgeführt im geschützten Bereich der DaTraV-Datenbank durch Mitarbeitende des DIMDI – analysiert werden. Alternativ konnte die Analyse selbst durch Mitarbeitende der Datentransparenzstelle beantragt werden. Der Datenkranz war gegenüber den regulären GKV-Routinedaten jedoch stark eingeschränkt, so gab es beispielsweise aus dem ambulanten und stationären Sektor keine Leistungs- oder OPS-Angaben, eine Unterscheidung nach Leistungserbringern fehlte vollständig, Krankenhausaufenthalte lagen lediglich monatsgenau ohne Liegedauern und durchlaufend nummeriert vor. Zudem standen die Daten erst mit einer zeitlichen Verzögerung von bis zu vier Jahren zur Verfügung. Diese Datenreduktion sowie weitere im Hinblick auf Konzept, Logistik und Administration einschränkende Rahmenbedingungen wirkten nicht förderlich für die Datennutzung.

## Struktur des FDZ Gesundheit (BfArM)

### Routinedaten im FDZ Gesundheit

„Mit dem Digitale-Versorgung-Gesetz (DVG) (Digitale-Versorgung-Gesetz 2019), das am 19. Dezember 2019 in Kraft getreten ist, wird die bisherige Datenaufbereitungsstelle zu einem Forschungsdatenzentrum weiterentwickelt“ (Datentransparenzverordnung 2020). Erklärtes Ziel dabei ist es, „Abrechnungsdaten der Gesetzlichen Krankenkassen deutlich schneller und in einem größeren Umfang als bisher u.a. der Versorgungsforschung zugänglich zu machen“ (ebenda). Dieses Ziel führt zum Forschungsdatenzentrum (FDZ) Gesundheit. Das FDZ Gesundheit, das wie oben erwähnt aus der Datenaufbereitungsstelle des ehemaligen DIMDI hervorgeht, wird einige Weiterentwicklungen mit sich bringen. Insbesondere soll der gesamte Prozess deutlich forschungsorientierter werden, wobei weiterhin größter Wert auf den Datenschutz gelegt wird. Neben einem neu aufgebauten digitalen Antragsverfahren erfährt auch die Datenbereitstellung einige



Innovationen, die dem iterativen Vorgehen bei der wissenschaftlichen Forschung gegenüber dem bisherigen Verfahren deutlich stärker entgegenkommen wird.

Zudem erweitert sich der Datenkranz erheblich. Der ursprüngliche Bestand der Datenaufbereitungsstelle wurde damals im Rahmen des morbiditätsorientierten Risikostrukturausgleichs (Morbi-RSA) aufgebaut und umfasst die Berichtsjahre 2008 (aufgrund methodischer Einschränkungen sinnvoll nutzbar ab 2009) bis 2015. Dieser Datenbestand wird vom FDZ übernommen. Die folgenden Berichtsjahre enthalten im Sinne einer Übergangsregelung gemäß § 12 der Datentransparenzverordnung zunehmend mehr Variablen. Mit dem Berichtsjahr 2023 wird dann der gesamte Datenumfang, wie er in § 3 der Datentransparenzverordnung (Datentransparenzverordnung 2020) vorgesehen ist, vorliegen. Dieser enthält neben Stammdaten wie Geburtsjahr, Geschlecht und Postleitzahl, Angaben zur Versicherung (beispielsweise die Anzahl Versichertentage), Daten aus der ambulanten und stationären Versorgung sowie zur Verordnung und Abgabe von Arzneimitteln. Im Bereich der ambulanten und stationären Abrechnungsdaten gibt es eine Reihe von entscheidenden Ergänzungen, so können in Zukunft beispielsweise auch OPS-Kodes zur Analyse von durchgeführten Prozeduren verwendet oder die Beatmungstunden der Patienten ausgewertet werden. Angaben zu den Leistungserbringern wie etwa die „Lebenslange Arztnummer“ werden ebenfalls enthalten sein, um die jeweilige Patientenversorgung zuordnen zu können, wobei im Sinne des Datenschutzes Pseudonyme verwendet werden. Auswertungsrelevante Angaben in den Originalidentifikatoren, wie beispielsweise der Fachgruppenschlüssel in der „Lebenslangen Arztnummer“, stehen jedoch für die Analysen zur Verfügung. Neben den im vorherigen Bestand bereits vorhandenen Profilen werden einige neue Profile für die Forschung verfügbar sein. Hierzu zählen Heil- und Hilfsmittel, Krankentransportleistungen, Hebammenhilfen, häusliche Krankenpflege und die Versorgung mit digitalen Gesundheitsanwendungen. So werden auch neuere Versorgungsformen zukünftig Eingang in Sekundärdatenanalysen finden können, was die Breite an Variablen erheblich erweitern wird.

Die genannten Daten werden für alle rund 73 Millionen gesetzlich Krankenversicherten in Deutschland verfügbar sein. Die damit möglichen Auswertungen anhand großer

Datensätze (Big Data) haben methodisch bedeutende Vorteile. So können beispielsweise durch die höheren Fallzahlen auch kleinere Effekte detektiert werden oder bei selteneren Erkrankungen überhaupt erst ausreichend Daten für eine statistische Analyse bereitgestellt werden. Da die Daten jährlich ergänzt werden, sind auch longitudinale Auswertungen möglich. Dabei wird die datenschutzkonforme Zusammenführung der Datenjahre auf Versichertenebene durch den Einsatz einer Vertrauensstelle ermöglicht. Diese ist beim Robert Koch-Institut (RKI) eingerichtet. Während die sensiblen Gesundheitsdaten, die nach Übermittlung durch die Krankenkassen in pseudonymisierter Form beim GKV-Spitzenverband gesammelt, aufbereitet und anschließend an das FDZ Gesundheit weitergeleitet werden, erhält die Vertrauensstelle lediglich ein Lieferpseudonym, das dort in ein periodenübergreifendes Pseudonym überführt und anschließend an das FDZ Gesundheit gesandt wird. Die Verbindung der Gesundheitsdaten mit dem periodenübergreifenden Pseudonym erfolgt über eine Arbeitsnummer, die jeweils bei jedem Übermittlungsprozess mitgeschickt wird (Datentransparenzverordnung 2020). Erst im FDZ Gesundheit können die verschiedenen Datenjahre miteinander verknüpft werden.

## Antragsprozess

Bei einer Umfrage unter Nutzenden der bisherigen Datenaufbereitungsstelle wurde die Planbarkeit der Auswertungen als größter Wunsch identifiziert. Gerade wenn Forschungsprojekte nur einen begrenzten Förderzeitraum haben oder die Bearbeitung der Forschungsfrage, wie beispielsweise zuletzt in der Corona-Pandemie, dringlich ist, ist es wichtig zu wissen, in welchem zeitlichen Rahmen mit der Bereitstellung der Daten gerechnet werden kann. Dieser Bedarf wird zukünftig im agilen Entwicklungsprozess, also in enger Absprache mit den Nutzenden, umgesetzt.

Die Antragstellung beim FDZ Gesundheit wird vollständig digital erfolgen. Das hierfür notwendige Formular wird derzeit erstellt. Bevor ein Antrag auf Datenbereitstellung möglich ist, muss sich die Institution zunächst registrieren und authentifizieren, da die Nutzungsberechtigung nach § 303e Absatz 1 SGB V (Datentransparenzverordnung 2020) nur auf einen bestimmten Kreis beschränkt ist. Hierzu zählen neben Hochschulen und anderen Forschungseinrichtungen unter anderem auch die Krankenkassen selbst sowie

verschiedene weitere Institutionen, Behörden und Patientenorganisationen. Auch die Nutzungszwecke sind im Sinne des Datenschutzes eingeschränkt (Sozialgesetzbuch 2022, § 303e Absatz 2) und müssen im Antrag angegeben werden. Mögliche Zwecke sind beispielsweise neben der Forschung selbst die Verbesserung der Versorgungsqualität oder die Planung von Leistungsressourcen, wie etwa die Krankenhausplanung. Weiterhin werden im Antrag die Auswertungsmethoden, die beteiligten Personen sowie die Struktur und der Umfang der benötigten Daten abgefragt. Die Erforderlichkeit und Eignung der Daten für die Fragestellung muss dabei nachvollziehbar beschrieben werden. Auch krankenkassenbezogene Auswertungen sind möglich, sofern eine Einwilligung der betroffenen Krankenkasse beigefügt wird. Antragstellende verpflichten sich, keinen Bezug zu Personen, Leistungserbringern oder -trägern herzustellen, dafür geeignete technische und organisatorische Maßnahmen umzusetzen und die Daten nur für die angegebenen Zwecke zu nutzen (ebenda).

Sobald ein Antrag im FDZ Gesundheit eingegangen ist, wird dieser zunächst formal (beispielsweise auf seine Vollständigkeit) und anschließend inhaltlich (beispielsweise bezüglich der Nutzungszwecke) geprüft. In der Vergangenheit wurde, wie bereits erwähnt, der Prüfzeitraum von Forschenden oft als zu langwierig und zu wenig transparent erlebt. Dem Wunsch nach Planbarkeit wird nun nachgekommen, da innerhalb von drei Monaten (maximal vier bei aufwendiger Prüfung) über den Antrag entschieden werden muss (Datentransparenzverordnung 2020). Mehr Planbarkeit soll auch durch mehr Informationen über den Antragsfortschritt und die nächsten Schritte erreicht werden. So können die Nutzungsberechtigten verlässlich einplanen, wann sie spätestens mit einer Datenbereitstellung rechnen können. Bereits im Aufbau des FDZ Gesundheit wird darauf hingearbeitet, Antrags- und Bearbeitungsprozesse nach Möglichkeit zu automatisieren und skalierbar zu implementieren, um auch bei erwarteter stetiger Zunahme der Anträge dem Forschungsbedarf bestmöglichst entsprechen zu können.

Um die Nutzung der Daten transparent zu gestalten, führt das FDZ Gesundheit ein öffentliches Register, das für jeden Antrag jeweils die Nutzungsberechtigten, den Titel, die Zwecke, das Jahr der Antragsentscheidung und eine kurze Ergebnisdarstellung enthält (ebenda). Für die Antragsbearbeitung wird eine Gebühr nach der Datentransparenz-

Gebührenverordnung je nach Prüfungs- und Beratungsaufwand und je nach Umfang der beantragten Daten fällig (Datentransparenz-Gebührenverordnung 2020).

## Datenbereitstellung

Fällt die Prüfung eines Antrags positiv aus, ist geplant, den Nutzungsberechtigten ein Scientific-Use-File für einen begrenzten Zeitraum bereitzustellen, mit dessen Hilfe diese Auswertungsskripte erstellen können. Hierbei handelt es sich um projektspezifische Datenzuschnitte. Dadurch soll ein iterativer Forschungsprozess ermöglicht werden. Die Bereitstellung der Scientific-Use-Files erfolgt in gesicherten, virtuellen Analyse-räumen, die beispielsweise das Herunterladen der sensiblen Gesundheitsdaten verhindern. Die anschließend mithilfe der Analyseskripte anhand des Originaldatensatzes generierten Ergebnismengen werden vor dem Export bezüglich des Re-Identifikationsrisikos durch das FDZ Gesundheit geprüft und können gegebenenfalls nach weiteren Maßnahmen zur Minimierung des Risikos an die jeweiligen Nutzungsberechtigten übermittelt werden.

Bei den Analysetools soll den Bedarfen der Forschenden zukünftig bestmöglich entgegengekommen werden. So sollen beispielsweise in weiteren Entwicklungsstufen auch intuitive Point-and-Click-Analysen möglich sein und auch Tools zur Nutzung von Methoden der Künstlichen Intelligenz zur Verfügung stehen. So können auch zukunftsorientierte Methoden Eingang in die Routinedatennutzung finden.

## Ausblick

### Erweiterungen (ePA)

Neben den Routinedaten der gesetzlichen Krankenkassen wird das FDZ Gesundheit im Verlauf auch Daten aus den elektronischen Patientenakten (ePA) bereitstellen können (Sozialgesetzbuch 2022, § 363 Verarbeitung von Daten der elektronischen Patientenakte zu Forschungszwecken). Hierfür sollen Versicherte ab dem Jahr 2023 die Daten ihrer eigenen ePA freiwillig und nach informierter Einwilligung freigeben können. Diese werden dann wie die Routinedaten in pseudonymisierter Form an das FDZ Gesundheit übermittelt. Die ersten Daten der ePA werden beispielsweise der Impfausweis und das

Zahnbonusheft sowie der Mutterpass und das Kindervorsorgeuntersuchungsheft sein. Diese werden als sogenannte „Medizinischen Informationsobjekte“ (MIOs) und damit in strukturierter Form vorliegen (Kassenärztliche Bundesvereinigung 2021). Der für die Aufbereitung und Bereitstellung der ePA-Daten erforderliche ETL-Prozess („Extract, Transform, Load“) wird derzeit erarbeitet. Durch die Ergänzung der Routinedaten mit den Informationen aus den elektronischen Gesundheitsakten erhöht sich die Datenvielfalt am FDZ Gesundheit erheblich, was wiederum die Möglichkeiten für medizinisch-wissenschaftliche Analysen deutlich erweitert.

## Europäische Einordnung

Um zukunftsfähig zu sein, führt das FDZ Gesundheit mit verschiedenen Kooperationspartnern Weiterentwicklungsprojekte durch und ist national und international in kontinuierlichem Austausch mit anderen Gesundheitsdateninfrastrukturen und Stakeholdern. Eine besondere Bedeutung kommt hierbei dem Europäischen Gesundheitsdatenraum, englisch „European Health Data Space“ (EHDS) (Europäische Kommission 2022a) zu, der derzeit von der Europäischen Kommission geplant wird. Dieser soll die grenzüberschreitende Primär- (beispielsweise der Zugriff auf die elektronische Gesundheitsakte bei Krankheitsfall in einem anderen EU-Mitgliedsstaat) und Sekundärnutzung von Gesundheitsdaten (beispielsweise für internationale Forschungsprojekte) regeln. Ein entsprechender Verordnungsentwurf wurde am 3. Mai 2022 von der EU-Kommission veröffentlicht (Europäische Kommission 2022b) und befindet sich aktuell in der Abstimmung. Im Bereich der Sekundärdatennutzung soll im Rahmen des EHDS eine gemeinschaftliche Gesundheitsdateninfrastruktur aufgebaut werden, die der Problematik der derzeit noch stark fragmentierten Datenlandschaft entgegenwirkt.

Unterstützt wird die EU-Kommission von der Joint Action Towards the European Health Data Space (TEHDAS 2020), an der sich auch das FDZ Gesundheit engagiert. Hier erarbeiten Kooperationspartner aus 25 europäischen Ländern gemeinschaftlich Konzepte für den EHDS. Für den Aufbau und die Vertestung einer ersten Version des EHDS ist ein EHDS2-Pilotprojekt geplant. Auch hier ist das FDZ Gesundheit als Co-Lead des Arbeitspaketes zum Aufbau der IT-Infrastruktur beteiligt.

Auch im europäischen regulatorischen Netzwerk ist das FDZ Gesundheit aktiv und ist Teil der „Big Data Steering Group“ und dem Advisory Board des Projekts „Data Analysis and Real World Interrogation Network (DARWIN EU®)“, bei dem eine datenschutzkonforme Infrastruktur zur Beantwortung von arzneimittelregulatorischen Fragestellungen mithilfe eines Koordinierungszentrums aufgebaut werden soll (Arlett et al. 2022). Dieses wiederum steht in enger Verbindung mit dem geplanten EHDS.

Mit Blick in die Zukunft stehen also bedeutende Weiterentwicklungen für die sekundäre Nutzung von Gesundheitsdaten an. Das FDZ Gesundheit wird völlig neue Möglichkeiten schaffen, große Datenmengen für wichtige wissenschaftliche Fragestellungen in einer sicheren und datenschutzkonformen Umgebung zu analysieren. Durch die europäische Vernetzung und den Aufbau des EHDS werden die großen und bisher noch zu wenig genutzten Potenziale von Sekundärdaten noch weiter ausgeschöpft werden können, um die Gesundheitsversorgung in Deutschland und Europa weiter zu verbessern.

## Start des FDZ Gesundheit

Das FDZ Gesundheit befindet sich derzeit noch im Aufbau. Bis zu seinem Start sind Anträge auf Datenbereitstellung noch nicht möglich. In der Zwischenzeit können sich Interessierte auf der Homepage über das FDZ Gesundheit (Bundesinstitut für Arzneimittel und Medizinprodukte 2022) und insbesondere auch über Neuigkeiten informieren. Auch der Livegang wird dort bekannt gegeben und der Zugang zum Antragsportal hier lokalisiert sein.

## Literatur

- Arlett, P., Kjær, J., Broich, K. und E. Cooke (2022). Real-World Evidence in EU Medicines Regulation: Enabling Use and Establishing Value. In: *Clinical Pharmacology & Therapeutics* 111 (1). S. 21–23.
- Bundesgesundheitsministerium (2017). Sondergutachten zu den Wirkungen des morbiditätsorientierten Risikostrukturausgleichs erstellt durch den Wissenschaftlichen Beirat zur Weiterentwicklung des Risikostrukturausgleichs im Auftrag des Bundesministeriums für Gesundheit. Bonn, 27. November 2017. Online unter [www.bundesgesundheitsministerium.de/fileadmin/Dateien/5\\_Publikationen/Gesundheit/Berichte/Sondergutachten\\_Wirkung\\_RSA\\_2017.pdf](http://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Gesundheit/Berichte/Sondergutachten_Wirkung_RSA_2017.pdf) (Download am 11. Juli 2022).
- Bundesgesundheitsministerium (2022). KM1-Statistik, Stand: Mai 2022, Gesetzliche Krankenversicherung – Mitglieder, mitversicherte Angehörige und Krankenstand (Monatswerte Januar-April 2022). Online unter [www.bundesgesundheitsministerium.de](http://www.bundesgesundheitsministerium.de) (Download am 9. Juli 2022).
- Bundesinstitut für Arzneimittel und Medizinprodukte (2022). Das FDZ Gesundheit. Online unter [www.forschungsdatenzentrum-gesundheit.de/das-fdz](http://www.forschungsdatenzentrum-gesundheit.de/das-fdz) (Download am 15. Juli 2022).
- Datentransparenz-Gebührenverordnung (DaTraGebV) (2020). Verordnung zur Erhebung von Gebühren und Auslagen für die Bereitstellung von Daten nach den Regelungen der Datentransparenzverordnung. Datentransparenz-Gebührenverordnung vom 30. April 2014. In: *Bundesgesetzblatt I.*, S. 458, die zuletzt durch Artikel 2 der Verordnung vom 19. Juni 2020 (*Bundesgesetzblatt I.*, S. 1371) geändert worden ist. Online unter [www.gesetze-im-internet.de/datragebv/](http://www.gesetze-im-internet.de/datragebv/) (Download am 30. August 2022).
- Datentransparenzverordnung (2020). Verordnung zur Neufassung der Datentransparenzverordnung und zur Änderung der Datentransparenz-Gebührenverordnung 2020. In: *Bundesgesetzblatt, Jahrgang 2020, Teil I, Nr. 29* Bonn. Online unter [www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/Gesetze\\_und\\_Verordnungen/GuV/D/VO\\_Datentransparenzverordnung.pdf](http://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/D/VO_Datentransparenzverordnung.pdf) (Download am 9. Juli 2022).
- Digitale-Versorgung-Gesetz [DVG] (2019). Gesetz vom 9. Dezember 2019. In: *Bundesgesetzblatt I.* S. 2562.

- Europäische Kommission (2022a). Verordnung zur Schaffung eines europäischen Raums für Gesundheitsdaten. Online unter [https://ec.europa.eu/health/ehealth-digital-health-and-care/european-health-data-space\\_de](https://ec.europa.eu/health/ehealth-digital-health-and-care/european-health-data-space_de) (Download am 15. Juli 2022).
- Europäische Kommission (2022b). Proposal for a regulation – The European Health Data Space. Online unter [https://ec.europa.eu/health/publications/proposal-regulation-european-health-data-space\\_en](https://ec.europa.eu/health/publications/proposal-regulation-european-health-data-space_en) (Download am 15. Juli 2022).
- Ferber, C. von (1997). Sekundärdatenforschung – Methodische und theoretische Aspekte. In: Ferber, L. von und J. Behrens (Hrsg.). Public Health Forschung mit Gesundheits- und Sozialdaten. Stand und Perspektiven. St. Augustin. S. 27–40.
- Hoffmann, F. und A. Icks (2012). Unterschiede in der Versichertenstruktur von Krankenkassen und deren Auswirkungen für die Versorgungsforschung: Ergebnisse des Bertelsmann-Gesundheitsmonitors. In: Gesundheitswesen 74. S. 291–297.
- Horenkamp-Sonntag, D., Schneider, U., Engel, S., Linder, R., Gerste, B. und P. Ihle (2014). Validität von GKV-Routinedaten: In welchem Umfang muss bei der wissenschaftlichen Nutzung von Sekundärdaten die Daten-Qualität geprüft werden? In: Zeitschrift für Palliativmedizin 15, PD293. ISBN/ISSN 1615–2921. doi: 10.1055/s-0034-1374464.
- Ihle, P., Köster, I., Herholz, H., Rambow-Bertram, P., Schardt, T. und I. Schubert (2005). Versichertenstichprobe AOK Hessen/KV Hessen – Konzeption und Umsetzung einer personenbezogenen Datenbasis aus der Gesetzlichen Krankenversicherung. In: Gesundheitswesen 67. S. 638–645.
- Ihle, P., Köster, I., Schubert, I., Ferber, C. von und L. von Ferber (1999). GKV-Versichertenstichprobe. In: Wirtschaft und Statistik 1999 (9). S. 742–749.
- Kassenärztliche Bundesvereinigung (2021). Medizinische Informationsobjekte (MIO). Online unter [www.kbv.de/html/mio.php](http://www.kbv.de/html/mio.php) (Download am 15. Juli 2022).
- Scholten, N., Ihle, P. und H. Pfaff (2012) und mit den kooperierenden Krankenkassen (AOK Rheinland/Hamburg, BARMER, DAK-Gesundheit, pronova BKK); für das CoRe-Net Konsortium. Nachhaltige Infrastruktur für die Versorgungsforschung: Der Aufbau einer regionalen, krankenkassenübergreifenden GKV-Routinedatenbank. In: Das Gesundheitswesen 83. S. 463–469.
- Schubert, I., Hammer, A. und I. Köster (2017). Möglichkeiten zur Einschätzung des Schweregrades einer Erkrankung auf der Basis von Routinedaten am Beispiel des Schlaganfalls [Severity assessment strategies based on administrative data using



- stroke as an example]. In: Z Evid Fortbild Qual Gesundhwes. 126. S. 66–75. German. doi: 10.1016/j.zefq.2017.06.008. Epub 2017 Aug 12. PMID: 28807634.
- Schwefel, D., John, J., Potthoff, P. und W. von Eimeren (1987). Diagnosenstruktur in der ambulanten Versorgung. Explorative Auswertungen. Berlin, Heidelberg.
- Sozialgesetzbuch (2011). Fünftes Buch (V): Gesetzliche Krankenversicherung – GKV-Versorgungsstrukturgesetz (GKV-VStG) Gesetz vom 22. Dezember 2011 BGBl. I. S. 2983 (Nr. 70).
- Sozialgesetzbuch (2022). Das Fünfte Buch Sozialgesetzbuch – Gesetzliche Krankenversicherung – (Artikel 1 des Gesetzes vom 20. Dezember 1988. In: Bundesgesetzblatt I. S. 2477–2482), zuletzt durch Artikel 8 des Gesetzes vom 28. Juni 2022 (Bundesgesetzblatt I, S. 969) geändert.
- Swart, E., Gothe, H., Geyer, S., Jaunzeme, J., Maier, B., Grobe, T. G. und P. Ihle (2015). Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen, 3. Fassung, Version 2012/2014. In: Gesundheitswesen 77 (02). S. 120–126. doi: 10.1055/s-0034-1396815
- Swart, E., Gothe, H., Hoffmann, F., Ihle, P., Semler, S. C. und S. March (2012). Jetzt die Weichen stellen für ein leistungsfähiges Forschungsdatenzentrum Gesundheit. In: Gesundheitswesen 83 (Suppl. 2). S. 139–141.
- Swart, E., Ihle, P., Gothe, H. und D. Matusiewicz (Hrsg.) (2014). Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. 2. vollständig überarbeitete Auflage, Bern.
- TEHDAS (2020). Joint Action Towards the European Health Data Space – TEHDAS. Online unter <https://tehdas.eu/> (Download am 15. Juli 2022).