

Potentiale von und Empfehlungen zur Nutzung von GKV Routinedaten in einer pandemischen Versorgungslage – Erfahrungen aus dem Projekt egePan-Unimed des Netzwerk Universitätsmedizin (NUM)

Recommendations for the Utilization of Claims Data During a Pandemic: Lessons Learned from the Project EgePan-Unimed of the Netzwerk Universitätsmedizin (NUM)

Autoren

Josephine Jacob¹ , Jochen Walker¹, Enno Swart², Fabian Baum^{3, 4}, Martin Rößler⁵, Falko Tesch³, Felix Walther³, Hauke Felix Wiegand⁶, Peter Ihle⁷, Jochen Schmitt⁸

Institute

- 1 InGef – Institut für angewandte Gesundheitsforschung Berlin GmbH, InGef - Institut für angewandte Gesundheitsforschung Berlin GmbH, Berlin, Germany
- 2 Institut für Sozialmedizin und Gesundheitssystemforschung, Otto-von-Guericke-Universität Magdeburg Medizinische Fakultät, Magdeburg, Germany
- 3 Zentrum für Evidenzbasierte Gesundheitsversorgung, Universitätsklinikum und Medizinische Fakultät Carl Gustav Carus, Dresden, Germany
- 4 Qualitäts- und Medizinisches Risikomanagement, Universitätsklinikum Carl Gustav Carus Dresden, Dresden, Germany
- 5 Zentrum für Evidenzbasierte Gesundheitsversorgung, Universitätsklinikum Carl Gustav Carus Zentrum für Evidenzbasierte Gesundheitsversorgung, Dresden, Germany
- 6 Klinik für Psychiatrie und Psychotherapie, Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Mainz, Germany
- 7 PMV forschungsgruppe, Universität zu Köln, Köln, Germany
- 8 Zentrum für Evidenzbasierte Gesundheitsversorgung, Medizinische Fakultät Carl Gustav Carus der Technischen Universität Dresden, Dresden, Germany

Schlüsselwörter

GKV Routinedaten, Covid-19 Pandemie, Handlungsempfehlungen, Evidenz

Key words

claims data, Covid-19 pandemic, recommendations for action, evidence

online publiziert 28.09.2022

Bibliografie

Gesundheitswesen 2023; 85 (Suppl. 2): 171–177

DOI 10.1055/a-1915-4526

ISSN 0949-7013

© 2022. Thieme. All rights reserved.

Georg Thieme Verlag, Rüdigerstraße 14, 70469 Stuttgart, Germany

Korrespondenzadresse

Josephine Jacob
InGef – Institut für angewandte
Gesundheitsforschung Berlin GmbH
InGef – Institut für angewandte
Gesundheitsforschung Berlin GmbH
Spittelmarkt 12
10117 Berlin
Germany
josephine.jacob@ingef.de

ZUSAMMENFASSUNG

In der COVID-19 Pandemie wurden zur Pandemiesteuerung und zur Evidenzgenerierung für die Beantwortung drängender Fragen empirische Datengrundlagen zu verschiedensten Fragestellungen benötigt, die mit Hilfe von GKV Routinedaten generiert werden können. Im Rahmen des Projektes egePan-Unimed des Netzwerk Universitätsmedizin (NUM) wurden fünf pandemierelevante Fragestellungen mit GKV Routinedaten bearbeitet. Untersucht wurden die Prävalenz und Relevanz von Risikofaktoren für einen schweren COVID-19 Verlauf, die Hintergrundinzidenz der Sinusvenenthrombose und Myokarditis, die Häufigkeit und Ausprägung von Post COVID sowie die Versorgung von Personen mit psychiatrischer Erkrankung. Auf Basis dieser Anwendungsfälle wurden kontextspezifische Handlungsempfehlungen zur Nutzung von GKV Routinedaten für

zukünftige Pandemielagen abgeleitet. Diese umfassen u. a. die Nutzung etablierter Kooperationspartnerschaften und interdisziplinär zu besetzende Projektteams. Die datenschutzkonforme Zusammenführung von Ergebnissen verschiedener Analysestellen wird durch das (meta-analytische) Poolen aggregierter Routinedatenanalysen der verschiedenen Datenhalter möglich. So können GKV Routinedaten auch im Kontext der Pandemiesteuerung eine schnell verfügbare und valide Datenquelle zur empirischen Begründung angemessener Public Health Maßnahmen sein.

ABSTRACT

For appropriate response to the COVID-19 pandemic, and for obtaining answers to various relevant research questions, empirical data are required. Claims data of health insurances are a valid data source in such a situation. Within the project egePan-Unimed of the Netzwerk Universitätsmedizin (NUM) we investigated five

COVID-19-related research questions using German claims data of statutory health insurances. We studied the prevalence and relevance of risk factors for a severe course of COVID-19, the background incidence of cerebral venous sinus thrombosis and myocarditis, the frequency and symptoms of post-COVID as well as the care of people with a psychiatric condition during the COVID-19 pandemic. Based on these cases, context-specific recommendations regarding the use of German claims data for future pandemics or other public health emergencies were derived, namely that the utilization of established and interdisciplinary project teams enables a timely project start and furthermore, meta-analytic methods are a valuable way to pool aggregated results of claims data analyses when data protection regulations do not allow a consolidation of data sets from different statutory health insurances. Under these circumstances, claims data are a readily available and valid data source of empirical evidence base necessary for public health measures during a pandemic.

Einleitung

Mit Beginn der COVID-19 Pandemie im Februar 2020 wurde deutlich, dass zeitnah empirische Datengrundlagen zur Pandemiesteuerung benötigt werden würden. Dabei mussten sowohl Fragen zu den Risikofaktoren, zur Behandlung und Versorgung von COVID-19 Patient_innen als auch zur Aufrechterhaltung der Gesundheitsversorgung von nicht an COVID-19 erkrankten Personen beantwortet werden. Das übergeordnete Ziel des Arbeitspaketes 0 (AP0) – *Daten für die Pandemiesteuerung*)¹ im BMBF-Verbundprojekt egePan-Unimed des Netzwerk Universitätsmedizin (Förderkennzeichen 01KX2021) war die Bereitstellung potenziell relevanter Test-, Verlaufs- und Outcomedaten und ihre Bewertung im Hinblick auf ihre Eignung für ein (regionales) Pandemiemanagement.

Insbesondere die Nutzung und die Rolle von Sekundärdaten im Rahmen der Pandemiebekämpfung und des Pandemiemonitorings war zum Zeitpunkt des Projektbeginns nur eingeschränkt erfolgt. Zu Sekundärdaten zählen insbesondere aber nicht ausschließlich Abrechnungsdaten der gesetzlichen Krankenkassen. Im Vergleich zu einer Primärdatenstudie ermöglicht die Analyse von Sekundärdaten die Untersuchung großer Populationen mit vergleichsweise geringem Aufwand. Die nachfolgend benannten fünf Anwendungsfälle, die zwischen Sommer 2020 und Dezember 2021, zumeist aufgrund akuter Relevanz, bearbeitet wurden, liefern dafür Prototypen:

1. Schätzung der Prävalenzen von Risikofaktoren schwerer COVID-19 Verläufe

2. Hierarchisierung von Risikofaktoren für schwere COVID-19 Erkrankungsverläufe
3. Bestimmung der Hintergrundinzidenz von Sinusvenenthrombose und Myokarditis
4. Analysen zu Häufigkeit und Verlauf von Post-COVID
5. Untersuchungen zu Veränderungen der Inanspruchnahme und von leitlinienorientierten, sektorübergreifenden Versorgungssequenzen psychiatrisch-psychosomatisch-psychotherapeutischer Leistungen in Folge der COVID-19 Pandemie (Covid-Psy).

Diese sog. use cases bildeten, zusammen mit Ergebnissen aus Workshops, Recherchearbeiten und Interviews aus AP0, die Grundlage für die Ableitung von Empfehlungen im Sinne eines Leitfadens für die zukünftige Nutzung von Routinedaten in einer pandemischen Versorgungslage. Ziel dieses Papiers ist es die Anwendungsfälle inhaltlich und hinsichtlich der gewählten methodischen und organisatorischen Vorgehensweise zu beschreiben. Die im Projekt erarbeiteten und hier beschriebenen allgemeinen und spezifischen Handlungsempfehlungen sollen einen Rahmen für die erweiterte Nutzung von GKV Routinedaten, auch über die aktuelle pandemische Versorgungslage hinaus, geben. Im Fokus stehen dabei prozedurale, methodische, technische und rechtliche Aspekte, die in zukünftigen Pandemien oder sonstigen Situationen mit mangelhafter Evidenzlage handlungsleitend für die Evidenzgenerierung mittels GKV Routinedaten sein können.

Methodik

Die Bearbeitung der Fragstellungen erfolgte auf Basis von GKV Routinedaten in verschiedenen Projektteams. Im Folgenden soll das prozedurale, methodische und technische Vorgehen für jedes Teilprojekt skizziert werden.

Prävalenzen Risikofaktoren schwerer COVID-19 Verläufe

In Anbetracht begrenzter Impfdosen zu Beginn der Impfkampagne im Jahr 2021 war das Ziel eine größtmögliche Zahl von COVID-

¹ <https://egepan.de/>. Das übergeordnete Ziel des Projektes war die Ressourcensteuerung während der COVID-19 Pandemie innerhalb regionaler Gesundheitsnetzwerke bestehend aus Universitätskliniken und verantwortlichen Einrichtungen. Das Projekt war in acht Arbeitspakete mit folgenden Schwerpunkten strukturiert: AP0 – Daten für die Pandemiesteuerung; AP1 – Regionale Versorgungsstrukturen; AP2 Vorhersagemodelle; AP3 Ressourcenmanagement; AP4 Risikostratifizierung; AP5 Interdisziplinäre Konzepte und Maßnahmen-Strategien; AP6 Qualitäts- und medizinisches Risikomanagement; AP7 Zusammenführung und Verstetigungskonzept.

19 bezogenen Hospitalisierungen und Todesfällen zu vermeiden. Eine vorab definierte Priorisierungsstrategie sah vor, zunächst ältere und vorerkrankte Personen sowie Berufstätige in Pflege und Gesundheitsversorgung zu impfen [1]. Zur Modellierung des Nutzens verschiedener Impfstrategien wurden vom Robert Koch-Institut (RKI) Prävalenzraten bekannter Risikofaktoren für einen schweren COVID-19 Verlauf benötigt. Für die Analysen konnten anonymisierte Daten von insgesamt etwa acht Millionen GKV Versicherten der AOK PLUS Sachsen und von etwa 60 Betriebskrankenkassen aus dem gesamten Bundesgebiet genutzt werden. Die Auswertungen erfolgten im Zentrum für Evidenzbasierte Gesundheitsversorgung (ZEGV) und dem Institut für angewandte Gesundheitsforschung Berlin GmbH (InGef). Ambulante und in der stationären Versorgung dokumentierte Diagnosen (ICD-10 Code) sowie indikationsspezifische Arzneimittelverordnungen wurden zur Falldefinition von 27 Erkrankungen, die mit einem moderatem bzw. hohem Risiko für einen schweren COVID-19 Verlauf assoziiert waren [2], herangezogen. Für das Jahr 2019 wurden die Prävalenzraten von Risikofaktoren für einen moderaten bzw. schweren COVID-19 Verlauf stratifiziert nach Alter, Geschlecht und Bundesland berichtet. Außerdem wurden direkt altersstandardisierte Prävalenzraten der Risikofaktoren berechnet. Als Standardpopulation wurde die deutsche Bevölkerung zum 31.12.2019 genutzt [3].

Hierarchisierung von Risikofaktoren für schwere COVID-19 Erkrankungsverläufe

Ergänzend zur vorherigen Fragestellung benötigte das RKI zur Weiterentwicklung der Impfkampagne im Pandemieverlauf empirische Daten zur Quantifizierung der Relevanz bekannter Risikofaktoren. Besonderes Interesse galt der Frage, ob das zeitgleiche Auftreten mehrerer Risikofaktoren bei multimorbiden Patient_innen mit einem höheren Risiko für einen schweren COVID-19 Verlauf assoziiert ist.

Ziele dieses Teilprojektes waren 1) die Umsetzung der aktuellen Impfpflicht im Versorgungsalltag durch eine einfache Identifikation von Personen mit einem hohen Risiko für einen schweren COVID-19 Verlauf zu unterstützen und 2) die Entwicklung eines Scores, der das zeitgleiche Auftreten von mehreren Vorerkrankungen berücksichtigt.

Zur Bearbeitung dieser Fragestellung wurde eine Kohorte mit Personen, die zwischen 27.01.2020 und 30.06.2020 bzw. 31.12.2020 an COVID-19 (ICD-10 Code U.07.1!) erkrankt waren, aus GKV Routinedaten der folgenden Krankenkassen selektiert: AOK Bayern, AOK PLUS Sachsen, 60 Betriebskrankenkassen, BARMER, DAK und Techniker Krankenkasse (TK). Als Endpunkt der Analysen wurden schwere COVID-19-bedingte Erkrankungsverläufe anhand von COVID-19-assoziierter intensivmedizinischer Behandlung, Beatmung oder Tod im zeitlichen Zusammenhang mit einer COVID-19-Erkrankung definiert. Durch die Beteiligung vieler Krankenkasse konnten Daten der ambulanten und stationären Versorgung sowie ambulante Arzneimitteldaten von ca. 38 Mio. Versicherten für die Analyse genutzt werden.

Das absolute Risiko für einen schweren Verlauf wurde über den Anteil an Personen mit dem entsprechenden Risikofaktor, die eine COVID-19-assozierte Intensivbehandlung erhielten oder verstarben, geschätzt. Mittels meta-analytischer Verfahren wurden die

auf den Datensätzen der einzelnen Krankenkassen erzielten Schätzergebnisse durch das ZEGV gepoolt.

Zur Ableitung eines Risikoscores wurden altersstratifizierte Poisson-Regressionsmodelle mit robusten Standardfehlern basierend auf zufällig selektierten Trainingsdatensätzen (90 % der Studienpopulation) der einzelnen Datenhalter berechnet. Der finale Score wurde in den Testdatensätzen hinsichtlich seiner Performance, das Outcome *schwerer COVID-19 Verlauf* vorherzusagen, evaluiert. Basis für die Analysen war ein in der Projektgruppe zuvor abgestimmtes Studienprotokoll in dem die Selektion der Studienpopulation, die Definition der Kovariaten und Outcomes sowie die statistische Analysemethoden beschrieben wurden. Außerdem wurde eine einheitliche Datensatzbeschreibung entwickelt, um Variablenamen und -ausprägungen zu homogenisieren. Die Analysen erfolgten getrennt durch die jeweiligen Datenhalter bzw. analyseberechtigten Institutionen und wurden im Anschluss mittels meta-analytischer Verfahren gepoolt. Die Auswahl relevanter Interaktionen von Alter und Risikofaktoren basierte auf den Empfehlungen klinischer Expert_innen. Die so ausgewählten Interaktionen wurden in Regressionsanalysen mit den ausgewählten Interaktionstermen berücksichtigt. Im Anschluss erfolgt eine finale meta-analytische Poolung der Ergebnisse am ZEGV. Die Regressionskoeffizienten wurden mittels min-max Normalisierung auf Werte zwischen 0 und 20 normalisiert [4].

Hintergrundinzidenz Sinusvenenthrombose und Myokarditis

Im Verlauf der Impfkampagne mit der Vektor-Impfung von AstraZeneca traten mehrere Fällen von Sinusvenenthrombosen (SVT) auf. Nach einer Impfung mit dem mRNA Impfstoff von Biontech entwickelten wenige geimpfte Personen eine Myokarditis. Um beurteilen zu können, inwiefern es sich bei den beobachteten Erkrankungsfällen um unerwünschte Arzneimittelwirkungen (UAW) handelt, wurden Informationen zur Hintergrundinzidenz beider Erkrankungen von der Ständigen Impfkommission (STIKO) benötigt.

Die Falldefinitionen zur Identifikation von SVT und Myokarditiden in den GKV Routinedaten wurden zusammen mit klinischen Expert_innen bestimmt. Ambulant und stationär dokumentierte Diagnosen (ICD-10 GM) sowie ambulant abgerechnete Leistungen (EBM; Einheitlicher Bewertungsmaßstab) und stationäre erbrachte Prozeduren (OPS; Operationen- und Prozedurenschlüssel) wurden zur Falldefinition genutzt. Zusätzlich zur Ermittlung der Hintergrundinzidenz von SVT wurde der Anteil der SVT Patient_innen mit begleitender Thrombozytopenie bestimmt. Die Inzidenz der Myokarditis wurde sowohl in der Gesamtbevölkerung als auch in einer Population von COVID-19 Erkrankten untersucht. Als Datengrundlage dienten anonymisierte Routinedaten von etwa sieben Millionen BKK Versicherten der InGef Forschungsdatenbank in den Jahren 2015 bis 2020. Die Ergebnisausgabe erfolgte stratifiziert nach Alter und Geschlecht.

Epidemiologie von Post-COVID in Deutschland

Im Sommer 2021 fehlte Evidenz zur Häufigkeit, Schwere und Risikofaktoren zum Krankheitsbild Post-COVID, insbesondere bei Kindern und Jugendlichen, zur Planung angemessener und zielgruppenspezifischer Public Health Maßnahmen. Das RKI, zu diesem Zeitpunkt Mitglied einer interministeriellen Arbeitsgruppe (IMA), die

durch das Bundesministerium für Gesundheit geleitet wurde und an der das Bundesministerium für Bildung und Forschung; das Bundesministerium für Arbeit und Soziales; das Bundesministerium für Familie, Senioren, Frauen und Jugend; das Bundesministerium für Ernährung und Landwirtschaft; das Bundesministerium des Innern, für Bau und Heimat sowie das Bundeskanzleramt beteiligt waren, trat deshalb an die Arbeitsgruppe heran, um gemeinsam Fragestellungen zu Post-COVID bei Kindern und Jugendlichen wie auch bei Erwachsenen in GKV Routinedaten zu analysieren.

Die Ziele des Projektes waren zu untersuchen 1) ob eine COVID-19 Erkrankung mit einer erhöhten Wahrscheinlichkeit für das inzidente Auftreten ausgewählter Erkrankungen (z. B. chronische Lungenerkrankungen und verschiedene psychische Erkrankungen) assoziiert ist und 2) welche COVID-19 Folgeerkrankungen in bestimmten Subgruppen (insb. Kinder und Jugendliche) auftreten und 3) inwiefern der Schweregrad der COVID-19 Erkrankung einen Einfluss auf das Risiko von COVID-19 Folgeerkrankungen hat.

Unter Koordination des ZEGV wurde gemeinsam mit der AOK Bayern, der AOK PLUS, der BARMER, der DAK, dem InGef und der TK eine gematchte Kohortenstudie auf Basis von GKV Routinedaten von ca. 38 Mio. Versicherten durchgeführt. Mittels exaktem Matching sowie Propensity Score Matching (PSM) wurden Personen mit COVID-19 Erkrankung und Personen aus der Allgemeinbevölkerung ohne COVID-19 Erkrankung verglichen. Poisson Regressionsmodelle wurden genutzt um Incidence Rate Ratios für das inzidente Auftreten von COVID-19 Folgeerkrankungen zu bestimmen.

Covid-Psy - Routinedatenanalysen zu Veränderungen der Inanspruchnahme im psychiatrisch- psychosomatisch-psychotherapeutischen GKV-Versorgungssystem in Folge der COVID-19 Pandemie

Beispielhaft für vielfältige Fragestellungen, die sich aus der eingeschränkten Gesundheitsversorgung und Inanspruchnahme während des ersten und zweiten bundesweiten Lockdowns im Frühjahr und Winter 2020 ergaben, wurde in Kooperation mit der Forschungsgruppe psychiatrische Versorgungsforschung der Universitätsklinik Mainz die Versorgung psychisch erkrankter Personen in diesem Zeitraum untersucht.

Das Ziel dieser Studie war es, Veränderungen von Angebot und Inanspruchnahme sowie von beispielhaften leitlinienorientierten, sektorübergreifenden Versorgungssequenzen des psychiatrisch-psychotherapeutischen Versorgungssystems in der COVID-19-Pandemie unter besonderer Berücksichtigung der beiden bundesweiten Lockdowns zu beschreiben.

Unter Koordination des ZEGV wurden anonymisierte GKV Routinedaten von insgesamt 8,8 Millionen GKV Versicherten der AOK PLUS Sachsen und von 60 Betriebskrankenkassen für die Analysen genutzt. Die Analyse der Daten erfolgte separat durch die nutzungsberechtigten Institutionen ZEGV und InGef. Die aggregierten Ergebnisse der Analysen wurden abschließend an das ZEGV zur Evidenzsynthese versandt.

Mittels einer Zeitreihenanalyse wurde innerhalb einer Kohorte von Personen mit psychischer Erkrankung oder Verhaltensstörung zwischen Januar 2016 und März 2021 der Einfluss des ersten (03/2020) und des zweiten (12/2020) Lockdowns der COVID-19 Pandemie auf das akute stationäre und ambulante Versorgungsangebot untersucht. Die Modellierung der Zeitreihen erfolgt mittels eines informed time series fore-casting Ansatzes, der auf Likelihood-basier-

te Schätzmethoden für die Analyse und Modellierung von Zeitreihen nach verallgemeinerten linearen Modellen zurückgreift [5].

Ergebnisse

Arbeitsmodus

Sofern mehr als eine Institution bzw. Krankenkasse an der Bearbeitung einer der oben beschriebenen Fragestellungen beteiligt war, wurde eine Projektmanagementplattform (Confluence) installiert, auf der das in der Projektgruppe a priori abgestimmte Studienprotokoll sowie Referenztabellen abgelegt wurden. Zusätzlich wurden für jedes Projekt Datensatzbeschreibungen konsentiert, in der einheitliche Variablennamen und Ausprägungen festgelegt wurden. Die Analyse der Daten erfolgte getrennt durch die jeweiligen Datenhalter bzw. zugriffsberechtigten Institutionen. Im Anschluss wurden die aggregierten Ergebnisse gepoolt und mit allen Projektbeteiligten diskutiert.

Die Gute Praxis Sekundärdatenanalyse der Deutschen Gesellschaft für Epidemiologie [6] sowie die aktuelle internationale wissenschaftliche Evidenz sowie die medizinischen Expertise und Kenntnis der GKV Routinedaten im Konsortium waren die Grundlage für die erfolgreiche Bearbeitung der skizzierten Fragestellungen.

(Wissenschaftliche) Verwertung und Nutzung der Projektergebnisse

Auf eine detaillierte Darstellung der Ergebnisse der einzelnen Anwendungsfälle wird an dieser Stelle verzichtet. Diese werden in einem detaillierten Abschlussbericht² veröffentlicht. Vielmehr wird allgemein die Verwendung und wissenschaftliche Verwertung der Ergebnisse beschrieben.

Die Prävalenzschätzungen von Erkrankungen mit einem erhöhten Risiko für einen schweren COVID-19 Verlauf wurden dem RKI zur mathematischen Modellierung verschiedener Impfstrategien zur Verfügung gestellt.

Die Ergebnisse zur Schätzung der Hintergrundinzidenz der SVT wurden in der 4. Aktualisierung der COVID-19 Impfpflicht vom 22. April 2021 berücksichtigt [7]. Außerdem wurden die Ergebnisse im Anschluss bei PLOS ONE als Short Report eingereicht und vorab als pre-print veröffentlicht [8]. Die Ergebnisse zur Hintergrundinzidenz der Myokarditis fanden Eingang in die 9. Aktualisierung der COVID-19 Impfpflicht [9].

Im Epidemiologischen Bulletin 19 | 2021 [10] wurde im Mai 2021 eine Rangfolge von 24 Erkrankungen bzw. Altersgruppen, die mit einem höheren absoluten Risiko für einen schweren COVID-19-Verlauf assoziiert waren, veröffentlicht und im Anschluss der STIKO vorgestellt.

Die Ergebnisse zum Vergleich von Personen mit dokumentierter COVID-19-Erkrankung und Kontrollpersonen aus der Allgemeinbevölkerung hinsichtlich des Auftretens von COVID-19 bedingten Folgeerkrankungen (Post-COVID) wurden als Preprint veröffentlicht und bei einem Peer-Review Journal eingereicht [11].

² Der Abschlussbericht wird auf den Webseiten der beteiligten Partner InGef, ZEGV, Institut für Sozialmedizin und Gesundheitssystemforschung, Medizinische Fakultät der Otto-von-Guericke-Universität Magdeburg und PMV forschungsgruppe an der Universität zu Köln verfügbar sein

Weitere Veröffentlichungen der Projektergebnisse zur Entwicklung des Scores zur Impfpriorisierung sowie zur Versorgung von psychisch erkrankten Personen, während der COVID-19 Pandemie werden aktuell bearbeitet.

Aus den pandemiespezifischen Anwendungsfällen abgeleitete Handlungsempfehlungen

Zusätzlich zu den in der Guten Praxis Sekundärdatenanalyse [3] konkretisierten, allgemeinen Empfehlungen zur Planung, Vorbereitung, Durchführung, Auswertung und Berichterstattung von GKV Routinedatenanalysen ergaben sich im Rahmen der aus der Bearbeitung der oben beschriebenen Anwendungsfälle folgende kontextspezifische Handlungsempfehlungen:

- Fragestellungen zur akuten Pandemiebewältigung bzw. mit akuter Public Health Relevanz, die mittels GKV Routinedaten beantwortet werden können, sollten auf Basis ihrer Public Health und klinischen Relevanz sowie dem vorhandenen Grad an Unsicherheit priorisiert werden.
- Durch Nutzung etablierter Kooperationspartnerschaften kann kurzfristig Evidenz zu Fragestellungen mit großer Public Health Relevanz zur Verfügung gestellt werden. Zusätzlich zur zügigen Zusammenstellung einer Projektgruppe, sollte geprüft werden, inwiefern bereits aufbereitete Daten, die in anderen Projekten generiert wurden, als Datengrundlage genutzt werden können. Voraussetzung dafür bleibt die Zustimmung der zuständigen Aufsichtsbehörde.
- Die interdisziplinäre Zusammenarbeit von Personen mit epidemiologischer, medizinischer und biometrischer Expertise im Konsortium und die direkte Einbindung von Mitarbeitenden der Krankenkassen hat sich bewährt.
- Das Poolen aggregierter, deskriptiver GKV Routinedatenanalysen verschiedener Datenhalter und der Einsatz meta-analytischer Verfahren ermöglicht es, Ergebnisse verschiedener Analysestellen datenschutzkonform zusammenzuführen. Auf diese Weise können die Vorteile einer breiten Datenbasis genutzt werden, ohne eine zeitaufwändige und rechtlich problematische Zusammenführung der Individualdaten zu erfordern.
- Standardisierte Datensatzbeschreibungen und Variablenformate sowie die Abstimmung auf die verwendete Statistiksoftware erlauben die Entwicklung, Prüfung und Anwendung von einheitlichen Analyseskripten. Das ermöglicht eine zeitnahe und einheitliche Analyse der Daten mit geringerer Fehleranfälligkeit.
- Perspektivisch sollte geprüft werden inwiefern die Zeiträume der Datenbereitstellung von GKV Routinedaten und damit die Erhöhung der Aktualität der Ergebnisse möglich ist und welche Voraussetzungen für eine schnellere Datenverfügbarkeit von Daten aus einzelnen Versorgungssektoren erfüllt sein müssen. Denkbar wäre beispielsweise, dass zur Beantwortung von Fragen mit hoher Public Health Relevanz ggf. (nichtgeprüfte) Daten der Kassenärztlichen Vereinigungen (KVen) genutzt werden könnten.
- Bei der Etablierung kurzfristiger und niedrigschwelliger Finanzierungsmöglichkeiten von Projekten mit akuter Public Health Relevanz, sollte die Aufwandsentschädigung der Dateneigner, in diesem Fall der kooperierenden gesetzlichen Krankenversicherungen, mit bedacht werden.

- Durch die (einfache und zeitnahe) Verknüpfung von GKV Routinedaten und anderen Datenbestände (z. B. Meldedaten der COVID-19, Impfinformationen, Labordaten) könnten weiterführende Fragestellungen beantwortet werden, die als Evidenzgrundlage für das Pandemiemanagement eine sehr hohe Relevanz haben können. Dies gilt insbesondere für die Ermöglichung der Verknüpfbarkeit von Impfdaten mit GKV Daten, um die u. a. die Wirksamkeit und Sicherheit der Impfungen im Versorgungskontext bewerten zu können.

Diskussion

Die Bearbeitung vielfältiger Fragestellungen sowie deren Verwendung und wissenschaftliche Nutzung während der COVID-19 Pandemie reflektiert den großen Nutzen und die umfassenden Nutzungsmöglichkeiten von GKV Routinedaten – auch im Rahmen eines (regionalen) Pandemiemanagements. Aufgrund ihres eindeutigen Bevölkerungs- und Personenbezugs sind GKV Routinedaten eine verlässliche Quelle zur Prävalenzschätzung von Risikofaktoren für einen schweren COVID-19 Verlauf oder von schweren Erkrankungen, die als mögliche Impfnebenwirkung diskutiert werden und für deren Prävalenz es in der Allgemeinbevölkerung keine oder keine ausreichende Datengrundlage gibt. Das Risiko von Verzerrungseffekten wie sie beispielweise durch Selbstselektion, Dropout, Recall- oder Interviewerbias entstehen könnten, kann durch die Nutzung von GKV Routinedaten insgesamt als gering eingestuft werden [12, 13]. Durch die Beteiligung verschiedener Krankenkassenarten und der damit einhergehenden großen Datenbasis kommt dieser grundsätzliche Vorteil von GKV Routinedaten in diesem Projekt besonders zum Tragen.

Die längsschnittliche und personenbezogene Abbildung von Krankheitsverläufen erlaubt die Nutzung von GKV Routinedaten zur Identifikation von Risikogruppen für einen schweren COVID-19 Verlauf und von COVID-19 assoziierten Spätfolgen (Post-COVID). Empfehlungen zur Versorgung von Personen, die nicht unmittelbar von einer COVID-19 Infektion betroffen sind, lassen sich aus dem Teilprojekt zur Behandlung psychiatrisch erkrankter Personen ableiten. Die im Teilprojekt APO des BMBF geförderten Projektes egePan-Unimed untersuchten Fragestellungen konnten sowohl einen Beitrag zur Beantwortung von Fragestellungen mit direktem Bezug (z. B. Risikofaktoren für und Folgen einer COVID-19 Infektion) als auch mit indirektem Bezug (veränderte Versorgung von psychiatrisch erkrankten Personen) zur COVID-19 Pandemie, leisten.

Limitationen von GKV Routinedaten im Rahmen des Pandemiemanagements

Die noch breitere Nutzung von Routinedaten zur Pandemiebewältigung ist beschränkt durch die generellen Limitationen, denen diese Daten unterliegen [12–14]. Besonders hervorzuheben ist in diesem Zusammenhang die Verzögerung, mit der die Abrechnungsdaten von ambulanten Leistungserbringern zur Verfügung stehen. Hier können die quartalsbezogenen Daten regelhaft erst ca. 4 bis 6 Monate nach Abschluss des Quartals genutzt werden, so dass beispielsweise eine Leistung, die im Januar in einer Praxis erbracht wurde, frühestens im dritten Quartal des Jahres bei der Krankenkasse verfügbar ist. Aufgrund dessen konnten für das Post-COVID Projekt ausschließlich Personen mit COVID-19 Infektion im ersten Halbjahr

2020 eingeschlossen werden, um einen Nachbeobachtungszeitraum von einem halben Jahr zu gewährleisten. Eine beschleunigte Bereitstellung von Routinedaten zur Bearbeitung wissenschaftlicher Fragestellungen mit akuter Public Health Relevanz wäre aus Sicht der Autor_innen wünschenswert.

Ferner enthalten GKV Routinedaten keine klinischen Informationen über die nach ICD-, OPS, ATC- und anderen Klassifizierungssystemen dokumentierten Entitäten hinaus. Die Identifikation von Personen mit COVID-19 Infektion aus den GKV Routinedaten hängt daher stark von der korrekten Dokumentation der Erkrankung von den Leistungserbringern ab. Hinzu kommt im Fall von COVID-19 die neue Klassifikationsziffer im ICD-Katalog, die sich erst schrittweise etablieren musste.

Auch zur Überwachung des Impfstoffsicherheit wäre die Nutzung von GKV Routinedaten denkbar. Wünschenswert wäre es, Informationen zum Impfstatus und den verabreichten Impfstoffen von Personen, die in den Impfzentren, einschließlich der bei ihnen angegliederten mobilen Impfteams, in Arztpraxen, die nicht an der vertragsärztlichen Versorgung teilnehmen und bei beauftragten Betriebsärzten oder die durch andere Stellen geimpft wurden, an die gesetzlichen Krankenkassen zu übermitteln. Aktuell liegen diese Informationen nur dem RKI vor. Das schränkt die Nutzbarkeit der GKV Routinedaten zur Impfsurveillance und zur Sicherheitsüberwachung der COVID-19 Impfstoffe erheblich ein. Hier bedürfte es einer konsentierten Ergänzung der gesetzlich an die Kostenträger zu übermittelnden Abrechnungsdaten.

Anzustreben ist die Schaffung datenschutzkonformer und praktikabel umsetzbarer Lösungen zur Verknüpfung von GKV Routinedaten mit anderen versorgungsnahen Daten wie Registerdaten, Klinikdaten und Daten des öffentlichen Gesundheitsdienstes (ÖGD). Die Ergänzung der GKV Daten um Outcomedaten und weitere klinische Daten hätte ein großes Potenzial, um zukünftig noch verlässlichere Evidenz in pandemischen wie auch in nicht pandemischen Zeiten aus den GKV Daten abzuleiten.

FAZIT FÜR DIE PRAXIS

GKV Routinedaten sind auch im Kontext des Pandemiemonitorings und der Evidenzgenerierung zur Pandemiesteuerung eine z. B. im Vergleich zu traditionellen Kohortenstudien schnell verfügbare und valide Datenquelle zur Ableitung angemessener Public Health Maßnahmen. Voraussetzung für die Bereitstellung valider Ergebnisse zur Pandemiesteuerung unter Nutzung von GKV Routinedaten erfordert Kenntnis der Stärken und Limitationen der Daten sowie umfassende Erfahrung in der Aufbereitung und Analyse der Daten. Die Kooperation von Wissenschaftler_innen an Universitäten, forschenden Instituten und bei gesetzlichen Krankenkassen hat sich bewährt, um bestmögliche Evidenz rasch verfügbar zu machen. Das Netzwerk Universitätsmedizin hat einen angemessenen Rahmen geboten, um effizient methodische Expertise mit klinischer Expertise der Universitätsmedizin zu verbinden und so rasch drängende Fragen zur Patientenversorgung in der COVID-19 Pandemie und zum Pandemiemanagement beantworten zu können.

Finanzielle Unterstützung

Das Projekt egePan-Unimed wurde vom Bundesministerium für Bildung und Forschung durch das „Nationale Netzwerk der Universitätsmedizin zu COVID-19“ gefördert (FK 01KX2021).

Danksagungen

Die Autorinnen und Autoren danken Jakob Peter Armann (Klinik für Kinder- und Jugendmedizin Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden), Pedro Ballesteros (bifg), Manuel Batram (Vandage GmbH), Stefan Baßler (AOK PLUS), Reinhard Berner (Klinik für Kinder- und Jugendmedizin Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden), Barbara Bertele (Techniker Krankenkasse), Dagmar Hertle (bifg), Roland Jucknewitz (AOK Bayern), Christina König (Techniker Krankenkasse), Friedrich Loser (Techniker Krankenkasse), Oliver Nagel (InGef), Uwe Repschläger (bifg), Nico Richter (AOK plus), Cordula Riederer (DAK), Christa Scheidt-Nave (Robert Koch-Institut), Stefan Scholz (Robert Koch-Institut), Anja Schramm (AOK PLUS), Claudia Schulte (bifg), Martin Seifert (ZEGV), Franziska Sobik (DAK), Stefan Stern (AOK Bayern), Nicole Toepfner (Klinik für Kinder- und Jugendmedizin Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden), Marina Treskova-Schwarzbach (Robert Koch-Institut), Annika Vivirito (InGef), Oliver Weidinger (AOK Bayern), Danny Wende (bifg) sowie Lothar Wieler (Robert Koch-Institut) für die Unterstützung, Begleitung und Durchführung des Vorhabens. Weiterhin möchten die Autorinnen und Autoren der AOK Bayern, dem bifg – BARMER Institut für Gesundheitssystemforschung, der DAK Gesundheit, dem Robert Koch-Institut sowie der Techniker Krankenkasse für die Unterstützung des Projektes danken.

Interessenkonflikt

JS erhielt institutionelle Fördermittel für wissenschaftlich-initiierte Forschung vom GB-A, dem BMG, BMBF, Freistaat Sachsen, Novartis, Sanofi, ALK und Pfizer. Er nahm als Berater an Advisory Board Meetings der Firmen Sanofi, Lilly und ALK teil und erhielt hierfür ein persönliches Honorar.

Die anderen Autor*innen erklären keine Interessenskonflikte.

Literatur

- [1] Bundesministerium für Gesundheit. Verordnung zum Anspruch auf Schutzimpfung gegen das Coronavirus SARS-CoV-2 (Coronavirus-Impfverordnung – CoronImpfV). 2021
- [2] Treskova-Schwarzbach M, Haas L, Reda S et al. Pre-existing health conditions and severe COVID-19 outcomes: an umbrella review approach and meta-analysis of global evidence. *BMC Med* 2021; 19: 212. doi:10.1186/s12916-021-02058-6
- [3] Destatis. Statistisches Bundesamt Deutschland – GENESIS-Online. Ergebnis 12411-0013 – Bevölkerung: Bundesländer, Stichtag, Geschlecht, Altersjahre 2019; Im Internet: <https://www-genesis.destatis.de/genesis/online?operation=abrufabelle&levelind ex=1&levelid=1642498201869&auswahloperation=abrufabelle&auspraeung=Auswaehlen&auswahlverzeichnis=ordnungsstruktur&auswahlziel=werteabruf&code=12411-0013&auswahltext=&werteabruf=Werteabruf#abreadcrumb> Stand: 18.01.2022

- [4] Patro SGK, Sahu KK. Normalization: A Preprocessing Stage. *International Advanced Research Journal in Science, Engineering and Technology* 2015; 20–22. doi:10.17148/IARJSET.2015.2305
- [5] Liboschik T, Fokianos K, Fried R. tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models. *Journal of Statistical Software* 2017; 82: 1–51. doi:10.18637/jss.v082.i05
- [6] Swart E, Gothe H, Geyer S et al. Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen. *Das Gesundheitswesen* 2015; 77: 120–126. doi:10.1055/s-0034-1396815
- [7] Koch-Institut R *Epidemiologisches Bulletin* 16/2021 2021; 76:
- [8] Jacob J, Klamroth R, Ploner T et al. Incidence of cerebral venous sinus thrombosis in adults in Germany – a retrospective study using health claims data. *Research Square* 2021. doi:10.21203/rs.3.rs-428469/v2
- [9] Koch-Institut R *Epidemiologisches Bulletin* 33/2021 2021; 56:
- [10] Rößler M, Jacob J, Risch L et al. Hierarchisierung von Risikofaktoren für schwere COVID-19-Erkrankungsverläufe im Kontext der COVID-19-Schutzimpfungen 2021. doi:10.25646/8405
- [11] Roessler M, Tesch F, Batram M et al. Post COVID-19 in children, adolescents, and adults: results of a matched cohort study including more than 150,000 individuals with COVID-19. *medRxiv* 2021. doi:10.1101/2021.10.21.21265133
- [12] Schubert I, Köster I, Küpper-Nybelen J et al. Versorgungsforschung mit GKV-Routinedaten: Nutzungsmöglichkeiten versichertenbezogener Krankenkassendaten für Fragestellungen der Versorgungsforschung. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 2008; 51: 1095–1105. doi:10.1007/s00103-008-0644-0
- [13] Ohlmeier C, Frick J, Prütz F et al. Nutzungsmöglichkeiten von Routinedaten der Gesetzlichen Krankenversicherung in der Gesundheitsberichterstattung des Bundes. *Bundesgesundheitsbl* 2014; 57: 464–472. doi:10.1007/s00103-013-1912-1
- [14] Schubert I, Ihle P, Köster I et al. Datengutachten für das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI). Daten für die Versorgungsforschung. Zugang und Nutzungsmöglichkeiten. Köln 2014