



egePan  
Unimed

Förderkennzeichen: 01KX2021

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

# *Erfahrungen, Potenziale und Empfehlungen zur Nutzung von Routinedaten zur Evidenzgenerierung in einer pandemischen Versorgungssituation*

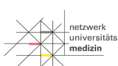
*Wissenschaftliche Ergebnisse aus dem NUM-Projekt  
egePan Unimed*

## **Autoren:**

Prof. Dr. med. Jochen Schmitt, MPH<sup>a</sup> | PD Dr. rer. biol. hum. Enno Swart<sup>b</sup> | Felix Walther, M. Sc.<sup>a, c</sup>  
Dr. med. Jochen Walker<sup>d</sup> | Dr. rer. pol. Martin Röbber, Dipl.-Volksw.<sup>a</sup> | Falko Tesch, M. Sc.<sup>a</sup>  
Dr. rer. nat. Fabian Baum Dipl.-Psych.<sup>a</sup> | Hauke Felix Wiegand M.D./Ph. D.<sup>e</sup> | Josephine Jacob, M. Sc.<sup>d</sup> | Peter Ihle<sup>f</sup>

## **Datum:**

März 2022



Universitätsklinikum  
Carl Gustav Carus



GOETHE  
UNIVERSITÄT  
FRANKFURT AM MAIN



ingef  
Evidenz-  
Basiertes  
Optimieren

---

#### **Affiliationen**

<sup>a</sup>Zentrum für Evidenzbasierte Gesundheitsversorgung, Universitätsklinikum und Medizinische Fakultät Carl Gustav Carus, Dresden

<sup>b</sup>Institut für Sozialmedizin und Gesundheitssystemforschung, Medizinische Fakultät der Otto-von-Guericke-Universität Magdeburg, Magdeburg

<sup>c</sup>Qualitäts- und Medizinisches Risikomanagement, Universitätsklinikum Carl Gustav Carus, Dresden

<sup>d</sup>InGef – Institut für angewandte Gesundheitsforschung Berlin GmbH

<sup>e</sup>Klinik für Psychiatrie und Psychotherapie, Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Mainz

<sup>f</sup>PMV forschungsgruppe an der Medizinischen Fakultät und Uniklinik Köln, Universität zu Köln, Köln

## Inhaltsverzeichnis

<b>1. Zusammenfassung / Executive Summary</b> .....	<b>4</b>
<b>2. Einleitung</b> .....	<b>7</b>
<b>3. Use Cases</b> .....	<b>10</b>
3.1 Use Case 1: Risikofaktoren für einen schweren COVID-19-Verlauf .....	10
3.2 Use Case 2: Hierarchisierung von Risikofaktoren für schwere COVID-19-Erkrankungsverläufe .....	13
3.3 Use Case 3: Hintergrundinzidenz Sinusvenenthrombosen und Myokarditis .....	17
3.4 Use Case 4: Post-COVID bei Kindern und Jugendlichen .....	21
3.5 Use Case 5: COVID-Psy-Routinedatenanalysen zu Veränderungen der Inanspruchnahme im psychiatrisch-psychosomatisch-psychotherapeutischen GKV-Versorgungssystem infolge der COVID-19-Pandemie .....	25
<b>4. Handlungsempfehlungen und Fazit</b> .....	<b>29</b>
4.1 Allgemeine Handlungsempfehlungen .....	29
4.2 Spezifische Handlungsempfehlungen in pandemischer Versorgungslage .....	31
4.3 Fazit .....	32
<b>5. Quellen</b> .....	<b>33</b>

## 1. Zusammenfassung / Executive Summary

Abrechnungsdaten der gesetzlichen Krankenversicherung (GKV-Routinedaten) werden in Deutschland seit über 20 Jahren für die wissenschaftliche Bearbeitung von Fragestellungen der Versorgungsforschung und der Epidemiologie genutzt. Da GKV-Routinedaten primär für Dokumentations- und Abrechnungszwecke erhoben werden, erfordert ihre wissenschaftliche Nutzung und die Generierung valider Ergebnisse insbesondere die Erfüllung datenschutzrechtlicher und administrativer Auflagen sowie spezifische methodische Verfahren bei Datenprüfung, -aufbereitung und -analyse. Hierfür existieren Leitlinien und methodische Standards wie die Gute Praxis Sekundärdatenanalyse, Gute Praxis Datenlinkage und der Berichtsstandard STROSA.

Im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Netzwerks Universitätsmedizin (NUM) beschäftigte sich das Projekt egePan Unimed mit den Potenzialen und Besonderheiten der Nutzung von GKV-Routinedaten zur Evidenzgenerierung in einer pandemischen Versorgungssituation. Im Rahmen des Projekts wurden zunächst wesentliche Stärken und Schwächen von GKV-Routinedaten hinsichtlich ihrer Nutzung im Kontext einer Pandemie bewertet, darauf aufbauend fünf Use Cases betrachtet und umgesetzt und schließlich Handlungsempfehlungen für die zukünftige Nutzung und Nutzbarmachung von GKV-Routinedaten in Bezug auf eine Pandemiebewältigung und auch Pandemievorsorge/ Pandemic Preparedness abgeleitet.

Eine wesentliche Stärke von GKV-Routinedaten ist, dass sie ein vollständiges, standardisiertes Abbild der Inanspruchnahme medizinischer Leistungen, Verordnungen und Diagnosen für die Versicherten einer oder mehrerer gesetzlicher Krankenkassen liefern. Die Versichertenpopulation lässt sich nach Alter, Geschlecht, Region und anderen versichertenbezogenen Merkmalen klassifizieren und bietet so – insbesondere bei Analysen von Daten mehrerer Krankenkassen – ein weitgehend repräsentatives Abbild der Gesamtbevölkerung. GKV-Routinedaten können sowohl fall- wie versichertenbezogen querschnittlich und längsschnittlich analysiert werden, somit lassen sich individuelle Krankheitsverläufe wie zeitliche Entwicklung inzidenter Ereignisse und prävalenter Zustände abbilden. Eine Limitation von GKV-Routinedaten ist, dass Informationen fehlen, die nicht in den Dokumentations- und Klassifizierungssystemen enthalten sind. Dies kann beispielsweise den Schweregrad einer Erkrankung, versichertenbezogene Gesundheitsdeterminanten wie Ernährungsgewohnheiten oder Umweltexpositionen betreffen. Ein großer Vorteil von GKV-Routinedaten gerade für Fragestellungen in einer Pandemie ist, dass sie relativ kurzfristig verfügbar sind: Daten der stationären Versorgung werden kontinuierlich

abgerechnet und stehen den Krankenkassen unmittelbar nach Übermittlung durch die Krankenhäuser zur Verfügung. Aus dem ambulanten Bereich sind quartalsbezogene Daten mit einer Latenz von vier bis sechs Monaten nach Abschluss des Quartals verfügbar. GKV-Routinedaten bieten daher gerade für Pandemiezeiten ein sehr hohes Potenzial zur raschen, kosteneffizienten Evidenzgenerierung und zum Monitoring von Pandemieeffekten bezogen auf die Gesundheitsversorgung.

Während der Projektlaufzeit wurden immer wieder von Institutionen des öffentlichen Gesundheitswesens wie dem Robert Koch-Institut (RKI) einerseits und von Klinikern aus dem NUM andererseits hochrelevante Fragen an die Projektgruppe herangetragen. In der Projektgruppe bestanden bereits Kooperationen mit Krankenkassen zur wissenschaftlichen Nutzung von GKV-Routinedaten sowie eine hohe methodische und praktische Expertise in der Durchführung von GKV-Routinedatenanalysen, sodass fünf Use Cases während der Projektlaufzeit bearbeitet werden konnten. Die Use Cases werden in diesem Bericht ausführlich beschrieben und bilden die Grundlage für die Empfehlungen zur Nutzbarmachung und Nutzung von GKV-Routinedaten für die Evidenzgenerierung in pandemischen Situationen. Diese Empfehlungen sollten unbedingt auch für eine zukünftige Pandemic Preparedness nach Bewältigung der derzeitigen COVID-19-Pandemie berücksichtigt werden.

### Die Use Cases beschäftigten sich mit Fragestellungen

1. prävalenter Risikofaktoren eines schweren COVID-19-Verlaufs,
2. der Hierarchisierung von Risikofaktoren als Grundlage für die Impfpriorisierung zu Beginn der Impfkampagne,
3. der Hintergrundinzidenz von Sinusvenenthrombosen und Myokarditis zur Bewertung der Sicherheit der COVID-19-Impfung,
4. Post-COVID bei Kindern und Jugendlichen und
5. der medizinischen Versorgung von Menschen mit psychischen Erkrankungen während der Pandemie.

Ein Teil der Use Cases wurde in Zusammenarbeit mit dem RKI durchgeführt und es konnte ein Konsortium von gesetzlichen Krankenversicherungen gebildet werden, das insgesamt Daten von rund 38 Millionen Versicherten für die Analysen einbrachte. Die Ergebnisse dieser Use Cases wurden mehrfach in der Ständigen Impfkommission (STIKO) am RKI vorgestellt und gingen in deren Empfehlungen ein.

Wichtige Erfolgsfaktoren für die Use Cases waren u.a. **(1)** die interdisziplinäre Zusammenarbeit von epidemiologischen, medizinischen und methodischen Expert:innen im

Konsortium, **(2)** die partnerschaftliche Zusammenarbeit von Krankenkassen, Versorgungsforscher:innen, Kliniker:innen und dem RKI, **(3)** die große Vorerfahrung der beteiligten Institute zu allen Belangen der Verarbeitung, Operationalisierung, Analyse und Interpretation von GKV-Routinedaten, **(4)** die Koordination der Auswertung an mehreren Analysestellen mit Einsatz metaanalytischer Verfahren als rasch umsetzbare, datenschutzkonforme Möglichkeit kassenübergreifender Analysen sowie (5) das Aufsetzen auf bestehenden Kooperationen und Abläufen in Bezug auf die notwendigen administrativen Voraussetzungen zur Nutzung dieser Daten.

GKV-Routinedaten haben sich als eine wesentliche Säule zur Evidenzgenerierung und zur Beantwortung dringender Fragen zum Pandemiemanagement erwiesen. Valide Analysen erfordern die Zusammenarbeit von Wissenschaftler:innen mit breiter Erfahrung im Umgang mit GKV-Daten, von Kliniker:innen mit Kenntnissen zur Kodierpraxis und klinischer Expertise sowie von Krankenkassen oder anderen Institutionen zur zeitnahen Bereitstellung von GKV-Routinedaten. Auch für eine zukünftige Pandemic Preparedness sollte an einer weiteren Verkürzung der Verfügbarmachung der Daten für Fragestellungen der Versorgungsforschung gearbeitet werden. Es sollte sichergestellt werden, dass auch weitere relevante Daten aus der Gesundheitsversorgung mit den GKV-Daten verknüpfbar sind. Aktuell fehlt eine direkte Verknüpfbarkeit der Daten von Impfungen gegen das SARS-CoV-2-Virus mit GKV-Abrechnungsdaten, sodass wichtige Fragestellungen, etwa zur Impfsurveillance, zur Wirksamkeit der Impfungen und zur Sicherheit der Impfungen in der Bevölkerung, nicht auf Basis der GKV-Routinedaten bearbeitbar sind. Hier sollte durch den Gesetzgeber möglichst rasch eine Möglichkeit zur Verknüpfung der Daten geschaffen werden. Weitere Empfehlungen finden sich im nun folgenden ausführlichen Bericht.

Das Netzwerk Universitätsmedizin bietet mit seiner breiten klinischen, methodischen und versorgungswissenschaftlichen Expertise die ideale Voraussetzung zur Bildung geeigneter Infrastrukturen für die Nutzung von GKV-Routinedaten zum Pandemiemanagement. Dadurch konnten in der gegenwärtigen Pandemie wichtige Fragestellungen mittels GKV-Routinedaten bearbeitet werden, so konnte zu einer möglichst guten Pandemiebewältigung beigetragen werden. Perspektivisch sollten durch das Netzwerk Universitätsmedizin im Sinne der Pandemic Preparedness strategische Partnerschaften mit gesetzlichen Krankenkassen und/oder mit Institutionen mit Zugriff auf GKV-Daten geschlossen und Infrastrukturen zur raschen Priorisierung und Umsetzung von GKV-Routinedatenanalysen etabliert werden.

Wir erhoffen uns, mit diesem Bericht Impulse für die zukünftige effektive und evidenzbasierte Pandemiesteuerung setzen zu können.

*Prof. Dr. med. Jochen Schmitt, MPH | PD Dr. rer. Biol. Hum. Enno Swart,  
Josephine Jacob, M. Sc. | Peter Ihle*

## 2. Einleitung

Routinedaten, hier insbesondere die Routinedaten der gesetzlichen Krankenversicherung (GKV-Routinedaten), haben sich in den letzten 20 Jahren zu einer etablierten Datengrundlage für die Versorgungsforschung entwickelt.<sup>1,2</sup> Auch in der Epidemiologie werden sie im Rahmen von sogenannten Sekundäranalysen zunehmend als systematische Ergänzung von Primärdaten genutzt.<sup>3-5</sup> Gefördert wird dies durch die Entwicklung und laufende Fortschreibung spezifischer methodischer Standards wie die Gute Praxis Sekundärdatenanalyse,<sup>6</sup> Gute Praxis Datenlinkage<sup>7</sup> und den Berichtsstandard STROSA,<sup>8</sup> unterstützt durch spezifische methodische Verfahren, die dem besonderen Charakter dieser Daten gerecht werden.<sup>9</sup>

Der größte Vorteil von GKV-Routinedaten erschließt sich aus ihrer Bezeichnung per se: Diese Daten werden zu Abrechnungszwecken routinemäßig generiert. Hieraus ergibt sich auch ihr Potenzial bei spontan auftretenden Herausforderungen in der Gesundheitsversorgung, beispielsweise in pandemischen Versorgungslagen. Vermehrte oder geänderte Inanspruchnahme in der ambulanten und stationären Versorgung bildet sich ohne aufwendige Modellierung direkt in den Routinedaten ab. Dieser grundsätzliche Vorteil muss jedoch im Kontext der Nutzungsmöglichkeiten von Routinedaten auch unter Einbeziehung der vorhandenen Limitation kritisch bewertet werden. Der vorliegende Projektbericht hat vor diesem Hintergrund Empfehlungen im Sinne eines Leitfadens für die Nutzung von Routinedaten in pandemischen Versorgungslagen erarbeitet. Grundlage für die Bewertungen waren vor allem fünf Anwendungsfälle / Use Cases, die im Rahmen des egePan-Projekts durchgeführt bzw. methodisch begleitet wurden. Zusätzlich gingen Ergebnisse aus Workshops, Rechercharbeiten und Interviews aus dem Arbeitspaket 0 (AP 0) des Projekts egePan Unimed ein.

### Potenziale und Limitation von Routinedaten

Zur Einordnung der weiteren Ausführungen müssen die grundsätzlichen Potenziale und Limitationen der infrage kommenden KV-Routinedaten beachtet werden. **Die Vorteile der GKV-Daten**, im Vergleich zu klinischen und epidemiologischen Daten, liegen darin, dass sie ein vollständiges, standardisiertes Abbild der Inanspruchnahme medizinischer Leistungen für eine definierbare und quantifizierbare Population, nämlich die Versicherten einer oder mehrerer gesetzlicher Krankenkassen, liefern. Vollständig in dem Sinne, dass alle zulasten einer Kasse erbrachten Leistungen nach detaillierten Vorschriften des Sozialgesetzbuches und Ausführungsbestimmungen des Spitzenverbandes Bund der Kassen von den Leistungserbringern dokumentiert und an die Kostenträger übermittelt werden. Die Versichertenpopulation einer Kasse lässt sich nach Alter, Geschlecht, Region und anderen versichertenbezogenen Merk-

malen eindeutig als Risikopopulation für die Ableitung epidemiologischer Maßzahlen bestimmen. Weiterhin lassen sich die Daten sowohl fall- wie versichertenbezogen querschnittlich und längsschnittlich analysieren, somit können individuelle Krankheitsverläufe wie zeitliche Entwicklung bezüglich inzidenter Ereignisse und prävalenter Zustände abgebildet werden. Mögliche kassenspezifische Unterschiede oder Besonderheiten hinsichtlich der Versichertenklientelen lassen sich i. d. R. durch das Poolen von Daten mehrerer Krankenkassenarten weitgehend ausgleichen.

Beim Einsatz der GKV-Daten aufgrund dieser Potenziale müssen jedoch auch **grundsätzliche Limitationen** beachtet und gegen die Vorteile abgewogen werden. So enthalten die Daten keine klinischen Informationen jenseits der nach ICD-, OPS-, ATC- und anderen Klassifizierungssystemen dokumentierten Entitäten. Es fehlen somit z. B. weitgehend dokumentierte Krankheitsschweregrade oder Befunde bzw. Ergebnisse von diagnostischen oder therapeutischen Leistungen. Die Daten enthalten zudem kaum Informationen zu epidemiologischen Risikofaktoren, ebenso sind die Angaben zu sozioökonomischen Faktoren begrenzt. Eine Risikostratifizierung der Patient:innen nach klinisch relevanten Merkmalen ist damit erschwert. Zu beachten ist, dass GKV-Routinedaten naturgemäß nur für gesetzlich krankenversicherte Personen zur Verfügung stehen (gut 73 Millionen, d. h. 88 % der etwas mehr als 83 Millionen Einwohner in Deutschland; Stand Ende 2020).<sup>10</sup> Analysen von GKV-Daten bedürfen also der Reflexion darüber, inwieweit deren Ergebnisse auf die Gesamtbevölkerung inkl. der privat krankenversicherten Personen übertragen werden können.

Daten der stationären Versorgung werden kontinuierlich abgerechnet, stehen bei den Krankenkassen unmittelbar nach Übermittlung durch die Abrechnungsstellen der Krankenhäuser zur Verfügung und sind dann für Analysen grundsätzlich nutzbar. Andere Sektoren, wie beispielsweise der ambulante ärztliche Sektor, sind abrechnungstechnisch mit größeren Verzögerungen behaftet. Hier stehen die quartalsbezogenen Daten ca. vier bis sechs Monate nach Abschluss des Quartals zur Verfügung, sodass beispielsweise eine Leistung, die im Januar in einer Praxis erbracht wurde, erst frühestens im dritten Quartal des Jahres bei der Krankenkasse verfügbar ist. Diese Verzögerung ist mitbedingt dadurch, dass auch die Kassenärztlichen Vereinigungen als erster Datenadressat die Daten erst nach Ablauf des Abrechnungsquartals, also im genannten Beispiel frühestens im April mit einer Verzögerung von drei Monaten, erhalten. Vor dem Hintergrund einer pandemischen Versorgungslage muss hier über andere Dokumentations- und Übermittlungswege an die in der Pandemie maßgeblichen Institutionen nachgedacht werden.



**Die Planung, Vorbereitung, Durchführung, Auswertung und Berichterstattung bei Analysen von GKV-Routinedaten lässt sich in Konkretisierung der Guten Praxis Sekundärdatenanalyse in nachfolgende praktische Schritte unterteilen:**

1. Erstellung eines Forschungsplans;
2. Prüfung und Erschließung potenzieller Daten;
3. Erstellung eines realistischen Zeitplans;
4. Einholung notwendiger Genehmigungen;
5. Ausarbeitung vertraglicher Grundlagen;
6. Schaffung ausreichender personeller und finanzieller Ressourcen;
7. Ausgestaltung eines methodischen Handlungsleitfadens;
8. Umsetzung spezifischer Methoden und Verfahren;
9. Erstellung einer Dokumentation der Datenaufbereitungs- und -analyseschritte;
10. Berichtserstattung unter Nutzung von STROSA.<sup>8</sup>

In dem während des Projekts egePan Unimed zusammengestellten und zweimal aktualisierten Reader\* wurden maßgebliche Grundlagen- und Übersichtsarbeiten sowie einige Originalarbeiten zur weiteren Vertiefung in diese allgemeine Thematik aufgenommen (siehe Anhang).

Die Nutzung von GKV-Routinedaten, zu denen primär, aber nicht ausschließlich Abrechnungsdaten der gesetzlichen Krankenkassen zählen, im Rahmen der Pandemiebekämpfung und des Pandemiemonitorings ist bislang nicht erfolgt. Die nachfolgend beschriebenen Use Cases liefern dafür Prototypen. Die daraus abgeleiteten allgemeinen und spezifischen Handlungsempfehlungen unter prozeduralen, methodischen, technischen und rechtlichen Aspekten stecken den Rahmen und die Möglichkeiten für zukünftige und gegenüber der aktuellen Pandemie erweiterte Nutzungen ab. Dabei werden die oben genannten allgemeinen Auswertungsschritte pandemiespezifisch ausgestaltet und differenziert.

---

\*<http://egepan.de/nutzung-von-routinedaten-fuer-das-pandemiemanagement/>



## 3. Use Cases

### 3.1 Use Case 1: Risikofaktoren für einen schweren COVID-19-Verlauf

#### 3.1.1 Ausgangssituation

Die deutsche Impfstrategie hatte das Ziel, die größtmögliche Zahl von COVID-19-Hospitalisierungen und COVID-19-assoziierten Todesfällen zu vermeiden. Daher sind ältere Personen, Personen, die in Pflegeeinrichtungen und der Gesundheitsversorgung arbeiten, sowie Menschen mit Vorerkrankungen, die einen schweren COVID-19-Verlauf begünstigen, zuerst geimpft worden.<sup>11</sup> Die Fachgruppe Impfprävention unter Leitung von PD Dr. Ole Wichmann und Dr. Thomas Harder sowie des Robert Koch-Instituts benötigte zur mathematischen Modellierung der Folgen verschiedener Impfstrategien Prävalenzraten von Risikofaktoren für einen schweren COVID-19-Verlauf in der deutschen Bevölkerung. Die Evidenz zur Prävalenz von Risikofaktoren für einen schweren COVID-19-Verlauf zu diesem Zeitpunkt war begrenzt und systematische, populationsbasierte Kennzahlen fehlten. Aufgrund ihres eindeutigen Personen- und Bevölkerungsbezugs bilden GKV-Routinedaten eine geeignete Grundlage für Morbiditätsschätzungen in der deutschen Bevölkerung.<sup>12-14</sup>

#### 3.1.2 Beteiligte Partner

Für den Erfolg des Use Cases entscheidend war die Konstellation der Arbeitsgruppe, bestehend aus

- mit GKV-Routinedatenanalysen erfahrenen Forschungsinstituten (ZEGV, InGef),
- dem Fachgebiet Impfprävention des Robert Koch-Instituts mit seiner epidemiologischen Expertise und Nähe zu gesundheitspolitischen Entscheidungen.

Dem Zentrum für Evidenzbasierte Gesundheitsversorgung (ZEGV) unter der Leitung von Prof. Dr. Jochen Schmitt stehen GKV-Routinedaten der AOK PLUS aus Sachsen zum Zweck der Versorgungsforschung zur Verfügung. Durch die InGef – Institut für angewandte Gesundheitsforschung Berlin GmbH können anonymisierte Routinedaten von etwa 6,5 Millionen Versicherten von Betriebskrankenkassen pro Jahr analysiert werden.

#### 3.1.3 Fragestellungen

Das Projekt adressierte die Fragestellung, wie häufig Erkrankungen mit moderatem und Erkrankungen mit hohem Risiko für einen schweren COVID-19-Verlauf in der deutschen Bevölkerung auftreten.

### 3.1.4 Methodik

Basierend auf einem Umbrella Review der Kolleg:innen des RKI, der die zur Verfügung stehende Literatur zu Risikofaktoren zusammengefasst hatte, wurden 27 Erkrankungen identifiziert, die mit einem schweren COVID-19-Verlauf assoziiert sind.<sup>15</sup> Zur Identifikation von Personen mit den entsprechenden Erkrankungen in den GKV-Routinedaten wurde ein Klassifikationsschema entwickelt. In der ambulanten und stationären Versorgung dokumentierte Diagnosen (ICD-10-Code) sowie indikationsspezifische Arzneimittelverordnungen wurden zur Falldefinition herangezogen. Für die Prävalenzberechnung wurden GKV-Routinedaten des Jahres 2019 genutzt. Um die Robustheit und Validität der Ergebnisse überprüfen zu können, wurden sowohl sensitive als auch spezifische Falldefinitionen festgelegt. Auf Basis der zu dem Zeitpunkt zur Verfügung stehenden Evidenz wurden außerdem Erkrankungen mit moderatem und hohem Risiko für einen schweren COVID-19-Verlauf gruppiert. Zur Plausibilisierung der genutzten Datenbasis wurden außerdem die COVID-19-Meldedaten, die dem RKI vorliegen, mit der Anzahl der COVID-19-Fälle in den Daten der AOK PLUS und der InGef-Forschungsdatenbank abgeglichen.

#### **Datenbasis und Studienpopulation:**

Für die Analysen konnten Daten von insgesamt etwa 8,6 Millionen GKV-Versicherten der AOK PLUS und von etwa 60 Betriebskrankenkassen genutzt werden. Für diese Versicherten wurden sämtliche bei den Krankenkassen abgerechneten Leistungen über einen Zeithorizont von mehreren aufeinanderfolgenden Jahren betrachtet. Dies umfasst sowohl ambulante als auch stationäre Diagnosen (inkl. COVID-19-Diagnosen), Prozeduren (EBM, OPS) und verschriebene Medikamente. Darüber hinaus sind Daten zu Arbeitsunfähigkeitsbescheinigungen (AU) und Sterbedaten verfügbar. Für die Prävalenzberechnung wurden GKV-Routinedaten des Jahres 2019 genutzt.

#### **Statistische Methodik:**

Es wurden direkt altersstandardisierte Prävalenzraten für das Kalenderjahr 2019 berechnet. Zur direkten Altersstandardisierung wurde die deutsche Bevölkerung laut Destatis-Bevölkerungsstatistik des Jahres 2019 gewählt. In die Risikopopulation wurden Personen in der ZEGV- und InGef-Forschungsdatenbank eingeschlossen, für die im Jahr 2019 vollständige Daten vorlagen. Als Prävalenz wurde die absolute Anzahl Personen mit der entsprechenden Erkrankung relativ zu 100 Personen unter Risiko berichtet. Die Ausgabe erfolgte stratifiziert nach Altersgruppen, Geschlecht und Bundesland.

### 3.1.5 Arbeitsmodus im Konsortium

In die Arbeitsgruppe waren erfahrene Wissenschaftler:innen mit Expertise im Bereich Statistik, Epidemiologie und Data Science von allen beteiligten Partnern involviert. Unter Koordination der InGef wurden a priori ein Studienprotokoll zwischen allen beteiligten Wissenschaftler:innen abgestimmt sowie eine entsprechende Referenztafel zur Defi-

dition der Outcomes entwickelt und im Konsortium geteilt und validiert. Hierfür wurde eine Projektmanagementplattform (Confluence) verwendet. In wöchentlichen Jour-fixe-Terminen wurden methodische Probleme unter den beteiligten Wissenschaftler:innen diskutiert und Lösungen erarbeitet. Zusätzlich wurde eine Datensatzbeschreibung konsentiert, in der Variablennamen und Ausprägungen festgelegt wurden, um den Austausch der aggregierten Daten zu vereinfachen. Die Analysen wurden getrennt durch das ZEGV und die InGef durchgeführt. Die aggregierten, deskriptiven Ergebnisse des ZEGV und der InGef wurden durch die InGef gepoolt, mit allen Projektbeteiligten diskutiert und dem RKI zur Verfügung gestellt.

Alle Analysen erfolgten im Einklang mit der Guten Praxis Sekundärdatenanalyse der Deutschen Gesellschaft für Epidemiologie unter Einbezug der aktuellen internationalen wissenschaftlichen Evidenz sowie der medizinischen Expertise und Kenntnis der GKV-Routinedaten im Konsortium.

### 3.1.6 Projektergebnisse *(Stand Dezember 2021)*

Die Ergebnisse des Projekts waren direkt altersstandardisierte Prävalenzraten von 27 Erkrankungen sowie für zusammengefasste Krankheitsgruppen mit moderatem bzw. hohem Risiko für einen schweren COVID-19-Verlauf. Die Prävalenzraten wurden außerdem stratifiziert nach Alter, Geschlecht und Bundesland berichtet. Nach der Analyse durch die jeweiligen Datenhalter ZEGV und InGef wurden die aggregierten, deskriptiven Ergebnisse zusammengeführt und dem RKI zur Verfügung gestellt.

### 3.1.7 Erkenntnisse

- GKV-Routinedaten bilden eine valide Grundlage zur Schätzung populationsbasierter Prävalenzraten von Risikofaktoren.
- Die für die Analysen genutzten GKV-Routinedaten erlauben eine regionale Stratifizierung der Ergebnisse. So können regionale Unterschiede schnell sichtbar gemacht und ggf. durch adäquate Public-Health-Maßnahmen zielgerichtet adressiert werden.
- Eine datenschutzkonforme Auswertung und Poolung (deskriptiver) GKV-Routinedatenanalysen von verschiedenen Datenhaltern ist möglich und sinnvoll. In diesem Projekt ergänzten sich die Daten des ZEGV, die überwiegend Versicherte aus Sachsen umfassen, und die Daten der InGef, die überwiegend Daten von Versicherten aus Süd- und Westdeutschland beinhalten. Allgemein lassen sich durch das Poolen von Daten spezifische Besonderheiten in Patientencharakteristika einzelner Kassen überwinden und eine höhere Generalisierbarkeit kann erzielt werden.
- Im Rahmen von bestehenden Partnerschaften können kurzfristig Ergebnisse zu relevanten und zeitkritischen Fragestellungen zur Pandemiesteuerung zur Verfügung gestellt werden.

## 3.2 Use Case 2: Hierarchisierung von Risikofaktoren für schwere COVID-19-Erkrankungsverläufe

### 3.2.1 Ausgangssituation

In Ergänzung zur Frage der Prävalenz bekannter Risikofaktoren für einen schweren COVID-19-Verlauf wurde im weiteren Pandemieverlauf Evidenz zur Relevanz der individuellen Risikofaktoren benötigt. Zur Weiterentwicklung der Modellierung der laufenden Impfkampagne war es nötig, das Ausmaß der Risikoerhöhung durch einzelne Risikofaktoren zu quantifizieren. Von besonderem Interesse war dabei die Frage, ob das zeitgleiche Auftreten von mehreren Risikofaktoren mit einem höheren Risiko für einen schweren COVID-19-Verlauf verbunden ist.

### 3.2.2 Beteiligte Partner

Für den Erfolg des Use Cases entscheidend war die Konstellation der Arbeitsgruppe, bestehend aus:

- mit GKV-Routinedatenanalysen erfahrenen Forschungsinstituten (ZEGV, InGef, bifg),
- Krankenkassen mit methodisch versierten Mitarbeiter:innen, hoher interner Priorisierung des Projekts und unentgeltlicher Zusammenarbeit (AOK Bayern, AOK PLUS, BARMER, DAK, TK) und
- dem Robert Koch-Institut mit seiner epidemiologischen Expertise und Nähe zu gesundheitspolitischen Entscheidungen.

### 3.2.3 Fragestellungen

Konkrete Fragestellungen und die methodische Umsetzung wurden unter Koordination der InGef und des ZEGV gemeinsam durch die Arbeitsgruppe festgelegt.

Das Projekt adressierte folgende Fragestellungen:

- Welche Vorerkrankungen gehen mit einem besonders hohen Risiko für einen schweren COVID-19-Verlauf einher?
- Inwiefern lassen sich diese Vorerkrankungen hinsichtlich ihres individuellen Risikos für einen schweren Verlauf hierarchisieren?
- Ist das zeitgleiche Auftreten mehrerer Vorerkrankungen mit einem nochmals erhöhten Risiko für einen schweren Verlauf von COVID-19 assoziiert?

Ziel des Projekts war es außerdem, einen Score zu entwickeln, der das zeitgleiche Vorliegen mehrerer Erkrankungen berücksichtigt und eine einfache Identifikation von Personen mit einem hohen Risiko für einen schweren COVID-19-Verlauf im Versorgungsalltag ermöglicht.

### 3.2.4 Methodik

Grundlage für die Analysen war eine Kohorte von Personen mit COVID-19-Diagnose (ICD-10-GM: U07.1! – COVID-19, Virus nachgewiesen), die zwischen 27.01.2020 und 30.06.2020 bzw. 31.12.2020 diagnostiziert wurden. Auf Basis ambulant und stationär abgerechneter COVID-19-bezogener Leistungen wurde ein Indexdatum als Beginn der COVID-19-Erkrankung definiert. Abrechnungsdaten der eingeschlossenen Personen aus dem Jahr vor dem Indexdatum wurden zur Erhebung von Risikofaktoren für einen schweren COVID-19-Verlauf genutzt. Da die Impfungen von Personen im Alter ab 80 Jahren zum Zeitpunkt der Analyse in Deutschland bereits weit fortgeschritten waren, wurden diese aus der Analyse ausgeschlossen. Als Endpunkt der Analysen wurden schwere COVID-19-bedingte Erkrankungsverläufe anhand von COVID-19-assoziiierter intensivmedizinischer Behandlung, Beatmung oder Tod im zeitlichen Zusammenhang mit einer COVID-19-Erkrankung definiert. Als Risikofaktoren wurden 35 Erkrankungen definiert, die basierend auf den Ergebnissen eines systemischen Reviews<sup>15</sup> und den Empfehlungen der STIKO mit einem erhöhten Risiko für einen schweren COVID-19-Verlauf assoziiert waren. Außerdem wurden Altersgruppen (18–24, 25–39, 40–49, 50–54, ..., 75–79) in der Analyse berücksichtigt.

#### **Datenbasis:**

Durch die breite Beteiligung von großen Krankenkassen konnten 38 Millionen Versicherte von AOK Bayern, AOK PLUS, BARMER, BKKen, DAK und TK in die Analysen eingeschlossen werden. Für diese Versicherten wurden sämtliche bei den Krankenkassen abgerechneten Leistungen zwischen 2019 und 2020 für die Analyse genutzt.

Daten der ambulanten und stationären Versorgung sowie ambulante Arzneimitteldaten wurden zur Bildung der entsprechenden Patientengruppen genutzt. Sofern mittels Routinedaten abbildbar, wurden zusätzlich zu ambulant oder stationär dokumentierten Diagnosen auch Arzneimittelverordnungen oder stationär durchgeführte Operationen und Prozeduren sowie ambulante Leistungsziffern zur Risikofaktordefinition genutzt.

#### **Statistische Methodik:**

Zunächst wurde zur Ableitung einer einfachen und möglichst effektiven Identifikation von Personen mit einem erhöhten Risiko für einen schweren COVID-19-Verlauf eine hierarchische Rangfolge der einzelnen Risikofaktoren erstellt. Diese wurde gebildet, indem iterativ Personen mit demjenigen Risikofaktor, der mit dem höchsten absoluten Risiko für schwere COVID-19-Erkrankungsverläufe im ersten Halbjahr 2020 assoziiert war, zur hypothetischen Impfung vorgesehen wurden. Das absolute Risiko für einen schweren Verlauf wurde hierbei über den Anteil an Personen mit dem entsprechenden Risikofaktor, die eine COVID-19-assoziierte Intensivbehandlung erhielten oder verstarben, geschätzt. Die auf den Datensätzen der einzelnen Krankenkassen erzielten Schätzergebnisse wurden in jedem Analyseschritt mittels metaanalytischer Verfahren durch das ZEGV gepoolt.

Zur Entwicklung eines Scores, der zur Impfpriorisierung genutzt werden kann, wurden die Datensets um das 2. Halbjahr 2020 erweitert und jeweils in einen zufällig selektierten Trainings- und einen Testdatensatz geteilt. Der Trainingsdatensatz enthielt jeweils 90 % der Studienpopulation, die Testdatensätze dementsprechend 10 % der Studienpopulation. Während der Trainingsdatensatz zur Entwicklung des Scores genutzt wurde, diente der Testdatensatz zur Validierung des Scores.

Der Einfluss der vorab definierten Risikofaktoren wurde anhand von Poisson-Regressionsmodellen mit robusten Standardfehlern berechnet. Als Outcome wurde ein schwerer COVID-19-Verlauf (Hospitalisierung mit Beatmung und/oder intensivmedizinischer Behandlung) oder Tod nach einer COVID-19 Infektion festgelegt.

Die Analysen erfolgten separat für die Altersgruppen 18–64, 65–79 und 80 Jahre und älter. Mittels metaanalytischer Verfahren wurden die Regressionsergebnisse aller Projektbeteiligten gepoolt.

Klinische Expert:innen haben anschließend auf Basis dieser Ergebnisse eingeschätzt, ob sich aus den Ergebnissen Interaktionen zwischen dem Alter und dem jeweiligen Risikofaktor ableiten lassen. Anschließend wurden die Regressionsanalysen mit den ausgewählten Interaktionstermen in allen Datensätzen wiederholt. Es folgte eine finale metaanalytische Poolung der Ergebnisse am ZEGV.

Die Regressionskoeffizienten wurden auf Werte zwischen 0 und 20 normalisiert. Der finale Score wurde im Testdatensatz hinsichtlich seiner Performance, das Outcome schwerer COVID-19-Verläufe vorherzusagen, evaluiert. Als Evaluationsparameter wurden ROC-Kurven sowie die AUC (Area under the Curve) ausgegeben.

### 3.2.5 Arbeitsmodus

Entscheidend für die erfolgreiche Bearbeitung des Use Cases war ein effizienter und kollegialer Arbeitsmodus aller Beteiligten. Im Konsortium wurde so ein effizienter und datenschutzkonformer Modus der Datenaufbereitung und -analyse etabliert. Erfahrene Statistiker:innen, Epidemiolog:innen und Data Scientists der beteiligten Krankenkassen und Forschungsinstitute waren Teil des Konsortiums. Unter Koordination der InGef und des ZEGV wurden a priori Studienprotokolle und statistische Analysepläne zwischen allen Beteiligten abgestimmt, entsprechende Skripte für die Analysen entwickelt und im Konsortium geteilt und validiert. Hierfür wurde eine Projektmanagementplattform (Confluence) verwendet. In wöchentlichen Jour-fixe-Terminen wurden methodische Probleme unter den beteiligten Analyst:innen diskutiert und Lösungen erarbeitet.

Die Analysen wurden getrennt für die Versicherten jeder beteiligten Krankenkasse innerhalb der einzelnen Krankenkassen selbst oder durch autorisierte Forschungs-

institute durchgeführt. Die aggregierten Ergebnisse der Analysen der einzelnen Krankenkassen-Datenbestände wurden abschließend an das ZEGV zur Evidenzsynthese versandt. Das ZEGV hat die deskriptiven Ergebnisse aller analysierenden Stellen zusammengefügt, um die Ergebnisse insgesamt interpretierbar zu machen.

Alle Analysen erfolgten im Einklang mit der Guten Praxis Sekundärdatenanalyse der Deutschen Gesellschaft für Epidemiologie unter Einbezug der aktuellen internationalen wissenschaftlichen Evidenz sowie der medizinischen Expertise und Kenntnis der GKV-Routinedaten im Konsortium.

### 3.2.6 Ergebnisse

#### **Rangfolge absoluter Risiken**

Als Projektergebnis konnte eine Rangfolge von 24 Erkrankungen bzw. Altersgruppen identifiziert werden, die mit einem höheren absoluten Risiko für einen schweren COVID-19-Verlauf assoziiert waren. Aktuell in Behandlung befindliche Personen mit hämatoonkologischen Erkrankungen bildeten die bedeutendste Risikogruppe, gefolgt von Personen mit metastasierten soliden Tumorerkrankungen mit Therapie und Demenzpatient:innen. Die Ergebnisse der Analysen belegen, dass eine rein am Alter orientierte Rangfolge nicht optimal ist.

Die Ergebnisse der Analysen wurden im Epidemiologischen Bulletin 19/2021 im Mai 2021 veröffentlicht und ebenfalls im Mai 2021 der STIKO vorgestellt.<sup>16</sup>

#### **Score**

Mittels des oben beschriebenen Scores wurde sowohl den Altersgruppen als auch den ausgewählten Risikofaktoren ein Scorewert im Sinne eines Gewichtes zugeteilt. Hohes Alter, hämatoonkologische Erkrankungen und immunsuppressive Therapien waren mit einem besonders hohen Risiko für einen schweren COVID-19-Verlauf verbunden. Der mittlere Risikoscore innerhalb der Testdatensätze lag zwischen 10,0 und 13,2. Die AUC variierte zwischen 0,87 und 0,92. Auch wenn der Parameter Alter bereits gut einen schweren COVID-19-Verlauf vorhersagt, können über die Berücksichtigung schwerer Erkrankungen Gerechtigkeitslücken geschlossen werden, die eine Priorisierung nach Alter mit sich bringt.

Neben einem potenziellen Einsatz in ambulanten Arztpraxen könnte dieses Scoreverfahren durch weitere Akteure mit Zugriff auf relevante Patientendaten (z. B. Krankenkassen, Kassenärztliche Vereinigungen) effizient implementiert werden. Geplant ist die Einreichung eines Manuskriptes bei einem internationalen wissenschaftlichen Journal, sodass die präsentierten Ergebnisse ggf. zur Planung zukünftiger (Auffrischungs-) Impfkampagnen oder von Verantwortlichen in Ländern, in denen COVID-19-Impfstoffe noch nicht für die gesamte Bevölkerung zur Verfügung stehen, genutzt werden können.



### 3.2.7 Erkenntnisse

- Durch die interdisziplinäre Zusammenarbeit von epidemiologischen, medizinischen und methodischen Expert:innen im Konsortium konnten valide und robuste Ergebnisse auf Basis von GKV-Routinedaten erzielt werden.
- Die Auswahl und Operationalisierung bzw. Definition von Vorerkrankungen und Endpunkten wurde innerhalb des Projektteams aus verschiedenen Perspektiven betrachtet und bedurfte intensiver Diskussionen bis zur finalen Festlegung. Die Formulierung eindeutiger Fragestellungen und Projektziele war elementare Grundlage für die Diskussion innerhalb des Projektteams.
- Das metaanalytische Poolen der Ergebnisse von Analysen auf separaten Datensätzen ermöglichte die datenschutzkonforme Auswertung von GKV-Routinedaten von fast 35 Millionen Versicherten.
- Durch Auswertung bei entsprechenden Datenhaltern oder berechtigten Stellen (InGef/ZEGV) und Austausch aggregierter Statistiken zwischen den Stellen entfielen zeitintensive Genehmigungsverfahren durch die Aufsichtsbehörden sowie die Klärung der technischen Bereitstellung von Daten mit den Datenhaltern.

## 3.3 Use Case 3: Hintergrundinzidenz Sinusvenenthrombosen und Myokarditis

### 3.3.1 Ausgangssituation

Im März 2021 traten mehrere Fälle von Sinusvenenthrombosen (SVT) innerhalb von vier bis 16 Tagen nach einer Vektor-Impfung gegen COVID-19 auf.<sup>17</sup> Bis Anfang April wurden 42 Fälle, die in einem entsprechenden zeitlichen Zusammenhang aufgetreten waren, an das Paul-Ehrlich-Institut (PEI) gemeldet. Bei 23 erkrankten Personen wurde zeitgleich eine Thrombozytopenie beobachtet, acht Personen waren bis zu dem Zeitpunkt aufgrund der SVT verstorben.

Das Auftreten von Myokarditiden in einem zeitlichen Zusammenhang mit dem mRNA-Impfstoff von Biontech machte weitere Analysen zur Hintergrundinzidenz dieser Erkrankung notwendig.

Da sowohl eine SVT als auch eine Myokarditis unabhängig von einer COVID-19-Impfung auftreten können, waren Informationen zur Hintergrundinzidenz der beiden Erkrankungen, also zur Häufigkeit der Erkrankungen, die auch in einer nicht geimpften Population innerhalb eines bestimmten Zeitraums auftreten würde, notwendig, um einordnen zu können, inwiefern es sich bei den beobachteten Fällen um unerwünschte Nebenwirkungen (UAW) der Impfung handelt. Besonders im Hinblick auf eine generelle Impfempfehlung, die zu diesem Zeitpunkt noch diskutiert wurde, waren valide Daten zur Hintergrundinzidenz notwendig.

Mitte März wurde der Bedarf zur Schätzung von Hintergrundinzidenzen der SVT an die Projektgruppe getragen. Die Anfrage zur Berechnung der Hintergrundinzidenz der Myokarditis folgte Ende April 2021.

### 3.3.2 Beteiligte Partner

Über eine bereits bestehende Kooperation (siehe Kapitel Risikofaktoren für einen schweren COVID-19-Verlauf) mit dem Fachgebiet Impfprävention am Robert Koch-Institut wurden die Analysen kurzfristig an der InGef durchgeführt.

Als medizinische Experten waren Herr PD Dr. Robert Klamroth, Chefarzt der Klinik für Innere Medizin – Angiologie und Hämostaseologie des Klinikums im Friedrichshain, und Prof. Dr. Dr. Andreas Bollmann, leitender Oberarzt der Abteilung für Rhythmologie am Herzzentrum Leipzig, an den Projekten beteiligt.

### 3.3.3 Fragestellungen

Das Projekt adressierte für diese beiden Krankheitsentitäten zwei Fragestellungen:

- Wie hoch war die Hintergrundinzidenz der SVT in den Jahren 2015 bis 2019 in der deutschen Bevölkerung?
- Wie hoch war die Hintergrundinzidenz der Myokarditis in den Jahren 2015 bis 2020 in der deutschen Bevölkerung?

### 3.3.4 Methodik

Zur Bestimmung der Hintergrundinzidenz von SVT und Myokarditis wurden jeweils jährliche querschnittliche Analysen durchgeführt. Durch die Einbindung von PD Dr. Robert Klamroth, Chefarzt der Klinik für Innere Medizin – Angiologie und Hämostaseologie des Klinikums im Friedrichshain, wurde eine geeignete Falldefinition mittels ICD-10-Codes zur Identifikation der SVT in den GKV-Routinedaten gefunden. In Zusammenarbeit mit Prof. Dr. Dr. Bollmann wurde eine klinisch sinnvolle Definition zur Identifikation von Myokarditispatient:innen gefunden.

#### **Datengrundlage:**

Grundlage für die Analyse zur Hintergrundinzidenz der Sinusvenenthrombose und der Myokarditis war die InGef-Forschungsdatenbank. Zum Zeitpunkt der Analyse enthielt die Datenbank anonymisierte, longitudinal verknüpfte Daten von ca. 6,5 Millionen Versicherten, die zwischen 2014 und 2019 bzw. 2020 in einer von 61 Betriebskrankenkassen in Deutschland versichert waren. Neben soziodemografischen Informationen enthält die Datenbank Informationen über ambulante Leistungen und Diagnosen; Krankenhausdaten einschließlich der Aufnahmezeiten, der Haupt- und Nebendiagnosen und der durchgeführten Operationen und Prozeduren; Daten über die Verschreibung von

Medikamenten; Informationen über Hilfsmittel und Heilmittel sowie über die Kosten, die in diesen Bereichen angefallen sind. Alle Daten auf Patienten- und Leistungserbringerebene sind entsprechend den deutschen Datenschutzbestimmungen und Bundesgesetzen anonymisiert.

Da die Daten in anonymisierter Form zur Verfügung standen und die Analysen keiner zusätzlichen (datenschutzrechtlichen) Genehmigung einer Aufsichtsbehörde bedurften sowie aufgrund der Dringlichkeit der Fragestellung konnten die Analysen sehr kurzfristig durchgeführt werden.

Für die Analysen zur Sinusvenenthrombose wurden Routinedaten der Jahre 2015 bis einschließlich 2019 genutzt. Da zum Zeitpunkt der Myokarditisanalysen bereits Abrechnungsdaten des Jahres 2020 vorlagen, wurden für diese Analysen Routinedaten bis einschließlich 2020 genutzt.

In die Studienpopulation eingeschlossen wurden Personen, die im jeweiligen Betrachtungsjahr vollständig beobachtbar waren bzw. verstorben sind. Anhand von ambulant oder stationär dokumentierten ICD-10- sowie ambulanten (EBM; einheitlicher Bewertungsmaßstab) und stationären Prozedurenschlüsseln (OPS; Operationen- und Prozedurenschlüssel) wurden SVT- und Myokarditispatient:innen identifiziert.

#### **Statistische Methodik:**

Die Hintergrundinzidenz wurde als Anzahl Personen mit SVT bzw. Myokarditis relativ zur Gesamtbevölkerung unter Risiko in der InGef-Forschungsdatenbank berechnet. Zur Überprüfung der Robustheit der Analysen wurden für beide Krankheitsentitäten verschiedene Falldefinitionen analysiert. Zusätzlich zur Bestimmung der Inzidenz der SVT wurde analysiert, wie hoch der Anteil der SVT-Patient:innen mit begleitender Thrombozytopenie in der Vergangenheit war. Zusätzlich zur Berechnung der Hintergrundinzidenz der Myokarditis in der Allgemeinbevölkerung wurde die Inzidenz der Myokarditis in der Population der an COVID-19 erkrankten Personen untersucht. Alle Erkrankungsdaten wurden stratifiziert nach Altersgruppen und Geschlecht berichtet.

#### **3.3.5 Arbeitsmodus**

Zur Bestimmung der Hintergrundinzidenz von SVT und Myokarditis wurde ein Kurzstudienprotokoll erstellt, in dem u.a. die Risikopopulation, die SVT- bzw. Myokarditisfalldefinition sowie die geplanten Stratifizierungen beschrieben wurden. Das Studienprotokoll wurde mit den Kolleg:innen des RKI und mit PD Dr. Klamroth bzw. Prof. Dr. Dr. Bollmann abgestimmt.

Nach der Analyse wurden die Ergebnisse beider Analysen jeweils hinsichtlich ihrer Plausibilität (Gesamthäufigkeit, Alters- und Geschlechtsverteilung) mit den Kolleg:innen des RKI und PD Dr. Klamroth bzw. Prof. Dr. Dr. Bollmann diskutiert.

### 3.3.6 Ergebnisse

#### SVT

Die mittlere jährliche Inzidenz der SVT lag zwischen 2015 und 2019 bei 1,9/100.000 Personen unter Risiko (96 %-CI 1,4–2,3), was etwa 1.300 SVT-Fällen in Deutschland pro Jahr entspricht. Die Inzidenz war deutlich höher bei weiblichen Personen zwischen 18 und 59 Jahren (2,4/100.000; 95 %-CI 1,5–3,4) als bei gleich alten Männern (1,0/100.000; 95 %-CI 0,4–1,7). Die Zahl der SVT-Patient:innen mit begleitender Thrombozytopenie war sehr klein (n = 0–2 pro Jahr).

Die finalen Ergebnisse des SVT-Projekts standen Ende März 2021 zur Verfügung und konnten so in die 4. Aktualisierung der COVID-19-Impfempfehlung vom 22.04.2021 einfließen.<sup>18</sup> Die Ergebnisse wurden im Anschluss außerdem bei PLOS ONE als Short Report eingereicht und vorab als Preprint veröffentlicht.<sup>19</sup>

#### Myokarditis

Die Einjahresinzidenz der Myokarditis lag im Jahr 2018 bei 21,1/100.000 Personen unter Risiko (95 %-CI 18,1–25,4) und bei 16,9/100.000 (95 %-CI 15,8–18,1) im Jahr 2020. Im Unterschied zur SVT erkrankten männliche Versicherte häufiger an einer Myokarditis als weibliche Personen. In der Gruppe der 18- bis 29-Jährigen war die Einjahresinzidenz im Jahr 2020 bei den Männern mehr als dreimal höher als bei den gleichaltrigen Frauen (25,7/100.000 vs. 8,1/100.000). In der Gruppe der Kinder und Jugendlichen, die an COVID-19 erkrankten, fanden sich keine Myokarditisfälle in den sechs Monaten nach COVID-19-Diagnose.

Die Ergebnisse zur Hintergrundinzidenz der Myokarditis wurden gemeinsam mit der Fachgruppe Impfprävention des RKI und Herrn Prof. Dr. Dr. Bollmann hinsichtlich ihrer klinischen Plausibilität diskutiert und fanden Eingang in die 9. Aktualisierung der COVID-19-Impfempfehlung.<sup>20</sup>

### 3.3.7 Erkenntnisse

- GKV-Routinedaten sind eine vergleichsweise schnell verfügbare und valide Datengrundlage zur Schätzung populationsbasierter Raten schwerer Erkrankungen.
- Durch bereits vorab bestehende Kooperationen der beteiligten Akteure konnte schnell Evidenz zur Pandemiesteuerung generiert werden.
- Die Daten in der InGef-Forschungsdatenbank liegen in anonymisierter Form vor. So konnten aggregierte Ergebnisse zügig zur Verfügung gestellt werden, da es keiner zusätzlichen Freigabe bzw. Bewilligung der Datennutzung durch eine Aufsichtsbehörde bedurfte.
- Input von Expert:innen, die in der medizinischen Versorgung tätig sind, unterstützt

Forschende dabei klinisch sinnvoll, Definitionen von Endpunkten zu finden, und ermöglicht eine klinische Einordnung und Plausibilisierung der Ergebnisse.

### 3.4 Use Case 4: Post-COVID bei Kindern und Jugendlichen

#### 3.4.1 Ausgangssituation

Im Frühjahr 2021 wurde international zunehmend Evidenz für anhaltende oder auch nach der Akutphase der COVID-19-Erkrankung neu auftretende Symptome beschrieben. Der Symptomkomplex ist variabel und kann Beschwerden der Atemwege, der körperlichen und psychischen Belastbarkeit im Sinne eines postviralen Erschöpfungssyndroms und chronische Schmerzzustände beinhalten. Bis zum Sommer 2021 gab es zu Häufigkeit, Schwere und Risikofaktoren von dem als Post-COVID (Symptome bestehen länger als zwölf Wochen) bezeichneten Krankheitsbild international kaum und deutschlandweit keine belastbare Evidenz. Dies betraf insbesondere die Existenz von Post-COVID bei Kindern und Jugendlichen. Es bestanden zu diesem Zeitpunkt erhebliche Unsicherheiten hinsichtlich der Definition, der Art und Häufigkeit von COVID-19-Folgeerkrankungen, der Risikofaktoren für deren Auftreten sowie der Auswirkungen von Post-COVID auf Ebene des Gesundheitssystems.

Vor diesem Hintergrund wurde im zweiten Quartal 2021 eine interministerielle Arbeitsgruppe (IMA) unter Beteiligung des Bundesgesundheitsministeriums und des Bundesfamilienministeriums gegründet, um die Folgen der Pandemie für Kinder und Jugendliche abzuschätzen und entsprechende Maßnahmen einzuleiten. Das Robert Koch-Institut als Mitglied der IMA ist vor diesem Hintergrund im Juli 2021 an die bestehende Arbeitsgruppe aus AP 0 des NUM-Projekts egePan Unimed herantreten, um gemeinsam Fragestellungen zu Post-COVID bei Kindern und Jugendlichen wie auch bei Erwachsenen anhand von GKV-Routinedaten zu analysieren und möglichst rasch Antworten auf die drängenden Fragen des öffentlichen Interesses auf dieser Datengrundlage zu erarbeiten.

#### 3.4.2 Beteiligte Partner

Für den Erfolg des Use Cases entscheidend war die Konstellation der Arbeitsgruppe, bestehend aus:

- mit GKV-Routinedatenanalysen erfahrenen Forschungsinstituten (ZEGV – Zentrum für Evidenzbasierte Gesundheitsversorgung, InGef – Institut für angewandte Gesundheitsforschung Berlin GmbH, bifg – BARMER Institut für Gesundheitssystemforschung),
- Krankenkassen mit methodisch versierten Mitarbeiter:innen, hoher interner Priorisierung des Projekts und unentgeltlicher Zusammenarbeit (AOK Bayern, AOK PLUS, BARMER, DAK-Gesundheit, Techniker Krankenkasse [TK]),

- dem Robert Koch-Institut mit seiner epidemiologischen Expertise und Nähe zu gesundheitspolitischen Entscheidungen sowie
- klinischer Expertise aus der Klinik und Poliklinik für Kinder- und Jugendmedizin des Universitätsklinikums Dresden.

Konkrete Fragestellungen und die methodische Umsetzung wurden unter Koordination des ZEGV gemeinsam durch die Arbeitsgruppe festgelegt.

### 3.4.3 Fragestellungen

Das Projekt adressierte folgende Fragestellungen:

- Bedingt eine SARS-CoV-2-Infektion eine erhöhte Wahrscheinlichkeit für das inzidente Auftreten ausgewählter Erkrankungen (insbesondere chronische Lungenerkrankungen, psychische Erkrankungen inkl. Fatigue, Erkrankungen des Zentralnervensystems, kardiovaskuläre Erkrankungen und Ereignisse, endokrinologische Erkrankungen wie Diabetes mellitus Typ 2)?
- Bedingt eine SARS-CoV-2-Infektion eine insgesamt erhöhte Erkrankungslast?
- Welche COVID-19-Folgeerkrankungen treten in bestimmten Subgruppen der Bevölkerung (insbesondere Kinder und Jugendliche) auf?
- Welche Rolle spielt die Schwere der COVID-19-Primärerkrankung für das Risiko von COVID-19-Folgeerkrankungen?
- Inwiefern unterscheiden sich Personen mit SARS-CoV-2-Infektion im Hinblick auf die oben genannten Endpunkte von Personen aus der Allgemeinbevölkerung ohne dokumentierte COVID-19-Erkrankung?

### 3.4.4 Methodik

Die Studie wurde als gematchte Kohortenstudie auf Grundlage von GKV-Routinedaten konzipiert. Es wurden Personen mit COVID-19-Diagnose mit Personen aus der Allgemeinbevölkerung hinsichtlich des Auftretens von Folgeerkrankungen und ihrer Inanspruchnahme von Leistungen des Gesundheitssystems verglichen. Hierbei wurde zwischen Post-COVID-Erkrankungen (in Anlehnung an NICE-Guideline mindestens zwölf Wochen nach COVID-19-Erkrankung) und COVID-assoziierten Erkrankungen (beliebiger Zeitpunkt nach COVID-Erkrankung) unterschieden.

#### **Datengrundlage und Studienpopulation:**

Durch die breite Beteiligung von großen Krankenkassen konnten 38 Millionen Versicherte von AOK Bayern, AOK PLUS (Region Sachsen), BARMER, BKKen, DAK und TK in die Analysen eingeschlossen werden. Für diese Versicherten wurden sämtliche bei den Krankenkassen abgerechneten Leistungen über einen Zeithorizont von mehreren Jahren betrachtet. Diese umfassten sowohl ambulante als auch stationäre Diagnosen (inkl. COVID-19- und Post-COVID-Diagnosen), Prozeduren (EBM, OPS) und verschriebene Medikamente. Darüber hinaus sind Daten zu Arbeitsunfähigkeitsbescheinigungen (AU)

und Sterbedaten (Todestag) verfügbar. Die Analyse wurde ab August 2021 durchgeführt. Es konnten dabei GKV-Routinedaten bis einschließlich des 4. Quartals 2020 in die Analysen einbezogen werden (Latenz sieben bis neun Monate).

#### **Statistische Methodik:**

Zur Sicherstellung der Vergleichbarkeit der Kohorten hinsichtlich relevanter Charakteristika wurden Matching-Verfahren eingesetzt. Hierbei wurde ein exaktes Matching auf Alter und Geschlecht vorgenommen. Zur Berücksichtigung weiterer Risikofaktoren für COVID-19-Folgeerkrankungen wurde ein Propensity-Score-Matching durchgeführt.<sup>21</sup> Dieser Ansatz erlaubt die Berücksichtigung einer Vielzahl relevanter Vorerkrankungen und weiterer Risikofaktoren, die den Einfluss einer COVID-19-Erkrankung konfundieren könnten. Die Schätzung von Unterschieden zwischen COVID-19- und Kontrollgruppe hinsichtlich der Inzidenz von Erkrankungen erfolgte über ein Poisson-Modell, sodass relative Risiken (Incidence Rate Ratios) abgeleitet werden konnten.

#### **3.4.5 Arbeitsmodus im Konsortium**

Das Projekt konnte innerhalb weniger Wochen Antworten auf drängende Fragen bereitstellen. Verantwortlich für den Erfolg des Use Cases war ein sehr effizienter, kollegialer Arbeitsmodus aller Beteiligten und deren Bereitschaft, rasch und einvernehmlich methodische und auch Management-Entscheidungen zu treffen. Im Konsortium wurde so ein effizienter und datenschutzkonformer Modus der Datenaufbereitung und -analyse etabliert. Die beteiligten Krankenkassen und Forschungsinstitute brachten erfahrene Statistiker:innen und Data Scientists in das Konsortium ein. Unter Koordination des ZEGV wurden a priori Studienprotokolle und statistische Analysepläne zwischen allen beteiligten Wissenschaftler:innen abgestimmt, entsprechende Skripte für die Analysen entwickelt und im Konsortium geteilt und validiert. Hierfür wurde eine Projektmanagementplattform (Confluence) verwendet. In wöchentlichen Jour-fixe-Terminen wurden methodische Probleme unter den beteiligten Mitarbeiter:innen diskutiert und Lösungen erarbeitet.

Die Analysen wurden getrennt für die Versicherten jeder beteiligten Krankenkasse innerhalb der einzelnen Krankenkassen selbst oder durch autorisierte Forschungsinstitute durchgeführt. Die aggregierten und weitgehend in standardisierter Form vorliegenden Ergebnisse der Analysen der einzelnen Krankenkassen-Datenbestände wurden abschließend an das ZEGV zur Evidenzsynthese versandt.

Alle Analysen erfolgten im Einklang mit der Guten Praxis Sekundärdatenanalyse der Deutschen Gesellschaft für Epidemiologie unter Einbezug der aktuellen internationalen wissenschaftlichen Evidenz sowie der medizinischen Expertise und Kenntnis der GKV-Routinedaten im Konsortium.

### 3.4.6 Gegenwärtiger Stand

Zum aktuellen Zeitpunkt (Dezember 2021) sind die Analysen zum Vergleich von Personen mit dokumentierter COVID-19-Erkrankung und Kontrollpersonen aus der Allgemeinbevölkerung weit fortgeschritten. Neben der Datenaufbereitung wurden Skripte zur Durchführung des Matchings, zur Erstellung deskriptiver Statistiken sowie zur Aggregation der Daten zwecks Evidenzsynthese erstellt, im Konsortium geteilt und validiert. Die Ergebnisse wurden im Preprint veröffentlicht und bei einem Peer-Review-Journal eingereicht.<sup>22</sup> Die Studie fand ein sehr breites Presse-echo und wurde von Prof. Dr. Wieler in der Bundespressekonferenz vom 03.11.2021 diskutiert. Die Ergebnisse wurden auch in der Sitzung der STIKO vorgestellt und diskutiert.

### 3.4.7 Erkenntnisse

Im Projektverlauf konnte eine Vielzahl relevanter organisatorischer, datenschutzrechtlicher und methodischer Probleme effizient gelöst werden. Zu den wichtigsten Erkenntnissen aus dem Projekt zählen:

- Durch die interdisziplinäre Zusammenarbeit von epidemiologischen, medizinischen und methodischen Expert:innen im Konsortium konnten valide und robuste Ergebnisse auf Basis von GKV-Routinedaten erzielt werden.
- Die Abstimmung in Bezug auf die verwendete (Statistik-)Software ermöglichte die gemeinsame Entwicklung, Prüfung und Anwendung von Auswertungsskripten. Der Austausch aggregierter Statistiken erlaubte die datenschutzkonforme und zeitnahe Zusammenführung der Ergebnisse einzelner Analysestellen. Auf diese Weise können die Vorteile einer breiten Datenbasis genutzt werden, ohne eine aufwendige und rechtlich problematische Zusammenführung der Individualdaten zu erfordern. Hierdurch können Analysen zeitnah und mit geringer Fehleranfälligkeit durchgeführt werden.
- Entscheidender Erfolgsfaktor war die partnerschaftliche Zusammenarbeit von Krankenkassen, Versorgungsforscher:innen, Kliniker:innen und dem RKI.
- Die bisherigen Erfahrungen und Ergebnisse demonstrieren die großen Potenziale des Projekts und die Stärken des in seinem Rahmen entwickelten Arbeitsmodus. Diese Potenziale sollten in künftigen Forschungsarbeiten weiter ausgeschöpft werden.



### **3.5 Use Case 5: COVID-Psy-Routinedatenanalysen zu Veränderungen der Inanspruchnahme im psychiatrisch-psychotherapeutischen GKV-Versorgungssystem infolge der COVID-19-Pandemie**

#### **3.5.1 Ausgangssituation**

Die Fragestellung dieses Use Cases stand beispielhaft für eine in vielen Bereichen der Gesundheitsversorgung beobachtete eingeschränkte Kapazität von Krankenhausaufnahmen sowie erbrachter medizinischer Leistungen infolge des bundesweiten Lockdowns während der COVID-19-Pandemie im Frühjahr und im Winter 2020. Dabei ging es darum, ex post facto den Grad einer temporären Unterversorgung infolge des Lockdowns in einer Pandemiesituation quantitativ abschätzen zu können.

Vor diesem Hintergrund ist die Forschergruppe „psychiatrische Versorgungsforschung“ der Universitätsmedizin Mainz an die bestehende Arbeitsgruppe aus AP 0 des NUM-Projekts egePan Unimed herangetreten. Übergreifendes Ziel der Studie waren die Charakterisierung von Veränderungen von Angebot und Inanspruchnahme im ambulanten und stationären psychiatrisch-psychotherapeutischen Versorgungssystem in der COVID-19-Pandemie sowie die Untersuchung von sektorübergreifenden Versorgungswegen.

Die Inanspruchnahme psychiatrischer Leistungen wurde dabei im ersten Projektteil separat für den stationären und ambulanten Sektor betrachtet. Der Setting-Übergang stationär-ambulant stellt auch bei den psychischen Erkrankungen eine große Schwachstelle im deutschen Gesundheitssystem dar, die möglicherweise in der Pandemie noch verstärkt wurde. Dazu wurden im zweiten Projektteil mit der Schizophrenie und den unipolaren majoren Depressionen zwei exemplarische Krankheitsbilder ausgewählt, bei denen eindeutige Leitlinienempfehlungen für die längerfristige, sektorübergreifende Behandlung existieren und andererseits Defizite in der Umsetzung im Versorgungssystem bekannt sind.

#### **3.5.2 Beteiligte Partner**

Der Use Case wurde in einer Arbeitsgruppe von Akteur:innen mit je unterschiedlichen inhaltlichen Schwerpunkten bearbeitet:

- mit GKV-Routinedatenanalysen erfahrene Forschungsinstitute (ZEGV, InGef) und
- klinische Expert:innen aus der Klinik für Psychiatrie und Psychotherapie des Universitätsklinikums Mainz.

Konkrete Fragestellungen und die methodische Umsetzung wurden unter Koordination des ZEGV gemeinsam durch die Arbeitsgruppe festgelegt.

### 3.5.3 Fragestellungen

Das Projekt adressierte folgende Fragestellungen:

1. Lassen sich Veränderungen der Inanspruchnahme stationärer psychiatrisch-psychotherapeutischer und psychosomatisch-psychotherapeutischer Behandlungen von Patient:innen der Erwachsenen-, Kinder- und Jugendpsychiatrie in der COVID-19-Pandemie feststellen? Gibt es Patientengruppen, Settings und Regionen, die davon besonders betroffen waren?
2. Lassen sich Veränderungen der Inanspruchnahme ambulanter psychiatrischer, psychosomatischer und psychotherapeutischer Leistungen in der COVID-19-Pandemie feststellen? Gibt es Patientengruppen, Settings und Regionen, die davon besonders betroffen waren?
3. Wie hat sich die Umsetzung von in Leitlinien empfohlener sektorenübergreifender Follow-up-Behandlung, spezifisch der Übergang von stationärer in ambulante Behandlung, während der COVID-19-Pandemie verändert?

### 3.5.4 Methodik

Die Studie wurde als longitudinale Beobachtungsstudie auf Grundlage von GKV-Routinedaten konzipiert. Es wurden innerhalb von vordefinierten Zeitintervallen und über einen Gesamtzeitraum von Januar 2016 bis März 2021 alle Versicherten mit psychischer und Verhaltensstörung hinsichtlich ihrer Inanspruchnahme von Leistungen des Gesundheitssystems in die Studie eingeschlossen. Besonderer Fokus lag in Veränderungen der Inanspruchnahme zum Zeitpunkt des ersten und zweiten Lockdowns (März 2020 und Dezember 2020) im Vergleich zum interventionsfreien Vorzeitraum (Januar 2016 bis Februar 2020).

#### **Datengrundlage und Studienpopulation:**

Es konnten 8,6 Millionen Versicherte der Krankenkassen AOK PLUS und der BKKen in die Analysen eingeschlossen werden. Für diese Versicherten können sämtliche bei den Krankenkassen abgerechneten Leistungen über einen Zeithorizont von mehreren Jahren nachvollzogen werden. Dies umfasst sowohl ambulante als auch stationäre Diagnosen, Prozeduren (EBM, OPS) und verschriebene Medikamente. Darüber hinaus sind Daten zu Arbeitsunfähigkeitsbescheinigungen (AU) und Sterbedaten verfügbar. Die Analyse wurde ab August 2021 durchgeführt. Es konnten dabei GKV-Routinedaten bis einschließlich des 1. Quartals 2021 in die Analysen einbezogen werden (Latenz sieben bis neun Monate).

#### **Statistische Methodik:**

Mittels einer Zeitreihenanalyse wurde innerhalb einer Kohorte von Personen mit psychischer Erkrankung oder Verhaltensstörung zwischen Januar 2016 und März 2021 der Einfluss des ersten (März 2020) und zweiten (Dezember 2020) Lockdowns der

COVID-19-Pandemie auf das akute stationäre und ambulante Versorgungsangebot untersucht. Die Modellierung erfolgte mittels eines autoregressiven Zeitreihenmodells, das auf likelihoodbasierte Schätzmethoden für die Analyse und Modellierung von Zählzeitreihen nach regularisierten verallgemeinerten linearen Modellen zurückgreift. Die Analyse wurde mithilfe des `tscout`-Packages<sup>23</sup> in der Statistiksoftware R<sup>24</sup> durchgeführt.

### 3.5.5 Arbeitsmodus im Konsortium

Verantwortlich für den Erfolg des Use Cases war ein sehr effizienter, kollegialer Arbeitsmodus aller Beteiligten und deren Bereitschaft, rasch und einvernehmlich methodische und auch Management-Entscheidungen zu treffen. Im Konsortium wurde so ein effizienter und datenschutzkonformer Modus der Datenaufbereitung und -analyse etabliert. Die beteiligten Krankenkassen und Forschungsinstitute brachten erfahrene Statistiker:innen und Data Scientists in das Konsortium ein. Unter Koordination des ZEGV wurden a priori Studienprotokolle und statistische Analysepläne zwischen allen beteiligten Wissenschaftler:innen abgestimmt, entsprechende Skripte für die Analysen entwickelt und im Konsortium geteilt und validiert. Hierfür wurde eine Projektmanagementplattform (Confluence) verwendet. In wöchentlichen Jour-fixe-Terminen wurden methodische Probleme unter den beteiligten Analyst:innen diskutiert und Lösungen erarbeitet.

Die Analysen wurden durch autorisierte Forschungsinstitute (ZEGV, InGef) durchgeführt. Die aggregierten Ergebnisse der Analysen der einzelnen Krankenkassen-Datenbestände wurden abschließend an das ZEGV zur Evidenzsynthese versandt.

Alle Analysen erfolgten im Einklang mit der Guten Praxis Sekundärdatenanalyse der Deutschen Gesellschaft für Epidemiologie unter Einbezug der aktuellen internationalen wissenschaftlichen Evidenz sowie der medizinischen Expertise und Kenntnis der GKV-Routinedaten im Konsortium.

Unterstützt wurden die spezifischen Aktivitäten innerhalb dieser Arbeitsgruppe durch zwei Workshops des egePan-Konsortiums zur wissenschaftlichen Nutzung von Sekundärdaten, besonders der Abrechnungsdaten der gesetzlichen Krankenversicherung im Februar und April 2021.\*

---

\*<https://egepan.de/workshop-routinedaten/>



### 3.5.6 Ergebnisse (Stand Dezember 2021)

Zum aktuellen Zeitpunkt sind die Analysen bezüglich der stationären und ambulanten Inanspruchnahme weit fortgeschritten. Neben der Datenaufbereitung wurden Skripte zur Erstellung deskriptiver Statistiken sowie zur Aggregation der Daten zwecks Evidenzsynthese erstellt. Es ist zu erwarten, dass erste umfassende Analyseergebnisse bis Ende Februar 2022 vorliegen.

Zusammenfassend zeigen die Analysen, dass die bundesweiten Lockdowns mit erheblichen Einschnitten im stationären Sektor der psychiatrischen Versorgung einhergingen. Sowohl vollstationäre als auch teilstationäre Kapazitäten wurden offensichtlich reduziert. Im Vergleich zum Basiszeitraum (Januar 2016 bis Januar 2020) war die Aufnahme- und Behandlungsdauer verringert. Im ambulanten Bereich kam es vor allem im ersten Lockdown zu einer Verringerung inzidenter Erstdiagnosen. Bei der Weiterbehandlung prävalenter Fälle wurden im Rahmen der Analyse keine Einschnitte in Form verringerter Therapieangebote oder psychopharmazeutischer Verschreibungen festgestellt.

### 3.5.7 Erkenntnisse

Im Projektverlauf konnte eine Vielzahl relevanter organisatorischer, datenschutzrechtlicher und methodischer Probleme effizient gelöst werden. Zu den wichtigsten Erkenntnissen aus dem Projekt zählen:

- Der bestehende Zugang zu großen Korpi an Routinedaten aufseiten erfahrener Institute (ZEGV, InGef) ermöglicht die Beantwortung akuter pandemiebedingter Fragestellungen.
- Durch die interdisziplinäre Zusammenarbeit von epidemiologischen, medizinischen und methodischen Expert:innen im Konsortium konnten valide und robuste Ergebnisse auf Basis von GKV-Routinedaten erzielt werden.
- Die Abstimmung in Bezug auf die verwendete (Statistik-)Software ermöglichte die gemeinsame Entwicklung, Prüfung und Anwendung von Auswerteskripten. Hierdurch können Analysen zeitnah und mit geringer Fehleranfälligkeit durchgeführt werden.
- Der Einsatz metaanalytischer Verfahren erlaubte die datenschutzkonforme Zusammenführung der Ergebnisse einzelner Analysestellen. Auf diese Weise können die Vorteile einer breiten Datenbasis genutzt werden, ohne eine aufwendige und rechtlich problematische Zusammenführung der Individualdaten zu erfordern.

## 4. Handlungsempfehlungen und Fazit

### 4.1 Allgemeine Handlungsempfehlungen

Eine wesentliche Herausforderung des Projekts bestand darin, eine in kurzer Zeit geeignete Datenbasis zu schaffen, mit der aktuell drängende Fragen zum Pandemiemanagement zeitnah und wissenschaftlich fundiert beantwortet werden können.

Die Erfahrungen aus diesem Projekt resultieren in erster Linie aus den oben vorgestellten Use Cases sowie aus mehreren Workshops im Netzwerk Universitätsmedizin (NUM) und aus semistrukturierten Einzelinterviews mit Kliniker:innen und Wissenschaftler:innen. Die Empfehlungen können untergliedert werden in allgemeine Handlungsempfehlungen, die eher dem Projektmanagement zuzuordnen sind, und spezifische Handlungsempfehlungen, die die Nutzung von GKV-Routinedaten in einer pandemischen Lage betreffen.

Die allgemeinen Handlungsempfehlungen betreffen Grundsätze des wissenschaftlichen Arbeitens/Projektmanagements und lassen sich wie folgt skizzieren:

- Benennung einer verantwortlichen Projektleitung
- Einbindung von Personen mit der notwendigen fachlichen Expertise (je nach Fragestellung): Epidemiologie, Statistik, Data Science, Medizin
- frühzeitige Abstimmung der Zielsetzung des jeweiligen Projekts mit allen kooperierenden Institutionen
- prospektive Verschriftlichung von Studienprotokollen und Analyseplänen
- Abhaltung regelmäßiger Projektmeetings zur Abstimmung praktischer und methodischer Vorgehensweisen
- Festlegung des Datenauswertungsmodus (statistischer Analyseplan)
- Definition der Adressat:innen der Ergebnisse
- Diskussion der Verwertung der Ergebnisse während der Planungsphase
- Nutzung einer gemeinsamen sicheren Online-Plattform zur Dokumentation aller projektrelevanten Festlegungen, Prozessschritte und Ergebnisse (z. B. Confluence) und damit Gewährleistung eines jederzeit einheitlichen Informationsstands aller Projektbeteiligten
- Definition und Operationalisierung der Ein- und Ausschlusskriterien und spezifischer Outcomes bei Beachtung aktueller Klassifikationen (z. B. ICD-Code für Post-COVID)

Wichtig ist, dass mit GKV-Routinedatenanalysen erfahrene Wissenschaftler:innen an Analysen beteiligt werden. Hier hat sich die Zusammenarbeit von wissenschaftlichen Instituten mit Schwerpunkt GKV-Routinedatenanalyse und Expert:innen von gesetzlichen Krankenkassen zum Umgang mit GKV-Rohdaten bewährt. Das Leistungsgeschehen

ist in Routinedaten durch die Verwendung verschiedener Klassifikationssysteme dokumentiert: ICD für die Dokumentation von Diagnosen; OPS und EBM für das Leistungsgeschehen im stationären und ambulanten Sektor; ATC für die medikamentöse Therapie; Heil- und Hilfsmittelkatalog für den sogenannten nicht ärztlichen Leistungserbringer.

Dabei muss im Fall von Pandemien geklärt werden, ob die relevanten Entitäten wie z. B. die maßgebliche Infektion oder deren klinische Symptomatik und Folgeerkrankungen sowie Gegenmaßnahmen bereits Aufnahme in diese Klassifikationssysteme gefunden haben. Zudem ist auch zu beachten, ob diese Klassifikationssysteme international gepflegt werden (ICD/ATC) oder, wie der EBM, nur innerhalb von Deutschland Anwendung finden.

In der Coronapandemie galt es also zu klären, inwieweit die Infektion mit SARS-CoV-2 und das klinische Erscheinungsbild von COVID-19 bzw. Post-COVID sowie die Coronaimpfungen kurzfristig Aufnahme gefunden haben. Für die Coronainfektion ist dies seit dem Frühjahr 2020 mit dem neu belegten ICD-Code U07.x der Fall, für die Krankheitsfolgen von COVID-19 und Post-COVID mit den Ziffern U08 bis U10. Für die Impfung mit Corona wird seit 2021 der Code U11.9 vergeben, in Verbindung mit einigen EBM-Pseudoziffern für die konkreten Impfstoffe. Somit ist die Abbildbarkeit in den Abrechnungsdaten inzwischen gegeben.

#### **Empfehlungen zur strukturellen Weiterentwicklung der Nutzung von Routinedaten für das Pandemiemanagement und die Versorgungsforschung**

Bei aktuell 97 Krankenkassen (Stichtag 01.01.2022) besteht die Herausforderung, für einzelne Fragestellungen kooperierende Krankenkassen zu gewinnen. Hierbei zeigt sich der Strukturvorteil, wenn mehrere Krankenkassen zu Analyseeinheiten zusammengefasst sind (z. B. InGef für BKKen). Die Anwendungsfälle Post-COVID und Impfpriorisierung haben gezeigt, dass durch eine verteilte Analyse in den Krankenkassen bzw. nutzungsberechtigten Stellen und die anschließende Zusammenführung der Ergebnisse diese Aufsplittung bewältigt werden kann. Inwieweit eine solche Struktur in zukünftigen Projekten als Auswertungsstandard genutzt werden kann, müssen weitere Anwendungsfälle zeigen. Bestrebungen, die Daten aller Kassen in dem im Aufbau befindlichen FDZ Gesundheitsdaten (BfArM, Bonn) ab dem Jahr 2022 zusammenzuführen, ist begrüßenswert. Die Daten werden allerdings über den Spitzenverband Bund der GKV mit einem Zeitverzug von zwölf Monaten übermittelt und sind damit für pandemische Versorgungssituationen nicht oder nur mit Einschränkungen geeignet. Inwieweit diese zeitliche Verzögerung minimiert werden könnte, sollte geprüft werden.

Komplementär zu den Krankenkassen werden in weiteren Institutionen Routinedaten gespeichert, allerdings oftmals getrennt nach Sektoren, wie beispielsweise die Daten der ambulanten ärztlichen Versorgung im Zentralinstitut (ZI) und bei der Kassenärztlichen Bundesvereinigung (KBV) oder Daten des stationären Sektors beim Institut für das Entgeltsystem im Krankenhaus (InEK). Im letzteren Fall sind die Daten aller Bundesbür-

ger:innen gespeichert, da auch stationäre Aufenthalte von privat versicherten Personen übermittelt werden. In der Coronapandemie wurden die Meldezeiten von ursprünglich einmal pro Jahr auf bis zu drei Meldezeitpunkte erweitert, um auch unterjährige Daten aktuell verfügbar zu haben. Allerdings wurden die Daten bislang nur für das Gutachten des Expertenbeirats<sup>25</sup> vom InEK im Rahmen einer Auswertungsanfrage genutzt, aggregierte Daten stehen auch über das InEK-Portal\* zur Verfügung, können und dürfen dort aber nur punktuell und manuell abgefragt werden. Auch dieser Datenweg sollte im Hinblick auf eine Routinenutzung geprüft werden. Ziel könnten tagesaktuell verfügbare Daten aus dem stationären Sektor sein, die auch im Sinne eines Frühwarnsystems genutzt werden können. Weiterhin sollte eine weitere Beschleunigung insbesondere der ambulanten Abrechnungsdaten für die wissenschaftliche Nutzung strukturell und normativ weiterentwickelt werden.

## 4.2 Spezifische Handlungsempfehlungen in pandemischer Versorgungslage

Die allgemeinen Handlungsempfehlungen werden hier ergänzt um spezifische Handlungsempfehlungen, die die konkrete Nutzung von GKV-Routinedaten im Rahmen einer Pandemie betreffen.

- (1) Prüfung der Verkürzung üblicher Datenbereitstellung und damit Erhöhung der Aktualität (z. B. Daten der Kassenärztlichen Vereinigung [KV] direkt nutzen, statt den üblichen Übermittlungsweg zur gesetzlichen Krankenversicherung abzuwarten) und ggf. Voraussetzung für schnellere Bereitstellung von GKV-Routinedaten aus einzelnen Versorgungssektoren diskutieren (Nutzung nicht durch KV-geprüfte ambulante Daten in Ausnahmesituationen). Falls eine Datenlieferung kurzfristig auch aufgrund regulatorisch notwendiger Vorarbeiten nicht möglich ist, kann die Möglichkeit der getrennten Datenverarbeitung durch die datenhaltenden Krankenkassen oder nutzungsberechtigte Institutionen mit anschließender (meta-)analytischer Poolung der (aggregierten) Ergebnisse in Betracht gezogen werden.
- (2) Nutzung bereits etablierter Kooperationspartnerschaften zur Initiierung eines kurzfristigen Projekts mit großer Public-Health-Relevanz. So kann geprüft werden, ob über bereits bestehende und bewährte Kooperationen zügig eine Projektgruppe zusammengestellt werden kann. Zusätzlich sollte geprüft werden, ob Daten, die in anderen Projekten bereits extrahiert wurden, als Grundlage genutzt werden können. Voraussetzung dafür bleibt die Zustimmung der zuständigen Aufsichtsbehörde.
- (3) Priorisierung von Fragestellungen, die zur akuten Bewältigung der Pandemie bearbeitet werden müssen und zu deren Beantwortung Routinedaten einen beträchtlichen Beitrag leisten können. Die Priorisierung sollte sich an den Aspekten Public-Health-Relevanz, klinische Relevanz und Grad an Unsicherheit orientieren.
- (4) Bei der Finanzierung von Projekten auch die Aufwandsentschädigungen für die Dateneigner mitberücksichtigen, sofern die kurzfristige Datenbereitstellung nicht

---

\*<https://datenbrowser.inek.org>



aus standardisierten Abläufen heraus möglich ist. Für Fragestellungen mit akuter Public-Health-Relevanz wie eine pandemische Versorgungslage sollten ggf. niedrigschwellige Finanzierungsmöglichkeiten für forschende Institutionen eingerichtet werden. Das Netzwerk Universitätsmedizin ermöglichte es hier, für die fünf Use Cases einen adäquaten Rahmen zu schaffen, um drängende Fragen des Pandemiemanagements mittels GKV-Routinedaten zu beantworten.

- (5) Die Möglichkeit der (einfachen und zeitnahen) Verknüpfung von Routinedaten mit anderen Datenbeständen (z. B. Meldedaten COVID-19, Impfinformationen, Labordaten) kann positive Auswirkungen auf das Pandemiemanagement haben.
- (6) In der gegenwärtigen Pandemie fehlt eine direkte Verknüpfbarkeit von Impfdaten mit GKV-Abrechnungsdaten, sodass wichtige Fragestellungen, etwa zur Impfsurveillance, zur Wirksamkeit der Impfungen und zur Sicherheit der Impfungen in der Bevölkerung, nicht auf Basis der GKV-Routinedaten bearbeitbar sind. Hier sollte durch den Gesetzgeber möglichst rasch die Nutzung der Krankenversicherungsnummer als Unique Identifier für die Dokumentation von Impfdaten ermöglicht werden.

### 4.3 Fazit

Die beschriebenen allgemeinen und spezifischen Handlungsempfehlungen im Sinne eines Leitfadens sollten beachtet werden, um auch in zukünftigen pandemischen Versorgungslagen Routinedaten als solide Datenbasis zu nutzen. Die Anwendungsfälle zeigen auf, welches Potenzial in den Daten steckt und welche Limitationen zu beachten sind.

Das Arbeitspaket „Nutzung von Routinedaten für das evidenzgeleitete Pandemiemanagement (AP 0)“ des Verbundprojekts egePan Unimed konnte auf Basis von GKV-Routinedaten einen substanziellen Beitrag zum Management der COVID-19-Pandemie leisten. Zusätzlich zu dem schnell geschaffenen finanziellen Projektrahmen haben bestehende Kooperationen, besonderes Engagement und hohe fachliche Expertise der beteiligten forschenden und datenhaltenden Institutionen und ihrer Mitarbeitenden maßgeblich zum Erfolg des Projekts und der oben beschriebenen Use Cases beigetragen.

Nur bei genauer Kenntnis der Stärken und Schwächen von GKV-Routinedaten und viel Erfahrung im Umgang mit komplexen Abläufen bei der Aufbereitung und Analyse dieser Daten sind fundierte Ergebnisse möglich. Wir empfehlen, dass sich möglichst mehrere Wissenschaftler:innen und Institute zusammentun, um bei der Bearbeitung von Forschungsfragen mögliche Fallstricke zu vermeiden und bestmögliche Evidenz aus den GKV-Routinedaten ableiten zu können. Das Netzwerk Universitätsmedizin (NUM) bietet nach Einschätzung der Autor:innen einen hervorragenden Rahmen dafür, effizient methodische Expertise mit klinischer Expertise der Universitätsmedizin zu verbinden, um bei Pandemien rasch drängende Fragen zur Patientenversorgung und zum Pandemiemanagement beantworten zu können. Hierzu bilden auch GKV-Routinedaten eine wesentliche Ressource.



## 5. Quellen

1. Pfaff H, Glaeske G, Neugebauer EA, Schrappe M. [Memorandum III: „Methods for Health Services Research“ (Part 1)]. *Gesundheitswesen*. 2009;71(8-9):505-510.
2. Neugebauer EA, Icks A, Schrappe M. [Memorandum III: Methods for Health Services Research (Part 2)]. *Gesundheitswesen*. 2010;72(10):739-748.
3. Schipf S, Schöne G, Schmidt B, et al. Die Basiserhebung der NAKO Gesundheitsstudie: Teilnahme an den Untersuchungsmodulen, Qualitätssicherung und Nutzung von Sekundärdaten. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*. 2020;63(3):254-266.
4. Stallmann C, Ahrens W, Kaaks R, Pigeot I, Swart E, Jacobs S. Individuelle Datenverknüpfung von Primärdaten mit Sekundär- und Registerdaten in Kohortenstudien: Potenziale und Verfahrensvorschläge. *Gesundheitswesen*. 2015;77(02):e37-e42.
5. Schmidt CO, Reber K, Baumeister SE, Schminke U, Völzke H, Chenot JF. Die Integration von Primär- und Sekundärdaten in der Study of Health in Pomerania und die Beschreibung von klinischen Endpunkten am Beispiel Schlaganfall. *Gesundheitswesen*. 2015;77(02):e20-e25.
6. Swart E, Gothe H, Geyer S, et al. Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen. *Gesundheitswesen*. 2015;77(02):120-126.
7. March S, Andrich S, Drepper J, et al. Gute Praxis Datenlinkage (GPD). *Gesundheitswesen*. 2019;81(08/09):636-650.
8. Swart E, Schmitt J. STandardized Reporting Of Secondary data Analyses (STROSA) – Vorschlag für ein Berichtsformat für Sekundärdatenanalysen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2014;108(8):511-516.
9. Swart E IP, Gothe H, Matusiewicz D. *Routinedaten im Gesundheitswesen – Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven*. 2 ed: Huber; 2014.
10. Statistisches Bundesamt. Das Informationssystem der Gesundheitsberichterstattung des Bundes. <https://www.gbe-bund.de>. Published 2022. Accessed 24/01/2022.
11. Bundesministerium für Gesundheit. Verordnung zum Anspruch auf Schutzimpfung gegen das Coronavirus SARS-CoV-2 (Coronavirus-Impfverordnung – CoronImpfV). [https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/C/Coronavirus/Verordnungen/Corona-ImpfV\\_BAnz\\_AT\\_11.03.2021\\_V1.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/C/Coronavirus/Verordnungen/Corona-ImpfV_BAnz_AT_11.03.2021_V1.pdf). Published 2021. Accessed 2021.
12. Abbas S, Ihle P, Köster I, Schubert I. Estimation of disease incidence in claims data dependent on the length of follow-up: a methodological approach. *Health Serv Res*. 2012;47(2):746-755.
13. Schubert I, Köster I, Küpper-Nybelen J, Ihle P. Versorgungsforschung mit GKV-Routinedaten. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*. 2008;51(10):1095-1105.

14. Swart E, Ihle P. Epidemiologische Studien auf Basis von Sekundärdaten. *Public Health Forum*. 2012;20(3):16-17.
15. 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(1):212.
16. Rößler M, Jacob J, Risch L, et al. Hierarchisierung von Risikofaktoren für schwere COVID-19-Erkrankungsverläufe im Kontext der COVID-19-Schutzimpfungen. 2021(19):3-12.
17. Paul-Ehrlich-Institut. Sicherheitsbericht – Verdachtsfälle von Nebenwirkungen und Impfkomplicationen nach Impfung zum Schutz vor COVID-19. [https://www.pei.de/SharedDocs/Downloads/DE/newsroom/dossiers/sicherheitsberichte/sicherheitsbericht-27-12-bis-02-04-21.pdf?\\_\\_blob=publicationFile&v=4](https://www.pei.de/SharedDocs/Downloads/DE/newsroom/dossiers/sicherheitsberichte/sicherheitsbericht-27-12-bis-02-04-21.pdf?__blob=publicationFile&v=4). *Published 2021. Accessed 2021.*
18. Robert Koch-Institut. Epidemiologisches Bulletin 16/2021 – Beschluss der STI-KO zur 4. Aktualisierung der COVID-19-Impfempfehlung und die dazugehörige wissenschaftliche Begründung. [https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2021/Ausgaben/16\\_21.html](https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2021/Ausgaben/16_21.html). *Published 2021. Accessed 10/02/2022.*
19. 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.
20. Robert Koch-Institut. Epidemiologisches Bulletin 33/2021: Beschluss der STI-KO zur 9. Aktualisierung der COVID-19-Impfempfehlung und die dazugehörige wissenschaftliche Begründung / Intensivierte genombasierte Surveillance von *Campylobacter*-Isolaten humaner Erkrankungsfälle in Deutschland. [https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2021/Ausgaben/33\\_21.html](https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2021/Ausgaben/33_21.html). *Published 2021. Accessed 10/02/2022.*
21. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
22. 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.
23. 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(5):1-51.
24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria 2018.
25. Bundesministerium für Gesundheit. Bericht des Expertenbeirats zu den Auswirkungen des COVID-19-Krankenhausentlastungsgesetzes <https://www.bundesgesundheitsministerium.de/ministerium/meldungen/2020/bericht-expertenbeirat.html>. *Published 2021. Accessed:10/02/2022*

## Nationales Forschungsnetzwerk der Universitätsmedizin zu Covid-19

Patient:innen optimal versorgen, Infektionen verhindern, Gesundheitsversorgung erhalten: Die Covid-19-Pandemie bringt Herausforderungen mit sich, die innerhalb kurzer Zeit neue Handlungsstrategien erfordern. Das Nationale Forschungsnetzwerk der Universitätsmedizin zu Covid-19, kurz Netzwerk Universitätsmedizin (NUM), bündelt und stärkt Forschungsaktivitäten zur Bewältigung der aktuellen Lage. Gefördert durch das Bundesministerium für Bildung und Forschung, koordiniert durch die Charité – Universitätsmedizin Berlin, arbeitet das Forschungsnetzwerk unter Beteiligung aller deutschen Universitätsklinika und weiterer Netzwerke an Lösungen für eine bestmögliche Patientenversorgung in der Pandemie. 13 umfängliche Verbundprojekte mit Leitungen an den verschiedenen Standorten der Universitätsmedizin sind hierfür in einer ersten Förderphase konzipiert worden. Das Programm ist auf schnelle, unmittelbare Unterstützungswirkungen ausgerichtet. Ein Akzent liegt auf der kliniknahen Forschung und Versorgungsforschung, deren Ergebnisse gemäß dem translationalen Ansatz direkt in Versorgung und Krisenmanagement einfließen. Dem Forschungsnetzwerk und den beteiligten Einrichtungen stehen zur Umsetzung dieser Aufgabe rund 150 Millionen Euro im ersten Jahr bereit, seit 2021 wird das Netzwerk bis zum Jahr 2024 mit weiteren 80 Millionen Euro jährlich bzw. zusätzlichen 240 Millionen Euro gefördert. Die gemeinsamen Entwicklungen in Forschung und Patientenversorgung, evidenzbasiertes Vorgehen sowie gegenseitiges Lernen sollen zu einem gemeinsamen Vorgehen bei der Pandemiebekämpfung und einer „Pandemic Preparedness“ führen.

Weitere Informationen: <https://www.netzwerk-universitaetsmedizin.de>



