

# 5 Leitfragen für die Versorgungsforschung – Eignen sich GKV-Routinedaten für Ihr Forschungsvorhaben?

## Five Key Questions for Health Services Research: are SHI Claims Data Suitable for Your Research Project?



Autorinnen/Autoren

Peter Ihle<sup>1‡</sup>, Udo Schneider<sup>2‡</sup>, Verena Vogt<sup>3,4‡</sup>

### Institute

- 1 PMV forschungsgruppe an der Medizinischen Fakultät und Uniklinik Köln, Universität zu Köln, Köln, Germany
- 2 Versorgungsmanagement, Techniker Krankenkasse, Hamburg, Germany
- 3 Health Care Management, Technische Universität Berlin, Berlin, Germany
- 4 Institut für Allgemeinmedizin Universitätsklinikum Jena, Germany

### Schlüsselwörter

Leitfragen, Routinedaten, Datenzugang, Datennutzung, Versorgungsforschung

### Key words

Guiding questions, claims data, routine data, use and access of data, health services research

online publiziert 2023

### Bibliografie

Gesundheitswesen

DOI 10.1055/a-2098-3039

ISSN 0941-3790

© 2023. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag, Rüdigerstraße 14,  
70469 Stuttgart, Germany

### Korrespondenzadresse

Dr. med. Peter Ihle  
PMV forschungsgruppe  
Universitätsklinikum Köln  
Herderstr. 52  
50931 Köln  
Germany  
[peter.ihle@uk-koeln.de](mailto:peter.ihle@uk-koeln.de)

### ZUSAMMENFASSUNG

Die Versorgungsforschung untersucht die Versorgungsstrukturen und -prozesse der Gesundheitsversorgung unter Alltagsbedingungen. Routinedaten der Gesetzlichen Krankenversicherung (GKV) – sogenannte versorgungsnahe Daten – bilden die reale Versorgung ab und sind daher eine wichtige Datenquelle für die Versorgungsforschung. Dieser Beitrag präsentiert 5 Leitfragen, mit deren Hilfe Forschende und datenhaltende Institutionen die Eignung dieser Datenquelle für die Beantwortung ihrer Fragestellung der Versorgungsforschung prüfen können. Ziel dieser Leitfragen ist es, zwischen Forschenden und datenhaltenden Institutionen ein gemeinsames Verständnis des Forschungsvorhabens, des Forschungsziels sowie der Umsetzbarkeit in der Versorgungsforschung zu generieren. Die 5 Leitfragen spannen den Bogen von der Formulierung der Forschungsfrage und der geplanten Methode, über die Zielpopulation, die relevanten Untersuchungszeiträume bis hin zu den erforderlichen Informationen aus den GKV-Routinedaten. Vier methodisch ausgerichtete Leitfragen werden um die Frage ergänzt, wie die Ergebnisse des Forschungsprojektes die Versorgung verbessern könnten. Diese 5 Leitfragen stellen somit für Forschende eine erste Strukturierung für Datenanfragen dar; für datenhaltende Institutionen liefern sie einen Rahmen, um eine mögliche Beteiligung oder Unterstützung einer Forschungsfrage in der Versorgungsforschung zu prüfen.

### ABSTRACT

Health services research examines the structures and processes of health care under everyday conditions. Routine data of the statutory health insurance (SHI) – the so-called routine practice data – represent real health care and are therefore an important data source for health services research. This paper presents 5 key questions that researchers and data-holding institutions can use to assess the suitability of this data source for answering their health services research question. The aim of these guiding questions is to generate a common understanding between researchers and data-holding institutions of the research project, the research objective, and the feasibility of implementation in health services research. The five guiding questions cover the formulation of the research question, the

‡ Diese Autorinnen/Autoren haben zu gleichen Teilen beigetragen

planned method, the target population, the relevant study periods, and the required information from SHI data. These methodologically oriented guiding questions are supplemented by the question of how the results of the research project could improve care. Thus, for researchers, the five guiding questions

provide an initial structuring for data requests; for data-holding institutions, they provide a framework for considering possible involvement in or support of a research idea in health services research.

## Einleitung

Versorgungsforschung untersucht „Versorgungsstrukturen und -prozesse der Gesundheitsversorgung“ unter Alltagsbedingungen [1]. Eine wichtige Datenquelle für die Versorgungsforschung bilden die Routinedaten der Gesetzlichen Krankenversicherung (GKV), die als sogenannte versorgungsnahen Daten die reale Versorgung abbilden [2]. Für die Nutzung dieser Daten bedarf es der Kooperation zwischen Versorgungsforschern und datenhaltenden Institutionen. An dieser Schnittstelle zwischen Versorgung und Forschung setzt dieser Beitrag an und präsentiert 5 Leitfragen, mit deren Hilfe Versorgungsforscher – idealiter unter Einbeziehung klinisch Tätiger – in den strukturierten Austausch mit datenhaltenden Institutionen gehen können.

Die Leitfragen fokussieren auf die Durchführung von Versorgungsforschungsprojekten auf Basis von GKV-Routinedaten. Diese Daten fallen bei der Abrechnung und Dokumentation medizinischer Leistungen an und werden daher auch als GKV-Abrechnungsdaten bezeichnet. Sie werden also nicht primär für Forschungszwecke generiert bzw. erhoben und sind daher sog. Sekundärdaten. GKV-Routinedaten werden in der Versorgungsforschung eingesetzt, um z. B. Krankheitskosten zu berechnen oder die Versorgungsqualität einer ausgewählten Erkrankung zu beschreiben. Auch zur Untersuchung kausaler Fragestellungen werden die Daten herangezogen wie etwa zum Zusammenhang zwischen dem erbrachten Leistungsvolumen und der erzielten Ergebnisqualität (siehe z. B. [3]) oder zu potenziellen Einflussfaktoren auf Leitlinienadhärenz (siehe z. B. [4]). Seit der Einführung des Innovationsfonds mit dem GKV-Versorgungsstärkungsgesetz im Jahr 2015 und den Schwerpunkten Neue Versorgungsformen und Versorgungsforschung (vgl. § 92a SGB V) kommen Routinedaten immer häufiger zum Einsatz.

Zu den Vorteilen der GKV-Routinedaten gehört, dass diese das gesamte zu Lasten der GKV abgerechnete Leistungsgeschehen aller GKV-Versicherten umfassen. Da diese Daten alle Leistungssektoren einschließen, kann mit ihnen das Versorgungsgeschehen sektorenübergreifend abgebildet werden. Somit können beispielsweise Versorgungsbrüche an den Schnittstellen der Versorgung (ambulant/stationär) untersucht werden. Neben den erbrachten Leistungen wird auch die Morbidität der Versicherten durch die Leistungserbringer (mittels ICD kodierter Diagnosen) dokumentiert. Der Personenbezug der Daten erlaubt es Forschenden, Krankheitsinzidenzen und -prävalenzen abzuschätzen, Krankheitsverläufe und Versorgungswege darzustellen (siehe z. B. [6]) oder Analysen unter Einbezug von GKV-Kosten durchzuführen [7, 8]. Durch die Vollerhebung – alle Krankenversicherten der jeweiligen Krankenkasse sind einbezogen – sind auch Informationen über die Inanspruchnahme von vulnerablen Gruppen enthalten, beispielsweise Hochbetagte, Schwerkranke, Heimbewohner oder Verstorbene, die i. d. R. in Studien mit aktiver Rekrutierung nicht oder nur sehr

schwer erreichbar sind. Die Daten stehen zudem über mehrere Jahre für die Analysen zur Verfügung. Die konkrete Datenverfügbarkeit ist dabei abhängig von im Sozialgesetzbuch genannten Löschrufen sowie der Datenhaltung innerhalb der angefragten Institutionen.

Daneben existieren Einschränkungen, die bzgl. der Eignung dieser Datenquelle für die Forschung berücksichtigt werden müssen. So gilt es zu beachten, dass die Daten zu Abrechnungszwecken generiert werden und damit deren Logik folgen. Diese abrechnungsspezifischen Datenstrukturen limitieren die Auswertungsmöglichkeiten. So werden Diagnosen im ambulanten Sektor nur auf Quartalsebene dokumentiert. Auch lässt sich im ambulanten Sektor kein direkter Zusammenhang zwischen Diagnose und abgerechneter Gebührenordnungsziffer bzw. der Verordnung eines Arzneimittels herstellen. Durch vorgegebene Löschrufen kann oftmals die Inzidenz einer Erkrankung nicht genau bestimmt werden, da der Blick in die Versorgungshistorie eingeschränkt ist. Zusätzlich ist auch der Lieferverzug der GKV-Daten zu beachten, der je nach Sektor unterschiedlich lang ausfällt (siehe hierzu auch Leitfrage 3). Klinische Parameter wie Laborwerte, Blutdruck oder auch Schweregrade einer Erkrankung sind nicht oder nur in Ansätzen aus den Daten ermittelbar. Mit wenigen Ausnahmen gilt dies auch für Arzneimittelgaben im stationären Sektor. Das Versichertenkollektiv einer einzelnen Kasse weicht in aller Regel von der gesamten GKV-Bevölkerung ab und ist somit für diese hinsichtlich bestimmter Einflussgrößen nicht repräsentativ [9]. Zudem bedarf es spezifischer methodischer Aufbereitungs-, Plausibilisierungs- und Analyse-schritte, um die Daten für die Beantwortung der Fragestellung nutzbar zu machen.

Zu den Vor- und Nachteilen der GKV-Routinedaten, ihren Inhalten und Information existieren bereits zahlreiche Publikationen (siehe zur Methodik unter anderem [10–12] oder zu leitliniengerechter Handhabung [13–15]). Die Kenntnis der Besonderheiten der Daten, insbesondere der Vor- und Nachteile der GKV-Routinedaten, ist notwendig, um sie für die Versorgungsforschung einsetzen zu können. Dabei ist der iterative Austausch zwischen (klinisch) Forschenden und datenhaltenden Institutionen essenziell.

Die Intention der Autoren ist es, 5 Leitfragen bereitzustellen, mithilfe derer beide Seiten beurteilen können, ob GKV-Routinedaten geeignet sind, die Forschungsfrage zu beantworten und welche Schritte und Methoden hierfür notwendig sind. Die Leitfragen ermöglichen eine strukturierte Beschreibung des Forschungsvorhabens und unterstützen damit den Einstieg in die Kommunikation zwischen Versorger und Forscher. Sie sollen dazu beitragen, ein gemeinsames Verständnis des geplanten Versorgungsforschungsprojektes zwischen (klinisch) Forschenden und den datenhaltenden Institutionen zu schaffen. Auf Basis einer solchen, gemeinsamen Sichtweise lassen sich Anpassungen an der ursprünglichen Forschungsfrage diskutieren oder auch die Entscheidung

► **Tab. 1** Leitfragen, um die Anwendungsmöglichkeiten von GKV-Routinedaten für die Versorgungsforschung zu überprüfen.

<b>Leitfrage 1: Was ist die Forschungsfrage und mit welcher Methode soll die Frage beantwortet werden?</b>
<ul style="list-style-type: none"> <li>▪ Was ist das Erkenntnisinteresse?</li> <li>▪ Welche Versorgungssektoren sollen untersucht werden?</li> <li>▪ Ggf. was sind die relevanten Endpunkte?</li> <li>▪ Welche Methodik soll angewandt werden?</li> <li>▪ Welche Ergebnismengen sollen generiert werden?</li> </ul>
<b>Leitfrage 2: Was ist die Zielpopulation?</b>
<ul style="list-style-type: none"> <li>▪ Mithilfe welcher Merkmale/Klassifikationsschlüssel lässt sich die Zielpopulation in den Daten aufgreifen?</li> <li>▪ Sind die erforderlichen Merkmale zum Aufgriff der Zielpopulation valide?</li> <li>▪ Soll die Zielpopulation aktiv oder passiv rekrutiert werden?</li> <li>▪ Wird eine Vergleichsgruppe zur Beantwortung der Fragestellung benötigt?</li> </ul>
<b>Leitfrage 3: Welche Zeiträume sind zur Beantwortung für die Forschungsfrage relevant? Retrospektiv versus prospektiv?</b>
<ul style="list-style-type: none"> <li>▪ Stimmen die Zeiträume mit der Dauer des Vorhaltezeitraums der entsprechenden Kasse überein?</li> <li>▪ Soll eine einmalige oder mehrmalige Datenlieferung erfolgen?</li> <li>▪ Ist die Auswertung im gegebenen Zeitrahmen umsetzbar (z. B. unter Berücksichtigung des Zeitverzugs der Lieferung der vertragsärztlichen Daten)?</li> </ul>
<b>Leitfrage 4: Welche Informationen aus GKV-Routinedaten sind für die Fragestellung relevant?</b>
<ul style="list-style-type: none"> <li>▪ Welche Variablen aus welchen Sektoren werden für die Beantwortung der Fragestellung benötigt? Sind diese vorhanden?</li> <li>▪ Kann die Wahl der Variablen begründet werden?</li> <li>▪ In welchem Detailgrad werden die Informationen benötigt (z. B. tagesgenaues Sterbedatum etc.)?</li> <li>▪ Welcher Datenhalter könnte die passenden Daten bereitstellen?</li> </ul>
<b>Leitfrage 5: Wie könnten die Ergebnisse des Forschungsprojektes die Versorgung verbessern?</b>
<ul style="list-style-type: none"> <li>▪ Wie könnten sich die Versorgungsqualität und/oder die Versorgungseffizienz durch das Forschungsvorhaben verbessern?</li> <li>▪ Welche Defizite in der Versorgung gesetzlich Versicherter sollen adressiert werden?</li> <li>▪ Wie können die Ergebnisse in die Versorgung übertragen werden?</li> </ul>

über eine Projektbeteiligung treffen. Dabei gilt es zu betonen, dass selbst nach einer Klärung der in den Leitfragen aufgeworfenen Aspekte weitere Punkte gemeinsam abzustimmen sind. Dazu gehören unter anderem Datenschutzaspekte und dabei im Besonderen notwendige aufsichtsrechtliche Genehmigungen, die vertragliche Vereinbarung der Kooperation oder auch die Klärung der Zuständigkeiten für die Projektaufgaben sowie zeitliche Aspekte in der Zusammenarbeit. Hinzu kommen noch technische Aspekte wie die Art und Weise der Bereitstellung der Daten. Diese Punkte sind jedoch explizit nicht Bestandteil der hier vertieften Leitfragen.

Die Zusammenstellung der 5 Leitfragen basiert auf den Erfahrungen der Autor:innen. Eingeflossen sind Erfahrungen aus zahlreichen Forschungsprojekten, bspw. Innovationsfondsprojekten des Gemeinsamen Bundesausschusses (G-BA), in denen die Autor:innen verschiedene Rollen innehatten. Damit decken die Leitfragen die Sicht einer datenhaltenden Institution, die datenschutzrechtliche Sicht aus Perspektive einer Vertrauens- und Datenaufbereitungsstelle und die analytische Sicht bei Fragen der Versorgungsforschung und Evaluation ab.

## 5 Leitfragen, um die Anwendungsmöglichkeit von GKV-Routinedaten zu überprüfen

Im Folgenden werden 5 Leitfragen erläutert, die den Dialog zwischen Forschenden und datenhaltenden Institutionen strukturieren können, um zu prüfen, ob die Analyse von GKV-Routinedaten sinnvoll für das Forschungsvorhaben ist. Die Leitfragen reichen von der Formulierung der Fragestellung über die Definition der Zielpopulation, die inhaltliche Abgrenzung der verfügbaren Datentabellen und Informationen, die zur Beantwortung der Fragestellung er-

forderlich sind bis hin zur Verwertung der Ergebnisse. Eine Übersicht über die Leitfragen ist in ► **Tab. 1** dargestellt.

### Leitfrage 1: Was ist die Forschungsfrage und mit welcher Methode soll die Frage beantwortet werden?

Die Formulierung der Forschungsfrage ist der zentrale und zumeist erste Schritt im Laufe eines Forschungsvorhabens. Aus der Forschungsfrage lassen sich das Erkenntnisinteresse, das Forschungsziel und der Forschungsgegenstand sowie die Zielpopulation ableiten. Damit liefert die Forschungsfrage die Grundlage für Entscheidungen zum Forschungsdesign und zur erforderlichen Datengrundlage.

Damit Forschende und/oder Datenhalter beurteilen können, ob sich die Fragestellung mithilfe von GKV-Routinedaten beantworten lässt, sollte diese den Forschungsgegenstand bzw. das Forschungsthema möglichst eng eingrenzen und präzise definieren. Aus der Fragestellung sollten folgende Informationen hervor gehen: i) das Erkenntnisinteresse, ii) die Zielpopulation und ggf. die Vergleichspopulation, iii) die zu untersuchenden Versorgungssektoren, iv) der zu betrachtende Zeitraum v) ggf. die relevanten Endpunkte. So eignet sich zum Beispiel die Formulierung der Fragestellung „Wie ist die Versorgungssituation von Diabetespatient:innen in Deutschland?“ nur begrenzt, um die Anwendungsmöglichkeiten von GKV-Routinedaten zu überprüfen. Hierzu ist es notwendig, das Konstrukt „Versorgungssituation“ zu operationalisieren und einen Betrachtungszeitraum festzulegen. Besser wäre daher zum Beispiel die Fragestellung „Wie häufig wurden Diabetespatient:innen in den vergangenen 5 Jahren im Krankenhaus behandelt?“ (ein Überblick zu Fragestellungen findet sich in [16, 17]). Idealerweise formuliert der Forschende bereits die entsprechenden Hypothesen zur Beant-

wortung der Fragestellung. Neben der Fragestellung kann auch das geplante methodische Vorgehen im Sinne einer Kurzfassung eines Studiendesigns hilfreich sein, um die Erforderlichkeit der Daten beurteilen zu können. Hierzu zählt auch, ob und in welchem Umfang externe Daten zu den GKV-Routinedaten verknüpft werden sollen – ein sogenanntes Datenlinkage. Dabei muss dargelegt werden, ob das Datenlinkage auf Aggregatebene – also beispielsweise eine Verknüpfung des Wohnortes mit dem Deprivationsindex [18] – oder auf Personenebene stattfindet. Insbesondere die Verlinkung mit Personenbezug erfordert besondere datenschutzrechtliche Beurteilungen und Vorkehrungen ggf. mit Implementierung einer Vertrauensstelle. Ebenso müssen die zusätzlichen Aufwände bei den datenhaltenden Institutionen für eine Datenbereitstellung mit Datenlinkage kalkuliert werden (siehe hierzu auch [19]).

Ein frühzeitiger Austausch über geplante Auswertungsergebnisse in Form von Tabellen und Abbildungen erleichtert das Verständnis über die Forschungsinhalte. Hierbei ist auch der gewünschte und benötigte Aggregationsgrad von Variablen erkennbar und damit der Detailgrad und der Umfang der Ergebnistabellen. Eine Ergebnistabellenübersicht visualisiert und konkretisiert das intendierte Forschungsziel auf Datenebene. Anhand des Detailgrads kann eingeschätzt werden, welche Maßnahmen ergriffen werden müssen, um den Identitätsschutz des Betroffenen (beispielsweise des Versicherten) zu wahren. Hierbei ist darauf hinzuweisen, dass im Laufe des Projekts der Detailgrad auch dynamisch an die gefundenen Mengenverhältnisse angepasst werden kann (im Sinne eines höheren Detailgrads bei großen Fallzahlen) oder muss (im Sinne eines geringeren Detailgrads bei kleineren Fallzahlen), was unter Umständen Rückkopplungen mit der Thematik des Datenschutzes auslösen kann.

## Leitfrage 2: Was ist die Zielpopulation und wie lässt sich diese aufgreifen?

Steht die Fragestellung fest, so sollte sich aus ihr auch die Zielpopulation des Forschungsvorhabens herleiten (zum Beispiel Patienten mit Diabetes Typ II). Relevant für GKV-Routinedatenanalysen ist die Frage, ob und wie die Zielpopulation in den Datenquellen basierend auf Ein- und Ausschlusskriterien definiert werden kann. Mögliche Aufgreifkriterien könnten das Vorliegen ausgewählter Diagnosecodes, die Abrechnung einer bestimmten Prozedur oder Medikation oder ein Aufenthalt im stationären Sektor sein. Um die Aufgreifkriterien näher zu spezifizieren, können die in den Abrechnungsdaten verwendeten Klassifikationsschlüssel herangezogen werden, z. B. die Gebührenordnungsposition (GOP), der Diagnosecode gemäß der Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10-GM), Codes der Anatomisch-Therapeutisch-Chemischen Klassifikation (ATC) oder der Operationen- und Prozedurenschlüssel (OPS).

Herausforderungen im Aufgriff der Zielpopulation bestehen in der Validität der von den Leistungserbringern dokumentierten Codes. Daher sind interne und gegebenenfalls externe Validierungsschritte notwendig [20]. Neben diesen Aufgreifmöglichkeiten auf Basis vorhandener Informationen in einer Datenquelle kann die Studienpopulation auch aktiv rekrutiert werden. Dabei werden die Teilnehmer durch direkte Ansprache vor Ort (z. B. in Arztpraxen, Krankenhäusern oder Pflegeheimen) rekrutiert. Im Rahmen

dieser Rekrutierung stimmt der Proband/die Probandin per informierter Einwilligung der Nutzung seiner GKV-Routinedaten zu.

Eine weitere Frage, die sich der Forschende bzgl. der Zielpopulation stellen sollte, ist, ob die gesamte Population mit dem interessierenden Merkmal aufgegriffen und untersucht werden soll oder ggf. eine Zufallsauswahl ausreichend ist. Mit dieser Frage sind sowohl datenschutzrechtliche Aspekte wie einer möglichen Re-Identifikation von Studienprobanden als auch Aspekte der Repräsentativität der Studienpopulation verknüpft.

Neben der Definition der eigentlichen Zielpopulation sollte der Forschende prüfen, ob sich die Fragestellung auf einen Vergleich zwischen Ziel- und ggf. Gesamt- oder einer anderen Population bezieht. So werden in der Versorgungsforschung häufig verschiedene Populationen gegenübergestellt, um Aussagen über Unterschiede in der Versorgungssituation ableiten zu können. Dies könnten zum Beispiel die Pro-Kopf-Kosten für Versicherte mit Typ 2 Diabetes im Vergleich zu Versicherten ohne Diabetes sein (vgl. [21]). Bei Evaluationsstudien wird häufig – beispielsweise im Rahmen von geförderten Projekten des Innovationsfonds – eine Vergleichsgruppe herangezogen, um den kausalen Effekt einer Intervention auf die relevanten Endpunkte schätzen zu können [5, 16]. Dabei besteht die Möglichkeit, die Vergleichsgruppe aktiv zu rekrutieren oder allein auf Basis bestimmter Merkmale aus GKV-Routinedaten aufzugreifen. Zudem sollte sich der Forschende Gedanken über die erforderliche Größe der Vergleichsgruppe im Sinne einer Fallzahlenberechnung machen.

## Leitfrage 3: Welche Zeiträume sind für die Forschungsfrage relevant? Retrospektiv versus prospektiv?

Ein weiterer relevanter Aspekt im Rahmen von Datenanfragen und deren Bereitstellung sind die für das geplante Vorhaben relevanten Zeiträume. Soll die Analyse retrospektiv ausgestaltet werden, d. h. auf bereits vorhandenen Daten aufbauen? Oder sollen für ein vorab zu bestimmendes Kollektiv die Abrechnungsdaten für einen bestimmten Zeitraum in der Zukunft zur Verfügung gestellt werden (prospektiv)? Beide Ansätze können in einem Projekt auch kombiniert werden. Dabei sind die bei den datenhaltenden Institutionen verfügbaren Beobachtungszeiträume zu beachten, da die den Kassen vorliegenden Daten in der Regel nur für eine bestimmte Zeit und für bestimmte Nutzungszwecke aufbewahrt werden. Diese Zeiträume können sich nach betrachtetem Leistungssektor unterscheiden. Daten aus der vertragsärztlichen Versorgung stehen den gesetzlichen Kassen in aller Regel frühestens sechs Monate nach Ende eines Abrechnungsquartals zur Verfügung, im stationären Sektor oftmals kurze Zeit nach Ende des Abrechnungsfalls.

Die benötigten Zeiträume werden durch die Fragestellung selbst und das methodische Vorgehen determiniert. So werden für die Bestimmung inzidenter Personen längere – teilweise mehrere Jahre – erkrankungsfreie Vorlaufzeiten benötigt [22]. Aus der geschätzten Anzahl an Zielpersonen oder Ereignissen ergibt sich auch die Notwendigkeit, den Aufgreifzeitraum auszudehnen oder mehrere Inzidenzjahre zu poolen, um die erforderliche Fallzahl zu erreichen. Und zuletzt sind je nach Fragestellung auch längere (Nach-)Beobachtungszeiträume erforderlich.

#### **Leitfrage 4: Welche Informationen aus GKV-Routinedaten sind für die Fragestellung relevant?**

Um auf eine datenhaltende Institution mit dem Wunsch nach Nutzung der dort vorliegenden Informationen zuzugehen, sollte der Forscher konkretisieren, welche Versorgungssektoren und welche der dort hinterlegten Informationen, Datenfelder bzw. Variablen ausgewertet werden sollen. In einem ersten Schritt sollten die für die Forschungsfrage relevanten Leistungssektoren identifiziert werden, bspw. ob Informationen aus dem ambulanten und/oder dem stationären Bereich benötigt werden. Im zweiten Schritt sollten für jeden betrachteten Leistungssektor die erforderlichen Informationen bzw. Variablen festgelegt werden (wie etwa Diagnosecodes, ATC-Codes bei Arzneimittelverordnungen oder Codes zur Abrechnung von Leistungen wie z. B. OPS-Codes).

Bei der Festlegung der relevanten Informationen gilt es zwei Aspekte zu berücksichtigen: Erstens bilden GKV-Abrechnungsdaten das komplette, zulasten der GKV abgerechnete Leistungsgeschehen ab. Somit sind z. B. Privatverordnungen oder Selbstmedikation (so genannte Over-the-Counter (OTC)-Präparate) nicht aus den Daten ersichtlich. Zweitens erfolgte die originäre Erhebung der Daten nicht zu wissenschaftlichen, sondern primär zu Abrechnungszwecken. Daher bilden die GKV-Routinedaten nur einen bestimmten Ausschnitt der Realität exakt ab. So können etwa Änderungen in der Spezifikation des Morbiditätsorientierten Risikostrukturausgleichs (Morbi-RSA) dazu führen, dass sich die Morbidität der Versicherten durch sog. Upcoding erhöht [23] und damit verzerrt. Aber auch die Abrechnungs- bzw. Erlösrelevanz (z. B. im Rahmen der Vergütung nach Diagnosis Related Groups (DRG)) kann das Koerverhalten der Leistungserbringer beeinflussen.

Differenziert betrachtet bedeutet dies, dass (i) Merkmale, welche die gewünschten Aspekte abdecken, potenziell vorhanden sein müssen und (ii) das entsprechende Merkmal den jeweiligen Aspekt auch zuverlässig abbildet. So ist beispielsweise im Aufnahmegrund der Daten aus der stationären Versorgung (§301 SGB V) zwar das Feld „Notfall“ hinterlegt. Die Praxis zeigt jedoch, dass dieses oftmals nicht valide auswertbar ist, da nicht alle Notfallbehandlungen auch dieses Kennzeichen tragen. Somit müssten für eine (indirekte) Operationalisierung der Notfallbehandlung im Krankenhaus weitere Faktoren herangezogen werden [24].

Vor dem Hintergrund der Datensparsamkeit müssen im Rahmen einer Datenbereitstellung die beantragten Variablen begründet werden. Die Begründung sollte nachvollziehbar sein und den Detaillierungsgrad der Variablen berücksichtigen. So ist es oftmals für Analysen ausreichend, wenn das Behandlungsdatum nicht tagesgenau, sondern beispielsweise auf Wochenebene dokumentiert ist. Andererseits ist zur Abbildung von Versorgungssequenzen ein vergrößertes Behandlungsdatum nicht ausreichend, da so unter Umständen die genaue Reihenfolge der Inanspruchnahme nicht ermittelt werden kann. Dabei ist jedoch zu beachten, dass mit zunehmenden Detailgrad der Variablen auch die Möglichkeit einer Re-Identifikation von Betroffenen tendenziell ansteigt.

Als weiteren zu beachtenden Punkt ist die Versichertenstruktur zu nennen, die zwischen den datenhaltenden Institutionen abweichen kann, abhängig unter anderem von regionalen Aspekten oder auch historisch bedingt. Dies kann die Repräsentativität des Analysekollektivs für die Wohnbevölkerung Deutschlands einschränken. Forschende sollten daher folgende Anforderungen an die

Daten klären: Inwiefern muss das Studienkollektiv repräsentativ für die Wohnbevölkerung Deutschlands sein? Sind angefragte Charakteristika bei den datenhaltenden Institutionen über- oder unterrepräsentiert? Sollen die Daten von selektierten Versicherten lediglich auf Bundesebene oder auch auf der Ebene der Bundesländer oder der Landkreise analysiert werden? Wenn ja, lassen die Zahlen an betroffenen Versicherten kleinräumige Analysen zu?

#### **Leitfrage 5: Wie könnten die Ergebnisse des Forschungsprojektes die Versorgung verbessern?**

Neben den eher methodischen Leitfragen 1 bis 4 thematisiert die letzte Leitfrage die Relevanz und das Potential des geplanten Forschungsvorhabens im Hinblick auf die Verbesserung der Versorgung als langfristiges Ziel der Versorgungsforschung. Die Auseinandersetzung mit der Frage der Versorgungsrelevanz des Forschungsvorhabens kann aber auch die Entscheidungsfindung bzgl. der Anwendbarkeit von GKV-Routinedaten unterstützen: Erstens, geht mit der Beurteilung der Versorgungsrelevanz einher, die Größe der Zielpopulation abzuschätzen. Dabei geht es auch um die Frage, ob sich bei einer datenhaltenden Institution ausreichend Fälle für eine Analyse der Forschungsfrage abbilden lassen (siehe hierzu Leitfrage 2). Zweitens kann die Darstellung der Versorgungsrelevanz den Forschenden dabei unterstützen, die datenhaltende Institution für eine Kooperation für das entsprechende Vorhaben zu gewinnen. Dabei gilt es zu berücksichtigen, welches Interesse ein Datenhalter an einer Umsetzung des Forschungsvorhabens haben könnte und wie dessen Funktionen im Rahmen einer solchen Umsetzung aussehen könnten.

Um die Versorgungsrelevanz des geplanten Vorhabens zu eruieren, kann sich der/die Forschende die folgenden weiterführenden Fragen stellen: Inwieweit können die erwarteten Ergebnisse aus dem angestrebten Projekt dazu beitragen, Defizite im Rahmen der Versorgung in der GKV zu identifizieren und zu beheben? Wie können die Ergebnisse die Versorgungsqualität oder die Versorgungseffizienz verbessern? Wie lassen sich die Ergebnisse in den Versorgungsalltag übertragen?

### **Zusammenfassung/Diskussion**

Forschende aus dem Bereich der Versorgungsforschung benötigen zur Analyse und Beantwortung ihrer Forschungsfragen Daten über die reale Versorgungssituation. Die Nutzung von GKV-Routinedaten in der Versorgungsforschung ist dabei, auch getrieben durch Projekte des Innovationsfonds, in den letzten Jahren gewachsen. Umso mehr erscheint es notwendig, Forschenden mit Hilfe der formulierten Leitfragen eine Hilfestellung für Anfragen an datenhaltende Institutionen zu geben. Hiervon profitieren auch letztere, da hierdurch ein gemeinsames Verständnis über die zu beantwortenden Fragestellungen erarbeitet werden kann.

Die vorgestellten Leitfragen stellen dabei nur einen möglichen Orientierungsrahmen für Forschende dar. Somit ist es nicht der Anspruch der Leitfragen, einen minimalen Anforderungskatalog zu liefern. Vielmehr sollten diese genutzt und auch kritisch hinsichtlich der Anwendbarkeit hinterfragt werden. Schließlich sind die aufgeführten Leitfragen nicht als statisches Konstrukt zu verstehen. Gesetzliche Änderungen mit Versorgungsbezug können Änderungen

in den Leitfragen ebenso nach sich ziehen wie datenschutzrechtliche Beurteilungen.

Viele Aspekte, die im Rahmen einer Kooperation von Forschenden mit datenhaltenden Institutionen bei wissenschaftlichen Projekten notwendig sind, wurden an dieser Stelle nur kurz skizziert oder sogar ausgelassen, da sie für eine Erstanfrage bei einer datenhaltenden Institution in aller Regel nur am Rande von Bedeutung sein dürften. Dazu zählt beispielsweise das umfangreiche Thema Datenschutz. Hierunter fallen Aspekte wie die Genehmigung zur Übermittlung von Sozialdaten durch eine Aufsichtsbehörde, die datenschutzkonforme Verarbeitung und Rechtmäßigkeit der Datennutzung ebenso wie Datenschutzaspekte bei der eigentlichen Übermittlung von Daten. Weiterhin sind rechtliche Aspekte zu nennen, wie Verträge (Kooperationsvereinbarungen oder auch Datenlieferungsvereinbarungen) oder die Regelung des Datenzugangs. Hier sind neben den oftmals differenzierten Vorgaben datenhaltender Institutionen auch die Vorschläge der Guten Praxis Sekundärdatenanalyse [13] einzubeziehen.

Selbst nach Vorliegen eines gemeinsamen Verständnisses zwischen Versorgenden und Forschenden und einer Kooperations- oder Unterstützungszusage von Seiten der datenhaltenden Institution(en) sind weitere Rahmenbedingungen der Projektdurchführung zu klären. Dazu zählen die Verantwortung des Projektmanagements, mögliche Finanzierungsquellen oder die Abstimmung und Durchführung der Arbeitsschritte. Bisherige Projekte liefern hierfür Orientierungspunkte.

Abschließend gilt es zu betonen, dass datenbasierte Ansätze in der Versorgungsforschung nicht nur ein gemeinsames Verständnis der Forschenden und der an der Bereitstellung von Daten Beteiligten benötigt. Auch die Bereitschaft, auf Basis der Ergebnisse wissenschaftlicher Projekte in einen kritischen Diskurs mit den beteiligten Leistungserbringern sowie der Politik [25] zu treten, ist wichtig, damit praxisrelevante Versorgungsforschung ihr Ziel einer Verbesserung der Versorgung von Patient:innen erreichen kann.

Die hier formulierten Leitfragen wurden vor allem im Hinblick auf Krankenkassen als datenhaltende Institutionen entwickelt. Sie gelten aber auch für die Kommunikation mit anderen Datenhaltern (beispielsweise dem Zentralinstitut für die kassenärztliche Versorgung). Zukünftig werden GKV-Routinedaten für Nutzungsberechtigte durch das Forschungsdatenzentrum Gesundheit (FDZ Gesundheit) bereitgestellt (SGB V §303 a-f). Auch hier sind die Leitfragen geeignet, den Abstimmungsprozess über Art und Umfang der bereitzustellenden Daten zu strukturieren. Bestehen bleibt die übergeordnete Botschaft, dass Projekte der Versorgungsforschung immer die gemeinsame Beteiligung von klinischer Expertise zusammen mit methodenerprobten Versorgungsforschern unter Kenntnis der genutzten Datenbasis erfordern.

## Danksagung

Die Autor:innen bedanken sich bei Dr. Ingrid Schubert für ihr scharfes Auge bei der Durchsicht des Manuskriptes und ihre wertvollen Vorschläge für prägnante Formulierungen und passenden Referenzen. Der Dank geht auch an die Gutachter:innen, die auf Basis der ersten Fassung Vorschläge für die Umstrukturierung des Manuskriptes und Schärfung der inhaltlichen Aspekte gegeben haben.

## Interessenkonflikt

Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

## Literatur

- [1] Pfaff H, Schrappe M. Einführung in die Versorgungsforschung. In: H. Pfaff, G. Glaeske, E. Neugebauer, M. Schrappe Hrsg. Lehrbuch Versorgungsforschung. Stuttgart: Schattauer Verlag; 2011: S 2–41
- [2] Klinkhammer-Schalke M, Kaiser T, Apfelbacher C et al. Manual für Methoden und Nutzung versorgungsnaher Daten zur Wissensgenerierung. *Gesundheitswesen* 2020; 82: 716–722. DOI: 10.1055/a-1237-4011
- [3] Maneck M, Köckerling F, Fahlenbrach C et al. Hospital volume and outcome in inguinal hernia repair: analysis of routine data of 133,449 patients. *Hernia* 2020; 24: 747–757. DOI: 10.1007/s10029-019-02091-8
- [4] Frank-Tewaag J, Bleek J, Horenkamp-Sonntag D et al. Use of guideline-recommended drug therapy in patients undergoing percutaneous coronary intervention for stable coronary heart disease in Germany: a multilevel analysis of nationwide routine data. *BMJ Open* 2020; 10: 1–9. DOI: 10.1136/bmjopen-2020-042886
- [5] Schubert I, Stelzer D, Siegel A et al. Ten-year evaluation of the population-based integrated health care system „Gesundes Kinzigtal“. *Dtsch Arztebl Int* 2021; 118: 465–472. DOI: 10.3238/arztebl.m2021.0163
- [6] Vogt V, Scholz SM, Sundmacher L. Applying sequence clustering techniques to explore practice-based ambulatory care pathways in insurance claims data. *Eur J Public Health* 2018; 28: 214–219
- [7] Blaschke K, Hellmich M, Samel C et al. Association between Periodontal Treatment and Healthcare Costs in Patients with Coronary Heart Disease: A Cohort Study Based on German Claims Data. *Dent J* 2022; 10: 133. DOI: 10.1016/j.diabres.2020.108641
- [8] Köster I, Schubert I, Huppertz E. Fortschreibung der KoDiM-Studie: Kosten des Diabetes mellitus 2000–2009. *Dtsch Med Wochenschr* 2012; 137: 1013–1016. DOI: 10.1055/S-0032-1304891
- [9] Hoffmann F, Icks A. Unterschiede in der Versichertenstruktur von Krankenkassen und deren Auswirkungen für die Versorgungsforschung: Ergebnisse des Bertelsmann-Gesundheitsmonitors. *Gesundheitswesen* 2012; 74: 291–297. DOI: 10.1055/S-0031-1275711
- [10] Swart E, Ihle P, Gothe H et al. Hrsg. Routinedaten im Gesundheitswesen. 2. Aufl. Bern: Hans Huber; 2014
- [11] Neubauer S, Zeidler J, Lange A et al. Prozessorientierter Leitfaden für die Analyse und Nutzung von Routinedaten der Gesetzlichen Krankenversicherung. Baden-Baden: Nomos; 2017. DOI: 10.5771/9783845281193
- [12] Schubert I, Köster I, Küpper-Nybelen J et al. Versorgungsforschung mit GKV-Routinedaten. *Bundesgesundheitsbl* 2008; 51: 1095–1105. DOI: 10.1007/s00103-008-0644-0
- [13] Swart E, Gothe H, Geyer S et al. Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen. 3. Fassung; Version 2012/2014 Third Revision 2012/2014 *Gesundheitswesen* 2015; 77: 120–126. DOI: 10.1055/s-0034-1396815
- [14] March S, Andrich S, Dreppner J et al. Gute Praxis Datenlinkage (GPD) Good Practice Data Linkage. *Gesundheitswesen* 2019; 81: 636–650. DOI: 10.1055/a-0962-9933
- [15] Swart E, Bitzer EM, Gothe H et al. Standardisierte BerichtsROUTINE für Sekundärdaten Analysen (STROSA) – ein konsentierter Berichtsstandard für Deutschland, Version 2. *Gesundheitswesen* 2016; 78: e145–e160. DOI: 10.1055/s-0042-108647
- [16] Busse R. Methoden der Versorgungsforschung. In: Hey M, Maschewsky-Schneider U, Busse R, Häußler B, Pfaff H, Rosenbrock R, Schrappe M, Hrsg. Kursbuch Versorgungsforschung. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2006: 244–251

- [17] Schubert I. Arzneimittelepidemiologie und Arzneimittelverbrauchs-  
forschung auf der Basis von Krankenkassendaten: zentrale  
Untersuchungsfragen und methodische Hinweise. *Präv Gesundheitsf*  
2022; 10. DOI: 10.1007/s11553-022-00968-8
- [18] Maier W, Fairburn J, Mielck A. Regionale Deprivation und Mortalität in  
Bayern. Entwicklung eines Index Multipler Deprivation\* auf  
Gemeindeebene. *Gesundheitswesen* 2012; 74: 416–425. DOI:  
10.1055/s-0031-1280846
- [19] March S, Antoni M, Kieschke J et al. Quo vadis Datenlinkage in  
Deutschland? Eine erste Bestandsaufnahme. *Gesundheitswesen* 2018;  
80: 20–31. DOI: 10.1055/s-0043-125070
- [20] Schubert I, Köster I. Krankheitsereignis: Operationalisierung und  
Falldefinition. In: Swart E, Ihle P, Holger G, Matusiewicz D, Hrsg.  
Routinedaten im Gesundheitswesen, 2. Aufl. Bern: Hans-Huber; 2014
- [21] Jacobs E, Hoyer A, Brinks R et al. Healthcare costs of Type 2 diabetes in  
Germany. *Diabet. Med* 2017; 34: 855–861. DOI: 10.1111/dme.13336
- [22] Abbas S, Ihle P, Köster I et al. Estimation of Disease Incidence in Claims Data  
Dependent on the Length of Follow-Up: A Methodological Approach. *HSR*  
2012; 47: 746–755. DOI: 10.1111/j.1475-6773.2011.01325
- [23] Bundesministerium für Gesundheit und Soziale Sicherheit/Reschke P,  
Sehlen S, Schiffhorst G et al. Klassifikationsmodelle für Versicherte im  
Risikostrukturausgleich – Endbericht (2004). Im Internet: [www.  
bundesgesundheitsministerium.de/fileadmin/Dateien/5\\_  
Publikationen/Ministerium/Klassifikationsmodelle\\_RSA.pdf](http://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Ministerium/Klassifikationsmodelle_RSA.pdf); Stand:  
13.03.2023
- [24] Greiner F, Slagman A, Stallmann C et al. Routinedaten aus  
Notaufnahmen: Unterschiedliche Dokumentationsanforderungen,  
Abrechnungsmodalitäten und Datenhalter bei identischem Ort der  
Leistungserbringung. *Gesundheitswesen* 2020; 82: 7282. DOI:  
10.1055/a-0996-8371
- [25] Häussler B, Hoffmann W, Stegmaier P. „Wir brauchen Experimentalräume  
für die Versorgungsforschung“. In: *Monitor Versorgungsforschung* (01/3)  
6–15. DOI: 10.24945/MVF.01.23.1866-0533.2470 (letzter Zugriff:  
28.02.2023)