Schätzung kleinräumiger Krankheitshäufigkeiten für die deutsche Bevölkerung anhand von Routinedaten am Beispiel von Typ-2-Diabetes

Jan Breitzkreuz · Gabriela Brückner · Jan Pablo Burgard ·
Joscha Krause · Ralf Münnich · Helmut Schröder · Katrin Schüssel

Eingegangen: 30. Januar 2019 / Angenommen: 15. März 2019 / Online publiziert: 3. April 2019
© Der/die Autor(en) 2019

Zusammenfassung Die Erfassung regionaler Krankheitsverteilungen ist ein zentrales Element in der Planung und Gestaltung adäquater Gesundheitsversorgung. In der öffentlichen Gesundheitsberichtserstattung liegen Schätzungen kleinräumiger Krankheitshäufigkeiten in der Regel nicht vor. Aufgrund begrenzter Ressourcen und zu wenigen ortsspezifischen Beobachtungen ist es meist nicht möglich regionale Krankheitsverteilungen mit hinreichender Genauigkeit zu quantifizieren. Hier repräsentieren Krankenkassendaten eine wichtige Alternative. Der Versichertenstamm einer Krankenkasse ist meist viel größer als der Stichprobenumfang einer Gesundheitserhebung. Zusätzlich handelt es sich bei den Daten einer Krankenkasse um Abrechnungsdaten, welche automatisch erfasst und nicht aufwändig erhoben werden müssen. Doch wenn regionale Krankheitshäufigkeiten anhand von Krankenkassendaten geschätzt werden, muss beachtet werden, dass der Versichertenstamm einer Krankenkasse keine zufällige Teilpopulation der Gesamtbevölkerung darstellt. Aufgrund verschiedener Eigenschaften des deutschen Gesundheitssystems ist die Kassenzugehörigkeit einer Person informativ für ihre Morbidität hinsichtlich verschiedener Krankheiten. Folglich muss eine Bias-Korrektur bei der Schätzung erfolgen, um valide Ergebnisse zu erhalten. Im Zuge des AOK-internen Projekts *Gesundheitsatlas* des Wissenschaftlichen Instituts der AOK (WiDO) wurde zusammen mit der Universität Trier eine Methodik entwickelt, welche diese Problematik behandelt. Sie ermöglicht die Quantifizierung kleinräumiger Krankheitshäufigkeiten anhand von anonymisierten Krankenkassendaten mit einer Multi-Source-Schätzung. Durch multivariate Analyseverfahren sowie modellbasierte Inferenz werden die anonymisierten AOK-Routinedaten und stationäre Behandlungshäufigkeiten al-

J. Breitzkreuz · G. Brückner · H. Schröder · K. Schüssel
Wissenschaftliches Institut der AOK, Rosenthaler Straße 31, 10178 Berlin, Deutschland

J. P. Burgard · J. Krause · R. Münnich (✉)
Wirtschafts- und Sozialstatistik, Universität Trier, Universitätsring 15, 54296 Trier, Deutschland
E-Mail: muennich@uni-trier.de

ler Krankenhausfälle zu einer Bias-korrigierenden Schätzmethodik kombiniert. In dem vorliegenden Beitrag wird diese Methodik beschrieben und ihre Effektivität am Beispiel von Diabetes Mellitus Typ 2 demonstriert. Ausgehend vom Versichertenstamm der AOK – *die Gesundheitskasse* sowie deren Leistungsdaten werden alters- und geschlechtsreferenzierte Krankheitshäufigkeiten auf Kreisebene für die gesamte deutsche Bevölkerung geschätzt. Es kann gezeigt werden, dass der Ansatz plausible Ergebnisse auf regionalen Ebenen liefert und somit die Darstellung der gesundheitlichen Lage in bisher nicht erreichter Detailtiefe ermöglicht.

Schlüsselwörter Diabetes Mellitus · Multi-Source-Schätzung · Prävalenz · Routinedaten · Selektivitätsverzerrung

JEL-Classification C31 · C51 · C52 · I13 · I18

Estimation of regional diabetes type 2 prevalence in the German population using routine data

Abstract The analysis of regional disease distributions is essential for planning and establishing comprehensive health care programs. Unfortunately, epidemiologic data provided by public health reporting often lack in geographic detail. Due to limited resources, corresponding data contain only few observations per region. Hence, it is usually not possible to quantify regional prevalence with sufficient accuracy. Health insurance records represent an important alternative in this area. The insuree population of a health insurance fund is usually much larger than the sample of a health survey. Also, the data of a health insurance company are billing data which are automatically recorded and do not have to be collected at great expense. Further, corresponding records are typically routine data that do not have to be collected in expensive surveys. However, the estimation of regional prevalence from health insurance records encloses a fundamental problem: the member population of a health insurance fund is no random subset of the total population. Due to specific characteristics of the German health care system, a person's health insurance membership can be informative for his or her morbidity with regard to various diseases. Accordingly, regional prevalence estimation from health insurance records must include a bias correction in order to obtain valid results. This methodological issue is solved in the course of the AOK internal research project *Gesundheitsatlas* that is conducted by the AOK Research Institute (WiDO) and the University of Trier. A statistical method was developed which allows the quantification of small-scale prevalence on the basis of anonymised health insurance data via multi-source estimation. Using multivariate analysis and model-based inference allows to combine AOK routine data as well as inpatient diagnosis frequencies for a bias-correcting estimation procedure. The presented paper describes this procedure and demonstrates its effectiveness using the example of diabetes mellitus type 2. The insuree population of the health insurance fund AOK—*die Gesundheitskasse* is used as basis to estimate the age and gender specific disease distribution of the entire German population at the district level. It can be shown that the approach produces plausible results at regional levels

and allows for health measurement with a previously unattained level of geographic detail.

Keywords Diabetes mellitus · Multi-source estimation · Prevalence · Routine data · Selection bias

1 Einleitung

Die Darstellung der gesundheitlichen Lage der Bevölkerung ist ein wichtiges Fundament der Gesundheitspolitik und Versorgungsforschung. Insbesondere die Erfassung kleinräumiger Gesundheitsdisparitäten ist von großer Bedeutung, etwa für die ärztliche Bedarfsplanung oder für Präventionsangebote von Krankenkassen. In Deutschland werden entsprechende Zahlen zur gesundheitlichen Lage der Bevölkerung selten in Registern erfasst und müssen deshalb geschätzt werden. Dies erfolgt in der Regel auf Basis nationaler Gesundheitserhebungen, wie etwa *Gesundheit in Deutschland aktuell* (GEDA) (Robert Koch-Institut 2014) oder der *Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland* (KiGGS) (Robert Koch-Institut 2017a). Doch aufgrund begrenzter Ressourcen und des enormen Befragungsaufwandes können im Rahmen solcher Erhebungen lediglich Stichproben gezogen werden, welche regionsspezifisch gegebenenfalls nur wenige Beobachtungen enthalten. Wird auf dieser Basis eine kleinräumige Schätzung der Krankheitshäufigkeit (Prävalenz) durchgeführt, beispielsweise auf Ebene der Landkreise, so sind die Ergebnisse mit einer inakzeptabel hohen Unsicherheit verbunden, da die entsprechenden Standardfehler zu groß sind. Folglich können lediglich Schätzwerte für größere Aggregate, etwa für Bundesländer, verlässlich ausgewiesen werden. Etwaige systematische Gesundheitsdisparitäten in den Aggregaten bleiben dann unberücksichtigt.

An dieser Stelle repräsentieren Krankenkassendaten eine wichtige Alternative. Der Versichertenstamm der meisten (gesetzlichen) Krankenkassen ist um ein Vielfaches größer als die Anzahl der Personen, welche im Rahmen einer Gesundheitserhebung befragt werden. Dies stellt einen Vorteil in der Beobachtungszahl dar, welcher insbesondere bei der Gesundheitsmessung auf regionalen Ebenen hilfreich ist. Darüber hinaus handelt es sich bei Krankenkassendaten um Routinedaten. Sie werden in versicherungsrelevanten Fällen automatisch erfasst und müssen nicht in zeit- bzw. kostenintensiven Interviews erhoben werden. Zusätzlich kann die Angabe entsprechender Daten nicht wie in Gesundheitserhebungen verweigert werden, da die vollständige Datenerfassung zu Abrechnungszwecken notwendig ist.

Bei der Schätzung kleinräumiger Krankheitshäufigkeiten anhand von Krankenkassendaten muss jedoch beachtet werden, dass der Versichertenstamm einer Krankenkasse nicht das Resultat einer zufallsbasierten Stichprobenziehung darstellt. Er ist somit keine zufällige Teilpopulation der Gesamtbevölkerung. Aufgrund verschiedener Eigenschaften des deutschen Gesundheitssystems, welche in Abschn. 2 genauer

erläutert werden, kann die Kassenzugehörigkeit einer Person informativ¹ für ihre Morbidität hinsichtlich verschiedener Krankheiten sein. Folglich liefert die Verallgemeinerung der Morbiditätsstruktur einer beliebigen Krankenkasse gegebenenfalls verzerrte Ergebnisse für die Gesamtbevölkerung. Da diese Problematik seit einiger Zeit bekannt ist, werden in der Praxis kassenspezifische Krankheitshäufigkeiten oft auf Alter und Geschlecht bedingt (Standardisierung). Dieser Ansatz ist häufig jedoch nicht ausreichend, und die systematischen Gesundheitsdisparitäten zwischen verschiedenen Versichertenstämmen bleiben auch nach der Standardisierung bestehen. Dies wurde mehrfach in empirischen Studien nachgewiesen, etwa von Hoffmann und Icks (2011, 2012) oder Hoffmann und Koller (2017).

Um regionale Krankheitshäufigkeiten der Gesamtbevölkerung auf Basis der Routinedaten einer Krankenkasse zu schätzen, muss aufgrund der informativen Kassenzugehörigkeit eine Bias-Korrektur vorgenommen werden. Im Rahmen des Projekts *Gesundheitsatlas*, welches vom Wissenschaftlichen Institut der AOK (WIdO) mit dem Kooperationspartner der Universität Trier durchgeführt wird, konnte eine Schätzmethodik entwickelt werden, welche diese Problematik auf Basis einer Multi-Source-Schätzung behandelt. Unter Verwendung multivariater Analyseverfahren (Feinstein 1996) und modellbasierter Inferenz werden Routinedaten der AOK und stationäre Behandlungshäufigkeiten aus der Fallpauschalenbezogenen Krankenhausstatistik (DRG-Statistik nach § 21 KHEntgG) kombiniert, um den Bias zu korrigieren und stabile Schätzungen von alters- und geschlechtsreferenzierten Krankheitshäufigkeiten auf Kreisebene durchzuführen. Hierbei wird innerhalb der AOK-Daten der bedingte Erwartungswert der kleinräumigen Krankheitshäufigkeiten gegeben regionaler Demographie sowie der stationären Behandlungshäufigkeiten verwandter Diagnosen anhand (gemischter) linearer Modelle (West et al. 2007) quantifiziert. Da regionale Unterschiede sowohl in der Krankheitshäufigkeit und als auch in der Versorgung a priori bekannt sind, erfolgt die dafür notwendige Schätzung der Modellparameter ebenfalls regionsspezifisch.

Der eben genannte bedingte Erwartungswert wird mit verschiedenen Modellspezifikationen geschätzt, da gängige Verfahren zur Bestimmung einer optimalen Spezifikation, wie etwa Analysis of Variance (ANOVA) (Gelman 2005) oder Informationskriterien (Burnham und Anderson 2004), aufgrund der fehlenden Stichprobenziehung nicht valide angewendet werden können. Stattdessen wird die Prädiktionskraft der einzelnen Modelle durch eine Bootstrap-basierte k -fold Cross Validation als Modifikation des Ansatzes von Efron und Tibshirani (1997) evaluiert, wobei instabile Modelle aussortiert werden. Anschließend werden pro Kreis für jede verbleibende Modellspezifikation alters- und geschlechtsreferenzierte Prädiktionen der Gesamtprävalenz erzeugt. Hierbei dienen die stationären Behandlungshäufigkeiten der Gesamtbevölkerung als Benchmarks, um den kassenspezifischen Bias zu korrigieren und kleinräumige Prädiktionen hinsichtlich der Krankheitshäufigkeiten zu erzeugen. Die hieraus erhaltenen Ergebnisse bilden dann eine Prädiktionsverteilung

¹ „Informative Kassenzugehörigkeit“ bedeutet hier, dass die Versichertenstämmen der Krankenkassen sich in der Verteilung gesundheitsrelevanter Merkmale systematisch unterscheiden. Die Konsequenz ist, dass ein Mitglied einer gegebenen Krankenkasse statistisch ein anderes Risiko für bestimmte Krankheiten aufweist, als das Mitglied einer anderen Krankenkasse.

über die regionale Gesamtprävalenz, von welcher dann ein plausibelster Schätzwert und ein plausibler Bereich abgeleitet werden.

Die Gliederung des Beitrags stellt sich wie folgt dar. Zunächst werden in Abschn. 2 die Bedeutung kleinräumiger Messwerte zur Krankheitsprävalenz und die Erfassung regionaler Gesundheitsdisparitäten thematisiert. Dann wird in Abschn. 3 die Schätzmethodik, anhand welcher auf Basis der Routinedaten der AOK kleinräumige Krankheitshäufigkeiten geschätzt werden, ausführlich beschrieben. In Abschn. 4 wird die Methodik anschließend angewendet, um anhand des Versichertenstamms der AOK von 2016 alters- und geschlechtsreferenzierte Krankheitshäufigkeiten für Diabetes Mellitus Typ 2 (im Folgenden Typ-2-Diabetes) für die deutsche Bevölkerung zu schätzen. Abschn. 5 schließt mit einem Fazit und einem Ausblick.

2 Hintergrund

2.1 Regionale Variation in der Gesundheit

Regionale Disparitäten sind in wohl allen gesellschaftlichen Feldern und Raumeinheiten feststellbar (Frankenfeld 2005). Auch im Gesundheitssektor sind sie offenkundig: So unterscheiden sich Regionen hinsichtlich ihrer Ausstattung mit Krankenhäusern, in der (Fach-) Arztdichte und Erreichbarkeit wie auch in der Verfügbarkeit weiterer Leistungserbringer, zum Beispiel Physiotherapeuten oder Apotheken. Darüber hinaus sind regional vor allem auch Unterschiede bei der Häufigkeit von Erkrankungen sowie der Inanspruchnahme von Leistungen erkennbar (Günster et al. 2019; Hecken 2018; Klauber et al. 2016; Bertelsmann Stiftung 2015). Um diese Disparitäten genauer beleuchten und Ursachen erhöhter Erkrankungsraten untersuchen zu können, bedarf es zunächst adäquater Analyseeinheiten. So sind geeignete Raumeinheiten und Datengrundlagen zu wählen, die es ermöglichen Indikatoren zu bestimmen, auf deren Basis genauere Untersuchungen erfolgen können (Frankenfeld 2005).

Der größte analytische Mehrwert von Gesundheitsdaten ergibt sich somit bei kleinräumiger und zugleich möglichst umfassender, also bundesweiter Betrachtungsweise. Denn erst im Vergleich können überhaupt regionale Variationen auffallen, und erst im Vergleich kann ein bestimmter Status quo einer Region als positiv oder negativ bewertet werden. Wird zum Beispiel ein westfälischer Landkreis hinsichtlich des Gesundheitszustandes von Kindern untersucht, so ist eine Bewertung für diese Region nur bedingt möglich, selbst wenn umfassende Gesundheitsindikatoren vorliegen. Fehlen Vergleichsinformationen zu ähnlichen oder auch andersartigen Regionen, so fehlt eine geeignete Bewertungsskala zur Einordnung des Gesundheitszustandes. Und selbst die Daten sämtlicher Landkreise des zugehörigen Bundeslandes Nordrhein-Westfalen sind für eine umfängliche Bewertung nicht hinreichend. Es fehlen überregionale Vergleiche, beispielsweise mit den Kindern in sächsischen Landkreisen, wo sich die Situation möglicherweise deutlich anders darstellt. Unterschiedliche Versorgungsstrukturen, Präventionsangebote oder aber Einflussfaktoren auf die Gesundheit wie der sozioökonomische Status können hier eine Rolle spielen.

Die Darstellung von Unterschieden ermöglichen einen Ansatzpunkt die Ursachen zu prüfen.

Bundesweite Gesundheitsinformationen auf kleinräumiger Ebene werden jedoch in Deutschland bisher kaum bereitgestellt. Das bedeutet, dass bei gesundheitsrelevanten Entscheidungen in der Politik und in den Kassen auf Kennzahlen zurückgegriffen wird, die den Durchschnitt für ein größeres geographisches Kollektiv mit inhomogener siedlungs- und infrastruktureller Prägung bilden (zum Beispiel für eine Kassenärztliche Vereinigung). Auf diese Weise wird der durchschnittliche Gesundheitszustand einer sehr großen Population dargestellt, die gegebenenfalls unter sehr heterogenen strukturellen Bedingungen lebt. Die darunterliegenden räumlichen Disparitäten führen letztlich dazu, dass es möglicherweise zu regionalen Fehlsteuerungen der Versorgung und damit einer suboptimalen Versorgung kommt.

2.2 Vom Nutzen einer öffentlichen Gesundheitsberichterstattung

Zunächst stellt sich die Frage, warum es überhaupt eine (öffentliche) Berichterstattung zur Gesundheit gibt. Zahlreiche Versicherungs-, Versorgungs- und Fürsorgeinstitutionen begründen sich über die Leitidee, dass es eine gesamtgesellschaftliche Aufgabe sei, für die Würde des Einzelnen und das Recht auf körperliche Unversehrtheit einzutreten. Humanität und Solidarität sind Werte, die in Deutschland nach erbitterten Kämpfen zwischen den verschiedenen gesellschaftlichen Kräften im 19. Jahrhundert zum Sozialstaatsprinzip und zu zahlreichen Systemen der sozialen Sicherheit geführt haben. Dieser Umstand mündete schließlich im Jahr 1883 im Gesetz betreffend der Krankenversicherung der Arbeiter und auch in der Gründung der Allgemeinen Ortskrankenkassen (Tennstedt 1997; Reiners 2008).

Das deutsche Sozialstaatsprinzip und die gesetzlichen Krankenkassen sind seither unauflöslich miteinander verbunden. Obwohl im Grundgesetz ein Recht auf Gesundheit nicht explizit aufgeführt wird, erfüllt dieses nicht verbriefte Recht die Prinzipien unseres Sozialstaates. Gleichzeitig wird dem breiten gesellschaftlichen Konsens entsprochen, der hinter Institutionen wie dem Kranken- und Pflegeversicherungssystem steht (Pestalozza 2007). Somit geht die Gesundheitsberichterstattung des Bundes und der Länder wie auch die Armuts- und Sozialberichterstattung auf die Wohlfahrtsfürsorge des Staates zurück. Möglich wird das Recht auf Wohlfahrt und Gesundheit jedoch nur, wenn der Gesundheitszustand der Bevölkerung bekannt ist. Nur mit einer Erfassung des Gesundheitszustands, bei welcher auch die regionalen Morbiditätsunterschiede in der Bevölkerung berücksichtigt werden, ist eine angemessene flächendeckende Gesundheitsversorgung realisierbar.

Im Mittelpunkt der Gesundheitsberichterstattung der Bundesländer in Deutschland steht die Beschreibung der gesundheitlichen Lage der Bevölkerung.² Im Zusammenhang mit der demographischen, sozialen und ökonomischen Entwicklung werden die Inanspruchnahme von Leistungen der Gesundheitsversorgung, die personellen Ressourcen sowie die Ausgaben dargestellt und interpretiert. In der aktuell gültigen Fassung von 2003 wurde durch die Arbeitsgemeinschaft der Obersten

² https://www.lzg.nrw.de/ges_bericht/ges_indi/indikatoren_laender/index.html.

Landesgesundheitsbehörden (AOLG) ein Indikatorensatz³ abgestimmt, der in der Gesundheitsberichterstattung der Länder, soweit es die entsprechende Datenlage ermöglicht, zur Anwendung kommen kann. Über einen Satz von Kernindikatoren, der von allen Bundesländern vorgehalten werden sollte, ist eine Vergleichbarkeit der Bundesländer gewährleistet. Zudem ist angestrebt, dass diese Kernindikatoren kompatibel zu Indikatoren der Europäischen Union sind. Sofern keine Länderdaten zur Verfügung stehen, werden für die Gesundheitsberichterstattung des Bundes vom Statistischen Bundesamt und dem Robert Koch Institut sowie weiteren Datenhaltern Bundesindikatoren berechnet. Die Gesundheitsberichterstattung (GBE) des Bundes⁴ wurde in den 1990er ins Leben gerufen und ist seitdem gemeinsame Aufgabe des Robert Koch-Instituts und des Statistischen Bundesamts. In der Gesundheitsberichterstattung des Bundes fließen im Wesentlichen Daten aus Befragungsstudien, Registerdaten und Daten des statistischen Bundesamtes ein. Das Leibniz-Institut für Sozialwissenschaften (GESIS) hat einen Indikatorsatz entwickelt, der beim Aufbau der Berichterstattung des Bundes zum Einsatz kam (Brecht 1990). Die gewonnenen Daten werden über das Informationssystem der Gesundheitsberichterstattung des Bundes zur Verfügung gestellt bzw. über die Journals of Health Monitoring des Robert Koch-Instituts sowie in Berichten zu Schwerpunktthemen angeboten. Zur Vergleichbarkeit sowohl auf Länderebene als auch mit Daten der Europäischen Union wird in der Regel in entsprechenden Publikationen auf eine definierte Europapopulation standardisiert (Zapf 1974, 1975, 1977; Robert Koch-Institut 2017b). Ein zunehmendes Interesse der Bundesländer die gesundheitliche Ebene auch regional abzubilden führt zur Entwicklung von Indikatoren auf kommunaler Ebene. Es werden Gesundheitsberichte entwickelt, die es erlauben die gesundheitliche Lage der Bevölkerung auf kommunaler Ebene zu bewerten und Handlungsfelder zu identifizieren⁵.

In das Forschungsspektrum rückten seit den 1990er Jahren im Zuge der Sozialberichterstattung vermehrt auch die Ungleichheiten beim Krankheits- und Sterberisiko, die sich auf die soziale Lage oder Schicht sowie das Geschlecht zurückführen lassen (Mielck 2008). Gesundheitliche Disparität, bedingt durch den soziokulturellen Status, ist grundsätzlich mit den Prinzipien des Sozialstaates nicht vereinbar. Insofern nutzt die öffentliche Gesundheitsberichterstattung Indikatoren, die (wenn auch nur implizit) gesundheitliche Ungleichheit aufgrund der sozioökonomischen Lage beschreiben. Die Anforderungen an jegliche Berichtssysteme umfassen deshalb auch Hinweise auf gesundheitspolitische Stellschrauben zur Nivellierung hochrelevanter Unstimmigkeiten. Die Deskription regionaler Unterschiede soll diejenigen Krank-

³ http://www.gbe-bund.de/pdf/Indikatorensatz_der_Laender_2003.pdf.

⁴ https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsK/2010_1_Ziele_Aufgaben.pdf?__blob=publicationFile.

⁵ Eine Auswahl..

<https://www.hamburg.de/sozialversicherung/4133760/morbiditaetsatlas-hamburg.html>.

<http://www.gesundheitsatlas-bw.de/>.

https://www.lgl.bayern.de/gesundheitsberichterstattung/gesundheitsatlas/ia_report/atlas.html.

https://www.lzg.nrw.de/ges_bericht/ges_indi/indikatoren_kreise/index.html.

heiten und Risikofaktoren identifizieren, bei denen die gesundheitlichen Ungleichheiten am größten sind (Kunst 2009).

2.3 Nutzen kleinräumiger Darstellungen des Gesundheitszustandes

Wie bereits erwähnt, bildet die Darstellung der gesundheitlichen Lage auf regionaler Ebene das Fundament für verschiedene relevante Handlungsfelder wie Gesundheitspolitik, Versorgungsforschung und die ärztliche Bedarfsplanung. Auch Präventionsangebote oder Selektivverträge von Krankenkassen können mit regionalisierten Daten deutlich zielgerichteter ausgestaltet werden, wovon schließlich die Versicherten profitieren. Unabdingbar dafür sind jedoch valide Kennziffern zur Häufigkeit von Krankheiten auf kleinräumiger Ebene (Mangiapane 2014; Nolting 2015; Klauber et al. 2012). Die so gewonnenen Informationen können beispielsweise für folgende Zwecke genutzt werden:

- Analyse und Bewertung der (infra-)strukturellen Rahmenbedingungen im Kontext von Krankheits- und Sterberisiko (Morbidität und Mortalität),
- Aufzeigen und Analysieren von Gründen für regionale Ausgabenunterschiede
- Identifikation von Regionen mit verbesserungswürdiger Versorgungsqualität oder erhöhten Versorgungsbedarfen,
- Entwicklung maßgeschneiderter indikationsspezifischer und regionalisierter Präventions-, Früherkennungs- und Versorgungsangebote,
- gezielte Steuerung der Versorgung anstelle des „Gießkannenprinzips“.

In einem ersten Schritt müssen also die regionalen Variationen des Gesundheitszustandes sichtbar gemacht und Zusammenhänge aufgezeigt werden. Basierend darauf können Unterschiede bewertet werden. Initiativen und Präventionsprogramme können anschließend problemorientiert und zielgerichtet hinsichtlich relevanter Risikogruppen oder spezifischer Risikogebiete eingesetzt werden (Altgeld 2008).

Die Anreicherung von regionalisierten Gesundheitsinformationen um kleinräumige sozioökonomische sowie weitere externe Datenquellen ermöglicht die tiefergehende Untersuchung des Zusammenhangs von Gesundheitsparametern mit der (medizinischen) Infrastruktur, den Umwelt- und Arbeitsbedingungen, sozioökonomischen Faktoren oder den kulturellen Mustern, die das gesundheitsbezogene Verhalten von Patienten und Ärzten beeinflussen. Einige deutsche Studien haben den Zusammenhang zwischen dem Wohnort (angereichert um einen regionalbasierten sozioökonomischen Index) und dem Gesundheitszustand der Bevölkerung kleinräumig analysiert. Beispielhaft seien hier Untersuchungen zum Einfluss von Umweltbedingungen, regionaler Deprivation, der Versorgungsstruktur und der Präferenzen von Patienten oder Ärzten erwähnt (Behrens und Noll 2006; Mielck 2008; Maier et al. 2012, 2013; Sundmacher und Busse 2012; Burgdorf und Sundmacher 2014; Kroll et al. 2016; Storz-Pfennig 2012).

Deutschland folgt damit auch einem internationalen Trend: Einige europäische Länder, etwa Großbritannien, Spanien oder die Niederlande, stellen seit Jahren im Internet den nationalen Gesundheitszustand in Form von Karten und Abbildungen zu speziellen Indikatoren dar. Dabei werden unter anderem die Säuglingssterblichkeit, die Prävalenz weitverbreiteter Krankheiten oder spezifische Eingriffshäufigkeiten

abgebildet. In Anlehnung an die erste kleinräumige und kartografische Darstellung gesundheitsrelevanter Merkmale, den Dartmouth Atlas of Health, werden bei diesen Projekten landesweite kleinräumige Analysen in einem Gesundheitsatlas zusammengestellt. Der Ansatz, der bei den angewendeten Kennzahlen verfolgt wird, ist immer raumorientiert (Wennberg und Gittelsohn 1973; Wennberg 2010 2014). Adressaten sind zum Beispiel Fachjournalisten, politische Entscheidungsträger oder Wissenschaftler (Jong 2008; van den Berg et al. 2010; Klemperer und Robra 2014; Mangiapane 2014; Böcken et al. 2015). Durch die Darstellung der Ergebnisse, bei der die Vorteile der visuellen Kommunikation genutzt werden, wird neben Akteuren des Gesundheitssystems aber auch die interessierte Öffentlichkeit erreicht. Somit wird eine umfassende, grundlegende Informationsbasis zur gesundheitlichen Lage einer Population zur Verfügung gestellt.

2.4 Limitierungen bestehender regionaler Darstellungen: mangelnde Kleinräumigkeit oder mangelnde Repräsentativität

Prinzipiell existiert bereits eine Vielzahl verschiedener Datenquellen und Gesundheitsinformationen, die Aussagen über den Gesundheitszustand der Bevölkerung zulassen. Ganz besonders hervorzuheben ist hier die Gesundheitsberichterstattung des Robert Koch-Instituts. Datenquellen zu ausgewählten Erkrankungen stehen zum Beispiel in Form von Befragungen oder Untersuchungserhebungen (Rommel 2016) oder Registerdaten zur Verfügung (Radespiel-Tröger und Meyer 2017). Jedoch erlaubt deren Stichprobenumfang häufig nur eine Darstellung auf Bundeslandebene, aber keine kleinräumigere regionale Gliederung (Kroll und Lampert 2011). Für eine Darstellung von Gesundheitsindikatoren bis auf Ebene der Kreise wurden Schätzverfahren unter Nutzung von Hilfsinformationen untersucht (Kroll et al. 2017). In der Regel fokussieren vorhandene Erhebungs- oder Registerdaten zudem auf nur wenige oder einzelne Krankheiten, sodass kein umfassendes Bild über verschiedene Krankheiten oder Gesundheitszustände erhoben werden kann. Weitere Limitierungen bei Befragungen mit freiwilliger Teilnahme ergeben sich aus den bekannten Problemen des Nonresponse-Bias (Cheung et al. 2017).

Aufgrund der genannten Einschränkungen der Befragungsdaten werden in den letzten Jahren auch vermehrt Routinedaten der gesetzlichen Krankenkassen (GKV) zur Erforschung des Gesundheitszustands verwendet. Über einen anonymisierten Versichertenbezug ist es möglich sektorübergreifende Kontakte des Versicherten mit dem Gesundheitswesen abzubilden. Hierrüber sind sowohl Quer- als auch Längsschnittanalysen möglich (Lange et al. 2014). So werden diese Daten zum Beispiel für die Diabetes-Surveillance des Robert Koch-Instituts (Schmidt et al. 2017), der regionalisierten Krankheitslastberechnung (Rommel et al. 2018) oder für regionalisierte Darstellungen des Versorgungsgeschehens (Mangiapane 2014) genutzt. In den Routinedaten der Krankenkassen sind neben den krankheitsbezogenen Informationen auch Angaben zum Wohnort der Versicherten enthalten, sodass kleinräumig sehr differenzierte Auswertungen möglich sind (Kauhl et al. 2016). Sofern eine Krankenkasse bundesweit flächendeckend tätig ist, können mit deren Routinedaten also kleinräumige Untersuchungen zum Gesundheitszustand durchgeführt werden.

Auch die Routinedaten der gesetzlichen Krankenkassen unterliegen jedoch wichtigen Limitierungen – am bedeutsamsten ist hier die Einschränkung, dass nur Informationen in den Daten enthalten sind, die kosten- oder erstattungsrelevant sind. Bei den Diagnose- und Leistungsdaten sind zudem Verzerrungen durch Rechtfertigungs- oder Abrechnungsstrategien zu berücksichtigen oder durch systembedingte Veränderungen (Glaeske 2010; Ohlmeier et al. 2014). Die auf Basis von Routinedaten gewonnenen Krankheitshäufigkeiten werden daher mit dokumentierter bzw. administrativer Prävalenz bezeichnet. Aus epidemiologischer Sicht fehlen zudem wichtige klinische Angaben, Labor- und Messwerte sowie Informationen zum Gesundheitsverhalten – wie zum Beispiel Blutdruckwerte, Körpergewicht oder sportliche Aktivität. Trotz dieser Limitierungen lassen sich mit Routinedaten Krankheitshäufigkeiten über Jahre und Regionen hinweg gut darstellen (Swart et al. 2008). Ein weiterer Vorteil in der Nutzung von Routinedaten ist die relativ kostengünstige Verfügbarkeit, da diese Daten für administrative Zwecke erhoben werden und der Datenanalyse somit leicht zugänglich sind. Zudem ist in Deutschland ein Großteil der Einwohner gesetzlich krankenversichert,⁶ sodass Verzerrungen durch Interviewer-Effekte, Recall-Bias oder den Ausschluss von Personengruppen, die in Befragungsdaten schwer erreichbar sind, ebenfalls nicht bestehen (Gerste et al. 2016). Somit stellen Routinedaten eine sinnvolle Ergänzung zu Primärdaten aus epidemiologischen Studien hinsichtlich einer vollständigen Erfassung von Gesundheitsindikatoren dar.

Das große Potenzial für Analysen anhand der Routinedaten der gesetzlichen Krankenkassen hat auch der Gesetzgeber mit der Datentransparenzverordnung (DaTraV) im Jahr 2012 anerkannt. Diese Verordnung ermöglicht es, die im Rahmen des Risikostrukturausgleichs an das Bundesversicherungsamt übermittelten Krankenkassendaten beim DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) bereitzustellen. Nutzungsberechtigte Institutionen können für fest definierte Auswertungszwecke Daten in aggregierter Form anfordern. Mit der Aktualisierung des SGB V § 303b im Jahr 2014 sind Angaben zum Regionalkennzeichen (Postleitzahl) enthalten. Somit sind auf diesem Datenbestand regionalisierte Auswertungen zu den mehr als 72,8 Mio. gesetzlich krankenversicherten Personen in Deutschland im Jahre 2018 möglich. Allerdings liegen Regionalkennungen derzeit nur für die Berichtsjahre 2009 und 2010 vor und werden voraussichtlich erst ab 2020 wiederum ab dem Berichtsjahr 2016 angeboten (DIMDI 2018). Nach einer zweijährigen Erprobungsphase wurde entsprechend der Datentransparenzverordnung das Informationssystem Versorgungsdaten 2015/2016 einer ersten Evaluation unterzogen. Die Einschätzungen vom DIMDI wie auch die der Nutzer aus den nutzungsberechtigten Institutionen zeigen zahlreiche Handlungsfelder: So scheint unter anderem eine Optimierung des Datenangebots hinsichtlich Umfang, Aktualität, Struktur und Qualität aber auch die Optimierung des Antragsverfahrens zur Reduzierung des Aufwands und der Bearbeitungsdauer notwendig zu sein (DIMDI 2016a). Die für die Nutzung der DaTraV-Daten zu bewältigenden hohen Aufwände sind einer explorativen Nutzung der Daten zur Generierung neuer Methoden nicht zuträglich. Darüber hinaus

⁶ Im Jahr 2016 waren ca. 87% der Bevölkerung über gesetzliche Krankenversicherungen (GKV) versichert (KM6 Statistik 2016), weitere 11% über private Krankenversicherungen (PKV) (PKV 2016).

kann die Wahrung des Identitätsschutzes der Versicherten ggf. eine Datenbereitstellung von granularen regionalen Daten zu Forschungszwecken verhindern.

Bundesweit regionalisierte Analysen sind auch im Wissenschaftlichen Institut der AOK (WiDO) anhand der anonymisierten Routinedaten von mehr als 26,5 Mio. AOK-Versicherten in 2018 möglich. Diese Daten werden für sektorenspezifische Analysen wie beispielsweise im Arzneiverordnungs-Report (Schwabe et al. 2018), Krankenhaus-Report (Klauber et al. 2019), Fehlzeiten-Report (Badura et al. 2018), Pflege-Report (Jacobs et al. 2018) oder Heilmittelbericht (Waltersbacher 2018) wie auch sektorenübergreifend im Versorgungs-Report (Günster et al. 2019) genutzt. Doch darüber hinaus können systematisch Erkrankungshäufigkeiten zu verschiedenen Krankheiten regional nach Kreisen und kreisfreien Städten ausgewertet werden. Hierüber können gezielt Regionen mit Handlungsbedarf detektiert werden und vertiefende zielgerichtete Analysen und Maßnahmen erfolgen. Diese Daten sind, im Vergleich zu den DaTraV-Daten, mit geringerem Zeitverzug nutzbar. Zudem liegen – über den DaTraV-Datensatz hinaus – weitere relevante Leistungsdaten wie etwa zu Krankenhausambulanzen, Heil- und Hilfsmittelverordnungen sowie Operationen bzw. Prozeduren vor (Nimptsch et al. 2014).

Die wesentliche Limitierung in der Nutzung von Routinedaten der gesetzlichen Krankenkassen ist jedoch die Tatsache, dass die Versichertenstruktur keine Zufallsstichprobe der deutschen Bevölkerung darstellt. Gesetzlich krankenversicherte Personen unterscheiden sich systematisch von privat versicherten Personen, da die private Krankenversicherung nur für spezielle Berufsgruppen oder bei einem die Versicherungspflichtgrenze überschreitenden, höheren Einkommen möglich ist (Sozialgesetzbuch SGB V, §§ 5 und 6). In vielen Untersuchungen wurden gesundheitliche Unterschiede zwischen privat und gesetzlich krankenversicherten Personen sowie zwischen verschiedenen gesetzlichen Krankenkassen gezeigt (Huber et al. 2010; Kriwy und Mielck 2006; Hoffmann und Icks 2012), wobei GKV-Versicherte in der Regel kränker waren – auch nach Kontrolle für demographische Faktoren. Diese Ergebnisse stehen im Einklang mit Untersuchungen, die einen Zusammenhang zwischen sozialem Status und Gesundheitszustand zeigen (Lampert et al. 2016). Dementsprechend ist davon auszugehen, dass neben rein demographischen Faktoren wie Alter und Geschlecht auch sozioökonomische Einflüsse wie Bildung, Einkommen und berufliche Stellung mit Morbiditätsunterschieden assoziiert sind – und gleichzeitig sind diese sozialen Faktoren des Einkommens und des Berufs mit der Zugehörigkeit zur gesetzlichen oder zur privaten Krankenversicherung assoziiert.

Die Mitgliedschaft einer Person bei einer Krankenkasse (hier: AOK – Die Gesundheitskasse) kann demnach statistisch informativ für ihr Risiko hinsichtlich bestimmter Krankheiten sein. Dies führt zu Verzerrungen, wenn (regionale) Krankheitshäufigkeiten (Prävalenzen) der Gesamtbevölkerung auf Basis der Versicherten einer spezifischen Krankenkasse geschätzt werden. Um regionale Prävalenzschätzungen für die Gesamtbevölkerung auf Basis der Mitglieder einer Krankenkasse durchzuführen, muss demnach die Informativität der Kassenzugehörigkeit hinsichtlich der Morbidität berücksichtigt werden. Die kassenspezifische Krankheitshäufigkeit nur nach Alter und Geschlecht zu adjustieren ist hierfür nicht ausreichend, da Morbiditätsunterschiede auch nach einer Alters- und Geschlechtsstandardisierung bestehen bleiben können, wie in Abschn. 1 bereits erwähnt wurde.

3 Methodik

3.1 Daten

3.1.1 Datenquellen

Die AOK-Gemeinschaft, bestehend aus elf regionalen AOKs, versicherte im Jahr 2016 rund 25 Mio. Personen. Dies entsprach knapp 35 % aller gesetzlich Krankenversicherten (KM6-Statistik 2016) und knapp 31 % der Bevölkerung Deutschlands. Im Wissenschaftlichen Institut der AOK (WiDO) liegen zu diesem Versichertenkollektiv anonymisierte Routinedaten zu Diagnosen, Prozeduren, Arznei- und Heilmittelverordnungen sowohl aus dem ambulanten wie auch dem stationären Versorgungsbereich im Längsschnitt vor (Nimptsch et al. 2014). Auf Basis dieser Daten können Krankheitshäufigkeiten (Prävalenzen und Inzidenzen) abgebildet werden. Die Datenverfügbarkeit, die über die Strukturen des Datensatzes entsprechend der Datentransparenzverordnung (DaTraV) hinausgeht, ermöglicht zudem umfangreichere Plausibilisierungen der verwendeten Daten zur Definition von Morbidität (Swart et al. 2014). Versichertenbezogene Informationen zu Alter, Geschlecht und Regionalkennzeichen des Wohnorts erlauben, die Alters- und Geschlechtsstruktur sowie die regionale Zuordnung zu Landkreisen und kreisfreien Städte abzubilden. Zur Alters- und Geschlechtsstruktur der Bevölkerung sind korrespondierende Referenzinformationen über das Statistische Bundesamt regional verfügbar (Statistisches Bundesamt 2018). Somit können strukturelle Unterschiede der AOK-Versicherten zur Bevölkerung hinsichtlich Alter und Geschlecht innerhalb der Regionen über eine Standardisierung ausgeglichen werden.

Als Datenquelle, die auch Morbiditätsaspekte enthält, stellt das Statistische Bundesamt die fallpauschalenbezogene Krankenhausstatistik (DRG-Statistik nach § 21 KHEntgG) zur Verfügung (Statistisches Bundesamt 2017). Diese Statistik umfasst jährlich die dokumentierten Diagnosen, Operationen und Prozeduren aller stationären Fälle in deutschen Krankenhäusern für die gesamte deutsche Bevölkerung (GKV, PKV, Touristen aus dem Ausland). Diese Daten nach dem Wohnort des Patienten regionalisiert bis auf Ebene der Kreise und kreisfreien Städte wurden dem WiDO durch eine entsprechende Datenbereitstellung durch das statistische Bundesamt bereitgestellt. Weiterhin sind die Diagnoseangaben bis auf Ebene des dreistelligen ICD-10-Codes vorhanden und werden nach Alters- und Geschlechtsgruppen differenziert ausgewiesen. Es werden sowohl die den Krankenhausaufenthalt bedingenden Diagnosen als auch relevante Nebendiagnosen übermittelt. Für die AOK-Versicherten können aus den AOK-eigenen Routinedaten analoge Statistiken erstellt werden. Die Daten der Bevölkerungsstatistik und der DRG-Statistik stellen somit eine geeignete Hilfsinformation dar, um strukturelle Informationen hinsichtlich der Morbidität sowohl nach Alter und Geschlecht als auch nach Kreisen und kreisfreien Städte abzubilden.

Für die Anwendung der Methodik im Rahmen des vorliegenden Beitrags werden dokumentierte Krankheitsprävalenzen für Typ-2-Diabetes⁷ betrachtet, wobei die Prävalenzen für die Wohnbevölkerung ausgehend von den AOK-Krankheitshäufigkeiten unter Nutzung der Daten zu stationären Behandlungshäufigkeiten und der Bevölkerungsdemographie geschätzt werden. Die dokumentierten AOK-Krankheitshäufigkeiten für Patienten mit Typ-2-Diabetes werden anhand der AOK-Leistungsdaten nach einem hierarchischen Algorithmus erfasst. Dieser umfasst eine interne Diagnosevalidierung. Es werden die relevanten Diagnosen für Typ-2-Diabetes sowohl in der stationären als auch in der ambulanten Versorgung geprüft. Die dokumentierten Diagnosen werden mitunter durch Berücksichtigung spezifischer Arzneimittelverordnungen sowie der Teilnahme im Disease Management Programm zu Typ-2-Diabetes zusätzlich validiert.

Datengrundlage zur Ermittlung der dokumentierten Prävalenz des Typ-2-Diabetes sind die in einem Jahr dokumentierten Diagnosen zu Diabetes (ICD: E10–E14 der ICD-10-GM Klassifikation (DIMDI 2016b)). Hierbei werden ausschließlich ambulante Diagnosen mit dem Zusatzkennzeichen „gesichert“, stationäre Haupt- und Nebendiagnosen sowie Diagnosen aus dem ambulanten Krankenhausbereich (wie zum Beispiel der ambulanten spezialfachärztlichen Versorgung im Krankenhaus oder der ambulanten integrierten Versorgung) berücksichtigt. Da davon auszugehen ist, dass die Diagnosen teilweise inkonsistent dokumentiert werden (bspw. gleichzeitig Diagnosen vom Typ 1 und Typ 2; Tamayo et al. 2016), erfolgt eine Einteilung in Typ-1- und Typ-2-Diabetes nach in einem mehrstufigen Prozess. Im ersten Schritt werden Patienten ohne Insulinverordnungen oder Patienten mit Verordnungen von oralen Antidiabetika (außer Metformin) dem Typ-2-Diabetes zugewiesen. So werden knapp 85 % der initial ermittelten Diabetes-Patienten dem Typ-2-Diabetes zugewiesen. Für die verbleibenden Patienten wird geprüft, ob ausschließlich eindeutige stationäre Hauptdiagnosen zu E11 „Nicht primär insulinabhängiger Diabetes mellitus [Typ-2-Diabetes]“ dokumentiert wurden. Auch hier kann von einer korrekten Dokumentation des Typ-2-Diabetes ausgegangen werden. Im nachfolgenden Schritt wird für die noch nicht zugewiesenen Patienten geprüft, ob eine eindeutige Einschreibung im Disease-Management-Programm (DMP) zu Typ-2-Diabetes vorliegt. Für die restlichen Patienten erfolgt die Zuweisung in Typ-2-Diabetes, wenn die ambulanten Diagnosen bzw. stationären Nebendiagnosen oder die ambulanten Krankenhausdiagnosen eindeutig einen Typ-2-Diabetes dokumentieren. Die nun verbleibende Patientengruppe wird über die relativen Häufigkeiten der dokumentierten Diabetesdiagnosen dem entsprechenden Diabetes zugewiesen, indem mindestens zwei Behandlungsfälle mehr mit einer Typ-2-Diagnose (ICD E11) als mit einer Typ-1-Diagnose (ICD

⁷ Diabetes mellitus (Zuckerkrankheit) ist eine häufig vorkommende Stoffwechselerkrankung. Es werden verschiedene Formen unterschieden. Der Typ-1-Diabetes tritt vorwiegend bei Kindern, Jugendlichen und jungen Erwachsenen auf und beruht auf einem absoluten Insulinmangel aufgrund einer Zerstörung der insulinproduzierenden Zellen in der Bauchspeicheldrüse. Der Typ-2-Diabetes ist die weitaus häufigste Form in Deutschland und ist meist durch relativen Insulinmangel aufgrund einer Insulinresistenz der Körperzellen verursacht. Diese Form des Diabetes tritt meist erst in höherem Erwachsenenalter auf, wobei Adipositas und Bewegungsmangel wichtige Risikofaktoren für die Entstehung eines Typ-2-Diabetes darstellen. Bei Frauen spielt zudem der Schwangerschaftsdiabetes eine Rolle, der sich zwar nach der Schwangerschaft in der Regel zurückbildet, aber das Risiko für die spätere Entwicklung eines Typ-2-Diabetes erhöht.

E10) oder sonstigen spezifischen Diabetes-Diagnose (ICD E13) vorgelegen haben müssen. Über diesen Zuordnungsalgorithmus wurden im Jahr 2016 96,7 % der Diabetiker in der AOK dem Typ-2-Diabetes zugeordnet, 2,8 % dem Typ 1 und 0,5 % der Diabetiker konnten keinem dieser beiden Diabetestypen zugeordnet werden.

Es sei nochmals darauf hingewiesen, dass es sich hierbei um die dokumentierte Prävalenz des Typ-2-Diabetes handelt, die näherungsweise die tatsächliche Typ-2-Diabetes-Prävalenz beschreibt. Unentdeckter bzw. nicht diagnostizierter Diabetes kann auf Basis dieser Daten nicht erfasst werden. Dies führt möglicherweise zu einer Unterschätzung tatsächlicher Typ-2-Diabetes-Prävalenz.

In den AOK-Diagnosedaten werden Diagnosen der Neugeborenen häufig über die Krankenversichertenkarte der Mutter dokumentiert. Da in den Daten der DRG-Statistik (§ 21) des Statistischen Bundesamtes zu erwarten ist, dass derartige Diagnosen dem Alter und Geschlecht des Kindes zugeordnet sind, werden sie aus allen Datenkörpern entfernt. Die Versichertenzählung der AOK erfolgt über Versichertenjahre (Versichertendauer/Anzahl Tage im Berichtsjahr), da es sich bei den AOK-Versicherten um eine offene Kohorte handelt, sodass Versicherte unterjährig die Kassen verlassen oder neu hinzukommen können. Somit wird die durchschnittliche Versichertenanzahl innerhalb eines Jahres berechnet. Die Zählung der Bevölkerung erfolgt ebenfalls über die durchschnittliche Bevölkerungsanzahl mit Stand zum 31.12. des Berichtsjahres und des Vorjahres.

3.1.2 Regionale Gliederung und Altersgruppeneinteilung

Wie in Abschn. 1 bereits erwähnt, erfolgt die Schätzung der regionalen Krankheitshäufigkeiten anhand linearer Modelle mit regionaler Parameterschätzung. Somit ist es erforderlich, die Zellen (Kreise und kreisfreie Städte, Altersklasse, Geschlecht) zu übergeordneten Einheiten zusammenzufassen. Dazu wurden zunächst die Kreise und kreisfreien Städte nach Regionen der Kassenärztlichen Vereinigungen (KV-Regionen)⁸ zusammengefasst. KV-Regionen sind administrative Gebiete, in denen das Vertragsgeschehen der Leistungserbringer mit den Kassen geregelt ist. So gibt es beispielsweise spezielle Arzneimittelvereinbarungen oder das Vergütungssystem der ambulanten kollektiven Versorgung auf Ebene der KV-Regionen. Die 401 Kreise und kreisfreie Städte des Jahres 2016 verteilen sich auf insgesamt 17 KV-Regionen. Die Stadtstaaten Berlin, Hamburg sowie Bremen/Bremerhaven stellen jeweils eine eigene KV-Region dar, wenngleich diese auch nur aus jeweils einem bzw. zwei Kreisen bestehen. Um eine ausreichende Beobachtungsanzahl auch in diesen KV-Regionen mit singulären Kreisen zu gewährleisten, wurden die Stadtstaaten zusammengefasst. In einer ersten Variante werden alle drei Stadtstaaten zu einer Region zusammengefasst. Aufgrund der Heterogenität der Städte wird zudem eine zweite Variante gewählt, bei der Bremen und Bremerhaven der benachbarten Region Niedersachsen zugeordnet werden. Die vier größten Kreise bzw. kreisfreien Städte

⁸ Die KV-Regionen sind administrative Einheiten der 17 Kassenärztlichen Vereinigungen, die für die vertragsärztliche Versorgung der GKV-Versicherten zuständig sind und die in Regel den Bundesländern entsprechen. Die Ausnahme stellt Nordrhein-Westfalen dar: Hier sind Nordrhein und Westfalen eigenständige KV-Regionen.

Berlin, Hamburg, München und die Region Hannover werden zu einem großstädtischen Regionstyp zusammengefasst.

Ein dritter Ansatz ist es, die Daten nicht nach administrativen Gebieten, sondern nach siedlungsstrukturellen Typen zusammenzufassen. Dies soll der Tatsache Rechnung tragen, dass Unterschiede betreffend Risikofaktoren und Gesundheitszustand der Bevölkerung zwischen ländlichen und städtischen Regionen beschrieben wurden, beispielsweise für das Rauchverhalten sowie für Depressionen (Völzke et al. 2006; Busch 2013). In dieser Variante werden die Kreise und kreisfreien Städte nach ihren siedlungsstrukturellen Kreistypen und Regionstypen entsprechend der Klassifikation des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR) zusammengefasst (BBSR 2015). Aus der Kombination von drei verschiedenen Raumordnungstypen (städtische Regionen, Regionen mit Verstärkeransätzen, ländliche Regionen) mit vier verschiedenen Kreistypen (Großstädte, städtische Kreise, ländliche Kreise mit Verdichtungsansätzen, dünn besiedelte ländliche Kreise) ergeben sich zunächst zwölf Gruppen, wobei jedoch die Kombination „dünn besiedelter ländlicher Kreis“ in „städtischer Region“ nur zweimal auftrat. Um eine ausreichende Anzahl an Beobachtungen zu gewährleisten, wurde diese Kombination der Gruppe „dünn besiedelter ländlicher Kreis“ in „Regionen mit Verstärkeransätzen“ zugewiesen, sodass elf verschiedene Kombinationen resultieren, die jeweils mindestens vier Kreise umfassen. Zusammenfassend liegen also drei verschiedene regionale Einteilungstypen vor, die in die verschiedenen Schätzungen eingehen.

Die Altersgruppeneinteilung basiert auf 19 Altersgruppen (0–4, 5–9, 10–14, ..., 80–85, ≥ 90 Jahre). Je nach Krankheit, welche im Zuge des Projekts bearbeitet wird, sind bestimmte Altersgruppeneinteilungen wenig oder nahezu gar nicht besetzt. So erscheint es zum Beispiel wenig sinnvoll, für Patienten mit Demenz die Altersgruppen der Kinder zu betrachten. Daher wird aus sachlogischen Vorüberlegungen und aufgrund der Häufigkeitsverteilung der Patientenraten eine Altersgruppeneinteilung vorgenommen. Ziel ist, eine Altersklasseneinteilung zu wählen, die die empirische Verteilung der Prävalenzraten nahezu normalverteilt abbildet. Hierzu werden die Häufigkeitsverteilungen (Histogramm und Kerndichteschätzer) geprüft und Altersgruppen entsprechend zusammengefasst. Die Altersklasseneinteilung ist somit krankheitsspezifisch. Je nach Krankheit werden zwischen vier und zehn Altersgruppen berücksichtigt. Die Altersgruppeneinteilung für Typ-2-Diabetes in diesem Beitrag ist: bis 54 Jahre (AK 1), 55–59 Jahre (AK 2), 60–64 Jahre (AK 3), 65–69 Jahre (AK 4), 70–74 Jahre (AK 5), 75–79 Jahre (AK 6), 80–84 Jahre (AK 7), 85–89 Jahre (AK 8), 89 Jahre und älter (AK 9).

3.2 Statistische Grundlagen

Zunächst wird der Inferenzschluss, welcher dem Schätzverfahren zugrunde liegt, skizziert. Ausgangspunkt ist der Versichertenstamm der AOK-Gemeinschaft inklusive seiner gesundheitsbezogenen Informationen. Hierfür wird der statistische Rahmen von Burgard et al. (2019) verwendet. Es sei U die Gesamtbevölkerung bestehend aus N Individuen, deren Krankheitshäufigkeiten einer relevanten Diagnose kleinräumig und referenziert nach Alter und Geschlecht geschätzt werden sollen. Die Gesamtbevölkerung U kann als die Vereinigung von Mitgliedern $\tilde{U}$ und Nicht-

Mitgliedern $U \setminus \widetilde{U}$ der AOK ausgedrückt werden. Zusätzlich wird U hinsichtlich seiner Geographie in drei hierarchische Aggregationsebenen unterteilt:

- Ebene 1: Regionen r der Größe N_r mit $r = 1, \dots, R$
- Ebene 2: Kreise d der Größe N_{rd} mit $d = 1, \dots, D_r$
- Ebene 3: Zellen c der Größe N_{rd}^c mit $c = 1, \dots, C$, die Alters- und Geschlechtskombinationen darstellen.

Dabei ist zu beachten, dass die Anzahl der Regionen R von der in Abschn. 3.1.2 dargestellten Einteilung abhängt. Zusätzlich wird die Anzahl der Zellen C innerhalb eines Kreises von der Einteilung der Alters- und Geschlechtskombinationen determiniert, welche auch in Abschn. 3.1.2 festgesetzt wurde. Die Zellgrößen N_{rd}^c und der Anteil an Mitgliedern innerhalb einer Zelle $\widetilde{N}_{rd}^c / N_{rd}^c$ sind aus den Krankenkassendaten und der Bevölkerungsstatistik bekannt. Darüber hinaus wird aus Gründen der Einfachheit zunächst angenommen, dass jede Zelle eine nicht-leere Teilmenge an Mitgliedern hat ($\widetilde{N}_{rd}^c > 0 \forall r, d, c$). Das Ziel ist nun die zellspezifischen Krankheitshäufigkeiten der relevanten Diagnose in der Gesamtbevölkerung U zu schätzen. Es sei y_{rd}^c die unbekannte Krankheitshäufigkeit der Gesamtbevölkerung in Zelle c des Kreises d , welcher sich in Region r befindet. Aus den Krankenkassendaten ist die kassenspezifische Krankheitshäufigkeit $\widetilde{y}_{rd}^c \leq y_{rd}^c$ für jede Zelle bekannt. Aufgrund der Informativität der Kassenzugehörigkeit für die Morbidität erhält man mit einer einfachen Skalierung der kassenspezifischen Krankheitshäufigkeiten durch Bedingung auf regionale Demographie eine verzerrte Schätzung der Zellprävalenz:

$$E(\widetilde{y}_{rd}^c | \widetilde{N}_{rd}^c) \neq E(y_{rd}^c | N_{rd}^c) \rightarrow E\left(\frac{N_{rd}^c}{\widetilde{N}_{rd}^c} \widetilde{y}_{rd}^c\right) \neq y_{rd}^c \forall r, d, c, \quad (1)$$

wobei $N_{rd}^c / \widetilde{N}_{rd}^c$ den demographischen Skalierungsfaktor darstellt. Folglich gibt es systematische Unterschiede in den Krankheitshäufigkeiten zwischen dem Versicherungstamm der AOK und der Gesamtbevölkerung auf der Zellebene, welche sich nicht durch Alter und Geschlecht erklären lassen. Die Idee ist nun zusätzliche Daten zu verwenden, welche die Morbiditätsunterschiede zwischen den Populationen besser erklären. Wie zuvor bereits beschrieben, werden hierfür stationäre Behandlungshäufigkeiten der DRG-Statistik herangezogen. Die Auswahl der für die Morbiditätsadjustierung geeigneten Diagnosen innerhalb der DRG-Statistik erfolgt auf Basis multivariater Analyseverfahren, welche in Abschn. 3.3 beschrieben werden. Zunächst wird davon ausgegangen, dass entsprechende Diagnosen im Vorfeld bekannt sind. Es sei X eine $[(C \cdot \sum_r D_r) \times p]$ -Matrix mit Häufigkeiten ausgewählter Diagnosen auf der Zellebene, welche sich auf die Gesamtbevölkerung beziehen. Daran anknüpfend sei $\widetilde{X}$ die gleiche Matrix, welche sich ausschließlich auf die AOK-Versicherten bezieht. Diese kann auf Basis der AOK-Routinedaten in Analogie zur DRG-Statistik erzeugt werden. Für die Häufigkeit in einer gegebenen Zelle werden $x_{rd}^c, \widetilde{x}_{rd}^c$ als $(1 \times p)$ -Vektoren notiert. Wenn diese Hilfsinformationen über eine hinreichend große Prädiktivkraft hinsichtlich der zellspezifischen Krankheitshäufig-

Abb. 1 Skizze des Inferenzschlusses

keiten verfügen, dann können die systematischen Morbiditätsunterschiede zwischen den Populationen durch sie erklärt werden und es gilt

$$E[y_{rd}^c | N_{rd}^c, \mathbf{x}_{rd}^c = (x_{rd}^{c1}, \dots, x_{rd}^{cp})] = E[\tilde{y}_{rd}^c | \tilde{N}_{rd}^c, \tilde{\mathbf{x}}_{rd}^c = (\tilde{x}_{rd}^{c1}, \dots, \tilde{x}_{rd}^{cp})]. \quad (2)$$

Dies impliziert, dass nach der Bedingung auf Alter und Geschlecht sowie den stationären Behandlungshäufigkeiten keine systematischen Unterschiede in zellspezifischen Krankheitshäufigkeiten zwischen AOK-Versichertenstamm und Gesamtbevölkerung mehr existieren. Wenn diese Annahme hält, dann kann der Vorteil hinsichtlich der regionalen Beobachtungszahlen der Krankenkassendaten für eine modellbasierte und korrigierte Quantifizierung von y_{rd}^c im Lichte der in Gl. 1 dargestellten Verzerrung mit großer Präzision verwendet werden. Der Inferenzschluss wird in Abb. 1 nochmals vereinfacht dargestellt.

Zunächst werden die kleinräumigen Morbiditätsstrukturen des AOK-Versichertenstammes auf AOK-Demographie sowie ausgewählte AOK-Routinedaten, welche analog zur DRG-Statistik erzeugt wurden, bedingt. Die Bedingung erfolgt dabei anhand von geeigneten statistischen Modellen. Anschließend werden die Daten der Bevölkerungsstatistik sowie der DRG-Statistik als Benchmarks verwendet, um die bedingten AOK-Morbiditätsstrukturen kleinräumig zu adjustieren. Auf dieser Basis können dann angepasste Prädiktionen der Morbiditätsstrukturen auf regionaler Ebene in der Gesamtbevölkerung erzeugt werden. Die Prädiktion erfolgt dabei auf Basis der statistischen Modelle, welche zuvor für die Bedingung der AOK-Morbiditätsstruktur verwendet wurden. Im Zuge dieses Verfahrens stellt Gl. 2 eine zentrale Annahme dar, deren Validität sorgfältig geprüft werden muss. Im Wesentlichen werden zwei Bedingungen für den Inferenzschluss benötigt. Erstens muss die Schätzung von regionalen Krankheitshäufigkeiten auf Basis von stationären Behandlungshäufigkeiten möglich sein. Zweitens muss die AOK-Morbidität nach der Adjustierung auf die Gesamtbevölkerung übertragbar sein. Diese Bedingungen wurden im Vorfeld der Studie durch Pretests in Form von multivariaten Analyseverfahren sowie verschiedener Simulationen, in welchen die Selektivität zwischen AOK-Versichertenstamm und Gesamtbevölkerung imitiert wurde, überprüft. Dabei konnte einerseits verifiziert werden, dass die stationären Behandlungshäufigkeiten über einen hinreichend großen Erklärungsgehalt hinsichtlich der regionalen Krankheitshäufigkeiten

verfügen. Darüber hinaus konnte gezeigt werden, dass auch unter Selektivität die Extrapolation der adjustierten Morbiditätsstruktur anhand der stationären Behandlungshäufigkeiten valide Ergebnisse liefert. Als zusätzliches Kontrollmittel um den Erklärungsgehalt der Modelle zu testen wird in Abschn. 3.4 noch eine Bootstrap-basierte Kreuzvalidierung durchgeführt, in welcher Spezifikationen mit unzureichender Vorhersagekraft aussortiert werden.

3.3 Variablenselektion

Im Folgenden wird beschrieben, wie Diagnosen innerhalb der DRG-Statistik ausgewählt werden, welche anschließend dann als Hilfsinformationen zur Morbiditätsadjustierung und zur Schätzung von y_{rd}^c verwendet werden. Die Variablenselektion erfolgt dabei regionsspezifisch, das heißt für jedes $r = 1, \dots, R$ wird ein eigenes Set bestehend aus zehn Diagnosen ausgewählt. Es werden zwei Verfahren der multivariaten Analyse (Feinstein 1996) eingesetzt: Korrelationsanalyse und Hauptkomponentenanalyse. Zunächst wird die Auswahl anhand der Korrelationsanalyse beschrieben. Als Korrelationskriterium wird der Korrelationskoeffizient nach Bravais-Pearson verwendet, welcher gegeben ist durch

$$\rho(y_{rd}^c, x_{rd}^{cj}) = \frac{Cov(y_{rd}^c, x_{rd}^{cj})}{\sqrt{Var(y_{rd}^c) \cdot Var(x_{rd}^{cj})}}, \quad (3)$$

wobei $x_{rd}^{cj} \in x_{rd}^c$. Dabei ist zu beachten, dass für die Bestimmung des Koeffizienten sowohl die Behandlungshäufigkeiten der DRG-Statistik als auch die Krankheitshäufigkeiten der zu schätzenden Krankheit auf Zellebene beobachtet sein müssen. Da aber y_{rd}^c unbekannt ist, lässt sich dies nicht auf den Gesamtbevölkerungsdaten umsetzen. Folglich muss die Korrelationsanalyse ausschließlich auf Basis der AOK-Routinedaten erfolgen. Die hier ausgewählten Diagnosen werden anschließend in der DRG-Statistik für die Gesamtbevölkerung extrahiert. Unter Annahme der Gleichheit der bedingten Erwartungswerte (Gl. 2) ist dies für die angestrebte modellbasierte Inferenz jedoch ausreichend. Die Korrelationsanalyse eignet sich besonders für die Auswahl der Diagnosen zur Morbiditätsadjustierung, wenn die zu schätzende Krankheit starke Prädiktoren innerhalb der DRG-Statistik hat. Dies ist entweder der Fall, wenn die betrachtete Krankheit hauptsächlich stationär behandelt wird⁹, oder wenn sie spezifische Komorbiditäten hat, welche typischer Weise gemeinsam mit ihr auftreten¹⁰.

Ist dies nicht gegeben, so können die benötigten Informationen zur Erklärung der Morbiditätsunterschiede auch als latente Variable in den stationären Behandlungs-

⁹ In diesem Fall existiert eine korrespondierende Hauptdiagnose in der DRG-Statistik, deren zellspezifische Behandlungshäufigkeiten stark mit den zellspezifischen Krankheitshäufigkeiten der betrachteten Diagnose korreliert sind.

¹⁰ In diesem Fall existieren mehrere Haupt- oder Nebendiagnosen in der DRG-Statistik, deren zellspezifische Behandlungshäufigkeiten stark mit den zellspezifischen Krankheitshäufigkeiten der betrachteten Diagnose korreliert sind.

Tab. 1 Zehn häufigsten Diagnosen zur Adjustierung

ICD-3-Code	Bezeichnung	Diagnoseart
E11	Diabetes mellitus, Typ 2	Hauptdiagnose
E11	Diabetes mellitus, Typ 2	Nebendiagnose
E78	Störungen des Lipoproteinstoffwechsels und sonstige Lipidämien	Nebendiagnose
E87	Sonstige Störungen des Wasser- und Elektrolythaushaltes sowie des Säure-Basen-Gleichgewichts	Nebendiagnose
I10	Essentielle (primäre) Hypertonie	Nebendiagnose
I48	Vorhofflimmern und Vorhofflattern	Hauptdiagnose
K29	Gastritis und Duodenitis	Nebendiagnose
M16	Koxarthrose [Arthrose des Hüftgelenkes]	Hauptdiagnose
T81	Komplikationen bei Eingriffen, anderenorts nicht klassifiziert	Nebendiagnose
Z92	Medizinische Behandlung in der Eigenanamnese	Nebendiagnose

häufigkeiten enthalten sein. Für diesen Fall eignet sich dann eine Variablenselektion auf Basis einer Hauptkomponentenanalyse. Dabei wird eine Hauptachsentransformation der regionsspezifischen Matrix der stationären Behandlungshäufigkeiten $\mathbf{X}_r$ derart durchgeführt, dass die entsprechende Kovarianzmatrix der Daten diagonalisiert wird. Dabei werden die zellspezifischen Behandlungshäufigkeiten aller Diagnosen in Linearkombinationen zusammengefasst. Im Zuge der Variablenselektion werden dann nicht mehr zehn tatsächliche Diagnosen aus der DRG-Statistik ausgewählt. Stattdessen werden jene zehn Linearkombinationen isoliert, welche den größten Anteil der Gesamtvarianz von $\mathbf{X}_r$ erklären. Diese stellen dann eine Approximation der gesamten Information dar, welche in der Matrix vorhanden ist. Für die angestrebte Morbiditätsadjustierung und kleinräumige Prävalenzschätzung ist dabei entscheidend, dass $\mathbf{X}_r$ und $\tilde{\mathbf{X}}_r$ exakt gleich rotiert werden. Für mehr technische Details über Hauptkomponentenanalyse, siehe Hastie et al. (2017).

Im Folgenden werden jene zehn Diagnosen aufgeführt, deren stationäre Behandlungshäufigkeiten innerhalb der AOK-Routinedaten am häufigsten auf Basis der Korrelationsanalyse regional in das Set der Prädiktoren aufgenommen wurden. Die Darstellung erfolgt tabellarisch in alphabetischer Reihenfolge nach dem ICD-3-Code. Dabei wird zwischen Haupt- und Nebendiagnose unterschieden (Tab. 1).

3.4 Modelle

Nach der Auswahl der Variablen erfolgt die Beschreibung der Modelle, anhand welcher der bedingte Erwartungswert $E[\tilde{y}_{rd}^c | \tilde{N}_{rd}^c, \tilde{\mathbf{x}}_{rd}^c = (\tilde{x}_{rd}^{c1}, \dots, \tilde{x}_{rd}^{cp})]$ quantifiziert wird. Diese Modelle werden wiederum zur Prädiktion von y_{rd}^c verwendet. Hierbei kommen Verfahren der Regressionsanalyse zum Einsatz. Es werden drei verschiedene lineare Modelltypen verwendet, welche sich hinsichtlich ihrer Effektstruktur und ihrer hierarchischen Aggregationsebene unterscheiden: regionale lineare Modelle, regionale gemischte lineare Modelle und nationale gemischte lineare Modelle. Bei den regionalen Modellen erfolgt die Schätzung der Modellparameter regionsspezifisch, wobei die Anzahl der Zellen (Kreis $\times$ Geschlecht $\times$ Altersklasse),

welche pro Region verfügbar sind, zwischen 72 (kleinste Region) und 1728 (größte Region) liegt. Bei den nationalen Modellen werden alle Zellen Deutschlands simultan berücksichtigt, insgesamt 7218. Die gleichzeitige Verwendung regionaler und nationaler Modelle ist insbesondere vor dem Hintergrund kleiner Regionen, welche nicht viele Zellen enthalten, als besonders wichtig hervorzuheben. In kleinen Regionen kann die Verwendung komplexer Modellspezifikationen gegebenenfalls zu einer vergleichsweise großen Varianz bei der Schätzung der Modellparameter führen. Dies ist auf die reduzierte Anzahl von Freiheitsgraden zurückzuführen. Die nationalen Modelle werden aufgrund der großen Anzahl an Zellen am stabilsten geschätzt, können jedoch weniger regionale Heterogenität abbilden. Hier ist also ein Trade-Off zwischen der Berücksichtigung regionaler Heterogenität und der Stabilität der Prädiktionen zu beachten, welcher die Verwendung beider Modelltypen nahelegt.

Für jeden Modelltyp werden zusätzlich verschiedene Modellspezifikationen berücksichtigt. Durch die Berechnung des bedingten Erwartungswerts auf Basis verschiedener Modelltypen/-spezifikationen auf verschiedenen Ebenen wird mögliche Unsicherheit beim modellbasierten Inferenzschluss antizipiert. Denn a priori ist weder absehbar, für welches Modell Annahme (Gl. 2) am ehesten erfüllt ist, noch ob der entsprechende Modellzusammenhang innerhalb der Gesamtbevölkerung bzw. eines Aggregats derselben homogen ist. Die in Abschn. 2 thematisierten regionalen Gesundheitsdisparitäten legen nahe, dass an dieser Stelle eher mit Heterogenität zu rechnen ist. Deswegen ist die Berücksichtigung verschiedener Modelltypen und Spezifikationen sinnvoll, um zu analysieren, inwiefern sie gleiche Implikationen hinsichtlich der Krankheitshäufigkeiten liefern. Zusätzlich wird parametrisches Bootstrapping (Shang und Cavanaugh 2008) verwendet, um eine Bootstrap-basierte Verteilung für jede Prädiktion einer zellspezifischen Krankheitshäufigkeit der Gesamtbevölkerung zu erzeugen und damit den zugrundeliegenden Inferenzschluss weiter zu verbessern. Darüber hinaus gibt das Bootstrapping einen wichtigen Überblick hinsichtlich der Unsicherheit der Modellparameterschätzung und der Stabilität der Prädiktionen.

3.4.1 Modelltyp 1: Regionale lineare Modelle

Der erste Modelltyp sind gängige multiple Regressionsmodelle, welche für jede Region zellspezifisch auf den AOK-Daten spezifiziert werden. Sie enthalten regionsspezifische Fixed Effects und haben die Form

$$\tilde{y}_{rd}^c = \tilde{x}_{rd}^c \beta_r + e_{rd}^c \forall r, d, c, \quad (4)$$

wobei $\beta_r = (\beta_r^1, \dots, \beta_r^p)^t$ der Vektor der Fixed Effect-Koeffizienten ist und t den Transpositions-Operator notiert. Der Term $e_{rd}^c \sim N(0, \sigma_r^2)$ ist ein stochastischer Fehler mit Erwartungswert 0 und regionalem Varianzparameter σ_r^2 . Innerhalb einer gegebenen Region r kann Gl. 4 in multivariater Form ausgedrückt werden:

$$\tilde{\mathbf{y}}_r = \tilde{\mathbf{X}}_r \beta_r + \mathbf{e}_r \forall r, d, \quad (5)$$

wobei $\tilde{\mathbf{y}}_r = (\tilde{y}_{r1}^1, \dots, \tilde{y}_{rD_r}^C)^t$ den Vektor aller kassenspezifischen Krankheitshäufigkeiten auf Zellebene in Region r darstellt. Analog dazu sind $\tilde{\mathbf{X}}_r = (\tilde{x}_{r1}^1, \dots, \tilde{x}_{rD_r}^C)^t$ und $\mathbf{e}_r \sim MVN(\mathbf{0}, \sigma_r^2 \mathbf{I}_{C \cdot D_r})$, wobei $\mathbf{I}$ die Einheitsmatrix notiert. Die Fixed Effect-Koeffizienten werden durch Ordinary Least Squares (OLS) anhand folgender Formel geschätzt:

$$\hat{\boldsymbol{\beta}}_r = (\tilde{\mathbf{X}}_r^t \tilde{\mathbf{X}}_r)^{-1} \tilde{\mathbf{X}}_r^t \tilde{\mathbf{y}}_r \forall r. \quad (6)$$

Die Prädiktion der zellspezifischen Krankheitshäufigkeit der Gesamtbevölkerung erfolgt dann durch die Kombination der Regressionskoeffizienten, welche auf Basis der AOK-Daten berechnet wurden, und den stationären Behandlungshäufigkeiten der Gesamtbevölkerung aus der DRG-Statistik: $\hat{y}_{rd}^c = \mathbf{x}_{rd}^c \hat{\boldsymbol{\beta}}_r$. Zusätzlich zur direkten Prädiktion auf Basis des Modells (Gl. 5) wird noch eine Prädiktive Verteilung für $\hat{y}_{rd}^c$ auf Basis eines parametrischen Bootstraps erzeugt. Hierzu werden zunächst $\tilde{y}_{rd}^{c*} := \tilde{x}_{rd}^c \hat{\boldsymbol{\beta}}_r$ und die Residualvarianz σ_r^{e2} definiert. Dann wird für alle Zellen in Region r eine Zufallszahl aus der Verteilung $\tilde{y}_{rd}^{cb} \sim N(\tilde{y}_{rd}^{c*}, \sigma_r^{e2})$ gezogen und anschließend $\hat{\boldsymbol{\beta}}_r^*$ anhand von Gl. 6 berechnet, wobei $\tilde{\mathbf{y}}_r$ durch $\tilde{\mathbf{y}}_r^* = (\tilde{y}_{r1}^{1*}, \dots, \tilde{y}_{rD_r}^{C*})^t$ ersetzt wird. Mithilfe von $\hat{\boldsymbol{\beta}}_r^*$ wird dann eine Bootstrap-Prädiktion $\hat{y}_{rd}^{c*} = \mathbf{x}_{rd}^c \hat{\boldsymbol{\beta}}_r^*$ erzeugt. Diese Prozedur wird $B = 1000$ mal wiederholt und man erhält eine Bootstrap-Verteilung für $\hat{y}_{rd}^c$.

3.4.2 Modelltyp 2: Regionale gemischte lineare Modelle

Der zweite Modelltyp sind gemischte lineare Regressionsmodelle, welche für jede Region zellspezifisch auf den AOK-Daten spezifiziert werden. Sie enthalten regionsspezifische Fixed Effects sowie kreisspezifische Random Effects und haben die Form

$$\tilde{y}_{rd} = \tilde{\mathbf{X}}_{rd} \boldsymbol{\beta}_r + \mathbf{Z}_{rd} \mathbf{b}_{rd} + \mathbf{e}_{rd} \forall r, d, \quad (7)$$

mit $\tilde{\mathbf{y}}_{rd} = (\tilde{y}_{rd}^1, \dots, \tilde{y}_{rd}^C)^t$, $\tilde{\mathbf{X}}_{rd} = (\tilde{x}_{rd}^1, \dots, \tilde{x}_{rd}^C)^t$, $\mathbf{e}_{rd} = (e_{rd}^1, \dots, e_{rd}^C)^t$ und $e_{rd}^c \sim N(0, \sigma_r^2)$ wie in Modelltyp 1. Dabei ist $\mathbf{Z}_{rd} = (\mathbf{z}_{rd}^1, \dots, \mathbf{z}_{rd}^q)^t$ die Matrix der Random Effect-Kovariaten. $\mathbf{b}_{rd} \sim MVN(\mathbf{0}, \Psi_r)$ notiert den Vektor der kreisspezifischen Random Effect-Koeffizienten mit regionsspezifischer $(q \times q)$ -Kovarianzmatrix Ψ_r , welche durch einen regionsspezifischen Vektor $\boldsymbol{\psi}_r \in R^q$ parametrisiert wird. Innerhalb einer gegebenen Region r kann Gl. 7 in multivariater Form ausgedrückt werden:

$$\tilde{\mathbf{y}}_r = \tilde{\mathbf{X}}_r \boldsymbol{\beta}_r + \mathbf{Z}_r \mathbf{b}_r + \mathbf{e}_r \forall r, \quad (8)$$

wobei $\tilde{\mathbf{y}}_r = (\tilde{\mathbf{y}}_{r1}, \dots, \tilde{\mathbf{y}}_{rD_r})^t$, $\tilde{\mathbf{X}}_r = (\tilde{\mathbf{X}}_{r1}, \dots, \tilde{\mathbf{X}}_{rD_r})^t$, $\mathbf{Z}_r = \text{diag}(\mathbf{Z}_{r1}, \dots, \mathbf{Z}_{rD_r})$, $\mathbf{b}_r = (\mathbf{b}_{r1}, \dots, \mathbf{b}_{rD_r})^t$ und $\mathbf{e}_r = (\mathbf{e}_{r1}, \dots, \mathbf{e}_{rD_r})^t$. Die Fixed Effect-Koeffizienten werden über Weighted Least Squares geschätzt:

$$\hat{\boldsymbol{\beta}}_r = \left(\tilde{\mathbf{X}}_r^t \hat{\mathbf{V}}_r^{-1} \tilde{\mathbf{X}}_r \right)^{-1} \tilde{\mathbf{X}}_r^t \hat{\mathbf{V}}_r^{-1} \tilde{\mathbf{y}}_r \quad \forall r, \quad (9)$$

wobei $\hat{\mathbf{V}}_r = \hat{\mathbf{V}}_r(\hat{\boldsymbol{\psi}}_r, \hat{\sigma}_r^2)$ ein Schätzer für $\mathbf{V}_r = \mathbf{V}_r(\sigma_r^2, \boldsymbol{\psi}_r) = \mathbf{Z}_r \boldsymbol{\Psi}_r \mathbf{Z}_r^t + \sigma_r^2 \mathbf{I}_{C \cdot D_r}$ ist. Die notwendigen regionalen Varianzparameter σ_r^2 , $\boldsymbol{\psi}_r$ werden mit Restricted Maximum Likelihood berechnet (REML; Rao und Molina 2015). Die Random Effects werden durch den Empirical Best Linear Unbiased Predictor geschätzt (EBLUP; West et al. 2007):

$$\hat{\mathbf{b}}_r = \hat{\boldsymbol{\Psi}}_r \mathbf{Z}_r^t \hat{\mathbf{V}}_r^{-1} (\tilde{\mathbf{y}}_r - \tilde{\mathbf{X}}_r \hat{\boldsymbol{\beta}}_r) \quad \forall r. \quad (10)$$

Die Prädiktion der zellspezifischen Krankheitshäufigkeiten der Gesamtbevölkerung erfolgt dann analog zu Abschn. 3.4.1, jedoch zusätzlich unter Berücksichtigung der geschätzten Random Effects: $\hat{\mathbf{y}}_{rd}^c = \mathbf{x}_{rd}^c \hat{\boldsymbol{\beta}}_r + \mathbf{z}_{rd}^c \hat{\mathbf{b}}_{rd}$. Für den anschließenden parametrischen Bootstrap wird zunächst $\tilde{\mathbf{y}}_{rd}^* := \tilde{\mathbf{X}}_{rd} \hat{\boldsymbol{\beta}}_r + \mathbf{Z}_{rd} \mathbf{b}_{rd}^* + \mathbf{e}_{rd}^*$ definiert, wobei $\hat{\boldsymbol{\beta}}_r$ aus Gl. 9 stammt. Die Varianzparameter werden aus den durch Restricted Maximum Likelihood geschätzten Verteilungen gezogen, nämlich $\mathbf{b}_{rd}^* \sim MVN(\mathbf{0}, \hat{\boldsymbol{\Psi}}_r)$ sowie $\mathbf{e}_{rd}^* \sim MVN(\mathbf{0}, \hat{\sigma}_r^2 \mathbf{I}_C)$. Anschließend werden alle Parameter des Modells (Gl. 7) erneut geschätzt, wobei $\tilde{\mathbf{y}}_{rd}$ durch $\tilde{\mathbf{y}}_{rd}^*$ ersetzt wird. Die neu geschätzten Modellparameter werden dann zur Erzeugung von Bootstrap-Prädiktionen erzeugt: $\hat{\mathbf{y}}_{rd}^{c*} = \mathbf{x}_{rd}^c \hat{\boldsymbol{\beta}}_r^* + \mathbf{z}_{rd}^c \hat{\mathbf{b}}_{rd}^*$. Diese Prozedur wird $B = 1000$ mal wiederholt und man erhält eine Bootstrap-Verteilung für $\hat{\mathbf{y}}_{rd}^c$.

3.4.3 Modelltyp 3: Nationale gemischte lineare Modelle

Der dritte Modelltyp sind gemischte lineare Regressionsmodelle, welche nicht wie in Abschn. 3.4.2 für jede Region einzeln spezifiziert werden, sondern über alle Regionen gleichzeitig. Es kann (zunächst auf Kreisebene) wie folgt formuliert werden:

$$\tilde{\mathbf{y}}_{rd} = \tilde{\mathbf{X}}_{rd} \boldsymbol{\beta} + \mathbf{Z}_{rd} \mathbf{b}_{rd} + \mathbf{e}_{rd} \quad \forall r, d, \quad (11)$$

wobei $\tilde{\mathbf{y}}_{rd}$, $\tilde{\mathbf{X}}_{rd}$, $\mathbf{Z}_{rd}$ wie in Abschn. 3.4.2 definiert. Es gilt jedoch $\mathbf{e}_{rd} = (e_{rd}^1, \dots, e_{rd}^C)^t$ mit $e_{rd}^c \sim N(0, \sigma^2)$, und $\mathbf{b}_{rd} \sim MVN(\mathbf{0}, \boldsymbol{\Psi})$, wobei $\boldsymbol{\Psi}$ durch $\boldsymbol{\psi} \in R^q$ parametrisiert wird. Es ist zu beachten, dass im Gegensatz zu Modelltyp 2 die Fixed Effect-Koeffizienten und die Varianzparameter hier nicht mehr regionalisiert sind. Spezifiziert man das Modell über die Kreise aller R Regionen, so erhält man

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}} \boldsymbol{\beta} + \tilde{\mathbf{Z}} \mathbf{b} + \tilde{\mathbf{e}}, \quad (12)$$

mit $\tilde{\mathbf{y}} = (\tilde{y}_{11}, \dots, \tilde{y}_{RD_R})^t$, $\tilde{\mathbf{X}} = (\tilde{X}_{11}, \dots, \tilde{X}_{RD_R})^t$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_{11}, \dots, \mathbf{Z}_{RD_R})$, $\mathbf{b} = (\mathbf{b}_{11}, \dots, \mathbf{b}_{RD_R})^t$ und $\mathbf{e} = (\mathbf{e}_{11}, \dots, \mathbf{e}_{RD_R})^t$. Definiert man nun $\mathbf{V} = \mathbf{V}(\sigma^2, \boldsymbol{\psi}) = \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^t + \sigma^2\mathbf{I}_{C \cdot \sum_r D_r}$, so werden die Fixed Effect-Koeffizienten über Weighted Least Squares anhand folgender Formel geschätzt:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^t \hat{\mathbf{V}}^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^t \hat{\mathbf{V}}^{-1} \tilde{\mathbf{y}}, \quad (13)$$

wobei $\hat{\mathbf{V}}(\hat{\boldsymbol{\psi}}, \hat{\sigma}^2)$ ein Schätzer für $\mathbf{V}(\sigma^2, \boldsymbol{\psi})$ ist. Die Varianzparameter werden analog zu Abschn. 3.4.2 über REML ermittelt. Die Random Effects werden, wie zuvor, anhand des EBLUP geschätzt:

$$\hat{\mathbf{b}} = \hat{\boldsymbol{\Psi}}\mathbf{Z}^t \hat{\mathbf{V}}^{-1} (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}). \quad (14)$$

Die Prädiktion der zellspezifischen Krankheitshäufigkeit erfolgt durch $\hat{y}_{rd}^c = \mathbf{x}_{rd}^c \hat{\boldsymbol{\beta}} + \mathbf{z}_{rd}^c \hat{\mathbf{b}}_{rd}$. Für den parametrischen Bootstrap wird $\tilde{\mathbf{y}}_{rd}^* := \tilde{\mathbf{X}}_{rd} \hat{\boldsymbol{\beta}} + \mathbf{Z}_{rd} \hat{\mathbf{b}}_{rd}^* + \mathbf{e}_{rd}^*$ definiert, mit $\hat{\boldsymbol{\beta}}$ aus Gl. 13. Die Varianzparameter werden aus ihren geschätzten Verteilungen gezogen: $\mathbf{b}_{rd}^* \sim MVN(\mathbf{0}, \hat{\boldsymbol{\Psi}})$ sowie $\mathbf{e}_{rd}^* \sim MVN(\mathbf{0}, \hat{\sigma}^2 \mathbf{I}_C)$. Die Parameter des Modells (Gl. 11) werden anschließend neu geschätzt, wobei $\tilde{\mathbf{y}}_{rd}$ durch $\tilde{\mathbf{y}}_{rd}^*$ ersetzt wird. Analog zu Abschn. 3.4.2 werden die Modellparameter dann zur Erzeugung von Bootstrap-Prädiktionen erzeugt: $\hat{y}_{rd}^{c*} = \mathbf{x}_{rd}^c \hat{\boldsymbol{\beta}}^* + \mathbf{z}_{rd}^c \hat{\mathbf{b}}_{rd}^*$. Die Prozedur wird $B = 1000$ mal zur Erzeugung der Bootstrap-Verteilung für $\hat{y}_{rd}^c$ wiederholt.

3.5 Plausibelster Schätzwert und plausibler Bereich

Aus Abschn. 3.4 geht hervor, dass nicht jeweils ein einziges Modell zur Prädiktion einer zellspezifischen Krankheitshäufigkeit berücksichtigt wird, sondern ein ganzes Prädiktions-Set. Dieses wird im Folgenden mit P_{rd}^c notiert. Es besteht jeweils aus verschiedenen Modelltypen und Spezifikationen, welche aus Gründen der Übersichtlichkeit mit dem Index $m = 1, \dots, M$ versehen sind. Zusätzlich werden die Ergebnisse des parametrischen Bootstraps – neben der Verwendung als Unsicherheitsmaß – dahingehend berücksichtigt, dass aus jeder Bootstrap-Verteilung für $\hat{y}_{rd}^c$ der Erwartungswert $E(\hat{y}_{rd}^{c*})$ als zusätzliche Prädiktion in das Prädiktions-Set miteingeht. Für jedes Modell existieren also eine direkte Prädiktion sowie eine entsprechende Bootstrap-Prädiktion innerhalb des Sets:

$$P_{rd}^c = \{\hat{y}_{rd}^c(1), E[\hat{y}_{rd}^{c*}(1)], \dots, \hat{y}_{rd}^c(M), E[\hat{y}_{rd}^{c*}(M)]\} \forall r, d, c. \quad (15)$$

Die Idee ist nun, die in Abschn. 3.4 thematisierte Unsicherheit beim modellbasierten Inferenzschluss über die Verteilung der einzelnen Prädiktionen innerhalb von P_{rd}^c zu antizipieren. Auf Basis des Prädiktions-Sets soll ein plausibelster Schätzwert so wie ein plausibler Wertebereich, in welchem die Prädiktion schwanken kann, bestimmt werden. Die Prädiktionen innerhalb von P_{rd}^c bilden eine Schätzverteilung über die unbekannte zellspezifische Krankheitshäufigkeit y_{rd}^c . Ist in der Schätzverteilung ein eindeutiger Massepunkt über einem bestimmten Wertebereich, so wird dieser Wertebereich als plausibel erachtet. Da aber die Modelle regional unterschied-

lich geeignet für die Prädiktion von y_{rd}^c sind oder ggf. ein zu großes Werteintervall aufspannen, muss das Prädiktions-Set um schwache bzw. instabile Modelle bereinigt werden. In Abschn. 1 wurde bereits thematisiert, dass klassische Methoden zur Modellselektion, wie etwa ANOVA (Gelman 2005) oder Informationskriterien (Burnham und Anderson 2004), in dem beschriebenen Setting nicht valide verwendet werden können, da keine klassische Stichprobenziehung vorliegt. Darüber hinaus ist der direkte Vergleich von linearen und gemischten linearen Modellen, welche teilweise auf unterschiedlichen Aggregaten spezifiziert werden, anhand dieser Maße ebenfalls nicht zulässig. Deshalb wird sich bei der Bestimmung des plausibelsten Schätzwerts und des plausiblen Bereichs auf die Betrachtung der prädiktiven Inferenz konzentriert. Diese wird auf Basis einer modifizierten k -fold Cross Validation innerhalb der AOK-Daten evaluiert, welche Ideen des nicht-parametrischen Bootstrapping adaptiert (Efron und Tibshirani 1993, 1997).

Hierfür wird bei den regionalen Modellen (Abschn. 3.4.1 und 3.4.2) die regionsspezifische AOK-Population $\tilde{U}_r$ in $k = 5$ Teilpopulationen $\tilde{U}_r^l$ mit $l = 1, \dots, k$ untergliedert. Bei den nationalen Modellen wird entsprechend die gesamte AOK-Population $\tilde{U}$ in fünf Teilpopulation $\tilde{U}^l$ unterteilt. Aus Notationsgründen wird sich auf die Darstellung bei den regionalen Modellen beschränkt. Für die nationalen Modelle wird jedoch analog verfahren. Es werden die Zellen einer Teilpopulation $\tilde{U}_r^l$ als Prädiktionsdaten bestimmt. Die verbleibenden Zellen in $\tilde{U}_r^{-l}$ sind dann Trainingsdaten. Aus $\tilde{U}_r^{-l}$ werden anschließend B Stichproben der Größe $|\tilde{U}_r^{-l}|$ durch einfache Zufallsauswahl mit Zurücklegen (SRSWR) gezogen. Auf diesen Stichproben werden dann die Modellparameter aller Modelle und Spezifikationen geschätzt. Die geschätzten Modellparameter werden schließlich für die Prädiktion der AOK-spezifischen Krankheitshäufigkeiten $\hat{y}_{rd}^c$ für alle Zellen in $\tilde{U}_r^l$ verwendet. Die Ergebnisse werden hinsichtlich des Relative Root Mean Squared Errors (RRMSE) verglichen. Der gesamte Algorithmus kann wie folgt beschrieben werden (Darstellung in Pseudo-Code):

Für die Teilpopulationen $\tilde{U}_r^1, \dots, \tilde{U}_r^k$

- Für die Modelltypen/-spezifikationen $m = 1, \dots, M$
 1. Teile $\tilde{U}_r$ in Trainingsdaten $\tilde{U}_r^{-l}$ und Prädiktionsdaten $\tilde{U}_r^l$
 2. Ziehe B Stichproben $\tilde{S}_r^1, \dots, \tilde{S}_r^B$ aus $\tilde{U}_r^{-l}$ via SRSWR
 3. Für die Stichproben $\tilde{S}_r^1, \dots, \tilde{S}_r^B$
 - a) Schätze die Modellparameter von m auf Basis von $\tilde{S}_r^b$
 - b) Erzeuge Prädiktionen $\hat{y}_{rd}^c(\tilde{S}_r^b, m)$ für die bekannten $\tilde{y}_{rd}^c$ für alle Zellen in $\tilde{U}_r^l$ unter Verwendung von $\tilde{x}_{rd}^c$
 4. Berechne die Abweichung anhand von

$$RRMSE[\hat{y}_{rd}^c(m)] = \frac{\sqrt{\frac{1}{B} \sum_b (\hat{y}_{rd}^c(\tilde{S}_r^b, m) - \tilde{y}_{rd}^c)^2}}{\tilde{y}_{rd}^c}$$

- Vergleiche die Ergebnisse für die Teilpopulation $\tilde{U}_r^l$

Die Ergebnisse der modifizierten k -fold Cross Validation werden anschließend dafür verwendet, das ursprüngliche Prädiktions-Set P_{rd}^c von schwachen oder instabilen Modelltypen-spezifikationen zu bereinigen. Hierzu werden jene Modelltypen/-spezifikationen inklusive ihrer Bootstrap-Prädiktionen aus P_{rd}^c entfernt, deren RRMSE einen bestimmten Grenzwert τ_{crit} überschreiten. Dieser Grenzwert ist dabei abhängig von der Krankheit, für welche zellspezifische Krankheitshäufigkeiten geschätzt werden sollen. Das finale Prädiktions-Set ist demnach:

$$\hat{P}_{rd}^c = \{\hat{y}_{rd}^c(m), E[\hat{y}_{rd}^{c*}(m)] \mid RRMSE[\hat{y}_{rd}^c(m)] < \tau_{\text{crit}}\} \quad \forall r, d, c. \quad (16)$$

Auf Basis von $\hat{P}_{rd}^c$ können nun ein plausibelster Schätzwert und ein plausibler Bereich für y_{rd}^c ausgewiesen werden. Als plausibelster Schätzwert wird der Median von $\hat{P}_{rd}^c$ verwendet. Als plausibler Bereich wird das 0,05-Quantil bzw. das 0,95-Quantil von $\hat{P}_{rd}^c$ gewählt.

4 Regionale Auswertung für Typ-2-Diabetes

4.1 Ergebnisse

Das Ziel ist nun auf Basis des Versichertenstamms der AOK aus dem Jahr 2016 alters- und geschlechtsreferenzierte Krankheitshäufigkeiten auf Kreisebene für Typ-2-Diabetes zu schätzen. Schauen wir zunächst auf die deutschlandweiten Ergebnisse. Die (nicht-standardisierte) AOK-spezifische Diabetes-Gesamtprevalenz im Jahr 2016 beträgt etwa 12,19 %. Auf Basis der Methodik wird für die Gesamtbevölkerung eine Prävalenz von 8,78 % geschätzt. Das plausible Intervall des Schätzwerts beträgt [8,03 %; 9,57 %]. Somit gelingt eine Anpassung der AOK-spezifischen Morbidität von −34 bis −21 %. In Abb. 2 ist die geschätzte Diabetes-Verteilung für die Bundesrepublik auf Kreisebene dargestellt.

Es werden deutliche Morbiditätsunterschiede zwischen Ost- und Westdeutschland sichtbar. Auf dem Gebiet der östlichen Bundesländer ist eine offenkundig höhere Diabetes-Prävalenz als im Rest des Landes zu verzeichnen. Dies lässt sich zum einen auf den höheren Altersdurchschnitt in diesen Gebieten zurückführen. Bekannt ist, dass das Risiko für Typ-2-Diabetes im Alter deutlich zunimmt. Zum anderen könnte aber auch der teilweise geringere sozioökonomische Status in diesen Gebieten zu der beobachteten Differenz beitragen (Maier 2017). So wurde gezeigt, dass Personen mit niedrigerem sozioökonomischen Status (gemessen anhand von Einkommen und Bildung) ein bis zu zweimal höheres Typ-2-Diabetes-Risiko haben als Personen mit einem hohen sozioökonomischen Status (Maier et al. 2013). Diese Diskrepanz ist sogar persistent wenn man zusätzlich auf Alter, Geschlecht und Body Mass Index bedingt. In Bayern und Baden-Württemberg werden tendenziell die geringsten Diabetes-Häufigkeiten auf Kreisebene gefunden. In den Gebieten des Saarlandes und Rheinland-Pfalz, aber auch in manchen Teilen des Nordrhein-Westfalens sind wiederum leicht erhöhte Häufigkeiten zu finden.

Um die modellbasierte Anpassung der kassenspezifischen Morbiditätsstrukturen durch die Schätzmethode zu verdeutlichen, werden im Folgenden die relati-

Abb. 2 Geschätzte Diabetes-Prävalenz, Kreisebene, nicht-standardisiert

von Krankheitshäufigkeiten des AOK-Versichertenstamms und die entsprechenden Schätzungen für die Gesamtbevölkerung auf Kreisebene verglichen.

Abb. 3 zeigt die Verteilung der relativen Krankheitshäufigkeiten in Prozent auf Kreisebene. Das obere Histogramm bezieht sich auf die AOK, das untere auf die Schätzungen für die Gesamtbevölkerung. Zur besseren Visualisierung wurde in Rot ein Kerndichteschätzer für die Verteilungen mitgegeben. Es wird zunächst deutlich, dass der Massepunkt der Kreisprävalenzen innerhalb der AOK deutlich höher liegt als in der Gesamtbevölkerung. Dies ist vor dem Hintergrund der informativen Kasenzugehörigkeit nachvollziehbar. Wäre der AOK-Versichertenstamm eine zufällige Teilpopulation aus der Gesamtbevölkerung, so wäre der Massepunkt der Verteilungen in einem ähnlichen Wertebereich zu erwarten. Der dargestellte Bias der AOK-Morbidität im Vergleich zur geschätzten Morbidität in der Gesamtbevölkerung ist auch vor dem Hintergrund vergangener Studien plausibel. So haben beispielsweise Hoffmann und Icks (2011, 2012) eine überdurchschnittlich hohe Diabetes-Morbidität in der AOK verglichen mit anderen Krankenkassen gefunden. Darüber hinaus fällt auf, dass die Verteilung innerhalb der AOK deutlich rechtsschief ist als die geschätzte Verteilung in der Gesamtbevölkerung. Dies verdeutlicht, dass die Schätzmethodik nicht nur eine einfache Skalierung der AOK-Prävalenz vornimmt, sondern

Abb. 3 Vergleich der kreisspezifischen Diabetes-Häufigkeiten

die Morbidität spezifischer anpasst. Andernfalls wäre die geschätzte Verteilung lediglich ein um wenige Prozentpunkte verschobenes Abbild der AOK-Verteilung.

Ein weiterer interessanter Aspekt ist zu untersuchen: Ist der Bias im Vergleich zur Gesamtbevölkerung in allen Kohorten des AOK-Versichertenstamms zu finden, und ist er hinsichtlich seines Ausmaßes persistent?

In Abb. 4 ist die bei der Schätzung erfolgte relative Anpassung der AOK-Morbidität auf Kreisebene pro Altersklasse dargestellt. Die relative Anpassung ist prozentual anhand folgender Formel quantifiziert:

$$\text{Relative Anpassung} = (\hat{p}_{rd}^c / \tilde{p}_{rd}^c - 1) \cdot 100\%, \hat{p}_{rd}^c = \hat{y}_{rd}^c / N_{rd}^c, \tilde{p}_{rd}^c = \tilde{y}_{rd}^c / \tilde{N}_{rd}^c.$$

Die Verteilung der relativen Anpassung ist in Rot dargestellt, der Mittelwert der Anpassung pro Altersklasse auf Kreisebene ist mit Blau gekennzeichnet. Es wird deutlich, dass der Bias je nach Altersklasse sehr unterschiedlich ausfällt. Während er in den jüngeren Kohorten stabil bei 20% und mehr ist, so nimmt er bei den älteren Kohorten systematisch ab. In der Kohorte AK 8 ist der Anpassungsfaktor bereits nahe Null. In der ältesten Kohorte AK 9 ist die Diabetes-Häufigkeit des AOK-Versichertenstamms sogar leicht unterhalb der Gesamtbevölkerung, da die durchschnittliche Anpassung größer Null ist. Die sichtbare Variation des Anpassungsfaktors über die Altersklassen hinweg signalisiert ebenfalls, dass die präsentierte Schätzmethode die AOK-Morbidität nicht einfach skaliert, sondern differenziert anpasst.

Abb. 4 Altersspezifische Morbiditätsanpassung

4.2 Externe Plausibilisierung der Ergebnisse

Die vorliegende Untersuchung ermittelte eine geschätzte Prävalenz des Typ-2-Diabetes in Deutschland von 8,78 %, mit einem plausiblen Intervall des Schätzwerts von 8,03 bis 9,57 %. Diese Größenordnung deckt sich mit anderen, vergleichbaren Untersuchungen, die im Folgenden genauer dargestellt werden. Da davon auszugehen ist, dass Typ-2-Diabetiker einen Anteil von über 95 % aller Diabetiker ausmachen (Jacobs und Rathmann 2018), werden sowohl Zahlen zum Diabetes insgesamt als auch zum Typ-2-Diabetes berücksichtigt. Allerdings sind methodische Unterschiede in der Datenerhebung (Untersuchungssurveys, Befragungen oder Routinedaten der gesetzlichen Krankenversicherung), dem betrachteten Zeitraum sowie den zugrundeliegenden Definitionen des Diabetes bei der Interpretation zu berücksichtigen.

In Tamayo et al. (2016) werden dokumentierte Prävalenzen für Versicherte der gesetzlichen Krankenkassen auf Basis der im DIMDI vorliegenden Routinedaten für die Datenjahre 2009/2010 über Diabetes-Diagnosen bestimmt. Die Studie weist für Diabetes insgesamt eine Prävalenz von 9,7 % im Jahr 2009 und 9,9 % im Jahr 2010 aus – entsprechend Alters- und Geschlechtsstandardisiert auf die deutsche Bevölkerung zum 31.12.2007. Die Typ-2-Diabetes Prävalenz wird in 2009 mit 6,9 % und in 2010 mit 7,1 % angegeben – wobei jedoch viele Diabetes-Fälle mit nicht eindeutiger Diagnosedokumentation hier nicht einbezogen sind. Das Zentralinstitut der kassenärztlichen Versorgung (Zi) hat in der Arbeit von Goffrier et al. (2017) auf Basis von vertragsärztlichen Abrechnungsdaten der kollektiven Versorgung eine dokumentierte Prävalenz des Diabetes von 9,8 % und des Typ-2-Diabetes von 9,5 % bei gesetzlich Krankenversicherten ausgewiesen – entsprechend Alters- und Geschlechtsstandardisiert auf die Mitglieder- und Versichertenstatistik mit Stand zum 01.07.2015. Die Ergebnisse aus der GEDA-Befragungsstudie 2014/2015 liefern eine 12-Monats-Prävalenz bei Erwachsenen ab 18 Jahren von 7,7 % für Diabetes (Typ 1 und Typ 2) (Heidemann et al. 2017). Die Angaben zum Diabetes sind Selbstangaben der befragten Personen, wobei Schwangerschaftsdiabetes ausgeschlossen wurde. Es kann möglicherweise eine Unterschätzung des Typ-2-Diabetes vorliegen, da in Befragungsstudien ältere Bevölkerungsgruppen ab 80 Jahren schwerer zugänglich sind (Gaertner et al. 2016). In Übersichtsarbeiten von Jacobs und Rathmann (2018) werden Schätzungen zur Häufigkeit des Typ-2-Diabetes mit 7 bis 8 % angegeben, in der Bestandsaufnahme zur Diabetes-Surveillance (Heidemann und Scheidt-Nave 2017) ist eine Spannbreite von 7,2 bis 9,9 % Diabetes-Prävalenz (Typ 1 und Typ 2) genannt. Zudem sind im Zeitverlauf steigende Prävalenzen berichtet worden (Heidemann und Schneidt-Nave 2017; Goffrier et al. 2017). Insgesamt liegen die Ergebnisse der vorliegenden Untersuchung mit einer Prävalenz des Typ-2-Diabetes von 8,78 % und dem plausiblen Intervall [8,03 %; 9,57 %] im Jahr 2016 in der zu erwartenden Größenordnung anderen Untersuchungen.

Für die Plausibilisierung der regionalen Unterschiede konnten die Arbeiten von Goffrier et al. (2017) und Heidemann et al. (2017) herangezogen werden. Die dort beschriebenen Muster mit höheren Diabetes-Prävalenzen insbesondere im Osten Deutschlands finden sich auch in der vorliegenden Untersuchung wieder. Insgesamt kann somit die Prävalenzschätzung im Vergleich zu externen Quellen unter Berück-

sichtigung der methodischen Unterschiede der verschiedenen Arbeiten als plausibel erachtet werden.

4.3 Diskussion der Methodik

Eine offensichtliche Stärke des Verfahrens ist die stabile Quantifizierung kleinräumiger Krankheitshäufigkeiten durch modellbasierte Inferenz. Klassische bzw. designbasierte Schätzverfahren auf Basis von Stichprobendaten könnten vergleichbar hochreferenzierte Schätzwerte lediglich mit einer prohibitiv hohen Varianz ausweisen. In der Literatur werden für solche Datensituationen oftmals Verfahren der Small-Area-Schätzung (Münnich et al. 2013; Rao und Molina 2015) vorgeschlagen. Siehe Tzavidis et al. (2018) für eine detaillierte Beschreibung der Anwendung entsprechender Methoden. Doch aufgrund der Selektivität des Versichertenstammes der AOK können zumindest gängige Small-Area-Verfahren, wie etwa das Fay-Herriot Modell (Fay und Herriot 1979), nicht auf die Routinedaten angewendet werden, da kein unverzerrter designbasierter Schätzer hergeleitet werden kann. Der von uns präsentierte Ansatz enthält zwar Elemente der Small-Area-Schätzung, insbesondere bei der Verwendung der gemischten linearen Modelle, unterscheidet sich jedoch hinsichtlich des Inferenzschlusses von den gängigen Verfahren. Die vorgestellte Verfahrensweise hinsichtlich Cross Validation, Bootstrap und der Konstruktion des plausiblen Werteintervalls sichern zusätzlich zumindest teilweise gegen Verletzungen der Modellannahmen ab. Der Inferenzschluss erfolgt nicht unter Berücksichtigung einer einzelnen Modellspezifikation, sondern einer Bandbreite von Spezifikationen, für welche im Einzelfall Abweichungen von den Annahmen erlaubt sind.

Ein weiterer Vorteil besteht in der großen Kosten- und Zeitersparnis, welche die Methodik bietet. Im Rahmen des Verfahrens werden ausschließlich Routine- und Registerdaten¹¹ verwendet, welche automatisch erhoben werden und nicht in Stichproben erfragt werden müssen. Zusätzlich verfügen Krankenkassen in der Regel über deutlich umfangreichere gesundheitsbezogene Informationen hinsichtlich ihres Versichertenstamms, als sie klassischerweise in einer Gesundheitserhebung abgefragt werden. Da es sich dabei um Routinedaten handelt, ist zudem eine Betrachtung über die Zeit hinweg möglich. Für eine solche Analyse müssten, wollte man ähnliche Informationen anhand von Stichproben erhalten, ansonsten aufwändige Panel-Erhebungen durchgeführt werden. Die Methodik bietet folglich große Potenziale für die Bewertung der gesundheitlichen Lage in den Regionen Deutschlands. Basierend hierauf ist eine problemorientierte und zielgerichtete Ausgestaltung der Versorgung in den Regionen Deutschlands möglich.

Das Wissenschaftliche Institut der AOK (WiDO) ist zusammen mit dem Robert Koch-Institut und dem Umweltbundesamt Partner im Projekt BURDEN 2020¹² (Rommel et al. 2018). Hierbei soll eine umfassende Krankheitslastberechnung nach internationalem Vorbild der Global Burden of Disease-Studie (GBD 2017) aufgebaut werden. Die Krankheitslast einer Population wird gemessen als verloren gegangene

¹¹ Die DRG-Statistik ist eine gesetzlich verpflichtende deutschlandweite Vollerhebung der in den Krankenhäusern behandelten Diagnosen.

¹² https://www.rki.de/DE/Content/Institut/Public_Health/BURDEN_2020.html.

gesunde Lebensjahre auf Grund von Krankheit oder vorzeitigem Versterben. Um ein vollständiges Bild der gesundheitlichen Lage in den Regionen Deutschlands zu erhalten, wird eine möglichst umfassende Abbildung aller Krankheiten angestrebt. Zusätzlich wird eine Attribution von Risikofaktoren für die verschiedenen Krankheiten berechnet, so dass regionale Unterschiede in der Krankheitslast auch hinsichtlich möglicher Ursachen bewertet werden können. Ein wesentlicher Baustein in der Krankheitslastberechnung ist die Datengrundlage zu den Prävalenzen verschiedener Erkrankungen und deren Folgezustände in der Bevölkerung auf regionaler Ebene. Die hier beschriebene Methodik kann daher auch genutzt werden, um Prävalenzen zur regionalen Krankheitslastberechnung für das BURDEN 2020 Projekt nutzbar zu machen. Dies erlaubt es, die Krankheitslastberechnung auch für Krankheiten zu ergänzen, die über die derzeit bestehenden Informationen der Gesundheitsberichterstattung hinausgehen.

Zu beachten ist jedoch, dass die Methodik primär für große Krankenkassen mit geographisch flächendeckend angesiedelten Versichertenstämmen geeignet ist. In Abschn. 3.3 wurde angenommen, dass in jedem Aggregat, für welches Krankheitshäufigkeiten geschätzt werden sollen, Mitglieder jener Krankenkasse sind, deren Versichertenstamm als Datengrundlage dient. Obwohl die vorgestellte Methodik es prinzipiell erlaubt, auch Krankheitshäufigkeiten für Aggregate zu schätzen, in denen keine Mitglieder vorhanden sind¹³, so ist dies nur in einem gewissen Rahmen unter zusätzlichen Annahmen möglich. Es muss an der Stelle angenommen werden, dass die systematischen Morbiditätsunterschiede zwischen jenen Gebieten, in denen Mitglieder sind, und jenen, in denen keine Mitglieder sind, sich (vollständig) aus den Hilfsinformationen erklären lassen. Zusätzlich muss der funktionale Zusammenhang zwischen Zellprävalenz und den Prädiktoren für diese Gebiete hinreichend gleich sein. Wenn eine Schätzung für eine beispielhafte Region A auf Basis eines Versichertenstammes, dessen Mitglieder ausschließlich in Region B sind, durchgeführt wird, dann ist die Strukturübertragung von B nach A mit entsprechend viel Unsicherheit verbunden. Der Inferenzschluss ist dann sowohl statistisch als auch inhaltlich schwer zu legitimieren. Deshalb ist für die valide Erfassung regionaler Gesundheitsdisparitäten die Beobachtung der lokalen Morbiditätsstrukturen von zentraler Bedeutung. Da die AOK mit etwa 26,5 Mio. Versicherten in 2018 die größte gesetzliche Krankenkassenart mit Versicherten in allen Regionen Deutschlands darstellt, ist diese Problematik im Zuge des vorliegenden Beitrags jedoch vernachlässigbar.

Eine weitere Einschränkung liegt in der betrachteten Krankheit, für welche kleinräumige Prävalenzen geschätzt werden sollen. In Abschn. 3.3 wurde thematisiert, dass sich die Methodik insbesondere für Krankheiten eignet, die entweder in der Regel stationär behandelt werden, oder typische Komorbiditäten vorweisen, die zur Dokumentation entsprechender Diagnosen im stationären Behandlungssetting führen. In dem vorliegenden Beitrag wurde mit Diabetes Mellitus Typ 2 eine Krankheit gewählt, welche typische Komorbiditäten (etwa mit anderen kardiovaskulären Erkrankungen) aufweist. Diese Komorbiditäten sind vor allem in den Nebendiagnosen

¹³ Modellbasierte Schätzverfahren erlauben theoretisch vollsynthetische Prädiktionen für Aggregate, in denen der Stichprobenumfang Null ist. Dies wird häufig in der Small-Area-Literatur unter dem Begriff „non-sampled area“ thematisiert.

der DRG-Statistik detailliert enthalten, was die Schätzung auf Basis der Methodik begünstigt. Sollen nun aber Schätzungen für eine Diagnose vorgenommen werden, auf welche beides nicht zutrifft, so kann ggf. die benötigte Information zur Morbiditätsadjustierung als latente Variable durch eine Hauptkomponentenanalyse extrahiert werden. Sollte auch dies die systematischen Gesundheitsdisparitäten zwischen den Populationen nicht hinreichend gut erklären, so muss auf weitere Datenquellen zurückgegriffen werden, die für die Verteilung der zu schätzenden Krankheit einen größeren Erklärungsgehalt haben.

So stehen zum Beispiel die Referenzdaten der Arzneimittelversorgung sowie Daten zu Heilmittelbehandlungen zur Verfügung. Aus der GKV-Arzneimittel-Schnellinformation (GAMSi) sind Informationen zu abgerechneten Arzneimitteln auf Produktebene (PZN) nach Menge (DDD¹⁴, Verordnungsanzahl) und Umsatz zugänglich¹⁵. Über die PZN können darüber hinaus alle Informationen aus der Arzneimittelstammdaten-Klassifikation¹⁶ (beispielsweise Wirkstoff, ATC-Code etc.) abgerufen werden. Die kleinste regionale Ebene ist hier die KV-Region. Weder ein Personenbezug noch eine Zuordnung nach Alter und Geschlecht sind möglich. Für die Heilmittelversorgung, d. h. Maßnahmen der Physiotherapie, Ergotherapie, Sprachtherapie oder medizinischen Fußpflege, können über das GKV-Heilmittel-Informationssystem (GKV-HIS) die Mengen der Heilmittelverordnungen von GKV-Versicherten nach Altersgruppen ausgewertet werden¹⁷. Der kleinste regionale Bezug ist auch hier die KV-Region. Ob ein Modell über einen hinreichend großen Erklärungsgehalt verfügt, kann im Zuge des Bootstraps und der Cross Validation festgestellt werden. Sie liefern bereits bei der Schätzung wichtige Hinweise darüber, ob die verwendeten Modelle über genügend Erklärungsgehalt verfügen. Sind die Modelle innerhalb dieser Schritte instabil, so müssen weitere Anpassungen vorgenommen werden.

Zusätzlich muss angemerkt werden, dass die für die Schätzung von Diabetes mellitus Typ 2 spezifizierten Modelle in Abschn. 3.4 sich nicht uneingeschränkt zur Modellierung seltener Krankheiten verwenden lassen. Die (gemischten) linearen Modelle eignen sich nur dann für seltene Krankheiten, wenn diese sehr starken Prädiktoren in den Behandlungshäufigkeiten der DRG-Statistik aufweisen. Ein Beispiel hierfür wären akute Herzinfarkte. Diese sind zwar im Vergleich zu Diabetes mellitus Typ 2 sehr selten, werden jedoch nahezu ausschließlich stationär behandelt. Soll nun aber eine seltene Krankheit modelliert werden, die keine starken Prädiktoren in den stationären Daten hat, wie etwa Multiple Sklerose, so muss auf andere Modelltypen zurückgegriffen werden. Im Zuge des Projekts *Gesundheitsatlas* des Wissenschaftlichen Instituts der AOK (WiDO) werden aktuell sowohl generalisierte lineare Modelltypen als auch aufwändigere Schätzverfahren mit nicht-linearen Nebenbedingungen sowie die Einbindung weiterer Datenquellen untersucht, um die Methodik auch für derartige Konstellationen zu erweitern.

¹⁴ Definierte durchschnittliche Tagesdosen.

¹⁵ <https://www.gkv-gamsi.de/>.

¹⁶ <https://www.wido.de/stammdaten.html>.

¹⁷ <https://www.gkv-heilmittel.de>.

5 Fazit

Im Rahmen des Projekts *Gesundheitsatlas* des Wissenschaftlichen Instituts der AOK (WiO) wurde in Kooperation mit der Universität Trier ein Verfahren entwickelt, welches die alters- und geschlechtsreferenzierte Schätzung regionaler Krankheitshäufigkeiten auf Basis von anonymisierten Routinedaten der AOK ermöglicht. Da die Kassenzugehörigkeit einer Person informativ für ihr Risiko hinsichtlich verschiedener Krankheiten sein kann, muss bei der Schätzung eine Selektivitätsverzerrung korrigiert werden. Die Methodik wurde für den vorliegenden Beitrag auf Diabetes Mellitus Typ 2 angewendet. Im Zuge des AOK-internen Projekts *Gesundheitsatlas* wird sie jedoch für die Schätzung zahlreicher weiterer Krankheiten benutzt. Es kann gezeigt werden, dass der präsentierte Ansatz eine gute Möglichkeit zur Schätzung kleinräumiger Gesundheitsdisparitäten darstellt. Eine Bewertung von regionalen Disparitäten in Bezug auf das Gesundheitsrisiko für die Bevölkerung Deutschlands wird somit ermöglicht und bietet den Akteuren im Gesundheitswesen eine weitere wesentliche Grundlage, Risiken genauer zu analysieren und Versorgung gezielt zu steuern. Die Daten können zudem in die BURDEN2020 Studie einfließen und stellen somit einen Bestandteil in der Krankheitslastberechnung für Deutschland dar. Die Nutzung der Routinedaten einer Kassenart bietet damit neben den Daten der Datentransparenzverordnung (DaTraV) eine weitere aufwandsarme Möglichkeit, regionale Krankheitshäufigkeiten unter Nutzung von selektiven aber umfassenderen Daten zu ermitteln.

Zukünftige Forschung wird unter anderem auf der Erstellung von Zeitreihen liegen. Die aktuellen Schätzungen belaufen sich ausschließlich auf Querschnittsdaten. Mittelfristig soll die Methodik jedoch dahingehend erweitert werden, dass die kleinräumige Entwicklung von Morbiditätsstrukturen auf Basis der AOK-Routinedaten im Zeitverlauf sichtbar gemacht werden können. Wie bereits in Abschn. 4.2 angedeutet, wird auch weitere Forschung zur Modellierung seltener Krankheiten erfolgen. Hierfür werden zum einen generalisierte lineare Modelle in Betracht gezogen, wie etwa Zero-Inflated Poisson oder negativ-binomiale Modelle. Zum anderen werden die bestehenden Verfahren zur Modellparameterschätzung erweitert, so dass sie für die spezielle Datensituation bei seltenen Krankheiten verwendbar sind.

Danksagung Die Autoren danken zwei anonymen Reviewern für ihre aufmerksame Durchsicht. Damit konnte eine noch bessere Lesbar- und Verständlichkeit des Artikels erreicht werden.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- Altgeld T (2008) Gesundheitsfördernde Settingarbeit als Schlüsselstrategie zur Reduktion von gesundheitlichen Ungleichheiten. In: Bauer U, Bittlingmayer UH, Richter M (Hrsg) Health Inequalities. Determinanten und Mechanismen gesundheitlicher Ungleichheit. VS, Wiesbaden, S 511–529
- Badura B, Ducki A, Schröder H, Klose J, Meyer M (2018) Fehlzeiten-Report 2018 – Sinn erleben – Arbeit und Gesundheit. Springer, Berlin, Heidelberg

- BBSR (2015) Laufende Raumbeobachtung – Raumabgrenzungen – Siedlungsstrukturelle Kreistypen. https://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Kreistypen4/kreistypen_node.html. Zugriffen: 30. Aug. 2018
- Behrens J, Noll H-H (2006) Gesundheit, Lebensstile und Ungleichheit: Indikatoren und Analysen im Kontext politischer Diskussionen. In: Rehberg K-S, Deutsche Gesellschaft für Soziologie (DGS) (Hrsg) Soziale Ungleichheit, kulturelle Unterschiede: Verhandlungen des 32. Kongresses der Deutschen Gesellschaft für Soziologie in München. Campus, Frankfurt am Main, S 973–975
- van den Berg M, Heijink R, Zwakhals L, Verkleij H, Westert G (2010) Health care performance in the Netherlands: Easy access, varying quality, rising costs. *Eurohealth* (Lond) 16(4):27–29
- Bertelsmann Stiftung (2015) Regionale Unterschiede zu (Fach-)Arztdichte. <https://faktencheck-gesundheit.de/de/faktenchecks/aerzte/dichte/ergebnis-ueberblick/>. Zugriffen: 11. März 2019
- Böcken J, Grote-Westrick M, Schwenk U (2015) Internationale Erfahrungen mit regionalen Variationen. Was kann Deutschland vom Ausland lernen? *G+g Wissenschaft* (ggw) 15(2):7–15
- Brecht JG (1990) Aufbau einer Gesundheitsberichterstattung in der Bundesrepublik Deutschland. *Informationsd Soziale Indik* 4:8–10. <https://doi.org/10.3238/arztebl.2014.0215>
- Bundesministerium für Gesundheit (2016) KM6 Statistik. <https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/mitglieder-und-versicherte.html>. Zugriffen: 28. Sept. 2018
- Burgard JP, Krause J, Münnich R (2019) Adjusting selection bias in German health insurance records for regional prevalence estimation. In Revision.
- Burgdorf F, Sundmacher L (2014) Potenziell vermeidbare Krankenhausfälle in Deutschland. Analyse von Einflussfaktoren auf die Raten ambulant-sensitiver Krankenhauseinweisungen. *Dtsch Arztebl Int* 111(13):215–223. <https://doi.org/10.3238/arztebl.2014.0215>
- Burnham KP, Anderson DR (2004) Multimodel inference: Understanding AIC and BIC in model selection. *Sociol Methods Res* 33:261–304
- Busch K (2013) Die Arbeitsunfähigkeit in der Statistik der GKV. In: Badura B, Ducki A, Schröder H, Klose J, Meyer M (Hrsg) Fehlzeiten-Report 2013. Verdammt zum Erfolg – die süchtige Arbeitsgesellschaft? Springer, Berlin, Heidelberg, S 447–458
- Cheung KL, Ten Klooster PM, Smit C, de Vries H, Pieterse ME (2017) The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health* 17(1):276. <https://doi.org/10.1186/s12889-017-4189-8>
- Deutsches Institut für Medizinische Dokumentation und Information (DIMDI) (2016a) Versorgungsdaten (Datentransparenz) – Evaluationsbericht 07/2013–02/2016 Teil 1, Version 01-01-000. <https://www.dimdi.de/static/de/downloads/deutsch/versorgungsdaten-evaluationsbericht2016-teil1.pdf>
- Deutsches Institut für Medizinische Dokumentation und Information (DIMDI) (2016b) Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme. 10. Revision. German Modification Version 2019. <https://www.dimdi.de/static/de/klassifikationen/icd/icd-10-gm/kode-suche/htmlgm2019/>. Zugriffen: 3. März 2019
- Deutsches Institut für Medizinische Dokumentation und Information (DIMDI) (2018) Strukturelle Beschreibung der DatraV-Daten. <https://www.dimdi.de/dynamic/de/weitere-fachdienste/versorgungsdaten/datensatzbeschreibung/>
- Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall/CRC, Boca Raton
- Efron B, Tibshirani R (1997) Improvements on cross-validation: The .632+ bootstrap method. *J Am Stat Assoc* 92(432):548–560
- Fay ER, Herriot RA (1979) Estimates of income for small places: An application of James-Stein procedures to census data. *J Am Stat Assoc* 74:269–277
- Feinstein AR (1996) Multivariable analysis. Yale University Press, New Haven
- Frankenfeld P (2005) Regionale Disparitäten. In: Akademie für Raumforschung und Landesplanung (Hrsg) Handwörterbuch der Raumordnung. Akademie für Raumforschung und Landesplanung, Hannover, S 185–190
- Gaertner et al (2016) Baseline participation in a health examination survey of the population 65 years and older: Who is missed and why? *BMC Geriatr* 16(1):21. <https://doi.org/10.1186/s12877-016-0185-6>
- GBD 2017 Disease and Injury Incidence and Prevalence Collaborators (2018) Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017. [https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736\(18\)32279-7.pdf](https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(18)32279-7.pdf). Zugriffen: 28. Nov. 2018 (A systematic analysis for the Global Burden of Disease Study 2017)
- Gelman A (2005) Analysis of variance? Why it is more important than ever. *Ann Stat* 33:1–53

- Gerste B, Drohan D, Günster C (2016) Diagnosehäufigkeit und Inanspruchnahme von Gesundheitsleistungen. In: Klauber J, Günster C, Gerste B, Robra B-P, Schmacke N (Hrsg) Versorgungs-Report 2015/2016. Schattauer, Stuttgart, S 391–445
- Glaeske G, Rebscher H, Willich S (2010) Versorgungsforschung: Auf gesetzlicher Grundlage systematisch ausbauen. Dtsch Arztebl 107(26):1295–1297
- Goffrier B, Bätzing J, Holstige J (2017) Entwicklung der administrativen Prävalenz des Diabetes mellitus von 2009 bis 2015. Monit Versorgungsforsch 10(5):46–49. <https://doi.org/10.24945/mvf.05.17.1866-0533.2040>
- Günster C, Klauber J, Robra BP, Schmacke N, Schmucker C (Hrsg) (2019) Versorgungs-Report. Schwerpunkt: Früherkennung. Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin
- Hastie T, Tibshirani R, Friedman J (2017) The elements of statistical learning. Data mining, inference, and prediction, 2. Aufl. Springer, Heidelberg, Berlin, New York
- Hecken J (2018) Beschluss des Gemeinsamen Bundesausschusses zur Abnahme des Endberichts „Gutachten zur Weiterentwicklung der Bedarfsplanung i. S. d. §§ 99 ff. SGB V zur Sicherung der vertragsärztlichen Versorgung“. https://www.g-ba.de/downloads/39-261-3493/2018-09-20_Endbericht-Gutachten-Weiterentwicklung-Bedarfsplanung.pdf
- Heidemann C, Scheidt-Nave C (2017) Prävalenz, Inzidenz und Mortalität von Diabetes mellitus bei Erwachsenen in Deutschland – Bestandsaufnahme zur Diabetes-Surveillance. J Health Monit 2(3):105–129. <https://doi.org/10.17886/RKI-GBE-2017-050>
- Heidemann C, Kuhnert K, Born S, Scheidt-Nave C (2017) Prävalenz, Inzidenz und Mortalität von Diabetes mellitus bei Erwachsenen in Deutschland – 12 Monats Prävalenz des bekannten Diabetes mellitus in Deutschland. J Health Monit. 2(1): 48–56. <https://doi.org/10.17886/RKI-GBE-2017-008>
- Hoffmann F, Icks A (2011) Diabetes “epidemic” in Germany? A critical look at health insurance data sources. Exp Clin Endocrinol Diabetes 120(7):410–415. <https://doi.org/10.1055/s-0032-1306331>
- Hoffmann F, Icks A (2012) Unterschiede in der Versichertenstruktur von Krankenkassen und deren Auswirkungen für die Versorgungsforschung: Ergebnisse des Bertelsmann-Gesundheitsmonitors. Gesundheitswesen 74(5):291–297. <https://doi.org/10.1055/s-0031-1275711>
- Hoffmann F, Koller D (2017) Different regions, differently insured populations? Sociodemographic and health-related differences between insurance funds. Gesundheitswesen 79(1):1–9. <https://doi.org/10.1055/s-0035-1564074>
- Huber J, Lampert T, Mielck A (2010) Unterschiede bei Gesundheitsrisiken, Morbidität und gesundheitlicher Versorgung zwischen Kindern GKV- bzw. PKV-versicherter Eltern: Ergebnisse aus dem Kinder- und Jugendgesundheitsurvey (KiGGS). Das Gesundheitswes 74(10):627–638. <https://doi.org/10.1055/s-0031-1301268>
- Jacobs E, Rathmann W (2018) Deutscher Gesundheitsbericht Diabetes 2018. https://www.diabetesde.org/system/files/documents/gesundheitsbericht_2018.pdf (Epidemiologie des Diabetes in Deutschland, S 10–22)
- Jacobs K, Kuhlmeier A, Greß S, Klauber J, Schwinger A (2018) Pflege-Report Schwerpunkt: Qualität in der Pflege. Springer, Heidelberg, Berlin, New York
- Jong JD (2008) Explaining medical practice variation. Social organization and institutional mechanisms. Nivel, Utrecht
- Kauhl B, Schweikart J, Krafft T, Keste A, Moskwyn M (2016) Do the risk factors for type 2 diabetes mellitus vary by location? A spatial analysis of health insurance claims in Northeastern Germany using kernel density estimation and geographically weighted regression. Int J Health Geogr 15(1):38. <https://doi.org/10.1186/s12942-016-0068-2>
- Klauber J, Geraedts M, Friedrich J, Wasem J (2012) Krankenhaus-Report 2012. Schwerpunkt: Regionalität. Schattauer, Stuttgart, New York
- Klauber J, Geraedts M, Friedrich J, Wasem J (Hrsg) (2019) Krankenhaus-Report 2019. Schwerpunkt: Das digitale Krankenhaus. Springer, Berlin
- Klauber J, Günster C, Gerste B, Robra B-P, Schmacke N (Hrsg) (2016) Versorgungs-Report 2015/2016. Schwerpunkt: Kinder und Jugendliche. Schattauer, Stuttgart
- Klemperer D, Robra B-P (2014) Regionale Versorgungsunterschiede. John Wennberg – Wegbereiter einer patientenorientierten Medizin. Dtsch Arztebl 111(4):118–120
- Kriwy P, Mielck A (2006) Versicherte der gesetzlichen Krankenversicherung (GKV) und der privaten Krankenversicherung (PKV): Unterschiede in Morbidität und Gesundheitsverhalten. Gesundheitswesen 68(5):281–288. <https://doi.org/10.1055/s-2006-926779>
- Kroll LE, Lampert T (2011) Regionalisierung von Gesundheitsindikatoren. Ergebnisse aus der GEDA-Studie 2009. Bundesgesundheitsblatt 55:129–140

- Kroll LE, Müters S, Lampert T (2016) Arbeitslosigkeit und ihre Auswirkungen auf die Gesundheit. Ein Überblick zum Forschungsstand und zu aktuellen Daten der Studien GEDA 2010 und GEDA 2012. Bundesgesundheitsblatt 59(2):228–237. <https://doi.org/10.1007/s00103-015-2282-7>
- Kroll LE, Schumann M, Hoebel J, Lampert T (2017) Regionale Unterschiede in der Gesundheit – Entwicklung eines sozioökonomischen Deprivationsindex für Deutschland. J Health Monit 2(2):103–120. <https://doi.org/10.17886/RKI-GBE-2017-035>
- Kunst A (2009) Herausforderungen bei der Beschreibung sozialer Ungleichheit. Methodische Überlegungen. In: Richter M, Hurrelmann K (Hrsg) Gesundheitliche Ungleichheit. Grundlagen, Probleme, Perspektiven. VS, Wiesbaden, S 367–383
- Lampert T, Richter M, Schneider S, Spallek J, Dragano N (2016) Soziale Ungleichheit und Gesundheit – Stand und Perspektiven der sozioepidemiologischen Forschung in Deutschland. Bundesgesundheitsblatt 59(02):153–165. <https://doi.org/10.1007/s00103-015-2275-6>
- Lange A, Neubauer S, Schulenburg J-M, Zeidler J (2014) Grundlagen und Methoden von GKV-Routinedatenstudien. http://diskussionspapiere.wiwi.uni-hannover.de/pdf_bib/dp-534.pdf
- Maier W (2017) Indizes multipler Deprivation zur Analyse regionaler Gesundheitsunterschiede in Deutschland. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 60(12):1403–1412. <https://doi.org/10.1007/s00103-017-2646-2>
- Maier W, Fairburn J, Mielck A (2012) Regionale Deprivation und Mortalität in Bayern. Entwicklung eines Index Multipler Deprivation* auf Gemeindeebene. Gesundheitswesen 74(7):416–425. <https://doi.org/10.1055/s-0031-1280846>
- Maier W, Holle R, Hunger M, Peters A, Meisinger C, Greiser KH, Kluttig A, Volzke H, Schipf S, Moebus S, Bokhof B, Berger K, Mueller G, Rathmann W, Tamayo T, Mielck A (2013) The impact of regional deprivation and individual socio-economic status on the prevalence of Type 2 diabetes in Germany. A pooled analysis of five population-based studies. Diabet Med 30(3):78–86. <https://doi.org/10.1111/dme.12062>
- Mangiapane S (2014) Lernen aus regionalen Unterschieden. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 57(2):215–223. <https://doi.org/10.1007/s00103-013-1889-9>
- Mielck A (2008) Regionale Unterschiede bei Gesundheit und gesundheitlicher Versorgung: Weiterentwicklung der theoretischen und methodischen Ansätze. In: Baue U, Bittlingmayer UH (Hrsg) Health inequalities. Determinanten und Mechanismen gesundheitlicher Ungleichheit. VS, Wiesbaden, S 167–190
- Münich R, Burgard JP, Vogt M (2013) Small Area-Statistik: Methoden und Anwendungen. AStA Wirtsch Sozialstat Arch 6(1):149–191. <https://doi.org/10.1007/s11943-013-0126-1>
- Nimptsch U, Bestmann A, Erhart M, Dudey S, Marx Y, Saam J, Schopen M, Schröder H, Swart E (2014) Zugang zu Routinedaten. In: Swart E, Ihle P, Gothe H, Matusiewicz D (Hrsg) Routinedaten im Gesundheitswesen – Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Huber, Bern, S 270–290
- Nolting H-D (2015) Regionale Variationen. Stand und Nutzungsperspektiven in Deutschland. G+g Wissenschaft (ggw) 15(2):16–22
- Ohlmeier C, Frick J, Prütz F, Lampert T, Ziese T, Mikolajczyk R, Garbe R (2014) Nutzungsmöglichkeiten von Routinedaten der Gesetzlichen Krankenversicherung in der Gesundheitsberichterstattung des Bundes. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 57(4):464–472. <https://doi.org/10.1007/s00103-013-1912-1>
- Pestalozza C (2007) Das Recht auf Gesundheit. Verfassungsrechtliche Dimensionen. Bundesgesundheitsblatt 50(9):1113–1118. <https://doi.org/10.1007/s00103-007-0310-y>
- PKV (2016) Zahlen und Fakten PKV Zahlenportal. <https://www.pkv-zahlenportal.de/>. Zugriffen: 28. Sept. 2018
- Radespiel-Tröger M, Meyer M (2017) Kleinräumige Analyse vermuteter Krebscluster anhand von Krebsregisterdaten. Bundesgesundheitsblatt 60(12):1393–1402. <https://doi.org/10.1007/s00103-017-2645-3>
- Rao JNK, Molina I (2015) Small area estimation, 2. Aufl. Wiley series in survey methodology. John Wiley & Sons, Inc, Hoboken
- Reiners H (2008) Was war wann? Gesundheitsgesetzgebung seit 1883. In: Stabsbereich Politik (Hrsg) Die Gesundheitskasse – 125 Jahre AOK – Eine Festschrift zum Jubiläum. AOK Bundesverband, Bonn
- Robert Koch-Institut (Hrsg) (2014) Daten und Fakten: Ergebnisse der Studie „Gesundheit in Deutschland aktuell 2012“. Beiträge zur Gesundheitsberichterstattung des Bundes. RKI, Berlin
- Robert Koch-Institut (Hrsg) (2017a) Methodische Studie zur Durchführung von Mixed-Mode-Befragungen zur Gesundheit von Kindern und Jugendlichen (Pilotstudie KiGGS Welle 2). Gesundheitsberichterstattung des Bundes gemeinsam getragen von RKI und Destatis. RKI, Berlin

- Robert Koch-Institut (Hrsg) (2017b) Krebs in Deutschland für 2013/2014. Robert Koch-Institut, Berlin <https://doi.org/10.17886/rkipubl-2017-007>
- Rommel A (2016) Nutzung des RKI Gesundheitsmonitorings für Forschungszwecke. https://webcache.googleusercontent.com/search?q=cache:vkoKKsw4s9oJ:https://www.tmf-ev.de/DesktopModules/Bring2mind/DMX/Download.aspx%3FMethod%3Dattachment%26Command%3DCore_Download%26EntryId%3D29391%26PortalId%3D0+&cd=1&hl=de&ct=clnk&gl=de. Zugriffen: 11. Aug. 2017
- Rommel A, von der Lippe E, Plaß D, Wengler A, Anton A, Schmidt C, Schüssel K, Brückner G, Schröder H, Porst M, Leddin J, Tobollik M, Baumert J, Scheidt-Nave C, Ziese T (2018) BURDEN 2020—Burden of disease in Germany at the national and regional level. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 61(9):1159–1166
- Schmidt C, Bätzing-Feigenbaum J, Bestmann A, Brinks R, Dreß J, Goffrier B, Hagen B, Laux G, Pollmanns J, Schröder H, Stahl T, Baumert J, Du Y, Gabry L, Heidemann C, Paprott R, Scheidt-Nave C, Teti A, Ziese T (2017) Integration von Sekundärdaten in die Nationale Diabetes-Surveillance. Hintergrund, Ziele und Ergebnisse des Sekundärdaten-Workshops am Robert Koch-Institut. Bundesgesundheitsblatt 60(6):656–661. <https://doi.org/10.1007/s00103-017-2552-7>
- Schwabe U, Paffrath D, Ludwig W-D, Klauber J (2018) Arzneiverordnungs-Report 2018. Springer, Heidelberg, Berlin, New York
- Shang J, Cavanaugh JE (2008) An assumption for the development of bootstrap variants for the Akaike information criterion in mixed models. Stat Probab Lett 78:1422–1429
- Statistisches Bundesamt (2017) Gesundheit – Fallpauschalenbezogene Krankenhausstatistik (DRG-Statistik). Diagnosen, Prozeduren, Fallpauschalen und Case Mix der vollstationären Patientinnen und Patienten in Krankenhäusern 2016. https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Krankenhaeuser/FallpauschalenKrankenhaus2120640167004.pdf?__blob=publicationFile. Zugriffen: 29. Aug. 2018
- Statistisches Bundesamt (2018) Bevölkerungsstand. <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Bevoelkerungsstand/Bevoelkerungsstand.html>. Zugriffen: 29. Aug. 2018
- Storz-Pfennig P (2012) Geografische Variationen in der stationären Versorgung: Internationale Erfahrungen. In: Klauber J, Geraedts M, Friedrich J, Wasem J (Hrsg) Krankenhaus-Report 2012. Schattauer, Stuttgart, S 33–43
- Sundmacher L, Busse R (2012) Der Einfluss der Ärztedichte auf ambulant-sensitive Krankenhausfälle. In: Klauber J, Geraedts M, Friedrich J, Wasem J (Hrsg) Krankenhaus-Report 2012. Schattauer, Stuttgart, S 183–204
- Swart E, Deh U, Robra BP (2008) Die Nutzung der GKV-Daten für die kleinräumige Analyse und Steuerung der stationären Versorgung. Bundesgesundheitsblatt 51(10):1183–1192. <https://doi.org/10.1007/s00103-008-0653-z>
- Swart E, Ihle P, Gothe H, Matusiewicz D (Hrsg) (2014) Routinedaten im Gesundheitswesen – Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Huber, Bern
- Tamayo T, Brinks R, Hoyer A, Kuß O, Rathmann W (2016) Prävalenz und Inzidenz von Diabetes mellitus in Deutschland – Auswertung von 65 Millionen Versichertendaten der gesetzlichen Krankenkassen aus den Jahren 2009 und 2010. Dtsch Arztebl 113(11):177–182. <https://doi.org/10.3238/arztebl.2016.0177>
- Tennstedt F (1997) Peitsche und Zuckerbrot oder ein Reich mit Zuckerbrot?. Der Deutsche Weg zum Wohlfahrtsstaat 1871–1881. Z Sozialreform 43(2):88–101
- Tzavidis N, Zhang L-C, Luna A, Schmid T, Rojas-Perilla N (2018) From start to finish: A framework for the production of small area official statistics. J Royal Stat Soc Ser A 181:927–979. <https://doi.org/10.1111/rssa.12364>
- Völzke H, Neuhauser H, Moebus S, Baumert J, Berger K, Stang A, Ellert U, Werner A, Döring A (2006) Rauchen: Regionale Unterschiede in Deutschland. Dtsch Arztebl 103(42):2784–2790
- Waltersbacher A (2018) Heilmittelbericht 2018. Wissenschaftliches Instituts der AOK (WiDO), Berlin
- Wennberg JE (2010) Tracking medicine. A researcher's quest to understand health care. Oxford University Press, New York
- Wennberg JE (2014) Forty years of unwarranted variation—And still counting. Health Policy (New York) 114(1–2):1–2. <https://doi.org/10.1016/j.healthpol.2013.11.010>
- Wennberg J, Gittelsohn A (1973) Small area variations in health care delivery. Science 182(4117):1102–1108. <https://doi.org/10.1126/science.182.4117.1102>
- West BT, Welch KB, Galecki AT (2007) Linear mixed models: A practical guide using statistical software. Chapman & Hall, New York

- Zapf W (1974) Soziale Indikatoren. Konzepte und Forschungsansätze Band I und II. Campus, Frankfurt am Main
- Zapf W (1975) Soziale Indikatoren. Konzepte und Forschungsansätze Band III. Campus, Frankfurt am Main
- Zapf W (1977) Einleitung in das SPES-Indikatorensystem. In: Zapf W (Hrsg) Lebensbedingungen in der Bundesrepublik Deutschland. Sozialer Wandel und Wohlfahrtsentwicklung. Campus, Frankfurt am Main, S 11–95

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.