

Stichprobendaten von Versicherten der gesetzlichen Krankenversicherung nach § 268 SGB V

I. Hintergrund

Seit 1994 gibt es den gesetzlich festgelegten Risikostrukturausgleich (RSA), an dem bis auf die Landwirtschaftlichen Krankenkassen alle Kassen der gesetzlichen Krankenversicherung (GKV) beteiligt sind.¹ Dieses Finanzausgleichsverfahren ermöglicht es, die durch unterschiedliche Versichertenstrukturen entstehenden Ungleichverteilungen in den Beitragseinnahmen und Leistungsausgaben der Krankenkassen auszugleichen. Hintergrund dieser politischen Diskussion ist die seit 1996 bestehende freie Kassenwahl, die zu einer erhöhten „Risikoselektion“ – einem Wettbewerb der Krankenkassen um „positive“ Risiken, also junge und gesunde Mitglieder – führte. Die Berechnung des Beitragsbedarfs erfolgt über die Merkmale Geschlecht, Alter und Nicht-Bezug einer Erwerbsminderungsrente. Ist der Beitragsbedarf größer als die Finanzkraft, besteht für die jeweilige Krankenkasse ein Ausgleichsanspruch, im umgekehrten Fall eine Ausgleichsverpflichtung.

Um eine höhere Zielgenauigkeit dieser finanziellen Zuweisungen zu gewährleisten und die Risikoselektion noch stärker zu minimieren, ohne den Krankenkassen die Motivation für ein wirtschaftliches Verhalten zu nehmen, ist am 10. Dezember 2001 das Gesetz zur Reform des Risikostrukturausgleichs in der Gesetzlichen Krankenversicherung in Kraft getreten. Ab 2007 soll dann „die Morbidität der Versicherten auf der Grundlage von Diagnosen, Diagnosegruppen, Indikationen, Indikationsgruppen, medizinischen Leistungen oder Kombinationen dieser Merkmale“ unmittelbar berücksichtigt werden.² Im Zusammenhang mit der Einführung des morbiditätsorientierten RSA wurde eine wissenschaftliche Untersuchung „zur Auswahl geeigneter Gruppenbildungen, Gewichtungsfaktoren und Klassifikationsmerkmale für einen direkt morbiditätsorientierten Risikostrukturausgleich“ durchgeführt, zu der das Bundesministerium für Gesundheit gesetzlich verpflichtet und das Institut für Gesundheits- und Sozialforschung (IGES) mit der Durchführung beauftragt wurde.³

Das Statistische Bundesamt und das Bundesministerium für Gesundheit sind darin übereingekommen, die aus dieser Untersuchung gewonnenen umfangreichen Daten über ambulante und stationäre Diagnosen, verordnete Arzneimittel, Arbeitsunfähigkeit und Kranken-

¹ Vgl. Sozialgesetzbuch Fünftes Buch - Gesetzliche Krankenversicherung - (2005), § 266.

² Vgl. Sozialgesetzbuch Fünftes Buch - Gesetzliche Krankenversicherung - (2005), §268.

³ Reschke, P./ Sehlen, S./ Schiffhorst, G./ Schröder, W. F./ Lauterbach, K. W. (2004) Klassifikationsmodelle für Versicherte im Risikostrukturausgleich – Endbericht, S. 5.

geld von etwa 2 Millionen Versicherten der gesetzlichen Krankenversicherung zunächst für die Berichtsperiode 2002 statistisch aufzubereiten, um diese der Wissenschaft für weitere Analysen zur Verfügung zu stellen. Im Folgenden werden die Datengrundlage, das Vorgehen der Aufbereitung sowie Möglichkeiten der weiteren Verwendung erläutert.

II. Datengrundlage

Die Daten von mehr als 2 Millionen Versicherten der gesetzlichen Krankenversicherung wurden speziell für die Analyse relevanter Modelle im Rahmen des morbiditätsorientierten Risikostrukturausgleichs für die Berichtsjahre 2001 und 2002 erfasst.⁴ An der umfangreichen Erhebung waren etwa 350 Krankenkassen, die 23 Kassenärztlichen Vereinigungen (KVen) und ihre Verbände, das Bundesversicherungsamt (BVA), die Bundesversicherungsanstalt für Angestellte (BfA) sowie das Deutsche Institut für medizinische Dokumentation und Information (DIMDI) beteiligt.⁵ Die Datenübermittlung erfolgte von den einzelnen Kassenärztlichen Vereinigungen an die Bundesverbände der Krankenkassen. Die Stichprobenauswahl basiert auf einer Drei-Prozent-Zufallsstichprobe in Form einer Geburtstagsstichprobe. Jede versicherte Person, die am 11. eines beliebigen Monats eines beliebigen Jahres geboren wurde, mindestens an einem Tag im Erhebungszeitraum in einer der beteiligten gesetzlichen Krankenkassen⁶ versichert war und nicht als Auftragsfall geführt wurde, ist in die Analyse einbezogen worden.⁷

Aufgrund der umfangreichen Daten über ambulante und stationäre Behandlungsfälle, Arbeitsunfähigkeit, Krankengeldbezug und Arzneimittelverordnungen von etwa 2 Millionen Versicherten der GKV aus dem Jahr 2002 ist ein hohes Analysepotenzial gegeben. Die demografischen Angaben wie Alter, Geschlecht und Region (Ost vs. West) ermöglichen einerseits eine sehr detaillierte Betrachtung sowie das Aufzeigen differenter Patientenstrukturen und anfallender Gesundheitskosten.

Das Datenmaterial setzt sich zusammen aus verschiedenen Datensätzen. Über eine Leitdatei mit den Merkmalen Versichertennummer (systemfrei), Geburtsjahr und Geschlecht sind die Daten verknüpfbar.

Als **Jahresdaten der Versicherten (SA_411)** stehen für das Berichtsjahr 2002 folgende Informationen für eine Analyse zur Verfügung:

- Versichertennummer (systemfrei)
- Berichtsperiode
- Ist die Person im Berichtsjahr verstorben

⁴ Vgl. Reschke et al. (2004) Klassifikationsmodelle für Versicherte im Risikostrukturausgleich – Endbericht, S.140 ff.

⁵ Zu zahnmedizinischen Fällen liegen keine Daten vor, da mit den Kassenzahnärztlichen Vereinigungen kein Datenaustausch erfolgt ist.

⁶ Folgende Krankenkassen sind in das Datenmaterial einbezogen: Allgemeine Ortskrankenkassen (AOK), Betriebskrankenkassen (BKK), Bundesknappschaft (BKN), Angestellten-Ersatzkassen (EAN), Arbeiter-Ersatzkassen (EAR), Innungskrankenkassen (IKK), See-Krankenkasse (SEE). Krankenkassen mit weniger als 30 Stichprobenversicherten oder weniger als 1000 Versichertenjahren sind nicht enthalten.

⁷ weiterführend Reschke et al. (2004), S. A-15 ff.

- Ausgaben für sonstige Leistungen (wie Heil- und Hilfsmittel, häusliche Pflege, Mutterschafts- bzw. Sterbegeld usw.)
- Dialyseleistung in der Berichtsperiode
- Rechtskreis (Ost/West)
- Anzahl der Versichertentage in der Berichtsperiode

Informationen über ambulante Behandlungsfälle sind insgesamt in den drei Satzarten SA_311 bis Sa_313 zu finden. Sie beinhalten sämtliche Fälle im ambulanten Sektor des Jahres 2002 mit Angaben über:

Ambulante Abrechnungen (SA_311)

- Versichertennummer (systemfrei)
- Fallzähler
- Zeitraum (Quartal und Jahr)
- Fachgruppe des Arztes
- Rechnerischer Ausgabenbetrag je Abrechnungsfall
- Sachkosten
- Leistungsquartal
- Punktzahlsumme je Abrechnungsfall

Diagnosen der ambulanten Behandlung (SA_312)

- Versichertennummer (systemfrei)
- Fallzähler
- Zeitraum (Quartal und Jahr)
- Diagnosenzähler
- Diagnose

Gebührenpositionen der ambulanten Behandlung (SA_313)

- Versichertennummer (systemfrei)
- Fallzähler
- Zeitraum (Quartal und Jahr)
- Gebührenposition der ambulanten Behandlung (nach EBM)
- Anzahl der Gebührenpositionen

Die Daten der stationären Versorgung gliedern sich in 4 Satzarten (SA_412-SA_415). Darin sind sämtliche stationäre Aufenthalte eines Stichprobenversicherten zu finden, wenn entweder das Aufnahme- oder das Entlassungsdatum des Falles in den Zeitraum des Jahres 2002 fällt bzw. wenn das Aufnahmedatum vor 2002 liegt und der Fall am Ende der Berichtsperiode noch nicht abgeschlossen ist. Merkmale des stationären Sektors sind:

Stationäre Behandlungsfälle (SA_412)

- Versichertennummer (systemfrei)
- Fallnummer
- Berichtsperiode
- Aufnahme­datum (Monat/ Jahr)
- Entlassungsdatum (Monat/ Jahr)
- Anzahl der Pfl­egetage
- Ausgaben
- Leistungsart

Diagnosen der stationären Behandlung (SA_413)

- Versichertennummer (systemfrei)
- Fallnummer
- Berichtsperiode
- Diagnosen­zähler
- Diagnose
- Art der Diagnosen­stellung

Entgelte der stationären Behandlung (SA_414)

- Versichertennummer (systemfrei)
- Fallnummer
- Berichtsperiode
- Positionsnummer
- Entgeltart (nach Entgeltartenschlüssel gemäß BPfIV)
- Entgeltanzahl

Im Rahmen der Krankenhausbehandlung durchgeführte Operationen (SA_415)

- Versichertennummer (systemfrei)
- Fallnummer
- Berichtsperiode
- Operations­zähler
- Operationsdatum (Monat/ Jahr)
- Operation (nach amtlichem OPS)
- zweiter Operations­schlüssel
- dritter Operations­schlüssel

Arzneimitteldaten (SA_416) bilden einen eigenen Bereich. Für jedes im Jahr 2002 von der Apotheke abgerechnete Medikament, Verband- oder Hilfsmittel für einen Versicherten ist ein Datensatz mit folgenden Merkmalen enthalten:

- Versichertennummer (systemfrei)
- Berichtsperiode
- Verordnungsdatum (Monat/ Jahr)

- Positionsnummer des Medikaments auf dem Rezept
- Kassenausgaben für das Rezept
- Pharmazentralnummer⁸
- Anzahl der verschriebenen Packungen pro Medikament

Auch die **Daten über Arbeitsunfähigkeit und Krankengeldbezug (SA_417)** sind in einem weiteren Zusatzfile zu finden. Er beinhaltet folgende Informationen, sofern der Versicherte zu Beginn der Arbeitsunfähigkeit Krankengeldanspruchsberechtigt war:

- Versichertennummer (systemfrei)
- Fallnummer
- Berichtsperiode
- Beginndatum der Arbeitsunfähigkeit (Monat/ Jahr)
- Endedatum der Arbeitsunfähigkeit (Monat/ Jahr)
- Anzahl der Arbeitsunfähigkeitstage
- Anzahl der Krankengeldbezugstage
- Krankengeldausgaben
- Diagnose 1
- Diagnose 2
- Diagnose3

III. Weiteres Vorgehen

Es ist geplant, die aufbereiteten Daten im Laufe des Jahres 2007 auf unterschiedlichen Zugangswegen der wissenschaftlichen Öffentlichkeit zur Verfügung zu stellen. Seit April 2007 steht zunächst der Zugang der kontrollierten Datenfernverarbeitung zur Verfügung. Auf diesem Weg wird die Nutzung des Informationspotenzials der Einzeldaten mittels einer selbst entwickelten Programmsyntax ermöglicht. Ein weiteres Ziel ist es, die in diesem Umfang bisher einzigartige Stichprobe von Versicherten der GKV der unabhängigen wissenschaftlichen Forschung in Form eines Scientific-Use-Files (SUF) zur Verfügung zu stellen.⁹ Das methodische Vorgehen der Anonymisierung von Einzeldaten umfasst dabei verschiedene Maßnahmen¹⁰; ein genaues Konzept für die hier vorliegende Stichprobe wird derzeit erstellt.

⁸ Für dieses Merkmal liegt kein Schlüsselverzeichnis vor. Der Bezug ist unter bestimmten Voraussetzungen und gegen Gebühr bei der IFA GmbH möglich.

⁹ Scientific Use Files sind faktisch anonymisierte Mikrodatensätze, deren Reidentifikation nur mit einem unverhältnismäßig hohen Aufwand möglich ist. Siehe hierzu Zühlke, S./ Zwick, M./ Scharnhorst, S./ Wende, T. (2003) Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. In: Statistisches Bundesamt (Hrsg.), *Wirtschaft und Statistik* 10, S. 906-911.

¹⁰ Siehe hierzu Müller, W./ Blien, U./ Knoche, P./ Wirth, H. (1991) Die faktische Anonymität von Mikrodaten, Band 19 der Schriftenreihe Forum der Bundesstatistik; siehe auch Merz, J./ Zwick, M. u. a. (2004) *MIKAS-Mikroanalysen und amtliche Statistik*, Band 1 der Schriftenreihe Statistik und Wissenschaft des Statistischen Bundesamtes.