

FDZ-Arbeitspapier  
Nr. 22

Patricia Lugert



**STATISTISCHE ÄMTER**  
DES BUNDES UND DER LÄNDER  
FORSCHUNGSDATENZENTREN

Stichprobendaten von  
Versicherten der  
gesetzlichen Kranken-  
versicherung -  
Grundlage und Struktur  
des Datenmaterials

2007

FDZ-Arbeitspapier  
Nr. 22

Patricia Lugert



**STATISTISCHE ÄMTER**  
**DES BUNDES UND DER LÄNDER**  
FORSCHUNGSDATENZENTREN

Stichprobendaten von  
Versicherten der  
gesetzlichen Kranken-  
versicherung -  
Grundlage und Struktur  
des Datenmaterials

2007

**Herausgeber:** Statistische Ämter des Bundes und der Länder  
**Herstellung:** Statistisches Bundesamt

**Fachliche Informationen**

zu dieser Veröffentlichung:

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 0611 75-4220  
Fax: 0611 72-4000  
forschungsdatenzentrum@destatis.de

Erscheinungsfolge: unregelmäßig  
Erschienen im Dezember 2007

**Informationen zum Datenangebot:**

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 0611 75-4220  
Fax: 0611 72-4000  
forschungsdatenzentrum@destatis.de

Forschungsdatenzentrum der  
Statistischen Landesämter  
– Geschäftsstelle –  
Tel.: 0211 9449-2876  
Fax: 0211 9449-8087  
forschungsdatenzentrum@lds.nrw.de

Diese Publikation wird kostenlos als **PDF-Datei** zum Download unter [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de) angeboten.

© Statistisches Bundesamt, Wiesbaden 2007  
(im Auftrag der Herausbergemeinschaft)

Für nichtgewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung. Alle übrigen Rechte bleiben vorbehalten.

**Bei den enthaltenen statistischen Angaben handelt es sich um eigene Arbeitsergebnisse des genannten Autors im Zusammenhang mit der Nutzung der bereitgestellten Daten der Forschungsdatenzentren. Es handelt sich hierbei ausdrücklich nicht um Ergebnisse der Statistischen Ämter des Bundes und der Länder.**

# Stichprobendaten von Versicherten der gesetzlichen Krankenversicherung – Grundlage und Struktur des Datenmaterials

Patricia Lugert<sup>1</sup>

## 1 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.<sup>2</sup> 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 - führt. Die Berechnung des Beitragsbedarfs der Krankenkassen, also des jeweiligen ausgleichenden Geldbetrages, 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 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 auch „die Morbidität der Versicherten auf der Grundlage von Diagnosen, Diagnosegruppen, Indikationen, Indikationsgruppen, medizinischen Leistungen oder Kombinationen dieser Merkmale“ unmittelbar berücksichtigt werden.<sup>3</sup> 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, mit deren Bearbeitung das Institut für Gesundheits- und

---

<sup>1</sup> FDZ des Statistischen Bundesamtes, E-Mail: patricia.lugert@destatis.de.

Der Beitrag wurde in ähnlicher Form zur Veröffentlichung in den DRV-Schriften Band 55/2007 eingereicht.

<sup>2</sup> Vgl. SGB V (2006), § 266.

<sup>3</sup> Vgl. SGB V (2006), §268.

Sozialforschung (IGES) vom Bundesministerium für Gesundheit beauftragt wurde.<sup>4</sup> 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 Krankengeld 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. Die Datengrundlage sowie ihre Struktur werden im Folgenden erläutert.

## 2 Datengrundlage

Die vorliegenden Stichprobendaten von Mitgliedern und Mitversicherten der gesetzlichen Krankenversicherung wurden speziell für die Analyse relevanter Modelle im Rahmen des morbiditätsorientierten Risikostrukturausgleichs für die Berichtsjahre 2001 und 2002 zusammengetragen.<sup>5</sup> 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.<sup>6</sup> Die Datenübermittlung erfolgte von den einzelnen Kassenärztlichen Vereinigungen an die Bundesverbände der Krankenkassen. Etwa 90 Prozent der deutschen Bevölkerung sind gesetzlich krankenversichert und bilden folglich die Grundgesamtheit. Die Stichprobenauswahl basiert auf einer Drei-Prozent-Zufallsstichprobe in Form einer Geburtstagsstichprobe. Jede (mit)versicherte Person, die am 11. eines beliebigen Monats eines beliebigen Jahres geboren wurde, mindestens an einem Tag im Erhebungszeitraum in einer der beteiligten gesetzlichen Krankenkassen<sup>7</sup> versichert war und nicht als Auftragsfall geführt wurde, ist in die Analyse einbezogen worden.<sup>8</sup> Das vom Forschungsdatenzentrum des Statistischen Bundesamtes aufbereitete Datenmaterial enthält Angaben von insgesamt 2.301.015 Personen, die im Jahr 2002 gesetzlich krankenversichert waren. Es setzt sich zusammen aus verschiedenen Satzarten mit fallbezogenen Informationen über die Jahresdaten der Versicherten, ambulante und stationäre Behandlungen, Operationen, Arzneimitteldaten sowie Arbeitsunfähigkeit und Krankengeld (vgl. Abbildung 1).

---

<sup>4</sup> Reschke u.a. (2005) S. 5.

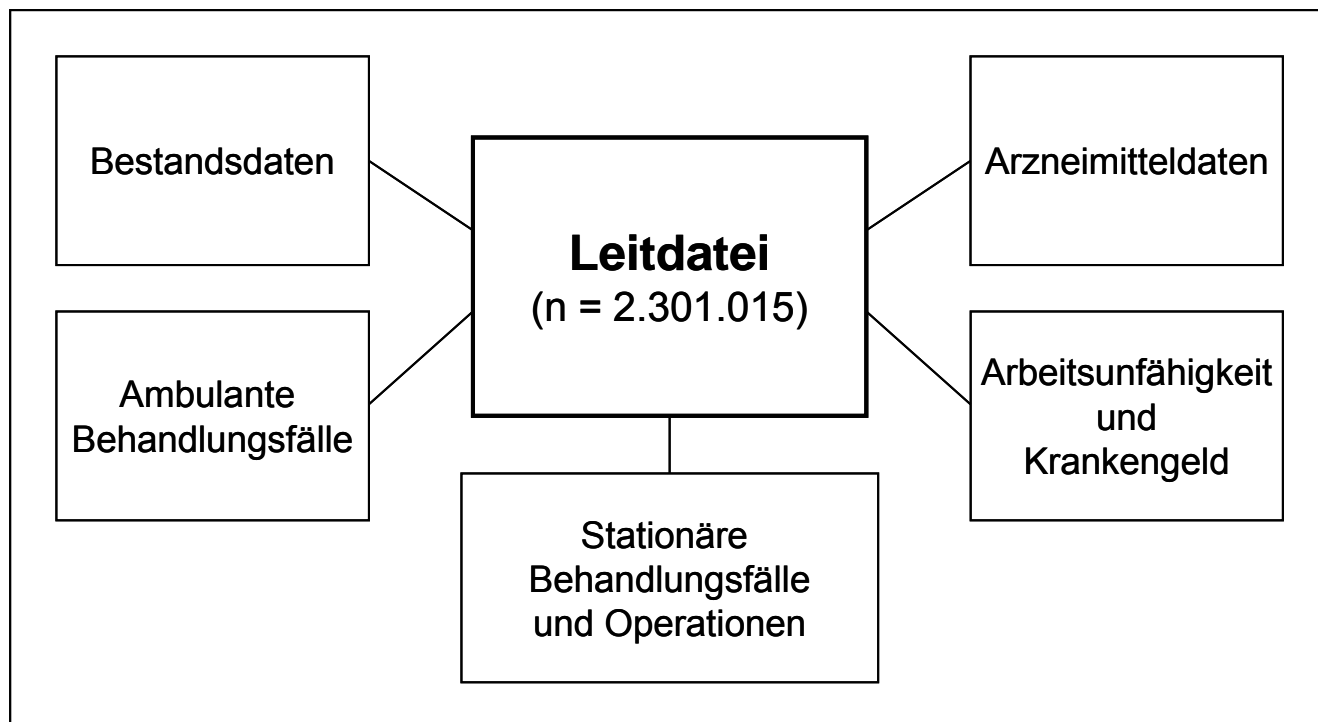
<sup>5</sup> Vgl. Reschke u.a. (2005), S.140 ff.

<sup>6</sup> Zu zahnmedizinischen Fällen liegen keine Daten vor, da mit den Kassenzahnärztlichen Vereinigungen kein Datenaustausch erfolgt ist.

<sup>7</sup> 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.

<sup>8</sup> Zum Stichprobenplan siehe Reschke u.a. (2005), S. A-16 ff (Anhang 2).

Abbildung 1: Struktur des aufbereiteten Datenmaterials im Statistischen Bundesamt



Quelle: Stichprobendaten 2002 nach § 268 SGB V; eigene Berechnungen

### 3 Analyse der Versichertenstruktur

Im Folgenden werden die Satzarten im Einzelnen näher erläutert. Ziel ist es, in rein deskriptiver Form die Stichprobe vorzustellen und die Struktur des Datenmaterials sowie das Analysepotenzial aufzuzeigen.

#### 3.1 Leitdatei

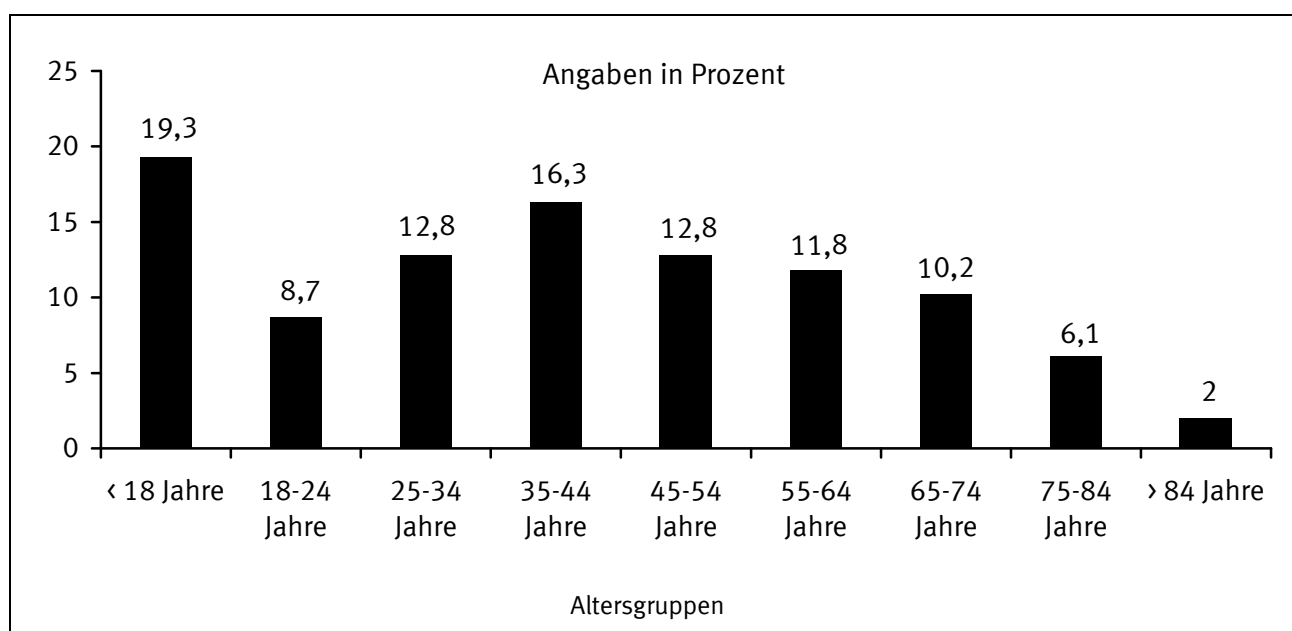
Die Leitdatei bildet den zentralen Datensatz, mit dem die anderen Dateien verknüpft werden. Sie enthält die pseudonymisierte Versichertennummer als Verknüpfungsmerkmal sowie Informationen über Geburtsjahr und Geschlecht der Versicherten.

Die Altersverteilung der Stichprobenversicherten entspricht in etwa der realen Bevölkerungsverteilung. Den größten Teil nimmt die Gruppe der Kinder und Jugendlichen unter 18 Jahren ein (vgl. Abbildung 2). Sie ist insgesamt mit fast einem Fünftel (19 Prozent) im Datensatz enthalten. Mit 16 Prozent bilden die 35- bis 44-Jährigen die zweitstärkste Gruppe in der Altersverteilung. Zu gleichen Teilen (je 13 Prozent) sind Menschen zwischen 25 und 34 Jahren sowie zwischen 45 und 54 Jahren vertreten. 9 Prozent der Versicherten zählen zu den jungen Menschen zwischen 18 und 24 Jahren. Die Zahl der Personen im regulären Rentenalter (ab 65 Jahre) umfasst insgesamt 18 Prozent; ein Zehntel von ihnen befindet sich im Alter von 65 bis 74

Jahren, weitere 6 Prozent der älteren Versicherten sind zwischen 75 und 84 Jahren alt und 2 Prozent gehören zur Gruppe der Hochbetagten (84 Jahre und älter).

In der Geschlechtsverteilung wird ein Frauenüberschuss erkennbar. Weibliche Versicherte sind mit 53 Prozent etwas stärker im Datensatz vertreten als die Männer. Hinsichtlich der Verteilung nach Altersgruppen und Geschlecht wird im mittleren Lebensalter (45-54 Jahre / 55-65 Jahre) eine relative Gleichverteilung deutlich. Zuvor – insbesondere im Kindes- und Jugendalter - dominiert die Zahl der männlichen Versicherten. Ab der Altersgruppe der 65- bis 74-Jährigen etwa kehrt sich diese Verteilung um, so dass bis zum hohen Lebensalter der Anteil der Frauen höher ist.

**Abbildung 2: Altersverteilung der Stichprobenversicherten (n = 2.301.015)**



Quelle: Stichprobendaten 2002 nach § 268 SGB V; eigene Berechnungen

### 3.2 Bestandsdaten

In diesem Modul sind die Bestandsdaten aller Stichprobenversicherten für das Jahr 2002 zu finden. Dafür stehen folgende Informationen für die Analyse zur Verfügung:

- Angabe, ob die/der Versicherter im Berichtsjahr verstorben ist
- Ausgaben für sonstige Leistungen (wie Heil- und Hilfsmittel, häusliche Pflege, Mutterschafts- bzw. Sterbegeld usw.)
- Dialyseleistung in der Berichtsperiode

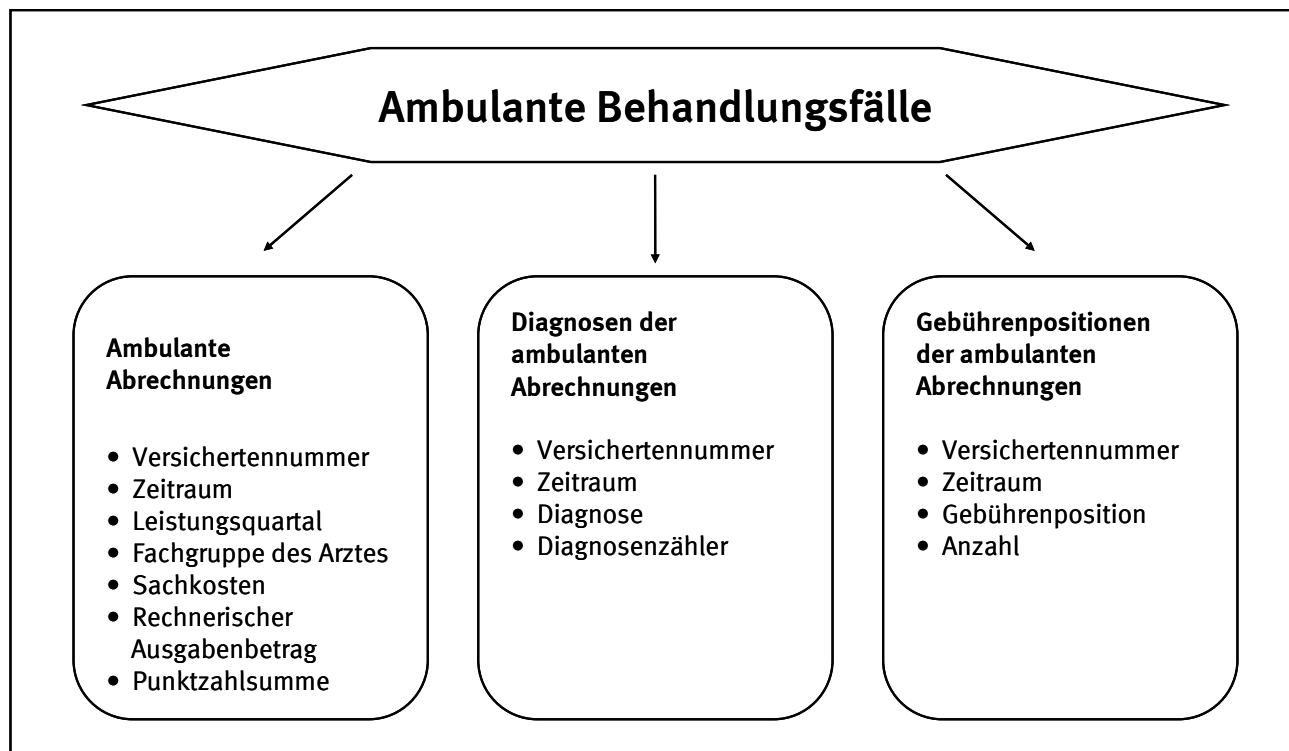
- Rechtskreis (Ost/West)
- Anzahl der Versichertentage in der Berichtsperiode
- RSA-Versichertengruppe.

Insgesamt ist innerhalb des Berichtsjahres 2002 etwa ein Prozent der Stichprobenversicherten verstorben. Sehr gering ist weiterhin die Zahl derjenigen, die in diesem Jahr eine Dialyseleistung in Anspruch genommen haben. Dabei handelt es sich um 1.025 Personen, also 0,04 Prozent. Dem Rechtskreis West (alte Bundesländer einschließlich Berlin) gehören 82 Prozent der Versicherten an, 18 Prozent sind im Rechtskreis Ost (neue Bundesländer) versichert.

### 3.3 Ambulante Behandlungsfälle

Die Daten der ambulanten Behandlungsfälle umfassen mehrere Satzarten. Sie beinhalten sämtliche Fälle im ambulanten Sektor des Jahres 2002. Es liegen Informationen zu den Abrechnungen, Diagnosen sowie Gebührenpositionen der ambulanten Behandlung vor (vgl. Abbildung 3). Insgesamt sind im Jahr 2002 etwa 12,9 Millionen Arztbesuche von den Stichprobenversicherten der GKV abgerechnet worden. Diese verteilen sich auf 84,6 Prozent der in die Untersuchung einbezogenen Versicherten.

Abbildung 3: Merkmalsinhalte der ambulanten Behandlungsfälle



Quelle: Stichprobendaten2002 nach § 268 SGB V; eigene Berechnungen



Im Bereich der ambulanten Abrechnungen sind Informationen über den Zeitraum der Behandlung (Monat und Jahr), das Leistungsquartal und die jeweilige Fachgruppe des Arztes dokumentiert. Weiterhin sind die Kosten dieser Behandlungen über die Merkmale Punktzahlsumme, Sachkosten sowie rechnerischem Ausgabenwert enthalten. Der rechnerische Ausgabenbetrag ergibt sich dabei aus dem Produkt von Punktzahlsumme des Abrechnungsfalles und vereinfachtem Punktwert zuzüglich der Sachkosten.

Ein anderes Modul innerhalb der ambulanten Behandlungsfälle sind die ICD-Diagnosen je Abrechnungsfall. Dieses beinhaltet über 44 Millionen Diagnosen, wobei je Diagnose eines Abrechnungsfalles aus der ambulanten Behandlung ein Datensatz vorliegt. Die Gebührenpositionen komplettieren die Daten im ambulanten Sektor. Sie basieren auf dem Einheitlichen Bewertungsmaßstab (EBM) für das Abrechnungsjahr 2002. Insgesamt sind über 96 Millionen Gebührenpositionen der ambulanten Behandlungsfälle im Datenmaterial enthalten.

### **3.4 Stationäre Behandlungsfälle**

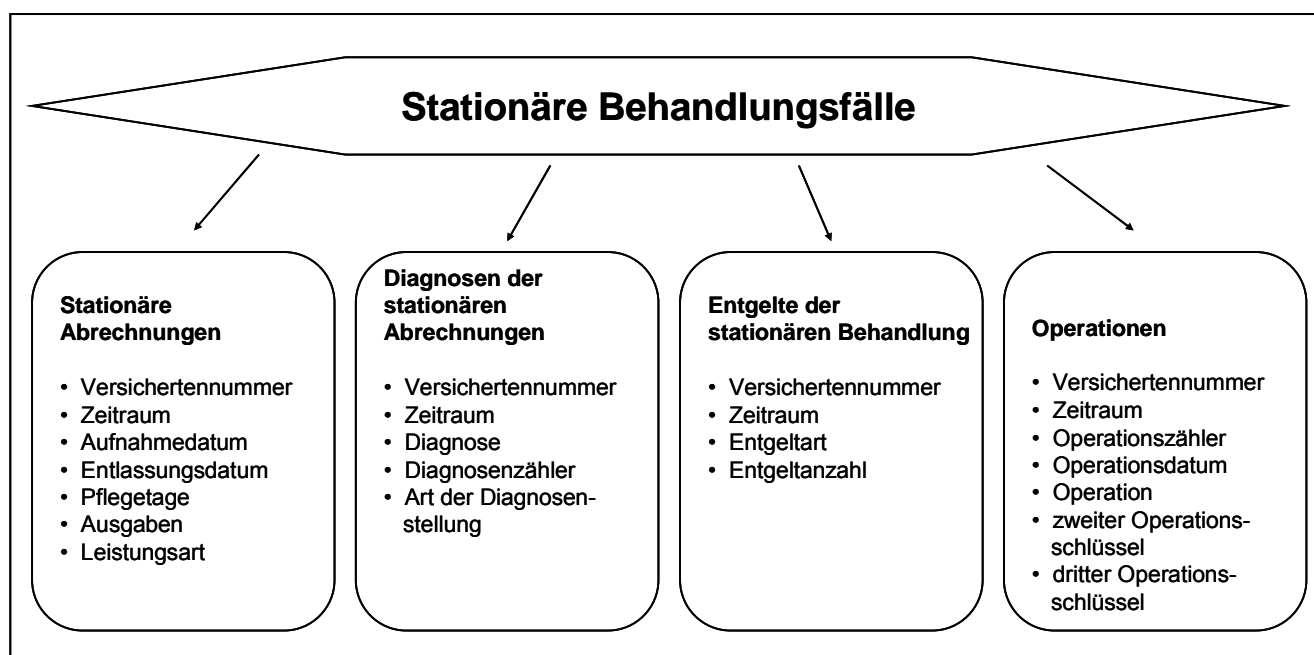
Die Daten der stationären Versorgung geben Auskunft zu den stationären Behandlungsfällen, Diagnosen und Entgelten dieser Behandlungen sowie Operationen im Rahmen der Krankenhausbehandlung (vgl. Abbildung 4). Darin sind sämtliche stationäre Behandlungen eines Stichprobenversicherten der GKV zu finden, 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.

Im Rahmen der Krankenhausbehandlung wurden im Jahr 2002 510.330 Operationen an insgesamt 9 Prozent der Stichprobenversicherten der GKV durchgeführt. Im Datenmaterial sind dazu Angaben über den Zeitpunkt (Monat und Jahr) sowie die Art der Operation, klassifiziert nach dem amtlichen Operationen- und Prozedurenschlüssel (OPS), zu finden.

### **3.5 Arzneimitteldaten**

Für etwa 79 Prozent der Stichprobenversicherten wurden im Jahr 2002 Arzneimittel abgerechnet. Mit insgesamt 23,2 Millionen Fällen ist das Datenmaterial diesbezüglich sehr umfangreich. Jedes von einer Apotheke abgerechnete Medikament, Verband- oder Hilfsmittel ist hier mit einem Datensatz abgebildet. Dazu liegen Informationen über das Verordnungsdatum (Monat und Jahr) und die Positionsnummer des Arzneimittels vor. Weiterhin sind die Pharmazentralnummer, die Anzahl der Packungen zu dieser Pharmazentralnummer sowie die Kassenausgaben für das gesamte Rezept enthalten.

Abbildung 4: Merkmale der stationären Behandlungsfälle



Quelle: Stichprobendaten 2002 nach § 268 SGB V; eigene Berechnungen

### 3.6 Arbeitsunfähigkeit und Krankengeld

Ebenfalls weist das Datenmaterial Informationen über Arbeitsunfähigkeit und den Bezug von Krankengeld auf. Von allen in die Untersuchung einbezogenen gesetzlich Versicherten waren im Jahr 2002 über 600.000 Personen mindestens einmal arbeitsunfähig – dies entspricht etwa jedem Vierten. Daraus ergeben sich insgesamt 1,2 Millionen Arbeitsunfähigkeitsfälle, die folgende Kriterien erfüllen: Beginn oder Ende der Arbeitsunfähigkeit liegen in der Berichtsperiode bzw. das Aufnahmedatum liegt vor 2002 und der Arbeitsunfähigkeitsfall ist am Ende des Jahres noch nicht abgeschlossen worden. Zudem haben 2,3 Prozent aller Versicherten bzw. 8,8 Prozent derjenigen, die in den Daten eine Arbeitsunfähigkeit aufweisen, an mindestens einem Tag im Berichtsjahr Krankengeld bezogen. Das bedeutet, von den 1,2 Millionen Krankheitsfällen mit Arbeitsunfähigkeit gehen fast 60.000 Fälle (4,9 Prozent) mit einem Leistungsbezug einher.

Das Datenmaterial enthält für die Analyse hinsichtlich Arbeitsunfähigkeit und Krankengeldbezug der GKV-Versicherten Merkmale wie Beginn und Ende der Arbeitsunfähigkeit (Monat und Jahr), die Anzahl der Arbeitsunfähigkeits- bzw. Krankengeldbezugstage, die Dauer des Krankengeldbezuges sowie die zugrunde liegenden Haupt- und weiteren Diagnosen.

## 4 Ausblick

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 über die kontrollierte Datenfernverarbeitung zur Verfügung. Auf diesem Weg wird die Nutzung des Informationspotenzials der Einzeldaten mittels einer Programmsyntax, die der Datennutzer selbst entwickelt, ermöglicht. Derzeit wird ein Konzept erstellt, welches die Nutzung dieser Daten an Gastwissenschaftlerarbeitsplätzen in den Statistischen Ämtern des Bundes und der Länder zulässt.<sup>9</sup> 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) für die Nutzung am eigenen Arbeitsplatz (OFF-Site) zur Verfügung zu stellen.<sup>10</sup> Das methodische Vorgehen der Anonymisierung von Einzeldaten umfasst dabei verschiedene Maßnahmen.<sup>11</sup>

Aufgrund der umfangreichen Daten über ambulante und stationäre Behandlungsfälle, Arbeitsunfähigkeit, Krankengeldbezug und Arzneimittelverordnungen von mehr als 2 Millionen Versicherten der GKV aus dem Jahr 2002 ist ein hohes Analysepotenzial gegeben. Die demografischen Angaben wie Alter, Geschlecht und Rechtskreis ermöglichen einerseits eine sehr detaillierte Betrachtung sowie das Aufzeigen differenter Patientenstrukturen und anfallender Gesundheitskosten. Ein wesentlicher Vorteil dieser Daten liegt zudem in der Zusammenführung von Informationen über die ambulante und stationäre Versorgung von gesetzlich Krankenversicherten.

## 5 Literatur

Merz, J./ Zwick, M. u.a. (2004): MIKAS – Mikroanalysen und amtliche Statistik. Statistisches Bundesamt (Hrsg.): Statistik und Wissenschaft, Bd. 1, Wiesbaden.

Reschke, P./ Sehlen, S./ Schiffhorst, G./ Schröder, W.F./ Lauterbach, K.W./ Wasem, J. (2005): Klassifikationsmodelle für Versicherte im Risikostrukturausgleich. Untersuchung zur Auswahl geeigneter Gruppenbildungen, Gewichtungsfaktoren und Klassifikationsmerkmale für einen direkt morbiditätsorientierten Risikostrukturausgleich in der gesetzlichen Krankenversicherung. Endbericht. Bundesministerium für Gesundheit und Soziale Sicherung (Hrsg.), Band 155, Nomos: Baden-Baden.

---

<sup>9</sup> Details zu den Zugangsformen finden Sie unter [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de).

<sup>10</sup> Scientific-Use-Files sind faktisch anonymisierte Mikrodatensätze, deren Reidentifikation nur mit einem unverhältnismäßig hohen Aufwand möglich ist. Siehe hierzu Zühlke/ Zwick/ Scharnhorst/ Wende (2003).

<sup>11</sup> Siehe hierzu Ronning, u.a. (2005); Willenborg, de Waal (2001); siehe auch Merz/ Zwick u.a. (2004).

- Ronning, G./ Sturm, R./ Höhne, J./ Lenz, R./ Rosemann, M./ Scheffler, M./ Vorgrimler, D. (2005): Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. Statistisches Bundesamt (Hrsg.): Statistik und Wissenschaft, Bd. 4, Wiesbaden.
- SGB V – Gesetzliche Krankenversicherung (2006). 14. überarbeitete Auflage. Beck-Texte. dtv: München.
- Willenborg, L./ de Waal, T. (2001): Elements of Statistical Disclosure Control. Lecture Notes in Statistics Bd. 155, Springer: New York, Berlin, Heidelberg.
- Zühlke, S./ Zwick, M./ Scharnhorst, S./ Wende, T. (2003): Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. In: Wirtschaft und Statistik. Statistisches Bundesamt (Hrsg.), Bd. 10/2003, Wiesbaden, S. 906-911.

## 6 Anhang

Datensatzbeschreibung der GKV-Stichprobe nach §268 SGB V für das Jahr 2002 (Stand: 21.11.2007)

<b>Leitdatei</b>					
<p>Dateiumfang (N = 2.301.015):            Je Stichprobenversichertem ein Datensatz (ohne Versicherte ohne Versicherungszeiten in 2002).            Kriterien für „Stichprobenversicherte“:</p> <ul style="list-style-type: none"> <li>• Versicherter hatte in 2002 an mindestens 1 Tag Krankenversicherungsschutz durch die Kasse.</li> <li>• wurde nicht als Auftragsfall (z.B. Sozialhilfe) geführt.</li> <li>• hat seinen Geburtstag am 11. eines beliebigen Monats eines beliebigen Jahres.</li> </ul>					
Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	systemfreie Nummerierung der Versicherten
ef2_leit	4	<b>Geburtsjahr</b>		numerisch	
ef3_leit	1	<b>Geschlecht</b> weiblich männlich	1 2	numerisch	

SA_311 Ambulante Abrechnungen					
Dateiumfang (N = 12.895.856): Je Fall eines Stichprobenversicherten ein Datensatz. (Die Satzarten 311, 312 und 313 sind nur über die Merkmalskombination ef1, ef2 und ef3 verknüpfbar, jedoch liegen nicht für jeden Versicherten alle Satzarten vor.)					
Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2	≤ 6	<b>Fallzähler</b>		numerisch	z. B. laufende Nummer, um Datensatz in Kombination mit ef1 und ef3 eindeutig zu identifizieren
ef3	5	<b>Zeitraum</b>		numerisch	Quartal und Berichtsjahr (QJJJ)
ef4_311	2	<b>Fachgruppe des Arztes</b> Anästhesie Augenheilkunde Chirurgie Gynäkologie Hals-Nasen-Ohren-Heilkunde Dermatologie Innere Medizin Kinderheilkunde Laboratoriumsmedizin Lungenheilkunde Mund-Kiefer-Chirurgie Nervenheilkunde Neurochirurgie Orthopädie Pathologie Psychotherapie Radiologie Urologie Nuklearmedizin Physikalische und rehabilitative Medizin Vorsorgemedizin Kliniken/ Institute Dialyseeinrichtungen Notfallmedizin Allgemeinmedizin/ praktischer Arzt nicht zuordenbar keine Angabe	01 02 03 04  05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24  25 99 leer	alphanumerisch	
ef5_311	≤ 7	<b>rechnerischer Ausgabenbetrag</b> keine Angabe	leer	numerisch	= Punktzahlsumme des Abrechnungsfalls, multipliziert mit dem vereinfachten Punktwert zzgl. Sachkosten in Cent (d.h. Kassenärztliche Vereinigungen übermitteln Punktzahlsumme je Patient, Arzt und Quartal; von den Bundesverbänden der Krankenkassen wurden die Punktzahlsummen in Ausgaben umgerechnet und mit Sachkosten der Ärzte zusammengefasst)
ef6_311	≤ 7	<b>Sachkosten</b>		numerisch	Summe der Sachkosten des Abrechnungsfalls inkl. Dialysekosten in Cent
ef7_311	5	<b>Leistungsquartal</b>		numerisch	Berichtsjahr und Quartal (JJJJQ)
ef8_311	≤ 7	<b>Punktzahlsumme</b> keine Angabe	leer	numerisch	Punktzahlsumme des Abrechnungsfalls in Zehntelpunkten (Fälle mit negativen Punktzahlen oder Sachkosten wurden nicht selektiert)

<b>SA_312 Diagnosen der ambulanten Behandlung</b>					
Dateiumfang (N = 44.208.439): Je Diagnose eines Abrechnungsfalles aus SA_311 ein Datensatz. (Die Satzarten 311, 312 und 313 sind nur über die Merkmalskombination ef1, ef2 und ef3 verknüpfbar, jedoch liegen nicht für jeden Versicherten alle Satzarten vor.)					
Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2	≤ 6	<b>Fallzähler</b>		numerisch	z. B. laufende Nummer; wie in Feld ‚Fallzähler‘ von SA_311
ef3	5	<b>Zeitraum</b>		numerisch	Quartal und Berichtsjahr (QJJJ); wie in Feld ‚Zeitraum‘ von SA_311
ef4_312	≤ 3	<b>Diagnosenzähler</b>		numerisch	Zähler für die einzelnen Diagnosen eines Falles
ef5_312	3	<b>Diagnose</b> ICD-10 keine (zulässige) Diagnosenangabe	A00-Z99  AAA	alphanumerisch	ohne Diagnosenzusätze; klassifiziert nach ICD-10; Grundlage: ICD-10 SGB-V Version 1.3; teilweise sind Diagnosen aus ICD-10 SGB-V Version 2.0 enthalten, die nicht in ICD-10 SGB-V Version 1.3 umgesetzt werden können; Klassifikationen unter <a href="http://www.dimdi.de">www.dimdi.de</a>

<b>SA_313 Gebührenpositionen der ambulanten Behandlung</b>					
Dateiumfang (N = 96.312.416): Für jede verschiedene Gebührenposition (GOP) eines Abrechnungsfalles aus SA_311 ein Datensatz. (Die Satzarten 311, 312 und 313 sind nur über die Merkmalskombination ef1, ef2 und ef3 verknüpfbar, jedoch liegen nicht für jeden Versicherten alle Satzarten vor.)					
Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2	≤ 6	<b>Fallzähler</b>		numerisch	z. B. laufende Nummer; wie in Feld ‚Fallzähler‘ von SA_311
ef3	5	<b>Zeitraum</b>		numerisch	Quartal und Berichtsjahr (QJJJ); wie in Feld ‚Zeitraum‘ von SA_311
ef4_313	≤ 7	<b>GOP</b> Klassifizierung gemäß EBM		alphanumerisch	Gebührenposition nach Einheitlichem Bewertungsmaßstab (EBM) für Abrechnungsjahr 2002
ef5_313	3	<b>Anzahl</b>		numerisch	Anzahl, mit der GOP aus Feld ef4_313 bei diesem Fall abgerechnet wurde

## SA\_411 Jahresdaten der Versicherten für das Berichtsjahr 2002

Dateiumfang (N = 2.302.427):

Je Stichprobenversichertem ein Datensatz (bei Wechsel des Rechtskreises bzw. der Versichertengruppe im Berichtsjahr mehr als ein Datensatz). (Die Jahresdaten sind über das Merkmal ef1 mit den anderen Satzarten verknüpfbar.)

Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2_411	4	<b>Berichtsperiode</b>		numerisch	Berichtsjahr (JJJJ)
ef3_411	1	<b>Verstorben</b> verstorben sonst	1 0	numerisch	Stand am Ende der Berichtsperiode (BP)
ef4_411	≤ 8	<b>Ausgaben sonstige Leistungen</b> Versicherte ohne Ausgaben in BP	0	numerisch	Ausgaben im HLB 5 (SL); Jahressumme in Cent
ef5_411	1	<b>Dialyse</b> mindestens eine Dialyseleistung in BP abgerechnet sonst	1 0	numerisch	
ef6_411	1	<b>Rechtskreis</b> Ostdeutschland Westdeutschland	0 W	alphanumerisch	zum 31.12.2002 bzw. zum Ende der Versicherungszeit; Ostdeutschland = neue Bundesländer ohne Berlin, Westdeutschland = alte Bundesländer einschließlich Berlin
ef7_411	≤ 3	<b>Versichertentage</b>		numerisch	in Berichtsperiode insgesamt
ef8_411	2	<b>RSA-Versichertengruppe</b> Mitglieder mit allg. Beitragssatz und KG-Anspruch nach mind. 6 Wochen Mitglieder mit erhöhtem Beitragssatz und KG-Anspruch nach max. 6 Wochen Mitglieder u. Familienversicherte ohne Anspruch auf KG mit ermäßigtem Beitragssatz Mitglieder mit allg. Beitragssatz u. KG-Anspruch nach mind. 6 Wochen, die Erwerbsminderungsrente beziehen Mitglieder mit erhöhtem Beitragssatz und KG-Anspruch nach max. 6 Wochen, die Erwerbsminderungsrente beziehen Mitglieder u. Familienversicherte ohne Anspruch auf KG mit ermäßigtem Beitragssatz, die Erwerbsminderungsrente beziehen	01 02 03 04 05 06	alphanumerisch	



## SA\_412 Stationäre Behandlungsfälle

Dateiumfang (N = 536.210):

Je Fall eines Stichprobenversicherten ein Datensatz, wenn eines der folgenden Kriterien erfüllt ist:

- Aufnahme­datum des Falles liegt in der Berichtsperiode.
- Entlassungsdatum des Falles liegt in der Berichtsperiode.
- Aufnahme­datum des Falles liegt vor der Berichtsperiode und der Fall ist am Ende der Berichtsperiode noch nicht abgeschlossen.

(Die Satzarten 412, 413, 414 und 415 sind nur über die Merkmalskombination ef1, ef2 und ef3 verknüpfbar, jedoch liegen nicht für jeden Versicherten alle Satzarten vor.)

Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2	≤ 6	<b>Fallnummer</b>		numerisch	z. B. laufende Nummer, um Datensatz in Kombination mit ef1 und ef3 eindeutig zu identifizieren
ef3	4	<b>Berichtsperiode</b>		numerisch	Berichtsjahr (JJJJ)
ef4_412	6	<b>Aufnahmedatum</b>		numerisch	Jahr und Monat (JJJMM)
ef5_412	6	<b>Entlassungsdatum</b> bei Tabellenerstellung offene Fälle	000000	numerisch	Jahr und Monat (JJJMM); Werte gehen bis ins 1. Quartal 2003 hinein, da Zeitpunkt der Tabellenerstellung erst in 2003; Werte ab April 2003 wurden als offene Fälle codiert
ef6_412	4	<b>Pflegetage</b> keine Angabe	leer	numerisch	Entlassungsdatum minus Aufnahmedatum in Tagen; bei Tabellenerstellung offene Fälle: Tabellenerstellungsdatum minus Beginndatum in Tagen plus 1
ef7_412	8	<b>Ausgaben</b> keine Angabe	leer	numerisch	nur für Berichtsperiode in Cent
ef8_412	1	<b>Leistungsart</b> vollstationäre Behandlung, Entbindung, Sterilisation, Schwangerschaftsabbruch teilstationäre Behandlung Anschluss-Rehabilitation vor- oder nachstationäre Behandlung, ambulante Operation am Krankenhaus sonstige oder keine Angabe möglich	1 2 3  4 9	numerisch	

SA_413 Diagnosen der stationären Behandlung					
Dateiumfang (N = 1.979.446): Je Aufnahme-, Einweisungs- oder Entlassungs- bzw. Verlegungsdiagnose eines Falles aus SA_412 ein Datensatz. (Die Satzarten 412, 413, 414 und 415 sind nur über die Merkmalskombination ef1, ef2 und ef3 verknüpfbar, jedoch liegen nicht für jeden Versicherten alle Satzarten vor.)					
Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2	6	<b>Fallnummer</b>		numerisch	z. B. laufende Nummer; wie in Feld ‚Fallnummer‘ von SA_412
ef3	4	<b>Berichtsperiode</b>		numerisch	Berichtsjahr (JJJJ); wie in Feld ‚Berichtsperiode‘ von SA_412
ef4_413	4	<b>Diagnosenzähler</b>		numerisch	Zähler für Diagnosen, um Datensatz eindeutig zu identifizieren
ef5_413	3	<b>Diagnose</b> ICD-10 ICD-9 keine (zulässige) Diagnosenangabe	A00 - Z99 001 - 999  AAA	alphanumerisch	Klassifiziert nach ICD-10; Grundlage: ICD-10 SGB-V Version 2.0; teilweise sind Diagnosen aus ICD-10 SGB-V Version 1.3, ICD-10 WHO Version 1.3 bzw. ICD-9 enthalten, die nicht in ICD-10 SGB-V Version 1.3 umgesetzt werden können; Klassifikationen unter <a href="http://www.dimdi.de">www.dimdi.de</a>
ef6_413	1	<b>Art der Diagnosenstellung</b> falls ef5_413 eine der Aufnahme- bzw. Einweisungsdiagnosen bei Diagnosen der Fachabteilungen bei Entlassungs-/ Verlegungsdiagnosen Sekundärdiagnose bei Entlassung bei Nebendiagnosen falls Angabe nicht möglich	A F E S N X	alphanumerisch	

SA_414 Entgelte der stationären Behandlung					
Dateiumfang (N = 1.592.749): Je Entgeltposition eines Falles aus SA_412 ein Datensatz (ohne Fälle, für die keine Entgeltart verfügbar ist). (Die Satzarten 412, 413, 414 und 415 sind nur über die Merkmalskombination ef1, ef2 und ef3 verknüpfbar, jedoch liegen nicht für jeden Versicherten alle Satzarten vor.)					
Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2	6	<b>Fallnummer</b>		numerisch	z. B. laufende Nummer; wie in Feld ‚Fallnummer‘ von SA_412
ef3	4	<b>Berichtsperiode</b>		numerisch	Berichtsjahr (JJJJ); wie in Feld ‚Berichtsperiode‘ von SA_412
ef4_414	4	<b>Positionsnummer</b>		numerisch	Zähler für die Entgeltposition, um Datensatz eindeutig zu identifizieren
ef5_414	8	<b>Entgeltart</b>  keine zulässige Angabe	AAA	alphanumerisch	nach Entgeltartenschlüssel gemäß BpflV für das Jahr 2002 (Quelle: VdAK)
ef6_414	4	<b>Entgeltanzahl</b>		numerisch	Anzahl, mit der das Entgelt in ef5_414 bei diesem Fall abgerechnet wurde; Entgeltpositionen ohne Angabe der Anzahl werden mit ‚0001‘ angegeben

## SA\_415 Im Rahmen der Krankenhausbehandlung durchgeführte Operationen

Dateiumfang (N = 510.330):

Je Operationstag eines Falles aus SA\_412 ein Datensatz. Durchläuft ein Patient mehrere Fachabteilungen mit mehreren OPs, sind entsprechend viele Datensätze enthalten. (Die Satzarten 412, 413, 414 und 415 sind nur über die Merkmalskombination ef1, ef2 und ef3 verknüpfbar, jedoch liegen nicht für jeden Versicherten alle Satzarten vor.)

Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2	6	<b>Fallnummer</b>		numerisch	z. B. laufende Nummer; wie in Feld ‚Fallnummer‘ von SA_412
ef3	4	<b>Berichtsperiode</b>		numerisch	Berichtsjahr (JJJJ)
ef4_415	4	<b>Operationszähler</b>		numerisch	laufende Nummer zum Operationssatz (z.B. innerhalb eines Falles), um Datensatz in Kombination mit ef1 und ef2 eindeutig zu identifizieren
ef5_415	6	<b>Operationsdatum</b>		numerisch	Jahr und Monat (JJJJMM)
ef6_415	3	<b>Operation</b> 110 - 999 keine zulässige Angabe	110 - 999 AAA	alphanumerisch	Klassifiziert nach amtlichem Operationen- und Prozedurenschlüssel (OPS); Grundlage: OPS-301 Version 2.1; teilweise sind Schlüssel aus OPS-301 Version 2.0 enthalten (199, 857, 895, 896, 899, 999) Klassifikationen unter <a href="http://www.dimdi.de">www.dimdi.de</a>
ef7_415	3	<b>zweiter Operationsschlüssel</b> 110 - 999 keine zulässige Angabe entfällt	110 - 999 AAA leer	alphanumerisch	Zusatzschlüssel 1
ef8_415	3	<b>dritter Operationsschlüssel</b> 110 - 999 keine zulässige Angabe entfällt	110 - 999 AAA leer	alphanumerisch	Zusatzschlüssel 2

## SA\_416 Arzneimitteldaten

Dateiumfang (N = 23.221.610):

Für jedes von einer Apotheke abgerechnete Medikament, Verband- oder Hilfsmittel bei einem Versicherten aus Tabelle ‚Leitdatei‘ ein Datensatz, wenn das Rezept in die Berichtsperiode fällt.

Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2_416	4	<b>Berichtsperiode</b>		numerisch	Berichtsjahr (JJJJ)
ef3_416	6	<b>Verordnungsdatum</b>		numerisch	Jahr und Monat (JJJJMM); ersatzweise Abgabedatum, ersatzweise Abgabemonat
ef4_416	1	<b>Positionsnummer</b>		numerisch	Position des Medikaments auf dem Rezept; Datensätze ohne Angabe der Position erhalten in diesem Feld den Wert ‚1‘
ef5_416	≤ 8	<b>Kassenausgaben Rezept</b> Folgemedikament(e) eines Rezeptes	0	numerisch	beim ersten Medikament eines Rezeptes (Positionsnummer = 1): Nettoausgaben der Kasse für das gesamte Rezept (d.h. Ausgaben abzüglich Patientenzuzahlungen und Rabatt); ‚0‘ bei allen folgenden Medikamenten eines Rezeptes (d.h. Positionsnummer > 1); Währung in Cent
ef6_416	≤ 10	<b>Pharmazentralnummer (PZN)</b> Klassifizierung gemäß Schlüsselverzeichnis		alphanumerisch	bei Fertigarzneimitteln (bzw. Verband- oder Hilfsmitteln); sonst Platzhalter für Rezepturen/ Homöopathika/ direkt abgerechnete Medikamente usw.;  <i>Hinweis:</i> ein Schlüsselverzeichnis liegt nicht vor, Bezug ist unter bestimmten Voraussetzungen bei der IFA GmbH möglich (die im Schlüsselverzeichnis für PZN vorhandenen Informationen beziehen sich auf Marktnamen, Packungsgröße, Darreichungsform)
ef7_416	≤ 3	<b>Anzahl Packungen</b>		numerisch	Anzahl der Packungen zu dieser PZN auf diesem Rezept; Datensätze ohne Angabe der Anzahl erhalten in diesem Feld den Wert ‚1‘

## SA\_417 Arbeitsunfähigkeit und Krankengeld

Dateiumfang (N = 1.211.818):

Für jeden Arbeitsunfähigkeitsfall (ohne Fälle mit Krankengeldbezug, die nicht mit Arbeitsunfähigkeit (AU) in Zusammenhang stehen, z.B. Erkrankung des Kindes) eines Versicherten aus Tabelle ‚Leitdatei‘ in der Berichtsperiode ein Datensatz, wenn der Versicherte zu Beginn der Arbeitsunfähigkeit Krankengeld-anspruchsberechtigt war.

Im Datensatz sind die Fälle enthalten, die eines der folgenden Kriterien erfüllen:

- das AU-Beginndatum des Falles liegt in der Berichtsperiode.
- das AU-Endedatum des Falles liegt in der Berichtsperiode.
- das Aufnahmedatum des Falles liegt vor der Berichtsperiode und der Fall ist am Ende der Berichtsperiode noch nicht abgeschlossen.

Merkmal	Satzstellen	Bezeichnung	Schlüssel	Feldeigenschaft	Erläuterungen
ef1	7	<b>Versichertennummer</b>		alphanumerisch	aus Feld ‚Versichertennummer‘ der Leitdatei
ef2_417	≤ 6	<b>Fallnummer</b>		numerisch	z.B. laufende Nummer, um Datensatz in Kombination mit ef1 eindeutig zu identifizieren
ef3_417	4	<b>Berichtsperiode</b>		numerisch	Berichtsjahr (JJJJ)
ef4_417	6	<b>AU-Beginndatum</b>		numerisch	Jahr und Monat (JJJJMM)
ef5_417	6	<b>AU-Endedatum</b> offene Fälle	000000	numerisch	Jahr und Monat (JJJJMM)
ef6_417	≤ 4	<b>AU-Tage</b>		numerisch	Endedatum minus Beginndatum in Tagen plus 1; bei offenen Fällen: Tabellenerstellungsdatum minus Beginndatum in Tagen plus 1
ef7_417	≤ 3	<b>Krankengeldbezugstage</b> Fälle ohne Krankengeldbezug	0	numerisch	nur auf Berichtsperiode beschränkt; bei offenen Fällen: 31.12. der Berichtsperiode minus Beginndatum in Tagen plus 1
ef8_417	≤ 7	<b>Krankengeldausgaben</b> Fälle ohne Krankengeldbezug	0	numerisch	nur auf Berichtsperiode beschränkt; nicht berücksichtigt sind KG-Zahlungen wegen Erkrankung des Kindes; Währung in Cent
ef9_417	3	<b>Diagnose 1</b> ICD-10 ICD-9 keine (zulässige) Diagnosenangabe	A00 - Z99 001 - 999 AAA	alphanumerisch	möglichst mit Hauptdiagnose gefüllt; klassifiziert nach ICD-10; teilweise sind Diagnosen aus ICD-9 enthalten, die nicht in ICD-10 umgesetzt werden können; Klassifikationen unter <a href="http://www.dimdi.de">www.dimdi.de</a>
ef10_417	3	<b>Diagnose 2</b> ICD-10 ICD-9 keine (zulässige) Diagnosenangabe entfällt	A00 - Z99 001 - 999 AAA leer	alphanumerisch	weitere Diagnose; klassifiziert nach ICD-10; teilweise sind Diagnosen aus ICD-9 enthalten, die nicht in ICD-10 umgesetzt werden können; Klassifikationen unter <a href="http://www.dimdi.de">www.dimdi.de</a>
ef11_417	3	<b>Diagnose 3</b> ICD-9 ICD-10 keine (zulässige) Diagnosenangabe entfällt	001 - 999 A00 - Z99 AAA leer	alphanumerisch	weitere Diagnose; klassifiziert nach ICD-10; teilweise sind Diagnosen aus ICD-9 enthalten, die nicht in ICD-10 umgesetzt werden können; Klassifikationen unter <a href="http://www.dimdi.de">www.dimdi.de</a>