

Krankenhaus-Diagnosestatistik – Scientific Use File und beispielhafte Analysen

Gabriele Philipp* und Dr. Olaf Schoffer**

* Statistisches Landesamt des Freistaates Sachsen

** FDZ der Statistischen Ämter der Länder

Einleitung

Gemäß dem Bundesstatistikgesetz (BStatG) „dürfen vom Statistischen Bundesamt und den statistischen Ämtern der Länder Einzelangaben an Hochschulen und sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden, wenn die Einzelangaben nur mit unverhältnismäßig großem Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können“¹. Auf Grundlage dieser Regelung zur „faktischen Anonymität“ ermöglichen die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder der Wissenschaft einen privilegierten Zugang zu Mikrodaten der amtlichen Statistik (vgl. Scheffler 2005).

In den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder werden verschiedene Statistiken fachlich zentralisiert aufbereitet und der Wissenschaft als Einzeldaten zur Verfügung gestellt.² Im Forschungsdatenzentrum der Statistischen Landesämter (FDZ/L) stehen dabei insbesondere regional tief gegliederte Einzeldaten für wissenschaftliche Analysen zur Verfügung.

Der regionale Standort Kamenz des FDZ/L bearbeitet als fachlichen Schwerpunkt die Statistiken des Gesundheits- und Sozialwesens. Hier werden die Einzeldaten aller Bundesländer aus den Bereichen Krankenhaus- und Todesursachen-Statistik sowie Sozialhilfe- und Jugendhilfestatistik gesammelt und aufbereitet. Dies schließt nicht zuletzt die Erstellung standardisierter faktisch anonymisierter Mikrodatenfiles (Scientific Use Files) ein.

Ein wichtiges Vorhaben des Standortes Kamenz ist daher die Konzeption eines Scientific Use Files der Krankenhaus-Diagnosestatistik. Die Arbeiten dafür erfolgen in Zusammenarbeit mit dem Statistischen Bundesamt im Rahmen eines Anonymisierungsprojektes. Zusätzlich werden für die Krankenhaus-Diagnosestatistik in Form eines Forschungsprojektes am Standort Kamenz beispielhaft deskriptive Analysen durchgeführt, um das Potenzial dieser Daten verdeutlichen zu können. Die Ergebnisse beider Projekte werden nachfolgend vorgestellt.

Weitere aktuelle Projekte am Standort Kamenz sind die Anonymisierung der Jugendhilfestatistik für die Arbeit am Gastwissenschaftler-Arbeitsplatz sowie die Aufbereitung der Kranken-

¹ § 16 Abs. 6 BStatG.

² Beim Workshop hat Herr Carsten Kuchler vom FDZ des Statistischen Bundesamtes zu den Aufgaben, Produkten und Zugangswegen zu den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder gesprochen. Für eine ausführliche Darstellung vgl. Zühlke et al. (2004). Die gemeinsame Internet-Präsenz der Forschungsdatenzentren unter www.forschungsdatenzentrum.de enthält darüber hinaus Informationen zum gemeinsamen Datenangebot sowie über aktuelle Projekte und Veranstaltungen.

hausstatistik (Grund- und Kostendaten) für die Analyse mittels kontrollierter Datenfernverarbeitung. In Vorbereitung ist außerdem die Aufbereitung der Einzeldaten für die Todesursachenstatistik aller Bundesländer, welche neben der Diagnosestatistik die zweite Schwerpunktstatistik für den Standort Kamenz bildet.

Diagnosestatistik als Teil der Krankenhausstatistik

Da die betrachtete Diagnosestatistik einen Teil der Krankenhausstatistik darstellt, folgt zunächst ein Überblick über deren allgemeine Eigenschaften. Die Krankenhausstatistik ist eine Totalerhebung der Krankenhäuser sowie der Vorsorge- oder Rehabilitationseinrichtungen. Sie enthält Informationen über organisatorische Einheiten, personelle und sachliche Ausstattung sowie erbrachte Leistungen in den jeweiligen Einrichtungen. Somit sind beispielsweise Aussagen über die Struktur der stationären Versorgung, die Zusammenhänge zwischen Morbiditäts- und Kostenentwicklung im Krankenhausbereich oder die regionale Häufung von Krankheitsarten möglich.

Damit bildet die Krankenhausstatistik die Planungsgrundlage vieler gesundheitspolitischer Entscheidungen. Beispielsweise dient sie zur Bedarfsermittlung an Krankenhausleistungen und technischer Ausstattung oder zur Untersuchung von Lücken in der medizinischen Versorgung. Ebenso ist sie Grundlage verschiedenster wissenschaftlicher Analysen, wie etwa Untersuchungen von Patientenströmen oder zur Entwicklung bestimmter Erkrankungen.

Die Krankenhausstatistik gliedert sich in drei Teile. Neben der Diagnosestatistik (Teil II) bilden so die Grunddaten (Teil I) mit Informationen zu technischer und personeller Ausstattung sowie erbrachten Leistungen der Einrichtungen und der Kostennachweis (Teil III) mit Informationen zu Sach-, Personal- und Ausbildungskosten ein umfassendes Bild der deutschen Krankenhauslandschaft.

Die Diagnosestatistik umfasst die Hauptdiagnosen sowie allgemeine Angaben zur Person entlassener oder verstorbener vollstationärer Patienten aus Krankenhäusern. Als Hauptdiagnose wird die Diagnose angegeben, welche hauptsächlich die Dauer der stationären Behandlung beeinflusst bzw. den größten Teil medizinischer Leistungen verursacht hat. Sie wird seit dem Berichtsjahr 2000 nach der 10. Revision der Internationalen Klassifikation der Krankheiten (ICD-10) verschlüsselt (vgl. DIMDI 2000), zuvor kam die ICD-9 zur Anwendung, (siehe hierzu auch den Beitrag von Himmelreicher im vorliegenden Band).

Die Diagnosestatistik wird zusammen mit den anderen Teilen der Krankenhausstatistik jährlich zum 31.12. erhoben. Mehrfach im Jahr vollstationär behandelte Patienten werden entsprechend mehrfach erfasst, aus jedem Krankenhausaufenthalt wird ein neuer Eintrag zusammengestellt. Die Diagnosestatistik ist somit fallbezogen - ein Rückschluss auf die Anzahl der Krankenhausaufenthalte eines bestimmten Patienten ist aufgrund fehlender eindeutiger personenidentifizierender Merkmale nicht möglich.

Die Daten der Krankenhauspatienten werden seit 1993 erfasst. Seit 2003 wird die Diagnosestatistik zusätzlich auch in Vorsorge- sowie Rehabilitationseinrichtungen mit 100 und mehr Betten erhoben.

Die beschriebenen Informationen über die behandelten Patienten werden um Angaben aus den Grunddaten der Krankenhausstatistik ergänzt, um leichter untersuchen zu können, in Einrichtungen welcher Art die Behandlungen erfolgten. Damit sind im Wesentlichen folgende Merkmale in der Diagnosestatistik enthalten:

Angaben zu den behandelten Patienten

- (Haupt-)Diagnose
- Wohnort, Geschlecht, Alter
- Verweildauer (mit Aufnahme- und Entlassungsdatum)
- Sterbefall (ja/nein)
- Operation im Zusammenhang mit Hauptdiagnose (ja/nein)
- Fachabteilung mit längster Verweildauer

Angaben zu den behandelnden Einrichtungen

- Standort der Einrichtung
- Art der Einrichtung, des Trägers und der Förderung
- Anzahl der Betten und der Fachabteilungen
- Anteil der Belegbetten.

Die Erarbeitung sowie Komplettierung von Metadaten für die Diagnosestatistik findet zurzeit statt. Der Qualitätsbericht zu dieser Statistik liegt bereits vor.³

Scientific Use File der Diagnosestatistik

In der ersten Nutzerumfrage zum Datenbedarf der Wissenschaft (vgl. Zühlke & Hetke 2003) zählt die Statistik der Krankenhäuser zu den 15 am häufigsten genannten Statistiken, welche weit gehend deckungsgleich zu den im Gründungsausschuss des Rates für Sozial- und Wirtschaftsdaten für den Aufbau des Forschungsdatenzentrums der statistischen Landesämter als prioritär genannten sind.

Daher wurde festgelegt, für die Diagnosestatistik als einem zentralen Bestandteil der Krankenhausstatistik ein Scientific Use File zu erstellen. Ein solches standardisiertes Mikrodatenfile ermöglicht es, auf Datenbedarf für wissenschaftliche Analysen schnell reagieren zu können. Voraussetzung für die Akzeptanz dieses Datenangebotes durch die Wissenschaft ist jedoch, dass das Analysepotenzial bei der dabei erforderlichen faktischen Anonymisierung der Einzeldaten bestmöglich erhalten bleibt.

Um dieses Ziel zu erreichen, erfolgt die Konzeption eines faktisch anonymisierten Mikrodatenfiles der Diagnosestatistik unter Einbeziehung eines wissenschaftlichen Beraterkreises. Auf Ba-

³ http://www.destatis.de/download/qualitaetsberichte/qualitaetsbericht_diagnosekhp.pdf.

sis der Vorschläge dieses Beraterkreises und von Überlegungen sowie Analysen zur faktischen Anonymisierbarkeit ist das nachfolgend beschriebene Konzept für das Scientific Use File der Diagnosedaten von Krankenhauspatienten und -patientinnen entstanden.

Zentrale Elemente des Anonymisierungskonzeptes sind

- der Erhalt der Vollerhebung,
- die Trennung der Angaben zu Patienten und zu Krankenhäusern in Form zweier Datensätze,
- die Beschränkung auf wenige „wichtige“ Variablen und
- die Vergrößerung der Merkmalswerte einiger Variablen.

Der Verzicht auf Ziehung einer Stichprobe zu Gunsten des Erhalts aller Fälle von im Berichtsjahr entlassenen Patienten ist ein wichtiger Schritt zur Erhaltung der Repräsentativität für die deutsche Krankenhauslandschaft. Im Unterschied zum Einzeldatenmaterial mussten aus Gründen der Geheimhaltung dennoch einige Fälle im Scientific Use File entfallen. Insbesondere sind das Fälle, deren Merkmale „Hauptdiagnose“, „Verweildauer“, „Alter“ oder „Geschlecht“ als Merkmalswerte „unbekannt“ aufweisen sowie Fälle, welche eine eindeutige Zuordnung zu bestimmten Personen erlauben. Der Anteil dieser Fälle, welche in Analysen ohnehin oft ausgeschlossen würden, liegt im Promillebereich. Das führt dazu, dass die mit dem Scientific Use File berechneten Ergebnisse von den Publikationen der statistischen Ämter abweichen. Anhand der dokumentierten Anzahl entfallener Fälle ist es jedoch möglich, die Größenordnung der dadurch verursachten Änderung abzuschätzen.

Bei der Anonymisierung ist zunächst zu bedenken, welche Informationen besonders schutzwürdig sind. Für Krankenhäuser ist insbesondere zu vermeiden, dass Angaben zu ihrer wirtschaftlichen Situation bekannt werden. Für die Patienten hingegen ist hauptsächlich die Hauptdiagnose als Grund des jeweiligen Krankenhausaufenthaltes zu schützen, da diese durch Dritte, etwa Arbeitgeber oder Nachbarn, zum Nachteil des Patienten verwendet werden kann.

Um dieser Anforderung bei möglichst geringem Verlust an Analysepotenzial gerecht zu werden, erfolgt die Aufteilung der Angaben zu Patienten und zu Krankenhäusern auf zwei getrennte Datensätze. Somit wird die Identifikation sowohl von Patienten als auch von Krankenhäusern über zusätzliche Informationen erschwert. Der erste Datensatz beinhaltet hauptsächlich patientenbezogene Daten und soll vornehmlich epidemiologischen Untersuchungen dienen. Der zweite Datensatz fokussiert dagegen krankenhausbefugene Daten und orientiert sich damit stärker an gesundheitsökonomischen und -politischen Fragestellungen.

Als weitere Maßnahme zur Anonymisierung werden – gemäß vom wissenschaftlichen Beraterkreis vorgeschlagener Prioritäten – nur die wichtigsten Merkmale in das Scientific Use File übernommen. Für wissenschaftliche Analysen sind die Anzahl der Patienten sowie deren Verweildauer nach Hauptdiagnose (vierstelliger ICD-Schlüssel) von zentraler Bedeutung. Daher sind die beiden Zielvariablen Verweildauer und Hauptdiagnose in beiden Datensätzen enthalten. Im patientenbezogenen Datensatz (DS I) werden sie ergänzt um Alter, Geschlecht und Wohnort des Patienten. Im krankenhausbefugenen Datensatz (DS II) werden die Zielgrößen er-

weitert um Alter, Behandlungsgebiet, Fachabteilung mit der längsten Verweildauer sowie Trägerschaft und Bettengrößenklasse des Krankenhauses.

Neben den bereits beschriebenen Anonymisierungsmaßnahmen erfordern einige Merkmale eine Vergrößerung, also eine Gruppierung der Merkmalsausprägungen. So wird der Patientenwohntort nur noch bis auf Kreisebene angegeben, der ursprünglich achtstellige Amtliche Gemeindeschlüssel (AGS) also auf fünf Stellen eingeschränkt. Für den Behandlungsort wird hingegen die Angabe des Bundeslandes aus dem AGS verwendet, um drei Gebiete (Nord, Süd und Ost) zu bilden. Die Merkmalsausprägungen von Fachabteilung mit der längsten Verweildauer und Bettengrößenklasse werden in Gruppen zusammengefasst. Die Merkmalsausprägungen von Verweildauer und Alter werden ebenfalls gruppiert, jedoch ist für diese die Gruppierung zwischen patienten- und krankenhausbegrenztem Datensatz verschoben.

Der aus den beschriebenen Anonymisierungsmaßnahmen resultierende Aufbau des Scientific Use Files mit Details bezüglich der Merkmalsausprägungen ist **Tabelle 1** zu entnehmen.

Merkmale, welche durch die beschriebene Anonymisierung im Scientific Use File nicht oder nur vergrößert berücksichtigt werden, können über andere Zugangswege des FDZ/L genutzt werden. Dies sind insbesondere die On-Site-Nutzung an einem Gastwissenschaftler-Arbeitsplatz in den Räumen eines Statistischen Landesamtes sowie die kontrollierte Datenfernverarbeitung.

Beispielhafte Analysen der Diagnosestatistik

Herr Dr. Schoffer hat im Februar 2005 für das Forschungsprojekt „Beispielhafte Analysen der Krankenhaus-Diagnosestatistik“ einen Antrag zur Nutzung der Einzeldaten beim FDZ/L gestellt. Ziel des Projektes ist es, einen Eindruck davon zu vermitteln, welche Analysemöglichkeiten die Daten bieten sowie Anregungen für Analyseansätze zu geben, ohne dabei zu sehr in die Tiefe zu gehen. Anhand des zuerst vollständig vorliegenden Einzelmateriale des Jahres 2001 wurden beispielhafte Analysen, Grafiken und Fragestellungen erarbeitet, die zur Öffentlichkeitsarbeit, insbesondere auf Nutzerkonferenzen und Themenworkshops des FDZ/L verwendet werden.

Einige der im Rahmen dieses Projektes erarbeiteten Ergebnisse werden nachfolgend vorgestellt.⁴ In der ersten Analyse wird die Wanderung von Krankenhaus-Patienten betrachtet. In **Tabelle 2** sind Prozentanteile der Patienten angegeben, welche sich in dem Bundesland behandeln lassen, in dem sie wohnhaft sind, sowie die Prozentanteile derer, die zur Behandlung ihr Bundesland verlassen haben.

Auf eine ausführliche Interpretation der Ergebnisse soll an dieser Stelle verzichtet werden. Besonders häufig gibt es jedoch Patientenströme zwischen den Stadtstaaten und den umgebenden Bundesländern sowie aus Rheinland-Pfalz in die umgebenden Bundesländer.

⁴ Die Analysen wurden anhand des Einzeldatenmaterials der Diagnosestatistik durchgeführt. Daher sei darauf hingewiesen, dass nicht alle der angesprochenen Fragestellungen mit dem Scientific Use File in der hier verwendeten Detailtiefe zu analysieren sind. Dafür stehen jedoch die erwähnten anderen Zugangswege des FDZ/L offen.

Tabelle 1: Merkmale und Merkmalsausprägungen im Scientific Use File

Merkmal	Merkmalsausprägungen	
	Patientenbezogener Datensatz	Krankenhausbezogener Datensatz
Hauptdiagnose	Vierstelliger ICD-10 Schlüssel	Vierstelliger ICD-10 Schlüssel
Verweildauer	0 Tage 1 Tag usw. in 1-Tages-Schritten bis 42 Tage 43 bis 49 Tage 50 bis 56 Tage 57 bis 63 Tage 64 bis 70 Tage 71 bis 84 Tage 85 bis 98 Tage 99 bis 140 Tage 141 bis 182 Tage mehr als 182 Tage	0 Tage 1 Tag usw. in 1-Tages-Schritten bis 42 Tage 43 bis 46 Tage 47 bis 53 Tage 54 bis 60 Tage 61 bis 67 Tage 68 bis 77 Tage 78 bis 91 Tage 92 bis 119 Tage 120 bis 161 Tage mehr als 161 Tage
Alter	0 bis unter 5 Jahre 5 bis unter 10 Jahre usw. in 5-Jahres-Schritten bis unter 95 Jahre 95 Jahre und älter	0 bis unter 3 Jahre 3 bis unter 8 Jahre usw. in 5-Jahres-Schritten bis unter 93 Jahre 93 Jahre und älter
Geschlecht	Männlich Weiblich	*
Patientenwohnort	AGS bis Kreisebene (5 Stellen) Ausland unbekannt	*
Behandlungsort	*	Nord (SH, HH, HB, NI, NW) Süd (BW, BY, HE, RP, SL) Ost (BE, BB, MV, SN, ST, TH)
Fachabteilung mit der längsten Verweildauer	*	Chirurgie Frauenheilkunde und Geburtshilfe Hals-Nasen-Ohren-Heilkunde Innere Medizin Kinderheilkunde Orthopädie Urologie Psychiatrie Sonstige unbekannt
Trägerschaft	*	öffentlich freigemeinnützig privat
Bettengrößenklasse	*	unter 100 Betten** 100 bis unter 200 Betten** 200 bis unter 500 Betten*** 500 und mehr Betten***

* Merkmal im Datensatz nicht enthalten

** für öffentliche Krankenhäuser in Ost werden diese zwei Klassen zu „unter 200 Betten“ zusammengefasst

*** für private Krankenhäuser werden diese zwei Klassen zu „200 und mehr Betten“ zusammengefasst

Quelle: eigene Arbeiten am Anonymisierungskonzept.

Analog zu der betrachteten Analyse kann die Wanderung auch „umgekehrt“ für die Frage „Aus welchen Bundesländern stammen die behandelten Patienten?“ betrachtet werden. Die Prozentanteile aus Sicht der Behandlungsorte sind beispielsweise im Statistischen Jahrbuch (Statistisches Bundesamt 2004: 244 f.) aufgeführt.⁵ Ebenso ist natürlich eine detailliertere Untersuchung, beispielsweise durch Verwendung kleinerer räumlicher Strukturen, denkbar.

Tabelle 2: Verbleib von Patienten im Bundesland im Jahr 2001

Wohnort (Bundesland)	Patient ist im Bundesland verblieben	
	ja	nein
Schleswig-Holstein	86.9 %	13.1 %
Hamburg	90.3 %	9.7 %
Niedersachsen	88.7 %	11.3 %
Bremen	91.1 %	8.9 %
Nordrhein-Westfalen	96.9 %	3.1 %
Hessen	90.1 %	9.9 %
Rheinland-Pfalz	85.9 %	14.1 %
Baden-Württemberg	95.2 %	4.8 %
Bayern	97.0 %	3.0 %
Saarland	92.7 %	7.3 %
Berlin	96.2 %	3.8 %
Brandenburg	83.8 %	16.2 %
Mecklenburg-Vorpommern	95.0 %	5.0 %
Sachsen	95.8 %	4.2 %
Sachsen-Anhalt	92.2 %	7.8 %
Thüringen	93.0 %	7.0 %

Quelle: eigene Forschungsergebnisse (Daten: Forschungsdatenzentrum der Statistischen Landesämter – Krankenhausstatistik, Teil II Diagnosen 2001).

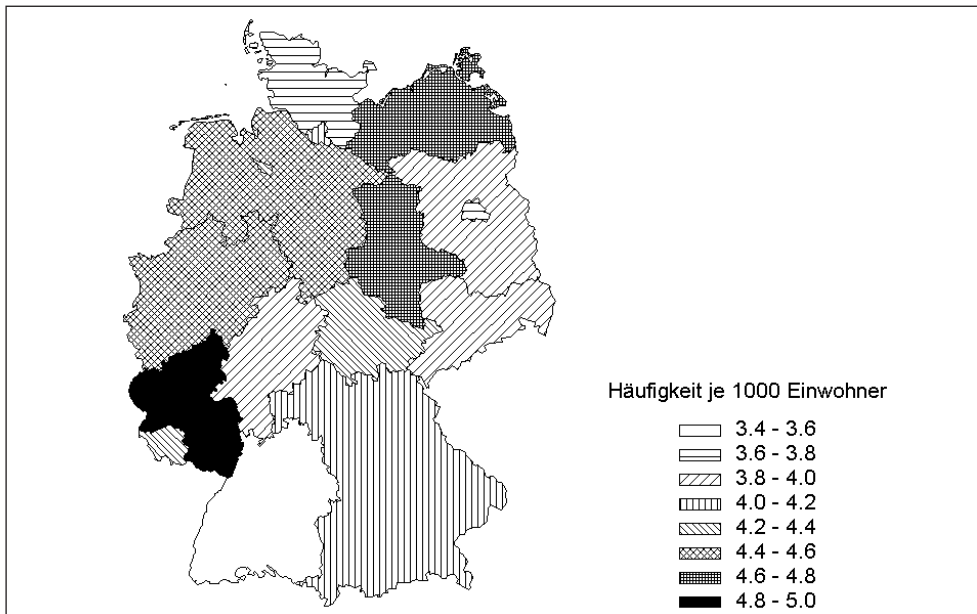
In einer weiteren Analyse werden die Häufigkeiten bestimmter Diagnosen nach dem Wohnort des Patienten bezogen auf die Einwohnerzahl berechnet und kartografisch dargestellt. Die Kartogramme in **Abbildungen 1 bis 3** stellen diese relativen Häufigkeiten in den einzelnen Bundesländern für drei ausgewählte Diagnosekapitel⁶ dar. Die Klassengrenzen in diesen Abbildungen wurden dabei so gewählt, dass die Bandbreite der jeweiligen empirischen Häufigkeiten bestmöglich wiedergegeben wird.

⁵ Alle Analysen im vorliegenden Artikel wurden einschließlich der Stundenfälle durchgeführt. Die Angaben im Statistischen Jahrbuch beziehen sich jedoch auf die Zahl der Behandlungen ohne Stundenfälle.

⁶ Die Diagnosekapitel I bis XXI bilden die oberste Gliederungsstufe der ICD-10.

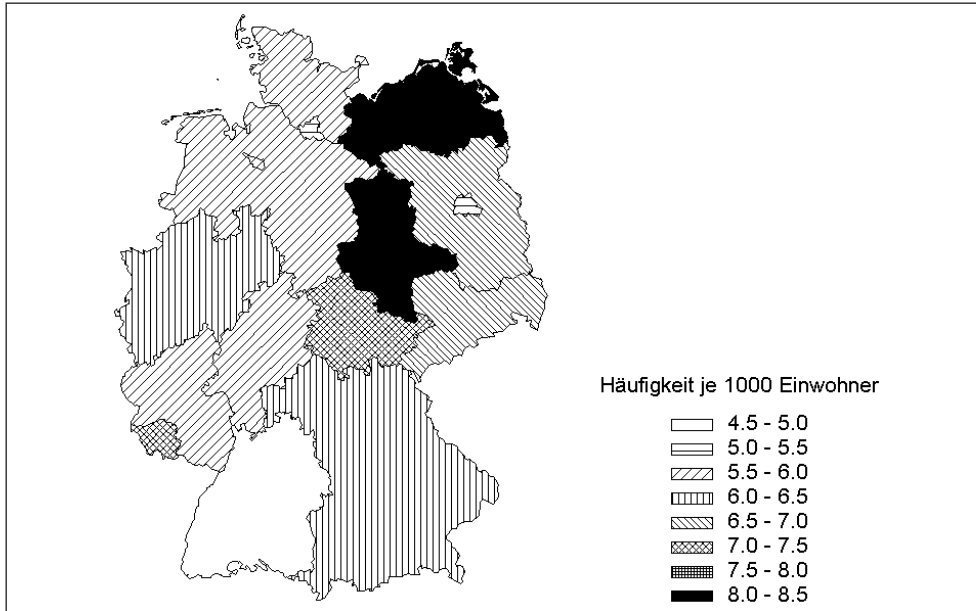
Während es im Diagnosekapitel I (Bestimmte infektiöse und parasitäre Krankheiten) bundesweit keine gravierenden Unterschiede in der Behandlungshäufigkeit gibt, treten im Diagnosekapitel IV (Endokrine, Ernährungs- und Stoffwechselkrankheiten) in den neuen Bundesländern, gemessen an der Bevölkerung, deutlich mehr Krankenhausfälle auf als im übrigen Bundesgebiet. Bei den psychischen und Verhaltensstörungen (Diagnosekapitel V) gibt es eher ein Süd-Nord-Gefälle. Baden-Württemberg zeigt in allen drei untersuchten Kapiteln sehr geringe Behandlungshäufigkeiten.

Abbildung 1: Kartogramm der relativen Fall-Häufigkeiten in Deutschland im Jahr 2001
(Diagnosekapitel I: Bestimmte infektiöse und parasitäre Krankheiten)



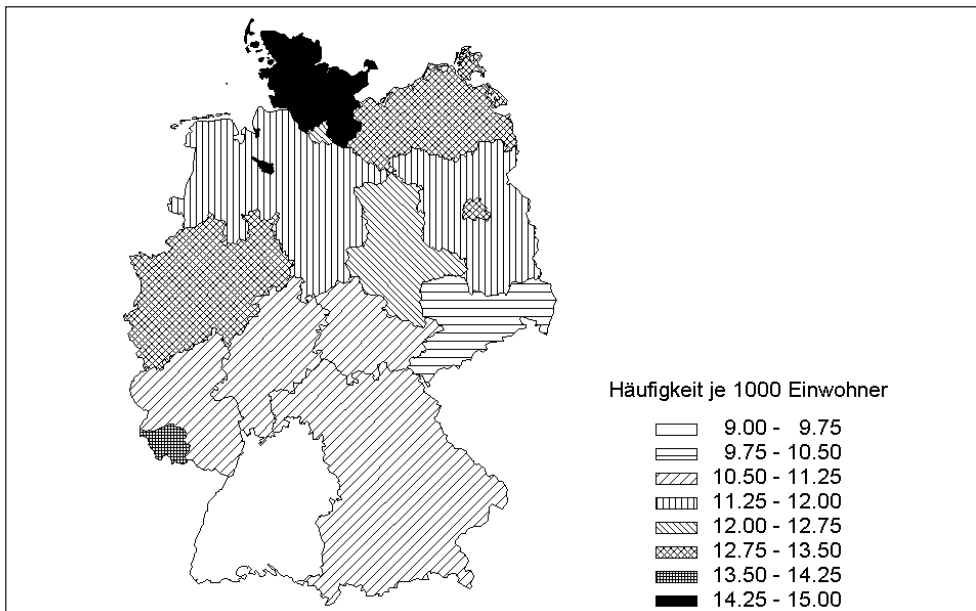
Quelle: eigene Forschungsergebnisse (Daten: Forschungsdatenzentrum der Statistischen Landesämter – Krankenhausstatistik, Teil II Diagnosen 2001).

Abbildung 2: Kartogramm der relativen Fall-Häufigkeiten in Deutschland im Jahr 2001
(Diagnosekapitel IV: Endokrine, Ernährungs- und Stoffwechselkrankheiten)



Quelle: eigene Forschungsergebnisse (Daten: Forschungsdatenzentrum der Statistischen Landesämter – Krankenhausstatistik, Teil II Diagnosen 2001).

Abbildung 3: Kartogramm der relativen Fall-Häufigkeiten in Deutschland im Jahr 2001
(Diagnosekapitel V: Psychische und Verhaltensstörungen)

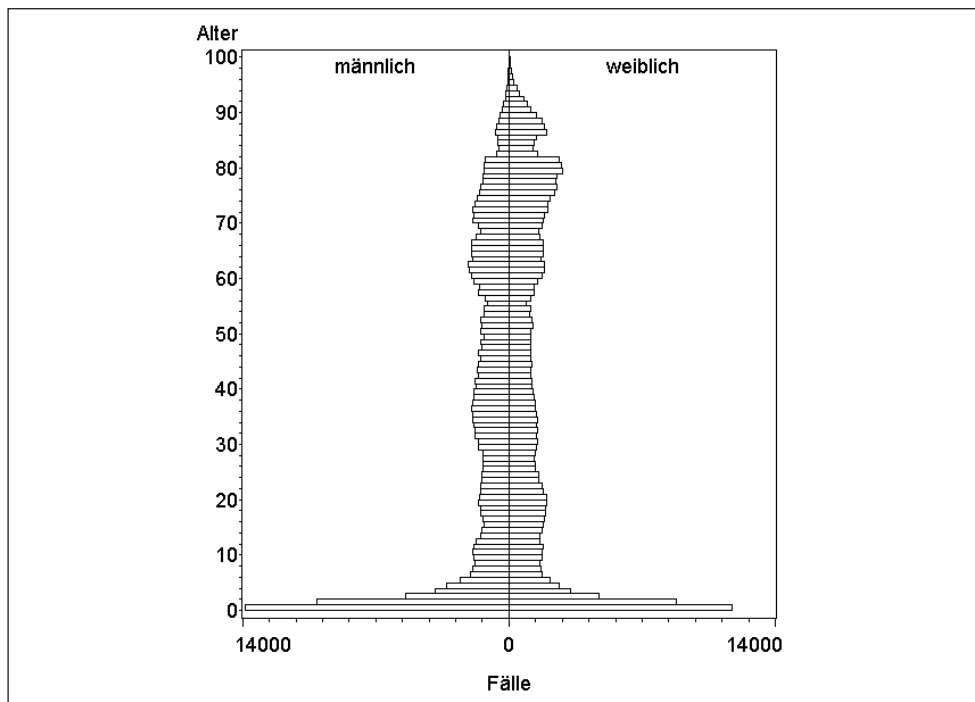


Quelle: eigene Forschungsergebnisse (Daten: Forschungsdatenzentrum der Statistischen Landesämter – Krankenhausstatistik, Teil II Diagnosen 2001).

Die letzte hier vorgestellte Analyse befasst sich mit der Altersstruktur der behandelten Patienten.⁷ Dabei ist jedoch die erwähnte Fallbezogenheit der Diagnosestatistik zu berücksichtigen – Patienten, welche im Jahr 2001 mehrfach in einem oder mehreren Krankenhäusern behandelt wurden, werden mehrfach gezählt. Daher sind die Alterspyramiden in den **Abbildungen 4 bis 6** bezogen auf die Fälle, nicht aber auf die Patienten. Die Auswahl der Diagnosekapitel ist hierbei dieselbe, wie bei der vorangegangenen Analyse.

Im Diagnosekapitel I ist eine Häufung für Säuglinge und Kleinkinder auffällig, da hier oft wegen Durchfallerkrankungen ein Krankenhausaufenthalt erforderlich wird. Bei älteren Patienten ist Wundrose eine häufige Einzeldiagnose.

Abbildung 4: Fallbezogene Altersstruktur von im Jahr 2001 in Krankenhäusern behandelten Patienten in Deutschland (Diagnosekapitel I: Bestimmte infektiöse und parasitäre Krankheiten)

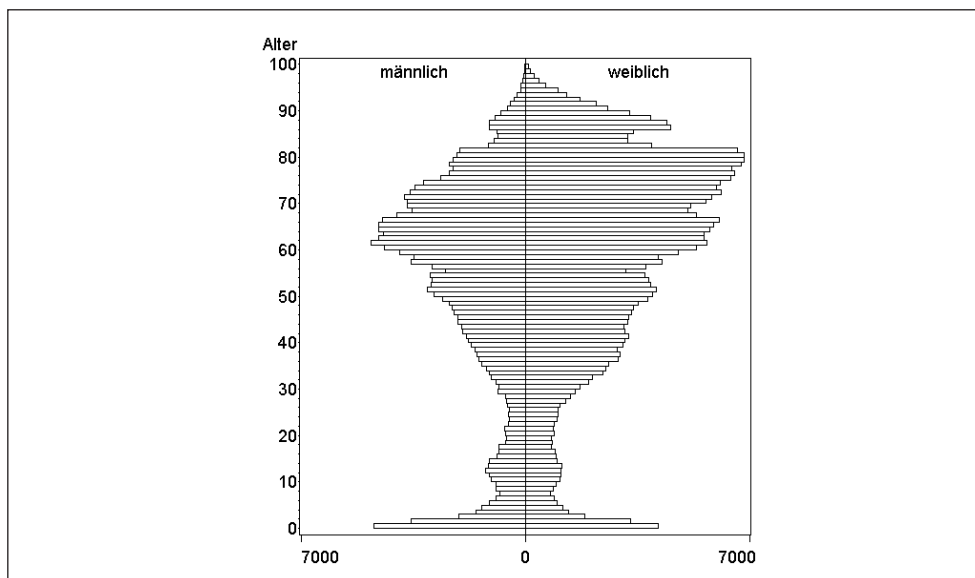


Quelle: eigene Forschungsergebnisse (Daten: Forschungsdatenzentrum der Statistischen Landesämter – Krankenhausstatistik, Teil II Diagnosen 2001).

Säuglinge und Kleinkinder werden auch im Diagnosekapitel IV überdurchschnittlich häufig behandelt, z.B. wegen Volumenmangel (Dehydratation). Bei den älteren Patienten spielen die Krankheitsklassen Diabetes mellitus und Krankheiten der Schilddrüse eine wichtige Rolle.

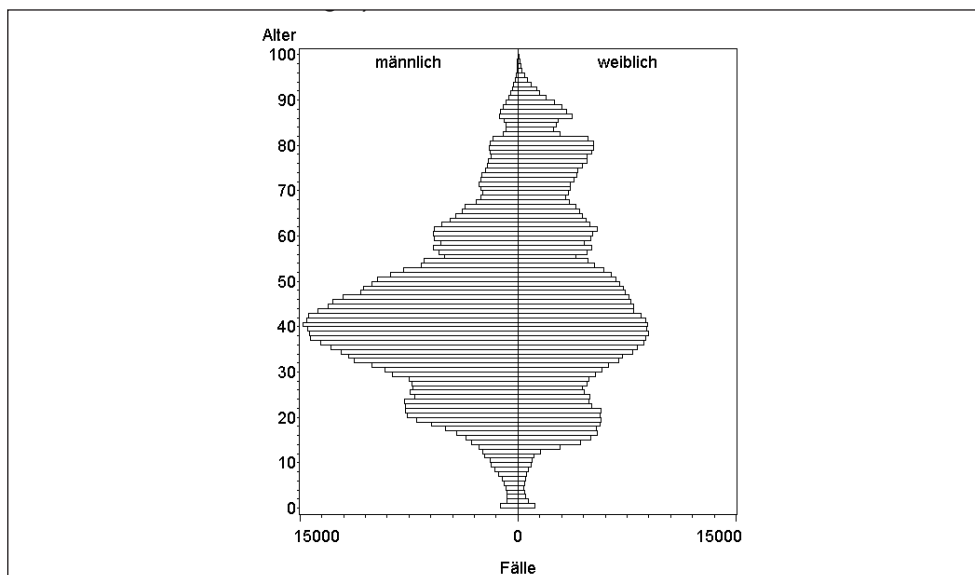
⁷ Eine gleichartige, etwas detailliertere Analyse für 1999 aus Berliner Krankenhäusern entlassene Patienten ist in Bömermann (2001) zu finden.

Abbildung 5: Fallbezogene Altersstruktur von im Jahr 2001 in Krankenhäusern behandelten Patienten in Deutschland (Diagnosekapitel IV: Endokrine, Ernährungs- und Stoffwechselkrankheiten)



Quelle: eigene Forschungsergebnisse (Daten: Forschungsdatenzentrum der Statistischen Landesämter – Krankenhausstatistik, Teil II Diagnosen 2001).

Abbildung 6: Fallbezogene Altersstruktur von im Jahr 2001 in Krankenhäusern behandelten Patienten in Deutschland (Diagnosekapitel V: Psychische und Verhaltensstörungen)



Quelle: eigene Forschungsergebnisse (Daten: Forschungsdatenzentrum der Statistischen Landesämter – Krankenhausstatistik, Teil II Diagnosen 2001).

Bei der Häufung der Krankenhausfälle im Diagnosekapitel IV bei Männern um das 40. Lebensjahr ist die Einzeldiagnose „Psychische und Verhaltensstörungen durch Alkohol“ überproportional häufig vertreten.

Die vorgestellten Analysen sind bewusst nicht allzu detailliert angelegt, um einer tiefergehenden Untersuchung der dargestellten Sachverhalte durch Nutzer des FDZ/L im Rahmen wissenschaftlicher Forschungsprojekte nicht vorzugreifen. Über die verschiedenen Nutzungswege können selbstverständlich auch weitere Fragestellungen bezüglich der Krankenhaus-Diagnosestatistik untersucht werden, beispielsweise:

- Gibt es für bestimmte Diagnosen auffällige Altershäufungen? Was können deren Ursachen und Wirkungen sein?
- Gibt es besondere regionale Auffälligkeiten bestimmter Krankheiten (etwa Häufungen von Krebserkrankungen in der Nähe von Atomkraftwerken oder anderer Erkrankungen in Gebieten mit ärztlicher Unterversorgung)?
- Wie lang ist die mittlere Verweildauer bestimmter Patientengruppen (gegliedert etwa nach Wohnort, Alter oder Diagnose)?
- Welche Änderungen in der Diagnosestellung haben sich durch die Einführung der Abrechnung mittels DRG ergeben?

Fazit

Mit dem geplanten Scientific Use File stehen der Wissenschaft erstmals faktisch anonymisierte Mikrodaten der Krankenhaus-Diagnosestatistik zur Off-Site-Nutzung zur Verfügung. Auf Grund der Zusammenarbeit mit dem wissenschaftlichen Beraterkreis bei der Erstellung des Anonymisierungskonzeptes eignet es sich für einen großen Teil möglicher wissenschaftlicher Datenanalysen bezüglich der Diagnosen von Krankenhauspatienten. Für Mikrodaten der Krankenhaus-Diagnosestatistik stehen darüber hinaus als Zugangswege die kontrollierte Datenfernverarbeitung sowie die On-Site-Nutzung am Gastwissenschaftler-Arbeitsplatz zur Verfügung.

Im Forschungsprojekt zur beispielhaften Analyse der Diagnosedaten wurde das Wanderungsverhalten von Krankenhauspatienten und die relative Häufigkeit bestimmter Erkrankungen gegliedert auf Ebene der Bundesländer sowie die Altersstruktur der Patienten untersucht. Das Projekt soll einerseits das Interesse zur Nutzung von Daten der Diagnosestatistik für wissenschaftliche Zwecke fördern und eine Vorstellung von deren Analysepotenzial vermitteln. Andererseits können die Analysen Wissenschaftlern als Vorlage bei der Formulierung eigener Fragestellungen dienen.

Abkürzungsverzeichnis

AGS	- Amtlicher Gemeindegchlüssel
BB	- Brandenburg
BE	- Berlin
BStatG	- Bundesstatistikgesetz
BW	- Baden-Württemberg
BY	- Bayern
DIMDI	- Deutsches Institut für medizinische Dokumentation und Information
DRG	- Diagnosis Related Groups
DS I	- Patientenbezogener Datensatz des Scientific Use Files
DS II	- Krankenhausbezogener Datensatz des Scientific Use Files
FDZ/L	- Forschungsdatenzentrum der Statistischen Landesämter
HB	- Bremen
HE	- Hesse
HH	- Hamburg
ICD-10	- 10. Revision der Internationalen Klassifikation der Krankheiten
ICD-9	- 9. Revision der Internationalen Klassifikation der Krankheiten
MV	- Mecklenburg-Vorpommern
NI	- Niedersachsen
NW	- Nordrhein-Westfalen
RP	- Rheinland-Pfalz
SH	- Schleswig-Holstein
SL	- Saarland
SN	- Sachsen
ST	- Sachsen-Anhalt
TH	- Thüringen

Literaturverzeichnis

- Bömermann, H.* (2001): Die Geschlechts- und Altersabhängigkeit des stationären Behandlungsaufkommens in den Berliner Krankenhäusern, *Berliner Statistik – Monatsschrift*, 8/01, 348-353.
- DIMDI* (Hg.) (2000): Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10), Version 2.0, Systematisches Verzeichnis. Kohlhammer: Stuttgart.
- Scheffler, M.* (2005): Ein Scientific-Use-File der Einzelhandelsstatistik 1999, *Wirtschaft und Statistik*, 3/2005, 197-200.
- Statistisches Bundesamt* (Hg.) (2004): Statistisches Jahrbuch 2004 für die Bundesrepublik Deutschland. Statistisches Bundesamt: Wiesbaden.
- Zühlke, S. & U. Hetke*, (2003): Datenbedarf und Datenzugang: Ergebnisse der ersten Nutzerbefragung des Forschungsdatenzentrums der Statistischen Landesämter, *Allgemeines Statistisches Archiv*, Band 87, 321-334.
- Zühlke, S., M. Zwick, S. Scharnhorst & T. Wende* (2004): The research data centres of the Federal Statistical Office and the statistical offices of the Länder, *Schmollers Jahrbuch*, Heft 4/2004, 567-578.

Gabriele Philipp, Dipl.-Ing. für Verkehrsingenieurwesen/Logistik, Studium und Forschungsstudium an der Verkehrshochschule „Friedrich List“ in Dresden (1984-1992), seit 1993 Referentin im Statistischen Landesamt Sachsen, seit 2004 zuständig für den Bereich Gesundheitsstatistik.

Dr. Olaf Schoffer studierte Statistik an der Universität Dortmund und promovierte im Dortmunder Graduiertenkolleg „Angewandte Statistik“. Er war wissenschaftlicher Mitarbeiter des Lehrstuhls Statistik und Ökonometrie am Fachbereich Statistik der Universität Dortmund. Seit September 2004 betreut er als Referent im Statistischen Landesamt Sachsen den regionalen Standort Kamenz des FDZ der Statistischen Landesämter.

Das **Forschungsdatenzentrum der Statistischen Landesämter** ist eine Arbeitsgemeinschaft der Statistischen Ämter der Länder mit regionalen Standorten. Es arbeitet eng mit dem Forschungsdatenzentrum des Statistischen Bundesamtes zusammen, um für die Wissenschaft ein abgestimmtes Dienstleistungsangebot der amtlichen Statistik zu entwickeln. Am regionalen Standort Kamenz im **Statistischen Landesamt des Freistaates Sachsen** werden Statistiken des Sozial- und Gesundheitswesens aufbereitet. Ausführliche Informationen zu den beiden Forschungsdatenzentren, zum Datenangebot und zu den Nutzungsmöglichkeiten sind im Internet unter www.forschungsdatenzentrum.de zu finden.