



Universität Potsdam



Hannes-Friedrich Ulbrich

Die Krankenhäuser Ostdeutschlands in Transition

Eine registerbasierte Analyse amtlicher Paneldaten

Potsdamer Schriften zu Statistik und Wirtschaft | 4
Hans Gerhard Strohe (Hrsg.)

Potsdamer Schriften zu Statistik und Wirtschaft
Hans Gerhard Strohe (Hrsg.)

Potsdamer Schriften zu Statistik und Wirtschaft | 4
Hans Gerhard Strohe (Hrsg.)

Hannes-Friedrich Ulbrich

Die Krankenhäuser Ostdeutschlands in Transition

Eine registerbasierte Analyse amtlicher Paneldaten

Universitätsverlag Potsdam

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.de/> abrufbar.

Universitätsverlag Potsdam 2012

<http://verlag.ub.uni-potsdam.de/>

Am Neuen Palais 10, 14469 Potsdam
Tel.: +49 (0)331 977 2533 / Fax: 2292
E-Mail: verlag@uni-potsdam.de

Die Schriftenreihe **Potsdamer Schriften zu Statistik und Wirtschaft** wird herausgegeben von Prof. Dr. em. Hans Gerhard Strohe.

ISSN (print) 2192-8061
ISSN (online) 2192-807X

Das Manuskript ist urheberrechtlich geschützt.
Umschlagbild: Hannes-Friedrich Ulbrich

Online veröffentlicht auf dem Publikationsserver der Universität Potsdam:
URL <http://pub.ub.uni-potsdam.de/volltexte/2012/6040/>
URN <urn:nbn:de:kobv:517-opus-60408>
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-60408>

Zugleich gedruckt erschienen im Universitätsverlag Potsdam
ISBN 978-3-86956-200-1

Kurzfassung

Vorliegende Arbeit untersucht – nach intensiver Datenanalyse und -klärung – die ersten eineinhalb Jahrzehnte Transition in Ostdeutschland ökonomisch. Im Mittelpunkt stehen dabei die Interaktionen zwischen der allgemeinen Wirtschaft und den Krankenhäusern sowie – auf Basis der Hauptdiagnose – den in Krankenhäusern behandelten Krankheiten. Verschiedene, in den politischen Diskussionen zur Effizienz der Krankenhäuser übliche Indikatoren werden verglichen und auf ihre Eignung zur adäquaten Beschreibung von Krankenhauskosten und -leistungen geprüft. Durch Einbeziehen der Daten aus Rheinland-Pfalz wird herausgearbeitet, welche der Veränderungen in Ostdeutschland transitionsbezogen sind und ob und wie sich die transitionsbezogenen von eher transitionsunabhängigen Veränderungen trennen lassen. Dazu werden den Daten – jährlich erhobenen Angaben zu einem jeden Krankenhaus bzw. jedem Kreis des Untersuchungsgebietes – Paneldatenmodelle angepasst. Ergänzt um deskriptive Statistiken, Graphen und Choroplethenkarten sind diese Modelle Grundlage einer wirtschaftswissenschaftlichen Interpretation der Veränderungen zwischen 1992 und 2005.

Größten Einfluss auf die Krankenhäuser in Ost- wie in Westdeutschland haben die sich verändernden Rahmenbedingungen, zuvörderst die etappenweise Ablösung einer budgetbasierten Krankenhausfinanzierung durch eine, die auf Fallpauschalen für die verschiedenen schweren Krankheitsfälle (per diagnose related groups – DRG) basiert. Bereits die um das Jahr 2000 unter den Ärzten und Krankenhausmanagern beginnende Diskussion um die Fallpauschalen führt zu erheblichen Veränderungen bei der ärztlichen Diagnosestellung (bzw. -kodierung), diese Veränderungen werden diskutiert. Vor allem den Besonderheiten der Finanzierung des deutschen Gesundheitswesens ist es geschuldet, dass die Transition für die ostdeutschen Krankenhäuser sehr schnell zu einem den westlichen Krankenhäusern ähnlichen Leistungs- und Kostenniveau führt und dass mikroökonomische Unterschiede in Wirtschafts- und Sozialparametern kaum Einfluss auf die Krankenhäuser haben.

JEL-Klassifizierung: C23, C41, C51, H44, H51, I11, I18, P36

Schlagnworte: Transition, Ostdeutschland, Krankenhäuser, Diagnosen, Panelanalysen, Amtliche Statistik

Abstract

This research investigates the first 15 years of East Germany's transition from an econometric perspective. The primary interest of the investigation lies with the potential interactions between the overall economic situation and the hospitals as well as all the hospital treated patients. Economic and legislative discussions of hospital efficiency are heavily based on economic indicators. Common indicators are compared for their usefulness in describing hospital efforts and costs. By incorporating data of one of the West German federal states, Rhineland-Palatinate, distinctions can be made between transition and non-transition related changes. Panel data regression models are applied to yearly hospital and district economic data. Their interpretation is supported by descriptive statistics, graphs and choropleth maps.

The hospitals of both East and West Germany are heavily influenced by the changes of the political and economic environment — most particularly by the transition from the budget-based financing to the lump compensation (DRG) based reimbursement. The data reveals that the year 2000 marks both the beginning of hospital physicians' and managers' discussion of the upcoming DRG system as well as the beginning of major changes in diagnosing (or diagnosis coding) — these changes are discussed. The most prominent reason for East Germany's hospitals having minimized the gap to their West German counterparts quite quickly with regard to performance and costs is due to the financing particularities of the German health care system. Microeconomic differences between districts have only minor influence on the hospitals.

JEL Classification: C23, C41, C51, H44, H51, I11, I18, P36

Keywords: Transition, East Germany, Hospitals, Diagnoses, Panel analyses, Official Statistics

Vorwort

Während meiner langjährigen Tätigkeit als Biostatistiker sowohl im akademischen als auch im Industrie-Umfeld wandelte sich mein Aufgabenspektrum – mal fließend, mal abrupt – und überdeckte Fragestellungen der Epidemiologie¹, der klinischen Entwicklung und – zunehmend – der präklinischen Forschung und Entwicklung sowie der Produktion von Pharmazeutika. Zu bearbeiten waren und sind die Aufgaben mit des Statistikers Expertise.

Auffällig bei (fast) allen diesen Arbeiten ist der Bezug zwischen der biologisch-medizinischen, pharmazeutischen oder epidemiologischen Fragestellung und den ökonomischen Rahmenbedingungen bzw. den wirtschaftlichen Implikationen der Antworten. Krankheiten, Medikationen, Heilungsaktivitäten und Prävention verursachen umfangs- und qualitätsabhängig Kosten, die Möglichkeiten, für diese Kosten aufzukommen, hängen auch von den ökonomischen Rahmenbedingungen ab. Während in Forschung und Entwicklung vor allem geplante Experimente Grundlage der Entscheidungen sind, sind in der Epidemiologie häufig Beobachtungsdaten die Basis für erkenntnis-, interpretations- und entscheidungsunterstützende Modellierungen.

Am Lehrstuhl Statistik und Ökonometrie der Universität Potsdam unter der Leitung von Herrn Prof. Dr. H. G. Strohe nahm von Beginn an die Modellierung ökonomischer Zusammenhänge einen wichtigen Platz ein (Strohe, 1993). Intensiv beschäftigte man sich mit der mikroökonomischen Sicht auf die Transition Ostdeutschlands (Strohe, 1999). Dabei pflegte der Lehrstuhl intensiven Kontakt zu den StLÄ der Länder Brandenburg und Berlin² sowie dem Forschungsdatenzentrum der Statistischen Landesämter und seinen lokalen Standorten. Reichhaltige Erfahrungen bestanden besonders bei der Bearbeitung von Wirtschaftsstatistiken

¹ Epidemiologie: wissenschaftliche Disziplin, die sich mit den Ursachen und Folgen sowie der Verbreitung von gesundheitsbezogenen Zuständen und Ereignissen in Populationen beschäftigt.

² seit 1. Januar 2007 vereinigt zum Amt für Statistik Berlin-Brandenburg.

(Strohe, 2000; Rambert und Strohe, 2001; Rambert, 2004; Mangelsdorf, 2007; Dietrich und Strohe, 2011a). Dietrich und Strohe (2011b) schenken dabei der Qualität der Daten besonders große Aufmerksamkeit.

Obwohl meine Idee, die Beziehungen zwischen Gesundheit und Wirtschaft und deren Veränderungen während der Transition empirisch zu beleuchten, eine deutliche Erweiterung der Forschungsthemen des Lehrstuhls darstellte, war man bereit, mich in meinem Vorhaben zu unterstützen.

Nach nunmehr 20 Jahren Transition Ostdeutschlands ist in Deutschland viel von den Veränderungen im Gesundheitswesen und besonders den Krankenhäusern die Rede, doch spezifische Veränderungen des ostdeutschen Gesundheitssystems bzw. der ostdeutschen Krankenhäuser werden kaum reflektiert, wie man aktuell in Neubauer und Beivers (2010), dem einführenden Kapitel zum Krankenhaus-Report 2010 (Klauber et al., 2010), sehen kann. Das geplante Vorhaben erschien lohnend.

Eine solche Arbeit bewältigt man nicht allein. Ich habe viel und unterschiedlichste Hilfe erfahren und möchte dafür herzlich danken. Besonderer Dank gebührt dabei meiner Schwester Maren für den entscheidenden Motivationsschub; Dres. Katrin und Christoph Gerlinger für das Anbahnen der Beziehungen zum Lehrstuhl Strohe und das unermüdliche Drängen nach sichtbarem Fortschritt; Herrn Prof. Strohe für seine Offenheit dem Thema gegenüber und seine fordernd fördernde Betreuung; Herrn Prof. Rasche für die befruchtenden Diskussionen; den Seminarteilnehmern am Lehrstuhl Strohe für die Diskussionsbereitschaft und die Geduld mit mir und meinen Ideen; Herrn Dr. Schoffer³ vom Forschungsdatenzentrum der Statistischen Landesämter, Standort Kamenz für sein unermüdliches Engagement, mir sehr viel seiner großen Expertise zur Krankenhausstatistik, den Vorgängen im Amt und möglichen Software-Lösungen weit jenseits der Grundausstattung zur Verfügung zu stellen. Und ich danke vor allem meiner Frau Christiane, ihre Inspiration und liebevoll-unendliche Geduld hat mich dieses Projekt auch in kritischen Phasen weiter verfolgen lassen.

Potsdam & Berlin, im Juli 2012

³ seit 2011 am Universitäts KrebsCentrum an der TU Dresden.

Inhaltsverzeichnis

Vorwort	i
Abkürzungsverzeichnis	xiii
Einleitung	1
1. Einführung: Transformationsprozesse Ostdeutschlands	5
1.1. Transition aus politikwissenschaftlicher Sicht	6
1.2. Transition aus Ökonomen-Sicht	7
1.3. Transition aus ökonometrischer Sicht	7
1.4. Sonderfall Transition Ostdeutschlands	8
1.5. Transition, ihre Dauer und parallele Veränderungen in Deutschland	10
2. Fragestellungen, Einordnung dieser Arbeit	11
2.1. Zielsetzung	14
2.2. Fragestellungen	14
2.3. Vergleichsland Rheinland-Pfalz	16
2.4. Berlin – Teil des Ostens?	18
2.5. Einordnung dieser Arbeit	19
3. Stand der Forschung	21
3.1. Definitionen und Begriffsabgrenzungen	21
3.1.1. Gesundheitssysteme als Forschungsgegenstand	25
3.1.2. Gesundheit und Wirtschaft, Wirtschaft und Gesundheit	26
3.1.3. Gesundheitssysteme in Europa und ihre Finanzierung	29
3.1.4. Transitionen von Gesundheitssystemen in Europa	31
3.1.5. Inlandskonzept und Inländerkonzept	33
3.2. Gesundheitswesen in Deutschland 1991–2005	33
3.2.1. Gesundheitsreformen	34
3.2.2. Krankenhäuser in Deutschland	36

3.2.3.	Der Markt eines (einzelnen) Krankenhauses	38
3.2.4.	Leistung und Leistungsmessung von Krankenhäusern . . .	41
3.2.5.	Die kumulative Krankenhausleistung eines Gebietes . . .	42
3.2.6.	Besonderheiten des Gesundheitsmarktes	42
3.3.	Themenbezogene Daten aus Deutschland	44
3.3.1.	Amtliche Daten	44
3.3.2.	Gesundheitsdaten	45
3.3.3.	Sozialdaten	48
3.3.4.	Wirtschaftsdaten	49
3.3.5.	Raumordnungsdaten	50
3.4.	Statistik in den empirischen Sozialwissenschaften	52
3.4.1.	Empirische Untersuchungen in den Sozialwissenschaften .	53
3.4.2.	Statistik für empirische Untersuchungen amtlicher Paneldaten	54
3.4.3.	Totalerhebungen und Modellierungen	55
3.4.4.	Paneldaten: Vor- und Nachteile	56
3.5.	Register und registerbasierte Statistik	59
3.6.	Analysen amtlicher Daten zum Gesundheitswesen	61
3.6.1.	Analysen der Krankenhausstatistik	61
3.6.2.	Quellenübergreifende Analysen	63
3.6.3.	Andere Aspekte der Transition Ostdeutschlands	64
3.7.	Stand der Forschung – Zusammenfassung	64
4.	Daten und Registererstellung	67
4.1.	Datenquellen und Datenzugang	68
4.1.1.	Amtliche Daten für die Wissenschaft: Bereitstellung durch die FDZ	69
4.1.2.	Anonymisierungen, GWAP und KDFV	71
4.1.3.	Daten aus dem FDZ/L – Krankenhausstatistik	72
4.1.4.	Weitere Daten aus dem Bereich der Amtlichen Statistik . .	74
4.1.5.	Primär- und Sekundärdaten	78
4.2.	Krankenhausstatistik – Inhalte und Beschreibung	79
4.2.1.	Beschreibung der Krankenhausstatistik	79
4.2.2.	Zur Abdeckung der Krankenhausstatistik	81
4.2.3.	Nicht verfügbare Merkmale und Informationen	82
4.2.4.	Veränderte Informationen – räumliche und zeitliche Vergleichbarkeit	84

4.2.5.	Auswahl der Daten	90
4.3.	Qualität der Statistiken	91
4.3.1.	Kriterien der Qualitätsbeurteilung von Statistiken	91
4.3.2.	Empirische Einschätzung der Datenqualität: Daten im offenen Zugriff	94
4.3.3.	Empirische Einschätzung der Datenqualität: Krankenhausstatistik	96
4.4.	Erstellung des Registers, Datenmodifikationen	106
4.4.1.	Struktur des Registers, Identifikatoren	107
4.4.2.	Aggregationen	109
4.4.3.	Datenmodifikationen	110
4.4.4.	Das Register unter GWAP-Bedingungen	114
4.5.	Daten und Register – Zusammenfassung	114
5.	Empirische Analysen: Methoden und Umsetzung	117
5.1.	Statistische Methoden und deren Quellen	117
5.1.1.	Aggregationen	118
5.1.2.	Kreise oder Krankenhäuser – zur statistischen Einheit der Analysen	118
5.1.3.	Klumpenbildung: klein aber ähnlich	120
5.1.4.	Signifikanz und Relevanz	124
5.1.5.	Regressionen	124
5.1.6.	Analysen von Verweildauern	135
5.1.7.	Modellierung unter KDFV-Anonymisierungsforderungen	138
5.1.8.	Verzerrungen und Modellfehlspezifikationen	139
5.1.9.	Exogene und endogene Variable und ihre Bedeutung für die Modellierung	140
5.1.10.	Räumliche Statistik – Landkarten	141
5.1.11.	Kompositionsdaten	144
5.1.12.	Visualisierung von Kompositionsdaten	148
5.1.13.	Das Kompositionsabweichungsdiagramm	151
5.2.	Programmtechniken und Algorithmen	153
5.2.1.	Modularer Aufbau	154
5.2.2.	Mantelskripte	154
5.2.3.	Einzelkripte	156
5.2.4.	Makros	157
5.2.5.	Register, SAS®-Datensätze und Skripte	157

5.2.6.	SAS®-Module und -Zusätze	158
5.2.7.	Anonymisierungsforderungen und Ergebnisdarstellungen	159
5.3.	Methoden und Umsetzungen – Zusammenfassung	160
6.	Empirische Analysen: Ergebnisse	161
6.1.	Deskriptionen	162
6.1.1.	Demografie – die Bevölkerung	162
6.1.2.	Wirtschafts- und Sozialparameter	163
6.1.3.	Grund- und Kostendaten – der Aufwand	169
6.1.4.	Diagnose-Statistik (Fälle) – die Leistung	178
6.2.	Ökonometrische Modellierungen	184
6.2.1.	Verweildauern und Hazardfunktionen	185
6.2.2.	Die Aufnahme vollstationärer und teilstationärer Fälle	187
6.2.3.	Verweildauer-Modelle	189
6.2.4.	Diagnose-Stellung und -Registrierung	193
6.2.5.	Wohn- und Behandlungsort	197
6.2.6.	Beschäftigte in den Krankenhäusern	200
6.2.7.	Bettendichte und Belegungsdichte	205
6.2.8.	Kosten der einzelnen Krankenhäuser	210
6.2.9.	Krankenhaus-Kosten in den einzelnen Kreisen	221
6.3.	Die Modellergebnisse aus ökonometrischer Sicht	226
6.3.1.	Krankenhäuser und Wirtschaft	226
6.3.2.	Wirtschaftlichkeit bei der Behandlung der Patienten	237
6.4.	Empirische Analysen – Zusammenfassung	244
7.	Antworten und Ausblick	249
7.1.	Arbeiten unter GWAP- und KDFV-Bedingungen	249
7.2.	Antworten	251
7.3.	Methoden-Entwicklungen	254
7.4.	Ausblick	255
A.	Anhänge	257
A.1.	Merkmalsliste Krankenhausstatistik	257
A.2.	Programmbausteine	260
A.3.	Tabellen	269
A.4.	Graphen	271
A.4.1.	Altersverteilungen	271

A.4.2. Monatliche Aufnahmezahlen vollstationärer und teilstationärer Fälle	273
Literatur	275

Abbildungsverzeichnis

2.1.	Veränderung der Einwohnerzahlen in den 16 Bundesländern . . .	12
2.2.	Krankenhausdichte in den 16 Bundesländern	12
2.3.	Bettendichte in den 16 Bundesländern	13
2.4.	Pro-Kopf Transferleistung des Länderfinanzausgleichs (2004) . . .	17
2.5.	Untersuchungsgebiet (mit Kreisgrenzen) in Deutschland	19
3.1.	Anzahl der Krankenhäuser in Deutschland (2005)	37
3.2.	Gesundheitswesen: Nachfrage, Angebot, Finanzierung und staatliche Eingriffe	43
3.3.	Siedlungsstrukturelle Gebietstypen des BBR	51
4.1.	Regionsgrundtypen und Kreistypen des BBR im Untersuchungsgebiet (2005)	76
4.2.	Brutto- und Nettoprinzip	85
4.3.	Diagnosen des ICD-9-Kapitels XII: Krankheiten der Haut (1993–2005)	99
5.1.	Abstandsmaßwerte der Klumpenbildung	123
5.2.	Gestapelte Balkendiagramme von Kompositionen	150
5.3.	Überlagerte Darstellung von Kompositionen	151
5.4.	Kompositionsabweichungsdiagramm	152
5.5.	Programmablauf	155
6.1.	Urbanität der Kreise des Untersuchungsgebiets (2005)	163
6.2.	Arbeitslosenquoten (1996, 2000, 2004 und 2005)	164
6.3.	Brutto-Erwerbstätigenquoten (1992, 1996, 2001 und 2005)	165
6.4.	Verfügbares Haushaltseinkommen je Einwohner (1995, 1998, 2002 und 2005)	166
6.5.	Arbeitnehmereinkommen je Arbeitnehmer (1996, 1999, 2002 und 2005)	167

6.6.	Brutto-Inlandsprodukt je Einwohner (1992, 1996, 2001 und 2005)	168
6.7.	Bettendichte, Belegbettendichte (1992–2005)	174
6.8.	Relative bereinigte Krankenhaus-Kosten (1992–2005)	175
6.9.	Personal-Sachkosten-Verhältnis (1992–2005)	176
6.10.	Frauenanteil am Krankenhauspersonal (1992–2005)	177
6.11.	Fallzahlen Ostdeutschland & Rheinland-Pfalz (1993–2005)	178
6.12.	Ereigniszeitverteilung zur Verweildauer (1993–2005)	183
6.13.	Hazardfunktion zur Verweildauer (1993–2005)	186
6.14.	Zeitreihen monatlicher Aufnahmezahlen (01/1993–10/2005)	188
6.15.	Diagnosen, ICD-9-Kapitel I–XVII (1993–2005)	193
6.16.	Diagnosen verschiedener Diagnose-Gruppierungen (1993–2005)	194
6.17.	Diagnosen der Diagnose-Gruppierung 650–669 (1993–2005)	195
6.18.	Krankenhausärztinnen (1992, 1997, 2002 und 2005)	204
A.1.	Merkmalsliste	259
A.2.	Geschlechts- und Altersverteilung (1993, 2005)	271
A.3.	Altersverteilung Entbindungen (1993, 2003)	272
A.4.	Altersverteilung ICD-9-Kapitel XV (1993, 2000, 2002, 2005)	272
A.5.	Saisonkomponenten monatlicher Aufnahmezahlen (01/1993–10/2005)	273

Tabellenverzeichnis

4.1. Registrierte Krankenhaus-Fälle (1993–2005)	98
4.2. Identifizierte Krankenhaus-Einträge in der Krankenhausstatistik (1992–2005)	103
5.1. Kompositionsdaten-Beispiel	149
6.1. Krankenhäuser je Bundesland (1992–2005)	170
6.2. Krankenhausanzahlen in den Kreisen (1992–2005)	171
6.3. Anteil Plankrankenhäuser (1992–2005)	172
6.4. Trägerschaft der Krankenhäuser (1992–2005)	173
6.5. Fallstatistiken: Anzahl und empirische Altersverteilung (1993–2005)	180
6.6. Verweildauermodellierung (Diagnose-Gruppierung 001–009) . . .	190
6.7. Modellierung der Anzahl der Ärzte (1992–2005)	202
6.8. Modellierung des Anteils der Ärztinnen (1992–2005)	203
6.9. Modellierung der Bettendichte (1992–2005) – Modell 1	206
6.10. Modellierung der Bettendichte (1992–2005) – Modell 2	207
6.11. Modellierung der Belegungsdichte (1992–2005) – Modell 1	208
6.12. Kostenmodellierung (Krankenhäuser) #1, Parameter	212
6.13. Kostenmodellierung (Krankenhäuser) #1, Prädiktionen	213
6.14. Kostenmodellierung (Krankenhäuser) #2, Parameter	214
6.15. Kostenmodellierung (Krankenhäuser) #2, Prädiktionen	215
6.16. Personal-Sachkosten-Verhältnis-Modellierung (Krankenhäuser) #1, Prädiktionen	219
6.17. Personal-Sachkosten-Verhältnis-Modellierung (Krankenhäuser) #2, Prädiktionen	220
6.18. Kostenmodellierung (Kreise) #4, Prädiktionen	223
6.19. Personal-Sachkosten-Verhältnis-Modellierung (Kreise) #4, Prädiktionen	224

6.20. Wichtungen #1	230
6.21. Wichtungen #2	231
A.1. Modellierung der Anzahl der nichtärztlichen Krankenhaus-Mitarbeiter (1992–2005)	269
A.2. Modellierung des Anteils der weiblichen an den nichtärztlichen Krankenhaus-Mitarbeitern (1992–2005)	269
A.3. Kostenmodellierung (Krankenhäuser) #3, Parameter	270

Abkürzungsverzeichnis

η	Linearer Prädiktor des jeweiligen Modells (siehe Abschn. 5.1.5: Verallgemeinerte lineare Modelle, Abschn. 5.1.6)
AGS	Amtlicher Gemeindeschlüssel (eine 8-stellige, eindeutig die Gemeinde identifizierende Nummer)
BA	Bundesagentur für Arbeit
BB	Brandenburg (Bundesland)
BBR	Bundesamt für Bauwesen und Raumordnung
BDSG	Bundesdatenschutzgesetz
BE	Berlin (Bundesland)
BerlinFG	Berlinförderungsgesetz (Gesetz zur Förderung der Berliner Wirtschaft)
BIP	Brutto-Inlandsprodukt
BPfIV	Bundespflegeverordnung (Verordnung zur Regelung der Krankenhauspflegesätze)
BStatG	Bundesstatistikgesetz (Gesetz über die Statistik für Bundeszwecke)
CMI	Case Mix Index – Maß für den Aufwand von Krankenhäusern, z. B. CMI von MediCare (USA)
CMW	Case Mix Weight – Maß für den Aufwand von Krankenhäusern (Herr, 2008, Kap. 2.3)
DRG	diagnosis-related group (diagnosebezogene Fallgruppe) – Maß für den Aufwand von Krankenhäusern; G-DRG – Version für Deutschland
ETR	Erwerbstätigenrechnung
FDZ	Forschungsdatenzentrum, Oberbegriff für FDZ/L und FDZ-RV
FDZ/L	Forschungsdatenzentrum der Statistischen Landesämter
FDZ-RV	Forschungsdatenzentrum der Rentenversicherung
GGh	Grundgesamtheit, (Gesamt-)Population
GKV	gesetzliche Krankenversicherung (Deutschland, SGB V)

GMG	GKV-Modernisierungsgesetz (Gesetz zur Modernisierung der gesetzlichen Krankenversicherung)
GSG	Gesundheitsstrukturgesetz (Gesetz zur Sicherung und Strukturverbesserung der gesetzlichen Krankenversicherung)
GSS	Gesundheitssystem
GSW	Gesundheitswesen, synonym gebraucht zu GSS
GWAP	Gastwissenschaftler-Arbeitsplatz (in den Räumen des FDZ)
KDFV	Kontrollierte Datenfernverarbeitung (vom FDZ durchgeführt)
KHEntgG	Krankenhausentgeltgesetz (Gesetz über die Entgelte für voll- und teilstationäre Krankenhausleistungen)
KHG	Krankenhausfinanzierungsgesetz (Gesetz zur wirtschaftlichen Sicherung der Krankenhäuser und zur Regelung der Krankenhauspflegesätze)
KHStatV	Krankenhausstatistik-Verordnung (Verordnung über die Bundesstatistik für Krankenhäuser)
KI	Konfidenzintervall
MV	Mecklenburg-Vorpommern (Bundesland)
Ostdt	Ostdeutschland (die Bundesländer BE, BB, MV, SN, ST und TH gemeinsam betrachtet)
PUF	Public-Use-Files, bereitgestellt von den FDZ
REML	residual maximum likelihood (auch: restricted ...)
RP	Rheinland-Pfalz (Bundesland)
SGB	Sozialgesetzbuch (Deutschland)
SH	Schleswig-Holstein (Bundesland)
SN	Sachsen (Bundesland)
ST	Sachsen-Anhalt (Bundesland)
StBA	Statistisches Bundesamt
StLA	Statistisches Landesamt
SUF	Scientific-Use-Files (bereitgestellt von den FDZ)
TH	Thüringen (Bundesland)
VGR	Volkswirtschaftliche Gesamtrechnung
VPI	Verbraucherpreisindex, relativ breit basierter deutscher Preisindex

Einleitung

Am 3. Oktober 1990 entsteht durch den Beitritt der Länder Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen sowie (bei gleichzeitiger Beendigung des Viermächtestatus) der Stadt Berlin die Bundesrepublik Deutschland in ihrer heutigen Form. Kurz bevor am 11. Januar des darauffolgenden Jahres mit der Übernahme der bis dahin nur in Berlin [West] geltenden Verfassung für das ganze Bundesland Berlin der letzte Schritt zur staatsrechtlichen Einheit Deutschlands vollzogen werden kann (Rott, 2009, S. 428), wird am 1. Januar 1991 in Deutschland die Krankenhaus-Statistik mit ihren Teilen Grund- und Kostendaten verbindlich, zwei Jahre später ergänzt man diese um die Diagnosestatistik.

In den ersten Jahren der deutschen Einheit ändert sich für die ostdeutschen Länder und ihre Bevölkerung – wie bereits in den zwei vorangegangenen Jahren auch – in historisch kurzer Zeit viel: ein Gesellschafts- und Sozialsystem geht unter und wird mittels Übernahme durch ein anderes, von vielen erträumtes, von wenigen erlebtes, in den sogenannten Altbundesländern funktionierendes System ersetzt. Auch auf die Menschen in den Altbundesländern kommen ungeahnte Veränderungen zu, wie z. B. der Solidaritätszuschlag, der von allen deutschen Steuerpflichtigen als Zuschlag zur Einkommensteuer, Kapitalertragsteuer und Körperschaftsteuer aufzubringen ist. Spätestens zu einer Bundestags- oder Europa-Wahl scheint es wieder so, dass immer noch Unterschiede zwischen ‚Ost‘ und ‚West‘ mit größerer Begeisterung thematisiert und analysiert werden als die anderer politisch-geografischer Einteilungen Deutschlands. Und wenig erinnert man sich außerhalb Berlins, dass die westlichen Berliner Stadtbezirke und deren Bewohner sich ganz spezifisch Westberliner Veränderungen stellen müssen, z. B. dem Wegfall von Sondervergünstigungen wie der seit dem Mauerbau gezahlten Berlin-Zulage von zuletzt 8 % des Bruttoeinkommens (Rott, 2009, S. 295; BerlinFG § 28).

Ost und West haben sich einander angenähert: politisch, ökonomisch, sozial. Und doch sind sie weiterhin zu großen Teilen das Ergebnis ihrer jeweiligen Geschichte. Alle 30 DAX-Unternehmen haben ihren Sitz in den Altbundesländern (Stand November 2010). Unterschiede in der Inzidenz bestimmter Infektionskrankheiten (Hellenbrand, 2007) werden auf die nach wie vor unterschiedliche Impfquote (gründend auch auf der Impfbereitschaft) der Bevölkerung zurückgeführt, selbst die Verfügbarkeit der entsprechenden Daten ist wegen einer an den DDR-Status angelehnten Gesetzeslage in den fünf ostdeutschen Flächenländern eine bessere als für andere Teile Deutschlands.

Gleichzeitig gibt es Veränderungen in Deutschland, die nicht unbedingt vereinigungsbedingt zu interpretieren sind. So wird aktuell (z. B. Diagramm in Rudzio und Uchatius, 2010) festgestellt, dass der Anteil psychischer Erkrankungen unter den Ursachen für die Frühverrentung von etwa 15 % 1993 auf 37,7 % im Jahr 2009 zugenommen hat und dieses mit einer allgemeinen Zunahme von Stress im Arbeitsumfeld erklärt. Die Bundesvereinigung der Deutschen Arbeitgeberverbände⁴ merkt an, dass der Rentenzugang wegen psychischer Erkrankungen in absoluten Zahlen hingegen zwischen 1997 und 2007 stabil geblieben sei, und weist damit auf eine andere Interpretationsmöglichkeit vorliegender Statistik hin.

Die Ost-West-Annäherung wird wissenschaftlich aus vielen Blickwinkeln heraus analysiert. Vieles ist bereits geschrieben zur ökonomischen Transition, einiges zur Transition des Gesundheitssystems und eher wenig zu den wechselseitigen Beeinflussungen von wirtschaftlicher Lage und Gesundheit im Verlauf der Transition.

Empirische Analysen basieren auf Daten, bereits verfügbaren oder für die jeweilige Untersuchung zu gewinnenden Daten. Die Amtliche Statistik als (in Demokratien der Neuzeit) allgemein verfügbares Instrument, das die für die politische Willensbildung in der Gesellschaft notwendigen statistischen Informationen liefert, stellt seit einigen Jahren in Deutschland zunehmend mehr Daten für die Wissenschaft bereit. Einen sehr weitreichenden Datenzugang für Wissenschaftler gewährt seit 2002 das neugeschaffene Forschungsdatenzentrum der Statistischen Landesämter (FDZ/L) in seinen Standorten am Sitz der Statistischen Landesämter; besonders interessant ist dabei die Nutzung von Mikrodaten entweder am Gastwissenschaftler-Arbeitsplatz oder als Kontrollierte Datenfernverarbeitung.

⁴ <http://www.bda-online.de/>.

Die Bereitstellung der Daten der Krankenhausstatistik durch das FDZ/L und deren Verbindung mit amtlichen Wirtschaftsdaten wie denen der Arbeitskreise VGR und ETR oder den Arbeitslosenzahlen der Bundesagentur für Arbeit (BA) ermöglicht es nun auf empirischer Grundlage, Veränderungen im Krankenhaussektor des Gesundheitssystems mit den gleichzeitigen wirtschaftlichen Veränderungen in Beziehung zu setzen. Hier sollen – im Rahmen des Projektes 606-2007 (Forschungsdatenzentrum der Statistischen Landesämter, 2007) – die ersten eineinhalb Jahrzehnte Transition in Ostdeutschland ökonometrisch hinsichtlich der Stärke dieser Beziehungen untersucht werden. Dabei gilt es herausfinden, welche der Veränderungen in Ostdeutschland transitionsbezogen sind und ob und wie sich die transitionsbezogenen von eher transitionsunabhängigen Veränderungen trennen lassen.

Kapitel 1.

Einführung: Transformationsprozesse Ostdeutschlands

DIE MÜHEN DER GEBIRGE LIEGEN HINTER UNS.
VOR UNS LIEGEN DIE MÜHEN DER EBENEN.
Bertolt Brecht⁵

Die Gesellschaft verändert sich ständig. Betrachten wir dabei kurze Zeiträume, so sind es vor allem die Zeiten spezifischer Krisen und Umbrüche, für die die Veränderungen allgemein und unmittelbar wahrnehmbar sind – charakterisierende Begriffe wie *Finanzkrise* (2009) oder *Friedliche Revolution* (1989) für den gesellschaftlichen Schock dienen der Beschreibung und Definition zugleich, es gibt jeweils ein *Davor* sowie ein *Danach*. Veränderungen in den dazwischen liegenden Zeitabschnitten erscheinen vergleichsweise weniger eindrucksvoll, vielleicht, weil sie nicht ganz so abrupt vonstatten gehen. In diesen Zeitabschnitten mag es sowohl zum Versuch eines Zurückführens auf den Vor-Schock-Zustand, einer Konsolidierung des Erreichten (bis hin zu Stagnation) als auch einer stetigen weiteren Veränderung kommen: in Politik, Wirtschaft, Sozialbeziehungen oder auch nur einzelnen dieser Bereiche der Gesellschaft.

Veränderungen der Grundlagen der gesamten Gesellschaft, des ganzen Gesellschaftssystems, werden häufig mit Begriffen wie Systemtransformation belegt, eine einheitliche Begriffsbildung hat sich bisher nicht durchgesetzt. Systemtransformationen lassen sich – z. B. hinsichtlich obiger Bereiche der Gesellschaft – aus verschiedenen Blickwinkeln betrachten, ihrer Analyse widmen sich verschiedene

⁵ Bertolt Brecht (1898–1956). *Wahrnehmung* (1949), Ausschnitt.

Wissenschaftszweige mit unterschiedlichen Schwerpunkten und häufig eigenen Begrifflichkeiten.

1.1. Transition aus politikwissenschaftlicher Sicht

In seinem Lehrbuch *Systemtransformationen: eine Einführung in die Theorie und Empirie der Transformationsforschung* bezieht sich Merkel (1999) vor allem auf die *moderne politikwissenschaftliche* (S. 15) Transformationsforschung, wenn er konstatiert (S. 69 ff.), dass eine Vielzahl von Begriffen für das Phänomen Transformation oft ohne vorherige Klärung durch den Nutzer oder Forscher verwendet werden; Transition, Transformation, Revolution, Zusammenbruch, Modernisierung, Liberalisierung, Demokratisierung, Konsolidierung, Regimewandel, Systemwandel sind nur einige davon. Diese Liste der ähnlich verwendeten Begriffe ließe sich besonders bei Einbeziehung der englisch- und spanischsprachigen Literatur verlängern. So spricht Huntington (1991) in seinem Buch *The Third Wave: Democratization in the Late Twentieth Century* von transplacement, im Spanischen wird der Übergang vom Franco-Regime zur parlamentarischen Monarchie unter König Juan Carlos als *Transición* bezeichnet.

Merkel (1999, I.3) fordert, bei Untersuchung von sozialen, wirtschaftlichen und politischen Transformationsprozessen zwei Dimensionen zu berücksichtigen und begrifflich exakt zu erfassen: (a) die Präzisierung des Analysegegenstands und (b) das Wie der Transformation, d. h. ihre Form, Geschwindigkeit und Etappen. Unter (a) stellt sich die Frage, was genau transformiert wird: die Regierung, das Regierungssystem, das politische Regime, der Staat oder das ganze soziopolitische System; während zu (b) die Frage zählt, welche politischen und gesellschaftlichen Akteure wie und mit welchem Einfluss beteiligt sind. Ebenfalls zu beantworten ist die Frage, ob es sich um Reformen, also einen Wandel, oder um Revolution und somit einen Wechsel des soziopolitischen Systems handelt; nach einem Systemwechsel sind Herrschaftszugang, -struktur, -anspruch und -weise grundsätzlich verändert.

Transition (transition bzw. transición) beschreibt laut Merkel (1999, I.3.2) fast ausschließlich den Übergang von autokratischen zu demokratischen Systemen. Für das Europa nach 1945 zählt man drei Demokratisierungswellen: unmittelbar nach

1945, Mitte der 70er Jahre (Portugal, Spanien, Griechenland) und die Demokratisierungswelle Osteuropas seit 1989 (vgl. Huntington, 1991). Die Begriffe *Transformation* bzw. *Systemtransformation* hingegen haben nach Merkel (1999) hier keine spezifische Bedeutung, sie sind Oberbegriffe für alle Formen, Zeitstrukturen und Aspekte von Systemwandeln oder -wechseln.

Die Transitionsforschung selbst widmet sich vor allem der Rolle der Akteure im Demokratisierungsprozess. Dabei entwickelt sie eine Periodisierung des Transitionsprozesses.

1.2. Transition aus Ökonomen-Sicht

Wirtschaftswissenschaftler haben eventuell eine andere Sicht bei der Präzisierung des Analysegegenstands. So schreiben Hall et al. (2000, auf S. 353): *We define the term 'transition' in a wide sense. It is not limited to the transition from centrally planned to competitive market structure alone, but also covers the transition from Keynesian to monetarist economic systems, which was observed during the 1980s in many European countries.* Nicht überraschend, es sind aus ihrer Sicht die größeren wirtschaftlichen Umbrüche, deren Gemeinsamkeiten sie unter dem Begriff *Transition* zusammenfassen wollen.

Diese Erweiterung des Begriffs *Transition* erscheint vielen Wissenschaftlern nicht nötig. Im Februar 1991 erscheint die erste Ausgabe von *MOCT-MOST: Economic Policy in Transitional Economies*, sie beginnt mit *The pace of change in Central and Eastern Europe* als erstem Artikel. Elf Jahre der wirtschaftswissenschaftlichen Betrachtung der *Transition* der dritten Demokratisierungswelle folgen, bevor die Zeitschrift 2002 in *Economics of Planning*, seit 2005 *Economic Change and Restructuring*, aufgeht.

1.3. Transition aus ökonometrischer Sicht

In der Ökonometrie, einer Teildisziplin der Wirtschaftswissenschaften (Abschn. 3.1), kann hingegen dem Begriff *Transition* eine ganz andere Bedeutung zugemes-

sen sein. So nennt Lancaster (1990) seine Monografie *The Econometric Analysis of Transition Data* und stellt darin die Analysemethoden vor für Daten, die die Zeit bis zum Eintreten eines Ereignis beschreiben; Transition ist hier die Veränderung des Status eines Objektes der Untersuchung, der Moment des Übergangs von einem Status in einen anderen. Beispiele dafür sind *unverheiratet* und *verheiratet*; *wirtschaftlich aktiv* und *insolvent* eines Unternehmens; oder *arbeitslos*, *angestellt*, *nicht verfügbar* eines Bürgers in Bezug zum Arbeitsmarkt. Diese statistischen Methoden – auch als Methoden der Überlebenszeitanalyse oder time-to-event analysis mit wiederkehrenden Ereignissen bekannt – werden auf spezifisch wirtschaftswissenschaftliche Fragestellungen angewendet.

1.4. Transition Ostdeutschlands – Sonderfall der dritten Demokratisierungswelle

Zeitlich gehört das Ende der DDR und der Beitritt der ostdeutschen Länder zur Bundesrepublik zur dritten Demokratisierungswelle in Europa (Huntington, 1991). Merkel (1999, IV.2.4) hingegen stellt heraus: *Da aber mit dem Kollaps des SED-Regimes die Existenzgrundlage des gesamten DDR-Staates wegfiel, soll die DDR als ‚Sonderfall‘ analysiert werden.* Als Ursachen für den Regimekollaps werden ein vorausgehender Legitimitätsverfall, außenpolitischer Druck und die Massenproteste der *Friedlichen Revolution* ausgemacht. Der Legitimitätsverfall umfasst alle Bereiche des gesellschaftlichen Lebens. Hervorzuheben sind hier aus der Liste in Merkel (1999) der Legitimitätsverfall in wirtschaftlicher und wohlfahrtsstaatlicher Hinsicht; wohlfahrtsstaatlich – mit den Zielen Vollbeschäftigung, soziale Sicherheit und sozialer Ausgleich – umfasst das staatliche Gesundheitswesen genauso wie das Recht auf Arbeit.

Wirtschaftswissenschaftler gehen ebenfalls von einem *Sonderfall* Ostdeutschland aus. *There are also more unusual forms of transition such as the unification of Germany, [...]*, konstatieren Hall et al. (2000) in ihrer Einführung in das Sonderheft von *Economic Modelling zu Inflation and Unemployment in Economies in Transition*.

Jedes der mittel- und osteuropäischen Länder der dritten Demokratisierungswelle bleibt innerhalb seines Staatsterritoriums bestehen oder separiert sein Territo-

rium für einen unabhängigen Staat einschließlich Bürgern und Wirtschaftsunternehmen aus Teilen des Vorgängerstaates. Während z. B. Polen, Ungarn, Albanien territorial unverändert bleiben, spaltet sich die Tschechoslowakei (nach 75 Jahren) zum 1. Januar 1993 friedlich in Tschechien und die Slowakei. Nur Slowenien gelingt es, ohne kriegerische Auseinandersetzung Jugoslawien zu verlassen, alle anderen Nachfolgestaaten Jugoslawiens sind in die Balkankriege der 90er Jahre des 20. Jahrhunderts verwickelt. Trotzdem: diese neuen Staaten organisieren ihre Transition weitgehend selbst, Ziele und vor allem das Tempo werden – unter Einbeziehung externer Modelle – in demokratischen Entscheidungsprozessen von den Bürgern selbst definiert.

Die Ausnahme von dieser Regel macht Ostdeutschland zum *Sonderfall*: nicht nur das System ändert sich, der selbständige Staat auf dem Territorium der DDR hört auf zu existieren, DDR-Staatsbürger werden automatisch und unmittelbar mit dem 3. Oktober 1990 zu Bürgern der Bundesrepublik. Für die Bundesrepublik ergeben sich Zunahmen: 25 % neue Bürger, 40 % Territorium, dazu Wirtschaftsunternehmen, Sozialeinrichtungen, Administration.

Brezinski und Fritsch (1995) verwenden den Begriff Transformation (ohne ihn zu definieren) im Sinne von *Transition* nach Merkel (1999). Sie spezifizieren aus wirtschaftswissenschaftlicher Sicht, vor welchen Arten von ökonomisch relevanten Schocks die ostdeutschen Firmen am Beginn der Transition standen: (a) dem Wettbewerbsschock, (b) dem mit einer deutlich geringeren Fertigungstiefe einhergehende Angebotsschock, (c) dem die Firmen stark belastenden Lohnschock und (d) dem Regulierungsschock – der Übernahme des gesamten westdeutschen Rechts- und Verwaltungssystems mit nur wenigen Übergangsregelungen. Diese ergeben für die Autoren folgerichtig einen mentalen Schock bei vielen Ostdeutschen. Der mentale Schock zeige sich vor allem darin, dass Einsicht in die fundamentale Bedeutung der Eigeninitiative zu langsam voranschreite und einen Hemmfaktor bei der schnellen wirtschaftlichen Entwicklung Ostdeutschlands darstelle. (Mögliche psychosomatische Folgen ihrer inhärenten Forderung nach schneller Adaptation der Menschen werden nicht erwähnt.) Aufbauend auf ihrer Schock-Klassifizierung analysieren die Autoren die ersten fünf Jahre der Transition und stellen gleich im ersten Kapitel klar, dass der Weg Ostdeutschlands mit dem anderer postkommunistischer Länder nicht vergleichbar ist.

1.5. Transition, ihre Dauer und parallele Veränderungen in Deutschland

Transition beginnt häufig mit einem Schock des politischen oder wirtschaftlichen Systems als erster Etappe. Für Ostdeutschland mag diese Etappe – ggf. als Summe von Teil-Etappen – eingegrenzt werden auf die Zeit vom Beginn der Montagsdemonstrationen bis zur Herstellung der staatlichen Einheit am 3. Oktober 1990. Der ökonomische Schock ist dabei vor allem mit der Währungsunion am 1. Juli 1990 und ihren Folgen verbunden. Der Schock und die daran anschließenden weiteren Etappen der Transition Ostdeutschlands werden seither intensiv wissenschaftlich-empirisch begleitet und analysiert.

Wann und ob eine Transition endet, scheint häufig offen, es sei denn, ein darauffolgender Schock läutet eine neue Transitionsphase ein. Nähme man das Aufgehen des vierteljährlichen Periodikums *MOCT-MOST: Economic Policy in Transitional Economies* (1991–2001) in der Zeitschrift *Economics of Planning* als Indikator, wäre für 2001 zumindest ein Ende des wissenschaftlichen Interesses an den wirtschaftlichen Veränderungen infolge der dritten Demokratisierungswelle anzusetzen.

Das wird der Situation in Deutschland definitiv nicht gerecht, auch 2010 werden wirtschaftliche, soziale und politische Disparitäten zwischen Ost und West mit weiterem Anpassungsbedarf erklärt. Dabei gilt es für alle Transitions-Analysen zu bedenken, dass, parallel zu den transitionsbedingten Entwicklungen, Ostdeutschland auch den allgemeinen Änderungen in Deutschland unterworfen ist. Gleichzeitig ist auch Westdeutschland, z. B. über Solidaritätszuschlag, Länderfinanzausgleich, Beamtentransfer, an den Veränderungen des Ostens beteiligt.

Für diese Arbeit wird davon ausgegangen, dass die Transitionsphase Ostdeutschlands 2005 noch nicht beendet ist.

Kapitel 2.

Fragestellungen, Einordnung dieser Arbeit

Im Jahr 1990, dem Jahr der Wiedervereinigung, leben in Deutschland 79,8 Mio. Einwohner ungleichmäßig verteilt auf einer Fläche von 357.046 km². 14,8 Mio. Ostdeutsche und 3,4 Mio. (Ost- und West-)Berliner kommen zu 61,6 Mio. Bürgern der alten Bundesrepublik.

16 Jahre später leben auf unverändertem Territorium bereits 82,3 Mio. Einwohner – eine Zunahme von 2,5 Mio.⁶ Diese Bevölkerungszunahme jedoch betrifft die einzelnen Bundesländer in unterschiedlichem Ausmaß. Abb. 2.1 zeigt die Bevölkerungsentwicklung in Deutschland standardisiert auf das Jahr 1990, sowohl für jedes Bundesland und aggregiert zu den 10 Altbundesländern, den 5 neuen Ländern und Berlin.

Die Dynamik der gesamtdeutschen Bevölkerungszunahme spiegelt sich in den Bundesländern unterschiedlich wieder. Während alle ostdeutschen Flächenländer 2006 eine geringere Bevölkerung aufweisen als 1990, gilt für die meisten Altbundesländer das Gegenteil, ihre Bevölkerung wächst. Die einzelnen Bundesländer zeigen stark voneinander unterschiedene Trends, Bremen, das Saarland, Berlin und Brandenburg bilden die Gruppe der Länder mit den geringsten Veränderungen. Kumuliert man die ost- sowie die westdeutschen Bundesländer jeweils, zeigt sich noch deutlicher: die Bevölkerung Ostdeutschland schrumpft in 16 Jahren um fast 10 %, die Bevölkerungsdichte der westdeutschen Bundesländer steigt von 248 auf 264, die der ostdeutschen Flächenländer sinkt von 137 auf

⁶ Daten sowie Berechnungen des Autors in der Kapiteleinleitung basieren auf Angaben aus: Statistisches Bundesamt (2006–2010), Statistisches Bundesamt (2008e) und Statistisches Bundesamt (1993–2005) – ergänzt und korrigiert mit Hilfe von Statistisches Landesamt Berlin (2002); Statistisches Landesamt Bremen (2009); Statistikamt Nord (2009).

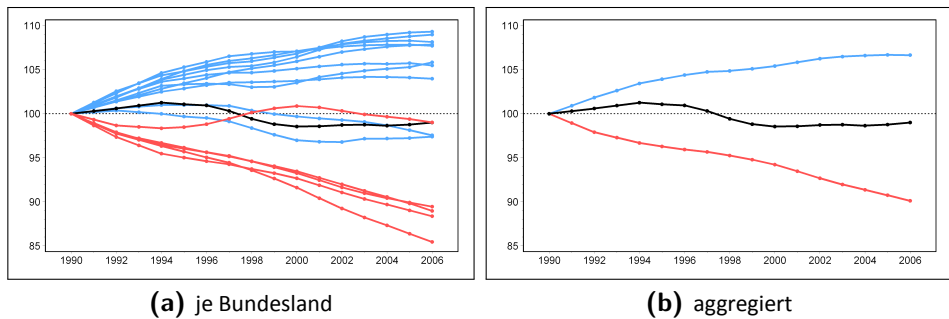


Abb. 2.1.: Veränderung der Einwohnerzahlen in den Bundesländern, standardisiert auf 1990 (= 100 %)
blau: Altbundesländer (10), rot: ostdeutsche Flächenländer (5), schwarz: Berlin (1)

123 Einwohner je km² – eine Angleichung der Trends der Bevölkerungsentwicklung findet zwischen Ost und West bisher nicht statt.

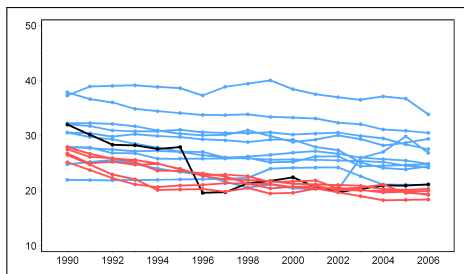


Abb. 2.2.: Krankenhausdichte in den 16 Bundesländern [KH/10⁶ EW]
blau: Altbundesländer (10), rot: ostdeutsche Flächenländer (5), schwarz: Berlin (1)

Krankenhäuser dienen der Gesundheitsversorgung der Bevölkerung. Setzt man mittels der Krankenhausdichte (Krankenhäuser je 1 Mio. Einwohner) die Anzahl der Krankenhäuser in Bezug zur zu versorgenden Bevölkerung ergibt sich für Deutschland zwischen 1990 und 2006 (Abb. 2.2), dass die dünner besiedelten ostdeutschen Flächenländer über den Gesamtzeitraum am unteren Rand des Krankenhausdichtespektrums zu finden sind. Es lässt sich vermuten, dass somit der Weg zum nächsten Krankenhaus für einen Bürger Ostdeutschlands im Mittel länger ist als für einen Bewohner der Altbundesländer.

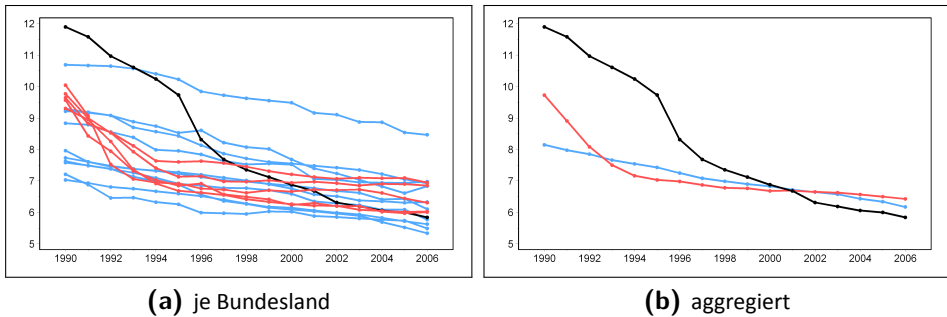


Abb. 2.3.: Bettendichte in den 16 Bundesländern [Betten/ 10^3 EW]
blau: Altbundesländer (10), rot: ostdeutsche Flächenländer (5), schwarz: Berlin (1)

Eine wichtige Kennzahl der durch Krankenhäuser bereitgestellten Gesundheitsleistung für die Bevölkerung ist in jenen Jahren die Anzahl der aufgestellten Betten. Bildet man auch hier die Bettendichte als bevölkerungsstandardisierte Kennzahl (Betten je 1000 Einwohner), ergibt sich für die Bundesländer (Abb. 2.3), dass neben dem allgemeinen Trend einer stetigen Verringerung der Bettenzahlen sich Ostdeutschland seit 1993 auf dem Niveau der Altbundesländer befindet, Berlin ist diesem Niveau seit 1998 angeglichen⁷.

Die Bevölkerung und die ihr zur Verfügung stehenden Krankenhäuser durchlaufen seit 1990 einen teils sehr dynamischen Veränderungsprozess, der weder für alle Bundesländer einheitlich noch allein durch Ostdeutschlands Angleichung an die Verhältnisse der Altbundesländer erklärt werden kann: Transition (Abschn. 1.1) und allgemeine Veränderungen überlagern sich, bewegen sich gegebenenfalls gegenläufig und summieren sich in ihren Wirkungen. Eine Trennung dieser wichtigen Einzeleffekte ist für die Analyse und Interpretation wirtschaftlicher und sozialer einschließlich gesundheitsbezogener Parameter anzustreben.

⁷ Bremen hat ob seiner geringen Größe mit seiner hohen Bettendichte wenig Einfluss auf die mittlere Bettendichte der Altbundesländer.

2.1. Zielsetzung

Ziel vorliegender Arbeit ist es, bisher wenig erforschte Aspekte des Transitionsprozesses Ostdeutschlands seit der Wiedervereinigung zu beschreiben und zu analysieren. Mittels empirischer Methoden sollen dabei die Beziehungen und Abhängigkeiten zwischen wirtschaftlichen und sozialen Prozessen anhand von Gesundheits- und Wirtschaftsparametern betrachtet werden. Basierend auf der amtlichen Krankenhausstatistik sowie ausgewählter amtlicher Wirtschafts- und Sozialstatistiken werden die Veränderungen der Parameter zwischen 1992 und 2005 und die Beziehungen zwischen diesen Parametern herausgearbeitet, sie werden hinsichtlich ihrer Bedeutung im Transitionsprozess interpretiert.

Für den betrachteten Zeitraum von 14 Jahren muss davon ausgegangen werden, dass nicht alle Veränderung ausschließlich auf die Transition zurückzuführen sind. Deshalb werden die Analysen um vergleichende Betrachtungen mit einem Bundesland außerhalb des Transitionsgebietes ergänzt.

2.2. Fragestellungen

Mit den Analysen dieser Arbeit wird versucht, vor allem auf folgende Fragen Antworten zu finden:

- i. Die Wiedervereinigung – der Beitritt Ostdeutschlands zur Bundesrepublik mit unmittelbarer Übernahme bereits etablierter Regeln des wirtschaftlichen und sozialen Miteinanders – macht Ostdeutschland zu einem Sonderfall (Abschn. 1.4) unter den osteuropäischen Ländern in Transition. Gilt dieser Sonderfall-Status auch bei Transition des Gesundheitssystems?
- ii. Empirische Analysen des Gesundheitswesens erfordern das System beschreibende Daten. Gibt es im Bereich der deutschen Amtlichen Statistik Daten zur Gesundheit und den damit verbundenen Aufwänden, mit denen sich die Leistungen des Gesundheitswesens beschreiben und beurteilen lassen?

- iii. Die Amtliche Statistik stellt verschiedene die Wirtschaft beschreibende Statistiken bereit. Wie lassen sich diese mit den Daten zu Gesundheit und Gesundheitsaufwänden für gemeinsame Analysen kombinieren?
- iv. Die Amtliche Statistik stellt Daten ihrer Statistiken zur wissenschaftlichen Auswertung durch externe Interessenten bereit. Inwieweit lassen sich diese Daten unmittelbar und qualitativ hochstehend nutzen?
- v. Veränderungen unter Transition erfolgen über mehrere Jahre (Abschn. 1.5) in einem dynamischen Prozess. Berichtsjahresweise werden dabei dieselben Wirtschaftssubjekte wiederholt beobachtet. Die Panelanalyse ist eine dafür sehr geeignete statistisch-ökonomische Methode, sie basiert auf einer adäquaten Datenstruktur. Lässt sich aus den Daten der Amtlichen Statistik ein Panel mit den Berichtsjahren des Interesses als Panelperioden⁸ generieren?
- vi. Wirtschaft und Gesundheitssystem sind als Bereiche derselben Gesellschaft nicht disjunkt, sie überlappen einander. Sind die (hier vorliegenden) Daten der Amtlichen Statistik ausreichend für Einblicke und Beurteilungen der Interaktionen zwischen dem Gesundheitswesen und Wirtschaft?
- vii. Das hier zugrunde gelegte Panel umfasst mit den Jahre 1992–2005 nicht die Gesamtzeitraum der bisher nicht als abgeschlossen angesehenen Transition. Lassen sich Strategie und Technik des Panel-Bildens und -Auswertens fortsetzen, lässt sich das generierte Panel selbst gegebenenfalls aufwandsarm fortsetzen?
- viii. Der Beitritt zur Bundesrepublik bringt für jeden einzelnen erhebliche wirtschaftliche und soziale Veränderungen mit sich, diese stellen teilweise hohe psychische Anforderungen an die Betroffenen und müssen verarbeitet werden; Kranksein kann die Folge sein. Kann die (Veränderung) der Häufigkeit bestimmter Krankheitsgruppen mit dem Transitionsgeschehen und der wirtschaftlicher Lage erklärt werden?
- ix. Transitionsgebiet ist primär Ostdeutschland, das Gebiet der Alt-Bundesrepublik ist dennoch mit betroffen. Lassen sich anhand der Daten Ost-West-Unterschiede für das Gesundheitswesen herausarbeiten und interpretieren?
- x. Auch im und über das Gesundheitswesen bevorzugt die Diskussion bestimmte Kennzahlen zur Positionsbestimmung und zum Vergleich – national und international. Nicht immer passen diese Kennzahlen zu den abgeleite-

⁸ oder Panelzeitpunkte.

ten Aussagen. Inwieweit müssen für die Aussagen dieser Arbeit Alternativen entwickelt werden?

Weiterführende und Teilfragen ergeben sich im Verlauf dieser Arbeit.

2.3. Vergleichsland Rheinland-Pfalz

Für die Analyse der Transitionseffekte unter Beachtung der allgemeinen (gesamt-) deutschen Veränderungen im Untersuchungszeitraum soll jeweils der Vergleich mit den kontemporären Veränderungen in den Altbundesländern einbezogen werden. Datenumfang, Dateneignerschaft und somit erforderliche Projektzustimmung durch das StLA eines jeden datengebenden Bundeslandes lassen eine Beschränkung auf einen Repräsentanten der zehn Altbundesländer ratsam erscheinen.

Für die Wahl von Rheinland-Pfalz als Repräsentant sprechen eine Reihe von Gründen. Rheinland-Pfalz ist ein Flächenland (19.847 km²), administrativ gegliedert in 24 Landkreise und 12 kreisfreie Städte. Gemäß Einstufung durch das Bundesamt für Bauwesen und Raumordnung sind für den Untersuchungszeitraum alle drei Regionsgrundtypen vertreten: neun der Kreise gehören zu einem *Agglomerationsraum*, 22 zu einem *Verstädterten Raum* und 5 zu einem *Ländlichen Raum*. Von den neun Kreistypen des Bundesamt für Bauwesen und Raumordnung (Abschn. 3.3.5 sowie Abb. 4.1 auf S. 76) ist nur der Typ *Ländlicher Kreis im Agglomerationsraum* nicht vertreten.⁹ 2005 kamen auf 4,059 Mio. Einwohner 97 Krankenhäuser, das sind 23,9 Krankenhäuser je 1 Mio. Einwohner, damit liegt Rheinland-Pfalz wie alle ostdeutschen Bundesländer unter dem deutschen Durchschnitt von 25,9; während Bayern mit 30,9 und Schleswig-Holstein mit 36,7 Krankenhäusern je 1 Mio. Einwohner die diesbezüglichen höchsten Raten aufweisen. Da Rheinland-Pfalz an keines der ostdeutschen Länder unmittelbar angrenzt, wird angenommen, dass unmittelbar durch einen täglichen Pendler-Strom generierte wirtschaftliche Einflüsse gering sind.¹⁰ Sowohl administrativ als auch de-

⁹ Die 8 unmittelbar an Berlin grenzenden Landkreise Brandenburgs werden, wie 4 Landkreise Sachsens auch, vom BBR als *Ländliche Kreise eines Agglomerationsraumes* eingestuft.

¹⁰ Auspendler des einen Gebietes verändern als Einpendler eines anderen Gebietes des letzteren Wirtschaftskraft und entziehen sich dem Arbeitsmarkt des ersteren.

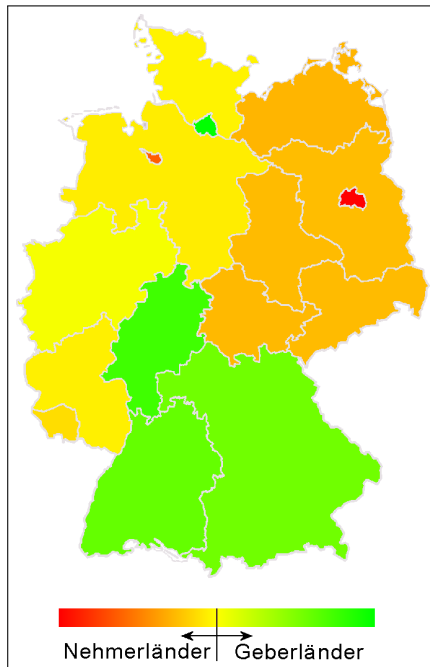


Abb. 2.4.: Pro-Kopf Transferleistung des Länderfinanzausgleichs (2004)
 Es nehmen: Berlin 795 €, ostdeutsche Flächenländer etwa 215 €,
 Rheinland-Pfalz 47 €
 (Zum Vergleich, es geben: Hamburg 329 €, Hessen 249 €)^x

^x Daten: eigene Berechnungen (basierend auf Statistisches Bundesamt, 2006e); Grafik: <http://de.wikipedia.org/w/index.php?title=Datei:Laenderfinanzausgleich.png&filetimestamp=20060709150501>, Zugriff am 12. 02. 2011

mografisch ähneln sich Rheinland-Pfalz und Sachsen des Jahres 2005, in Sachsen leben 4,284 Mio. Einwohner auf 18.415 km² in 22 Landkreisen und 7 kreisfreien Städten, es gibt 19,8 Krankenhäuser je 1 Mio. Einwohner. Im Gegensatz zu Rheinland-Pfalz wird jedoch keiner der sächsischen Landkreise einem *Ländlichen Raum* zugeordnet; alle anderen ostdeutschen Flächenländer besitzen hingegen Kreise des *Ländlichen Raumes*.

In Rheinland-Pfalz leben 2005 6,2% der Einwohner der Altbundesländer (ohne Berlin), 5,4% der 1806 Krankenhäuser der Altbundesländer befinden sich in Rheinland-Pfalz. Legt man die Pro-Kopf-Transferleistungen des Länderfinanzaus-

gleichs (Abb. 2.4) für 2004 zugrunde, so ist die Wirtschaftskraft von Rheinland-Pfalz derjenigen der ostdeutschen Flächenländer ähnlicher als z. B. die von Baden-Württemberg, ein ebenfalls nicht unmittelbar an ein ostdeutsches Bundesland angrenzendes Flächenland.

2.4. Berlin – Teil des Ostens?

Administrativ werden in Deutschland Stadtstaaten als Bundesland ohne nachgeordnete Kreisstruktur geführt¹¹; weder weist der Amtliche Gemeindegliederungsschlüssel (AGS) für Berlin Substrukturen aus, noch sind die Stadtbezirke den Kreisen in Flächenbundesländern gleichgestellt. Demzufolge werden auch im Bereich der Amtlichen Statistik Angaben, die für Flächenländer kreisbezogen bereitgestellt werden, für Stadtstaaten nur Bundesland-aggregiert ausgegeben.

Mit dem Ende der DDR und dem Beitritt der Länder Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen zur Bundesrepublik endet auch der Viermächtestatus der geteilten Stadt Berlin. Das vormalige Ost-Berlin vereinigt sich mit West-Berlin und wird somit Teil des Bundeslandes Berlin. West-Berlins letztmals durch das Viermächteabkommen¹² bestimmter besonderer Status endet damit; Berlin wird konstitutiver Bestandteil der Bundesrepublik.

West-Berlins Inseldasein endet, wirtschaftliche und soziale Verbindungen zum Umland werden reaktiviert oder sind neu zu erschaffen. Gleichzeitig jedoch ist es nötig, den bis dahin besonderen Status West-Berlins auch in wirtschaftlichen und sozialen Fragen zu beenden, Sondervergünstigungen fallen weg. Für Arbeitnehmer wie Arbeitgeber spürbar wird dies beim Wegfall der seit dem Mauerbau gezahlten Berlin-Zulage von zuletzt 8 % des Bruttoeinkommens (vgl. Rott, 2009, S. 295; BerlinFG § 28), dem öffentlichen Dienst wird Personalabbau verordnet – alles mit dem Ziel, Berlins starke Abhängigkeit von Zuwendungen des Bundes zu vermindern. Das vereinigte Berlin muss sich zwei Aufgaben gleichzeitig stellen: Transition im Ostteil und Rückkehr zu mehr Bundesfinanzen-unabhängiger Marktwirtschaft im Westteil.

¹¹ Ausnahme: Bremen besteht aus zwei kreisfreien Städten.

¹² am 3. Juni 1972 in Kraft getreten.

Für diese Arbeit ergibt sich die Frage, ob das vereinigte Berlin dem Transitionsgebiet Ostdeutschland oder dem Vergleichsgebiet der Altbundesländer zugeordnet oder aus den Analysen gänzlich ausgeschlossen werden soll. Ein Ausschluss verbietet sich, da Brandenburg seiner wirtschaftlichen Mitte beraubt würde, den Kreisen des Agglomerationsgebietes Berlin fehlte das Agglomerationszentrum. Obwohl andere Autoren (z. B. Kuchinke und Wübker, 2009, leider ohne Argumentation) Berlin im Ost-West-Vergleich den westlichen Bundesländern zuschlagen, wird in dieser Arbeit davon ausgegangen, dass die notwendigen Veränderungen in beiden Teilen des vereinigten Berlin sich als zum bedeutenden Teil transitions- und transitionsfolgebefehdigt beschreiben lassen.

2.5. Einordnung dieser Arbeit

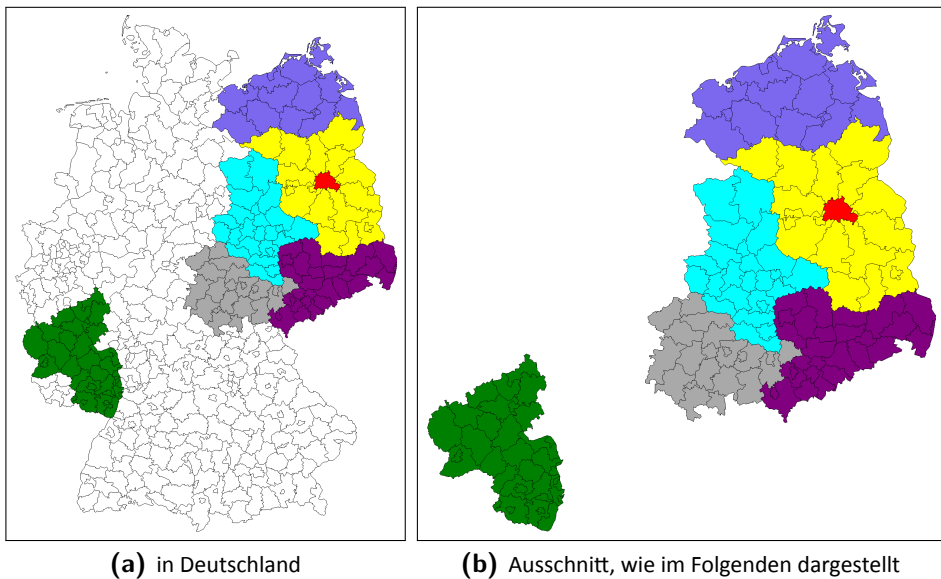


Abb. 2.5.: Untersuchungsgebiet mit Kreisgrenzen (2005):
Rheinland-Pfalz, die 5 ostdeutschen Flächenländer und die Stadt Berlin

Wirkungen der Transition Ostdeutschlands auf das Gesundheitssystem und die Interaktion zwischen GSS und wirtschaftlichen Rahmenbedingungen sollen hier

mit Hilfe von Daten der einzelnen Krankenhäuser und Wirtschaftsdaten der Kreise aus den ersten eineinhalb Jahrzehnten der Transition beschrieben werden. Hauptaugenmerk liegt dabei auf Ostdeutschland mit den Bundesländern Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen. Rheinland-Pfalz als Bundesland ohne Transition wird zum Vergleich mit den ostdeutschen Ländern und zur Diskriminierung zwischen transitionsbedingten und allgemeinen zeitlichen Veränderungen im Untersuchungszeitraum herangezogen.

Nur sieben der 16 deutschen Bundesländer gehen in die Untersuchungen ein. Diese Einschränkung hat sich als notwendig erwiesen, denn eine geringere Anzahl an datenliefernden Bundesländern verringert die Kosten und – mehr noch – senkt das Risiko der Projektablehnung durch eines der zu beteiligenden StLÄ.

Bei der Bearbeitung der Fragestellungen benutzte (kombinierte und entwickelte) Methoden für die Analyse von (ausgewählten) wirtschaftsrelevanten Gesundheitsparametern sollen so dargestellt werden, dass eine Wiederholung oder eine Erweiterung dieser Analysen mit zukünftig aktualisierten oder um andere Parameter oder Regionen erweiterte Daten der Amtlichen Statistik möglich ist. Die entwickelten Methoden lassen sich auch auf regionale Unterschiede innerhalb der Regionen Ostdeutschlands oder Deutschlands insgesamt anwenden.

Kapitel 3.

Stand der Forschung

Mehr als zwei Dekaden sind seit Beginn der Transition Ostdeutschlands vergangen mit vielen Veränderungen in Wirtschaft, Gesellschaft und Gesundheit. Wissen um diese Veränderungen speist sich aus eigenem Erleben; allgemeingültigere Erkenntnisse sind das Ergebnis von wissenschaftlich-empirischen Analysen, von Metaanalysen – den Zusammenfassungen mehrerer Primäranalysen – und theoretischen Überlegungen. Diese widmen sich sowohl dezidiert Fragen in den Teilgebieten als auch themenübergreifenden Problemen. Auf dieses Vorwissen besonders zu wirtschaftlichen und gesundheitsbezogenen Veränderungen in der Transition wird Bezug genommen werden.

3.1. Definitionen und Begriffsabgrenzungen

Wie bereits für den Begriff der Transition / Transformation (Abschn. 1.1) dargestellt worden ist, muss für Schlüsselbegriffe mit allgemein nicht eindeutiger Verwendung vorab deren Reichweite und Begrenzung für vorliegende Arbeit bestimmt werden.

Ein Parameter (griech.: $\pi\alpha\rho\alpha$ = neben und $\mu\epsilon\tau\rho\omicron\nu$ = Maß) ist im engeren Sinne für die mathematische und die angewandte Statistik eine *Konstante zur Charakterisierung einer Wahrscheinlichkeitsverteilung oder einer empirischen Verteilung in einer Grundgesamtheit* (Rönz und Strohe, 1994, S. 274 f.). Im weiteren Sinne wird in vielen Wissenschaftszweigen ein Parameter als eine charakterisierende Eigenschaft, eine Kenngröße oder Kennzahl eines zu beschreibenden Phänomens verstanden. Allgemein akzeptierte volkswirtschaftliche Kennzahlen sind z. B. das

Brutto-Inlandsprodukt (BIP), der Schuldenstand der öffentlichen Haushalte, die Zahl der Arbeitslosen oder die Kaufkraft.

Wirtschaftsparameter als Parameter in jenem weiteren Sinne seien im Folgenden Kenngrößen, die die wirtschaftlichen Bedingungen der Krankenhäuser oder der Landkreise und kreisfreien Städte (einschließlich der Stadtstaaten) meist auf jährlicher Basis darstellen. Dazu zählen u. a. die (bereinigten) Kosten und das Personal-Sachkosten-Verhältnis eines Krankenhauses, während (hier) das BIP, die Anzahl der Erwerbstätigen, die Brutto-Erwerbstätigenquote, die Arbeitslosenquote, das Arbeitnehmereinkommen je Arbeitnehmer oder das verfügbare Haushaltseinkommen je Einwohner Wirtschaftsparameter auf Kreisniveau sind.

Da kreisfreie Städte (in Baden-Württemberg Stadtkreise) und Landkreise (in Nordrhein-Westfalen und Schleswig-Holstein Kreise) der Flächenländer in der Verwaltungsstruktur Deutschlands auf derselben Stufe stehen und im Bereich der Amtlichen Statistik Angaben für diese gleichrangig neben Angaben für die Stadtstaaten Berlin, Bremen und Hamburg gestellt werden, werden in dieser Arbeit alle Vorgenannten als *Kreise* bezeichnet.

Gesundheitsparameter seien im Folgenden Kenngrößen, die auf jährlicher Basis ausgewählte Aspekte des Gesundheits- bzw. Krankheitszustands der Patienten der Krankenhäuser und der Ressourcen für deren Behandlung beschreiben. Diese Kenngrößen können auf einzelne Krankenhäuser oder – aggregiert – auf Kreise bezogen sein. Beispiel dafür sind die Anzahl der aufgestellten Betten, die Anzahl der Fälle allgemein und bestimmter Diagnose-Gruppierungen oder auch abgeleitete Größen wie die Anzahl der Fälle je Einwohner. Weitere, für die Beschreibung des Gesundheitswesens wichtige Kennzahlen finden sich mit Definitionen z. B. in Statistisches Landesamt Berlin (2002)¹³.

Ökonometrie, makro- und mikroökonomisches Modellieren

Die Ökonometrie beschäftigt sich als Teilgebiet der Wirtschaftswissenschaft mit der *quantitative[n] Analyse ökonomischer Phänomene, der komplexen, jedoch nicht immer exakt erfassbaren ökonomischen Zusammenhänge und Prozessab-*

¹³ dort als Kennziffern bezeichnet.

läufe mit dem Ziel der empirischen Verifizierung theoretisch begründeter Lehrmeinungen, der Stützung wirtschaftspolitischer Entscheidungen bzw. der Ermittlung von Alternativen und der fundierten Prognose wichtiger ökonomischer Größen (Rönz und Strohe, 1994, S. 266 f.); sie ist eine Vereinigung von Wirtschaftstheorie, angewandter Stochastik und Wirtschaftsstatistik und als solche Teilgebiet der Wirtschaftswissenschaften. Ökonometrischer Erkenntnisgewinn lässt sich nach Rönz und Strohe gut in 5 Phasen vorstellen: (a) Spezifikation des Modells, (b) Identifikation – die Beantwortung der Frage, ob sich die Parameter des Modells aus den Daten eindeutig bestimmen lassen, (c) Schätzung der Parameter des Modells, (d) Hypothesenprüfung – zur Prüfung sowohl der Annahmen des Modells als auch der Bedeutung der einzelnen Parameter, (e) Interpretation – Bewertung der Ergebnisse aus wirtschaftswissenschaftlicher Sicht. Makroökonometrische Modellierungen beschäftigen sich dabei mit Volkswirtschaften insgesamt oder deren Sektoren, mikroökonometrisches Modellieren soll die Verhaltensweisen einzelner Wirtschaftssubjekte wie Haushalte oder Unternehmen analysierbar machen; die Gesamtheit der einzelnen Krankenhäuser ist im Folgenden Grundlage mikroökonometrischer Betrachtungen, während solch eine Zuordnung für Analysen auf der Basis der kreisweisen (kumulierten) Krankenhausleistungen nicht ganz so eindeutig möglich scheint.

Lancaster (1990) meint: ... *econometrics is not statistics and the problems posed by an econometric analysis are not identical to those that a statistician would see*. Dem kann zugestimmt werden, wenn Statistik nur als mathematische Statistik mit dem Hauptziel der Weiterentwicklung von Theorie und Verfahren verstanden wird. Ökonometrie ist zu einem großen Teil angewandte Statistik, ökonometrische Fragestellungen befördern oft die Entwicklung statistischer Theorie und Verfahren oder initiieren diese sogar, so wie dies aus anderen Statistik-Anwendungsgebieten heraus ebenfalls geschieht. Parallelentwicklungen führen dabei oft zu unterschiedlichen Begriffssystemen für dasselbe, eine Begriffskonsolidierung über Fachgebiete hinweg ist meist schwierig.

Hierarchische, Mehrebenen- und Gemischte Modelle in Statistik und ihren Anwendungen

Beck (2005) stellt (am Beispiel der Politologie) fest, dass verschiedene Wissenschaftler-Arbeitsgruppen für ein und dasselbe statistische Modell unterschiedlichen Begriffswelten anhängen. Aus Sicht der Gesundheitswissenschaft halten Palta und Seplaki (2002) dieses fachgebietsübergreifend für ein gravierendes Problem, mangelnder fachlicher Austausch ist Ursache und Ergebnis. Beck plädiert sowohl für die Anerkennung der jeweils anderen Terminologie und empfiehlt deren Vereinheitlichung. Speziell trifft dies auf die unter dem Begriff des Mehrebenenmodells (Goldstein, 1995), des Hierarchischen (Bryk und Raudenbush, 1992) oder des Gemischten Modells/ Modells mit gemischten Effekten bekannten linearen und nicht linearen Modelle zu. Bereits sieben Jahre vor Beck zeigt Singer (1998) über diese Begriffsgrenzen hinweg die Gemeinsamkeiten auf und erläutert die Analysemöglichkeiten für die Gemischten Modelle mit ihren festen und zufälligen Faktoren mittels der Prozedur `proc mixed` in SAS® (SAS Institute Inc., 2004b). Ihre Anwendungsbeispiele entstammen der Bildungs- und Verhaltensforschung.

Weite Verbreitung finden diese Modelle besonders bei der Analyse von Veränderungen über die Zeit anhand von an denselben Subjekten (oder Objekten) wiederholten Messungen. Für vorliegende Arbeit werden in Anlehnung an Brown und Prescott (2006) Begriffe wie *Gemischtes lineares Modell*, *Varianzkomponentenmodell* und *Zufallskoeffizientenmodell* (letzteres ebenfalls in Beck, 2005) speziell für die longitudinalen Aspekte der Modellierung und Analyse von Paneldaten (Abschn. 3.4.1 und 3.4.2) verwendet.

Gesundheitswesen und Gesundheitssystem

Die Begriffe Gesundheitswesen (GSW) und Gesundheitssystem (GSS) werden sowohl in der Fachliteratur (z. B. Maes, 2000) als auch im täglichen Sprachgebrauch synonym verwendet. Die EU-Kommission tut dieses ebenso, wenn sie als die drei langfristigen gemeinsamen *EU-Ziele für Gesundheitswesen ...* (European Union, 2001) Zugänglichkeit für jeden, hohe Qualität und langfristige Finanzierbarkeit formuliert und postuliert, dass dieses *...nur dann verwirklicht werden [kann]*,

wenn es sämtlichen Akteuren innerhalb der Gesundheitssysteme gelingt, starke Partnerschaften aufzubauen.

Das Gesundheitswesen besteht aus den Organisationen, Einrichtungen, Personen, Regelungen und Prozessen, deren Aufgabe Erhalt und Förderung der Gesundheit von Einzelpersonen als auch der Bevölkerung sowie Vorbeugung und Behandlung von Krankheiten und Verletzungen ist. Leistungsempfänger sind vor allem Einzelpersonen als Patienten oder Vorsorgeteilnehmer, Leistungsfinanzierer (in Deutschland) die verschiedentlich Versicherten, die Arbeitgeber oder auch Direktzahler. Die Beteiligung des Staates erfolgt sowohl über die gesetzgeberische Funktion als auch durch Regulierung und Überwachung z. B. durch die Gesundheitsämter. Leistungserbringer sind Ärzte, Pflegepersonal und andere im Rahmen des GSW tätige Personen. Krankenhäuser, Rehabilitationseinrichtungen, Ambulanzen, Praxen und Apotheken sind Einrichtungen, in denen die Leistungserbringer dabei organisiert tätig sind.

Die Art und Weise, wie Gesundheitssysteme in unterschiedlichen Staaten organisiert sind, variiert im internationalen Vergleich recht deutlich, stellen Busse und Wörz (2009) fest und führen weiter aus, dass [j]edes Gesundheitssystem ... vier Funktionen erfüllen [muss]: Leistungen sind zu erbringen; sie müssen finanziert werden; Technologien, Infrastruktur und Humanressourcen sind bereitzustellen; die Verhältnisse der Leistungserbringer und der Kostenträger (und anderer Akteure) müssen definiert werden.

3.1.1. Gesundheitssysteme als Forschungsgegenstand

Busse und Wörz fahren fort: *Hinsichtlich der Finanzierung unterscheiden sich Gesundheitssysteme danach, ob Steuern, Sozialversicherungsbeiträge oder private Mittel herangezogen werden. Analog hierzu kann die Regulierung des Gesundheitssystems als überwiegend durch den Staat, durch gesellschaftliche Akteure – wie Sozialversicherungsträger – oder den Markt beschrieben werden. Leistungen werden typischerweise von staatlichen, privat-gemeinnützigen oder privat-erwerbswirtschaftlichen Einrichtungen erbracht.*

Mit Gesundheitssystemen als Forschungsgegenstand beschäftigen man sich (in Deutschland) aus ganz unterschiedlichen Einrichtungen heraus. Als Beispiele seien genannt: Das *Fachgebiet Management im Gesundheitswesen* im Institut für Technologie und Management der TU Berlin sieht seinen Schwerpunkt in *ökonomischen Fragen im Gesundheitswesen*¹⁴; das *Institut für Gesundheitssystemforschung* (IGFo) der Fakultät für Medizin an der Universität Witten-Herdecke hingegen legt seinen *Schwerpunkt auf [die] Analyse von Einflussfaktoren auf die Qualität der medizinischen Versorgung*¹⁵, während das gemeinnützige ISEG – *Institut für Sozialmedizin, Epidemiologie und Gesundheitssystemforschung e.V.* Gesundheitssystemforschung als einen Teil der Gesundheitswissenschaften / Public Health versteht¹⁶.

Krappweis et al. (2000) erklären die in diesem Kontext wichtigen Begriffe Gesundheitsökonomie, Public Health, Pharmakoökonomie, Pharmakoepidemiologie und Krankheitskostenstudie.

Die Gesundheitsökonomie beschäftigt sich explizit mit der wirtschaftlichen Situation des Gesundheitssystems, z. B. am *Institut für Epidemiologie, Sozialmedizin und Gesundheitssystemforschung* der Medizinischen Hochschule Hannover. *Der Arbeitsschwerpunkt Gesundheitsökonomie wendet die Methoden und Erkenntnisse der ökonomischen Theorie auf den Bereich des Gesundheitswesens an. Forschungsfelder der Gesundheitsökonomie sind: ökonomische Analyse des Gesundheitssystems, ökonomische Evaluation von Gesundheitstechnologien, gesamtwirtschaftliche Wirkungen des Gesundheitssektors und Management in Institutionen des Gesundheitswesens.*¹⁷

3.1.2. Gesundheit und Wirtschaft, Wirtschaft und Gesundheit

Gesamtwirtschaftliche Wirkungen des Gesundheitssystems werden vor allem aus zwei Blickwinkeln heraus empirisch untersucht: einer Betrachtung des *Wirt-*

¹⁴ http://www.mig.tu-berlin.de/menue/ueber_uns/, Zugriff am 03. 06. 2010.

¹⁵ <http://medizin.uni-wh.de/humanmedizin/institute/uebersicht/institut-fuer-gesundheitssystemforschung/>, Zugriff am 03. 06. 2010.

¹⁶ <http://www.iseg.org/allgemeines.html/>, Zugriff am 03. 06. 2010.

¹⁷ Forschungsschwerpunkt Gesundheitsökonomie und Gesundheitspolitik, <http://www.mh-hannover.de/1567.html>, Zugriff am 03. 06. 2010.

schaftsfaktors Gesundheitswesen (Blöß, 2005) bzw. der Sicht auf einen *Gesundheitsmarkt* (Henke et al., 2003). Dabei stehen vor allem der Einfluss des GSWs auf die Anzahl der Arbeitsplätze und das Wirtschaftswachstum einer Region im Mittelpunkt des Interesses.

Gleichzeitig beschäftigt man sich empirisch und theoretisch mit der *zunehmende[n] Diskussionen hinsichtlich Qualität und Wirtschaftlichkeit der medizinischen Versorgung. Die Ressourcen des Gesundheitssystems sind seit längerem nicht mehr ausreichend, um alles medizinisch Machbare auch finanzieren zu können. Daher ist es notwendig, medizinische Versorgung so effizient wie möglich zu gestalten.* (Willich und Busse, 2006). Das führt dazu, dass (neue) Medikamente und Therapien nicht nur hinsichtlich ihrer medizinischen Wirksamkeit und Sicherheit gegenüber unerwünschten Nebenwirkungen untersucht und beurteilt werden sondern zunehmend auch hinsichtlich ihres Kosten-Nutzen-Verhältnisses. Ziel dieser Betrachtungen ist das Ableiten von Entscheidungen über die Rückerstattungsfähigkeit der Kosten eines Medikamentes oder einer Therapie. Das Kosten-Nutzen-Verhältnis wiederum ist schwer zu definieren, aktuell existieren in den europäischen Ländern sehr unterschiedliche Ansätze der Bestimmung des Kosten-Nutzen-Verhältnisses. So favorisiert das NICE¹⁸ eine Bewertung neuer Arzneimittel und Therapien mit Hilfe der Kosten für die Zunahme von 1 QALY (*Quality Adjusted Life Year*) gegenüber der bisherigen Standardpraxis (Campbell, 2010), während das IQWiG¹⁹ in seinem aktuellen *Arbeitspapier Modellierung* (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2009b) sehr viele konkurrierende Methoden und Zielparameter vorstellt, ohne QALY überhaupt zu erwähnen.²⁰

Dieses unterschiedliche Vorgehen mag mit dem jeweils national vorherrschenden Gesundheitssystem und seiner jeweiligen Finanzierung (Abschn. 3.1.3) zusammenhängen.

Mit Veränderungen des Gesundheitszustandes und Veränderungen der Gesundheitskosten beschäftigen sich aus makroökonomischer Perspektive Day und Tou-

¹⁸ National Institute for Health and Clinical Excellence (Großbritannien), gegründet 1999.

¹⁹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Deutschland), gegründet 2004 in der Folge des Gesetzes zur Modernisierung der gesetzlichen Krankenversicherung (GMG).

²⁰ Im Arbeitspapier *Kostenbestimmung* (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2009a) findet sich der Begriff QALY allein im Glossar.

signant (2005). Sie untersuchen anhand der Situation in Kanada zwischen 1926 und 1999 das Verhältnis zwischen Einkommen, Gesundheitsstatus und Gesundheitsausgaben. Anhand verschiedener gemessener Merkmale – mehr oder minder guter Näherungsvariabler (Proxy), standardisiert für Einwohnerzahl und Altersverteilung – interpretieren sie ihre empirischen Ergebnisse im Lichte sich ändernder Gesetze und gesellschaftlicher Programme. Ein (mikroökonomischer) Vergleich zwischen den administrativen Einheiten Kanadas (Gemeinden, Counties, Provinzen und Territorien) erfolgt nicht.

Röhmel (2009) untersucht die Veränderungen der Effizienz ostdeutscher Krankenhäuser über mehrere Jahre. Anhand einer wohl *bewussten Auswahl* (Rönz und Strohe, 1994, S. 349 f. Stichprobenverfahren) von 100 Krankenhäusern mit vollständiger Biografie, d. h. unter Auslassung aller Krankenhausschließungen, -neugründungen und -fusionen, wird mittels Data Envelopment Analysis (DEA) geschlossen, dass zwischen den verschiedenen Krankenhäusern keine wesentlichen Effizienzunterschiede bestehen.

Aus Sicht der Wirtschaft, der öffentlichen Arbeitgeber und, infolge dessen, der Politik besteht ein großes Interesse an Gesundheit und Leistungsfähigkeit der Arbeitnehmer. Der vielleicht unmittelbarste – auf jeden Fall jedoch in der Öffentlichkeit, Politik, Wissenschaft und den Betrieben selbst heißest und kontrovers diskutierte – wirtschaftsrelevante Gesundheitsparameter ist der betriebliche Krankenstand. Der Krankenstand aus Sicht der Politik und der Krankenkassen ist der Anteil aller Arbeitnehmer in der gesetzlichen Krankenversicherung, der wegen Krankheit arbeitsunfähig gemeldet ist. Erfasst werden dabei nur die Pflichtmitglieder ohne Berücksichtigung der privat und freiwillig Versicherten. Berechnet wird der Krankenstand als Verhältnis der gemeldeten Arbeitsunfähigkeitstage zu den Versicherungstagen. Der Jahresdurchschnittswert des Krankenstands ist auf eine Stichtagerhebung der gesetzlichen Krankenkassen zurückzuführen. Aus den Stichtagswerten (jeweils am Monatsersten) wird anhand des arithmetischen Mittels der prozentuale Anteil der arbeitsunfähigen Pflichtmitglieder ermittelt.

Aus der Sicht eines Einzelbetriebes lässt sich der Krankenstand gewöhnlich nur und besser mittels der Erhebung von Fehlzeiten bestimmen. Amend und Lachtrup (1991) unterscheiden unentschuldigte und entschuldigte Fehlzeiten, Krankheit und Arbeitsunfälle sind wichtige Unterkategorien der letzteren. Der betriebliche Krankenstand berechnet sich dabei in Deutschland aus den Krankmeldungen

der Arbeitnehmer mit oder – bei kurzzeitigen Erkrankungen – ohne Arbeitsunfähigkeitsbescheinigung (AU-Bescheinigung) des behandelnden Arztes oder Zahnarztes. Ortlieb (2003) untersucht anhand der Daten aus 40 Jahren eines württembergischen Metallbetriebes die Zusammenhänge zwischen der betriebswirtschaftlichen Situation, Belegschaftsveränderungen durch Einstellung und Entlassung und des Gesundheitszustandes der Arbeitnehmer. Sie arbeitet anhand dieses einen Betriebes heraus, dass übliche Sichtweisen auf das Verhältnis zwischen Krankenstand und Betriebssituation sowie die zeitlichen Abfolgen der Veränderungen beider einer empirischen Überprüfung häufig nicht standhalten. Mit diesen Daten modellieren Kauermann und Ortlieb (2004) den Effekt einer Entlassungswelle auf den betrieblichen Krankenstand in einer Periode vom Beginn der wirtschaftlichen Schwierigkeiten bis zur Konsolidierung des Betriebes mit etwa $\frac{1}{3}$ der ursprünglich Beschäftigten. Die lokale Arbeitslosenquote ist dynamische Kovariate in ihrem Modell mit penalisierter Spline-Schätzung.

Eine vergleichbare empirische Betrachtung des Verhältnisses zwischen Gesundheitsstatus und wirtschaftlicher Leistungsfähigkeit innerhalb und zwischen administrativen Regionen – Gemeinden, Kreisen, Bezirken – insgesamt fehlt bisher. Desgleichen lassen sich Betrachtungen, ob und welche Unterschiede es gibt bei der Gesundheitsversorgung zwischen wirtschaftlich stärkeren und schwächeren Regionen, bisher kaum finden.

3.1.3. Gesundheitssysteme in Europa und ihre Finanzierung

Für das Europa der vergangenen 100 Jahre sind im Wesentlichen drei verschiedene Gesundheitssystem-Modelle zu betrachten, ein jedes üblicherweise mit dem Namen eines Politikers verbunden.

Das älteste Modell, zuerst 1883 in Deutschland unter Reichskanzler Otto von Bismarck²¹ eingeführt, basiert auf einer für (fast) alle Einwohner bindenden Sozialversicherungspflicht, deren Beiträge (ggf. mit einer Kappungsgrenze) einkommensbezogen sind. Die Finanzierung des Gesundheitssystems erfolgt über die verschiedenen Sozialversicherungsträger (in Deutschland heute: Allgemeine Ortskrankenkassen, Betriebskrankenkassen, Innungskrankenkassen, Landwirt-

²¹ Graf Otto von Bismarck, 1815–1898, deutscher Politiker, 1871–1890 Reichskanzler.

schaftliche Krankenkassen, Knappschaft und Ersatzkassen); die Finanzierung ist genauso reglementierenden gesetzlichen Festlegungen unterworfen wie die Aktionen der Leistungserbringer. Ergänzt werden kann dieses (sozialversicherungs-basierte) System durch Privatversicherungen.

Der Arzt Nikolai A. Semaschko²² entwickelte ein zentralstaatlich geplantes, organisiertes und finanziertes Gesundheitssystem und führte dieses während seiner Amtszeit als erster Gesundheitsminister in der UdSSR ein. Für alle Einwohner sind die Leistungen kostenfrei, alle Kosten des Systems werden vom Staat getragen. Dieses System wurde nach 1945 in den meisten Ländern des Ostblocks modifiziert übernommen.

1942 wurde in Großbritannien von einer Regierungskommission unter Leitung des Ökonomen William Beveridge²³ ein Bericht (Beveridge Report) vorgelegt, der u. a. als Grundlage für die Einführung des NHS (National Health Service) im Jahr 1948 diente. Das NHS-Gesundheitssystem bietet – wie das Semaschko-System – für jeden Einwohner kostenfreien Zugang zu allen medizinischen Leistungen (einer staatlicherseits definierten Maximalversorgung). Der Staat organisiert das System, es wird durch Steuern finanziert. Leistungserbringer sind vor allem Einrichtungen der öffentlichen Hand; Privatanbieter und Privatversicherungen können das System ergänzen. Das Beveridge- (oder NHS-basierte) System wurde in Europa sehr schnell von Irland und den skandinavischen Staaten übernommen.

Mit den Umwälzungen in den Staaten des früheren Ostblocks stellte sich die Frage nach den Grundlagen des nationalen Gesundheitssystems neu.

²² Николай Александрович Семашко (engl.: *Semashko*), 1874–1949, russischer Arzt und Politiker, Professor für Sozialhygiene, 1918–1930 Gesundheitsminister, ab 1945 Mitglied der Akademie der Wissenschaften der RSFSR.

²³ William Henry Beveridge, 1879–1963, seit 1946 Baron Beveridge of Tuggal, britischer Ökonom und Politiker, 1919–1937 Direktor der London School of Economics and Political Science, später Anführer der Liberalen im House of Lords.

3.1.4. Transitionen von Gesundheitssystemen in Europa seit den 70-er Jahren des 20. Jahrhunderts

Während sich Portugal (1979), Spanien (1986) und Griechenland (1983) bei ihrem nachdiktatorischen Übergang zu einer demokratischen Gesellschaft für das Beveridge-Modell eines Gesundheitssystems entschieden und Italien 1978 vom Bismarck- zum Beveridge-System wechselte (van der Zee und Kroneman, 2007, Tab. 2), optierten rund ein Jahrzehnt später die europäischen Staaten und Nachfolgestaaten des früheren Ostblocks mehrheitlich für die Etablierung eines Gesundheitssystems basierend auf dem Bismarck-Modell, wie in Grielen et al. (2000) für Bulgarien, Estland, Kroatien, Lettland, Litauen, Polen, Rumänien, die Slowakei, Slowenien, Tschechien, Ungarn und die Ukraine dargelegt wird. Das Ziel der Untersuchung von Grielen et al. ist, basierend auf Fragebogenerhebungen in den Jahren 1993 und 1994 Unterschiede in der Transition der Gesundheitssysteme der zentral- und osteuropäischen *former communist countries* (Grielen et al., 2000, Abstract Background) herauszuarbeiten. Dabei fällt auf, dass neben Albanien, Armenien, Aserbaidschan, Belarus, Georgien, Moldawien, Mazedonien und dem serbisch dominierten Rest-Jugoslawien auch Ostdeutschland nicht in die Untersuchung einbezogen worden ist – leider ohne Erwähnung oder gar Begründung.

Bonilla-Chacin et al. (2005) untersuchen die Transition im GSW für die sieben ärmsten der früheren Sowjetrepubliken – darunter Moldawien, Armenien, Aserbaidschan und Georgien als Länder in Europa. Alle diese Länder nehmen mehr oder minder langsam Abschied vom Semaschko-System, doch ist (2005) immer noch nicht klar offensichtlich, ob jeweils ein Beveridge- oder ein Bismarck-System Grundlage der zukünftigen Gesundheitsversorgung sein wird.

In Slowenien hingegen fiel die Entscheidung für einen Übergang des Gesundheitssystems zum Bismarck-Modell schnell. Mit der Loslösung von der Föderativen Republik Jugoslawien wurde 1992 per Gesetz (konkurrenzlos) ein Versicherer als Träger der Pflichtversicherung geschaffen, der für alle Einwohner die Gesundheitsgrundversorgung garantiert (Albrecht et al., 2009). Für darüber hinausgehende Leistungen wurde ein System von konkurrierenden Krankenkassen zugelassen. Es dauerte mehrere Jahre, bis 1999 mit einem Gesetz dieser Übergang auf eine adäquate gesetzlichen Grundlage gestellt werden konnte (Albrecht und Klazinga,

2002), ein bindendes System des Risikoausgleichs zwischen den Kassen ergänzte das System ab dem Jahr 2005. Für Polen wird für 1998 berichtet, dass der staatliche Gesundheitssektor (in Fortschreibung des Semaschko-Modell-Gesundheitssystems der Jahre der Volksrepublik) *die überwiegende Mehrheit der polnischen Bevölkerung versorgte* (Sowada, 1999, S. 3). Erst Ende 1998 wurden ein Gesetz zur Einführung einer sozialen Krankenversicherung als Pflichtversicherung für (fast) alle Bevölkerungsgruppen beschlossen, hauptsächlich regional operierende Krankenkassen bildeten das erste Gerüst des Systems.

Im Gegensatz zu denen aller anderen Ostblockstaaten kann das Gesundheitssystem der DDR nicht als ein Semaschko-System betrachtet werden (vgl. S. 26 f. in Busse und Riesberg, 2005). Neben den staatlichen Institutionen der Gesundheitsversorgung existierten parallel gemeinnützige und private Einrichtungen mit (1989) 7 % aller aufgestellten Betten. Das System wurde für die Gesamtbevölkerung über die Sozialversicherungspflicht (mit Kappungsgrenze) finanziert; zwei große Krankenkassen – eine für Arbeiter und Angestellte (89 % der Versicherten, seit 1951 in einheitsgewerkschaftlicher Trägerschaft) und eine für Mitglieder von Produktionsgenossenschaften, Künstler, Selbständige und andere Berufsgruppen (11 % der Versicherten) – ergaben eine Interpretation des Bismarck-Gesundheitssystems ohne Konkurrenz zwischen den Kostenträgern.

Das Sozialversicherungssystem der DDR wird über die Währungs-, Wirtschafts- und Sozialunion zwischen der Bundesrepublik Deutschland und der DDR am 1. Juli 1990 und den Einigungsvertrag in das System der Bundesrepublik überführt, d. h. in Kranken-, Unfall- und Rentenversicherung, erweitert um die Arbeitslosenversicherung. Eigene Strukturen für Ostdeutschland müssen nicht entwickelt werden, vom 1. Januar bis zum 31. Dezember 1991 übernimmt eine Überleitungsanstalt, danach unmittelbar die in der Bundesrepublik tätigen Kranken- und Rentenkassen. Die Modalitäten der Finanzierung des ostdeutschen Gesundheitssystems sind mit dem Jahr 1992 strukturell dem funktionierenden System der Altbundesländer angeglichen, die Transition der Administration des Gesundheitssystems in Ostdeutschland ist somit schneller als in jedem anderen Land des ehemaligen Ostblocks abgeschlossen.

Die Transition des Gesundheitssystems Ostdeutschlands unterscheidet sich gravierend von dem aller anderen ehemaligen Ostblockländer, sie ist somit Teil des Sonderfalls Ostdeutschland (Abschn. 1.4).

Insofern erscheint es weniger verwunderlich, dass sich bisher keine vergleichenden Studien zur Transition von Gesundheitssystemen des ehemaligen Ostblocks finden lassen, die die DDR und Ostdeutschland einschließen.

3.1.5. Inlandskonzept und Inländerkonzept

Das Statistische Bundesamt definiert: *Das Inlandskonzept wird auch Arbeitsortkonzept genannt. Es misst alle in einem Wirtschaftsgebiet erbrachten wirtschaftlichen Leistungen, unabhängig davon, wer sie erbracht hat.*²⁴ Im Gegensatz dazu steht: *Das Inländerkonzept wird auch Wohnortkonzept genannt. Es misst die wirtschaftlichen Leistungen aller inländischen Wirtschaftseinheiten, unabhängig davon, an welchem Ort sie erbracht wurden.*²⁵

Ob ein Wirtschaftsparameter nach dem Inlands- oder dem Inländerkonzept erhoben wird, hängt u. a. von der zugrunde liegenden Fragestellung ab. So erklärt das StBA für den Begriff des Erwerbstätigen: *Je nach Fragestellung werden die Erwerbstätigen nach dem Inländerkonzept (Wohnortkonzept) oder dem Inlandskonzept (Arbeitsortkonzept) dargestellt.*²⁶

3.2. Gesundheitswesen in Deutschland 1991–2005

Dem World Health Report 2000²⁷ der Weltgesundheitsorganisation WHO ist zu entnehmen, dass in Deutschland mehr als 10 % des Brutto-Inlandsprodukts²⁸ in

²⁴ <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/VolkswirtschaftlicheGesamtrechnungen/Begriffserlaeuterungen/Inlandskonzept,templateId=renderPrint.psml>, Zugriff am 01. 05. 2011.

²⁵ <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/VolkswirtschaftlicheGesamtrechnungen/Begriffserlaeuterungen/Inlaenderkonzept,templateId=renderPrint.psml>, Zugriff am 01. 05. 2011.

²⁶ <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/VolkswirtschaftlicheGesamtrechnungen/Begriffserlaeuterungen/Erwerbstaetige,templateId=renderPrint.psml>, Zugriff am 01. 05. 2011.

²⁷ http://www.who.int/entity/whr/2000/en/whr00_en.pdf, S. 193, Daten von 1997, Zugriff am 28. 05. 2011.

²⁸ Total expenditure on health as % of GDP.

das Gesundheitswesen fließen, Deutschland belegt damit den vierten Platz unter 191 Staaten. Seit 1995, dem ersten Jahr der Dokumentation²⁹ durch die WHO, steigt diese Quote leicht an, 2009 beträgt sie 11,3 %. Das mag verwundern, ist doch ein Hauptziel einer jeden Gesundheitsreform in Deutschland, die Effizienz des Gesundheitssystems, d. h. die Leistungen im Verhältnis zu den eingesetzten Mitteln, zu erhöhen.

3.2.1. Gesundheitsreformen

Die Geschichte des Gesundheitswesens in Deutschland ist auch eine Geschichte seiner Reformen. Der Staat als ein wichtiger Akteur im Gesundheitssystem (Abschn. 3.1) nimmt mit seinen gesetzgebenden Körperschaften und seinen Regulierungs- und Überwachungsstellen, wie z. B. den Gesundheitsämtern, dem Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) oder dem Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Eingriff in die Rahmenbedingungen, in denen die anderen Akteure des GSS handeln können. Nicht jede Veränderung betrifft dabei die Krankenhäuser und ihre Finanzierung gleichermaßen.

Nach dem Kostendämpfungsgesetz von 1977 und dem Kostendämpfungs-Ergänzungsgesetz von 1982 werden seit der Gesundheitsreform 1989 (Gesundheitsreformgesetz, GRG) u. a. die gesetzliche Krankenversicherung separat im SGB V behandelt, eine *Negativliste* unwirtschaftlicher Arzneimittel³⁰ nach § 34 Abs. 3 SGB V durch das Bundesministerium herausgegeben und periodisch aktualisiert sowie Festbeträge für Arzneimittel festgesetzt.

Am 10.04.1991 erfolgt die Neubekanntmachung des erstmals am 29.06.1972 beschlossenen Krankenhausfinanzierungsgesetzes (KHG) mit dem Zweck, Krankenhäuser wirtschaftlich zu sichern und so *eine bedarfsgerechte Versorgung der Bevölkerung mit leistungsfähigen, eigenverantwortlich wirtschaftenden Krankenhäusern zu gewährleisten und zu sozial tragbaren Pflegesätzen beizutragen* (§ 1 (1)). Investitionskosten werden durch die öffentliche Förderung übernom-

²⁹ <http://apps.who.int/ghodata/?vid=9200&theme=country> » Germany, Zugriff am 28. 05. 2011.

³⁰ <http://www.g-ba.de/institution/sys/glossar/98/>; für von der Verordnung in der gesetzlichen Krankenversicherung ausgeschlossene Arzneimittel siehe z. B. http://www.g-ba.de/downloads/83-691-2/RL_AMR3-2003-10-18.pdf, Zugriff am 29. 05. 2011.

men, leistungsgerechte Erlöse sollen aus den Pflegesätzen beglichen werden (§ 4). Krankenhauspläne³¹ und Investitionsprogramme liegen in Verantwortung der Länder (§ 6).

1992 wird das Gesundheitsstrukturgesetz (GSG, gültig ab 01.01.1993) verabschiedet. Mit diesem werden sowohl eine Budgetierung der Leistungsausgaben und Verwaltungskosten sowie Arznei- und Heilmittelbudgets eingeführt. Diese Budgets und parallel begrenzte Regelleistungsvolumina beschränken für den ambulanten Sektor den Trend zur Mengenexpansion. Den Versicherten ermöglicht das Gesetz eine freie Wahl der Krankenkasse, der Beitragssatz wird für die Krankenkassen in der Konkurrenz um Mitglieder zum Wettbewerbsparameter, während die Krankenkassen sich an einem Risikostrukturausgleich untereinander beteiligen müssen. In § 14 wird ein Krankenhausinvestitionsprogramm *Zur zügigen und nachhaltigen Verbesserung des Niveaus der stationären Versorgung der Bevölkerung* für die neuen Bundesländer aufgelegt. Und es wird ein neues Entgeltsystem für Krankenhäuser eingeführt, welches in den folgenden Jahren mehrfach weiterentwickelt wird. Das neue Entgeltsystem soll für die Krankenhäuser einen Anreiz zur Ausgabenbegrenzung schaffen und die Transparenz hinsichtlich der Leistungen erhöhen, indem die Entgelte zunehmend den realen Behandlungsaufwand widerspiegeln. Von nun an gibt es neben den tagesbezogenen Pflegesätzen auch leistungsbezogene Pauschalentgelte und Sonderentgelte für chirurgische Eingriffe und aufwändige diagnostische Maßnahmen, die Vergütung für den Fall wird von der Verweildauer entkoppelt.

Die Bundespflegesatzverordnung (BPfIV) tritt zwei Jahre später, am 01.01.1995, in Kraft. Sie regelt die Vergütungen für die stationären und teilstationären Pflegeleistungen in Krankenhäusern neu. Im Interesse der Beitragssatzstabilität werden in § 6 Verkürzungen der Verweildauern, Fehlbelegungsprüfungen und Leistungsverlagerungen in die ambulante Versorgung gefordert.

Das GKV-Finanzstärkungsgesetz (GKV-FG, gültig ab 24.03.1998) soll auf eine langfristige Entschuldung und die finanzielle Stabilisierung der Krankenkassen hin wirken.

Das Fallpauschalengesetz (FPG – Gesetz zur Einführung des diagnose-orientierten Fallpauschalensystems für Krankenhäuser) vom 23.04.2002 zielt auf die Erset-

³¹ auch: Krankenhausfinanzierungspläne.

zung der bisherigen Mischfinanzierung – tagesbezogene Pflegesätze, leistungsbezogene Pauschalentgelte sowie Sonderentgelte für chirurgische Eingriffe und aufwändige diagnostische Maßnahmen – durch ein diagnosebezogenes Fallpauschalensystem basierend auf den Diagnose Related Groups (DRG, Abschn. 3.3.2). Eine Konvergenzphase, ursprünglich bis 2008 geplant, wird in zwei Fallpauschalenänderungsgesetzen später neu terminiert. Die Vergütung der Leistungen der Krankenhäuser, die unter DRG-Bedingungen arbeiten, regelt speziell das ebenfalls am 23. 04. 2002 verabschiedete Krankenhausentgeltgesetz (KHEntgG).

Mit dem 01. 01. 2004 tritt das GKV-Modernisierungsgesetz (auch Gesundheitsmodernisierungsgesetz, GMG) in Kraft. U. a. ist damit festgelegt, dass bei einem Krankenhausaufenthalt von Seiten des Patienten maximal 28 Tage lang täglich 10,- € Zuzahlung zur stationären Behandlung zu erbringen ist.

Für die Krankenhäuser führen diese wettbewerbszentrierten Reformen zu einer Abkehr vom budgetbasierten Wirtschaften hin zu deutlichen Forderungen nach mehr Kosten- oder Aufwandstransparenz sowie mehr Konkurrenz. Einen kondensierten Überblick über diese Entwicklungen geben u. a. Breckenkamp et al. (2007).

3.2.2. Krankenhäuser in Deutschland

Das Statistische Jahrbuch (Statistisches Bundesamt, 2007) dokumentiert: Zum Jahresende 2005 gibt es in Deutschland 2139 Krankenhäuser (Abb. 3.1) mit 523.824 aufgestellten Betten. Im Vergleich zur Situation 10 Jahre früher haben sich diese Zahlen um 8,0 bzw. 14,0 % verringert. Während das Personal im nicht-ärztlichen Dienst um 10,1 % auf 938.275 Beschäftigte abnimmt, steigt die Anzahl der Beschäftigten im ärztlichen Dienst im selben Zeitraum um 12,4 % auf 132.380 an. Unter den Gesundheitsausgaben 2005 von 239.357 Mio. € betreffen 25,9 % (62.107 Mio. €) die Krankenhäuser, 25,1 % der im Gesundheitswesen 4,264 Mio. Beschäftigten arbeiten in den Krankenhäusern.

Das Krankenhausfinanzierungsgesetz definiert in § 3 (3) 1 Krankenhäuser als *Einrichtungen, in denen durch ärztliche und pflegerische Hilfeleistung Krankheiten, Leiden oder Körperschäden festgestellt, geheilt oder gelindert werden sollen oder*

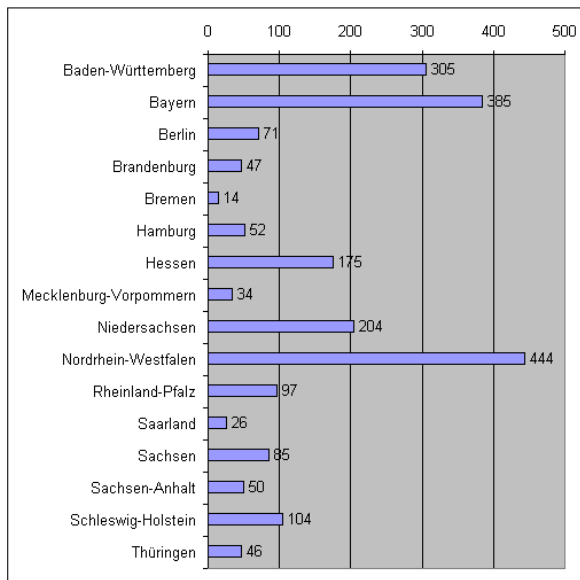


Abb. 3.1.: Anzahl der Krankenhäuser in den Bundesländern (2005)^x

^x Zahlen aus: Statistisches Bundesamt (2007, S. 233), ohne Bundeswehrkrankenhäuser

Geburtshilfe geleistet wird und in denen die zu versorgenden Personen untergebracht und gepflegt werden können (Bundesministerium für Justiz, 1991).

Krankenhäuser lassen sich klassifizieren. Für Berichte werden Krankenhäuser häufig nach der Größe – per Anzahl der (pro Jahr) behandelten Fälle oder nach der Anzahl der Planbetten –, nach ihrem Tätigkeitsschwerpunkt (1846 Allgemeine Krankenhäuser, 293 sonstige Krankenhäuser und zusätzlich 8 Bundeswehrkrankenhäuser) oder dem Träger des Krankenhauses unterteilt. Nach KHG § 1 (2) ist die Vielfalt der Träger zu beachten. Träger sind dabei öffentlich (Gebietskörperschaften³² oder Parafisci³³), freigemeinnützig (hauptsächlich Wohlfahrtsverbände³⁴) oder privat. Der Anteil der Krankenhäuser in privater Trägerschaft nimmt

³² z. B. Gemeinden bzw. Gemeindeverbände, Bundesländer – für Universitätsklinika, der Bund – für Bundeswehrkrankenhäuser.

³³ z. B. Knappschaft u. a. Sozialversicherungsträger.

³⁴ u. a. Deutscher Caritasverband (DCV), Diakonisches Werk der Evangelischen Kirche in Deutschland (DW), Arbeiterwohlfahrt (AWO), Deutsches Rotes Kreuz (DRK), Der Paritätische Wohlfahrtsverband (DPWV).

zu: 26,6 % der Krankenhäuser sind 2005 in privater Trägerschaft gegenüber 14,8 % im Jahr 1991 – bei nahezu gleichbleibendem Anteil der in freigemeinnütziger und sinkendem Anteil derer in öffentlicher Trägerschaft (von 46,0 % auf 35,1 %, Statistisches Bundesamt, 2007). Dabei gibt es gravierende regionale Unterschiede. Für 2003 berichtet Rolland (2005a) einen überdurchschnittlich hohen Anteil privater Kliniken für Berlin (42,0 %), für Nordrhein-Westfalen betrug dieser nur 8,7 %. Während der Anteil freigemeinnütziger Krankenhäuser mit 71,7 % in Nordrhein-Westfalen am höchsten ist, hat Bayern den höchsten Anteil öffentlicher Krankenhäuser (56,8 %), Berlin mit 5,8 % den geringsten. 2005 (Statistisches Bundesamt, 2007) beträgt deutschlandweit die mittlere Größe der Krankenhäuser 115 Betten für die in privater, 226 für die in freigemeinnütziger und 364 für die in öffentlicher Trägerschaft. 2005 sind 34 Krankenhäuser Hochschulkliniken (1,8 %) mit 44.161 Betten (9,1 %); 1568 Krankenhäuser sind in den Krankenhausplan eines Bundeslandes aufgenommen (*Plankrankenhäuser*, 84,9 %) mit 430.202 Betten (88,7 %); von den verbleibenden sind 101 Krankenhäuser mit Zulassung zur Behandlung von GKV-Versicherten (*Krankenhäuser mit Versorgungsvertrag*, 5,5 %) mit 8.052 Betten (1,7 %) und 143 sonstige (7,7 %) mit 2.540 Betten (0,5 %).

Seit der Erstfassung des Krankenhausfinanzierungsgesetzes 1972 werden die Kosten der Krankenhäuser zweigeteilt in Investitionskosten und pflegesatzfähige Kosten. Die pflegesatzfähigen Kosten sind dabei von den Versicherten bzw. deren Krankenkassen zu tragen. Investitionen tragen die Bundesländer gemäß ihrem jeweiligen Krankenhausplan. *Hat ein Krankenhaus auch für die Versorgung der Bevölkerung anderer Länder wesentliche Bedeutung, so ist die Krankenhausplanung insoweit zwischen den beteiligten Ländern abzustimmen (§ 6 (2)).*

3.2.3. Der Markt eines (einzelnen) Krankenhauses

Die Ökonomisierung des Gesundheitssektors hat dazu geführt, dass die Gesundheitseinrichtungen als Wirtschaftsunternehmen gesehen werden, die ihre Leistungserstellung unter dem Gesichtspunkt der Wirtschaftlichkeit und der Qualität erbringen müssen (Haubrock, 2005). Wettbewerb wird auch im Krankenhaussektor als Mittel angesehen, das Leistungsangebot stärker an den Präferenzen der Patienten auszurichten, die Qualität der Leistungen zu erhöhen, die Kosten zu senken und die Mittelzuteilung fairer zu gestalten (Robra et al., 2002).

Während für die öffentlichen Träger der Versorgungsauftrag ihrer Krankenhäuser im Mittelpunkt steht (oder stehen sollte) und die gemeinnützigen – steuerlich begünstigt – bestimmte bedarfswirtschaftliche sowie karitative, diakonische, humanitäre oder weltanschauliche Ziele mit ihren Krankenhäuser verfolgen, stehen für die Krankenhäuser in privatwirtschaftlicher Trägerschaft erwerbswirtschaftlich-kommerzielle Ziele im Vordergrund. Nicht nur durch die staatlichen Vorgaben sondern auch durch die Leistungsfähigkeit und den zunehmenden Marktanteil der privatwirtschaftlichen Krankenhäuser steigt der Druck auf die Krankenhäuser der anderen Trägerschaften, sich ebenfalls aktiv im Markt zu positionieren und zu behaupten.

Mühlbauer (2005) stellt fest, dass Menschen, die als Patienten in ein Krankenhaus kommen, in der Regel durch niedergelassene Ärzte eingewiesen sind. Dabei können die Empfehlungen des Arztes durchaus mit denen von Freunden oder Verwandten als ehemaligen Patienten konkurrieren. *Die Qualität der Beziehung zwischen niedergelassenem Arzt und einem Krankenhaus ist für die Belegungssituation, für die Quantität und Qualität der notwendigen und angebotenen Diagnose- und Therapiemöglichkeiten, für die Vorhaltung und zeitliche Beanspruchung des Personals und für den Einsatz von Materialien (z. B. Medikamente, Verbandstoffe) von besonderer Bedeutung* (Mühlbauer, 2005). Deshalb sind die Krankenhäuser gut beraten, ihre Marktposition auch anhand folgender Fragen zu bestimmen: (a) Welche Ärzte weisen in das Krankenhaus ein, welche nicht? (b) Gibt es Schlüsseleinweiser mit überproportional hohen Anteilen? (c) Werden einzelne Fachabteilungen von bestimmten Ärzten bevorzugt? (d) Wie ist die Entwicklung des Einweisungsverhaltens? (e) Wie positionieren sich die Mitbewerber? Ist der Mensch bereits Patient, so gilt es, ihn als Kunden zu pflegen, denn mit einem positiven Behandlungsverlauf erstrecken sich des Patienten Interessen zunehmend auf Service- und Hotelfaktoren, er ist als Nachfrager gegebenenfalls bereit, gesondert zu bezahlen.

Die Menschen als potenzielle Patienten oder Kunden sind der Markt eines jeden Krankenhauses – sehr in Abhängigkeit von der Krankheit des Patienten und vom Leistungsspektrum des jeweiligen Krankenhauses. *die Mobilität der Bevölkerung nimmt zu. Es ist aber nicht geprüft, welche Patienten welche zusätzlichen Wege für welche Leistungen in Kauf nehmen mögen und welche Informationen und Anreize sie dafür brauchen. Für gut planbare elektive Leistungen (...) lässt sich am ehesten ein Vorsorgetourismus erwarten, Für akute Probleme und solche mit*

wohnnaher Dauerbetreuung ist er schwer vorstellbar. Angehörige unterschiedlicher sozialer Schichten und Altersgruppen werden unterschiedlich mobil sein. So die Einschätzung von Robra et al. (2002). Dazu kommt: Wenig überlappende Krankenhaus-Einzugsgebiete führen zu einer lokalen Marktmacht regionaler Krankenhäuser. Diese nimmt zu, wenn sie (wie informell auch immer) in überregionalen Krankenhausketten zusammengeschlossen sind. Marktmacht bekommen mit den Krankenhäusern auch die Spezialisten, die den wesentlichen Teil der Einnahmen erwirtschaften und gewährleisten.

Den Markt eines Krankenhauses, die potenziellen Patienten, zu bestimmen, ist deshalb eine herausfordernde Aufgabe. Gresenz et al. (2004) vergleichen unterschiedliche Marktregionsdefinitionen: geopolitisch-administrative Einteilungen, Krankenhausumkreise fester vorgegebener Radien (wie 25 km) und solche mit Radien, in denen ein bestimmter Prozentsatz (z. B. 75 oder 90 %) der – bisherigen – Patienten des Krankenhauses leben. Gresenz et al. kommen zu dem Schluss, dass, obwohl keine Methode universell einsetzbar ist, die meisten ökonomisch-wissenschaftlichen Fragen gut mit einem prozentsatzbasierten Radius-Ansatz zu bearbeiten sind. Sie wählen den Herfindahl-Hirschmann-Index als Maß der Konkurrenz zwischen den Krankenhäusern.

Ob und inwieweit sich die USA-basierten Untersuchungen von Gresenz et al. (2004) und deren Referenzautoren auf deutsche Verhältnisse übertragen lassen, wäre genauso zu prüfen, wie der Einfluss des Vorhandenseins einer oder mehrerer Spezialkliniken in einem Krankenhaus oder in Krankenhausverbänden.

Ähnliche Marktüberlegungen lassen sich für sowohl für Arztpraxen, Gemeinschaftspraxen und Medizinische Versorgungszentren³⁵ als auch für Rehabilitations- und Vorsorgeeinrichtungen anstellen. Dabei wird den einzelnen Teilthemen unterschiedlich Gewicht beizumessen sein, für den Markt des vom Patienten selbst gewählten Hausarztes ist die Nähe ein bedeutendes Argument, für eine Rehabilitations- oder Vorsorgeeinrichtungen dürfte ein überregional guter Ruf die Marktposition erheblich verbessern. Wenig überraschend ist, dass Corporate Design für Krankenhäuser sowie Rehabilitations- oder Vorsorgeeinrichtungen ein immer wichtigerer Teil der Wahrnehmung im Markt wird (Mühlbauer, 2005).

³⁵ seit 2004 mit dem GKV-Modernisierungsgesetz zugelassene Einrichtung zur ambulanten Versorgung mit angestellten Fachärzten mehrerer Spezialisierungen.

3.2.4. Leistung und Leistungsmessung von Krankenhäusern

Durch die Gegenüberstellung von Erlösen und Kosten lässt sich die Wirtschaftlichkeit (Effizienz) einer Leistungserbringung ermitteln. Die Effizienz ist somit das Ergebnis der Division zweier Geldbeträge (Haubrock, 2005) – für die Krankenhäuser aktuell: Fallpauschalen und Pflegesätze gegen die entstandenen Kosten. Diese monetäre Sicht kann den derzeitigen wirtschaftlichen Erfolg des Krankenhauses repräsentieren, allumfassend die Leistung eines Krankenhauses beschreiben kann sie nicht.

Die Beschreibung der Leistungen von Krankenhäusern ist schwierig, da deren Output inhomogen und nur schwer messbar erscheint. Als Ziel der – personalintensiven – Produktion eines Krankenhauses lässt sich die positive Beeinflussung des Gesundheitszustandes – im günstigen Fall die Heilung, sonst auch die Eindämmung bzw. Beherrschbarkeit des Krankheitszustands einschließlich der Schmerzlinderung – ansehen. Bei der wirtschaftlichen Betrachtung von Krankenhäusern werden u. a. die (kumulative) Arbeitszeit der Ärzte, der Pfleger und des sonstigen Personals, der Bedarf an medizinischen Verbrauchsgütern, oder die Sachkosten insgesamt als Input-Indikatoren, die Anzahl der Operationen, der Untersuchungen oder der Behandlungsfälle und Pfl egetage als Output-Indikatoren herangezogen (z. B. Breyer et al., 2005). Das (Plan-)Bett wird ebenfalls als Output-Indikator verwendet – das leere Bett verstanden als Leistung des Krankenhauses zur Versorgungssicherheit der Bevölkerung. Will man Patienten (bzw. Fälle) für Leistungseinschätzung differenzieren – nach ihrer Schwere und dem nötigen Aufwand des Krankenhauses bei der Behandlung, so sollen die Fälle in eine möglichst überschaubare Anzahl möglichst homogener Gruppen eingeteilt werden (Breyer et al., 2005, S. 335), genau das ist auch Ziel bei der Entwicklung von Fallpauschalensystemen.

2002 erscheint unter dem Titel *Qualität im Krankenhaus – Überlegungen zu Begriffen und Methoden der Leistungsbewertung* (Matthes und Wiest, 2002) ein Statusbericht, der ausführlich aufzeigt, was bei einer Bestimmung der Leistung eines Krankenhauses zu bedenken ist: von möglichen Indikatoren, ihrer Reliabilität, Validität und der Notwendigkeit von Risikoadjustierungen bis hin zu den nötigen Daten, -aufbereitungstechniken und Analyseeinheiten. Nur allgemein akzeptierbare und anerkannte Maße (oder Instrumente) können dann sowohl zur kranken-

hausinternen Prozesskontrolle als auch zum Vergleich zwischen Krankenhäusern herangezogen werden.

Weltweit wird an der Entwicklung solcher Maße gearbeitet, doch es gibt sie bisher nicht, wie auch 6 Jahre später Smith et al. (2008) zu entnehmen ist. Das Ziel bleibt, denn der *Zugang zu zuverlässigen Vergleichsdaten von Qualität und Wirtschaftlichkeit medizinischer Leistungserbringer ist Voraussetzung für die Wahl einer effektiven und wirtschaftlichen Versorgung* (Schneeweiss und Sangha, 2001).

3.2.5. Die kumulative Krankenhausleistung eines Gebietes

Krankenhäuser wirken primär im und für ein Gebiet, das in seiner Größe von Einwohnerdichte, Konkurrenzkrankenhäusern und dem eigenen Leistungsspektrum stark beeinflusst sein kann. Die Größe des Einflusses für das unmittelbare Umfeld und ferner liegende Gebiete ist gewöhnlich unbekannt.³⁶

Aus Sicht der Patienten sind konkurrierende lokale Krankenhäuser Optionen; aus Sicht der jeweiligen Gebietskörperschaft stellen die Krankenhäuser gemeinsam das verfügbare Angebot an Krankenhausleistungen dar. Leistungsmessung der Krankenhäuser muss nach Matthes und Wiest (2002, S. 167) für verschiedene Analyseeinheiten (Individuen wie Patienten oder Ärzte; Gruppen wie Fachgebiete oder Stationen; ganze Organisationen wie Krankenhäuser oder Krankenhausketten) durchgeführt werden; kumuliert ergibt sich die Krankenhausleistung (Abschn. 5.1.2: Gesamtkrankenshausaktivität) eines Gebietes – über alle verschiedenen Krankenhausträger hinweg.

3.2.6. Besonderheiten des Gesundheitsmarktes

Der Gesundheitsmarkt verhält sich nicht nach den Regeln anderer Märkte. Mehr noch als in andere Märkte wird seitens der Politik in den Gesundheitsmarkt insgesamt und in die Beziehungen zwischen den Marktteilnehmern eingegriffen, der

³⁶ Einflussgebiete von Krankenhäusern sind nicht nur vom (guten) Ruf – verbreitet durch Ärzte und vormalige Patienten –, sondern auch von lokaler, fachgebietsspezifischer Nichtverfügbarkeit von Alternativen abhängig.

Markt wird in gesamtgesellschaftlichem Interesse stark reguliert. Es gibt eine Triade von Marktteilnehmern bestehend aus den Patienten als Kunden (Nachfrage), den Versicherern und anderen Kostenträgern (Finanzierung) und dem Angebot der Leistungserbringer (Abb. 3.2). Das Besondere dabei ist die Einbindung der Kostenträger zwischen den Kunden und die Leistungserbringern (von Arztpraxen bis Rehabilitationseinrichtungen). Der Kunde möchte Qualität, dennoch wird *das Verhältnis zwischen Preis und Qualität ... zwischen Versicherern und Leistungserbringern ausgehandelt* (de Pouvourville, 2002).

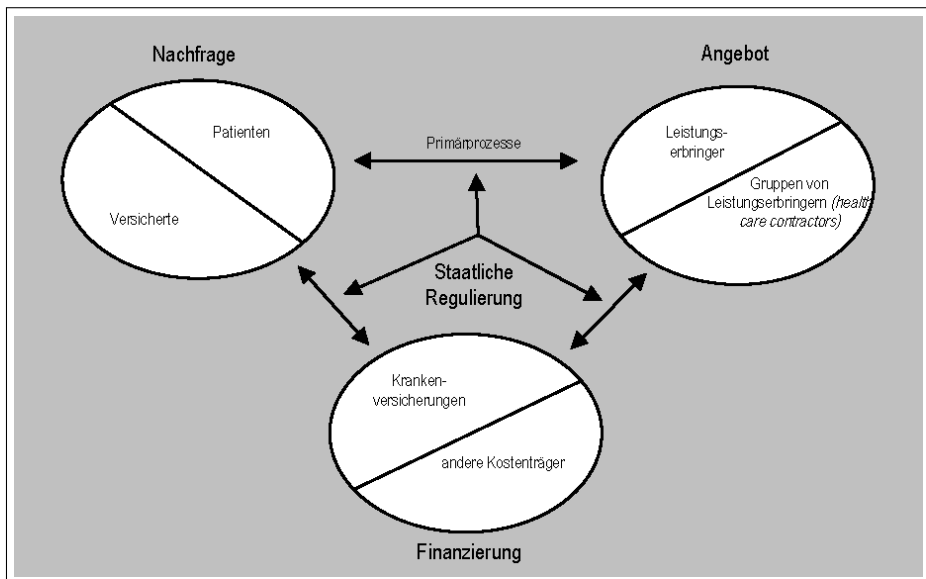


Abb. 3.2.: Gesundheitswesen: Nachfrage, Angebot, Finanzierung und staatliche Eingriffe^x

^x Aachener Kompetenzzentrum Medizintechnik, AGIT mbH, Deutsche Gesellschaft für Biomedizinische Technik im VDE und Konsortialpartner. Situation der Medizintechnik in Deutschland im internationalen Vergleich. Studie im Auftrag des BMBF. http://www.gesundheitsforschung-bmbf.de/_media/MTStudiell_Titel-Inhalt.pdf und http://www.gesundheitsforschung-bmbf.de/_media/MTStudiell_Kap11.pdf, Zugriff am 14. 08. 2011

Unabhängig davon, ob der jeweilige Leistungserbringer intern ein System der Leistungseinschätzung etabliert hat, stehen im Gesundheitswesen bisher vergleichende Informationen zur Qualität und Wirtschaftlichkeit weder den Patienten noch den überweisenden Ärzten oder den Kostenträgern in ausreichendem Maße zur Verfügung. Schneeweiss und Sangha (2001) führen aus: *Durch das bestehende*

Informationsungleichgewicht können sich Angebot und Nachfrage nicht ausreichend nach Qualität und Wirtschaftlichkeit orientieren. Daraus leitet sich neben dem Bedarf der Patienten auch ein gesellschaftlicher Bedarf an Leistungsvergleichen ab.

Verbesserte Mittel der Leistungseinschätzung und des Leistungsvergleichs ermöglichen der Politik und allen Beteiligten eine bessere Beurteilung des Erfolgs oder Misserfolgs regulierender Maßnahmen.

3.3. Themenbezogene Daten aus Deutschland

Für ein objektiviertes Bild der Wirklichkeit werden häufig Einzelinformationen zusammengetragen, kumuliert und in ihrer Gänze interpretiert. Daten sollten dabei in wohlstrukturierter Form als Statistiken gehalten, verarbeitet und herausgegeben werden können. Struktur und Inhalt der Daten und Statistiken werden hauptsächlich vom Eigner und seinen Zielen bestimmt. So werden behandelndes Krankenhaus, zahlende Krankenkasse und Statistisches Bundesamt durchaus unterschiedliche Daten in ihren jeweiligen Statistiken zu den Behandlungsfällen im Krankenhaus zusammentragen (lassen).

3.3.1. Amtliche Daten

Amtliche Daten seien hier als Daten und Statistiken aus dem Bereich der Amtlichen Statistik verstanden. Gablers Wirtschaftslexikon definiert Amtliche Statistik als *Gesamtheit der von Bund, Ländern, Gemeinden oder von speziellen Trägern durchgeführten Statistik.*³⁷ und erklärt dazu: *Neben dieser ‚ausgelösten‘ Statistik der Statistischen Ämter steht die ‚Ressortstatistik‘ solcher Behörden, bei denen die Statistik nicht aus dem Bereich der fachlichen Verwaltung herausgelöst ist, u. a. bei Bundesministerien, ..., Bundesagentur für Arbeit, ...*³⁸

³⁷ <http://wirtschaftslexikon.gabler.de/Archiv/959/amtliche-statistik-v7.html> – Gabler Wirtschaftslexikon Online, Zugriff am 12. 04. 2011.

³⁸ *ibid.*

Amtliche Daten benötigen eine Rechtsgrundlage; der Gesetzgeber beauftragt die Amtliche Statistik mit der Vorbereitung, Erhebung, Verarbeitung und Speicherung einzelner Statistiken. Gemäß Bundesstatistikgesetz (BStatG, § 9 Abs. 1) muss die Rechtsgrundlage einer Statistik deren Erhebungsmerkmale, die für die Durchführung der Erhebung erforderlichen Hilfsmerkmale, die Art der Erhebung, den Berichtszeitraum, den Berichtszeitpunkt der Datenlieferung an das StLA, die Periodizität der Erhebung und den Kreis der Befragten benennen. Wissenschaftliche (oder andere) Partikularinteressen sind weder für eine Erweiterung oder Veränderung einer bestehenden noch für die Einführung einer neuen amtlichen Statistik ausreichend.

Die Amtliche Statistik in Deutschland erhebt Daten in Statistiken zu einer Vielzahl von Themenkreisen:³⁹ (a) Gebiet, Bevölkerung, Erwerbstätigkeit, Wahlen; (b) Bildung, Soziales, Gesundheitswesen, Rechtspflege; (c) Gebäude und Wohnen, Flächennutzung, Umwelt; (d) Landwirtschaft, verarbeitendes Gewerbe, Bergbau, Energie- und Wasserversorgung, Baugewerbe, Handel, Gastgewerbe, Tourismus, Verkehr, Dienstleistungen; (e) Außenhandel, Handwerk, Unternehmen und Betriebe, Gewerbeanzeigen, Insolvenzen; (f) Preise, Verdienste, Arbeitskosten; (g) Öffentlichen Haushalte, Personal im öffentlichen Dienst; (h) Volkswirtschaftliche Gesamtrechnungen. Einige dieser sind für eine Analyse der Beziehungen zwischen Wirtschaft und Gesundheit im Transitionsprozess Ostdeutschlands von größerem Belang.

3.3.2. Gesundheitsdaten

Die Krankenhausstatistik gehört zu den umfangreichsten amtlichen Statistiken über das Gesundheitswesen. Sie erfasst jährlich alle Krankenhäuser und hält Daten über die Art und die Trägerschaft des Krankenhauses, die dort Tätigen und Auszubildenden, die organisatorischen Einheiten, die sachliche Ausstattung, die Kosten und die im Krankenhaus in Anspruch genommenen Leistungen. Daneben enthält sie zu jedem Patientenaufenthalt die Hauptdiagnose, demografische Angaben und die Verweildauer. Patienten mit wiederholtem Aufenthalt selbst in demselben Krankenhaus erscheinen als neuer Fall, eine Patientenzuordnung erfolgt nicht – die Diagnosedaten bilden eine Fallstatistik, keine Personenstatis-

³⁹ <http://www.statistikportal.de/>, Zugriff am 10. 04. 2011.

tik. Rehabilitations- und Vorsorgeeinrichtungen berichten einen eingeschränkten Merkmalsumfang an die Krankenhausstatistik, Diagnosen müssen nur von Einrichtungen mit 100 oder mehr Betten erhoben werden. Die Krankenhausstatistik enthält somit Angaben zum gesellschaftlich notwendigen Aufwand für die Bekämpfung von Krankheit und die (nachhaltige) Verbesserung des Gesundheitsstatus der Behandelten. Medizinisch wird der einzelne Behandlungsfall über seine (Haupt-)Diagnose abgebildet, der wirtschaftliche Aufwand erscheint aggregiert, er lässt sich nicht einem jeden Einzelfall zuordnen.

Seit 2005 erhebt die amtliche Fallpauschalenbezogene Krankenhausstatistik⁴⁰ (DRG-Statistik) für alle Krankenhäuser, die nach dem DRG-Vergütungssystem abrechnen je Fall Angaben zu Erkrankungsart, Operationen und Prozeduren sowie Verweildauer und Fachabteilung. Diese neuere Statistik beschreibt den fallspezifischen wirtschaftlichen Aufwand der Krankenhäuser für jeden Behandelten, indem auf den letztendlich bundesweiten Daten vorangehender Jahre beruhend einem jedem DRG-Code ein akzeptierter und in Rechnung stellbarer Aufwand zugeordnet wird. Die Gesamtleistung des Krankenhauses ergibt sich als Summe.

Desweiteren gibt es die jährliche amtliche Todesursachenstatistik als Totalerhebung mit Angabe der Diagnose, die zum Tod führte. Diese Statistik beschreibt den Krankheitsstatus der Bevölkerung zu einen äußerst spezifischen Moment, dem des Todes, und ist für die Beschreibung des Gesundheitszustandes der Gesamtbevölkerung nur bedingt geeignet.

Im Rahmen der Gesundheitsberichterstattung des Bundes⁴¹ – einer gemeinsamen Aufgabe des Robert-Koch-Instituts und des StBA – wird versucht, amtliche (u. a. die Krankenhausstatistik) und nichtamtliche Statistiken (Statistiken von kasernenärztlichen Vereinigungen, Ärztekammern, etc. – aus etwa 60 Datenquellen⁴²) zusammenzufassen.

⁴⁰ Die Fallpauschalenbezogene Krankenhausstatistik ist kein Teil der Krankenhausstatistik; sie wird nicht über die StLÄ sondern über das Institut für das Entgeltsystem im Krankenhaus (InEK) erhoben und dann dem StBA zur Verfügung gestellt.

⁴¹ <http://www.gbe-bund.de/>.

⁴² http://www.gbe-bund.de/gbe10/hrecherche.prc_datensquellen?p_aid=95646734&p_uid=gast&p_sprache=D&p_knoten=ANDERE&tk=51310&tk2=51312&cnt_ut=1&ut=51312, Zugriff am 19. 04. 2011.

Eingedenk der Rechtslage und des zu vermutenden hohen Aufwandes ist es wenig überraschend, dass es keine jährliche Gesundheits-Totalerhebung gibt. Das Robert-Koch-Institut führt im Auftrag des Bundesministeriums für Gesundheit wiederholt Gesundheitssurveys durch, seit 2009 als telefonische Befragungen im Rahmen des Gesundheitsmonitorings (*Gesundheit in Deutschland Aktuell* (GEDA)⁴³). GEDA 2009 befragte 21.000 deutschsprachige Erwachsene zum subjektiven Gesundheitsempfinden und einer Reihe von spezifischen Krankheitsbildern (Robert-Koch-Institut, 2011). Auch das Sozio-Ökonomische Panel (SOEP), die größte stichprobenbasierte wiederkehrende Bevölkerungsbefragung in Deutschland, enthält neben einer Frage zur Zufriedenheit mit der eigenen Gesundheit eher Fragen zur Nutzung des Gesundheitssystems im Krankheitsfall und zu Krankheit als Grund für den Verlust des Arbeitsplatzes⁴⁴. Böhm (2005) gibt einen weiterhin gültigen allgemeinen Überblick über die *Gesundheitsdaten in Deutschland* unter Ausschluss der Daten in kommerzieller, privatwirtschaftlicher Eignerschaft.

Es fällt auf, dass es eine die Krankenhausstatistik ergänzende Statistik zu den Leistungen des Gesundheitswesens im ambulanten Bereich im Bereich der Amtlichen Statistik nicht gibt.⁴⁵

Gesundheitsdaten sind – im Bereich der Amtlichen Statistik ebenso wie bei den anderen Dateneignern – hauptsächlich Krankheitsdaten. Den umfassendsten Überblick zu Aufwand und Leistungen bezüglich der Gesundheit der Bevölkerung gibt die amtliche Krankenhausstatistik, auch wenn die Krankenhausstatistik keinen allgemeinen Gesundheitsüberblick sondern eine Beschreibung der eher schwersten Krankheitsfälle liefert.

⁴³ http://www.gbe-bund.de/gbe10/ergebnisse.prc_tab?fid=9774&suchstring=bundesgesundheitsurvey_Koch-Institut&query_id=&sprache=D&fund_typ=DQM&methode=2&vt=1&verwandte=1&page_ret=0&seite=&p_lfd_nr=6&p_news=&p_sprachkz=D&p_uid=gast&p_aid=52324990&hlp_nr=3&p_janein=J, Zugriff am 20. 04. 2011.

⁴⁴ http://www.diw.de/de/diw_02.c.240864.de/soep_codebooks.html, Zugriff am 07. 04. 2011.

⁴⁵ Ob die Daten der Krankenkassen, der kassenärztlichen Vereinigungen und der Privatversicherer gemeinsam eine Statistik der ambulanten Versorgung in Deutschland ergäben, muss hier offen bleiben.

3.3.3. Sozialdaten

Arbeitslosigkeit ist für die meisten Betroffenen unerwünscht und wird u. a. des Verdienstwegfalls wegen für das eigene Wohlergehen als problematisch empfunden. Das Einkommen wiederum spiegelt in den meisten Fällen den Wohlstand und die soziale Position in der Gesellschaft wieder.

Arbeitslosenzahlen werden von der Bundesagentur für Arbeit (BA, bis Ende 2002⁴⁶ Bundesanstalt für Arbeit) für jeden Kreis ermittelt und als amtliche Statistik zusammengestellt, die Arbeitslosenquoten daraus abgeleitet⁴⁷. Die Veränderungen der Arbeitslosenzahlen und -quoten eines jeden Kreises werden als Zeitreihen verstanden und stehen als solche im offenen Zugriff über den Internetauftritt der BA zur Verfügung.

Die Deutsche Rentenversicherung Bund ist Eigner wesentlicher Daten zur gesetzlichen Rentenversicherung in Deutschland. Der Rentenanspruch eines jeden Versicherten wird dabei monatlich kumulativ mittels der individuellen Entgeltpunkte festgehalten. Die Entgeltpunkte eines Jahres⁴⁸ berechnen sich dann aus den rentenversicherungspflichtigen Bruttoeinkünften aus unselbständiger Arbeit, sie entsprechen somit dem Jahresarbeitseinkommen. Nicht erfasst sind Beamte und Einkünfte, die nicht aus unselbständiger Arbeit erzielt werden. Wegen der Beitragsbemessungsgrenze unterliegen die Daten einer Rechtszensurierung; aufgrund gesetzlicher Regelungen für langjährige Einzahler auch eine Linkzensurierung. Trotzdem gilt die Summe der Entgeltpunkte als ein *exzellentes Maß für die tatsächliche Einkommensposition von überwiegend abhängig Beschäftigten* (von Gaudecker, 2006). Gemittelt über alle in einem Kreis wohnhaften (abhängig) Beschäftigten können die jährlichen Entgeltpunkte eine Näherungsvariable für den einkommensbasierten Wohlstand dieses Kreises sein.

Die Rentenstatistiken der Deutschen Rentenversicherung und die Arbeitslosenzahlen und -quoten der Bundesagentur für Arbeit (BA) sind somit interessante

⁴⁶ http://www.arbeitsagentur.de/nn_27200/zentraler-Content/A01-Allgemein-Info/A015-Oeffentlichkeitsarbeit/Allgemein/Geschichte-der-BA.html, Zugriff am 14. 09. 2010.

⁴⁷ Die Bezugsgrößen für die Berechnung der Arbeitslosenquoten werden einmal jährlich aktualisiert, <http://www.pub.arbeitsamt.de/hst/services/statistik/000000/html/start/arbeitslosenquote.shtml>, Zugriff am 09. 05. 2007.

⁴⁸ Die Entgeltpunkte werden monatsbezogen in der Statistik gehalten (Himmelreicher, 2006).

Sozialdaten aus dem Bereich der Amtlichen Statistik. Gleichzeitig sind sie Wirtschaftsdaten, die eine Form notwendiger Ausgaben der Gemeinschaft für Einzelne (Arbeitslose) und andererseits über den Wohlstand Kaufkraft und potenziell möglichen Konsum beschreiben.

3.3.4. Wirtschaftsdaten

Für zwei wichtige Bereiche der die Wirtschaft beschreibenden Statistiken gibt es im Bereich der StLÄ Arbeitskreise, von denen Wirtschaftsdaten regelmäßig wiederkehrend zusammengestellt und veröffentlicht werden. Diese Statistiken sind Sekundärstatistiken, denn ihre Ergebnisse entstehen nicht durch unmittelbar zugeordnete Erhebungen sondern durch Aufbereitung der Daten von Primärstatistiken im Bereich der StLÄ.

Der 1954 gegründete Arbeitskreis Volkswirtschaftliche Gesamtrechnung der Länder⁴⁹ erarbeitet die regionalen Ergebnisse der Volkswirtschaftlichen Gesamtrechnungen für Bundesländer und Kreise und stellt sie der Öffentlichkeit zur Verfügung. *Die Volkswirtschaftlichen Gesamtrechnungen stellen das umfassendste statistische Instrumentarium der Wirtschaftsbeobachtung dar.*⁵⁰ Alle Berechnungen werden nach dem Europäischen System Volkswirtschaftlicher Gesamtrechnungen (ESVG 1995) durchgeführt; dabei werden für die Kreise nur ausgewählte Aggregate der VGR berechnet: Brutto-Inlandsprodukt (BIP), Bruttowertschöpfung nach Wirtschaftsbereichen, Arbeitnehmerentgelt bzw. Bruttolöhne und -gehälter (Inlandskonzept) nach Wirtschaftsbereichen sowie Primär- und verfügbares Einkommen der privaten Haushalte⁵¹. Daten liegen seit 1996 jährlich vor, für die Jahre 1992, 1994 und 1995 nur zu einem eingeschränkten Merkmalsatz.

Seit 1984 ermittelt und veröffentlicht der Arbeitskreis Erwerbstätigenrechnung der Länder⁵² die durchschnittliche Zahl der Erwerbstätigen sowie deren geleistete Arbeitsstunden in einem bestimmten Berichtszeitraum für die Bundesländer und Kreise. Die Daten enthalten sowohl die Anzahlen der Erwerbstätigen als auch

⁴⁹ http://www.vgrdl.de/Arbeitskreis_VGR/, Zugriff am 16. 04. 2011.

⁵⁰ *ibid.*

⁵¹ Erwerbstätige bzw. Arbeitnehmer (Inlandskonzept) nach Wirtschaftsbereichen sind als Übernahme vom Arbeitskreis ETR, Einwohnerzahlen als Übernahme vom StBA hinzugefügt.

⁵² <http://www.aketr.de/>, Zugriff am 16. 04. 2011.

der Arbeitnehmer in den Kreisen für die Jahre seit 1991 – insgesamt und nach Wirtschaftsbereichen⁵³.

Bisher wird das Brutto-Inlandsprodukt (BIP) häufig als Indikator für Wohlstand und sozialen Fortschritt verwendet. Dieses wird seit längerem infrage gestellt. Im Auftrag der französischen Regierung hat eine internationale Kommission unter Leitung von Stiglitz, Sen und Fetoussi herausgearbeitet, dass das BIP eher durch das Einkommen der Haushalte als Wohlstandsmaß und weitere, die Lebensqualität und die Nachhaltigkeit einschließlich der Umweltbelastungen beschreibende Indikatoren ergänzt oder ersetzt werden sollte (Stiglitz et al., 2009). Mit den Daten des Arbeitskreises Volkswirtschaftliche Gesamtrechnung der Länder kann zumindest dem Einkommen Beachtung geschenkt werden.

Während BIP und Bruttowertschöpfung Nahrungsvariablen für die wirtschaftliche Leistungsfähigkeit, Bruttolöhne und Anzahlen der Erwerbstätigen bzw. Arbeitnehmer die zugehörige Aufwandseite abbilden, lassen sich Primär- und verfügbares Einkommen der privaten Haushalte sowohl den Wirtschafts- als auch den Sozialdaten zuordnen.

Die öffentlich zugänglichen Daten beider Arbeitskreise werden rückwirkend dem jeweils aktuellen Gebietsstand angepasst, Kreisreformen führen somit zu Veränderungen in den Daten bereits früher berichteter Jahresergebnisse.

3.3.5. Raumordnungsdaten

Das Bundesamt für Bauwesen und Raumordnung (BBR) teilt Deutschland in siedlungsstrukturelle Gebietstypen⁵⁴ ein. Jede Gemeinde wird dabei einem von 17 Gemeindetypen (Abb. 3.3) zugeordnet, jeder Kreis gehört zu einem von 9 Kreistypen, zu einem von 7 Differenzierten Regionstypen und zu einem von

⁵³ <http://www.statistik-hessen.de/erwerbstaetigenrechnung/Veroeffentlichungen.htm>, Zugriff am 02. 01. 2007.

⁵⁴ Angaben dieses Abschnitts beruhen auf http://www.bbsr.bund.de/cln_016/nn_21360/BBSR/DE/Raumb Beobachtung/Werkzeuge/Raumabgrenzungen/SiedlungsstrukturelleGebietstypen/Kreistypen/kreistypen.html und http://www.bbsr.bund.de/cln_016/nn_340582/BBSR/DE/Raumb Beobachtung/Werkzeuge/Raumabgrenzungen/SiedlungsstrukturelleGebietstypen/Regionstypen/regionstypen.html, Zugriffe am 06. 04. 2011.

3 Regionsgrundtypen. Die Regionsgrundtypen sind dabei: Agglomerationsräume mit einem Oberzentrum mit mehr als 300.000 Einwohnern oder einer Einwohnerdichte um 300 EW/km², Verstädterte Räume mit einer Einwohnerdichte von mehr als 150 EW/km² oder einem Oberzentrum über 100.000 Einwohnern bei einer Mindestdichte von 100 EW/km², während Ländliche Räume entweder kein Oberzentrum über 100.000 Einwohner haben oder aber eine Einwohnerdichte von weniger als 100 EW/km² aufweisen.



Abb. 3.3.: Siedlungsstrukturelle Gebietstypen des BBR^x

^x http://www.bbsr.bund.de/nn_103086/BBSR/DE/Raubeobachtung/Werkzeuge/Raumabgrenzungen/SiedlungsstrukturelleGebietstypen/PDF__Download,templateId=raw,property=publicationFile.pdf/PDF_Download.pdf, Zugriff am 06. 04. 2011

Als Abgrenzungskriterien für die 9 Kreistypen dienen dabei sowohl Einwohnerzahlen als auch Einwohnerdichten: Kernstädte haben mehr als 100.000 Einwohner, ein Kreis gilt als hochverdichtet bei einer Einwohnerdichte über 300 EW/km², als verdichtet bei einer Einwohnerdichte über 150 EW/km² und als ländlich bei geringerer Einwohnerdichte. Für die ländlichen Räume liegt die Grenze zwischen höherer und geringerer Dichte bei 100 EW/km² (Stand 2011).

Die siedlungsstrukturellen Gebietstypen stellen somit Ähnlichkeiten zwischen entfernt liegenden Gemeinden bzw. Kreisen her. Sie *dienen dem intraregionalen Vergleich*⁵⁵.

Einwohnerzahlen und -dichten ändern sich langsam, die Klassifizierung nach siedlungsstrukturellen Gebietstypen könnte als relativ zeitkonstant in jegliche weitergehende Analyse einfließen. Da die Klassifizierung jedoch auf den aktuellen administrativen Einheiten (Gemeinden, Kreisen) beruht, ist diese zusätzlich mit jeder Gemeinde- oder Kreis-Reform zu überarbeiten.

Die aktuellen Zuordnungen der Gemeinden und Kreise zu siedlungsstrukturellen Gebietstypen sind öffentliche Daten und werden als Flächenkartogramm (Choroplethenkarte, Abschn. 5.1.10) veröffentlicht, aktuelle Tabellen der Kreiszuordnung versendet das BBR auf Anfrage.

3.4. Statistik in den empirischen Sozialwissenschaften

Eingangs seiner Monografie *Econometric Analysis of Cross Section and Panel Data* behauptet Wooldridge (2002, S. 3): *The goal of most empirical studies in economics and other social sciences is to determine whether a change in one variable, say w , causes a change in another variable, say y .*

Für gut etablierte und theoretisch fundierte Fragestellungen kann dem zugestimmt werden; allein für bisher eher wenig bearbeitete Fragestellungen gilt es zuvörderst, bemerkenswerte Beziehungen zwischen Variablen herauszuarbeiten, sie auf ihre sozialwissenschaftliche Relevanz hin zu untersuchen und zu beurteilen, bevor für diese Beziehungen ein kausaler Zusammenhang hergeleitet wird.

Wie bereits angeführt beginnt nach Rönz und Strohe (1994, S. 266 f.) die ökonomische Arbeit mit der Spezifikation eines Modells passend zur ökonomischen Aufgabenstellung. Allgemeiner gilt, dass die substanzwissenschaftliche Frage in ein adäquates Modell zu übersetzen ist, bevor Daten zur Beantwortung der gestellten Frage herangezogen werden können.

⁵⁵ *ibid.*

3.4.1. Empirische Untersuchungen in den Sozialwissenschaften

In den empirischen Sozialwissenschaften (von Erziehungswissenschaften über Politikwissenschaften, Sozialpsychologie bis hin zu den Wirtschaftswissenschaften) werden Erkenntnisse meist durch gezielte Beobachtungen und seltener durch geplante Experimente gewonnen. Kish (1985) gibt einen vergleichenden Überblick der wichtigsten Designkategorien: Beobachtungsstudien auf Stichprobenbasis und andere kontrollierte Beobachtungen, Totalerhebungen (auch Zensus oder Vollerhebung: alle Elemente einer Grundgesamtheit werden statistisch erhoben, vgl. Rönz und Strohe, 1994, S. 368, S. 415) und registerbasierte Untersuchungen, Beobachtungen über die Gesamtpopulation oder lokale Beobachtungen (geografisch definierter Teilpopulationen). Population – die Grundgesamtheit in der Terminologie der Stochastik (Rönz und Strohe, 1994, S. 146 f.) – kann hierbei sowohl die Bevölkerung, Teile derselben wie Rentner oder Migranten, Haushalte, Staaten oder aber auch die Menge aller Betriebe sein, ein einzelner Betrieb der Stichprobe (oder jeder einzelne Betrieb bei Totalerhebung) wäre dann ein Subjekt der Untersuchung.

Hinsichtlich der Zeitkomponente lassen sich Beobachtungsstudien grob in 3 Kategorien einteilen. In Querschnittsstudien (engl.: *cross sectional studies*) wird jedes Subjekt einmal – idealerweise zum selben Zeitpunkt – gemessen, d. h. im einfachsten Fall, dass das interessierende Merkmal an jedem Subjekt erhoben wird, z. B. die in der Nachwahlumfrage vor dem Wahllokal angegebene gewählte Partei. Eine Zeitreihe wiederum besteht aus der häufigen wiederholten Beobachtung des Merkmales an einem Subjekt, z. B. des DAX zu Handelsschluss oder des BIP eines bestimmten Landes zum Ende eines jeden Geschäftsjahres. Panel-Studien (auch als Längsschnitt-Studien bezeichnet) lassen sich als gemeinsame Verallgemeinerung der beiden vorgenannten betrachten, denn es werden je nach Betrachtungsweise mehrere Zeitreihen im Querschnitt untersucht oder jedes Subjekt einer Querschnittuntersuchung wird zu mehreren aufeinanderfolgenden Zeitpunkten gemessen. Die beiden Dimensionen – die Zeit- und die Subjektkomponente – können je Panel sehr unterschiedlich ausfallen, von wenigen Subjekten mit vielen Messungen je Subjekt bis hin zu vielen Subjekten mit wenig Wiederholungen jeweils. Das Sozio-Ökonomische Panel (SOEP) in Deutschland gehört eher in die zweite, ein Vergleich mehrerer nationaler Aktienleitindizes zum Handelsschluss

erhoben über einen Zeitraum von einem Jahr oder mehr ist Beispiel für die erste Kategorie.

Für die Zeitkomponente von Paneldaten haben sich verschiedene Begriffe eingebürgert. So spricht man bei prospektiv geplanten Paneldaten oft von Welle (z. B. beim SOEP), wobei der zeitliche Abstand zwischen zwei Wellen ist nicht unmittelbar sichtbar ist. Bei (lückenlosen) Berichtsjahren, -quartalen oder -monaten ist die zeitliche Äquidistanz bereits im Begriff gegeben.

3.4.2. Statistik für empirische Untersuchungen amtlicher Paneldaten

Betrachtet man jedoch die Panel-Studien aus dem Blickwinkel der statistischen Auswertung und Modellierung, so zeigt sich schnell, dass gute Techniken der Parameterschätzung für Panel mit vielen Subjekten – beobachtet zu im Vergleich wenigen Zeitpunkten – andere sind, als für die gemeinsame Betrachtung relativ weniger längerer Zeitreihen. Erstmals in Beck und Katz (1995) vorgeschlagen und heute in Fachzyklopädien (Beck, 2008) eingegangen ist deshalb die begriffliche Trennung in *Time-Series Cross-Section Methods* für letztere und *Paneldaten-Analysen* für erstere. Einen guten aktuellen Überblick über Theorie und Praxis der Paneldaten-Analyse in Wirtschaftswissenschaften und Ökonometrie geben Mátyás und Sevestre (2008, 1996), spezifischere Probleme der Paneldaten-Analyse werden an anderer Stelle diskutiert (Abschn. 3.4.4).

Sinnvolle Auswertung und Modellierung der Daten sind nur auf Basis und unter Nutzung adäquater Planung einer empirischen Untersuchung möglich. Bei Nutzung amtlicher Daten sind gewisse Teilaspekte des Studiendesigns bereits vorgegeben. Der maximale Datenumfang – Merkmale, Erhebungsfrequenz, Erfassungszeitraum, Erfassungseinheiten (Behandlungsfälle, Personen, Familien, Betriebe, Kreise, ...) und Gesamtumfang (Stichprobe oder Totalerhebung) – der Beobachtungen ist bereits definiert und kann für einzelne Projekte nicht einfach erweitert werden, Auswertungen erfolgen häufig retrospektiv. Trotzdem bleibt in diesem Rahmen zu entscheiden, ob mit dem vollen Datensatz oder einem Teil gearbeitet werden soll, und auf welcher Grundlage (Zufallsstichprobe, geschichtete Zufallsstichprobe, begründeter anderer Auswahlmechanismus) dieser Teil auszuwählen

ist. Es ist davon auszugehen, dass eine Analyse amtlicher Daten (fast) immer Sekundäranalyse (im Sinne von z. B. Rosenberg et al., 2006) ist, die Daten sind bereits (mindestens) einmal analysiert und interpretiert.

3.4.3. Stichprobenraum, Totalerhebungen, Verteilungsanpassungen, Gültigkeit der Modellierungen

Die hier vorliegenden Untersuchungen beruhen auf den Daten der Krankenhausstatistik sowie Teilen der Volkswirtschaftlichen Gesamtrechnung, der Erwerbstätigenrechnung und der Arbeitslosenquoten der Bundesagentur für Arbeit. Bei jeder dieser Statistiken handelt es sich um eine Totalerhebung der jeweils zugrunde liegenden Grundgesamtheit: eines jeden Krankenhauses bei Grund- und Kostendaten, eines jeden Krankenhaus-Falles für die Diagnosedaten sowie eines jeden Kreises für VGR, ETR und der Arbeitslosenquoten. Im Gegensatz zur Stichprobenerhebung liegt somit – Datenausfall ausgeschlossen – alle die komplette Grundgesamtheit beschreibende Information vor.

Ökonometrisches Modellieren hingegen bedient sich der Techniken der deskriptiven, explorativen und konfirmatorischen Statistik mit dem Ziel, über Verteilungs- und Repräsentativitätsannahmen von stichprobenbasierten Informationen auf die Verhältnisse in der Grundgesamtheit zu schließen, Parameter mit Konfidenzintervallen für die Grundgesamtheit zu schätzen und deren Bedeutung mittels Teststatistiken und den zugehörigen Ablehnungswahrscheinlichkeiten zu beurteilen. Da Totalerhebungen keine Stichprobenauswahl darstellen, erscheint dieses Vorgehen nicht unbedingt plausibel; auch neuere Publikationen zu empirisch-statistischen Methoden einschließlich Totalerhebungen allgemein (Kish, 1985), Registern und Officialstatistik (Wallgren und Wallgren, 2007) oder Ökonometrie (Wooldridge, 2002) reflektieren dieses häufig nicht.

Hingegen bot bereits Haavelmo (1944) eine Lösung dieses Dilemmas an in seinem Versuch, die damals unter Wirtschaftswissenschaftlern verbreitete Ablehnung stochastischer Modelle zu hinterfragen. Haavelmo (1944, S. 51, S. 56) postuliert, dass die wirtschaftswissenschaftlich interessierende Grundgesamtheit nicht einfach die (als potenziell unendlich groß betrachtete) Menge aller Individuen sei, sondern hingegen die unendlich große Menge von deren wirtschaftlichen Ent-

scheidungen, die zu dem beobachtbaren und als Zufallsvariable verstandenen, wirtschaftlichen Ergebnis führen können. Haavelmos Sicht ist heute weithin anerkannt (Hoover, 2006, III. Do Probability Models Apply to Economic Data?) und wird u. a. von Frees (2004, S. 77) explizit auf Analyse und Modellierung von Paneldaten angewendet; diese Interpretation ermöglicht es, die Daten von Totalerhebungen zur Grundlage statistisch-ökonomischer Modellierung zu machen.

Haavelmo folgend gehen vorliegende Untersuchungen davon aus, dass die auf Totalerhebung(en) basierenden Daten der Krankenhausstatistik, der Volkswirtschaftliche Gesamtrechnung, der Erwerbstätigenrechnung und der Arbeitslosenquoten für die Jahre 1992 bis 2005 und die Bundesländer Berlin, Brandenburg, Mecklenburg-Vorpommern, Rheinland-Pfalz, Sachsen, Sachsen-Anhalt sowie Thüringen eine Stichprobe aus der Grundgesamtheit möglicher (wirtschaftlicher) Entscheidungen ist, die von Patienten (sich einer Behandlung unterziehen, Wahl eines Krankenhauses), Ärzten (Überweisung in Krankenhaus, Festlegung der Hauptdiagnose), Klinikleitungen (Angebotsspektrum des jeweiligen Krankenhauses) und Kreistagen (wirtschaftliche Rahmenbedingungen) getroffen wurden – um nur einige Beispiele zu nennen.

3.4.4. Paneldaten: Vor- und Nachteile

Paneldaten und Methoden ihrer Analyse (Abschn. 3.4.2) werden seit über 50 Jahren ökonometrisch untersucht und angewendet. Dabei sind eine Reihe von Vorteilen als auch Nachteilen der Paneldaten-Analyse herausgearbeitet und diskutiert worden. Verglichen wird dabei mit den beiden anderen sehr üblichen Designs für Beobachtungsstudien – der Querschnittsstudie bzw. der Zeitreihenanalyse (Abschn. 3.4.1).

Zieht man Frees (2004, 1.2 Benefits and Drawbacks of Longitudinal Data), Baltagi (2005, S. 4–9) und Hsiao (2007, mit Diskussion) zu Rate, erscheinen als Nachteile bedeutend: (N.1) die größere Komplexität des Designs und die damit verbundenen höheren Datenerhebungsanforderungen, (N.2) das durch das Fortschreiben von Messfehlern bei den wiederholten Erhebungen erhöhte Risiko einer Verzerrung der Ergebnisse, (N.3) das erhöhte Verzerrungsrisiko durch die Auswahl von Panelteilnehmern gemäß ihrer Bereitschaft zu wiederholter Datenerhebung so-

wie das vorzeitige Verlassen der Panel-Studie wegen sinkender Teilnahmebegeisterung (*panel attrition*), (N.4) die Gefahr, dass Argumente der Asymptotik wegen der geringen Anzahl an Wiederholungen in der Zeitkomponente bei der statistischen Modellierung unerkannt nicht greifen, und (N.5) die Gefahr, dass durch in die Analyse nicht einbindbare Abhängigkeitsstrukturen zwischen Panelsubjekten (z. B. Ländern oder Verwandten ohne gemeinsamen Haushalt) die Modellierung inadäquat bleiben muss. Die Punkte (b) und (c) sind bei Surveys mit Freiwilligen als Teilnehmern von größerer Relevanz als bei administrativ erhobenen Daten mit Auskunftspflicht für die Teilnehmer.

Die vorgenannten Autoren führen als Vorteile an, dass: (V.1) individuelle Heterogenität beim Modellieren kontrolliert werden kann, (V.2) informativere Daten mit mehr erklärbarer Variabilität, weniger Kollinearität und mehr Freiheitsgraden für letztendlich mehr Effizienz beim Modellieren vorliegen, (V.3) Dynamik – die Veränderung über die Zeit – vorteilhafter analysierbar ist, (V.4) Effekte schätzbar sind, die im Querschnitt nicht schätzbar sein müssen, (V.5) komplexere – hoffentlich wirklichkeitsnähere – Modelle schätzbar sind, und dass (V.6) Mikro-Niveau-Daten von Individuen, Firmen oder Haushalten exakter messbar und somit von besserer Qualität sein können als ihre Pendants auf Makro-Niveau.

Lineare wie zunehmend auch nichtlineare Regressionsmethoden für Paneldaten werden angewendet, und es wird versucht, Methoden für die gültige Einbindung von endogenen Variablen in die Modellgleichungen zu entwickeln. Die Vorteile sind so überzeugend, dass nach Ausweitungen der Methode gesucht wird. So entwickelt Deaton (1985) ein Modell, in dem nicht vorhandene Paneldaten durch eine Abfolge von (voneinander unabhängigen) Querschnittsdaten angenähert werden können. Genutzt wird dabei eine Kohortenstruktur: jeder der Querschnitte hat Daten zu (anderen) Subjekten der jeweiligen Kohorten, die Kohorten sind somit die Einheiten der wiederholten Beobachtung.

Effekte: fest oder zufällig?

Eine sehr wichtige Frage (auch) beim Modellieren von Paneldaten ist die nach dem Typ des zu wählenden Modells. Churchill Eisenhart war der erste, der eine klare Unterscheidung zwischen Modellen *mit festen Effekten* und Modellen *mit*

zufälligen Effekten (Class I bzw. Class II, Eisenhart, 1947) hinsichtlich ihrer Eigenschaften forderte. Sind alle Regressoren Merkmale, deren Werte vor einem Versuch festgelegt werden können, dann ist das zugehörige Modell aus Class I; bei Class II-Modellen sind alle Regressoren zufällig, nicht vor dem Versuch fixierbar und somit Beobachtungswerte. Beide Modelle unterscheiden sich sowohl in der Modellierung als auch der Interpretation der Ergebnisse.

Für Modelle mit vielen Regressoren ist es häufig sinnvoll, sowohl feste Effekte (Class I) als auch zufällige Effekte (Class II) in die Modellierung einzubeziehen; für diese hat sich der Begriff *gemischte Modelle* eingebürgert. Gardiner et al. (2009) geben unter dem Titel *Fixed effects, random effects and GEE: What are the differences?* einen aktuellen Überblick zu den Begriffen und Modellen – einschließlich Paneldaten-Modellen –, zu deren Unterschieden und Ähnlichkeiten sowie den zugrunde liegenden Annahmen und Auswirkungen auf die mögliche Interpretation. Zur Verdeutlichung wählen sie die Daten einer Studie zur Betreuung von Schwangeren in einem County in Michigan. Diese Daten werden mit verschiedenen der Modelle ausgewertet.

Berrington et al. (2006) diskutieren speziell für Paneldaten-Analyse die Unterschiede zwischen einer Modellierung der Zeitkomponente als fester oder als zufälliger Effekt. Im ersten Fall wird der zeitliche Verlauf als eine Folge diskreter Übergänge zwischen aufeinanderfolgenden Zeitpunkten betrachtet, während Berrington et al. (Kap. 5 Random Effects Models for Repeated Measures Data) zufolge der zweite Fall die Betrachtung der Zeitpunkte als Querschnitte eines kontinuierlichen (Entwicklungs-)Prozesses zulässt. Einschluss der Zeitkomponente nicht nur linear sondern auch quadriert (oder mit anderen Exponenten) erweitert den Kreis der möglichen Modelle.

Eine wesentliche Voraussetzung für die Anwendung von Modellen mit zufälligen Effekten ist die Unkorreliertheit zwischen den zufälligen Effekten und allen zeitkonstanten und zeitveränderlichen Kovariaten im Modell. Der Hausman-Test (Hausman, 1978) adressiert diese Frage mit der Nullhypothese der Unkorreliertheit.

Speziell in der ökonometrischen Literatur wird oft anhand der Ergebnisse des Hausman-Test entschieden, ob das Modell einen speziellen Effekt als zufällig enthalten sollte oder nicht. Argumente gegen ein zu automatisiertes Anwenden die-

ser Strategie finden sich u. a. in Fielding (2004). Einer sachlichen Begründung für ein gewähltes Modell sollte häufiger Vorzug gegeben werden gegenüber einem rein datengesteuerten Ergebnis eines statistischen (Vor-)Tests, speziell wenn dieser Test bei großen Fallzahlen ein ‚signifikantes‘ Ergebnis zeigt. Baltagi (2005, S. 19 Fußnote 6 und S. 29) stellt die Hausman-Test-basierte Entscheidung einer Modellwahl unter Nutzung des Informationskriteriums nach Akaike (AIC) oder seines Bayesianischen genannten Pendant (BIC) zugunsten letzterer gegenüber.

Für Sekundärdatenanalysen (Abschn. 4.1.5) ist die begründete Auswahl des Typs der in das Modell einbezogenen Regressoren ein wichtiger Schritt; das gilt für Querschnittsdaten und mehr noch für die Analyse von Paneldaten.

3.5. Register und registerbasierte Statistik

Gewöhnlicherweise wird für die Analyse von Daten aus administrativen Quellen davon ausgegangen, dass es sich um amtliche Daten entweder einer Stichprobe oder eines Zensus handelt, die behördlicherseits für einen bestimmten Zweck, zu einem (mehr oder minder klar umrissenen) Thema erhoben wurden. Für beide Typen (Stichprobe, Zensus) gibt es entwickelte Theorien zur Planung, Durchführung und statistischen Analyse.

Eine registerbasierte Untersuchung – der dritte Typ in diesem Kontext – fußt hingegen auf (gewöhnlicherweise bereits vorliegenden und ausgewerteten) Daten, die verschiedenen Themen zugeordnet in unterschiedlichen Datenbanken vorgehalten werden. Obwohl, wie Wallgren und Wallgren (2007) gleich im einleitenden Kapitel anmerken, eine solche datenquellenübergreifende Analyse eine sehr alte und häufig benutzte Technik ist, sind weder registerbasierte Statistik noch registerstatistische Methoden theoretisch gut fundiert; ad-hoc-Methoden dominieren im Umgang mit diesem Thema.

Nach Wallgren und Wallgren sind (statistische) Register das Ergebnis einer strukturellen und inhaltlichen Zusammenführung von Daten aus unterschiedlichen Primärquellen – nicht nur – der Amtlichen Statistik mit dem Ziel der gemeinsamen Analyse jenseits der spezifischen Primärziele der einzelnen Datenbanken. Subjekte (Personen, Betriebe, Landkreise) der einzelnen Primärdatenbanken sind im

(statistischen) Register weiterhin identifizierbar, aggregierte Zusammenstellungen aus mehreren Primärdatenbanken ergeben kein Register.

Dem Umgang mit Registern und registerbasierter Statistik versuchen Wallgren und Wallgren (2007) in ihrem wegweisenden Buch eine solide Grundlage zu geben – sowohl theoretisch als auch praktisch-anwendungsorientiert. Sie widmen sich den Fragen der Strukturierung eines Registers, den Problemen der logischen und physischen Datenzusammenführung, adäquaten Schätzmethode, dem Umgang mit fehlenden Informationen und – dem Charakter amtlicher Daten entsprechend – dem Schutz der Vertraulichkeit der Daten. Ausführlich widmen sich Wallgren und Wallgren (S. 68 ff.) den Phasen des Zusammenstellens eines neuen Registers, von den wissenschaftlichen Zielen, der Analyse des Vorhandenen, bis hin zu den nötigen Variablen-Übernahmen und -Veränderungen und identifizieren dabei drei Prozesse, die parallel laufen sollten: (a) Schaffung und Nutzung des Registers, (b) Qualitätssicherung und (c) Dokumentation. Den besonderen Anforderungen beim Generieren eines longitudinalen Register (mit Angaben zu mehreren aufeinanderfolgenden Erhebungszeitpunkten, z. B. Jahre) widmet sich ein Teilkapitel (Wallgren und Wallgren, Kap. 6.4), ohne auf die spezifischen Schwierigkeiten der eindeutigen Identifikation von Registerobjekten über den Betrachtungszeitraum gesondert einzugehen. Interessanterweise sucht man eine Aussage zur statistischen Grundlage für Modellierungen im Falle einer Totalerhebung (Abschn. 3.4.3) in Wallgren und Wallgren (2007) vergeblich.

Für vorliegende Arbeit gilt es, aus den unterschiedlichen Quellen amtlicher Daten (Krankenhausstatistik der Länder, der VGR und der ETR aus den Arbeitskreisen der StLÄ und des StBA, der Arbeitslosen-Statistik der Bundesagentur für Arbeit sowie der siedlungsstrukturellen Kreistypen des Bundesamt für Bauwesen und Raumordnung) ein Register in oben genanntem Sinne zu definieren, datentechnisch und datenflussmäßig zusammenzustellen und auszuwerten. Dieses Register soll, als Kombination jener totalerhebungsbasierten Statistiken die Jahre 1991–2005 und die Bundesländer Berlin, Brandenburg, Mecklenburg-Vorpommern, Rheinland-Pfalz, Sachsen, Sachsen-Anhalt sowie Thüringen umfassen. Ähnliches ist bisher (Stand September 2010) mit dieser Datenbasis nicht publiziert.

3.6. Analysen amtlicher Daten zum GSW: Vergleichbare und themenverwandte Arbeiten

Die Krankenhausstatistik gibt es seit Beginn der 1990-er Jahre, die aktuellen Formen des Zugriffs – SUF, GWAP und KDFV – auf amtliche Daten durch die Wissenschaft bewähren sich. Es bleibt die Frage nach ersten veröffentlichten Ergebnissen.

3.6.1. Analysen der Krankenhausstatistik

Das StBA veröffentlicht jeweils etwa 15 Monate nach Ende des Berichtsjahres die jährlichen Ergebnisse der Krankenhausstatistik in der Fachserie 12 getrennt nach Grunddaten (Reihe 6.1)⁵⁶, Diagnosedaten (Reihe 6.2)⁵⁷ und Kostendaten (Reihe 6.3)⁵⁸. Diese Publikationen enthalten Erläuterungen zu den Teilen der Krankenhausstatistik, den Merkmalen und den Daten sowie im Hauptteil Tabellen mit Auszählungen. Werte vorangegangener Jahre sind teilweise hinzugefügt, Interpretationen gibt es keine.

Basierend auf den Daten der Krankenhausstatistik werden in den Monatsschriften der StLÄ Qualitätsbeurteilungskriterien für Statistiken allgemein (Bömermann und Hauswirth, 2002) erläutert und (in)adäquate statistische Maßzahlen (Bömermann, 2003) erklärt. Es wird ausgezählt und tabellarisch zusammengefasst, Kenngrößen wie die bereinigten Kosten je aufgestelltem Krankenhausbett, je Pflage-tag und je vollstationärem Fall werden berechnet und präsentiert (z. B. Reh, 2001). Kleinsorge (1998) geht darüber hinaus, indem sie Vergleiche zieht zwischen den Häufigkeiten bestimmter Hauptdiagnosen in Sachsen mit denen im gesamten Bundesgebiet. Diese Analysen werden verfeinert, indem der Bezug zur jeweiligen Altersverteilung und zum Geschlecht der Patienten hergestellt wird. All diese Analysen erfolgen beschreibend, Modellierungen der Daten werden nicht angestrengt. Domanja et al. (2007) widmen sich gesundheitlichen Störungen durch Alkohol, vergleichen verschiedene Regionen und belegen sowohl eine steigen-

⁵⁶ seit 2004 getrennt für Krankenhäuser sowie Rehabilitations- und Vorsorgeeinrichtungen.

⁵⁷ seit 2000 getrennt für Krankenhäuser sowie Rehabilitations- und Vorsorgeeinrichtungen.

⁵⁸ Reihe 6.4 publiziert seit 2005 Ergebnisse der DRG-Statistik.

de Tendenz bei der Anzahl der Behandlungsfälle als auch die wegen dieser Erkrankungen entstehenden hohen mittleren Kosten, zu bezahlen durch die Gesellschaft. Richter (2008) stellt die Wahrnehmung des Gesundheitswesens als Kostenfaktor an den Beginn seiner Überlegungen und plädiert für die Entwicklung einer *gesundheitsökonomischen Gesamtrechnung*. Diese Arbeit, basierend u. a. auf den Kosten und den Diagnosedaten der Krankenhausstatistik, kann – wie die folgende – als Grundlage politischer Entscheidungen herangezogen werden. Ausgehend von Diagnosedaten, Einwohnerzahlen und Altersverteilung der Bevölkerung wird von Knödler (2005) eine Status-Quo-Voraussrechnung zur Abschätzung des zukünftigen Versorgungsbedarfs angestrengt.

Es fällt auf, dass die Daten der Krankenhausstatistik im Bereich der StLÄ und des StBA regelmäßig im Querschnitt der einzelnen Jahre aggregiert werden, Entwicklungen über die Zeit werden mit den aggregierten Daten betrachtet, Modellierungen oder Paneldaten-Analysen finden sich kaum.

Auch in Publikationen außerhalb der StLÄ werden die Daten der Krankenhausstatistik erst einmal im Querschnitt analysiert, wie in den Überblicksarbeiten zum Datenstand von Rolland (2005b) und Schulte (2005) – die Autoren selbst sind Mitarbeiter des StBA.

Analysen unabhängiger Wissenschaftler

Gemessen am Datenreichtum gibt es Analysen der Daten der Krankenhausstatistik durch von der Amtlichen Statistik unabhängigen Wissenschaftler bisher nur wenige.

In der bereits erwähnten Arbeit von Röhmel (2009) wird ein Panel von 100 ostdeutschen Krankenhäusern bezüglich der Effizienz der Krankenhäuser betrachtet. Das untersuchte Panel überdeckt die Dekade 1994–2003 und nutzt für Analysen mittels Data Envelopment Analysis vor allem Merkmale aus den Grunddaten: Träger des Krankenhauses, Anzahl aufgestellter Betten, Anzahl an Ärzten, Pflegepersonal und sonstigem Personal⁵⁹ und Sachkosten als Input-Indikatoren sowie als

⁵⁹ jeweils als Vollkräfte im Jahresdurchschnitt bestimmt.

Output-Indikatoren die Anzahl der Pflēgetage, der vollstationären Fälle und der Stundenfälle.

Herr (2009) und Folgepublikationen (Herr, 2008; Herr et al., 2011) nutzt Daten der Krankenhausstatistik für ihre Untersuchungen zur Konkurrenz zwischen Krankenhäusern und der Effizienz der Krankenhäuser in Abhängigkeit von der Trägerschaft⁶⁰. Für das erste Teilthema, die Betrachtung der Kosteneffizienz, wird auf die Daten der Jahre 2000 bis 2003 zurückgegriffen. Von den zwischen 2242 und 2197 (2000 bzw. 2003, jeweils zum Jahresende) deutschen Krankenhäusern (Statistisches Bundesamt, 2006e, S. 229) sollen *1556 to 1635 general hospitals* in den erwähnten Jahren herangezogen werden, $\frac{1}{4}$ der Krankenhäuser wird somit nicht betrachtet⁶¹. Herr berichtet von erheblichen Problemen mit unvollständigen Daten, die dazu führen, dass 11 % der Beobachtungen ausgeschlossen werden – der Anteil erscheint hoch für Daten der Amtlichen Statistik. Die Daten werden mit Methoden der Paneldaten-Analyse ausgewertet. Wegen der Kürze des Panels wird jedoch in der Interpretation den Ergebnissen der Querschnittsanalysen mehr Bedeutung beigemessen. Für die Analyse der Profiteffizienz der Krankenhäuser, dem zweiten Teilthema, nutzt Herr (2009) Daten der Krankenhausstatistik der Jahre 2002 bis 2005 und ergänzt diese um (externe) Daten aus dem Rheinisch-Westfälischen Institut für Wirtschaftsforschung (RWI) in Essen. Damit gelingt Herr eine quellenübergreifende Datenanalyse mit Daten der Amtlichen Statistik.

3.6.2. Quellenübergreifende Analysen

Wie Herr (2009) greift auch Schwierz (2009) auf die Daten der Krankenhausstatistik zurück – speziell auf die Jahre 1996 und 2006. Schwierz schließt dabei die rein psychiatrischen Kliniken, Bundeswehrkrankenhäuser und reine Tages-/Nachtkliniken aus⁶². Unter KDFV-Bedingungen können die Gemeinde des Krankenhauses genauso wie die Wohngemeinde der Patienten in die Analysen eingehen. Diese Daten werden mit Daten der Einwohnerstatistik auf Ge-

⁶⁰ Ob der Zugriff auf die Daten unter GWAP und/oder KDFV erfolgt, wird nicht erwähnt.

⁶¹ Für 2003 gibt das StBA 1868 allgemeine Krankenhäuser an (Statistisches Bundesamt, 2005, Tab. 2.1.1).

⁶² somit werden die allgemeinen Krankenhäuser betrachtet.

meindeebene verbunden, die Einwohnerzahl wird als Näherungsvariable für den Bedarf an Krankenhausleistung genutzt.

Für alle Vorgenannten (Röhmel bis Schwierz) ist das Krankenhaus Analyseeinheit.

Mertens und Haas (2006) hingegen verbinden Personendaten von 3000 Personen aus der Deutschen Lebensverlaufsstudie – einer *Reihe einmaliger Retrospektivbefragungen von Personen ausgewählter Geburtsjahrgänge* – mit amtlichen Daten der Bundesagentur für Arbeit zur Arbeitslosenquoten auf Kreisebene und untersuchen den Einfluss regionaler Arbeitslosenquoten auf Arbeitsplatz- und Betriebswechsel. Hier sind die Kreise Deutschlands die Analyseeinheiten.

3.6.3. Andere Aspekte der Transition Ostdeutschlands

An auf amtlichen Daten basierenden empirischen Forschungsarbeiten zu anderen Aspekten der Transition Ostdeutschlands seien hier nur stellvertretend Strohe (1999) zur wirtschaftlichen Entwicklung Brandenburgs, Rambert (2004) mit einer Untersuchung zur Beschäftigungsentwicklung und Mangelsdorf (2007, 2009) mit Analysen zu den Veränderungen im verarbeitenden Gewerbe genannt. Beginnend mit Strohe (2000) beruhen diese Arbeiten zunehmend auf Mikrodaten der Amtlichen Statistik.

3.7. Stand der Forschung – Zusammenfassung

Das Ende der DDR und die Transition zu einer demokratischen Gesellschaft gehen einher mit dem Beitritt (unter Einbeziehung West-Berlins) in die bundesrepublikanische Wirklichkeit mit ihren gefestigten und bewährten Strukturen. Das unterscheidet die Transition Ostdeutschlands von der aller anderer Staaten des Ostblocks. Parallel zur Transition verändern sich Wirtschaft und Gesellschaft im vereinigten Deutschland der Alt- und Neubundesländer. Diese Veränderungen werden aus verschiedenen Blickwinkeln wissenschaftlich reflektiert. Mit der Bereitstellung amtlicher Daten für die Wissenschaft sind erweiterte empirische Analysen

möglich geworden, dabei kommen zunehmend moderne Methoden der Ökonometrie zur Anwendung.

Steigende Gesundheitsausgaben veranlassen die Politik, neue Regeln für das Gesundheitswesen zu formulieren und erhöhte Kostentransparenz und -effizienz einzufordern. Die Wissenschaft begleitet diesen Prozess der Veränderungen im Gesundheitssystem sowohl beratend als auch empirisch analysierend, so wie sie dieses für wirtschaftliche und soziale Veränderungen der Gesellschaft schon seit langem tut.

Empirische Analysen weitergehender Interaktionen zwischen dem Gesundheitssystem, seinen Leistungen und seinen Aufwendungen mit der allgemeinen wirtschaftlichen Lage, der Verfügbarkeit der wirtschaftlichen Leistungserbringer (abhängig Beschäftigte), Arbeitslosigkeit, Krankenstand, Genesungsdauern oder Diagnoseverteilungen lassen sich hingegen sowohl bei der Analyse der Transition Ostdeutschlands als auch der Analyse für größere Wirtschaftsräume kaum finden.

Kapitel 4.

Daten und Registererstellung

DIE QUALITÄT DER EMPIRISCHEN FORSCHUNG HÄNGT NICHT NUR VON DER
QUALITÄT DER ÖKONOMISCHEN UND STATISTISCHEN THEORIE AB, SONDERN AUCH VON DER
QUALITÄT DER ZUR VERFÜGUNG STEHENDEN DATEN.
von Auer (2007, Kap. 1.5)

Grundlage einer jeden empirischen Forschung sind Daten – erhobene und gemessene Angaben zu Phänomenen der Wirklichkeit.

Daten können unmittelbar für eine Untersuchung gewonnen werden, oder es werden bereits vorliegende Daten für den weiteren Erkenntnisgewinn herangezogen (Abschn. 3.4.1 und Kish, 1985). Im ersteren Fall können Datenstruktur und gewünschte Dateninhalte vorab in Hinsicht auf die Auswertung definiert werden. Im Fall der (Nach-)Nutzung bereits vorliegender Daten ist die Struktur bereits vorgegeben, sie muss gegebenenfalls den Zielen der Auswertung angepasst werden.

Problematisch für die Sekundärauswertung vorliegender Daten ist, dass ihre Qualität häufig nicht ausreichend dokumentiert ist (Rosenberg et al., 2006). Wallgren und Wallgren (2007, Kap. 10) weisen darauf hin, dass Datenqualität kein absoluter Begriff ist, sondern jeweils im Kontext der Ziele der ursprünglichen Datenzusammenstellung betrachtet werden sollte. Wang und Strong (1996) geben neben der Fehlerfreiheit (*accuracy*) mit der Relevanz (*relevancy*), dem Bezug der Daten zur zu beschreibenden Wirklichkeit (*representation*) und der Erreichbarkeit (*accessibility*) drei weitere Hauptkategorien für die sinnvolle (Nach-)Nutzbarkeit von Daten an.

Bedenkt man, dass die Sicherung von Datenqualität aufwändig und somit teuer ist, sollte besonders vor einer Sekundärauswertung bereits vorliegender Daten eine Qualitätseinschätzung anhand der aktuellen Auswertungsprioritäten vorgenommen werden. Im ersten Satz ihrer Arbeit *How to Lie with Bad Data* postulieren de Veaux und Hand (2005): *Schlechte Daten können jegliche Analyse ruinieren.*⁶³ Sie weisen in auf die allgemeingültige Erfahrung hin, dass der größere Teil der für die Datenauswertung eines Projektes vorgesehen Zeit mit Datenüberprüfung und Datenbereinigung zu verbringen ist, allgemein wäre dabei von 60–95 % auszugehen. Erst danach kann mit der eigentlichen Datenanalyse begonnen werden. Denn immer gilt, dass unabhängig von der Art und den ursprünglichen Zielen der Datengewinnung die Qualität der Datenauswertungen und der Dateninterpretation von der Qualität der Daten in großem Maße abhängt.

Zur Beschreibung der Transitionsphase Ostdeutschlands und zum Vergleich mit Rheinland-Pfalz als Repräsentanten für die Altbundesländer (Abschn. 2.5 und Abb. 2.5) sind Daten der Amtlichen Statistik interessant und relevant, die sich auf die Jahre 1991 bis 2005 beziehen.

4.1. Datenquellen und Datenzugang

Erhebung, Aufbewahrung, Verarbeitung und Nutzungsbereitstellung amtlicher Daten obliegt der Amtlichen Statistik, die in Deutschland größtenteils in der Verantwortung der einzelnen Bundesländer liegt. Der Rechtsrahmen für den Umgang mit amtlichen Daten wird dabei im Wesentlichen durch das Bundesstatistikgesetz (BStatG) gesetzt.

Mit dem Beginn des Jahres 1992 und der Überführung des seit 1990 bestehenden Gemeinsamen Statistischen Amtes der neuen Länder (GeStAL) in Einrichtungen der einzelnen Länder selbst gibt es 16 Statistische Landesämter (StLÄ) in bundeslandspezifischer Benennung⁶⁴. Gesamtnationale Daten, wie sie das Statistische Bundesamt publiziert, werden zum größten Teil durch Aggregation der Da-

⁶³ Übersetzung des Autors.

⁶⁴ Nach Fusionen zum Statistikamt Nord – Statistisches Amt für Hamburg und Schleswig-Holstein (01. 01. 2004) und zum Amt für Statistik Berlin-Brandenburg (01. 01. 2007) sind es aktuell 14 StLÄ.

ten der StLÄ gewonnen. Um diese Aggregation zu ermöglichen und die Vergleichbarkeit der dezentral erhobenen und gehaltenen Daten zu gewährleisten, definieren StBA und StLÄ gemeinsam Struktur und Inhalte der einzelnen Statistiken (BStatG § 3 (1) und (2) sowie § 3a (1)).

Einblick in die vorhandenen Daten gewähren regelmäßige Publikationen des StBA wie das jährlich herausgegebene Statistische Jahrbuch für die Bundesrepublik Deutschland und themenspezifische Publikationen sowie die Publikationen der StLÄ, die zunehmend online verfügbar sind⁶⁵.

Wichtiger Eingangsschritt einer Analyse amtlicher Daten ist die Klärung von Daten-Verfügbarkeit und -Zugriffsmöglichkeiten. BStatG und Bundesdatenschutzgesetz (BDSG) definieren dabei den Rahmen, wie bei der Nutzung amtlicher Mikrodaten die Rechte der Bürger und Unternehmen – der Wirtschaftssubjekte – auf Schutz der sie betreffenden Daten zu beachten sind; die Anonymität eines jeden Einzelperson ist in jedem Fall zu wahren, für Daten von Unternehmen und öffentlichen Einrichtungen gelten aggregatsniveaubezogene Einschränkungen.

4.1.1. Amtliche Daten für die Wissenschaft: Bereitstellung durch die FDZ

Laut BStatG § 3a (2) *gehört auch die Bereitstellung von Daten für die Wissenschaft zu den Aufgaben der StLÄ und des StBA, den Schutz der amtlichen Mikrodaten unter dem sogenannten *Wissenschaftsprivileg* regelt § 16 (6) BStatG.*

Im Herbst 2001 wird das Forschungsdatenzentrum (FDZ) des StBA und ein halbes Jahr später das FDZ/L der StLÄ gegründet mit dem Ziel eines für die Wissenschaft verbesserten Zugangs zu den Mikrodaten der Amtlichen Statistik. Gleichzeitig wurde bei der Deutschen Rentenversicherung Bund das FDZ-RV etabliert; alle drei FDZ beginnen ihre Tätigkeit im Rahmen eines vom Bundesministerium für Bildung und Forschung geförderten Projekts.

⁶⁵ z. B. über das gemeinsame Statistik-Portal von StLÄ und StBA <http://www.statistikportal.de/>.

Das FDZ/L ist bundesweit in den Standorten der einzelnen StLÄ angesiedelt⁶⁶, die Geschäftsstelle befindet sich in den Räumen des Geschäftsbereichs Statistik des Landesbetriebes Information und Technik Nordrhein-Westfalen (IT.NRW) in Düsseldorf. Neben dem in allen Standorten des FDZ/L vorgehaltenen Leistungsangebot entwickeln die einzelnen Standorte jeweils besondere Expertise in einigen speziellen Arbeitsschwerpunkten⁶⁷. Für den Nutzer soll trotzdem der jeweils geografisch nächste Standort des FDZ/L der bevorzugte Partner sein.

Neben dem BStatG ist das Bundesdatenschutzgesetz eine wesentliche Grundlage für die Nutzung amtlicher Daten durch die Wissenschaft. Das BDSG bestimmt in § 3 (6) das Anonymisieren als einen wesentlichen Begriff beim Schutz des Einzelnen vor Einschränkungen des Persönlichkeitsrechts bei der automatisierten Verarbeitung seiner personenbezogenen Daten (BDSG § 1 (1) und (2)). Die FDZ unterscheiden dabei absolute, faktische und formale Anonymisierung. Zühlke et al. (2003) führen dazu aus: *Absolut anonymisierte Daten werden durch Aggregation oder durch die Entfernung einzelner Merkmale so weit verändert, dass eine Identifizierung der Auskunftgebenden nach menschlichem Ermessen unmöglich gemacht wird. ... Dagegen werden Mikrodaten als faktisch anonym bezeichnet, wenn die Deanonymisierung zwar nicht gänzlich ausgeschlossen werden kann, die Angaben jedoch nur mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft dem jeweiligen Merkmalsträger zugeordnet werden können.* Formal anonymisierte Daten entstehen durch das Entfernen aller direkten Identifikatoren wie z. B. des Namens und der Adresse einer Person oder eines Wirtschaftssubjekts (*Pseudonymisieren* gemäß BDSG § 3 (6a)).

Passend zu diesen definierten Anonymisierungsgraden werden Mikrodaten der Amtlichen Statistik durch das FDZ/L in folgenden Formen bereitgestellt (Zühlke et al., 2003, 2004; Schoffer, 2005): (a) Public-Use-Files (PUF) sind absolut anonymisiert und für jedermann zugänglich. (b) Scientific-Use-Files (SUF) sind (relativ stark) faktisch anonymisiert und werden zur Nutzung außerhalb der FDZ an Wissenschaftler übergeben. (c) Daten für eine On-Site-Nutzung am Gastwissenschaftler-Arbeitsplatz (GWAP) vor Ort im FDZ sind projektbezogene, faktisch anonymisierte Daten zur Nutzung mit direkter Beratung durch die Vorort-Themenspezialisten des FDZ/L. (d) Daten für eine Kontrollierte Datenfernverarbeitung

⁶⁶ <http://www.forschungsdatenzentrum.de/kontakt-laender.asp>.

⁶⁷ z. B. Berlin-Brandenburg: Dienstleistungen, Handel und Gastgewerbe, Tourismus, Öffentliche Finanzen; Rheinland-Pfalz: Bevölkerung; Sachsen: Gesundheitswesen, Sozialleistungen.

(KDFV) sind formal anonymisiert und ermöglichen das Abarbeiten von vom Wissenschaftler erarbeiteten Software-Skripte durch das FDZ/L, die Ergebnisse werden vor der Übergabe an den Wissenschaftler auf Einhaltung der entsprechenden gesetzlichen Regelungen geprüft. Die Nutzung von SUF, eine On-Site-Nutzung am GWAP und Teilnahme an der KDFV müssen mit Angabe der Ziele des Forschungsvorhabens beantragt und vom FDZ/L (zentral in Düsseldorf) genehmigt werden.

4.1.2. Anonymisierungen, GWAP und KDFV

Ende 2005 beginnt die Arbeit an diesem Projekt mit ersten Kontakte zu den Standorten Berlin und Sachsen des FDZ/L, ersteres der geografischen Nähe, zweiteres des thematischen Schwerpunkts Gesundheitswesen wegen. Entgegen der allgemeinen FDZ-Empfehlung der geografischen Nähe wird dem fachlichen Direktkontakt mit dem Themenspezialisten der Vorzug gegeben (Schoffer, 2005, und Folgejahre)⁶⁸. Ziel ist es, die Daten der Krankenhausstatistik als Primärdaten zu nutzen und diese für die Analysen als Paneldaten aufzuarbeiten. Diese Paneldaten sollen dann mit Wirtschaftsdaten – gleichfalls im Panelformat – in Bezug gesetzt werden.

Da die Bundesländer (über ihre StLÄ) Eigner der jeweiligen Daten ihres Bundeslandes bleiben, erfordert ein wissenschaftliches Projekt mit dem FDZ/L und Daten aus mehreren Bundesländern die Zustimmung der Themen-Referenten aller betroffenen Länder; mit jeder weiteren gewünschten Statistik erhöht sich die Anzahl der Zustimmungspflichtigen nochmals. Alternativ lässt sich der Antrag stellen, externe Daten zu den Daten aus dem FDZ/L hinzu zu spielen. Diese Daten werden Teil der Projektabsprachen und als solche vom FDZ/L auf Regelkonformität geprüft, der Anonymisierungsgrad der Projekt-Daten muss auch mit diesen Ergänzungen erhalten bleiben.

Projektbezogene faktische Datenanonymität am GWAP wird durch Regeln gewährleistet, deren wichtigste für die Datenübergabe an den Wissenschaftler sind für vorliegende Analysen:

⁶⁸ Vertragsbeginn ist der 01. 06. 2007 (Forschungsdatenzentrum der Statistischen Landesämter, 2007) mit Datenzugriff am GWAP, plangemäß wird der Vertrag später um die Nutzung unter KDFV erweitert.

- (a) Das FDZ/L stellt eine Zufallsstichprobe von ca. 80 % der Krankenhäuser bereit.
- (b) Jedes Krankenhaus ist nur durch eine kontextfreie Nummer identifiziert.
- (c) Um die Identifizierung einzelner Krankenhäuser ausschließen zu können, darf die geografisch-administrative Zuordnung eines Krankenhauses weder auf Gemeinde- noch auf Kreisebene erfolgen; kleinste Einheiten der Zuordnung müssen jeweils mindestens 3 Kreise⁶⁹ umfassen (Dreier-Regel).
- (d) Für Wohnorte der Patienten (Fälle) gilt dieselbe Dreier-Regel.

Welche Kreise zusammengefasst und wie diese Zusammenfassungen identifiziert werden sollen, obliegt dem Wissenschaftler und kann von ihm im Sinne seiner Analyseziele beeinflusst werden.

Für die Nutzung unter KDFV-Bedingungen werden die FDZ/L-Daten formal anonymisiert bereitgestellt. Ein Ausschluss von Krankenhäusern erfolgt nicht, die Krankenhäuser sind wieder durch eine kontextfreie Nummer identifiziert, wiederum ist es möglich, geprüfte externe Daten hinzuspielen zu lassen. Analysen können mit weder durch Selektion noch durch andere datenverändernde Anonymisierungsmaßnahmen modifizierten Daten durchgeführt werden. Erst vor Übergabe an den Wissenschaftler ist jedes durch KDFV erzielte Analyseergebnis vom Fachpersonal des FDZ/L auf Einhaltung der Anonymisierungsanforderungen zu prüfen. Es ist somit für den Wissenschaftler und FDZ/L effizient, von Beginn an anforderungskonforme Ausgabe-Schritte in den Software-Skripten festzulegen.

Implikationen dieser Forderungen für vorliegende Analysen und die Umsetzung während der Datenvorbereitung für GWAP und KDFV werden im weiteren Verlauf (Abschn. 4.4: Erstellung des Registers) beschrieben.

4.1.3. Daten aus dem FDZ/L – Krankenhausstatistik

Die Krankenhausstatistik⁷⁰ beschreibt die strukturelle Situation der Krankenhäuser und die Nutzung dieser durch Patienten. Sie ist in drei Erhebungsteile un-

⁶⁹ Einzige Ausnahme hier: Berlin als Bundesland ohne Kreis- oder Gemeinde-Substruktur und mit einer gegen unbeabsichtigte Re-Identifikationen ausreichend großen Anzahl an Krankenhäusern muss nicht dieser Regel unterworfen werden.

⁷⁰ <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/abisz/Krankenhausstatistik,templated=renderPrint.psml>, Zugriff am 04. 04. 2011.

tergliedert: Grunddaten und Kostendaten werden seit 1991⁷¹ und Diagnosedaten seit 1993 erhoben. Rechtsgrundlage ist dabei die Krankenhausstatistik-Verordnung vom 10. 04. 1990 in Verbindung mit dem Bundesstatistikgesetz vom 22. 01. 1987 in jeweils aktueller Änderung.

Die Krankenhausstatistik als Bundesstatistik ist eine jährliche Totalerhebung mit Auskunftspflicht für alle Krankenhäuser in Deutschland. Die 1991 novellierte Fassung des Krankenhausfinanzierungsgesetzes (KHG) zählt die zu erfassenden Sachverhalte auf und regelt die Auskunftspflicht der Krankenhausträger gegenüber den Statistischen Landesämtern (Bömermann und Hauswirth, 2002, 1.1). In den Publikationen der Fachserie 12, Reihe 6.1.1 (Statistisches Bundesamt, 2006–2010) werden im Abschnitt 1.6 *Erhebungsgesamtheit, Zuordnungsprinzip der Erhebungsgesamtheiten* Krankenhäuser einschließlich deren Ausbildungsstätten unter Hinweis auf die KHStatV und das Sozialgesetzbuch V definiert; auf den umfassenden Krankenhausbegriff des Krankenhausfinanzierungsgesetzes wird Bezug genommen. Im Wesentlichen handelt es sich bei Krankenhäusern um Einrichtungen, die der Krankenbehandlung einschließlich Geburtshilfe dienen, fachlich-medizinisch unter ständiger ärztlicher Leitung stehen, dabei über ausreichende diagnostische und therapeutische Möglichkeiten verfügen und nach wissenschaftlich anerkannten Methoden arbeiten. Patienten müssen in diesen Einrichtungen untergebracht und gepflegt und mit Hilfe von jederzeit verfügbarem ärztlichem, Pflege-, Funktions- und medizinisch-technischem Personal versorgt werden können (nach Statistisches Bundesamt, 1993–2005).

Eingeschlossen in die Krankenhausstatistik sind von Beginn an ebenfalls die Rehabilitations- und Vorsorgeeinrichtungen. Sie haben jedoch Auskunftspflicht nur zu Grunddaten bei einem zusätzlich eingeschränkten Merkmalskatalog. Seit einer Änderung der KHStatV im Jahr 2001 haben Rehabilitations- und Vorsorgeeinrichtungen mit mehr als 100 Betten beginnend mit dem Jahr 2003 auch Auskunftspflicht bezüglich der Diagnosedaten, kleinere Häuser sind davon ausgenommen (Statistisches Bundesamt, 2005b).

Ausdrücklich ausgenommen aus der Krankenhausstatistik sind Krankenhäuser im Straf- oder Maßregelvollzug, Polizeikrankenhäuser sowie reine Tages- oder Nachtkliniken ohne vollstationäre Betten. Bundeswehrkrankenhäuser sind nur mit ih-

⁷¹ Für die Altbundesländer sind Grunddaten (ohne Personal) und Kostennachweis bereits für das Berichtsjahr 1990 zu erheben.

ren Leistungen für Zivilpatienten einbezogen, Grund- und Diagnosedaten sind zu berichten.

Für die Krankenhausstatistik als dezentraler Statistik obliegen Erfassung, Pflege und Weiterverarbeitung der Daten den einzelnen Bundesländern. Die Daten werden dann von den StLÄ sowohl dem FDZ/L als auch dem StBA zur Verfügung gestellt. Jährlich publiziert das StBA in der Reihe 6 der Fachserie 12 Ergebnisse der Krankenhausstatistik. Das FDZ/L stellt diese Daten für die Wissenschaft zur Nutzung am GWAP und unter KDFV bereit⁷².

Die Krankenhausstatistik – eingeschränkt auf die Krankenhäuser der Jahre 1991 bis 2005 – mit ihren Informationen zu diagnostizierten und behandelten Krankheiten, zu dazu nötigem Aufwand, verursachten Kosten sowie zum vorgehaltenen Personal mit seinen Leistungen bilden den Primärdatenpool vorliegender Analysen⁷³.

4.1.4. Weitere Daten aus dem Bereich der Amtlichen Statistik

Ergänzt werden sollen die Daten der Krankenhausstatistik um amtliche Wirtschafts- und Sozialdaten aus verschiedenen Quellen mit dem Ziel, die spezielle Aspekte um Gesundheit (Krankheit) und wirtschaftlicher Situation während der Transition Ostdeutschlands zu beschreiben.

FDZ-RV

Daten der Deutschen Rentenversicherung Bund werden über das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) bereitgestellt, die Regeln des Zugriffs entsprechen im Wesentlichen denen der FDZ/L. Mittels der Daten der Rentenversicherung kann die Einkommensposition und somit der (einkommensabhängige) Wohlstand beschrieben werden (Abschn. 3.3.3). Allerdings wird dabei für jeden erfassten Bürger nur der aktuelle Wohnort in den Daten vorgehalten, ei-

⁷² <http://www.forschungsdatenzentrum.de/bestand/krankenhaus/index.asp>, Zugriff am 03.04.2011; Diagnosedaten von 2003 sind auch als SUF verfügbar.

⁷³ Zu weiteren Einschränkungen siehe Folgeabschnitte.

ne rückwirkende Zuordnung, wo die Entgeltpunkte früherer Jahre erworben wurden, ist mit diesen Daten nicht möglich (Himmelreicher, 2006).

Deshalb wird von einer Einbeziehung der Daten der Deutschen Rentenversicherung Bund Abstand genommen.

Bundesagentur für Arbeit – Internetangebot

Arbeitslosenzahlen und -quoten stehen im offenen Zugriff über den Internetauftritt der Bundesagentur für Arbeit zur Verfügung. Die *Zeitreihe Arbeitslose und Quoten nach Kreisen ab 1985*⁷⁴ bietet für jeden Kreis für die Jahre 1985 bis 2004 die entsprechenden Angaben in monatlicher Genauigkeit, für die Zeiträume vor Januar 1987 und Juli 1993 bis Dezember 1997 gibt es diese Angaben nur als quartalsweise Daten – unter den Spaltenüberschriften März, Juni, September und Dezember. Daten für Ostdeutschland liegen erst mit dem Beginn des Jahres 1996 vor, die Daten der *Zeitreihe Arbeitslose* für Berlin beziehen sich bis Ende 1995 augenscheinlich nur auf das vormalige West-Berlin. Die Daten für 2005 finden sich in ähnlicher Form⁷⁵.

Arbeitslose werden am jeweiligen Wohnort registriert, Arbeitslosenzahlen und -quoten folgen dem Wohnortkonzept und somit dem Inländerkonzept (Abschn. 3.1.5).

Bundesamt für Bauwesen und Raumordnung

Von den vier siedlungsstrukturellen Gebietstypen des BBR (Abschn. 3.3.5) sind für dieses Projekt der Regionsgrundtyp mit den Ausprägungen Agglomerationsraum, Verstädterter Raum und Ländlicher Raum und der spezifischere Kreistyp mit 9 verschiedenen Ausprägungen interessant. Die Zuordnung dieser siedlungsstrukturellen Gebietstypen zu den Kreisen des Gebietsstandes 2005 sind vom BBR

⁷⁴ <http://www.pub.arbeitsamt.de/hst/services/statistik/detail/q.html> » Zeitreihe Arbeitslose und Quoten nach Kreisen ab 1985 » zr_alo_kreise_1984d.zip, Zugang am 08. 05. 2007.

⁷⁵ <http://www.pub.arbeitsamt.de/hst/services/statistik/detail/q.html> » Arbeitslose nach Kreisen (monatlich, Auszug aus dem Kreisreport), Zugang am 13. 11. 2008.

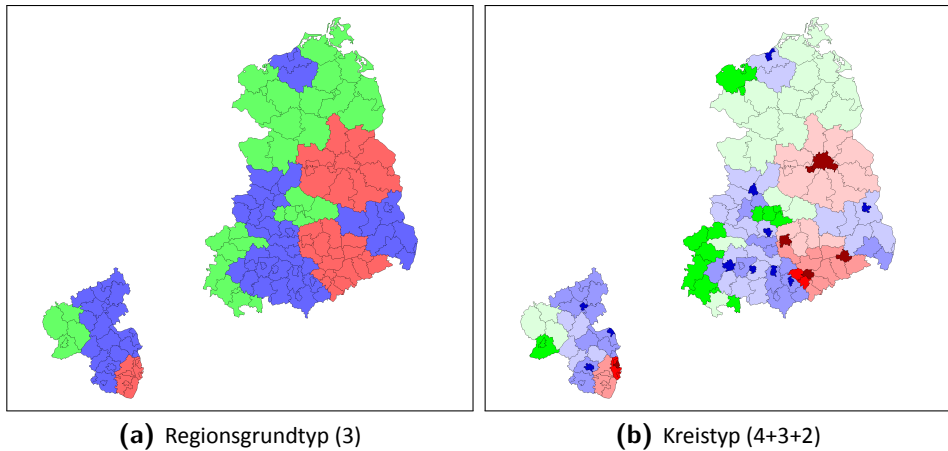


Abb. 4.1.: Siedlungsstrukturelle Gebietstypen – BBR-Klassifikation der Kreise (2005): Agglomerations- (*rot*), Verstädterte (*blau*) und Ländliche Räume (*grün*), (hellere Farbtöne [4 rot, 3 blau, 2 grün] für jeweils geringeren Verdichtungsgrad)

in tabellarischer Form zur Verfügung gestellt worden⁷⁶. Ergänzend enthält die Tabelle Angaben zur Einwohnerzahl sowie der Einwohnerdichte in den Jahren 1992 und 2004 und somit implizit zur Fläche der Kreise in km².

Abb. 4.1 zeigt die Zuordnung der Kreise des Untersuchungsgebietes zum Gebietsstand 2005. Während Ostdeutschland alle 9 Kreistypen aufweist, gibt es in Rheinland-Pfalz keinen Ländlichen Kreis eines Agglomerationsraumes.

Arbeitskreise VGR und ETR – Internetangebot

Daten der Arbeitskreise Volkswirtschaftliche Gesamtrechnung und Erwerbstätigenrechnung stehen im Internet im freien Zugriff bereit.

Vom Angebot des Arbeitskreises Volkswirtschaftliche Gesamtrechnung⁷⁷ sollen die Daten der Kreise Verwendung finden. Diese werden jährlich als *Reihe 2* in 3 Bänden veröffentlicht: Band 1 *Bruttoinlandsprodukt, Bruttowertschöpfung in*

⁷⁶ als Excel-Datei, email an den Autor, 08. 01. 2007.

⁷⁷ http://www.vgrdl.de/Arbeitskreis_VGR/, Zugriff am 16. 04. 2011.

den kreisfreien Städten und Landkreisen Deutschlands; Band 2 Arbeitnehmerentgelt in den kreisfreien Städten und Landkreisen Deutschlands; Band 3 Einkommen der privaten Haushalte in den kreisfreien Städten und Landkreisen Deutschlands. Alle Daten sind (auch rückwirkend) in der Version der Revision der Volkswirtschaftlichen Gesamtrechnungen 2005 aufbereitet.

Die Daten liegen ab dem Jahr 1996 für jedes Jahr und alle Merkmale vor, für die Jahre davor jedoch nur für ausgesuchte. Soweit hier von Interesse sind das: (a) das Brutto-Inlandsprodukt zu Marktpreisen (in jeweiligen Preisen) – 1992, ab 1994; (b) Arbeitnehmerentgelt / Bruttolöhne und -gehälter (Inlandskonzept) insgesamt – ab 1996; (c) Primär- und Verfügbares Einkommen der privaten Haushalte (einschließlich privater Organisationen ohne Erwerbszweck) – ab 1995; (d) Erwerbstätige / Arbeitnehmer (Inlandskonzept) insgesamt – 1992, ab 1994; und (e) Einwohnerzahlen – 1992, ab 1994. Die Angaben zu Erwerbstätigen / Arbeitnehmern sind dabei Übernahmen der Ergebnisse des Arbeitskreises Erwerbstätigenrechnung; die Einwohnerzahlen werden mittels Fortschreibung des Bevölkerungsstandes im Statistischen Bundesamt ermittelt und als solche den Daten der Volkswirtschaftlichen Gesamtrechnung hinzugefügt.

Brutto-Inlandsprodukt, Arbeitnehmerentgelt / Bruttolöhne und -gehälter sowie Erwerbstätige / Arbeitnehmer werden nach dem Inlandskonzept, Primär- und Verfügbares Einkommen der privaten Haushalte nach dem Inländerkonzept (Abschn. 3.1.5) ermittelt.

Die Daten des Arbeitskreises Erwerbstätigenrechnung selbst werden im Internet jährlich aktualisiert als Excel-Dateien bereitgestellt⁷⁸. Unter *Reihe 2: Kreisergebnisse* findet sich der Zugang zu Band 1 *Erwerbstätige in den kreisfreien Städten und Landkreisen der Bundesrepublik Deutschland 1991 bis [...]*. Für ein aktuell interessierendes Jahr sind etwa ab Mai des übernächsten Jahres die Daten verfügbar. Interessierende Merkmale für vorliegendes Projekt sind sowohl die jährlichen Anzahlen der Erwerbstätigen als auch die der Arbeitnehmer in den Kreisen des Untersuchungsgebietes; diese Daten liegen für ausnahmslos jedes Jahr seit 1991 vor.

⁷⁸ <http://aketr.de/index.php/veroeffentlichungen.html>, Zugriff am 24. 04. 2011.

Da die Daten beider Arbeitskreise auch rückwirkend dem jeweils gültigen Gebietsstand angepasst werden, erscheinen die Tabellen des aktuellen Jahres nicht unbedingt als Fortschreibung derer zum Vorjahresstand.

Im Juli 2007 tritt in Sachsen-Anhalt erneut eine Kreisreform in Kraft, aus 3 kreisfreien Städten und 21 Landkreisen werden 3 kreisfreie Städte (veränderten Zuschnitts) und nur noch 11 Landkreise.

Beide Arbeitskreise – VGR und ETR – haben hingegen bereits in ihrer Berichterstattung für die Jahre 1991 bis 2006 für Sachsen-Anhalt den Gebietsstand Juli 2007 zugrunde gelegt. Deshalb beruhen vorliegende Auswertungen auf ETR_R2B1_2006_j.xls⁷⁹, sowie R2B1.xls, R2B2.xls und R2B3.xls⁸⁰, ergänzt um die freundlicherweise vom Statistischen Landesamt Sachsen-Anhalt bereitgestellten Daten für die Kreise im Zuschnitt vor jener Kreisreform.

Statistische Jahrbücher, Fachserie 12 des StBA

Neben den Einzeldaten der erwähnten Statistiken und Aggregationen derselben werden aggregierte Angaben aus den Statistischen Jahrbüchern für die Bundesrepublik Deutschland (z. B. Statistisches Bundesamt, 2006e), aus den Statistischen Jahrbüchern und Berichten Berlins, Hamburgs und Bremens (Statistisches Landesamt Berlin, 2002; Statistikamt Nord, 2009; Statistisches Landesamt Bremen, 2009) genauso zum Vergleich herangezogen wie die Publikationen der Fachserie 12 *Gesundheitswesen* des StBA (speziell Statistisches Bundesamt (1993–2005) und Statistisches Bundesamt (2006–2010)).

4.1.5. Primär- und Sekundärdaten

Unter dem Titel *Secondary Data Analyses* erklären Rosenberg et al. (2006) die Unterschiede zwischen Primär- und Sekundärdaten in Abhängigkeit ihrer Stel-

⁷⁹ <http://www.statistik-hessen.de/erwerbstaetigenrechnung/Veroeffentlichungen.htm>, Zugriff am 13. 11. 2008.

⁸⁰ http://www.statistik-bw.de/Arbeitskreis_VGR/ergebnisse.asp?lang=de-DE, Zugriff am 13. 11. 2008.

lung zur den aktuellen Analysen. Primärdaten sind die Daten, die für genau diese Analyse prospektiv erhoben wurden, die in all ihren Eigenschaften wie Umfang, Struktur, Aktualität, Genauigkeit der Erfassung sowie räumliche und zeitliche Vergleichbarkeit durch die Initiatoren bestimmt sind.

Dieselben Daten werden zu Sekundärdaten, wenn sie erneut und mit anderen als den ursprünglichen Fragestellungen analysiert werden, Kombinationen von Daten verschiedener Primärquellen und speziell Register (Abschn. 3.5) gehören ebenfalls in diese Kategorie.

Für vorliegendes Projekt handelt es sich somit ausschließlich um Sekundärdaten und Sekundärdatenanalysen.

4.2. Krankenhausstatistik – Inhalte und Beschreibung

Die Daten der Arbeitskreise Volkswirtschaftliche Gesamtrechnung und Erwerbstätigenrechnung, der Bundesagentur für Arbeit sowie aus dem Bundesamt für Bauwesen und Raumordnung sind im vorhergehenden Abschnitt hinsichtlich ihres (hier zu nutzenden) Inhalts und ihrer zeitlichen Verfügbarkeit beschrieben. Die Krankenhausstatistik (Abschn. 3.3.2 und 4.1.3) mit ihren Teilen, verschiedenen Erfassungsniveaus (Krankenhaus, Fall) ist wesentlich umfangreicher als jede der vorgenannten und vergleichsweise komplex.

4.2.1. Beschreibung der Krankenhausstatistik

KHG § 28 gibt als Sachverhalte der Krankenhausstatistik an: (a) Art des Krankenhauses und der Trägerschaft; (b) im Krankenhaus tätige Personen nach Geschlecht, Beschäftigungsverhältnis, Tätigkeitsbereich, Dienststellung, Aus- und Weiterbildung; (c) sachliche Ausstattung und organisatorische Einheiten des Krankenhauses; (d) Kosten nach Kostenarten; (e) in Anspruch genommene stationäre und ambulante Leistungen; (f) Patienten nach Alter, Geschlecht, Wohnort, Erkrankungen nach Hauptdiagnosen und (g) Ausbildungsstätten am Kranken-

haus. Die Erhebungs- und Hilfsmerkmale, die Periodizität, der Berichtszeitpunkt und Regeln der Datenweitergabe sind in der KHStatV §§ 3 bis 6 festgelegt.

Die Krankenhausstatistik ist eine jährliche Statistik mit dem Berichtszeitraum 01. 01. bis 31. 12. und dem Erhebungstichtag 31. 12. Offiziell handelt es sich bei der Krankenhausstatistik um eine Statistik in drei Teilen:

- (a) den Grunddaten – Angaben zur sachlichen und personellen Ausstattung sowie zur Patientenbewegung in den Krankenhäusern und ihren organisatorischen Einheiten,
- (b) den Kostendaten – Angaben zu Sach- und Personalkosten sowie Zinsen und Steuern der Krankenhäuser, Kosten der Ausbildungsstätten an Krankenhäusern, Abzüge für nicht-stationäre Leistungen, und
- (c) den Diagnosedaten – Angaben zu Erkrankungsart, Verweildauer, Fachabteilung sowie soziodemographische Merkmale (z. B. Alter, Geschlecht, Wohnort) für jeden Fall, d. h. für jeden Behandlungsaufenthalt eines Patienten.

Die Unterschiede zwischen den Teilen sind so groß, dass die Teile häufig in getrennten Dokumenten betrachtet werden (z. B. die Qualitätsberichte des StBA: Statistisches Bundesamt, 2008c,d,b). Meldetermin für Grund- und Diagnosedaten ist der 01. 04., für die Kostendaten der 30. 06. des dem Berichtsjahr folgenden Jahres. Kostendaten beziehen sich auf das Geschäftsjahr eines Krankenhauses, der Erhebungstichtag bezieht sich auf den letzten Tag des Geschäftsjahres und kann somit vom 31. 12. abweichen.

Die Diagnosedaten (auch Diagnosestatistik, z. B. Bömermann und Hauswirth, 2002) sind eine Fallstatistik: Merkmalsträger ist der Patient als Behandlungsfall, jeder weitere Behandlungsaufenthalt ein und desselben Patienten stellt einen weiteren Fall dar, eine Patientenidentifikation erfolgt in der Statistik nicht.

Erfasst und gehalten werden die Daten der Krankenhausstatistik in 7 logischen Strukturen (genannt Satzart mit Nummer): (1) Grunddaten des Krankenhauses – 102 Merkmale, (2) Bettenausstattung und Patientenbewegung in Fachabteilungen – 16 Merkmale, (3) Ärzte nach Spezialisierungsrichtungen (Gebiets- / Teilgebietsbezeichnung) – 14 Merkmale, (4) nichtärztliches Personal nach Personalgruppe / Berufsbezeichnung – 6 Merkmale, (5) Leistungen des Krankenhauses bei teil-, vor- und nachstationären Behandlungen nach Fachabteilungen – 5 Merkma-

le, (6) Kostendaten des Krankenhauses – 47 Merkmale, und (7) Diagnosedaten als Fallstatistik (ohne Satzart-Nummer) – 12 Merkmale. Hinzu kommen die Merkmale zur eindeutigen Identifikation der jeweiligen Dateneinträge (Krankenhaus, Fachabteilung, ..., Fall). Für die 202 Erfassungsmerkmale ist festgelegt, welche Angaben sich auf den Stichtag 31. Dezember, auf das abgelaufene Kalender- bzw. auf das abgelaufene Geschäftsjahr beziehen.

Auf den ersten Blick finden sich manche der Angaben mehrfach in den Daten. So gibt es in Satzart 1 die Angabe der Betten insgesamt, in Satzart 2 wird die Anzahl der aufgestellten Betten je Fachabteilung erhoben, wobei auch für das gesamte Krankenhaus (als Fachabteilung ‚insgesamt‘) ein Eintrag zu erstellen ist.

Einen erweiterten Überblick über die zu erhebenden Merkmale geben die Veröffentlichungen des Statistischen Bundesamtes⁸¹.

4.2.2. Zur Abdeckung der Krankenhausstatistik

Die Krankenhausstatistik ist für alle Krankenhäuser verbindlich⁸², auskunftspflichtig sind die Träger der jeweiligen Einrichtungen. Rehabilitations- und Vorsorgeeinrichtungen berichten ebenfalls, allerdings in einem erheblich eingeschränkten Merkmalsumfang. Die Krankenhausstatistik ist damit eine Totalerhebung.

Allerdings wird wiederholt angemerkt: *Krankenhäuser, die innerhalb des Erhebungsjahres oder zwischen dem Erhebungsstichtag und dem Meldetermin schließen, können in der Statistik u. U. nicht erfasst werden* (Statistisches Bundesamt, 2006–2010). Ähnliches dürfte im Falle von Zusammenschlüssen oder Eingliederungen kleinerer in größere Krankenhäuser gelten.

Während sich die inhaltliche Abdeckung der Krankenhausstatistik in den Erhebungsmerkmalen widerspiegelt und die zeitliche Abdeckung durch den Jahresbezug (Stichtag, Kalenderjahr, Geschäftsjahr) klar vorgegeben ist, bedarf die Frage der räumlichen Abdeckung einiger Zusatzüberlegungen.

⁸¹ <http://www.forschungsdatenzentrum.de/bestand/krankenhaus/index.asp> » Datenangebot Krankenhausstatistik » Metadaten, Zugriff am 02. 04. 2011.

⁸² zu den Ausnahmen siehe Abschn. 4.1.3.

Die Krankenhausstatistik ist eine Statistik der Einrichtungen (Krankenhäuser, Rehabilitations- und Vorsorgeeinrichtungen) und der behandelten Fälle. Festgehalten wird dabei der Sitz der Einrichtung, dieser kann dann einer Gemeinde, einem Kreis und einem Bundesland zugeordnet werden. Nicht erfasst wird dabei, ob und zu welchem Anteil die Leistungen des Krankenhauses am Ort des Sitzes erbracht werden, Außenstellen werden in der Krankenhausstatistik nicht ausgewiesen. Für die Behandlungsfälle gilt ähnliches: der registrierte Behandlungsort ist der Sitz des Krankenhauses. Mit dem Zusammenlegen von Krankenhäusern oder dem Eingliedern kleinerer in größere Einrichtungen ist anzunehmen, dass es vereinzelt solche Diskrepanzen gibt.

Für dieses Projekt wird davon ausgegangen, dass der Sitz der Einrichtung in dem Ort der Behandlung entspricht.

Die Diagnosedaten enthalten einen jeden Behandlungsfall unabhängig seiner Versicherung oder Herkunft, den deutschen Staatsbürger genauso wie Personen mit Aufenthaltsrecht, Touristen oder die Ausländer, die allein wegen der Behandlung in ein deutsches Krankenhaus kommen. Über den erfassten Wohnsitz des Patienten lässt sich jeweils der Anteil der lokalen und ortsfremden Behandlungsfälle ermitteln.

Deutsche (Inländer), die im Ausland behandelt werden, enthält die Krankenhausstatistik nicht. Gleichfalls enthält die Statistik eines jeden Bundeslandes nur die in diesem Bundesland behandelten Fälle (genauer, die in einem Krankenhaus mit Sitz in diesem Bundesland behandelten Fälle), die Krankenhausstatistik genügt somit dem Inlandskonzept (Abschn. 3.1.5).

4.2.3. Nicht verfügbare Merkmale und Informationen

Wird eine Statistik als Grundlage einer Sekundärauswertung herangezogen, stellen sich in Abhängigkeit der Fragestellung die Sekundärdaten (Abschn. 4.1.5) irgendwann als unvollständig, die Fragestellung nicht abdeckend heraus. Rosenberg et al. (2006, 6.1) sehen dies als Hauptnachteil bei der Nutzung von Sekundärdaten. Von ihnen benannte Gründe sind fehlende Merkmale, fehlende Angaben

oder wegen Mess- oder Kodierfehlern inkorrekte Werte – letztere beide werden mit der Qualität der Statistiken (Abschn. 4.3) diskutiert.

Im hier betrachteten Zeitraum beginnt sich das Krankenhausfinanzierungssystem zu ändern, aus einer budgetabhängigen Zuordnung der Mittel für den Betrieb des Krankenhauses wird eine fallpauschalenbezogene Vergütung der Leistungen, die sich ab 2005 in der DRG-Statistik (Abschn. 3.3.2) widerspiegelt.

Alternativ zur oder basierend auf der DRG-Statistik werden krankenhausspezifische Indizes zur Beurteilung der Leistung eines Krankenhauses herangezogen. Der bekannteste dieser Indizes, der *Case Mix Index* (CMI) von Medicare (USA), bildet dabei den mittleren Aufwand eines Krankenhauses über all seine (Medicare-relevanten) Behandlungsfälle ab, der Index multipliziert mit der Fallzahl wird dann als Maß für den akzeptierten Aufwand bei Krankenhausvergleichen herangezogen⁸³. Ein CMI für Krankenhäuser in Deutschland liegt jedoch nicht vor.

Deshalb hat Herr (2008, 2009) ein *Case Mix Weight* (CMW) genanntes Maß aus den mittleren Verweildauern je Hauptdiagnose konstruiert und angewendet. Herr setzt voraus, dass die Verweildauer eine gute Näherungsvariable für den Aufwand sei, und zweifelt dies sofort an (Herr, 2008, 2.3). Gleichzeitig fällt sie damit auch hinter die mit dem GSG seit 01. 01. 1993 gelebte Praxis einer Mischvergütung der Krankenhäuser durch tagesbezogene Pflegesätze und leistungsbezogene Pauschalentgelte zurück (Abschn. 3.2.1), der anerkannt höhere Aufwand für chirurgische Eingriffe und spezielle diagnostische Maßnahmen wird seither durch Sonderentgelte vergütet. Das Maß CMW von Herr wird als Summe über alle Krankenhäuser gewonnen; unerklärt bleibt, wieso Krankenhäuser mit sehr vielen Fällen einer Hauptdiagnose mit demselben Gewicht von 1 eingehen wie Krankenhäuser mit sehr wenigen. CMW wird auf ein und demselben Datensatz entwickelt und zu Aussagen über die Daten herangezogen, eine unabhängige Untersuchung der Eigenschaften wäre unabdingbare Voraussetzung für eine Einschätzung des Einsatzspektrums des CMW⁸⁴.

⁸³ Wie viele Ressourcen das Krankenhaus wirklich verbraucht, kann mit dem CMI nicht ermittelt werden.

⁸⁴ Für Deutschland wird dieses wohl nicht mehr angestrengt werden, die DRG-Statistik liefert seit 2005 eine feinere Einschätzungsmöglichkeit, ein CMI lässt sich leicht darauf aufbauen.

In den Daten der Krankenhausstatistik findet sich für die Jahre vor 2005 weder ein allgemein anerkanntes direktes Maß für die Beurteilung der Schwere der Fälle und somit den zu erwartenden Aufwand eines Krankenhauses, noch lässt sich ein solches aus den Daten unmittelbar ableiten.

4.2.4. Veränderte Informationen – räumliche und zeitliche Vergleichbarkeit

Statistiken ändern sich mit der Zeit, so wie sich die Rahmenbedingungen verändern. Veränderte Interessenlagen oder modifizierte Gesetzesforderungen lassen sich mit neuen oder modifizierten Erhebungsmerkmalen abbilden, Kodierungen werden zugunsten interner Konsistenz⁸⁵ geändert oder internationalen Gepflogenheiten⁸⁶ angepasst, Erfassungsbögen sind zu modifizieren oder werden zugunsten von direkter elektronischer Datenübermittlung verkleinert. So geben z. B. Rosenow und Steinberg (2002) als Gründe für die mit dem Jahr 2002 wirksam werdenden Neuerungen der KHStatV an: Anpassung an das geänderte Pflegesatzrecht und das Leistungsgeschehen in den Krankenhäusern sowie Harmonisierung mit anderen Erhebungen. Rosenow und Steinberg fahren fort: *Gleichzeitig stellen diese Umstellungen eine Vereinfachung für die Krankenhäuser dar. Der Umfang der Erhebung bleibt insgesamt nahezu gleich.*

Die Veränderungen der Krankenhausstatistik lassen sich im Wesentlichen auf jeweils einen der genannten Gründe zurückführen. Eine vollständige Übersicht aller Veränderungen im Untersuchungszeitraum kann mit Hilfe der jährlich erscheinenden Fachserie 12, Reihen 6.1 bzw. 6.1.1, 6.2 bzw. 6.2.1 und 6.3 entsprechend der Teile der Krankenhausstatistik gewonnen werden; wichtige Beispiele folgen hier.

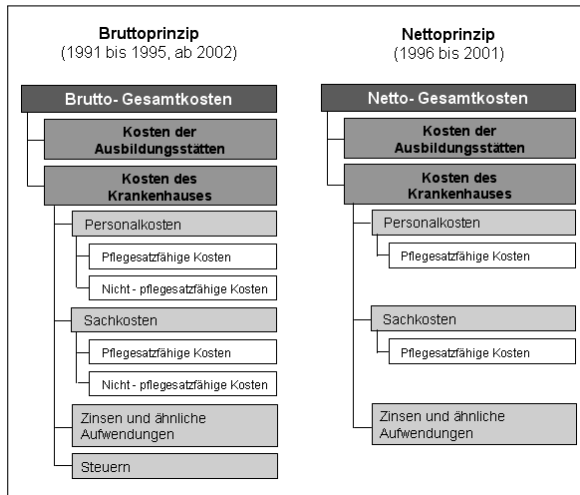


Abb. 4.2.: Brutto- und Nettoprinzip bei der Berechnung der Gesamtkosten eines Krankenhauses mit Gültigkeitszeiträumen^x

^x aus: Statistisches Bundesamt (2006a)

Kostendaten

Infolge der Bundespflegesatzverordnung (BPfIV) wurde mit dem Jahr 1996 die Ermittlung der Kosten der Krankenhäuser vom Bruttoprinzip auf das Nettoprinzip umgestellt, nach der Novellierung der Krankenhausstatistik-Verordnung (KHStatV) vom 13. August 2001 kehrte man mit dem Jahr 2002 wieder zum Bruttoprinzip zurück; Abb. 4.2 zeigt die wesentlichen Unterschiede.

Wie andere Autoren auch stellen Rosenow und Steinberg (2002) dar, dass nach dem Bruttoprinzip erhobene Kosten nicht mit denen nach Nettoprinzip verglichen werden können, eine longitudinale Sicht auf die Jahre über die Veränderungen hinweg ist somit nicht sinnvoll möglich. Einzige Ausnahme sind die bereinigten Kosten – die *pflegesatzfähigen Kosten* (Richter, 2008) bzw. *die vollstationären Kosten für allgemeine Krankenhausleistungen ... [sie] ergeben sich aus der Diffe-*

⁸⁵ eher gleiche statt ähnliche Codes für ähnliche Merkmale.

⁸⁶ z. B. Diagnosen nach jeweils international gültiger Version des ICD (International Classification of Diseases).

renz zwischen den Gesamtkosten und den Abzügen⁸⁷. Zu den nicht-pflegesatzfähigen Kosten der Krankenhäuser zählen dabei die Kosten für Personalunterkunft und -verpflegung oder wissenschaftliche Forschung und Lehre, die Kosten ambulanter Leistungen von Ärzten des Krankenhauses, die Kosten der Ambulanz des Krankenhauses selbst sowie – bei Lehrkrankenhäusern – die Kosten der Arztausbildung. Die bereinigten Kosten *sind über alle Jahre hinweg vergleichbar*⁸⁸.

Werblow und Schoffer (2010, Kap. 5.3.4) gehen davon aus, dass neben den bereinigten Kosten auch das Personal-Sachkosten-Verhältnis von den Veränderungen der Merkmalsdefinitionen unbeeinflusst ist und begründen dieses damit, dass beide Teile – Zähler und Nenner – gleichartig beeinflusst werden. Eine longitudinale Analyse des Personal-Sachkosten-Verhältnisses ist über die Definitionsbrüche hinweg sinnvoll möglich.

Krankenhaustyp, Planbetten, Großgeräte, Arbeitslosenquoten

Als weitere Definitionsveränderung eines Erhebungsmerkmals sei der Krankenhaustyp erwähnt: für die Jahre 2002 bis 2004 wurden Krankenhäuser mit ausschließlich neurologischen Betten zu den sonstigen Krankenhäusern gezählt, bevor man 2005 wieder zur vorherigen Definition – Krankenhäuser mit ausschließlich neurologischen Betten sind allgemeine Krankenhäuser – zurückkehrte. Longitudinale Betrachtungen werden damit erschwert.

Seit 2002 wird die Anzahl der Planbetten⁸⁹ nicht mehr erfasst, während die zu zählenden medizinisch-technischen Großgeräte um Digitale-Subtraktions-Angiographiegeräte, Gammakameras, Herz-Lungen-Maschinen und Dialysegeräte erweitert wurden; letzterer Anzahlen liegen somit für die Jahre bis 2001 nicht vor. Gleichzeitig wird der Begriff *Nieren- und Gallensteinzertrümmerer* durch *Stoßwellenlithotripter* ersetzt.

Entsprechend ändert sich jeweils der Erfassungsbogen für die jeweiligen Erhebungsjahre.

⁸⁷ http://www.gbe-bund.de/glossar/Bereinigte_Kosten.html, Zugriff am 21. 05. 2011.

⁸⁸ *ibid.*

⁸⁹ maximale Anzahl an Betten, die ein Krankenhaus laut Landeskrankenhausplan aufstellen darf.

Mit dem Inkrafttreten des SGB II (Hartz IV-Gesetz) 2005 wird die Definition von Arbeitslosigkeit verändert (SGB III, § 16). Damit einher geht eine definitionsbedingte Veränderung der Arbeitslosenquote, für longitudinale Analysen der Arbeitslosenquoten ist dieses zu beachten.

Vollstationär, teilstationär, ambulant oder Stundenfall?

Krankenhausfälle werden mit dem Ziel typisiert, die Typen voneinander abgrenzen zu können. Üblich Begriffe dabei sind der ambulante, der teilstationäre, der vollstationäre und der Stundenfall. Der Stundenfall wird bis 2002 in relevanten Publikationen der Amtlichen Statistik jeweils separat ausgewiesen⁹⁰; der Begriff *Stundenfall* selbst wird in den Erläuterungen der Grunddaten der Krankenhäuser des Statistischen Bundesamtes häufig verwendet und doch nicht erklärt (Statistisches Bundesamt, 2005c).

Das Bundessozialgericht hingegen hat sich mehrfach mit dem Unterschied zwischen vollstationärer, teilstationärer und ambulanter Behandlung befasst, die *geplante* Behandlungsdauer trennt dabei zwischen voll- und teilstationär (Anonym, 2005; Schulenburg, 2005).

In den Daten der Krankenhausstatistik gibt es kein Merkmal zur Separierung von Stundenfällen, teilstationären oder ambulanten Fällen von den Fällen mit vollstationärer Behandlung. Deshalb wird hier ein solches Merkmal generiert: *vollstationär* sind alle Fälle, deren Entlassung nicht am Aufnahmetag erfolgt, und alle Sterbefälle.⁹¹ Alle anderen in der Krankenhausstatistik dokumentierten Fälle werden hier ohne weitere Differenzierung als *teilstationär* betrachtet.

Kodierungsbrüche

Kodierungsbrüche, Veränderungen der Kodierung eines ansonsten definitions-konstanten Merkmals, dürften immer mit einer Erhöhung der Fehlerrate einher-

⁹⁰ seit 2003 wird zunehmend darauf verzichtet (Statistisches Bundesamt, 2005c).

⁹¹ Für Sterbefälle wird davon ausgegangen, dass wegen der Lebensbedrohlichkeit des Patienten-zustands die am Aufnahmetag geplante Behandlungsdauer mehr als einen Tag beträgt.

gehen. Dieser Nachteil eines veränderten Kodierungsschemas ist den Vorteilen der neuen Kodierung gegenüberzustellen.

Mit der ersten Novellierung der KHStatV modifiziert man bei gleichbleibenden Merkmalen die Kodierungen bestimmter Einträge. So ändert sich der Code für Gesamtkrankenhaus von 990 auf 999 bei der Erfassung der mittleren Beschäftigten-Anzahl ausgedrückt in Vollzeitäquivalenten, während die Codes für die Fachabteilungen gleich bleiben; diese Änderungen betreffen Satzart 3 (ärztliches Personal) und Satzart 4 (nichtärztliches Personal) gleichermaßen.

Für ein und dasselbe Merkmal *Träger des Krankenhauses* ändert sich das Kodierungsschema von 1 = öffentlich, 2 = freigemeinnützig, 3 = privat zu 1 = öffentlich unselbständig, 2 = öffentlich selbständig, 3 = öffentlich privatrechtlich, 4 = freigemeinnützig, 5 = privat – gleiche Codes mit unterschiedlicher Bedeutung und unterschiedliche Codes mit derselben Bedeutung sind bei longitudinalen Auswertungen zu beachten.

Seit 1994 beginnen die ersten Mitgliedsstaaten der WHO (World Health Organization) Krankheiten bzw. deren Diagnosen mit der Version 10 des ICD- (International Classification of Diseases) Schlüssels zu kodieren. Das Bundesministerium entschied Mitte 1999, dass mit dem Jahr 2000 u. a. die Diagnosen der Krankenhausstatistik mit der 10. (ICD-10) statt der bis dahin gültigen 9. Revision (ICD-9) des ICD-Schlüssels zu erfassen sind. Diagnosen der deutschen Krankenhausstatistik sind damit international unmittelbar vergleichbar.

Die ICD-10 stellt dabei ein gegenüber ICD-9 völlig verändertes Kodierungsschema dar, ein rein numerischer Schlüssel wird durch einen alphanumerischen ersetzt. Für beide gilt, dass die vierstellige Version jeweils eine feinere Spezifizierung der dreistelligen Diagnosen darstellt. Für dieses Projekt werden vom FDZ/L die Diagnosen in der dreistelligen ICD-9- bzw. ICD-10-Kodierung bereitgestellt.

Graubner (2000) hält dazu fest: *Für Zeitreihen von Diagnosenstatistiken wird generell die Überleitung von ICD-10-Schlüsselnummern in die der ICD-9 empfohlen, weil die umgekehrte Umschlüsselung zu fehlerhaft ist und zu wenig Informationen liefert.* Das DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) bietet ICD-Überleitungstabellen zwischen verschiedenen Revisionen der

ICD-Schlüssel⁹² zur freien Nutzung an. Mit der Konkordanzliste (Überleitungstabelle) zwischen ICD-9-WHO V6.0 und ICD-10-WHO V1.3 ist es möglich, abwärtskompatibel Kodierungen der ICD-10 auf ICD-9 zurückzuführen.

Administrative Zuordnungen

Die Grunddaten der Krankenhäuser enthalten identifizierende Merkmale zur eindeutigen Krankenhaus-Zuordnung (Name des Krankenhauses, Adresse des Krankenhaussitzes). Für die Arbeit an diesem Projekt werden daraus eineindeutig abgeleitete kontextfreie Krankenhausnummern bereitgestellt.

Die Krankenhausstatistik enthält neben den identifizierenden Merkmalen daraus abgeleitete Angaben wie die Gemeindenummer des Sitzes des Krankenhauses gemäß Amtlichem Gemeindeschlüssel (AGS) – eine 8-stellige Nummer, deren erste zwei Stellen das Bundesland, die folgenden drei den Kreis im Bundesland und die letzten drei die Gemeinde im Kreis identifizieren. Die Gemeindenummer ist entsprechend Gebietsstand im Berichtsjahr zu erfassen⁹³. Für die Datennutzung unter KDFV stehen die ersten fünf Stellen zur Kreis- und Bundeslandidentifizierung zur Verfügung.

Der AGS ändert sich – bundeslandspezifisch – (mindestens) mit jeder Kreisreform. Im Untersuchungsgebiet und -zeitraum betrifft das manche Bundesländer mehrfach: Brandenburg (06. 12. 1993), MV (12. 06. 1994), SN (01. 08. 1994, 01. 01. 1996, 01. 01. 1999, 01. 01. 2000), ST (01. 07. 1994) und TH (01. 07. 1994, 01. 01. 1998). Eine unmittelbare longitudinale Zuordnung des Sitzes eines Krankenhauses ist somit nicht möglich – der Kodierungsbrüche wegen –; die räumliche Vergleichbarkeit der Daten der Krankenhausstatistik über mehrere Jahre hinweg erfordert als erstes eine Anpassung der Zuordnungen zum Gebietsstand eines Referenzzeitpunktes. Referenzzeitpunkt für diese Arbeit ist der 31. 12. 2005.

⁹² <http://www.dimdi.de/static/de/klassi/diagnosen/ueberl/index.htm>, Zugriff am 01. 12. 2008.

⁹³ Eine rückwirkende Aktualisierung auf den zum Berichtszeitpunkt aktuellen Gebietsstand erfolgt – im Gegensatz zu VGR oder ETR – nicht.

4.2.5. Auswahl der Daten

Dieses Projekt zur Transition Ostdeutschlands legt Hauptaugenmerk auf die ostdeutschen Bundesländer einschließlich Berlins, das westdeutsche Bundesland Rheinland-Pfalz wird zum Vergleich herangezogen (Abschn. 2.5 und Abb. 2.5).

Obwohl die Krankenhausstatistik neben Krankenhäusern auch Rehabilitations- und Vorsorgeeinrichtungen erfasst, setzt sich die Erkenntnis durch, dass die Unterschiede zwischen beiden (Krankenhäuser: Heilung, Totalerhebung, voller Merkmalsumfang; Rehabilitations- und Vorsorgeeinrichtungen: Prävention und Nachbehandlung, Erfassung in Abhängigkeit von der Einrichtungsgröße⁹⁴, eingeschränkter Merkmalsumfang) eine getrennte Ergebnis-Betrachtung erfordern, wie die Aufteilung der Fachserie 12, Reihen 6.1⁹⁵ und 6.2⁹⁶ durch das StBA zeigt.

Für das Jahr 1991 – das erste Jahr der Krankenhausstatistik, das letzte Jahr vor Etablierung der StLÄ in Ostdeutschland – ist von Startschwierigkeiten bei der Einführung der neuen Statistik auszugehen (Schoffer, 2006). Ein Indiz dafür ist, dass die Grund- und Kostendaten nur für etwa 50 % der Krankenhäuser Thüringens vorliegen. Von einer Auswertung der Daten des Jahres 1991 wird Abstand genommen.

Grundlage vorliegenden Projektes sind deshalb die Daten der Krankenhausstatistik der erwähnten Bundesländer zwischen 1992 und 2005 ausschließlich der Daten zu Rehabilitations- und Vorsorgeeinrichtungen. Ergänzt werden diese um die entsprechenden kreisbezogenen Daten der Volkswirtschaftlichen Gesamtrechnung, Erwerbstätigenrechnung, der Arbeitslosenquoten und der siedlungsstrukturelle Gebietstypen.

Vertragsgemäß ergeben sich Einschränkungen bei den nutzbaren, durch das FDZ/L bereitgestellten Merkmalen aus der Krankenhausstatistik; eine Auflistung

⁹⁴ Unter den Rehabilitations- und Vorsorgeeinrichtungen sind die mit ≥ 100 Betten berichtspflichtig, das sind 53 % der Einrichtungen und 85,9 % der Fälle (2004).

⁹⁵ Grunddaten – seit 2004 getrennt in 6.1.1 (Krankenhäuser) und 6.1.2 (Rehabilitation und Vorsorge).

⁹⁶ Diagnosedaten – seit 2000 getrennt in 6.2.1 (Krankenhäuser) und 6.2.2 (Rehabilitation und Vorsorge).

der eingeschlossenen Merkmale findet sich im Anhang (Abschn. A.1)⁹⁷. Das Merkmal EF14 der Diagnosestatistik steht so nicht zur Verfügung, stattdessen wird ein daraus abgeleitetes binäres Merkmal zur Frage, ob der Patient (Fall) aus dem Kreis des Sitzes des Krankenhauses stammt (*Kreis des Krankenhauses = Wohnkreis?*) bereitgestellt. Auch dieses Merkmal erfordert vorab eine Anpassung der Zuordnungen zum Gebietsstand auf den Referenzzeitpunkt 31. 12. 2005.

4.3. Qualität der Statistiken

Datenqualität, speziell die Qualität von Statistiken, hat einen enormen Einfluss auf die Glaubwürdigkeit jeglicher Schlussfolgerungen sowohl für die Wissenschaft als auch für die breite Öffentlichkeit. Fehler bei der Bereitstellung des Datenmaterials ... *können durch keine noch so gute ökonomische Theorie vollständig wettgemacht werden* (von Auer, 2007, S. 11). Mögen auch Politik und Öffentlichkeit gelegentlich absolute Fehlerlosigkeit von Daten erwarten, ist diese doch nie zu garantieren.

4.3.1. Kriterien der Qualitätsbeurteilung von Statistiken

Für amtliche Daten kann man sich an der *Definition of Quality in Statistics* von Eurostat⁹⁸ und deren 6 Kriterien orientieren: Relevanz (*relevance*), Akkuratessse oder Korrektheit (*accuracy*), Aktualität und Pünktlichkeit (*timeliness and punctuality*), Zugänglichkeit und Übersichtlichkeit (*accessibility and clarity*), Vergleichbarkeit (*comparability*) und Kohärenz bzw. Zusammenhang (*coherence*). Ergänzend wird darauf hingewiesen, dass die Kosten für die Erstellung der Statistik ebenso wie die Belastung, die diese Erstellung für die Befragten darstellt, qualitätsrelevante Parameter sind, obwohl diese selbst keine Maße für Qualität selbst darstellen. Da die Erklärung von Eurostat für Akkuratessse (*Accuracy in the general statis-*

⁹⁷ Einige der Merkmale erweisen sich im Laufe des Projekts als weniger interessant, wie z. B. die einzelnen Kostenarten, die alle dem Definitionswechsel Brutto-Netto-Brutto (Abschn. 4.2.4: Kostendaten) unterliegen.

⁹⁸ <http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf>, Zugriff am 11. 06. 2011.

tical sense denotes the closeness of computations or estimates to the exact or true values.) von Nähe zum exakten oder wahren Wert spricht, ist anzunehmen, dass Eurostat Fehlerlosigkeit der Daten nicht für ein realistisches Ziel hält; ganz im Unterschied zur Deutschen Gesellschaft für Informations- und Datenqualität e. V. (DGIQ), deren *15 Dimensionen der Informationsqualität* Fehlerfreiheit einschließt: *Fehlerfreiheit (free of error): Informationen sind fehlerfrei, wenn sie mit der Realität übereinstimmen*⁹⁹. Diese 15 Dimensionen der DGIQ stimmen fast mit den 16 *Data quality dimensions* von Pipino et al. (2002) überein, jedoch erklären Pipino et al. *free-of-error* als *the extent to which data is correct and reliable* und dabei anerkennend, dass die Fehlerfreiheit von Daten nur graduell erreicht sein mag. Eine konsolidierte Sicht auf Datenqualität und ihre Dimensionen gibt es bisher nicht; Batini und Scannapieco (2006) geben in *2. Data Quality Dimensions* einen Überblick über eine Reihe von Begriffssystemen.

Pipino et al. (2002) weisen darauf hin, dass die Frage der Korrektheit von den Umständen bzw. der Bedeutung des jeweiligen Datums abhängig ist. Das wiederum ist auch die Sicht von Wallgren und Wallgren (2007, S. 11), wenn sie sagen, dass Fehler vorkommen können und es aus Statistikers Perspektive darauf ankommt, diese unter Kontrolle zu behalten und (möglichst) zu vermeiden. Fehler in den Daten sind für Wallgren und Wallgren akzeptabel, wenn davon ausgegangen werden kann, dass ihr Einfluss auf das Ergebnis gering ist.

Die Kriterien von Eurostat auf Sekundärdatenanalysen (Abschn. 4.1.5) anzuwenden bedeutet: die Fragen der Relevanz sind themenspezifisch vorab zu klären; auf Aktualität und Pünktlichkeit sowie Zugänglichkeit und Übersichtlichkeit kann aus Sicht der Sekundäranalyse (rückwirkend) kein Einfluss mehr genommen werden. Vergleichbarkeit und Kohärenz sollten bei der Nutzung von Sekundärdaten genauso vorab geprüft werden wie die Korrektheit der Daten. Zusätzlich zu bereits vorliegenden Fehlern können Qualitätseinschränkungen durch Datenintegration und -verknüpfung entstehen (de Veaux und Hand, 2005), auch führen das Fehlen oder Irrtümer bei der Interpretation der zugehörigen Metadaten zu einer weiteren Minderung der Datenqualität.

Die nachträgliche Prüfung von Daten auf Korrektheit ist in Gänze ein unmögliches Unterfangen. Ein Einzeldatum außerhalb des gültigen (eine Schulnote von 8) oder

⁹⁹ <http://www.dgiq.de/> » Service » Downloads » 15. IQ Dimensionen.pdf, Zugriff am 11. 06. 2011.

des sinnvoll anzunehmenden Wertebereichs (eine Körperhöhe eines Mannes von 3,43 m) mag mit endlichem Aufwand gefunden werden. Dieses Vorgehen nennt man Signierkontrolle, während eine Kombinationskontrolle die logische Abhängigkeit zwischen Einzeldaten nutzt (ein Krankenhaus ohne geburtshilfliche Station jedoch mit mehreren Geburten im Jahr könnte einen Datenfehler bedeuten). Kombinationskontrollen erfordern ein deutlich komplexeres Herangehen – vom Festlegen der logischen Beziehungen bis zur Algorithmierung derselben – für eine effiziente Datenprüfung.

Jegliche Verbesserung der Datenqualität ist (auch) eine Kostenfrage (Radermacher und Körner, 2006). Dass sich ein als falsch herausstellender Wert sinnvoll korrigieren lässt, ist nur von Fall zu Fall gegeben; für Sekundärdatenanalysen sind die Chancen dafür nochmals geringer.

Im Rahmen der Qualitätssicherungsmaßnahmen der Amtlichen Statistik werden die Daten vor der Freigabe zur Nutzung geprüft. Auch wenn *10 Jahre bundeseinheitliche Krankenhausstatistik* von Rosenow und Steinberg (2002) keine Aussage zur Qualität der Daten enthält, wird für die Krankenhausstatistik die Qualität der Daten wiederkehrend in Qualitätsberichten dokumentiert (z. B. Statistisches Bundesamt, 2006b,c,d). Da es sich bei der Krankenhausstatistik um eine Totalerhebung handelt, sind stichprobenbedingte Fehler ausgeschlossen; überhaupt seien – so der Tenor der Qualitätsberichte – die Fehlermöglichkeiten sehr gering. Allerdings könne Untererfassung vorkommen, da im Berichtsjahr neu eröffnete Krankenhäuser nicht zwingend gemeldet werden müssten und innerhalb des Erhebungsjahres schließende Krankenhäuser nicht mehr meldeten. Bömermann und Hauswirth (2002) beschreiben den Erhebungs- und Aufbereitungsprozess für Diagnosedaten im StLA Berlin ausführlicher und verweisen auf die Ausführungen von Wein (2002), die unter dem Titel *Datenqualität und Datenprüfungen* aus gemeinsamer Sicht der StLÄ und des StBA eine *kurze Vorstellung des Qualitätsbegriffs der deutschen amtlichen Statistik* anbieten.

Bei Wein wird darauf hingewiesen, dass zu einer Plausibilitätskontrolle auch vorhergehende Statistiken¹⁰⁰ herangezogen werden. Zeitliche Vergleichbarkeit bezieht sich hingegen nur auf Veränderungen von Merkmalen durch Definitionsveränderung (Statistisches Bundesamt, 2006c,d,b); nicht vergleichbar sind z. B. die

¹⁰⁰ zu verstehen als: dieselbe Statistik zu vorangehenden Erfassungszeiträumen.

Kosten zwischen den Perioden 1991–1995, 1996–2001 und seit 2002¹⁰¹. Die für eine Panel-Sicht auf die Daten notwendige Konstanz der Messinstrumente und Maße wird nicht betrachtet.

Für die Diagnosedaten wird amtlicherseits (Statistisches Bundesamt, 2006b) festgestellt: *... ist stets zu bedenken, dass die Vergabe des Diagnoseschlüssels im Krankenhaus erfolgt und bereits dort Fehler auftreten können. Dadurch, dass die Angabe der Hauptdiagnose für die Krankenhäuser entgeltrelevant ist, kann davon ausgegangen werden, dass die Qualität der Diagnoseangabe hoch ist.*

4.3.2. Empirische Einschätzung der Datenqualität: Daten im offenen Zugriff

Amtliche Daten sind plausibilisiert und qualitätsgeprüft; eine systematische Prüfung dieser Daten für vorliegendes Projekt ist deshalb nicht vorgesehen.

Trotzdem ergeben sich mit der Anbindung der amtlichen Daten der BA, des BBR und der Arbeitskreise VGR und ETR zu den Daten der Krankenhausstatistik Auffälligkeiten, deren Bedeutung für die weitere Arbeit entschieden werden muss. Diese Auffälligkeiten sind Ergebnis eines reaktiven Prozesses, der sich aus Betrachtungen der gültigen Wertebereiche, der Häufigkeitsverteilungen mit Minimum und Maximum (nur am GWAP vor Ort zulässig), weiteren grafischen Darstellungen und Reaktionen auf unerwartete oder unerklärliche Modellierungsergebnisse zusammensetzt. Misstrauen gegenüber den Daten führt sowohl zu Klärungs- und Korrekturanstrengungen als auch zum Versuch der Einschätzung, welchen Einfluss ein jeder auffälliger Wert auf die Ergebnisse haben könnte.

Berichtet werden die Datenauffälligkeiten sowie die Entscheidungen zur Datenqualitätsverbesserung hier in ihrer thematischen Zuordnung und nicht in der Reihenfolge des Auffindens.

¹⁰¹ wegen der Definitionswechsel Brutto-Netto-Brutto (Abschn. 4.2.4: Kostendaten).

Bundesagentur für Arbeit – Arbeitslosenquoten

Obwohl mit der Wiedervereinigung die Einführung des bundesdeutschen Rechts- und Sozialsystems einhergeht, stehen die Arbeitslosenzahlen und -quoten für Ostdeutschland erst ab März 1996 (Quartalswert) zur Verfügung. Obwohl nicht dokumentiert ist anhand der Daten zu vermuten, dass die angegebenen Werte für *Berlin, Stadt* sich bis einschließlich Dezember 1995 nur auf die westlichen Stadtbezirke beziehen.

Für den Kreis *Eisenach, Stadt* in Thüringen fehlen die Angaben für die Jahre 1996 und 1997 komplett. Da die Stadt Eisenach bis zum 31. 12. 1997 dem *Wartburgkreis* angehört, wird hier für 1996 und 1997 die Arbeitslosenquote Eisenachs der des Wartburgkreises gleich gesetzt. Für die Kreise *Brandenburg an der Havel, Stadt; Cottbus, Stadt; Potsdam, Stadt* und *Potsdam-Mittelmark* werden für März 1996 Arbeitslosenquoten von 0,0% angegeben, obwohl jeweils Arbeitslosenzahlen von mehr als 5000 dokumentiert sind. Unter der Annahme, dass der Nenner – die Anzahl der Erwerbspersonen als Summe aus den Anzahlen von Erwerbstätigen und Arbeitslosen – jeweils zwischen zwei aufeinanderfolgenden Quartalen (nahezu) unverändert bleibt, wird die Arbeitslosenquote für März 1996 als Quotient aus den Arbeitslosenzahlen März 1996 und den Erwerbspersonen im Juni 1996 berechnet.

Mit der Rückkehr zu monatlichen Arbeitslosenzahlen und -quoten im Januar 1998 wird jeder Prozentwert mit der unglaublichen Anzahl von 13 Nachkommastellen angegeben; für dieses Projekt werden Arbeitslosenquoten einheitlich auf den Standard bis Dezember 1997 – eine Nachkommastelle – gerundet.

Bundesamt für Bauwesen und Raumordnung – siedlungsstrukturelle Gebietstypen

Die Angaben zu siedlungsstrukturellen Gebietstypen als Regionsgrundtyp und als spezifischer Kreistyp sind plausibel – dargestellt in Abb. 4.1 –, sie werden somit als fehlerfrei angenommen und unverändert den Analysen zugeführt.

Arbeitskreise VGR und ETR – Wirtschaftsdaten

Die Daten für Sachsen-Anhalt für das Jahr 2005 zum Gebietsstand 2005 sind individuell nachgefordert und zu den Daten der anderen Bundesländer hinzugefügt. Für die Jahre bis 1995 liegen Angaben für einige Merkmale nicht vor, doch das ist für alle Kreise konsistent (Abschn. 4.1.4).

Ausnahme bilden die fehlenden Angaben zum Brutto-Inlandsprodukt der Kreise *Eisenach*, *Stadt* und *Wartburgkreis* für die Jahre bis 1995 einschließlich¹⁰². Die aggregierten Werte beider Kreise gemeinsam werden jeweils als Differenz der Landessumme Thüringen und der Summe aller anderen Kreise von Thüringen gewonnen. Einzelwerte ließen sich daraus proportional zu den Verhältnissen 1996 – dem ersten Jahr lückenloser Berichterung – gewinnen, doch das ist für vorliegendes Projekt nicht nötig (Abschn. 4.4).

4.3.3. Empirische Einschätzung der Datenqualität: Krankenhausstatistik

Auffälligkeiten in den Daten der Krankenhausstatistik beziehen sich hier auf die kumulierten Totalerhebungen für die Krankenhäuser der Bundesländer Berlin, Brandenburg, Mecklenburg-Vorpommern, Rheinland-Pfalz, Sachsen, Sachsen-Anhalt und Thüringen – das Untersuchungsgebiet – der Jahre 1991 bis 2005 einschließlich, so wie sie vom FDZ/L für das Projekt FDZ-Nutzung 606-2007 (Forschungsdatenzentrum der Statistischen Landesämter, 2007) zur Verfügung gestellt werden.

Diagnosedaten

Für die Nutzung amtlicher Daten im medizinischen und speziell Krankenhaus-Umfeld sind für Iezzoni (1997) die Datenqualität hinsichtlich folgender Attribute besonders wichtig: der klinische Inhalt der Daten, die Akkuratessse der Kodierungen,

¹⁰² Die Ausgliederung der Stadt Eisenach aus dem Wartburgkreis erfolgt hingegen erst mit dem 1. Januar 1998.

die Vollständigkeit der Kodierungen sowie Unterschiede zwischen Erfassenden-Einheiten. Aus der US-amerikanischen Sicht von Iezzoni gelten speziell Kodierungen, die in direkter Beziehung zur Vergütung von Krankenhausleistungen stehen, als besonders kritisch für die Datenqualität.

Laut Merkmalsdefinition der Krankenhausstatistik wird die Hauptdiagnose als ... *die Diagnose angegeben, die nach Analyse als diejenige festgestellt wurde, die hauptsächlich für die Veranlassung des stationären Krankenhausaufenthaltes des Patienten verantwortlich ist. Der Begriff ‚nach Analyse‘ bezeichnet die Evaluation der Befunde am Ende des stationären Aufenthalts* (Statistisches Bundesamt, 2005a). Das mag im Falle multimorbider Patienten gelegentlich zu einem Auswahl dilemma führen, diese Definition hat sich jedoch über die Jahre als brauchbar durchgesetzt, sie ist auch nach der Einführung der DRG-Statistik gültig. Die Hauptdiagnose ist – als das einzige unmittelbar die Krankheit beschreibende – das bedeutendste Merkmal in den Diagnosedaten, sie wird zum Abschluss eines jeden Falles in den Krankenhäusern¹⁰³ nach ICD-9 und seit 2000 nach ICD-10 kodiert; ohne Hauptdiagnose ist ein Fall nicht vorstellbar.

Trotzdem gibt es in den Daten Fälle ohne die Angabe einer Hauptdiagnose. Tab. 4.1 zeigt für das Untersuchungsgebiet und die Jahre ab 1993¹⁰⁴ die Anzahl der Fälle ohne Diagnoseangabe. Obwohl der Qualitätsbericht Diagnosen (Statistisches Bundesamt, 2006b) meint: *Sofern es zu Antwortausfällen auf Ebene der Krankenhäuser kommt und diese ein spezifisches Behandlungsspektrum aufweisen, können bestimmte Hauptdiagnosen unterrepräsentiert sein. In diesem Fall wäre durch die Unvollständigkeit der Daten die Vergleichbarkeit beeinträchtigt*, ist dieser Anteil mit maximal 0,162 % (1993) gering; hier wird davon ausgegangen, dass sich dieser unsystematisch über die Krankenhäuser verteilt. Bemerkenswert ist, dass bei allgemein abnehmender Tendenz in der Anzahl der Fälle ohne Hauptdiagnose das Jahr 2000 mit einem Wiederanstieg auf 0,143 % einen Bruch darstellt. Dieser Bruch fällt zeitlich mit der Umstellung der Diagnoseverschlüsselung auf ICD-10 zusammen.

Fälle ohne Hauptdiagnose sind von der weiteren Analyse ausgeschlossen.

¹⁰³ ob in den einzelnen Fachabteilungen oder mehr zentralisiert, muss hier offen bleiben.

¹⁰⁴ Beginn der jährlichen Erhebung von Diagnosedaten.

Tab. 4.1.: Anzahl registrierter Krankenhaus-Fälle im Untersuchungsgebiet 1993–2005

	Ausschluss ^w	valide ^x	unvollst. ^y
1993 ^z	5.408	3.338.898	(20.408)
1994	5.250	3.826.898	(16.714)
1995	4.646	3.912.165	(9.783)
1996	2.649	4.045.624	(6.974)
1997	1.962	4.138.203	(2.616)
1998	1.716	4.274.849	(3.743)
1999	1.347	4.359.276	(1.756)
2000	6.335	4.412.355	(2.033)
2001	4.243	4.400.251	(859)
2002	2.646	4.379.841	(1.269)
2003	578	4.417.006	(702)
2004	984	4.310.315	(376)
2005	375	4.298.398	(1.587)
Σ	38.139	54.114.079	(68.820)

^w Fälle ohne Angabe der Hauptdiagnose – als invalide betrachtet

^x Anzahl auswertbarer Fälle

^y (davon) Fälle mit Fehlangaben bei Geschlecht, Sterbefall oder Wohnort

^z Für 1993 fehlen die Diagnosedaten von Mecklenburg-Vorpommern.

Wichtige weitere Merkmale der Diagnosedaten sind das Geschlecht und der Wohnort¹⁰⁵ des Patienten sowie die Frage, ob der Patient das Krankenhaus als Sterbefall verlässt. Mindestens eine dieser Angaben fehlt bei einer Reihe von Fällen, so dass diese Fälle in bestimmte Analysen als Fehlwerte eingehen müssen. Diese Anzahl dieser Fälle sinkt über die Jahre, das Maximum liegt bei 0,611 % im Jahr 1993.

Die verschlüsselten Hauptdiagnosen selbst entziehen sich einer inhaltlichen Nachprüfung, von einer formalen Übereinstimmung mit gültigen Codes wird generell ausgegangen. Bezüglich des Wechsels auf ICD-10 stellt das Statistische Bundesamt (2004) dar: *Einige wenige Krankenhaufälle (2 %) wurden im Berichtsjahr 2000 von den Einrichtungen noch nach der alten ICD-9-Systematik nachgewiesen.* Allerdings ist dabei stark zu vermuten (Schoffer, 2009), dass dieses lokal gehäuft auftritt, stationsweise, wohl abhängig vom Kodierenden.

¹⁰⁵ für das Projekt als Binärmerkmal: *Wohnhaft im Kreis des Krankenhauses?* bereitgestellt.

Leistungsfähigkeit der ICD-Konkordanzlisten

Um eine zeitliche Vergleichbarkeit der Hauptdiagnosen herzustellen, werden die mit ICD-10 verschlüsselten Diagnosen mit der Konkordanzliste (Überleitungstabelle, Abschn. 4.2.4) nach ICD-9 umgeschlüsselt. Die klar getrennten Wertebereiche für ICD-9 (3 Ziffern) und ICD-10 (1 Buchstabe gefolgt von 2 Ziffern) ermöglichen die automatische, eindeutige Rekodierung auch ohne Jahreseinschränkung¹⁰⁶.

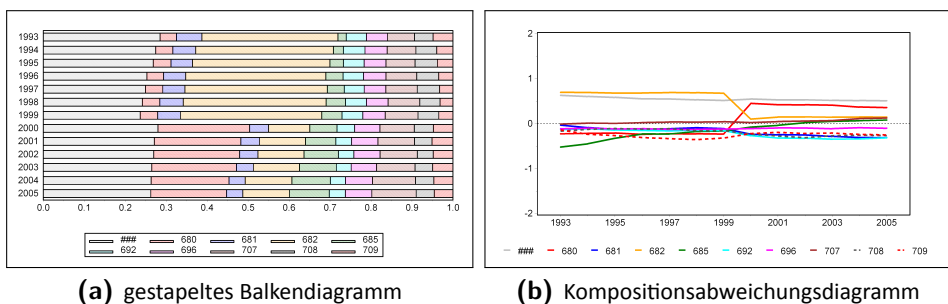


Abb. 4.3.: ICD-9-Kapitel XII Krankheiten der Haut^x (30 Diagnosen^y, 748.948 Fälle)

^x Krankheiten der Haut und des Unterhautzellgewebes = Diagnose-Gruppierung 680–709

^y hier: 9 häufigste Diagnosen und Aggregat (###, per Amalgamation) aller anderen

Schaut man sich die Daten zu den Diagnosen des ICD-9-Kapitels XII Hautkrankheiten (Abb. 4.3) an, so fällt auf, dass die gestapelten Balkendiagramme (a) einen sprunghaften Verlauf zwischen 1999 und 2000 für drei Komponenten zeigen: für die Diagnosen 680 (rötlich) und 682 (beige) sowie die Aggregat-Komponente aller verbleibenden Diagnosen dieses Kapitels. Im Kompositionsabweichungsdiagramm (b)¹⁰⁷ wird ersichtlich, dass die sprunghaften Veränderungen der Diagnosen 680 und 682 viel bedeutsamer sind als die der Aggregat-Komponente. Mehr noch, die Veränderungen der Diagnosen 680 und 682 kompensieren sich gegenseitig in etwa; alle anderen Veränderungen bleiben im Vergleich dazu unbedeutend.

¹⁰⁶ ICD-9-Codes werden unverändert übernommen.

¹⁰⁷ für eine Definition siehe Abschn. 5.1.13.

Hält man sich vor Augen, dass mit dem Jahreswechsel 1999 zu 2000 die Kodierung nach ICD-10 die bis dahin geltende nach ICD-9 ablösen sollte, kann man vermuten, dass das ICD-Rekodierungsprogramm von ICD-10 auf ICD-9 an dieser Stelle einen (systematischen) Fehler aufweist. Eine Analyse der Kodierungsvorschriften ergibt, dass ICD-9-Code 680 (*Karbunkel und Furunkel*) nicht völlig mit ICD-10-Code L02 (*Hautabszess, Furunkel und Karbunkel*) gleichzusetzen ist. ICD-10-Code L03 (*Phlegmone*) hingegen umfasst die Abszesse (der Haut) nicht, diese finden sich jedoch in ICD-9-Code 681 (*Phlegmone und Abszeß der Finger und Zehen*) und 682 (*Phlegmone und Abszeß sonstigen Sitzes*). Völlige Konkordanz scheint nur in der Zusammenfassung der ICD-9-Codes 680 bis 682 (entsprechend dann ICD-10-Code L02 und L03) möglich zu sein, natürlich auf Kosten einer feinen Aussagegüte und unter Klärung der Frage, ob medizindiagnostisch eine Kategorie *Abszess, Furunkel, Karbunkel und Phlegmon* sinnvoll interpretierbar ist.

Empirische Analysen dieses und aller anderen Diagnose-Kapitel ergibt zwischen den Jahren 1999 und 2000 keine weiteren offensichtlichen Abruptveränderungen in den Diagnose-Codes. Eine feinere Unterteilung (Abschn. 4.4.3: Diagnose-Gruppierungen) hingegen zeigt ähnlich komplexe Verschiebungen z. B. unter den Diagnosen 580–589 (Nephritis, Nephrotisches Syndrom und Nephrose): der sprunghafte Anstieg in der relativen Häufigkeit der Diagnose 583 (Nephritis und Nephropathie, ohne Angabe, ob akut oder chronisch) zwischen 1999 und 2000 fällt in absoluten Zahlen zusammen mit einem sprunghaften Anwachsen der Fallzahl für die ganze Diagnose-Gruppierung von 20.034 (1999) auf 28.478 im darauffolgenden Berichtsjahr.

Verweildauern

Die Differenz zwischen aus Entlassungsdatum eines Patienten und seinem zugehörigen Aufnahmedatum ergibt die Verweildauer dieses Falles in Tagen¹⁰⁸; Fälle werden dem Berichtsjahr des Entlassungsdatums zugeordnet. Probleme ergeben sich bei fehlenden Datumsangaben oder implausiblen Beziehungen zwischen den Daten. Fällen ohne Angabe eines gültigen Entlassungs- oder Aufnahmedatums und Fällen, bei denen die Entlassung vor der Aufnahme erfolgt, kann keine Ver-

¹⁰⁸ Fälle, also vollstationäre Patienten, mit einer Verweildauer von weniger als einem Tag sind möglich.

weildauer zugeordnet werden. Fälle, bei denen der 29. Februar registriert ist, obwohl es sich nicht um ein Schaltjahr handelt, werden hier auf den 28. Februar korrigiert.

Für die ersten Jahre der Krankenhausstatistik werden Verweildauern von über 20 Jahren berichtet, die Anzahl dieser Fälle nimmt mit den Jahren ab. Diese Abnahme entspricht den Vorgaben, die (hauptsächlich psychischen) Langzeitkranken zeitiger als vordem in Pflegeheime zu überweisen. In den vorliegenden Daten fällt jedoch auf, dass Fälle großer Verweildauern sehr wohl in Berlin, Rheinland-Pfalz und Sachsen nachweisbar sind, die Daten Thüringens hingegen keine solche Fälle enthalten – ein bundeslandspezifischer Umgang mit dem Problem lässt sich nur vermuten. Obwohl in der Definition der Krankenhausstatistik ausgeschlossen, ist es nach Informationen aus dem StBA (Schoffer, 2008) auch möglich, dass es sich bei einigen um *forensische Fälle*, d. h. psychiatrische Fälle mit gerichtlicher Entscheidung, handelt.

Eine Datenklärung dieser Fälle ist hier nicht möglich, große Verweildauern werden deshalb für Auswertung in einer Kategorie als rechts-zensierte Werte zusammengefasst.

Fehlende Daten einzelner Bundesländer

Jedes Datum in den Daten der Krankenhausstatistik ist dem jeweiligen Krankenhaus und somit dem Kreis und Bundesland des Krankenhaus-Sitzes zugeordnet, das Bundesland ist dann Eigner dieser Daten.

Für das erste Jahr der Krankenhausstatistik, 1991, ist von einer deutlichen Verringerung der Anzahlen der Krankenhäuser in den ostdeutschen Bundesländern auszugehen. Das Land Thüringen hat im Januar 79 Krankenhäuser, am 31. 12. 1991 sind es 10 weniger (Ossen, 2011). Anzunehmen ist, dass diese Krankenhäuser nicht in die Krankenhausstatistik eingegangen sind. Desweiteren ist für dieses erste Jahr der Krankenhausstatistik allgemein von einer eher geringen Datenqualität auszugehen (Schoffer, 2008); für das Land Thüringen z. B. liegen nur für die Hälfte aller Krankenhäuser Angaben zu den Grund- und Kostendaten vor.

Deshalb wird für dieses Projekt auf die Daten des Jahres 1991 komplett verzichtet.

Für 1993 – das Jahr der erstmaligen Erhebung der Diagnosedaten – fehlen aus Mecklenburg-Vorpommern jegliche Diagnosedaten¹⁰⁹. Größenordnungsmäßig ist dabei von 330.000 Fällen auszugehen¹¹⁰. Obwohl weder dieser massive Datenausfall noch sein Komplement, die Fälle der anderen Bundesländer, als Zufallsauswahl angesehen werden können, bleibt Mecklenburg-Vorpommern Teil aller Analysen.

Für die vier aufeinander folgenden Jahre 1998 bis 2001 fehlen die gesamten Grund- und Kostendaten aus Mecklenburg-Vorpommern¹¹¹. Auch wenn die Frage nach dem Bundesland für die zugehörigen Diagnosedaten beantwortbar ist, ist damit die originale Zuordnung zum jeweiligen Krankenhaus und somit dem Kreis des Sitzes des Krankenhauses verloren gegangen. Diese Angaben werden für dieses Projekt durch Abgleich der jeweiligen Krankenhausnummer (Abschn. 4.2.4) mit den Jahren vor 1998 bzw. nach 2001 erzeugt; das Bestehen eines Krankenhauses in einem Jahr jener Zeitperiode wird angenommen, wenn in den Diagnosedaten jenes Jahres (mindestens) ein Fall dem Krankenhaus zugeordnet wird¹¹².

Dieser Datenausfall – definitiv nicht als zufälliges Fehlen von Werten interpretierbar – hat vor allem Implikationen für die gemeinsame Deskription aller Krankenhäuser des Untersuchungsgebietes; bei der Modellierung mit Regressionsverfahren über alle Jahre wird diesen Fehlstellen eine Teilnahme am mittleren Trend unterstellt. Mecklenburg-Vorpommern bleibt Teil aller Analysen.

Zwei der von Wallgren und Wallgren (2007, S. 13) postulierten Vorteile von Registern und registerbasierter Statistik (Abschn. 3.5) – die (fast) vollständige Abdeckung der Grundgesamtheit und die komplette Abdeckung des untersuchten Zeitraumes – realisieren sich für dieses Projekt nicht.

¹⁰⁹ Sowohl Grund- als auch Kostendaten sind vorhanden.

¹¹⁰ Rückrechnung bei Proportionalitätsannahme zum MV-Anteil an allen im Folgejahr 1994.

¹¹¹ eines der 1 bis 2 Bundesländer, für die Daten im Datenangebot des StBA zur Krankenhausstatistik fehlen: <http://www.forschungsdatenzentrum.de/bestand/krankenhaus/index.asp>, Zugriff am 03. 04. 2011.

¹¹² Diese Zuordnung deckt 100 % der fraglichen Fälle ab.

Identifikation der Krankenhäuser – Paneldaten

Das einzelne Krankenhaus ist die wichtigste Erfassungseinheit der Krankenhausstatistik. Zielt eine Analyse der Krankenhausstatistik auf mehr als jährliche Querschnittsanalysen, so ist die eindeutige jahresübergreifende Identifikation eines jeden Krankenhauses notwendige Voraussetzung¹¹³ für Longitudinal- oder Panelanalysen (Frees, 2004, S. 8).

Tab. 4.2.: Anzahl unterschiedlich identifizierter Krankenhaus-Einträge in der Krankenhausstatistik über alle Jahre 1992–2005, jedes Jahr einzeln zum Vergleich^x

	RP	BB	MV	SN	ST	TH	BE	Σ
1992–2005	146	75	47	118	76	79	142	683
1992	116	66	43	104	71	66	100	566
1993	119	64	41	102	69	65	99	559
1994	119	60	37	97	69	63	102	547
1995	119	60	37	96	66	60	98	536
1996	118	59	37	97	62	57	69	499
1997	118	58	36	98	60	57	69	496
1998	117	55	36	97	59	56	74	494
1999	119	53	34	96	58	53	75	488
2000	119	54	34	93	56	53	77	486
2001	114	53	34	93	55	53	71	473
2002	112	51	36	89	54	49	68	459
2003	103	49	35	88	53	49	70	447
2004	99	47	34	87	51	50	72	440
2005	98	47	34	86	50	46	72	433
2005 ^y	97+1	47	34	85+1	50	46	71+1	430+3

^x vgl. Tab. 6.1 auf S. 170

^y Statistisches Bundesamt (2007, S. 246), ergänzt um Bundeswehrkrankenhäuser

Tab. 4.2 zeigt für 1992 bis 2005 je Jahr und Bundesland die Anzahlen der unterschiedlichen Krankenhausnummern (Abschn. 4.2.4). Diese sind in guter Übereinstimmung mit den Angaben der jeweiligen Statistischen Jahrbücher für Deutschland, kleine Differenzen können sich durch die unterschiedlichen Berichtsmoda-

¹¹³ Wiederholte Querschnittsanalysen über Zeitverläufe von Aggregaten werden hier nicht betrachtet.

litäten ergeben: Das StBA berichtet zum Stichtag 31. 12. des jeweiligen Jahres und tabelliert ohne Bundeswehrkrankenhäuser. Die Krankenhausstatistik hingegen enthält alle Krankenhäuser, die für das Berichtsjahr Daten übergeben haben, strukturell also auch alle Neugründungen und diejenigen, die im Berichtsjahr geschlossen oder fusioniert werden.

Im Untersuchungszeitraum und -gebiet berichten maximal 566 Krankenhäuser (1992) an die Krankenhausstatistik. Für diese werden 683 verschiedene Krankenhausnummern verwendet. Das hieße bei eindeutiger Zuordnung von Krankenhäusern zu Identifikationsnummern, dass es sich bei $\frac{1}{6}$ der nachgewiesenen Krankenhäuser um Neugründungen handeln muss – das erscheint nicht realistisch glaubhaft¹¹⁴.

Es ist davon auszugehen, dass das hier verwendete Panel der Daten der Krankenhausstatistik weiterhin Fälle von Panel-Attrition (Schließung von Krankenhäusern) und Panel-Zuwachs (Neugründungen) enthält, die allein auf Identifikations-Diskontinuitäten zurückzuführen sind.

Grund- und Kostendaten

Krankenhäuser werden administrativ der Gemeinde ihres Sitzes zugeordnet, diese Angabe ist Teil der ein Krankenhaus identifizierenden Merkmale¹¹⁵. Erfasst ist der 8-ziffrige Amtliche Gemeindeschlüssel (AGS), für dieses Projekt werden die ersten 5 Stellen als Identifikation von Bundesland und Kreis bereitgestellt. Kreisreformen bedeuten immer eine Veränderung des AGS mit zeitlich klarer Definition, welcher AGS zum Erhebungszeitpunkt der korrekte ist – alle Kodierungsbrüche sind terminiert. Trotzdem finden sich in den Daten der Krankenhausstatistik auch

¹¹⁴ Erste Analysen der Daten ließen weit mehr als die Hälfte aller Krankenhäuser als Neugründungen erscheinen, mit Häufungen zu bestimmten, bundeslandspezifischen Jahreswechseln. Namentlich Dr. Schoffer vom FDZ/L in Kamenz ist es zu verdanken, dass diese Auffälligkeiten in den kontextfreien Krankenhausnummern rückübersetzt und mit den sonstigen Krankenhaus-identifizierenden Merkmalen abgeglichen wurden. Die Datenqualität konnte damit entscheidend verbessert werden.

¹¹⁵ Eine gegebenenfalls vorhandene Außenstelle in einer anderen Gemeinde wird nicht separat identifiziert.

Jahre später noch längst obsoletere AGS-Angaben genauso wie Mischungen gültiger und ungültiger Kodierungen¹¹⁶.

Wenig überraschend erscheint deshalb, dass sich in den Diagnosedaten ähnlich fehlerhafte AGS-Angaben zum Wohnort häufen.

Deshalb hat das FDZ/L¹¹⁷ für dieses Projekt vorab eine jahresbezogene Liste aller verwendeten AGS-Codes bereitgestellt. Diese Liste ist die Grundlage des vom Autor für dieses Projekt entwickelten AGS-Rekodierungsprogramms, welches das FDZ/L in den Prozess der Datenbereitstellung eingebunden hat. Das AGS-Rekodierungsprogramm wird sowohl auf die Angaben zum Sitz des Krankenhauses in den Grunddaten als auch auf die Wohnortangaben der Diagnosedaten angewendet. Wegen diesem, den Anonymisierungsforderungen bei der Arbeit am GWAP und unter KDFV geschuldeten Vorgehen ist wohl die Anzahl der zu korrigierenden Codes, nicht jedoch die Anzahl der korrigierten Werte bekannt.

Mit dem AGS-Rekodierungsprogramm werden gleichzeitig die nötigen Umschlüsselungen zum Gebietsstand 2005 vorgenommen (Abschn. 4.2.4: Administrative Zuordnungen).

Die Grunddaten enthalten Angaben zu Merkmalen, die sich als Aggregationen aus anderen Merkmalen der Krankenhausstatistik verstehen lassen. Das ist nicht in jedem Fall korrekt. So sind die aggregierten Angaben zum Personal und einzelnen Personalgruppen als mittlere Zahlen über das Berichtsjahr oder als Vollzeitäquivalente zu verstehen, während die Einzelangaben den Istzustand an einem Stichtag oder die Anzahl aller beschäftigten Personen unabhängig der Frage der Teilzeitbeschäftigung betreffen. Gelegentlich lassen sich dabei Kombinationen feststellen, die nicht glaubhaft erscheinen¹¹⁸; diese Daten werden unverändert übernommen.

Diskrepanzen zwischen aggregierten Fallzahlen der Grunddaten und den Anzahlen berichteter Fälle können auch auf Veränderungen in der IT-Landschaft zurückzuführen sein, Systemwechsel im Berichtsjahr werden als eine Ursache für Minderberichtigungen angesehen (Schoffer, 2008). Vorliegendes Projekt stützt sich bei

¹¹⁶ Für Berlin allein finden sich über die Jahre mehr als 35 verschiedene Kodierungen, obwohl nur genau eine – 11000 – korrekt ist.

¹¹⁷ Wiederum ist Dr. Schoffer für sein Engagement ausdrücklich zu danken.

¹¹⁸ z. B. mehr Vollzeitäquivalente als Personen.

Fallzahlbetrachtungen ausschließlich auf die Angaben der Diagnosedaten, bei Geschlechtervergleichen von Personalgruppen auf die Personen unabhängig von der jeweils vertraglich vereinbarten Arbeitszeit.

Fehlwerte und Null-Einträge

Bundesland- (und damit hier auch StLA-) sowie Berichtsjahr-spezifische Daten-deskriptionen lassen ein stark unterschiedliches Verständnis vom Umgang mit Null- und Fehlwerten vermuten. So gibt es in einem Bundesland mehrere Krankenhäuser mit 0 Großgeräten eines bestimmten Typs und keines, bei dem die Angabe zu diesem Großgerät fehlt; ein anderes Bundesland hingegen hat nur Krankenhäusern ohne Angabe oder mit mindestens einem dieser Großgeräte. Solch ein Muster lässt sich bei verschiedenen Merkmalen finden, von Personenanzahlen bis zu Kostenangaben. Der Umgang mit Fehlwerten und Null-Einträgen scheint einer Bearbeiter-spezifischen Interpretation der Merkmalsdefinitionen der Krankenhausstatistik zu unterliegen.

Für dieses Projekt werden deshalb Null-Einträge wie Fehlwerte behandelt und damit offen gelassen, ob dieser Wert hätte vorhanden sein müssen. Für Daten-deskriptionen bedeutet das, dass die Merkmale als bedingt bezüglich positiver Werte zu verstehen sind. Summierende Aggregationen liefern vernünftige Werte, die Anzahlen zugrundeliegender Beobachtungen werden unterschätzt. Bezüge der Aggregate auf andere Aggregate (z. B. je Kreis die Anzahl der krankenhäuslichen Geburten je Einwohner) sind davon unbeeinflusst.

4.4. Erstellung des Registers, Datenmodifikationen

Wallgren und Wallgren (2007, Kap. 5.1) sehen drei Themenkomplexe bei und für die Strukturierung registerbasierter Analysen: (T.1) Man entwickle eine Struktur für das Register, die das Schaffen von neuen und besseren Statistiken ermöglicht. (T.2) Bei der Entwicklung des Registers biete man dieses in das existierende System (bereits existierender Register) ein. (T.3) Bereits vorhandene Register und Statistiken sollten genutzt werden. Themenkomplex (T.2) ist bei der externen

Nutzung amtlicher Daten durch die Wissenschaft¹¹⁹ nicht realisierbar. Für Themenkomplex (T.3) steht nicht das Ob, sondern das kontrollierte Wie im Mittelpunkt der Fragestellung.

Für die Erstellung und Auswertung des Registers sehen dieselben Autoren mit (P.1) Schaffung und Nutzung des Registers, (P.2) Qualitätssicherung und (P.3) Dokumentation drei parallel zu bedienende Prozesse als gleich bedeutend an, allerdings liegt für sie der Schwerpunkt der Dokumentation mehr auf der Nutzbarkeit des Registers als auf dem Verständnis und der Glaubhaftigkeit der gewonnenen Ergebnisse.

4.4.1. Struktur des Registers, Identifikatoren

Ziel des Registers sind panelbasierte Analysen der Krankenhäuser und der Kreise des Untersuchungsgebietes, sie sind die Merkmalsträger des Interesses, die über den Zeitraum 1992 bis 2005 in ihren jährlichen Veränderungen verfolgt werden sollen¹²⁰. Die wichtigsten Identifikatoren sind somit die (kontextfreie) Krankenhausnummer und die Kreiszuordnung – beide eineindeutig und unveränderlich über den Zeitraum – sowie die Angabe des Berichtsjahres. Die Zuordnung eines jeden Krankenhauses zu genau einem Kreis ist ebenfalls Voraussetzung für das Register.

Vermutete Inkonsistenzen in den Krankenhausnummern und die möglichen Auswirkungen auf die Panelanalysen sind bereits mit Tab. 4.2 diskutiert; es wird im Folgenden vom Status quo ausgegangen.

Mit Ausnahme der Arbeitslosenzahlen und -quoten sowie der siedlungsstrukturellen Gebietstypen des BBR (Abschn. 3.3.5) sind alle hier einbezogenen Statistiken jahresbezogen. Um den Strukturerefordernissen des Registers zu genügen, werden der Regionsgrundtyp und der spezifischere Kreistyp jeweils allen Jahren zugeordnet; die Arbeitslosenzahlen und -quoten müssen zu Jahreswerten aggregiert werden.

¹¹⁹ und somit für dieses Projekt.

¹²⁰ Eine Verfolgung einzelner Unterstrukturen (z. B. Fachabteilungen) der Krankenhäuser wird nicht angestrebt.

Hinsichtlich der administrativen Gliederung des Untersuchungsgebietes in Kreise wird der Gebietsstand am 31. 12. 2005 zugrunde gelegt (Abschn. 4.2.4: Administrative Zuordnungen). Dazu ist es nötig, die administrativen Gliederungen der Bundesländer Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen¹²¹ vor und zwischen den Kreisreformen auf den Gebietsstand 2005 abzubilden.

Zuordnungen auf den Gebietsstand 2005

Bei den Kreisreformen der ostdeutschen Flächenländer handelt es sich zum überwiegenden Anteil um Zusammenlegungen oder Eingliederungen von früheren Kreisen zu neuen, größeren Einheiten. Dabei werden vereinzelt Gemeinden aus ihrem bisherigen Kreis herausgelöst und einem anderen angrenzenden Kreis zugeordnet – diese Gemeindeverschiebungen werden hier sowohl bezüglich der Krankenhauszuordnung als auch bezüglich der Wirtschaftsparameter der Kreise als nicht relevant angesehen.

Die Kreise Chemnitz-Land (AGS₁₉₉₂ 14018) und Dresden-Land (AGS₁₉₉₂ 14022) hingegen werden jeweils auf mehrere Kreise aufgeteilt, dabei werden die Hauptteile den Kreisen Chemnitzer Land (AGS₂₀₀₅ 14173) bzw. Kamenz (AGS₂₀₀₅ 14192) zugeordnet. Analog wird hier bei der Registererstellung verfahren. Wie groß sowohl der Anteil an Wirtschafts- als auch an Krankenhausleistung jener Altkreise ist, der inkorrekt zugeordnet wird, kann anhand vorliegender Daten nicht abgeschätzt werden und muss hier offen bleiben.

Mit dem Beginn der Jahre 1996 bzw. 1998 werden die Stadtkreise Hoyerswerda (Sachsen) und Eisenach (Thüringen) aus den umliegenden Kreisen herausgelöst, Daten der Krankenhaus- und aller anderen Statistiken liegen für die Jahre davor für diese Kreise nicht vor. Für dieses Projekt wird entschieden, die Stadtkreise Hoyerswerda und Eisenach als Teil der umgebenden Kreise Kamenz bzw. Wartburgkreis zu betrachten¹²². Dies ist möglich, da der Regionsgrundtyp und der spezifischere Kreistyp der jeweiligen Paare gleich ist; alle anderen Statistiken sind

¹²¹ In Rheinland-Pfalz gibt es im Untersuchungszeitraum keine Kreisreform, zwei Kreise werden umbenannt (01. 01. 1997 Landkreis Pirmasens in Landkreis Südwestpfalz, 01. 01. 2004 Landkreis Ludwigshafen in Rhein-Pfalz-Kreis); der Stadtstaat Berlin gliedert sich nicht in Kreise.

¹²² Diese Entscheidung gilt nur für das Register und die Analysen unter KDFV.

geeignet zu aggregieren. Die projektintern vereinigten Kreise bekommen eine im AGS (bisher) nicht verwendete Pseudo-AGS-Nummer zugeordnet (Hoyerswerda und Landkreis Kamenz im Regierungsbezirk Dresden: 14299; Eisenach und Wartburgkreis in Thüringen: 16099).

Die nötigen Umschlüsselungen der Krankenhausdaten zum modifizierten Gebietsstand 2005 werden bereits bei der Datenbereitstellung durch das FDZ/L mit dem AGS-Rekodierungsprogramm (Abschn. 4.3.3: Grund- und Kostendaten) vorgenommen; die extern vom Autor zusammengestellten Daten im offenen Zugriff (Abschn. 4.3.2) entsprechen dem modifizierten Gebietsstand 2005 von vornherein.

4.4.2. Aggregationen

Unterjährige Angaben und Angaben, die sich auf Substrukturen von Krankenhäusern oder Kreisen (die Subjekte der jeweiligen Panelanalysen, Abschn. 3.4.1) beziehen, sind gegebenenfalls vor der Übernahme in das Register zu aggregieren.

Zeitliche Aggregationen

Arbeitslosenzahlen und -quoten werden von der Bundesagentur für Arbeit für unterjährige Berichtszeiträume (je nach Berichtsjahr monats- oder quartalsweise) bereitgestellt. Diese Daten werden bei Übernahme in das Register zu Arbeitslosenquoten für jeweils ein Jahr aggregiert, indem das arithmetische Mittel der Arbeitslosenzahlen eines Kreises im jeweiligen Jahr durch die Zahl der Erwerbspersonen dividiert wird¹²³. Die Anzahl der Erwerbspersonen – der Nenner – wird auch von der Bundesagentur für Arbeit nur einmal jährlich für jeden Kreis aktualisiert¹²⁴.

¹²³ Für Jahre mit Quartals- neben Monatswerten wird ein gewichtetes Mittel angewendet.

¹²⁴ <http://www.pub.arbeitsamt.de/hst/services/statistik/000000/html/start/arbeitslosenquote.shtml>, Zugriff am 09. 05. 2007.

Aggregationen in der Subjektkomponente der Panel

Die Daten der Krankenhausstatistik zu Satzart 2 bis 4 beziehen sich auf Substrukturen des Krankenhauses (Fachabteilungen, Ärzte, nichtärztliches Personal), müssen aber jeweils die entsprechenden Merkmalswerte für das Gesamtkrankenhaus ebenfalls enthalten, auf eine Aggregation kann hier verzichtet werden¹²⁵. Die Diagnosedaten werden ohne Aggregation dem Register zugeführt.

Für kreisbezogene Daten des freien Datenzugriffs ist die Entscheidung bezüglich der Stadtkreise Hoyerswerda (Sachsen) und Eisenach (Thüringen) umzusetzen. Jedem Merkmal wird eine Aggregationsregel zugeordnet; neben einfachen Summen (z. B. der Einwohnerzahlen) sind gewichtete Mittel (z. B. beim BIP je Einwohner) die am häufigsten verwendeten.

4.4.3. Datenmodifikationen

Neben den bereits im Datenbestand des FDZ/L vor Übergabe an das Projekt erfolgten Modifikationen auf den Gebietsstand 2005 und die Diagnoseverschlüsselung nach ICD-9 müssen weitere Vereinheitlichungen in den Daten vorgenommen werden.

Kodierungsbrüche, Fehlwerte und Null-Einträge

Veränderungen der verwendeten Schlüssel für bestimmte Merkmale (Abschn. 4.2.4: Kodierungsbrüche) zwischen den Berichtsjahren der Statistiken werden bei der Registererstellung zu einem für den Untersuchungszeitraum einheitlichen neuen Schlüssel verändert, so dass das Maximum an Information erhalten werden kann. Für das Merkmal *Träger des Krankenhauses* z. B. wurde 1 = öffentlich, 4 = freigemeinnützig, 5 = privat gewählt.

¹²⁵ Eine Konsistenzprüfung wurde nicht vorgenommen.

Gleichzeitig mit der Registererstellung werden die Daten nach den in Abschn. 4.3 vorgestellten Regeln für den Umgang mit Fehlwerten und Null-Einträgen vereinheitlicht.

Deflationierung der Kostendaten

In Geldeinheiten ausgedrückte Merkmale zu messen bedeutet immer, dass sich neben den Veränderungen des Merkmals selbst auch die Inflation (bzw. Deflation) in diesen Werten ausdrückt. Um den Effekt der Inflation bereinigte Daten erhält man mittels Deflationierung, der Division des erhobenen Merkmalswertes mit einem relevanten Preisindex. Der Wahl des Preisindex ist von Bedeutung, branchenspezifische Preisindizes können branchenspezifische Kostendaten besser adjustieren als allgemeine.

Da die Kostendaten der Krankenhäuser eine Zusammenfassung sehr unterschiedlicher Branchen und Themen (von Bauwesen über Pharmaindustrie bis Personal) sind, wird hier mit dem Verbraucherpreisindex (VPI) ein gut verfügbarer¹²⁶ und relativ breit basierter Index gewählt: alle Kostendaten werden auf das Jahr 2000 (= 100 %) deflationiert. Gleichfalls deflationiert werden die auf Geldeinheiten basierenden Daten der anderen Datenquellen. Die entstehenden VPI-korrigierten Angaben gehen in alle weitere Berechnungen ein.

Abgeleitete Größen

Abgeleitete Merkmale werden durch arithmetische Operationen aus anderen Merkmalen gewonnen, wie z. B. die Einwohnerdichte einer administrativen Einheit als Quotient aus der Einwohnerzahl und der Fläche gebildet werden kann. Das Halten von abgeleiteten Merkmale in einer Datenbank parallel zu ihren Grundmerkmalen ist allgemein als kritisch zu sehen, denn jede Korrektur eines der Grundmerkmale erfordert die Neuberechnung aller daraus abgeleiteten. Allgemein qualitätssicherer ist die Generierung von abgeleiteten Merkmale jeweils im ersten Analyseschritt.

¹²⁶ <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/abisz/VPI.psm1>, Zugriff am 21. 02. 2008.

Da für vorliegendes Projekt (sowohl am GWAP als auch unter KDFV-Bedingungen) Datenkorrekturen mit der Übergabe der Daten durch das FDZ/L und der Erstellung des Registers abgeschlossen sind, können abgeleitete Merkmale ohne erhöhtes Qualitätsrisiko in das Register aufgenommen werden, wiederholte, zeitintensive Berechnungen unmittelbar vor Auswertung können unterbleiben.

Das Binärmerkmal (*Kreis des Krankenhauses = Wohnkreis?*) wird aus dem Wohnort und dem Ort des Sitzes des Krankenhauses abgeleitet, der Anonymisierungsforderungen wegen geschieht dieses, wie bereits beschrieben, bereits vor der Datenübergabe durch das FDZ/L.

Die Erwerbstätigenquote, der Anteil der Erwerbstätigen an der Wohnbevölkerung, lässt sich nur näherungsweise bestimmen, denn die vorliegenden Angaben zu den Erwerbstätigen entsprechen dem Inlandskonzept (Erfassung am Arbeitsort), die zur Wohnbevölkerung (Einwohnerzahlen) dem Inländerkonzept (Abschn. 3.1.5).¹²⁷

Das von Werblow und Schoffer (2010) vorgeschlagene Personal-Sachkosten-Verhältnis (Abschn. 4.2.4: Kostendaten) wird ebenfalls bei der Registererstellung für jedes Krankenhaus und Berichtsjahr bestimmt.

Diagnose-Gruppierungen

Auswertungen aller Fälle der Krankenhausstatistik gemeinsam geben einen Gesamtüberblick ohne Beachtung von medizinisch-epidemiologische Zusammenhängen. Feinere Analysen sollen sich auf Gruppen zueinander ähnlicher Krankheiten beziehen. Die 17 Diagnose-Kapitel der ICD-9 (bzw. die 21 der ICD-10) sind dabei nur teilweise hilfreich: ‚große‘ Kapitel wie das Kapitel XVII *Verletzungen und*

¹²⁷ Der als Näherungsvariable (Proxy) dienende Quotient aus Erwerbstätigen am Arbeitsort und Einwohnern am Wohnort wird allgemein als Brutto-Erwerbstätigenquote bezeichnet.

Vergiftungen umfassen immer noch sehr unterschiedliche Krankheitsbilder¹²⁸ decken sehr unterschiedliche Krankheitsbilder ab.

Deshalb werden folgende Diagnose-Gruppierungen¹²⁹ anhand ihrer ICD-9-Schlüssel in ihrem Bezug zu den Kapiteln definiert und ausgewählt:

- 001–009: Infektiöse Krankheiten des Verdauungssystems (aus Kapitel I)
- 020–027: Zoonosen durch Bakterien (aus I)
- 090–099: Syphilis (Lues) und sonstige Geschlechtskrankheiten (aus I)
- 140–208: Bösartige Neubildungen (aus II)
- 240–279: Endokrinopathien, Ernährungs- und Stoffwechselkrankheiten sowie Störungen im Immunsystem (III komplett)
- 280–289: Krankheiten des Blutes und der blutbildenden Organe (IV komplett)
- 320–326: Entzündliche Krankheiten des Zentralnervensystems (aus VI)
- 360–379: Affektionen des Auges und seiner Anhangsgebilde (aus VI $\hat{=}$ ICD-10 VII komplett)
- 380–389: Krankheiten des Ohres und des Warzenfortsatzes (aus VI $\hat{=}$ ICD-10 VIII komplett)
- 390–459: Krankheiten des Kreislaufsystems (VII komplett)
- 460–519: Krankheiten der Atmungsorgane (VIII komplett)
- 520–543: Krankheiten der Organe des Verdauungstraktes (aus IX)
- 580–589: Nephritis, Nephrotisches Syndrom und Nephrose (aus X)
- 650–669: Normale Entbindung, Komplikationen im Verlauf der Wehen und der Entbindung (aus XI)
- 680–709: Krankheiten der Haut und des Unterhautzellgewebes (XII komplett)
- 800–829: Frakturen des Schädels, der Wirbelsäule, des Rumpfskeletts, der oberen und unteren Extremitäten (aus XVII)
- 870–897: Offene Wunden an Kopf, Hals, Rumpf, oberen und unteren Extremitäten (aus XVII)

¹²⁸ z. B. Kapitel XVII: Frakturen, Luxationen, Verstauchungen, innere Verletzungen, offene Verletzungen, Verbrennungen, Nerven- und Rückenmarksverletzungen sowie Vergiftungen durch Drogen, Medikamente oder nichtgebräuchlichen medizinischen Substanzen. Komplikationen nach chirurgischen Eingriffen und ärztlicher Behandlung werden ebenfalls diesem Kapitel zugerechnet.

¹²⁹ Der Begriff Diagnose-Gruppe ist bereits sowohl in ICD-9 als auch ICD-10 für eine noch feinere Unterteilung der Diagnose-Kapitel verwendet.

Damit sind 456 von 961 (47,5 %) der Diagnoseschlüssel der 17 Kapitel der ICD-9 abgedeckt.

4.4.4. Das Register unter GWAP-Bedingungen

All die beschriebenen Schritte zur Erstellung des Registers sind am GWAP entwickelt und getestet. Besonderheiten dabei ergeben sich u. a. daraus, dass die Information zur administrativen Zuordnung der Krankenhäuser nicht kreisgenau erfolgen darf (Dreier-Regel, Abschn. 4.1.2). Die geforderten Kreiszusammenfassungen erfordern zusätzliche Aggregationen nach den beschriebenen Regeln. Da die Merkmale der siedlungsstrukturellen Gebietstypen (Abschn. 3.3.5) als nominal skalierte Merkmale über Kreise hinweg nicht summierend aggregierbar sind, wird aus allen in den Daten auftretenden Kombinationen jeweils ein neuer Schlüssel geschaffen.

4.5. Daten und Register – Zusammenfassung

Amtliche Daten zur Gesundheit sind überwiegend Daten zu Krankheit, Krankheitsbekämpfung und -prävention. In Deutschland ist die Krankenhausstatistik dabei die umfangreichste und bedeutendste Amtliche Statistik, eine ähnliche Statistik zur Versorgung in den Arztpraxen, Gemeinschaftspraxen und Medizinischen Versorgungszentren existiert nicht. Mit amtlichen Daten lassen sich somit die Behandlungen der schweren und schwersten Krankheiten in Krankenhäusern, Rehabilitations- und Vorsorgeeinrichtungen beschreiben, dabei findet nur für die Krankenhäuser eine Totalerhebung mit einheitlichem Merkmalsumfang und Berichtspflicht für alle Krankenhäuser mit äußerst geringen gesetzlichen Ausnahmen statt.

Für die Wissenschaft können die Daten der Krankenhausstatistik durch das FDZ/L bereitgestellt werden; hier erfolgt der vertraglich definierte Zugriff unter den Bedingungen von GWAP und KDFV. Die Daten werden um amtliche Wirtschaftsdaten aus freiem Zugriff ergänzt.

Amtliche Daten gelten als geprüft, dieses ist dokumentiert. Erfahrungen mit den Daten sprechen jedoch dafür, dass die Interpretation einzelner Merkmale in den Bundesländern als den jeweiligen Dateneignern unterschiedlich ist. Aussagen zur Konsistenz der Daten über Berichtsjahre hinweg finden sich nur spärlich in den Qualitätsdokumenten zur Krankenhausstatistik. Vorliegende Daten lassen vermuten, dass sich die Identifikation ein und desselben Krankenhauses zwischen Berichtsjahren ändert, so dass eine eindeutige longitudinale Zuordnung der Krankenhäuser nicht gesichert ist. Mit diesen Erkenntnissen muss man schlussfolgern, dass sich die Daten nicht unmittelbar für eine bundeslandübergreifende und longitudinale Auswertung eignen, eine Revalidierung vor Auswertung im Rahmen eines wissenschaftlichen Projekts wäre zu empfehlen.

Veränderungen im Schlüsselsystem für bestimmte Merkmale führen häufig dazu, dass die ältere, offiziell abgelöste Version eines Schlüssels auch in mehreren Folgejahren noch Verwendung findet. Bei Schlüsseln mit vollständig getrennten Wertebereichen (z. B. ICD-9 und ICD-10) wäre eine algorithmenbasierte Umschlüsselung im Rahmen der Qualitätsverbesserung vorstellbar und wünschenswert; bei überlappenden Wertebereichen (z. B. beim Merkmal *Träger des Krankenhauses*) muss der Einfluss der Schlüsselveränderung offen bleiben. Das Problem potenziert sich mit der schnellen Abfolge von Kreisreformen und AGS-Modifikationen (besonders in Sachsen) in den 90er Jahren.

Jahres- und bundeslandübergreifende Revalidierungen könnten nur im Rahmen des StBA oder des Verbundes des StLA erfolgen; für einen Wissenschaftler mit begrenzten Zugriffsrechten im Rahmen von GWAP oder KDFV sind diese wegen der Anonymisierungsforderungen nicht hinreichend durchführbar. Für dieses Projekt sind dennoch viele Anstrengungen zur Verbesserung der Datenqualität unternommen worden; der von de Veaux und Hand (2005) für Datenbereinigung als nötig erachtete Anteil von 60 bis 95 % der Projektzeit scheint zumindest mit der Untergrenze des Intervalls zunehmend realistisch.

Aus den Daten der unterschiedlichen Quellen wird ein Register erstellt, dieses enthält die Daten der Berichtsjahre 1992 bis 2005 zu Rheinland-Pfalz, den ostdeutschen Flächenländern und Berlin. Das Register ist die Grundlage aller Analysen.

Folgende Einschränkungen gelten für das Register:

- ▷ Es muss vermutet werden, dass einzelne Krankenhäuser fehlerhafterweise als Schließungen bzw. Neugründungen in den Daten erscheinen, obwohl es sich um ein und dasselbe Krankenhaus handelt.
- ▷ Für Merkmale mit nichtunterscheidbaren Null- und Fehl-Werten werden diese auf Fehl-Wert gesetzt.
- ▷ Alle Kosten sind nach VPI auf das Jahr 2000 hin deflationiert, dennoch gelten a priori nur die bereinigten Kosten als über den Gesamtzeitraum vergleichbar.
- ▷ Diagnosen, die nach ICD-10 kodiert in den Originaldaten vorliegen, sind auf ICD-9 umgeschlüsselt.
- ▷ Alle Daten sind (meist durch geeignete Aggregation) dem modifizierten Gebietsstand 31. 12. 2005 angepasst¹³⁰.
- ▷ Die Wirtschaftsparemeter der Kreise liegen erst ab 1996 vollständig vor.
- ▷ Die Arbeitslosenquote für 2005 entspricht der geänderten Definition der Arbeitslosigkeit und ist nicht unmittelbar mit der früherer Jahre vergleichbar.

Das Register beginnt mit dem Jahr 1992, es ist ab dem Jahr 1996 merkmalsvollständig. Eine Fortsetzung des Registers über das Jahr 2005 hinaus wäre problematisch, denn die mit dem Jahr 2005 geänderte Definition der Arbeitslosigkeit stellt bereits einen nichtkompensierbaren Bruch in den Daten dar. Das Jahr 2007 bringt die nächste Kreisreform in einem Bundesland (Sachsen-Anhalt), einige amtliche Statistiken mit längerer Bearbeitungszeit sind bereits für das Erhebungsjahr 2005 primär in Kreisstrukturen von Mitte 2007 verfügbar. Mit dem Jahr 2006 ändert das Bundesamt für Bauwesen und Raumordnung seine Zuordnung eines Regionsgrundtyps für Trier und Umgebung in Rheinland-Pfalz, folglich ändern sich auch alle Kreistypen in jenem Gebiet.

Lässt man die fraglichen Merkmale weg oder findet Lösungen zur Anpassung der Daten über erwähnten Brüche hinweg, so lässt sich mit den vorgestellten Techniken ein modifiziertes bzw. neues Register sowohl über einen größeren Untersuchungszeitraum als auch über alle Bundesländer erstellen.

¹³⁰ Dieser ist zwischen 1996 und 2005 als unverändert anzusehen, der Kreis-Wechsel einzelner Gemeinden in diesem Zeitraum wird nicht abgebildet.

Kapitel 5.

Empirische Analysen: Methoden und Umsetzung

ALL MODELS ARE WRONG BUT SOME ARE USEFUL.
Box (1979, S. 202)

Ökonometrische Modelle sind stochastische Modelle mit einem systematisch-funktionalen Teil und Störtermen. Für jedes zur Beantwortung einer ökonomisch-substanzwissenschaftlichen Frage gewählte ökonometrische Modell ist zu prüfen, ob sich die Parameter des spezifizierten Modells aus den Daten eindeutig bestimmen lassen; ist diese Frage zu verneinen, so kann das oft nur entweder durch das Heranziehen weiterer Daten oder durch Wahl eines neuen – meist sparsameren – Modells geändert werden.

5.1. Statistische Methoden und deren Quellen

Typischerweise bilden ökonometrische Modelle eine (oder mehrere) endogene Variable als bekannte Funktion von erklärenden Größen ab. Diese Abbildung wird um Annahmen ergänzt, z. B. zur Verteilung über die Störterme in der Regressionsbeziehung. Die unbekannt Parameter des Modells werden geschätzt, häufig mittels Methoden der Kleinsten Quadrate oder des Maximum-Likelihood. Falsche Parametrisierungen des systematischen Teils des ökonometrischen Modells können genauso zu inkonsistenten Schätzfunktionen führen wie fehlerhafte Vertei-

lungsannahmen über die stochastische Restgröße. Eine wichtige Frage aller statistischen Modellierungen ist die nach der statistischen Einheit.

5.1.1. Aggregationen

Aggregation, das Zusammenfassen von Einzelmessungen zu einem gemeinsamen Wert (Rönz und Strohe, 1994, S. 7), bedeutet immer eine Verringerung der Komplexität und meist auch der Variabilität; Unterschiede zwischen den Einzelmessungen sind nicht mehr analysierbar. Das gilt für kumulative Aggregationen (Summenbildung) genauso wie für (gewichtete) Mittel oder jegliche andere zusammenfassende Statistik. Aggregationen können sowohl über mehrere Einzelmessungen ein und derselben statistischen Einheit erfolgen, wie beim Generieren eines Jahreswertes aus monatlichen Angaben¹³¹, oder Zusammenfassungen mehrerer statistischer Einheiten zu einer neuen sein, so wie es bei der Gewinnung von Angaben zum Kreis aus denen aller Gemeinden des Kreises erfolgt. Mischformen beider sind ebenfalls üblich.

5.1.2. Kreise oder Krankenhäuser – zur statistischen Einheit der Analysen

Wichtigste Beobachtungseinheit der Krankenhausstatistik ist jährlich jedes einzelne Krankenhaus. Zu jedem Krankenhaus werden jahresweise Daten auf unterschiedlichen Ebenen¹³² erhoben: (a) je Krankenhaus selbst, z. B. bereinigte Kosten, (b) je Fachabteilung, z. B. Anzahl der Betten in der Augenheilkunde, (c) je Spezialisierungsrichtungen (Gebiets- / Teilgebietsbezeichnung), z. B. Anzahl der Kardiologinnen, (d) je Personalgruppe / Berufsbezeichnung, z. B. Anzahl der männlichen Zytologieassistenten, (e) je Fall, z. B. Hauptdiagnose. Die Ebenen (b)–(d) sind dabei unmittelbar unterhalb der Krankenhausebene (a) angesiedelt; die Ebene der Fälle (e) kann als unterhalb der Fachabteilungsebene (b) betrachtet werden, denn mit jedem Fall ist die Fachabteilung mit der längsten Verweildauer zu erfassen.

¹³¹ hier z. B. die Arbeitslosenzahlen und -quoten für jedes Berichtsjahr generiert aus den monatlichen oder Quartalsangaben der Bundesagentur für Arbeit.

¹³² Die Ebenen entsprechen den Satzarten 1 und 6, 2 und 5, 3, 4 bzw. 7 der Erfassungsstruktur der Krankenhausstatistik (Abschn. 4.2.1).

Für eine vergleichende Analyse der Krankenhäuser ist jedes Krankenhaus – mit den wiederholten Beobachtungen über die Zeit auf den verschiedenen Beobachtungsebenen – Einheit der statistischen und ökonometrischen Analysen. Ob man für jedes Krankenhaus des Untersuchungsgebietes annimmt, es könne dabei völlig unabhängig von allen anderen Krankenhäusern agieren, hat Einfluss auf die Wahl eines passenden ökonometrischen Modells. Der Einfluss der wirtschafts- und gesundheitspolitischen Entscheidungen der Kreistage (sowie die übergeordnete Krankenhausplanung der Länder) lassen die Annahme einer größerer Abhängigkeit der Krankenhäuser innerhalb eines Kreises gegenüber Krankenhäusern unterschiedlicher Kreise realistisch erscheinen.

Für die Analyse der Beziehungen des Gesundheitssystems mit dem wirtschaftlichen Umfeld ist das einzelne Krankenhaus hingegen nicht die geeignete statistische Einheit, da einem Krankenhaus nicht ein-eindeutig ein Wirtschaftsraum zugeordnet werden kann. Hier bietet es sich an, jedem Wirtschaftsraum alle einliegenden Krankenhäuser zuzuordnen. Betrachtet man dann die einzelnen Wirtschaftsräume als Einheit der statistischen und ökonometrischen Analysen, so können die Krankenhäuser eines Wirtschaftsraumes im Sinne des Mehrebenenmodells (Abschn. 3.1) als geklumpt (engl: *clustered*, z. B. Palta und Seplaki, 2002) betrachtet werden. Die Annahme dabei ist: Krankenhäuser eines Wirtschaftsraumes unterscheiden sich von denen eines anderen Wirtschaftsraumes auch dadurch, dass die Unterschiede der Wirtschaftsräume zu Unterschieden bei den Krankenhäusern führen (können), wirtschaftliche Veränderungen des Wirtschaftsraumes haben Einfluss auf die Krankenhäuser. Die wirtschafts- und gesundheitspolitischen Entscheidungen der Kreistage definieren für den jeweiligen Kreis das gemeinsame Umfeld.

Gesamtkrankenhausaktivität

Alternativ zum Modell aller einzelnen einliegenden Krankenhäuser lässt sich aus der Aggregation der Krankenhäuser die Gesamtkrankenhausaktivität des jeweiligen Wirtschaftsraumes bestimmen, das Modell vereinfacht sich durch Verminderung um eine Ebene. Die Gesamtkrankenhausaktivität beschreibt die Aufwände, Ressourcen und Leistungen aller Krankenhäuser des Wirtschaftsraumes gemeinsam; die Krankenhäuser stehen den Menschen des Wirtschaftsraumes (In-

ländern, Einpendlern und Gästen gleichermaßen) zur Verfügung. Notwendig dafür ist die räumliche Aggregation im Sinne von Rönz und Strohe (1994, S. 7). Die Merkmalswerte der einzelnen Krankenhäuser eines Kreises werden durch einfache Summation (jegliche Zähl- und Kostendaten) und darauffolgende Neuberechnung von Verhältniszahlen (z. B. Anzahl der Fälle je Einwohner) gebildet. Nominale Angaben wie die Trägerschaft oder der Typ des Krankenhauses entziehen sich der Aggregation durch Summation.

Die Kreise des Untersuchungsgebietes sind die Wirtschaftsräume, die in ihrer Entwicklung ebenso durch die Entscheidungen der lokalen Autoritäten (Kreistage) beeinflusst werden wie das Gesundheitssystem und somit die Krankenhäuser¹³³. Als solche sind die Kreise wichtigste statistische Einheit der folgenden Analysen.

Modifikationen für den GWAP: Klumpen von Kreisen

Für die Arbeit am Gastwissenschaftler-Arbeitsplatz gelten die Regeln zur Sicherung der projektbezogenen faktischen Datenanonymität (Abschn. 4.1.2), Kreise dürfen nicht einzeln betrachtet sondern müssen gemäß Dreier-Regel zu Gruppen von mindestens 3 Kreisen zusammengefasst werden. Diese Gruppen von mindestens 3 Kreisen werden im Folgenden *Klumpen* genannt; Berlin als Bundesland ohne Kreis- oder Gemeinde-Substruktur und mit einer ausreichend großen Anzahl an Krankenhäusern darf ein eigener Klumpen sein. Eingedenk des Teilzieles, zwischen den Altbundesländern – hier repräsentiert durch Rheinland-Pfalz – und den ostdeutschen Bundesländern zu vergleichen, sollen aus den 149 Kreisen des Untersuchungsgebietes¹³⁴ separat Klumpen ost- und westdeutscher Kreise entstehen.

5.1.3. Klumpenbildung: klein aber ähnlich

Die üblichsten Verfahren zur Einteilung unbekannter Objekte so, dass die jeweils ähnlichsten Objekte in einer Klasse zusammengefasst werden und sich die Klas-

¹³³ Krankenhausbedarfsplanung ist in Deutschland Ländersache, ein Kreis muss für seine Interessen teilweise im Landesrahmen eintreten.

¹³⁴ Hoyerswerda und Eisenach sind hier nicht mit den umgebenden Kreisen vereinigt.

sen voneinander möglichst gut unterschieden, sind Clusteranalysen (Backhaus et al., 2003; Bortz, 1999, u. a.). Die Klassen werden Cluster genannt. Man unterscheidet zwei Hauptarten der Clusteranalyse. Das hierarchische Clustern fasst – aufsteigend – sequenziell jeweils die zwei ähnlichsten Objekte zusammen oder teilt – absteigend – mit dem einen Cluster aller Objekte beginnend jeweils so, dass aus einem bestehenden Cluster zwei neue maximal voneinander verschiedene Cluster entstehen. Das partitionierende Cluster hingegen versucht, eine Eingangseinteilung in eine gewünschte Anzahl von Clustern durch Verschieben einzelner Objekte zwischen den Clustern zu verbessern.

Ähnlichkeit

Wichtigste Voraussetzung für das Clustern ist die Definition eines Abstands- oder eines Ähnlichkeitsmaßes, das für jedes Paar von Objekten und bereits vorliegenden Clustern sagen kann, wie unähnlich / verschieden bzw. ähnlich sich diese sind. Das Abstandsmaß muss dabei den Skalenniveaus (nominal, ordinal, metrisch) der relevanten Merkmale entsprechend gewählt werden, verschiedene Maße führen zu unterschiedlichen Cluster-Ergebnissen.

Für die GWAP-basierten Vor-Analysen der Gesamtkrankenhausaktivität wird angestrebt, wirtschaftlich über die Jahre ähnlichen Kreisen des Untersuchungsgebietes einen geringen Abstand zuzuordnen. Zur Beschreibung der wirtschaftlichen und sozialen Situation der Kreise werden als Merkmale herangezogen: (a) die Einwohnerdichte 1994 und 2004, (b) die Erwerbstätigenquote 1994 und 2004, (c) das Bruttoinlandsprodukt je Einwohner 1992–2004, (d) das Arbeitnehmerereinkommen je Arbeitnehmer 1996–2004, (e) das verfügbare Haushaltseinkommen je Einwohner 1996–2004 und (f) die vierteljährlichen Arbeitslosenquoten 1996–2004¹³⁵. Die gewählten Merkmale mit Jahresangaben werden je Jahr als metrisch skalierte Variable betrachtet – die Arbeitslosenquoten quartalsweise –, so dass ein Abstandsmaß über 71 Dimensionen zu definieren ist. Jede der Dimensionen wird $N(0, 1)$ -standardisiert, die Unterschiede der Messskalen werden somit nivelliert. Jeder Kreis repräsentiert sich als Punkt durch seine Koordinaten im 71-dimensionalen Euklidischen Raum. Als Abstandsmaß sei die Summe der Abweichungsquadrate vom Mittelpunkt zweier Kreise gewählt, wobei – we-

¹³⁵ Zum Zeitpunkt dieser Analysen lagen die Daten für 2005 u. a. noch nicht vor.

gen der größeren Informationsdichte durch die quartalsweise Erfassung – die Dimensionen der Arbeitslosenquoten nur mit einem Gewicht von $\frac{1}{4}$ in die Summe eingehen. Zwei Kreise haben den minimalen Abstand von 0 genau dann, wenn sie in allen Dimensionen und somit Merkmalswerten exakt übereinstimmen.

Clustern

Für beide Clusterverfahrensarten – hierarchisch oder partitionierend – gilt, dass die Anzahl der Cluster relativ gut vorgegeben werden kann, die Anzahl der Mitglieder eines jeden Clusters hingegen unausweichliches Ergebnis des Prozesses ist. Unter Anwendung von Verfahren wie Single Linkage, Complete Linkage, Average Linkage, Centroid oder Ward (hierarchisch) bzw. Austausch- oder Iteriertem Minimaldistanzverfahren (partitionierend) entstehen häufig Cluster mit sehr unterschiedlich großen Mitgliederzahlen. 149 Kreise so zu clustern, dass das kleinste Cluster mindestens 3 Mitglieder besitzt führt in einem sehr günstigen Fall dazu, dass 14 Cluster entstehen, deren größtes 22 Mitglieder besitzt. Eine mittlere Clustergröße von fast 11 Mitgliedern erscheint für die Analyse am GWAP doch sehr fern von den 147 Kreisen, auf denen die Analysen unter KDFV durchgeführt werden.

Partitionierung zu vorgegebenen Klumpengrößen

Wünschenswert erscheint eine Partitionierung so, dass den Forderungen der Dreier-Regel genüge getan wird und gleichzeitig nur die geringstmögliche Aggregation der Ausgangsdaten erforderlich ist. Die 36 Kreise in Rheinland-Pfalz sollen zu 12 Klumpen, die 112 ostdeutschen¹³⁶ zu 37 Klumpen zusammengefasst werden¹³⁷. Mit Berlin ergeben sich dann 50 Klumpen.

Dafür wird folgendes Verfahren vorgeschlagen:

Die Kreise in Rheinland-Pfalz und getrennt davon die Kreise Ostdeutschlands werden zufällig auf die beschriebenen 12 bzw. 37 Klumpen verteilt. Diese Zufalls-

¹³⁶ ohne Berlin.

¹³⁷ Einer der ostdeutschen Klumpen besteht somit aus 4 Kreisen.

zuteilung wird sehr häufig wiederholt, die zugehörigen Konfigurationen werden festgehalten. Für jede dieser Konfigurationen wird das Abstandsmaß berechnet. Gewählt wird dann jeweils die Konfiguration mit dem minimalen Wert des Abstandsmaßes.

Für diese Ähnlichkeitspartitionierung erforderte die Ermittlung des absoluten Minimums der Werte des Abstandsmaßes die vollständige Enumeration aller Zuteilungsmöglichkeiten der Kreise in die Klumpen. Für Rheinland-Pfalz bedeutete dies, $3,57 \times 10^{23}$ Möglichkeiten¹³⁸ durchzurechnen, für Ostdeutschland erhöhte sich die Anzahl auf $2,14 \times 10^{111}$. Der Rechenaufwand dürfte das Ergebnis kaum rechtfertigen.

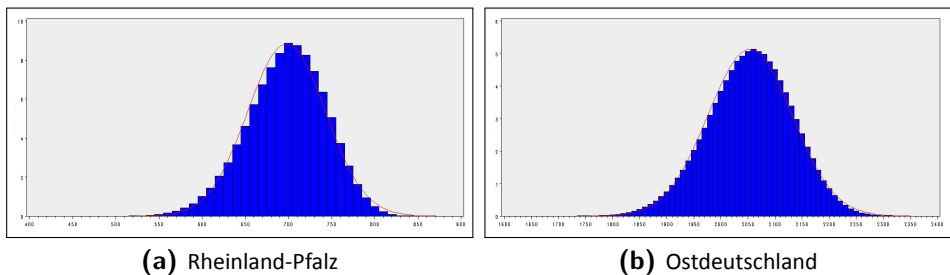


Abb. 5.1.: Abstandsmaßwerte der Klumpenbildung (siehe Text)

Die Entscheidung beruht hier auf 384.065 Zufallszuteilungen für Rheinland-Pfalz mit einem Minimum von 464,9 (arithmetisches Mittel = 697,9; Maximum = 865,8) und auf 1.763.938 Zufallszuteilungen für Ostdeutschland mit dem Minimum 1641,5 (arithmetisches Mittel = 2054,0; Maximum = 2393,1). Die Häufigkeitsverteilungen der Abstandsmaßwerte zeigt Abb. 5.1, beide sind leicht linksschief und rechtssteil. Es wird davon ausgegangen, dass bei der Anzahl der Zufallszuteilungen und der Form der empirischen Dichtefunktion der Verteilungen die gefundenen zufälligen Minima den wahren Minima relativ nahe sind.

Allein wegen des zeitlichen Umfangs dieser Rechnungen¹³⁹ und Dank der Tatsache, dass die nötigen Wirtschaftsdaten der Kreise öffentlich zugänglich sind, ist

¹³⁸ Rheinland-Pfalz: $\frac{36!}{(31)^{12} \cdot 12!} = 3,57 \cdot 10^{23}$, Ostdeutschland: $\frac{112!}{(31)^{36} \cdot (41)^4 \cdot 36! \cdot 1!} = 2,14 \cdot 10^{111}$.

¹³⁹ kumulative Rechenzeit: ca. 60 h (auf Intel® Core 2 Duo mit 2 GHz für die unter SAS® in mehreren Rechenläufe geteilte Aufgabe).

die gewählte Klumpung vorab unabhängig von GWAP und KDFV bestimmt und dem FDZ/L zur Anbindung an die FDZ-eigenen Daten übergeben worden.

5.1.4. Signifikanz und Relevanz

Im Rahmen der Modellierung werden statistische Testverfahren häufig für Gruppenvergleiche oder zur Beurteilung der Magnitude von Parameterschätzungen herangezogen. Ceteris paribus sind Testentscheidungen über einen in der Grundgesamtheit gegebenen Unterschied dabei von der zugrundeliegenden Fallzahl abhängig, kleine Fallzahlen sprechen häufiger für das Beibehalten, größere für die Ablehnung der Nullhypothese einer eher unbedeutenden Magnitude. Die fachwissenschaftliche (sozialwissenschaftliche, ökonometrische, biologische, etc.) Relevanz hingegen ändert sich mit der Fallzahl nicht.¹⁴⁰

Somit sichert weder ein hochsignifikantes Ergebnis automatisch die Relevanz noch kann ein nichtsignifikantes Ergebnis einen fachwissenschaftlich begründeten Zusammenhang als nichtig erklären. Gelman und Stern (2006) reflektieren diesen Problemkreis unter dem Titel *The Difference Between 'Significant' and 'Not Significant' is not Itself Statistically Significant*. Ihren Warnungen ist mit striktem Kontextbezug in der Interpretation jeglicher statistischer Ergebnisse zu begegnen.

5.1.5. Regressionen

Regressionen als ökonometrische Modelle zielen entweder auf eine Prognose als Vorhersage des Wertes der endogenen abhängigen Variablen für bestimmte Kombinationen der unabhängigen Variablen, oder dienen zur Analyse der Stärke des funktionalen Zusammenhangs zwischen den unabhängigen und den abhängigen Variablen des Modells. Letzteres kann zur sukzessiven Modellverbesserung durch Einschluss weiterer relevanter und Ausschluss der zusammenhangsschwächsten Variablen genutzt werden. Wichtigste Bindeglieder zwischen den unabhängigen

¹⁴⁰ Konfidenzintervall-basierte Entscheidungen zur Frage der Überdeckung des Nullwertes durch das Intervall ändern das Problem nicht.

Variablen als Regressoren und einem oder mehreren Zielparametern (Regressanden) sind der angenommene funktionale Zusammenhang, der jeweilige Skalentyp des Merkmals (kategorial, ordinal oder metrisch; Rönz und Strohe, 1994, S. 241 f.) und die Annahmen über die Verteilung der Residuen – der Abweichungen der einzelnen Datenpunkte des Regressanden von ihren modellierten Pendants.

Although classic linear regression assumes a continuous dependent variable, later incarnations of the technique allowed the response to take on a variety of more limited forms: binary, multinomial, truncated, censored, strictly integer, and others. Increasingly, regression texts are incorporating some limited-dependent-variable techniques—typically, binary response models—along with classic linear regression in their coverage.

DeMaris (2004, S. xv) betont hier aus Sicht der empirischen Sozialwissenschaft sowohl die Bedeutung der klassischen linearen Regressionsverfahren mit potenziell unbeschränktem metrischem Zielparameter als auch die der Regressionsverfahren für Regressanden mit beschränktem Wertebereich¹⁴¹.

Verallgemeinerte lineare Modelle

Verallgemeinerte lineare Modelle sind hilfreiche Erweiterungen der klassischen linearen Regression für Regressanden mit einer Verteilung aus der Klasse der exponentiellen Familie (Hedeker, 2005; Wooldridge, 2002, S. 649 ff.). Logistische Regressionen für Bernoulli- oder binomialverteilte Regressanden gehören genauso wie Poisson- und Negativ-Binomial-Modelle für Zähldaten als Regressanden zu den verallgemeinerten linearen Modellen.

Generell setzt sich ein verallgemeinertes lineares Modell aus drei Komponenten zusammen:

¹⁴¹ Wooldridge (2002, S. 451) definiert eine *limited dependent variable* als ein Merkmal, dessen Wertebereich in besonderem Maße eingeschränkt ist. Aus Sicht des Ökonometers sind Merkmale mit strikt positivem Wertebereich (Löhne, Kosten, ..., sowie jegliche Zähldaten), Merkmale mit einer endlichen Anzahl geordneter oder ordnungsfreier möglicher Werte, Merkmale mit Werten im Intervall [0,1] (Quoten, Wahrscheinlichkeiten) sowie Binärmerkmale wichtige Beispiele. Einen etablierten deutschen Fachbegriff für *limited dependent variable* gibt es nicht.

- (a) einer Wahrscheinlichkeitsverteilung aus der exponentiellen Familie für den Regressanden und seine unabhängig identisch verteilten Realisierungen $Y = (y_1, \dots, y_n)'$,
- (b) einer linearen Funktion $\eta = \mathbf{X} \cdot \boldsymbol{\beta}$ als systematischem Zusammenhang zwischen den Regressoren und dem linearen Prädiktor η , sowie
- (c) einer monotonen und differenzierbaren Linkfunktion g , so dass die Beziehung $E(Y) = \mu = g^{-1}(\eta)$ gilt.

Mit der Gauss-Verteilung in (a) und der Identität als Linkfunktion (c) gehört die lineare Regression zu den verallgemeinerten linearen Modellen. Die logistische Regression ergibt sich mit der Binomialverteilung und der Logit-Funktion $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$ als Link; Zähldaten lassen sich mit der Log-Funktion als Link und Poisson- oder negativer Binomialverteilung verallgemeinert linear modellieren. Die Parameter $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ können dabei mit iterativen Maximum-Likelihood-Methoden (z. B. IWLS: iteratively weighted least squares) geschätzt werden.

SAS® bietet für mit den Prozeduren `proc logistic` und `proc genmod` (beide SAS Institute Inc., 2004b) Möglichkeiten der Analyse verschiedener verallgemeinerter linearer Modelle.

Poisson- oder Negativ-Binomial-Modell

Poisson- und Negativ-Binomial-Modell werden beide zur Modellierung von Zähldaten herangezogen. Für Poisson-Modelle ist dabei vorauszusetzen, dass Varianz und Erwartungswert gleich sind ($E(y) = \text{Var}(y)$), eine bei Beobachtungsdaten gewöhnlich wenig realistische Forderung. Hingegen hat die negative Binomialverteilung einen Parameter für Überdispersion (engl.: *overdispersion*, $\text{Var}(y) > E(y)$); das Negativ-Binomial-Modell hat sich deshalb zur wie selbstverständlich bevorzugten Lösung bei Problemen mit Überdispersion entwickelt. Völlig zu Recht weisen Berk und MacDonald (2008) auf die Gefahren einer kritiklosen, die verschiedenen Gründe der Überdispersion nicht analysierenden, Anwendung des Negativ-Binomial-Modells hin.

Verallgemeinerte lineare gemischte Modelle und marginale Modelle

Verallgemeinerten linearen Modellen liegt die Annahme der Unabhängigkeit zwischen den Einzelbeobachtungen zugrunde. Ist diese nicht erfüllt, muss dem Umstand bei der Modellierung Rechnung getragen werden. Für longitudinale Daten beschreiben Diggle et al. (1994, Kap. 7) drei Klassen von Modellen: Marginale Modelle, Modelle mit Zufallseffekten (*random effects models*) und Modelle mit Vergangenheitseinfluss (*transition models*¹⁴²). Hedeker (2005) hingegen unterscheidet für die allgemeinere Situation von Mehrfachmessungen nur *Generalized Linear Mixed Models*¹⁴³ von den marginalen Modellen. Verallgemeinerte lineare gemischte Modelle modellieren den Regressanden bedingt auf die Zufallseffekt, sie werden ebenfalls konditional genannt.

Eine Gruppe von miteinander korrelierten Daten wird allgemein als (engl.:) *cluster* bezeichnet. Die erwähnten Modelle bieten die notwendigen Erweiterungen verallgemeinerter linearer Modelle für die Analyse von Clustern, wie sie bei Panel- und longitudinalen Daten oder allgemeiner jeder Form hierarchisch strukturierter Daten auftreten. Während Parameterschätzungen für verallgemeinerte lineare Modelle iterative Pseudo-Likelihood-Techniken und häufig geschickte Startwertvorgaben erfordern, werden die marginalen Modelle mittels verallgemeinerter Schätzgleichungen (engl.: *Generalized Estimating Equations [GEE]*), die mittels Quasi-Likelihood¹⁴⁴ gelöst werden, bearbeitet.

Das Modell mit Zufallseffekten ermöglicht die Modellierung von clusterspezifischen Effekten¹⁴⁵. Neben Maximum-Likelihood-basierten Schätzmethoden stehen verschiedene Pseudo-Likelihood-basierte Methoden, unter diesen kommt REML (residual maximum likelihood) besonders für die linearen gemischten Modelle größte Bedeutung zu (Gurka, 2006). Ist man nur an einer Schätzung der Erwartungswerte über die Cluster hinweg interessiert, also an der Randverteilung für die Population, so bieten sich beide Modelle, das marginale (mit GEE) oder das

¹⁴² etwas unspezifisch auch als konditionale Modelle – bedingt auf zeitlich vorhergehende Werte des Regressanden – bezeichnet.

¹⁴³ auch: *Generalized Linear Hierarchical Models* oder *Generalized Linear Multilevel Models*.

¹⁴⁴ Statt der (unbekannten) tatsächlichen Kovarianzmatrix geht eine relativ frei wählbare Arbeitskovarianzmatrix in die Gleichungen ein.

¹⁴⁵ Da solch ein Cluster oftmals mit einer Person gleichzusetzen ist, werden diese Modelle auch als subjektspezifische (z. B. Grömping, 1996) bezeichnet.

mit Zufallseffekten, an. Für den Spezialfall der Gauss-Verteilung mit Identität als Linkfunktion liefern diese Modelle gleiche Parameterschätzungen, leichte Abweichungen ergeben sich lediglich durch die unterschiedlichen Schätzverfahren. Für andere Verteilungen und/oder Linkfunktionen gilt dies nicht unbedingt, wie u. a. in Diggle et al. (1994, Kap. 7.4) beispielhaft für die logistische Regression (Binomialverteilung, Logit-Funktion als Link) gezeigt wird. Während das marginale Modell die mittlere Schätzung über alle Cluster liefert, lässt sich mit dem Modell mit Zufallseffekten die Mittelung aller clusterspezifischen Schätzungen bestimmen. Da jedoch allgemein $E[g(y)] \neq g[E(y)]$ gilt, ergeben sich für die Modelle sowohl unterschiedliche Parameterschätzungen als auch Interpretationen. Lee und Nelder (2004, mit Diskussionen) stellen diese Differenzen in den Mittelpunkt ihrer Reflexionen. In Gardiner et al. (2009, 2.5.3) werden für wichtige verallgemeinerte lineare gemischte Modelle die Beziehungen zwischen dem marginalem und dem Modell mit Zufallseffekten dargestellt. Die Auswahl des günstigsten Schätzverfahrens muss in Abhängigkeit vom Hauptziel der Modellierung erfolgen.

Primär geben Gardiner et al. in ihrem Artikel einen aktuellen Überblick über marginale Modelle mit GEE, gemischte Modelle und Modelle mit ausschließlich festen Effekten; Hedeker (2008) widmet sich ausführlicher den Mehrebenenmodellen für ordinale und nominale Regressanden. Dass die Begriffsbildung für diese Modelle noch nicht abgeschlossen ist, zeigen Skrondal und Rabe-Hesketh (2008) sowohl explizit als auch beim Vergleich mit den anderen hier erwähnten Publikationen.

Für die Analyse von GEE-marginalen Modellen stellt SAS® zuvörderst die Prozedur `proc genmod` zur Verfügung. Bei der Analyse von verallgemeinerten linearen (gemischten) Modellen ist zwischen mehreren SAS®-Prozeduren zu wählen, von `proc logistic` bis `proc glimmix`. In der klassischen Situation – Regressanden mit Gauss-Verteilung und der Identität als Linkfunktion – bieten sich spezifischer `proc mixed` aus Sicht der gemischten Modelle (Abschn. 3.1) oder `proc panel` aus Sicht der Paneldaten-Analyse (Abschn. 3.4.4) an (SAS Institute Inc., 2004b, 2005, 2006).¹⁴⁶

¹⁴⁶ Die Analyse nichtlinearer Modelle kann mit `proc nlmixed` (SAS Institute Inc., 2004b) durchgeführt werden.

Modellvergleiche, Modellselektion und Auswahl geeigneter Regressoren

Häufig lässt sich ein und derselbe Sachverhalt mit mehreren Modellen verschieden beschreiben, sparsamere und an Regressoren reichere Modelle konkurrieren miteinander. Sparsamere Modelle sind meist leichter interpretierbar, reicher parametrisierte Modelle sind vielleicht besser den Daten angepasst. Die Güte der Anpassung wird dabei mit verschiedenen Methoden bestimmt, viele dieser Methoden beruhen auf Informationskriterien: AIC, AICC, CAIC und BIC¹⁴⁷ sind die gängigsten. Mit Hilfe der Anpassungsgüte werden Modelle verglichen und die Auswahl des besten unter ihnen unterstützt. Singer und Willett (2003, Kap. 4.6.4) erläutern die Anwendung von AIC und BIC für Modelle mit unterschiedlichen Regressoren (*nonnested models*) über genau derselben Datenbasis; Gurka (2006) weist auf die Schwierigkeiten im Fall von REML-geschätzten Modellen hin. Im Folgenden wird das von Schwarz¹⁴⁸ vorgeschlagene Bayes'sche Informationskriterium in der Form $BIC = -2 \log(L) + q \log(N - p)$ mit dem Wert der Likelihood-Statistik L und den Anzahlen an festen Effekten p , an Kovarianzparametern q sowie an Beobachtungen N (Brown und Prescott, 2006, S. 224, modifiziert in von SAS® berichteter Form) zur Beurteilung der Modellgüte herangezogen: kleinere Werte sprechen für eine bessere Anpassung.

Der Prozess der Modellselektion besteht aus Teilschritten. Der Wahl einer geeigneten Klasse von Modellen, z. B. der multiplen linearen Regressionsmodelle für Paneldaten oder von Cox-Modellen für Verweildauern, folgen die – hier BIC-basierten – Entscheidungen, welche Merkmale sinnvoll als Regressoren dienen können¹⁴⁹ und zusammen ein gutes Modell ergeben. Die Wald-Statistik W kann helfen, die Bedeutung der einzelnen Merkmale im Modellbildungsprozess sowie das letztendlich gewählte Modell zu bewerten. Für eine beliebige Linearkombination $L'\beta$ des Parametervektors $\beta = (\beta_0, \dots, \beta_n)'$ betrachtet man dabei das Hypothesenpaar

$$H_0: L'\beta = 0 \quad \text{gegen} \quad H_A: L'\beta \neq 0.$$

¹⁴⁷ Akaike's information criterion, corrected AIC, consistent AIC, Schwarz's Bayesian information criterion.

¹⁴⁸ Schwarz G. Estimating the dimension of a model. Annals of Statistics 6: 461–464, 1978.

¹⁴⁹ z. B. durch Hinzunahme weiterer mittels step-up- oder Ausschluss bereits vorhandener Merkmale im step-down-Verfahren.

Speziell lässt sich mit der Wahl der entsprechenden Matrix L die Frage klären, ob ein einzelner Parameter β_i von 0 verschieden ist.

Die Wald-Statistik

$$W = (L'\hat{\beta})' \cdot V^{-1} \cdot (L'\hat{\beta})$$

hat für die interessierende Linearkombination $L'\hat{\beta}$ der Parameterschätzungen $\hat{\beta}$ und der entsprechenden Varianz-Kovarianz-Matrix V unter der Nullhypothese asymptotisch eine χ_n^2 -Verteilung mit Freiheitsgraden n entsprechend dem Rang der Matrix L . Die Indikatorvariablen zur Charakterisierung eines mehrstufigen Einflussfaktors lassen sich so gemeinsam testen. Für den Test einzelner Parameter β_i auf Verschiedenheit von 0 reduziert sich W zu einer χ_1^2 -verteilten Größe, diese Situation lässt sich äquivalent über $z = \sqrt{W}$ – gleichfalls als Wald-Statistik bezeichnet – mit Hilfe der Standardnormalverteilung beurteilen. Wegen der asymptotischen Gültigkeit beider Statistiken empfehlen sich besonders für kleine Fallzahlen der Vergleich der Wald-Statistik mit der F- bzw. t-Verteilung (McCulloch und Searle, 2001, S. 24, S. 240; Brown und Prescott, 2006, S. 77). Ablehnungswahrscheinlichkeiten für die Wald-Statistik werden im Folgenden als p-Werte angegeben.

Signifikanz zu einem vorgegebenen Niveau¹⁵⁰ bedeutet, dass sich der Parameter im Verhältnis zur Unsicherheit seiner Schätzung ausreichend von 0 unterscheidet. Wie an anderer Stelle (Abschn. 5.1.4) ausgeführt, kann die Frage der ökonomischen Bedeutung eines Parameters nicht durch Signifikanz bzw. Nichtsignifikanz allein beantwortet werden.

Verteilungs-Annahmen und -Anpassungen

Adäquate Modellierung von Daten basiert u. a. auf Annahmen zur Verteilung der einzelnen Merkmale.¹⁵¹ Explorativ werden Daten dahingehend untersucht, welche (für die folgenden Analyseschritte gut nutzbare) Verteilung den Daten am besten entspricht: Normal-, Lognormal-, Weibull- oder Exponential-Verteilungen sind dabei übliche Kandidaten. Kolmogorov-Smirnov-, Anderson-Darling-, Cramér-von-Mises- oder χ^2 -Anpassungs-Tests werden neben grafischen Methoden

¹⁵⁰ konventionellerweise oft auf 0,05 = 5 % gesetzt.

¹⁵¹ Hansen (2005) jedoch plädiert für semiparametrische Verfahren basierend auf weniger strikten Annahmen zur Verteilung der Merkmale.

für den Vergleich der empirischen mit einer (angepasst) gewählten theoretischen kumulativen Verteilungsfunktion genutzt.

Viele Wirtschaftsparameter werden als Merkmale auf einer metrischen Skala mit positiven Werten gemessen; durch Logarithmieren transformieren sich diese Werte in einen unbeschränkten reellzahligen Wertebereich. Für die Analyse so transformierter Daten ergeben sich drei bedeutende Vorteile: (a) der Einfluss einzelner extrem groß erscheinender Werte¹⁵² wird verringert, (b) die Chancen auf Homoskedastizität der Residuen steigt, (c) die Ergebnisse lassen sich gut als Veränderungs- oder Wachstumsrate – Steigerung bzw. Verringerung um Anteile vom vorherigen Ganzen – interpretieren. Gaynor und Anderson (1995), Crémieux und Ouellette (2001) sowie Dranove und Lindrooth (2003) verwenden alle in ihren Analysen zur Effizienz von Krankenhäusern eine (*standard*) *translog cost function* bzw. modellieren in *translog functional form* basierend auf der Annahme einer lognormalen Verteilung für die Kosten-Daten. Nunes Amaral et al. (1997) bzw. Buldyrev et al. (1997) gehen für die Beschreibung des Wachstums von Firmen ähnlich vor. Diese Überlegungen gelten für Regressanden und Regressoren gleichermaßen. Die Ergebnisse lassen sich in einfacher Weise als Elastizitäten (Rönz und Strohe, 1994, S. 101 ff.) interpretieren.

Eigenen Untersuchungen folgend werden in diesem Projekt alle Kostendaten sowie einkommensbezogene Merkmale¹⁵³ als lognormalverteilt in die Modellierungen einbezogen.

Quoten¹⁵⁴ mit Werten aus dem geschlossenen Intervall $[0, 1]$ (bzw. $[0\%, 100\%]$) werden in ihren Logit

$$\text{logit}(p) = \ln \left(\frac{p}{1-p} \right), \quad p \in [0, 1]$$

mit unbeschränkt reellzahligem Wertebereich transformiert.

¹⁵² z. B. vereinzelte ‚Spitzen‘-Gehälter in einer Untersuchung zum Einkommen abhängig Beschäftigter.

¹⁵³ z. B. Arbeitnehmereinkommen je Arbeitnehmer, verfügbares Haushaltseinkommen je Einwohner.

¹⁵⁴ z. B. Arbeitslosenquoten.

Die Brutto-Erwerbstätigenquote – der Quotient aus Erwerbstätigen (Inlandskonzept) und Einwohnern (Inländerkonzept, Abschn. 3.1.5) und als Näherungsvariable (Proxy) für die Erwerbstätigenquote herangezogen – ist selbst keine Quote, da Werte von über 100 % angenommen werden können (z. B. Bayern, Kreis Schweinfurt, 2005: 60.800 Erwerbstätige, 54.600 Einwohner).¹⁵⁵ Wie für Kostendaten so führt auch für die Brutto-Erwerbstätigenquote die Annahme einer Lognormal-Verteilung zu einer Verringerung der Heteroskedastizität.

Extremwerte, Multikollinearität und numerische Überlegungen

Schätzungen der Regressionsparameter sind numerisch zu gewinnen. Dabei können verschiedene Umstände einer stabilen numerischen Lösung entgegenstehen, wobei besonders iterative Verfahren mit numerischen Approximationen anfällig für numerische Instabilitäten sind. Als auslösende Faktoren kommen vor allem Regressoren mit einzelnen Extremwerten, Regressoren mit einer gegenüber dem Mittelwert um ein vielfaches kleineren (oder verschwindenden) Variabilität sowie Multikollinearität zwischen den Regressoren in Betracht.

Extremwerte sind einzelne Merkmalswerte, die sich deutlich von der übergroßen Mehrzahl der Beobachtungswerte eines Merkmales unterscheiden.¹⁵⁶ Der Größe des Einflusses von Extremwerten auf die Modellanpassung – das Regressionsergebnis – ist jeweils im Rahmen der Modellvalidierung zu untersuchen.

Vittinghoff et al. (2005, S. 148) definieren: *Multikollinearität*¹⁵⁷ nennt man die *Korrelation zwischen Regressoren, wenn sie stark genug ist für einen substantiellen Präzisionsverlust bei der Schätzung der Regressionskoeffizienten einiger oder aller der korrelierten Regressoren*¹⁵⁸. Multikollinearität führt ebenfalls zu numerisch

¹⁵⁵ Für Untersuchungsgebiet und -zeitraum hat Koblenz mit Werten zwischen 84,0 und 88,6 % mit Abstand die höchste Brutto-Erwerbstätigenquote, alle anderen Werte liegen zwischen 22,0 % für den Rhein-Pfalz-Kreis 1992 und 75,0 % für Mainz im Jahr 2000 mit einem Median von 38,4 % repräsentiert durch den Burgenlandkreis in Sachsen-Anhalt 1996.

¹⁵⁶ Für amtliche Daten wird hier davon ausgegangen, dass Extremwerte im Plausibilisierungsprozess der Datenerhebung auf Korrektheit geprüft sind.

¹⁵⁷ Belsley et al. (1980, S. 85) merken an, dass *collinearity*, *multicollinearity* und *ill conditioning* synonym gebraucht werden. Unter diesen ist *ill conditioning* – von Seiten der Numeriker – am besten definiert und wäre deshalb nach Belsley et al. als Begriff zu bevorzugen.

¹⁵⁸ Übersetzung des Autors.

instabilen Lösungen, im eher theoretischen Fall perfekter Korrelation ist eine Lösung wegen Überspezifikation unmöglich. Multikollinearität vergrößert die Varianz der Schätzungen der Regressionskoeffizienten und beeinträchtigt (somit) das Einschätzen der Bedeutung einzelner Regressanden für die Regression. Deshalb ist der Varianz-Inflationsfaktor ein wichtiges Instrument der Detektion von Multikollinearität. Für klassische lineare Regressionsmodelle beschreiben Belsley et al. (1980, unveränderter Nachdruck 2004) mit dem Varianz-Inflationsfaktor, der Toleranz und den *condition indices* verschiedene Möglichkeiten der Diagnose von Multikollinearität und unterbreiten Vorschläge für die Modellmodifizierung. Lesaffre und Marx (1993) reflektieren die zunehmende Komplexität der Multikollinearitätsdiagnose für verallgemeinerte lineare Modelle, eine grundlegende Arbeit zu Multikollinearitäten in verallgemeinerten linearen gemischten Modelle steht wohl noch aus.

Multikollinearitätsdiagnosen können als Teil der jeweiligen Modellbildung nur unmittelbar im Modellbildungsprozess vorgenommen werden. Trotzdem bietet es sich für dieses Projekt an, die Daten aus frei zugänglichen Quellen (Volkswirtschaftliche Gesamtrechnung, Erwerbstätigenrechnung, Arbeitslosenquoten und siedlungsstrukturelle Gebietstypen) vorab zu untersuchen. Der Panelstruktur der geplanten Analysen Rechnung tragend werden dazu sowohl die Korrelationsmatrix aller Jahre (1992–2005) gemeinsam als auch die jahrgangsweisen Korrelationsmatrizen betrachtet. Gleichzeitig werden, Belsley et al. folgend, die Merkmale ersatzweise als Regressanden in einer (ansonsten uninteressanten) Regression verwendet; die *condition indices* werden aus den Eigenwerten der auf 1 als Hauptdiagonale normierten Matrix $\mathbf{X}'\mathbf{X}$ als Quotienten des größten Eigenwertes mit einem jeden Eigenwert abgeleitet. Nach Belsley et al. (S. 153) sprechen *condition indices* größer 10 bereits für schwache Abhängigkeiten, *condition indices* über 100 sprechen für eine erhebliche Varianz-Inflation und ein hohes Risiko der Multikollinearität.

Für die Daten dieses Projekts zeigt sich, dass mehrere interessante Kombinationen potenzieller Regressanden *condition indices* größer als 10 aufweisen. Beispielsweise lässt die Kombination der Indikatorvariablen *Ostdeutschland?* mit dem Erhebungsjahr, der Einwohnerdichte, der Brutto-Erwerbstätigenquote, der Arbeitslosenquote und den relativen Arbeitnehmerinkommen, Brutto-Inlandspro-

dukten sowie verfügbaren Haushaltseinkommen¹⁵⁹ genauso wie das Pendant, bei dem die Einwohnerdichte durch den Kreistyp (Abschn. 3.3.5) ersetzt wird, Multikollinearitätsprobleme erwarten. Bezieht man ein Absolutglied in die Betrachtungen mit ein, ergeben sich Werte von weit über 100 für die *condition indices*.

Eine Erhöhung der numerischen Stabilität lässt sich durch Modell-Reparametrisierung erreichen, bei denen Regressoren, deren Variabilität klein ist gegenüber dem Mittelwert, geeignet transformiert werden. Üblich ist dabei die Standardisierung¹⁶⁰ mit dem arithmetischen Mittel (\bar{x}) und der empirischen Standardabweichung (s) zu $z_i = \frac{x_i - \bar{x}}{s}$, wobei der unmittelbare Bezug zum Ausgangswert auf der Original-Messskala datenabhängig und wenig intuitiv ist. Oft mag auch die einfache Subtraktion eines einprägsamen Wertes ausreichend sein. Für diese Arbeit werden die Jahreszahlen in Abstände zum Jahr 2000 transformiert.¹⁶¹

Panelanalysen mit dem verallgemeinerten linearen Modell

Die Verteilung des Regressanden in einem verallgemeinerten linearen Modell ist eine aus der exponentiellen Familie; Erwartungswert $E(y_{ti})$ und Varianz $D(y_{ti})$ der verallgemeinerten linearen Paneldaten-Modelle genügen hier

$$\begin{aligned} E(y_{ti}) &= E(y|\mathbf{x}_{ti}) = \mu_{ti} = g^{-1}(\eta_{ti}) = g^{-1}(\mathbf{x}_{ti}'\beta) \\ D(y_{ti}) &= \frac{\phi V(\mu_{ti})}{w_{ti}} \end{aligned} \quad (5.1)$$

mit t als Index der Zeit- und i als Index der Subjektkomponente. Die monotone und differenzierbare Linkfunktion $g(\cdot)$ verbindet den linearen Prädiktor $\eta_{ti} = \mathbf{x}_{ti}'\beta$ mit den Erwartungen $E(y_{ti})$. Die Varianz $D(y_{ti})$ setzt sich zusammen aus einer Konstante ϕ , einem beobachtungsspezifischen Gewicht w_{ti} und der Varianzfunktion $V(\cdot)$ mit μ_{ti} als Argument. ϕ – der konstante Dispersionsparameter – ist für manche Verteilungen bekannt¹⁶² oder muss aus den Daten geschätzt werden.

¹⁵⁹ bezogen auf die Arbeitnehmer bzw. Einwohner.

¹⁶⁰ richtiger: Studentisierung.

¹⁶¹ 1991, ..., 2005 ergeben -9, ..., 5.

¹⁶² Für Binomial-, Exponential-, Multinomial-, negative Binomial-, Poisson- und andere Verteilungen gilt $\phi \equiv 1$.

Die im Folgenden angenommenen Verteilungen des Regressanden mit zugehörigen Linkfunktionen sind: (a) die Binomialverteilung mit der Logit-Funktion, (b) die Gauss'sche Normalverteilung (Gauss-Verteilung) mit der Identität als Link, (c) die Lognormalverteilung mit der Identität, und (d) die negative Binomialverteilung mit der Log-Funktion zur Basis e . Für die Gauss-Verteilung entspricht das natürlicherweise einer Panelanalyse mit dem allgemeinen linearen Modell.

Für die verallgemeinerten linearen Paneldaten-Modelle wird hier davon ausgegangen, dass die Panelsubjekte voneinander (statistisch) unabhängig sind, bezüglich der Zeitkomponente sind die m -mal wiederholt erhobenen Werte miteinander korreliert, sie werden als Varianzkomponenten mit autoregressiven Kovarianzstrukturen der Ordnung 1 modelliert – entweder mit AR(1) oder heterogenen mit ARH(1), wobei bei ersterer davon ausgegangen wird, dass die Varianz für alle Zeitpunkte t gleich, bei letzterer jedoch zeitpunktspezifisch ist:

$$AR(1) : \quad \Sigma = \sigma^2 \cdot \begin{pmatrix} 1 & \varrho & \cdots & \varrho^{m-1} \\ \varrho & 1 & \cdots & \varrho^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho^{m-1} & \varrho^{m-2} & \cdots & 1 \end{pmatrix},$$

$$ARH(1) : \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho & \cdots & \sigma_1\sigma_m\varrho^{m-1} \\ \sigma_2\sigma_1\varrho & \sigma_2^2 & \cdots & \sigma_2\sigma_m\varrho^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_m\sigma_1\varrho^{m-1} & \sigma_m\sigma_2\varrho^{m-2} & \cdots & \sigma_m^2 \end{pmatrix} \quad (5.2)$$

Neben dem Dispersionsparameter ϕ sind hier, wie im allgemeinen linearen Modell, die Parameter der Kovarianzstruktur Σ (Gl. 5.2) und mit $\beta = (\beta_1, \dots, \beta_k)'$ die Parameter des linearen Prädiktors $\eta_{ti} = \mathbf{x}_{ti}' \beta$ zu schätzen. Modelle werden im Folgenden durch Angabe ihres linearen Prädiktors η_{ti} angegeben.

5.1.6. Analysen von Verweildauern

Obwohl sehr häufig angewendet, ist der Mittelwert kein geeignetes Maß zur Beschreibung von Verweildauern. Bömermann (2003) zeigt dieses für die Diagnose-

daten der Krankenhausstatistik. Verweildauern haben eine unsymmetrische, sehr schiefe Verteilung, einige extrem lange Verweildauern haben einen sehr großen Einfluss auf einen – dann erhöhten – Mittelwert, ohne dass sich an der Verteilung der übergroßen Mehrheit etwas ändert. Konsequenterweise wird auf den Median als Alternative verwiesen, den Wert, für den 50 % der Verweildauern geringer und 50 % größer sind¹⁶³. Der Median bleibt unbeeinflusst, wenn bei großen Verweildauern sich der genaue Wert nicht mehr exakt bestimmen lässt. Ist wenigstens bekannt, wie groß ein solcher unexakter Wert mindestens gewesen sein muss, dann kann der unexakte durch diesen Mindestens-Wert ersetzt werden, die unexakte Verweildauer nennt man dann am Ersatzwert (*rechts*) *zensiert*.

Als Maß für die Streuung wird von Bömermann der Quartilsabstand herangezogen, d. h. die Differenz zwischen dem 75. Perzentil (3. Quartil) und dem 25. Perzentil (1. Quartil) der empirischen Verteilung der Verweildauern. Eine noch bessere Beschreibung lässt sich durch die Angabe weiterer ausgewählter Perzentile (z. B. 5., 10., 90., und 95.) erreichen¹⁶⁴.

Verweildauern als Zeitdifferenzen zwischen einem End- und einem Anfangszeitpunkt sind immer nichtnegative Zahlen, sie stellen die Zeit von der Aufnahme bis zum Ereignis der Entlassung aus dem Krankenhaus dar und lassen sich als nichtnegative Zufallsvariable verstehen. Ereigniszeitanalysen sind dafür adäquate statistische Verfahren¹⁶⁵, einen umfassenden Überblick geben Singer und Willett (2003, Teil II) sowohl für kontinuierlich als auch diskret erhobene Ereigniszeiten. Häufigst verwendete Methode zur Schätzung von Ereigniszeitverteilungen ist die Kaplan-Meier-Schätzung $S(t)$, sie ist für die gemeinsame Analyse beobachteter und zensierter Ereigniszeiten t geeignet und entspricht dem Komplement $S(t) = 1 - F(t)$ der empirischen Verteilungsfunktion $F(t)$.

Neben der Kaplan-Meier-Schätzung ist die Hazard¹⁶⁶-Funktion $\lambda(t)$ ein bewährtes Mittel, Ereigniszeitverteilungen zu analysieren. Sie beschreibt über $\lambda(t)dt =$

¹⁶³ genauer: Der Median ... ist der Wert des mittelsten Elements bei ungeradem Stichprobenumfang bzw. das arithmetische Mittel der beiden mittelsten Elemente bei geradem Stichprobenumfang (Rasch, 1987, S. 368 f.).

¹⁶⁴ Jedes dieser Perzentile ist nur dann verzerrungsfrei, wenn alle rechts-zensierten Werte größer als das gewählte Perzentil sind.

¹⁶⁵ Erste Anwendungen dieser Methoden betrafen die Analyse der (Verlängerung der) Lebenszeit nach Tumorbehandlung, deshalb wird sie häufig Überlebenszeitanalyse genannt.

¹⁶⁶ engl: *hazard* $\hat{=}$ Gefahr, Risiko, Wagnis, Zufall.

$\Pr(t \leq T \leq (t + dt) | T \geq t)$ für jeden (nichtnegativen) Zeitpunkt die Wahrscheinlichkeit für das Eintreten des Ereignisses genau zu diesem Zeitpunkt unter der Voraussetzung, dass es bis dahin noch nicht eingetreten ist. Für Analysen genutzt werden dabei die kumulative Hazard-Funktion $\Lambda(t)$ oder geglättete (z. B. *kernel smoothed*) Hazard-Funktionen (Singer und Willett, 2003, S. 498).

Cox (1972) entwickelt ein Regressionsmodell (bekannt als *proportional hazards regression model* oder *Cox-Modell*), mit dem sich der Einfluss verschiedener exogener Variabler auf die Hazardrate durch Modellierung der kumulativen Hazard-Funktion analysieren lässt. Das Cox-Modell

$$\Lambda(t_{ij}) = \Lambda_0(t_j) \cdot e^{(\mathbf{x}_{ij}' \boldsymbol{\beta})} = \Lambda_0(t_j) \cdot e^{(\beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \dots + \beta_p X_{p,ij})} \quad (5.3)$$

modelliert für alle Subjekte i und Ereigniszeiten j den Einfluss der Regressoren X_1 bis X_p auf die kumulative Hazard-Funktion $\Lambda(t_{ij})$. $\Lambda_0(t_j)$ ist die für alle Subjekte gemeinsame Baseline kumulative Hazard-Funktion. Die Regressoren X_p , $p \in \{1, \dots, P\}$ selbst können subjekt- und zeitveränderliche Werte $x_{p,ij}$ annehmen; $\eta_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta}$ ist wiederum ein – hier absolutgliedfreier – linearer Prädiktor η . Entlogarithmiert man (zur Basis e) die Parameter $\boldsymbol{\beta}$, so ergeben sich mit den entsprechenden Hazard-Verhältnissen leichter zu interpretierende Größen.

Ereigniszeitverteilungen $S_i(t)$ mehrerer Gruppen ($i = 1, \dots, k$, $k \geq 2$) lassen sich mit dem Logrank-Test (Mantel-Haenszel-Test) vergleichen. Dieser prüft mit dem Hypothesenpaar

$$H_0: S_1(t) = \dots = S_k(t) \quad \text{gegen} \quad H_A: \exists i, j: S_i(t) \neq S_j(t).$$

die Gleichheit aller gruppenspezifischen Ereigniszeitverteilungen gegen die Alternative, dass mindestens zwei dieser Ereigniszeitverteilungen voneinander verschieden sind. Der Logrank-Test und der dem Cox-Modell inhärente *score test* sind im Falle fehlender Bindungen identisch, mit Bindungen führen ihre verschiedenen Varianzschätzungen zu marginal unterschiedlichen Ergebnissen (Therneau und Grambsch, 2000, Kap. 3.4).

Erweiterte Cox-Modellvarianten für feste und/oder zufällige Effekte finden zunehmend Anwendung (z. B. Therneau und Grambsch, 2000, Kap. 9 Frailty Models). Nutzt man die Verfahren für Zählprozesse, so lässt sich unter Nutzung von

Ramlau-Hansen (1983) ein Glättungsverfahren für die Hazard-Funktion entwickeln und anwenden¹⁶⁷.

5.1.7. Modellierung unter KDFV-Anonymisierungsforderungen

Verschiedene Autoren (z. B. Allison, 2005, Kap. 1 Introduction to Fixed Effects Methods) stellen begründet die Vorteile einer Modellierung von longitudinalen und Paneldaten mit den Subjekten als feste Effekte dar. Im vorliegenden Fall sind entweder Krankenhäuser oder Kreise die Subjekte. Modelliert als feste Effekte wäre jeder Kreis bzw. jedes Krankenhaus seine eigene Referenz bei der Betrachtung der Veränderungen über die Zeit.¹⁶⁸ Gleichzeitig bedeutete dies für die Modellierung das Beherrschen einer großen Anzahl von Freiheitsgraden: 147 mit Kreisen bzw. 683 mit Krankenhäusern¹⁶⁹ als Subjekte.

Modellierungen sind – wie der Datenzugriff – unter KDFV-Anonymisierungsforderungen sehr weitreichend möglich, erst der Ergebniszugriff ist den Anonymisierungsforderungen gemäß nur eingeschränkt gestattet (Abschn. 4.1.2). Modelle, in die die Krankenhäuser als feste Effekte eingehen, sind anpassbar, können jedoch nicht überprüft werden, denn eine Überprüfung der Modellanpassung erforderte die Betrachtung der krankenhausspezifischen Parameterschätzungen und Residuen. Da einige Kreise weniger als drei Krankenhäuser haben, gilt das gemäß Dreier-Regel auch für Modellierungen mit Kreisen als feste Effekte.

Deshalb muss auf die Panel-Modellierung mit Krankenhäusern oder Kreisen als festen Effekten verzichtet werden. Kreise (und Krankenhäuser) werden somit als zufällige Realisierungen von ökonomischen Entscheidungen innerhalb Ost- bzw. Westdeutschlands (Rheinland-Pfalz) im Sinne von Haavelmo (Abschn. 3.4.3) verstanden und als zufällige Effekte (Abschn. 3.4.4) modelliert.

¹⁶⁷ Hier wird ein vom Autor weiterentwickeltes SAS®-Makro von Allison verwendet, <http://www.soc.upenn.edu/courses/soc613/SMOOTH.SAS>, Zugriff am 31. 03. 2008.

¹⁶⁸ Beziehungen zwischen Krankenhäusern / Kreisen mit anderen, zeitunveränderlichen Faktoren wie z. B. der Frage, ob in Ostdeutschland gelegen, sind mit festen Effekten nicht modellierbar.

¹⁶⁹ genauer: 683 verschiedene Krankenhausidentifikationen, vgl. Tab. 4.2 auf S. 103.

5.1.8. Verzerrungen und Modellfehlspezifikationen

Verzerrungen oder systematische Fehler (engl.: *bias*) nennt man jeden nicht auf der Zufallsauswahl einer Stichprobe beruhenden Fehler einer Schätzung (Rönz und Strohe, 1994, S. 363). Unter den vielen Arten möglicher Verzerrungen sind *selection bias* und *omitted variable bias* bei Beobachtungsstudien besonders relevant. Wie bei den statistischen Modellen (Abschn. 3.1: Hierarchische, Mehrebenen- und Gemischte Modelle in Statistik und ihren Anwendungen) gibt es auch bei den Verzerrungen anwendungsgebietspräferierte Terminologien (Imlach Gunasekara et al., 2008; Maciejewski et al., 2011).

Selection bias – die Verzerrung aufgrund einer nichtrepräsentativen Stichprobenauswahl – sollte bei Totalerhebungen (Abschn. 3.4.1) eigentlich ohne Bedeutung sein, in vorliegendem Projekt kann jedoch das jahresweise Fehlen der Daten kompletter Bundesländer zu einem *selection bias* führen.¹⁷⁰

Omitted variable bias – die Verzerrung wegen unberücksichtigter Merkmale – kann entstehen, wenn beim Modellieren Merkmale mit relevantem Informationsgehalt unberücksichtigt bleiben, weil sie entweder nicht erhoben oder in ihrer Bedeutung für die Modellierung verkannt sind.¹⁷¹ Ist die unberücksichtigte Variable mit dem Regressanden oder einem der Regressoren korreliert, muss *omitted variable bias* angenommen werden. Unberücksichtigte Merkmale und Störvariable (engl.: *confounders* in epidemiologischer Terminologie) fallen besonders dann als Problem bei der Modellierung zusammen. Ist ein Merkmal nicht erhebbar, so mag eine geeignete Näherungsvariable (Proxy) die Modellierung ermöglichen.¹⁷² Unter bestimmten Annahmen können mit Paneldaten trotz unberücksichtigter Variabler konsistente Parameterschätzungen gewonnen werden. Speziell gilt das, wenn das unberücksichtigte Merkmal für jedes Panelsubjekt (Abschn. 3.4.1) genau einen (zeitkonstanten) Wert aufweist (Wooldridge, 2002, Kap. 10.1).

¹⁷⁰ Mit dem Fehlen der Daten aus Mecklenburg-Vorpommern sind besonders die ländlichen Räume (Abschn. 3.3.5) für die betroffenen Berichtsjahre weniger präsent.

¹⁷¹ Die Trennung der Begriffe *omitted variable bias* und *selection bias* scheint in der Literatur allgemein deutlicher als in Maciejewski et al. (2002, Tab. 1) angegeben.

¹⁷² Bis zu welcher Güte eine Näherungsvariable als geeignet betrachtet werden kann, mag im konkreten Fall anhand bereits vorliegender Daten schwer überzeugend zu klären sein.

A priori ist für vorliegendes Projekt u. a. anzunehmen, dass die (hier) unbekannte Altersverteilung der Bevölkerung eines jeden Kreises Quelle eines *omitted variable bias* ist: für Kreise mit jüngerer Altersstruktur ist ein anderes Krankheitsmuster – und somit ein anderer Bedarf an Krankenhausleistungen – anzunehmen als für Kreise mit im Mittel älterer Wohnbevölkerung. Inwieweit die ebenfalls unbekannte Geschlechtsverteilung eine weitere Quelle für bedeutsamen *omitted variable bias* ist, muss hier offen bleiben.

Neben den im Modellierungsprozess unberücksichtigten Variablen stellen irrelevante Variable ein weiteres Problem dar. Irrelevante Variable haben keinen (oder nur sehr geringen) Einfluss auf den Regressanden und sind somit für die Modellbeziehung verzichtbar. Da ihr Einschluss in das Modell jedoch zu einer Erhöhung der unerklärten Variabilität führt und das Risiko für Multikollinearität steigt, wird durch diese der Modellbildungsprozess behindert. Die Entscheidung, ob ein Merkmal irrelevant oder *omitted variable bias*-generierend ist, sollte Teil des Modellbildungsprozesses sein (Wooldridge, 2003, Kap. 3).

5.1.9. Exogene und endogene Variable und ihre Bedeutung für die Modellierung

Die Raumordnungsmerkmale (Abschn. 3.3.5) Kreistyp bzw. Regionsgrundtyp charakterisieren einen jeden Kreis bezüglich seiner Urbanität unter Bezugnahme auf die angrenzenden Kreise. Diese zwei Merkmale sind die einzigen, deren Werte für einen Kreis unmittelbar in Bezug zu den Werten der Nachbarkreise stehen. Beide Merkmale ändern ihre Werte nur selten; sie sind für den Untersuchungszeitraum zeitkonstant.

Die Einwohnerdichte – der Quotient aus Einwohnerzahl und Fläche – beschreibt ebenfalls Urbanität, jedoch ohne Bezug zu umgebenden Kreisen. Stetigen Veränderungen unterliegt dabei die Einwohnerzahl; die Fläche ändert sich ausschließlich mit administrativen Entscheidungen¹⁷³.

Während die (mittlere jährliche) Arbeitslosenquote Auskunft über die soziale Lage der Erwerbsbevölkerung gibt, ist das Brutto-Inlandsprodukt eine allgemein ge-

¹⁷³ z. B. durch Kreisreformen, Gemeindeausgliederungen oder Eingemeindungen.

nutzte Näherung für die Wirtschaftskraft eines Kreises¹⁷⁴. Um eine Vergleichbarkeit auch zwischen ‚großen‘ und ‚kleinen‘ Kreisen herzustellen, wird das BIP auf die Einwohnerzahl (Inlandskonzept) bezogen. Derselbe Bezug wird für das Arbeitnehmerinkommen als Näherung für die Wirtschaftskraft der Beschäftigten und das verfügbare Haushaltseinkommen als Maß für die Konsumkraft der Einwohner.

Alle bisher genannten werden als exogene Variable in verschiedene Modelle eingehen.

Eher endogene Bedeutung haben die Variablen, die das einzelne Krankenhaus bzw. kumulativ die Gesamtkrankenhausesaktivität eines Kreises (Abschn. 5.1.2: Gesamtkrankenhausesaktivität) beschreiben. Das gilt für Kostendaten genauso wie für Merkmale zur personellen und materiellen Ausstattung der Krankenhäuser. Vergleichbarkeit zwischen Krankenhäusern bzw. Kreisen wird durch Bezug auf die Einwohnerzahl (Inlandskonzept) oder die Anzahl der behandelten Fälle (aus den Diagnosedaten) hergestellt.

5.1.10. Räumliche Statistik – Landkarten

Mit einer Choroplethenkarte (auch Flächenkartogramm, engl.: *choropleth map* oder *shaded map*)¹⁷⁵ kann flächenbezogen die Verteilung eines qualitativen oder quantitativen Merkmales dargestellt werden, wenn sich das darzustellende Merkmal auf die vorselektierten Teilflächen und nicht auf einzelne Punkte der Karte bezieht. Die Teilflächen werden in Abhängigkeit des zugehörigen Merkmalswertes gekennzeichnet (eingefärbt und/oder schraffiert). Teilflächen können dabei kleinere administrative Einheiten wie Gemeinden, Kreise oder Bezirke innerhalb größerer Einheiten sein, sie werden durch ihre Grenzen definiert.

Basierend auf wahrnehmungsspezifischen (Farb- oder Musterwahl, z. B. Cleveland, 1994, Kap. 4: Graphical Perception) bis zu statistischen Überlegungen wurde für die grafische Kennzeichnung der Flächen eine Reihe von Kriterien vorgeschlagen und in ihren Eigenschaften charakterisiert. Zu den wichtigsten Fragen gehö-

¹⁷⁴ jedoch nicht für den Wohlstand (Abschn. 3.3.4).

¹⁷⁵ Gegensatz: *contour map* bzw. *isopleth map* – eine Landkarte mit Höhenlinien.

ren die nach der Anzahl und der Definition der Klassen (Intervalle), in die das darzustellende quantitative Merkmal eindeutig und vollständig eingeteilt werden soll. Jenks und Caspall (1971) entwickelten ein statistisches Verfahren zur datengesteuerten Klassierung, das die Variabilität innerhalb aller Klassen minimiert, diejenige zwischen den Klassen hingegen maximiert. Gale und Halperin (1982) empfehlen die klassenlose Choroplethenkarte, Einfärbung bzw. Schraffierung der Teilflächen erfolgen entsprechend einer stetigen Farb- oder Musterskala, während Evans (1977) eine Einordnung der verschiedenen Verfahren der Klassenbildung vornimmt. Monmonier (2005) warnt, dass mittels willkürlicher Klassenteilungen (und entsprechend signalisierender Einfärbung) sowohl interessante Aspekte der Daten versteckt als auch weniger bedeutsame dramatisiert werden können. Exogene Klassenteilungen erfolgen zu themenrelevanten Klassengrenzwerten (z. B. ein Geschlechterverhältnis von 1), ihre Festlegung erfolgt datenunabhängig. Idiografische Einteilungen hingegen sind von den jeweils vorliegenden Daten abhängig, hierzu gehören das Verfahren von Jenks und Caspall (1971) sowie alle Quantileinteilungen.

Mit Choroplethenkarten lassen sich sowohl die Flächengrößen als auch nachbarschaftlichen Verhältnisse zwischen den Teilflächen maßstabsgetreu abbilden, wobei die Flächengrößen wahrnehmungstechnisch hervorstechen: große Flächen sind deutlich auffälliger. Dieses ist ein Nachteil der Choroplethenkarten, denn die Flächengröße selbst steht in den meisten Fällen nicht mit dem zugehörigen Wert des darzustellenden Merkmals in Beziehung. Walter (1993) zeigt dies beispielhaft an standardisierten Inzidenzraten bestimmter Tumore (das quantitative Merkmal) in den administrativen Einheiten der kanadischen Provinz Ontario (Statistics Canada, 2010). Die größte unter den 50 Einheiten (Kenora District, 407.192 km²) ist hier fast 400 mal größer als die kleinste (Prince Edward County, 1.050 km²); Kenora District allein nimmt knapp 38 % der Fläche Ontarios ein. Verglichen damit sind die Verhältnisse im Untersuchungsgebiet vorliegender Arbeit deutlich homogener: 2005 ist Deutschlands flächengrößter Kreis, der Landkreis Uckermark in Brandenburg (3.058 km²), etwa 80 mal so groß wie die kreisfreien Städte Stralsund in Mecklenburg-Vorpommern (39 km²) oder Speyer in Rheinland-Pfalz (41 km²), die Fläche des Landkreises Uckermark stellt nur knapp 2,4 % der Fläche des aus 147 Einheiten bestehenden Untersuchungsgebietes (128.432 km²) dar.

Gelman und Price (1999) weisen darauf hin, dass Choroplethenkarten keinen Hinweis auf die jeweils teilflächenbezogene Stichprobengröße zulassen. Sie zeigen, dass größere Unterschiede in den Fallzahlen zwischen den Teilflächen zu verzerrten Ergebnissen führen, die räumliche Beziehung zwischen den betrachteten Teilflächen hat einen Einfluss auf die Schätzung für jede einzelne Teilfläche. Ersteres spielt bei vorliegender Totalerhebung eher eine geringere Rolle, zweiteres hingegen ist zu bedenken.

Bereits Evans (1977) thematisiert die Möglichkeit, dass Gruppen von Karten zum Vergleich mehrerer Merkmale in ein und derselben oder eines Merkmales in verschiedenen geografisch-administrativen Regionen herangezogen werden können. Brewer und Pickle (2002) untersuchen sieben verschiedene Klassenbildungs-Methoden und dazu angepasste Techniken zur Darstellung von Kartenserien. Die Beispiele in Gale und Halperin (1982) betrachtend wird deutlich, dass das Lesen einer klassenlosen Choroplethenkarte dem Lesen einer Choroplethenkarte mit zu vielen Klassen sehr ähnlich sein kann; Evans (1977) schlägt zwischen 4 und 10 Klassen vor, Brewer und Pickle (2002) betrachten Karten-Serien mit 5 bzw. 7 Klassen.

Unter den sieben Klassenbildungs-Methoden in Brewer und Pickle (2002) ist die idiografische Einteilung mittels Quantil¹⁷⁶-Methode die einzige, die sichert, dass jeder Klasse genauso viele Teilflächen angehören. In Kartenserien wird jede Karte mit spezifischen Quantilswerten als Klassengrenzen aufwarten, jedoch sind Klassen mit extrem geringer Belegung oder leere Klassen ausgeschlossen.

Unter Beachtung der die Anonymität der Krankenhäuser (und die der Fälle) sichernden Anforderungen (Abschn. 4.1.2) werden in vorliegender Arbeit Quintilkarten, d. h. Choroplethenkarten mit 5 Klassen, zur Darstellung der räumlichen Verteilung einzelner Parameter über die 147 Kreise des Untersuchungsgebietes herangezogen. Die Werte des 1. bis 4. Quintils werden als *Quintilgrenzen* angegeben, das n-te Quintil trennt dabei den n-ten vom (n+1)-ten Quintilsbereich. Wegen der Anonymitätsanforderungen muss auf die Angabe von Minimum und Maximum genauso verzichtet werden wie auf die Ergänzung der Karten um ein

¹⁷⁶ In geordneten Stichproben bezeichnet man [...] als *k*-Quantile *k*-1 Zahlenwerte, die diese Stichprobe in *k* Gruppen teilen, die je ein *k*-tel der Beobachtungswerte enthalten (Rasch, 1987, S. 491). Quantile mit eigener Benennung sind: Quartile (*k*=4), Quintile (*k*=5), Dezile (*k*=10) und Perzentile (*k*=100).

univariates Scatterplot-Histogramm oder ein kumulatives Häufigkeitsdiagramm (Monmonier, 2005; Cromley und Cromley, 2009). Um der Forderung nach der Nicht-Identifizierbarkeit einzelner Krankenhäuser Genüge zu leisten, sind diejenigen Kreise von der kartografischen Darstellung auszunehmen, in denen (im jeweiligen Berichtszeitraum) weniger als 3 Krankenhäuser ihren Sitz haben. (Die Quintilberechnungen selbst beziehen sich jedoch auf alle Kreise mit mindestens einem Krankenhaus.)

5.1.11. Kompositionsdaten

Für nichtnegative Daten, die so in innerer Abhängigkeit zueinander stehen, dass ihre Summe für jede Beobachtungseinheit denselben konstanten Wert ergibt, scheint sich im deutschen Sprachraum bisher kein eindeutiger Begriff eingebürgert zu haben. So spricht van den Boogaart 2008 von *Zusammensetzungsdaten*, im Folgejahr an gleicher Stelle über *Statistik von Zusammensetzungen* und von *Zusammensetzungen / Kompositionen*¹⁷⁷ als Daten spezieller Skalen. Hron und Filzmoser (2008) nutzen explizit den Begriff der *Kompositionsdaten*, dieser Begriff wird im Folgenden verwendet werden.

Nichtnegative Daten konstanter Summe sind Kompositionsdaten. Allgemeiner jedoch definiert man Kompositionsdaten als Daten, die nur relative Informationen als Teil eines Ganzen tragen, oder bei denen nur diese relativen Informationen zueinander von Interesse sind (Filzmoser et al., 2009). Die gewonnenen relativen Informationen lassen sich bezüglich des jeweiligen Ganzen standardisieren – auf eine Konstantsumme.¹⁷⁸ Häufig hat die Summe den Wert 1 (oder 100%), wenn es sich um Aussagen über die Anteile an einem Ganzen handeln soll.¹⁷⁹ Bereits vor mehr als 110 Jahren erkannte Pearson (1897) das diesen Daten innewohnende Problem einer rein der Beziehung zwischen den einzelnen Komponenten

¹⁷⁷ http://www.stat.boogaart.de/interests_de.php » Arbeits- und Interessengebiete, Zugriffe am 07. 05. 2008 bzw. 21. 03. 2009.

¹⁷⁸ Regressionen mit Kompositionsdaten als Regressanden sind somit *limited dependent variable*-Modelle.

¹⁷⁹ Andere Konstantsummenwerte sind vorstellbar, wenn z. B. in Psychologie oder Marketing eine vorgegebene Menge von Punkten der eigenen Präferenz entsprechend auf eine vorgegebene Anzahl von Wertekategorien vergeben werden soll, diese Daten werden auch als ipsative Daten bezeichnet.

geschuldeten Verzerrung der Korrelation, er nannte diese unberechtigt-störend (*spurious*). Für Kompositionsdaten ergibt sich diese Verzerrung, da die *letzte* Komponente einer Komposition aus D Komponenten immer den Ergänzungswert zur Konstantsumme darstellt, mindestens eine der Kovarianzen und somit auch der Korrelationen zwischen Komponenten ist negativ – für $D = 2$ ist diese Korrelation $\rho \equiv -1$. Die Korrelationskoeffizienten zwischen Komponenten können somit nicht alle frei zwischen -1 und $+1$ schwanken, sie sind verzerrt hin zu negativen Werten. Einen validen Lösungsansatz für den Umgang mit dieser Art Daten gab es für die auf Pearson (1897) folgenden 80 Jahre nicht, obwohl besonders in der Geologie (Mineral-Kompositionen von Gesteinen, Kompositionen verschiedener Erden) als auch in der Archäologie (Mineral-Kompositionen von Artefakten) Erkenntnisse häufig in Form von Kompositionsdaten vorliegen.

Zu Beginn der 80er Jahre des vergangenen Jahrhunderts stellte Aitchison (1982, 1986) eine völlig neue Herangehensweise an die Analyse von Kompositionsdaten vor. Kompositionsdaten lassen sich als reellwertige Vektoren mit positiven Komponenten darstellen, die ein Simplex

$$S^D = \{x \in \mathbb{R}^D \mid x_i > 0, x_1 + \dots + x_D = \text{const}\}$$

der Dimension $D-1$ im D -dimensionalen Raum aufspannen. Ein Simplex S^D ist somit ein begrenzter $(D-1)$ -dimensionaler Teilraum des reellen D -dimensionalen Raumes \mathbb{R}^D . Die Punkte $\vec{x} = (x_1, \dots, x_D)'$ eines Simplex S^D wiederum kann man durch geeignete ein-eindeutige Transformationen in Punkte eines $(D-1)$ -dimensionalen unbegrenzten Raumes Euklidischer Struktur überführen, nennen wir diesen einen Koordinatenraum. Drei Transformationstypen wurden dafür bisher vorgeschlagen: **alr** (additive logratio), **clr** (centered logratio) und **ilr** (isometric logratio).

Die Transformationen **alr** und **clr** (mit dem geometrischen Mittel aller Komponentenwerte $g(\vec{x}) = g(x_1, \dots, x_D) = \sqrt[D]{x_1 \cdots x_D}$ als Bezug) gehen auf Aitchison (1982) zurück, **ilr** wurde erstmals von Egozcue et al. (2003) vorgestellt. Alle drei Transformationen basieren auf dem Logarithmieren; zum dabei möglicherweise auftretenden Problem mit Nullwerten in einzelnen Komponenten der Kompositionen siehe z. B. Martín-Fernández und Thío-Henestrosa (2006).

Während **alr** und **ilr** ein Simplex S^D direkt in den \mathbb{R}^{D-1} als Koordinatenraum abbilden, wird mittels **clr** in eine durch den Ursprung des \mathbb{R}^D verlaufende Hyperebene der Dimension $D-1$ transformiert, so dass jeder Punkt im Simplex in Koordinaten des \mathbb{R}^D überführt werden, deren Summe 0 ergibt. Im Unterschied zu **alr** ist **clr** eine isometrische Transformation, sie führt das Abstandsmaß der Aitchison-Geometrie auf dem Simplex S^D in den gewöhnlichen Euklidischen Abstand des \mathbb{R}^D (eingeschränkt auf die zu betrachtende $(D-1)$ -dimensionale Hyperebene) über.

Weder **ilr** noch **alr** sind eindeutig definiert, es handelt sich eher um Klassen von Transformationen. Der Vorteil der **ilr**- gegenüber der **alr**-Transformation besteht darin, dass mittels **ilr** in einen \mathbb{R}^{D-1} mit orthonormalen Basis überführt wird.

Egozcue et al. (2003) zeigen, dass sich eine beliebige **ilr**-Transformation mittels einer speziellen $(D-1) \times D$ -dimensionalen Matrix **U** aus der **clr**-Transformation bilden lässt:

$$\mathbf{ilr}(\vec{x}) = \mathbf{U} \cdot \mathbf{clr}(\vec{x})$$

Jede orthonormale Rotation einer **ilr**-Transformation mittels einer $(D-1) \times (D-1)$ -dimensionalen orthonormalen Matrix **A**

$$\mathbf{ilr}_A(\vec{x}) = \mathbf{U}_A \cdot \mathbf{clr}(\vec{x}) = \mathbf{A} \cdot \mathbf{U} \cdot \mathbf{clr}(\vec{x})$$

ist selbst wieder **ilr**-Transformation.

In der Klasse der **alr**-Transformationen kann jede Komponente x_i von \vec{x} die Rolle der Referenzkomponente einnehmen.

Jede konkret ausgewählte Transformation aus der Klasse der **ilr**- bzw. **alr**-Transformation ist wie die **clr**-Transformation ein-eindeutig: Zu jeder der Transformationen existiert eine inverse Transformation vom \mathbb{R}^{D-1} zurück in das Simplex S^D (bzw. die jeweilige Äquivalenzklasse, vgl. Ulbrich, 2010).

Statistische Analysen von Kompositionsdaten – von Stichprobendeskription mit dem Zentrum der Verteilung und Variabilitätsmaßen bis hin zu Hypothesentests und statistischer Modellierung – können (auf den transformierten Daten) im Koordinatenraum mit Hilfe der (üblichen) multivariaten Verfahren durchgeführt, Modell-Prädiktionen mittels der Invers-Transformationen in das Simplex des Stichprobenraumes zurückgeführt werden. Im Stichprobenraum des Simplex

selbst ist jede Modellierung (z. B. Regression) eine *limited-dependent-variable technique* im Sinne von Abschn. 5.1.5.

Eine Subkomposition einer Komposition $\vec{x} \in S^D$ besteht aus d ($1 < d \leq D$) Komponenten, während Amalgamation die additive Zusammenfassung zweier (oder mehrerer) Komponenten einer Komposition bedeutet (Aitchison, 1986, 2003, Kap. 2.5 bzw. 2.6). In beiden Fällen sind die Ergebnisse Kompositionen in einem Simplex S^d ($1 \leq d \leq D$). Eine wichtige Eigenschaft dabei ist *subcompositional coherence* (Pawlowsky-Glahn et al., 2007, Kap. 2.2.3): die Subkomposition entspricht einer Projektion in einen niedriger-dimensionalen Teilraum des ursprünglichen Simplex, der Abstand zweier projizierter Punkte ist (nicht negativ und) nicht größer als der Abstand dieser Punkte im ursprünglichen Raum.

Obwohl bereits Aitchison (1986) (mikro-)ökonometrische Kompositionsdaten (Anteile der Ausgaben der Privathaushalte in Hongkong für verschiedene Bedarfsgütergruppen) untersucht, ist die Kompositionsdatenanalyse in der ökonometrischen Literatur eher selten zu finden. Beispiele sind Fry et al. (2001), die sich der Modellierung von Mikrodaten mit 0-Werten widmen, und Graf (2006) mit einer Analyse zu den Anteilen am Einkommen der Schweizer Erwerbsbevölkerung. Weiterentwicklungen der Methode fanden zu einem großen Teil mit und für Anwendungen in der Geologie statt (Aitchison und Egozcue, 2005, und viele weitere Artikel in *Mathematical Geology / Mathematical Geoscience*).

Die aktuell umfassendste Darstellung zum Stand der Entwicklung der Kompositionsdatenanalyse findet sich in Pawlowsky-Glahn et al. (2007); kurze Einführungen in deutscher Sprache geben Pawlowsky-Glahn und Egozcue (2007) sowie Ulbrich (2010). Filzmoser et al. (2009) zeigen – quasi als Weiterführung der Überlegungen von Pearson (1897) – die Bedeutung der Kompositionsdatenanalyse für die Fälle, wenn nur eine Teilmenge der Komponenten einer konstanten Summe analysiert werden sollen.

Die lineare Logit-Regression

$$y = \text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon, \quad \varepsilon \sim N(0, 1),$$

genutzt über $p = \left(\frac{e^y}{1+e^y}\right)$ zur Modellierung der Quote p , kann als Spezialfall der Kompositionsdatenanalyse für den Simplex S^2 (in \mathbb{R}^2) mit Konstantsumme von 1 (100%) angesehen werden: die **alr**-Transformation vereinfacht sich zur logistischen Transformation $y = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$, die Korrelation zwischen p und $(1-p)$ ist $\varrho(p, 1-p) = -1$ und somit selbst konstant.

5.1.12. Visualisierung von Kompositionsdaten

Adäquate grafische Darstellungen von Kompositionsdaten¹⁸⁰ sollten widerspiegeln, dass es sich um Datenpunkte in einem Simplex S^D handelt. Darstellungen im unbeschränkten \mathbb{R}^D erscheinen hier generell nicht als erstrebenswerte Lösung.

Für den Zwei-Komponenten-Fall S^2 verstanden als Raten existieren etablierte Darstellungsformen, z. B. gestapelte Säulen- oder Balkendiagramme und Kurven der logistischen Regression. Beide Komponenten ergänzen sich zu Balken konstanter Länge und werden darstellungstechnisch gleichwertig behandelt; tauscht man die Komponenten, ergibt sich ein Diagramm welches sich symmetrisch von ersterem unterscheidet.

Für Kompositionsdaten von 3 oder mehr Komponenten dargestellt im Balkendiagramm gilt diese Symmetrie nicht.

Kompositionen dreier Komponenten lassen sich in Dreiecksdiagrammen (engl.: *ternary plot*) visualisieren (z. B. Egozcue und Pawlowsky-Glahn, 2006), die gleichseitigen Dreiecke garantieren eine visuelle Gleichwertigkeit der Komponenten bei Drehungen um 120° . Möchte man die Darstellung von Kompositionen dreier Komponenten um eine zusätzliche (Einfluss-)Variable auf kategorialer, ordinaler oder metrischer Skala erweitern, so ist bereits das nur sehr eingeschränkt möglich. Aitchison (1986, Fig. 1.3) zeigt eine perspektivische Abfolge von Dreiecksdiagrammen, die man sich auch als eine Sequenz nebeneinanderliegender Diagramme vorstellen kann – in beiden Fällen sind kleine Veränderungen kaum und in ihrer Richtung nicht sicher wahrnehmbar.

¹⁸⁰ Eine ausführlichere Darstellung mit Beispielen findet man in Ulbrich (2010).

Für Komponentenzahlen größer als 3 nehmen die Schwierigkeiten beim Versuch der Visualisierung zu. Während für vier-komponentige Kompositionen eine dynamisch-räumliche Darstellung des Tetraeders als Abbild des Simplex S^4 zumindest vorstellbar erscheint, ist dieser Weg für Komponentenzahlen von 5 und darüber nicht gangbar.

Einzel-Kompositionen können jederzeit mit einem Kuchendiagramm veranschaulicht werden, für eine Sequenz von Kuchendiagrammen gilt jedoch das für eine Sequenz nebeneinanderliegender Dreiecksdiagramme bereits gesagte: kleine Veränderungen sind kaum und in ihrer Richtung nicht sicher wahrnehmbar. Außerdem ist in einer Folge von Kompositions-Kuchendiagrammen immer nur eine Richtung (die 0°-Achse) fixierbar, nur die beiden anliegenden Komponenten haben im Diagramm eine von den anderen Komponenten unbeeinflusste Position. Somit sind, obwohl kreisrund, Kuchendiagramme in einer Sequenz bezüglich der einzelnen Komponenten nicht gleichwertig.

Häufiger als Kuchendiagramme werden (Sequenzen von) gestapelten Säulen- oder Balkendiagrammen zur Veranschaulichung von Kompositionsdaten herangezogen. Hier ist ebenfalls bereits ab 3 Komponenten eine völlige Gleichbehandlung der Komponenten nicht mehr möglich. Abb. 5.2 zeigt das Problem anhand der Daten aus Tabelle 5.1: sechs Kompositionen zu 100 %, dabei seien z1 bis z6 beliebige unterschiedliche Ereignisse, möglicherweise über ein metrisches Merkmal miteinander verknüpft. Die Teilbilder unterscheiden sich nur in der Ordnung der

Tab. 5.1.: Kompositionen zu 100 % (Beispieldaten)

#	Komponenten [%]				
z1	12,0	1,4	3,5	20,0	63,1
z2	15,0	2,2	3,5	17,0	62,3
z3	20,0	2,8	3,5	11,0	62,7
z4	17,0	2,6	3,5	13,0	63,9
z5	15,0	2,5	3,5	12,0	67,0
z6	11,0	2,3	3,5	19,0	64,2

Komponenten zueinander. Teilbild (a) ist zu entnehmen, dass der Anteil der größten Komponente (5) ein wenig um 63 % mit einem Maximum bei z5 schwankt; Komponenten 2 und 3 sind in ihrer jeweiligen Veränderung kaum beurteilbar. Das

ändert sich für Komponente 2 in Teilbild (b) durch den Positionswechsel an den (linken) Rand des Diagramms – hier zuungunsten der Sichtbarkeit der Komponente-5-Veränderungen. Jedoch zeigt erst Teilbild (c), dass die Werte für Komponente 3 keinerlei Veränderungen unterliegen. Mit dem Anstieg der Anzahl der Komponenten vergrößert sich das Problem, dass nur die randständigen Komponenten wegen ihrer fixierten Lage gut hinsichtlich ihrer Veränderungen beurteilbar sind. Handelt es sich bei den Komponenten um ein System mit innerer Rangordnung von Komponente 1 ‚klein‘ zu Komponente 5 ‚groß‘ verletzt ein Positionswechsel ziemlich die Forderung nach Klarheit eines Diagramms.

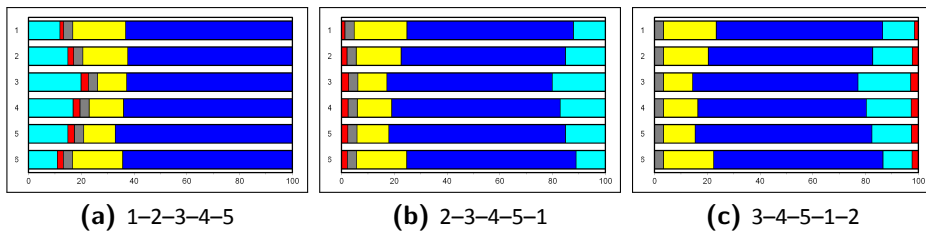


Abb. 5.2.: Kompositionen zu 100 % (Tab. 5.1)
bei unterschiedlicher Reihenfolge der Komponenten

Alternativ zum gestapelten Säulen- oder Balkendiagramm findet man vielfach die überlagerte Darstellung einer jeden Komponente wie in Abb. 5.3 für den Beispieldatensatz. Handelt es sich bei z1 bis z6 um beliebige Kategorien zur Identifikation, so können die Verbindungen lediglich als Hilfe für die optische Zuordnung verstanden werden. Häufiger handelt es sich bei z1 bis z6 jedoch um Werte eines metrischen Merkmals (z. B. in äquidistantem Abstand), geeignete Modellierung wie z. B. in Bucciante und Pawlowsky-Glahn (2005, Fig. 12) führt dann zu glatten Funktionen als Anpassung an die Daten.

Abb. 5.3 zeigt deutlich, dass Komponente 5 wenig Veränderung erfährt, Komponenten 1 und 4 zeigen gegenläufiges Verhalten. Komponenten 2 und 3 werden kaum wahrgenommen, obwohl Komponente 2 ihren Anteil zwischenzeitlich ‚verdoppelt‘ (von 1.4 % auf 2.8 %) – natürlich auf ‚Kosten‘ aller anderen, deren gemeinsamer Anteil von 98.6 % auf 97.2 % sinkt.

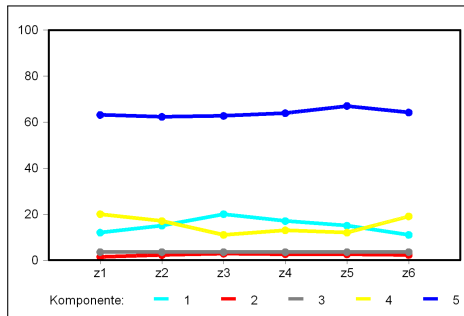


Abb. 5.3.: Kompositionen zu 100 % (Tab. 5.1)
[Verbindungen zwischen den Punkten für bessere Lesbarkeit]

5.1.13. Das Kompositionsabweichungsdiagramm – eine neue Darstellungsmöglichkeit für Kompositionsdatenanalysen

Grafische Darstellung von Kompositionsdaten in einer Art, dass jede der Komponenten gleich behandelt wird, ist somit für bis zu 3 Komponenten möglich. Da nicht jede auf Kompositionsdaten basierende Fragestellung durch entweder Amalgamation auf 3 Komponenten oder durch (sukzessive) Analyse von 3er Subkompositionen ausreichend gut beantwortet werden kann, bleibt die Frage nach Darstellungen für Kompositionsdaten höherer Komponentenanzahl.

Der Autor vorliegender Arbeit schlägt dafür erstmals in Ulbrich (2010) ein Diagramm auf Basis der **clr**-Transformation vor

$$\mathbf{clr}(\vec{x}) = \mathbf{clr} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{pmatrix} = \frac{1}{D} \cdot \begin{pmatrix} D-1 & -1 & \cdots & -1 \\ -1 & D-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & D-1 \end{pmatrix} \cdot \ln(\vec{x});$$

und nennt dieses Kompositionsabweichungsdiagramm (engl.: *compositional deviation chart*).

Das Kompositionsabweichungsdiagramm nutzt die Eigenschaften der **clr**-Transformation: je Komponente aus S^D gibt es eine Koordinate im \mathbb{R}^D und diese Ko-

ordinaten-Werte addieren sich zu 0. Die Koordinaten-Werte liegen auf der durch $\mathbf{1}' \cdot \mathbf{clr}(\vec{x}) = 0$ mit dem D -dimensionalen Einsvektor $\mathbf{1} = (1, \dots, 1)'$ spezifizierten Hyperebene der Dimension $(D-1)$. Das Diagramm zeigt für jede Kompositionen die \mathbf{clr} -transformierten Werte einer jeden Komponenten auf der Ordinate, während die Abszisse Raum lässt für die Anordnung mehrere Kompositionen über einem interessanten Einflussmerkmal, sei dieses eine (Einfluss-)Variable auf kategoriel-ler, ordinaler oder metrischer Skala. Der ‚Verlauf‘ der einzelnen Komponenten über dem Einflussmerkmal lässt sich leicht hinsichtlich zweier Eigenschaften be-gutachten: der mittleren Lage bezogen auf das (jeweilige) geometrische Mittel und der Veränderungen (Anstieg, Abfall oder ‚Konstanz‘) gegenüber den anderen Komponenten. Abb. 5.4 zeigt, dass Komponente 5 die anteilmäßig bedeutendste

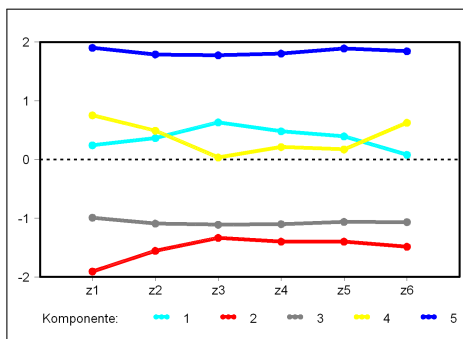


Abb. 5.4.: Kompositionsabweichungsdiagramm (zu Tab. 5.1)

[Verbindungen zwischen den Punkten für bessere Lesbarkeit]

Komponente ist und dass die Veränderungen in Komponente 5 relativ unbedeutend scheinen gegenüber dem deutlichen Anstieg Komponente 2 zwischen z_1 und z_2 . Die Veränderungen in Komponente 2 zeigen sich so bedeutsam, wie die in den Komponenten 1 und 4. Bemerkenswert ist auch, dass die ‚Konstanz‘ der Werte der Komponente 3 im Bezug auf die anderen Komponenten keine Konstanz ist.

Was auch immer man Abb. 5.3 entnommen hat, das Kompositionsabweichungsdiagramm in Abb. 5.4 zeigt, was die Kompositionsdatenanalyse analysiert: in einen linearen Raum der Dimension $D-1$ transformierte Punkte des Simplex S^D .

Da sich, wie in Abschn. 5.1.11 ausgeführt, jede Transformation aus der Klasse der \mathbf{ilr} -Transformationen als orthonormale Drehung der \mathbf{clr} -Transformations-Hyperebene $\mathbf{1}' \cdot \mathbf{clr}(\vec{x}) = 0$ erweist, spiegeln die \mathbf{clr} -Koordinaten die Kompositionen

im Simplex genauso unverzerrt wider wie die **ilr**-Koordinaten. Auf diesen Gedanken aufbauend lassen sich nach Modellierungen zum jeweiligen Kompositionsabweichungsdiagramm Residuendiagramme ableiten – der Informationsfülle wegen ggf. für jede Komponente separat gezeichnet.

Nachteile des Kompositionsabweichungsdiagramms könnten in den Augen des (substanzwissenschaftlichen) Betrachters vor allem darin liegen, dass transformierte Daten auf arbiträr erscheinender Skala dargestellt werden. Dennoch, das menschliche Auge sieht (linear) additiv, die Situation im Simplex kann wegen der Konstanzsummeneigenschaft additiv nicht sein. Datendeskription und -Modellierung im Koordinatenraum lassen jedoch sinnvollerweise additiv gestalten, das Kompositionsabweichungsdiagramm kann hilfreich sein bei der adäquaten Darstellung der den Daten innewohnenden Beziehungen, die datengesteuert willkürlich erscheinende Skalierung der Ordinate nimmt man in Kauf.

5.2. Programmtechniken und Algorithmen am GWAP und unter KDFV

Arbeiten mit amtlichen Daten dieses Umfangs erfordert intensive – auch zeitintensive – Einarbeitung in Datenumfang und Datenstrukturen. Das gewählte zweistufige Vorgehen – Arbeit am GWAP und zunehmend Umsetzung der erarbeiteten Auswertestrategien unter Nutzung der KDFV geht auf eine intensive Diskussion des Autors mit dem FDZ/L (Schoffer, 2005, 2006) zurück.

Effizienz und Transparenz liegen dabei im Interesse beider, des Auswertenden sowie des Betreuenden vom FDZ/L. Transparenz-Forderungen des FDZ/L betreffen die klare Identifikation und Dokumentation wiederholter Analyseläufe sowohl am GWAP als auch unter KDFV.

Gemäß Vertrag (Forschungsdatenzentrum der Statistischen Landesämter, 2007) basieren alle softwarebasierten Arbeitsschritte und -ergebnisse auf dem Statistik-Programmsystem SAS® in der Version 9.1.3¹⁸¹.

¹⁸¹ SAS Institute Inc., SAS Campus Drive, Cary, NC 27513, USA, <http://www.sas.com/software/sas9/>.

5.2.1. Modularer Aufbau

Da alle KDFV-Auswertungen von FDZ-Mitarbeitern durchgeführt werden und durch das Abarbeiten von vom Wissenschaftler erarbeiteten Software-Skripten (*Syntax-Skripte*, Zühlke et al., 2003, S. 910) entstehen, bestimmen klar strukturierte Skripte mit eindeutig definiertem Beginn und Ende ganz wesentlich die Güte der Zusammenarbeit zwischen dem FDZ und dem Wissenschaftler.

Für den nutzenden Wissenschaftler ist die KDFV-Situation durchaus mit einer *Black Box* vergleichbar, jeder Programmierfehler droht einen das Skript abarbeitenden Analyselauf ergebnislos oder fehlerhaft-nutzlos enden zu lassen. Gelingt es vorab, sich am GWAP von der Fehlerfreiheit des Skripts zu überzeugen, steigt die Chance auf einen erfolgreichen Analyselauf unter KDFV.

Für einen modularen Aufbau sprechen bei der Komplexität der geplanten Auswertungen eine Vielzahl von bekannten Gründen, wie Übersichtlichkeit und Testbarkeit der Einzelmodule. Bei wiederholten Analyseläufen können im weiteren Verlauf bereits erfolgreiche Teilschritte genauso von der Wiederholung ausgenommen werden wie besonders zeitintensive.

5.2.2. Mantelskripte

Für vorliegende Arbeit ist ein Mantelskript `_start_.sas` als SAS®-Steuerung in zwei Varianten programmiert worden, einer für alle Analyseläufe des Autors direkt am GWAP in den Räumen des FDZ/L, Standort Sachsen in Kamenz, und die zweite für die Berechnungen unter den Bedingungen der KDFV. Beide haben denselben logischen Aufbau (Abb. 5.5) und teilen lange Passagen des Programmcodes (Anhang A.2). Sie unterscheiden sich hauptsächlich hinsichtlich der gesetzten Referenzvariablen (zur Identifikation des GWAP/KDFV-Unterschiedes und für die jeweiligen Verzeichniszuordnungen) und der Aufrufe der GWAP- bzw. KDFV-spezifischen SAS®-Steuerungen zum Einlesen – und Korrigieren – der zugehörigen Daten.

Alle darauffolgenden Passagen des Mantelskripts sind in beiden Varianten nahezu identisch, sie organisieren – wie in Kapiteln geordnet – den Aufruf der einzelnen

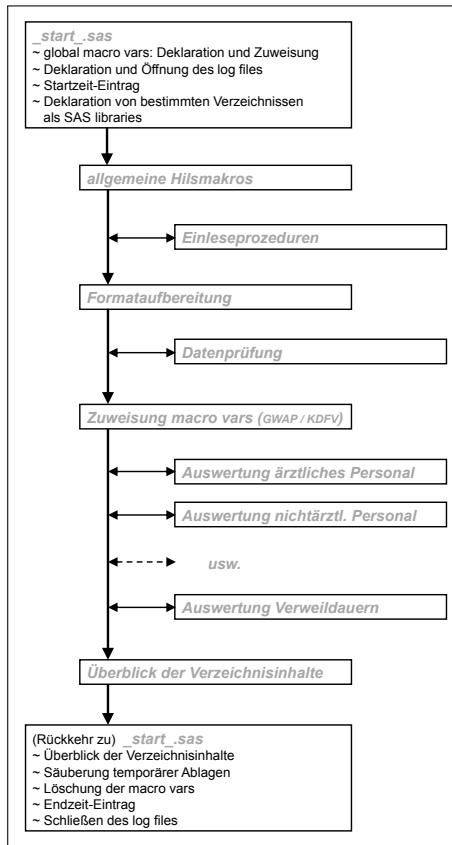


Abb. 5.5.: Programmablauf der SAS®-Skripte aus dem Mantelskript `_start_.sas` heraus

Evaluierungsschritte, die selbst in separaten Dateien als SAS®-Steuerungsskripte vorliegen.

Jeder Analyselauf am GWAP als auch unter KDFV-Bedingungen bedient sich seiner Mantelskript-Variante, ein Analysieren der Daten ohne Mantelskript findet nicht statt. Die Uhrzeiten von Anfang und Ende eines jeden Laufes sowie aufgerufene Einzelskripte werden – neben anderem – in der Logdatei des Laufes dokumentiert.

5.2.3. Einzelskripte

Jedes Einzelskript wird vom Mantelskript `_start_.sas` aufgerufen und übergibt an seinem Ende wieder an das aufrufende Mantelskript. Das Einzelskript beginnt mit der Zuordnung einer individuellen Logdatei und zweier Ergebnisdateien – eine für die Übergabe an den Autor, eine (im rtf-Format) zur einfacheren Kontrolle durch das FDZ/L. Log- und Ergebnisdateien erhalten einen jeweils aktuellen Zeitsempel (Datum und sekundengenaue Urzeit); für die Logdatei wird dieser Teil des Dateinamens. Ein Überschreiben der Logdatei durch einen darauffolgenden Analyselauf ist somit nicht möglich, das FDZ kann jederzeit nachvollziehen, was gerechnet wurde.

Am Ende eines jeden Einzelskripts werden alle temporären Dateien und Variablenlisten gelöscht. Dieses Vorgehen ist die beste Versicherung gegen unbemerktes Beeinflussen der Schritte aus nachfolgenden Einzelskripten.

Die Einzelskripte für das Dateneinlesen enthalten jeweils vom FDZ/L erstellte Subskripte zur Datenübernahme und zur Übernahme der verwendeten Datencodes. Ergänzt wird das Dateneinlesen um die notwendigen Schritte der Code-Angleichung; so werden hier alle Diagnosen von ICD-10 nach ICD-9 unter Nutzung der Konkordanzliste (Überleitungstabelle, Abschn. 4.2.4) rekodiert, die bundesland-spezifischen Präferenzen für 0-Wert- oder Fehlwert-Angaben vereinheitlicht.

Nach der Abarbeitung der Skripte für das Dateneinlesen liegen die Daten permanent im SAS®-eigenen Dateiformat vor, auf sie kann unmittelbar oder in späteren Analyseläufen ohne erneutes Dateneinlesen zurückgegriffen werden. Für überschaubare Rechenzeiten erweist es sich als gut, dass die Daten der Diagnosestatistik je Erfassungsjahr (1993–2005) separat im SAS®-komprimierten Dateiformat gehalten werden. Ergebnisse der Datencheck-Skripte fließen in das letzte Skript zur Datenaufbereitung ein: Korrektur.

Neben den GWAP/KDFV-spezifischen Einzelskripten für das Dateneinlesen gibt es Analyseskripte (gegebenenfalls mit Subskripten) zur Analyse von Grund- und Kostendaten der Krankenhäuser sowie der kumulierten Krankenhäuser der Kreise, zur Analyse des ärztlichen sowie des nichtärztlichen Personals. Zwei Einzelskripte widmen sich der Diagnosestatistik, speziell den Diagnoseverteilungen und den Verweildauern der Fälle im Krankenhaus.

Diese Analyseskripte werden unverändert für GWAP- und KDFV-Analyseläufe genutzt. Fallunterscheidungen nach GWAP bzw. KDFV finden in Ausnahmefällen innerhalb der Skripte statt. So ist z. B. die Darstellung von Choroplethenkarten für die GWAP-Situation unnützlich, denn die Klumpen bestehen im allgemeinen nicht aus geografisch zusammenhängenden Gebieten, eine Kartierung widerspricht somit der Anonymisierungsforderungen nach der Nichtidentifizierbarkeit einzelner Kreise.

5.2.4. Makros

Wiederholt auszuführende Teilschritte der Datenmodifikation, -umstrukturierung, -deskription, -modellierung und -ausgabe sowie der grafischen Darstellung sind als Makros (SAS Institute Inc., 2004c) programmiert. Auf eine klare Trennung zwischen dem Inneren eines Makros und seiner Umgebung wird dabei im Interesse der sicheren Anwendung Wert gelegt. Dazu ist es in SAS[®] nötig, lokale Makrovariablen als solche zu deklarieren, denn nur so können diese vom System als verschieden von (zufällig) gleichnamigen Makrovariablen in der Umgebung erkannt werden. Mit dem Verlassen eines Makros sind die lokalen Makrovariablen nicht länger existent. Temporär innerhalb eines Makros entstehende Dateien werden makrospezifisch benannt und vor dem Austritt aus dem Makro gelöscht.

5.2.5. Register, SAS[®]-Datensätze und Skripte

Alle Auswertungen beruhen auf dem Register (Abschn. 4.4). Das Register wird mit den Schritten Einlesen der Daten der Krankenhausstatistik, Datenmodifikation einschließlich -korrektur, Datentransformation und Datenergänzung um die extern kompilierten Daten der kreisbezogenen Wirtschafts- und Raumordnungsmerkmale erstellt. Das Register wird in Dateien im SAS[®]-Datendateiformat festgehalten, die Schritte der Registererstellung müssen somit nicht bei jedem Analyselauf durchlaufen werden.

Das Register hat die Struktur einer relationalen Datenbank (Codd, 1970, 1982) und besteht (in der KDFV-Version) aus 7 verschiedenen Typen von Relationen (Tabellen, realisiert in SAS®-Datendateien):

- (a) einer Tabelle mit einer Zeile je vorhandener Krankenhausidentifikation mit zeitunveränderlichen Angaben (z. B. Kreis des Sitzes des Krankenhauses) und zum zeitlichen Muster vorliegender Daten für dieses Krankenhaus (723 Zeilen),
- (b) einer Tabelle mit jährlichen Angaben zu jedem Krankenhaus – Grund- und Kostendaten (7.320 Zeilen),
- (c) einer Tabelle mit jährlichen Angaben zu jeder Fachabteilung – Betten, Belegungstage, Patientenbewegungen (56.859 Zeilen),
- (d) einer Tabelle mit jährlichen Angaben zum ärztlichen Personal für jede Gebiets- / Teilgebietsbezeichnung – Anzahl, Geschlecht, Position (92.292 Zeilen),
- (e) einer Tabelle mit jährlichen Angaben zum nichtärztlichen Personal für jede Personalgruppe – Anzahl, Geschlecht (238.460 Zeilen),
- (f) 13 Jahrestabellen der Diagnosen (jeweils zwischen 3,3 und 4,4 Mio. Zeilen, im SAS®-komprimierten Dateiformat gehalten) und
- (g) einer Tabelle mit den kreisbezogenen Zusatzdaten (2.205 Zeilen für 15 Jahre, 147 Kreise).

Schlüssel für die Beziehungen zwischen den Tabellen sind die Identifikatoren des Registers (Abschn. 4.4) mit (kontextfreier) Krankenhausnummer, Kreiszuordnung und Berichtsjahr als den wichtigsten.

Auf dieses Register¹⁸² greifen alle Analyseskripte zu.

5.2.6. SAS®-Module und -Zusätze

SAS® als Software ist selbst modular aufgebaut. Base SAS® – das nötige Kernsystem – kann nutzerspezifisch um Module ergänzt werden, von SAS/GRAPH® über SAS/STAT® (SAS Institute Inc., 2004b) bis hin zu SAS/ETS® (SAS Institute Inc., 2004a), dem Modul für Zeitreihen und weitere ökonomische Verfahren. Im

¹⁸² Für zeiteffizientere Analyseschritte werden spezielle Modifikationen der 13 Jahrestabellen ebenfalls permanent vorgehalten.

FDZ/L, Standort Kamenz, stehen diese Module neben anderen zur Verfügung. Das Fehlen des Moduls SAS/QC® lässt sich für vorliegendes Projekt programmieretechnisch kompensieren.

Einzelne Datenanalyseschritte sind bei SAS® in Prozeduren (`proc`) gebündelt, mit `proc freq` lassen sich z. B. Kreuztabellen (Kontingenztafeln) auswerten. Neben den zur jeweiligen Version gehörenden Modulen bietet der Hersteller SAS Institute Inc. gegebenenfalls Vorabversionen zukünftiger Module oder Prozeduren unter Vorbehalt an. Für dieses Projekt werden `proc panel` (SAS Institute Inc., 2005) und `proc glimmix` (SAS Institute Inc., 2006) in Vorabversionen genutzt.

5.2.7. Anonymisierungsforderungen und Ergebnisdarstellungen

Minima und Maxima einer Gruppe von Werten oder einer empirischen Verteilung stellen Einzelwerte dar und unterliegen als solche dem Schutz durch die Anonymisierungsregeln. Werte unterhalb des 1. und oberhalb des 99. Perzentils werden deshalb bei Datendeskriptionen automatisch von der Ausgabe ausgeschlossen.

Für Zähldaten sind aus demselben Grund Angaben von 1, 2 oder 3 gemäß Dreier-Regel (Abschn. 4.1.2) nur gemeinsam in einer Kategorie zulässig, deren Auftreten bei der Auswertung unter KDFV-Bedingungen jedoch a priori unbekannt. Per `format`-Definition wird in den SAS®-Steuerungen gesichert, dass diese Werte als ‚1-3‘ maskiert in Auflistungen erscheinen; Summierungen sind, da sie einen Rückschluss auf die maskierten Einzelwerte zulassen könnten, folglich ausgeschlossen.

Zum Schutz der Anonymität der einzelnen Krankenhäuser erscheinen in den Quintil-Choroplethenkarten Kreise mit weniger als 3 Krankenhäusern immer ungefärbt, obwohl diese Kreise in die Berechnungen der Quintilswerte eingehen. Minima und Maxima werden auch hier nicht angegeben.

5.3. Methoden und Umsetzungen – Zusammenfassung

Ökonometrische Analysen vorliegender Paneldaten erfolgen hauptsächlich mit Regressionsmodellen unter Beachtung der longitudinalen Struktur der Daten. Panelsubjekte sind einerseits die einzelnen Krankenhäuser als auch die Kreise mit der kumulativen Krankenhausleistung als Gesamtkrankenhausaktivität. Die Regressionen als Anwendung des verallgemeinerten linearen Modells (einschließlich Modelle für *limited dependent variables*) sind überwiegend linear in den Regressoren, die Panelstruktur wird durch Kombinationen von festen und zufälligen Effekten und/oder Varianzkomponenten modelliert. Ergänzend werden Kompositionsdaten nach geeigneter Transformation mit multivariaten Regressionen und Zeiten bis zu einem Ereignis wie der Entlassung aus dem Krankenhaus mit Methoden der Ereigniszeitanalyse untersucht. In allen Fällen gilt es, Verzerrungen, Multikollinearitäten und numerische Instabilitäten zu vermeiden. Mit dem Kompositionsabweichungsdiagramm wird ein neuer Graph zur Visualisierung der Veränderungen von Kompositionsdaten vorgeschlagen.

Die Analysen erfolgen in zusammenhängenden Rechenläufen unter SAS® beginnend mit der Datenübernahme. Jeder Rechenlauf wird aus mehreren Rechner-Skripten für Einzelaufgaben zusammengestellt, ein Mantelskript – gleichartig für Läufe unter GWAP und KDFV – verbindet die jeweiligen Einzelaufgaben. Bei Wiederholungen von Rechenläufen können durch Modifikation des Mantelskripts Einzelaufgaben ausgelassen werden. Die Datenstruktur entspricht (weitgehend) dem relationalen Datenmodell.

Die logischen Schritte zur (sequenziellen) Lösung der Einzelaufgaben sind auch in anderen Programmsystemen wie Stata¹⁸³, SPSS¹⁸⁴ oder der open source Software R¹⁸⁵ umsetzbar. Ob diese Systeme so flexibel wie SAS® mit seiner Makrosprache programmierbar sind und wie diese Systeme speziell mit großen (Diagnose-) Datensätzen umgehen, kann hier nicht erörtert werden.

¹⁸³ <http://www.stata.com/>.

¹⁸⁴ <http://www.spss.com/>.

¹⁸⁵ <http://www.r-project.org/>.

Kapitel 6.

Empirische Analysen: Ergebnisse

Ökonometrisches Modellieren bedient sich der Techniken der deskriptiven, explorativen und konfirmatorischen Statistik (Abschn. 3.4.3).¹⁸⁶ Erster Schritt ist dabei die Beschreibung vorliegender Daten.

Sofern nicht anders erwähnt, beziehen sich alle Aussagen auf die Analysen unter den Bedingungen der KDFV. Die Aussagen betreffen das Untersuchungsgebiet – die ostdeutschen Bundesländer Berlin¹⁸⁷, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen sowie das westdeutsche Bundesland Rheinland-Pfalz – für die Jahre 1992 bis 2005 als Untersuchungszeitraum. Das Untersuchungsgebiet wird für den gesamten Untersuchungszeitraum mit seinen Kreisen zum Gebietsstand 31. 12. 2005 als Referenzzeitpunkt betrachtet (Abschn. 4.2.4: Administrative Zuordnungen), wobei die erst spät vom Umland administrativ separierten kreisfreien Städte Hoyerswerda (Sachsen, 1996) und Eisenach (Thüringen, 1998) als Teil der sie umgebenden Landkreise Kamenz bzw. Wartburgkreis (Abschn. 4.4.1: Zuordnungen auf den Gebietsstand 2005) verstanden werden. Somit ergeben sich 147 Kreise: 36 Kreisen in Rheinland-Pfalz stehen 111 Kreise (Brandenburg 18, Mecklenburg-Vorpommern 18, Sachsen 28, Sachsen-Anhalt 24, Thüringen 22 und Berlin 1) in Ostdeutschland gegenüber.

Alle in Geldeinheiten ausgedrückte Merkmale sind auf das Jahr 2000 (= 100 %) deflationiert (Abschn. 4.4.3: Deflationierung der Kostendaten). Für eine bessere numerische Stabilität gehen statt der Jahreszahlen deren Abstände zum Jahr 2000¹⁸⁸ als Zeitvariable $jahr$ in die Modellrechnungen ein.

¹⁸⁶ Die Daten-Qualitätskontrolle bedient sich vorab oft ähnlicher Techniken (Abschn. 4.3.2 f.).

¹⁸⁷ siehe Abschn. 2.4.

¹⁸⁸ 1992 = -8, ..., 2000 = 0, ..., 2005 = 5.

6.1. Deskriptionen

Beschreibungen mittels statistischer Maßzahlen geben einen ersten, meist univariaten Einblick in die vorliegenden Daten. Einwohner, Wirtschaftskraft, soziale Lage und ausgewählte Kenngrößen der Krankenhäuser sind hier von besonderem Interesse. Es wird versucht, dabei der longitudinalen bzw. Panelstruktur der Daten zu entsprechen.

6.1.1. Demografie – die Bevölkerung

Die Bevölkerung Deutschlands wächst im Untersuchungszeitraum, doch sind die ostdeutschen Länder davon ausgenommen (Abschn. 2 und Abb. 2.1). So wächst die Bevölkerung von Rheinland-Pfalz von 3,85 Mio. 1992 streng monoton auf 4,06 Mio. Einwohner (+ 5,5 %) im Jahr 2005 an. Während Berlin (von 3,45 auf 3,39 Mio.) und Brandenburg (2,54 auf 2,56 Mio.) 2005 jeweils ungefähr so viele Einwohner haben wie 13 Jahre zuvor, sinkt die Einwohnerzahl aller anderen ostdeutschen Länder stetig: Mecklenburg-Vorpommern von 1,87 auf 1,71 Mio., Sachsen von 4,66 auf 4,28 Mio., Sachsen-Anhalt von 2,81 auf 2,48 Mio. und Thüringen von 2,55 auf 2,35 Mio. Einwohner. Für Ostdeutschland bedeutet dies für die Jahre 1992 bis 2005 einen Negativsaldo der Einwohnerzahl von 1,12 Mio. oder 6,25 %.

Die Einwohnerzahlen der Kreise liegen dabei zwischen 35.000 (Stadt Zweibrücken) und 210.000 (Kreis Mayen-Koblenz) mit dem Median bei 108.000 für Rheinland-Pfalz, zwischen 43.000 (Stadt Suhl) und 540.000 (Stadt Leipzig) mit 115.000 als Median für die ostdeutschen Flächenländer. Berlin als Bundesland ohne Kreisstruktur mit seinen 3,4 Mio. Einwohnern ist somit sechs mal so groß wie der nächstkleinere Kreis.

Die Beschreibung der Urbanität der Kreise kann sowohl auf der Einwohnerdichte als auch dem Kreistyp der Raumordnungsmerkmale (Abschn. 3.3.5) erfolgen. Stellt man eine Landkarte mit dem Kreistyp und eine Quintilkarte (Abschn. 5.1.10) gegenüber, lässt sich die Bedeutung der Kreise mit höchster Einwohnerdichte als Zentren der Urbanität erkennen (Abb. 6.1 und 4.1).

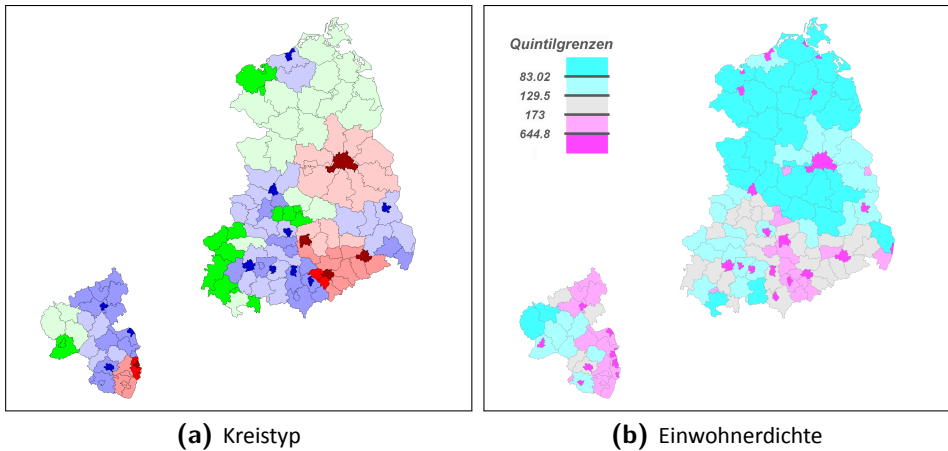


Abb. 6.1.: Urbanität der Kreise des Untersuchungsgebiets (2005):

(a) BBR-Klassifikation Kreistyp

Agglomerationsräume (rot), Verstädterte (blau) und Ländliche Räume (grün), hellere Farbtöne (4 rot, 3 blau, 2 grün) für jeweils geringeren Verdichtungsgrad

(b) Einwohnerdichte [EW/km², Quintilkarte]

Anders als in Rheinland-Pfalz gilt für jedes der ostdeutschen Flächenländer, dass besonders die Kreise mit niedrigerer Einwohnerdichte an Einwohnern verlieren: sowohl Minimum, 1. Dezil als auch 1. Quartil – letzteres mit Ausnahme Brandenburgs – der empirischen Verteilung nehmen stetig ab. So sind es eher die ländlich dünn besiedelten Kreise, deren Bevölkerung in besonderem Maße schrumpft.

Im Gegensatz zu den Krankenhausfällen (Tab. 6.5) liegen für die Wohnbevölkerung der Kreise¹⁸⁹ keine Angaben zu den potenziellen Störvariablen (*confounder*) Geschlecht und Alter vor.

6.1.2. Wirtschafts- und Sozialparameter

Arbeitslosenquoten der Kreise liegen für die Jahre 1996 bis 2005 vor. Abb. 6.2 zeigt mit Quintilkarten die Veränderungen im Vierjahresabstand ergänzt um die Daten für 2005, dem ersten Jahr der mit Hartz IV-Gesetz veränderten Defi-

¹⁸⁹ im Rahmen diese Projektes.

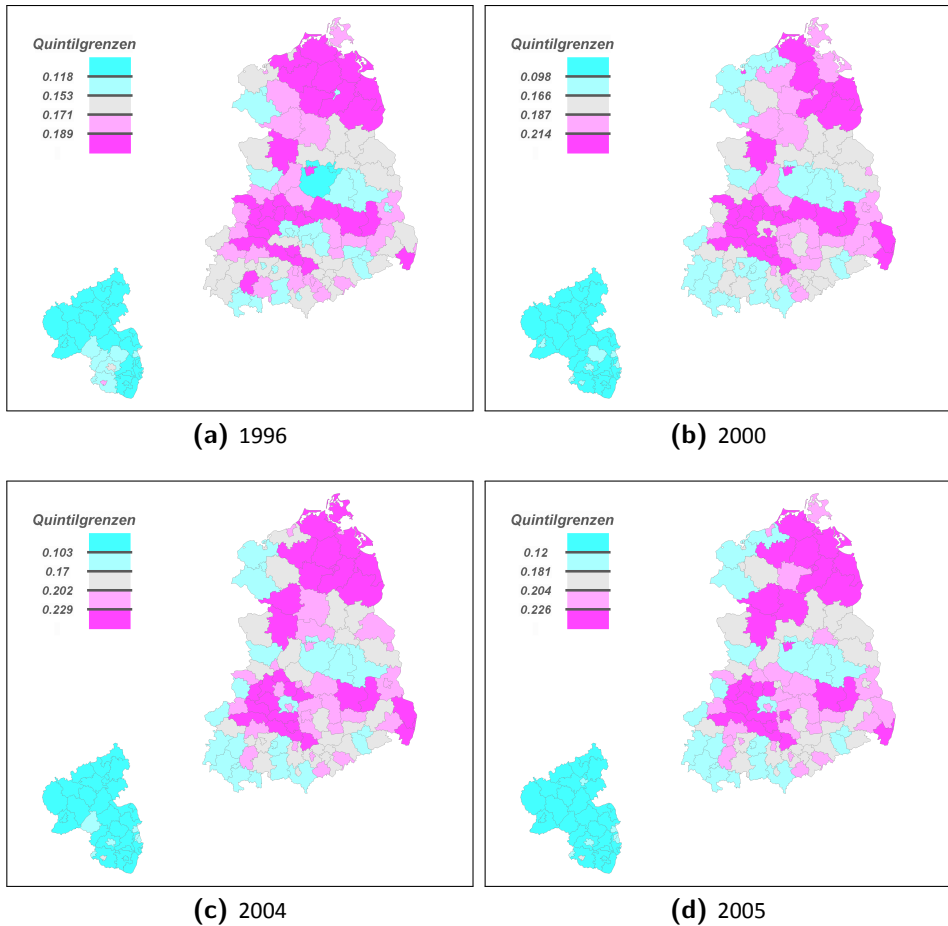


Abb. 6.2.: Arbeitslosenquoten in den Kreisen [Anteil, Quintilkarten]

nition von Arbeitslosigkeit (Abschn. 4.2.4: Krankenhaustyp, Planbetten, medizinisch-technische Großgeräte, Arbeitslosenquoten). Die mediane Arbeitslosenquote verändert sich von 0,163 (16,3 %) über 0,188 (1997), 0,178 (2000) auf 0,190 (2004) und unter neuer Definition auf 0,193 (2005). Mit Ausnahme der Kreise Potsdam und Potsdam-Mittelmark im Jahr 1996 und Potsdam im Jahr 1997 befinden sich im untersten Quartil ausschließlich Kreise von Rheinland-Pfalz, die Arbeitslosenquote allein wäre somit ein guter Indikator für die Trennung zwischen ost- und westdeutschen Bundesländern im Untersuchungszeitraum. Die Definiti-

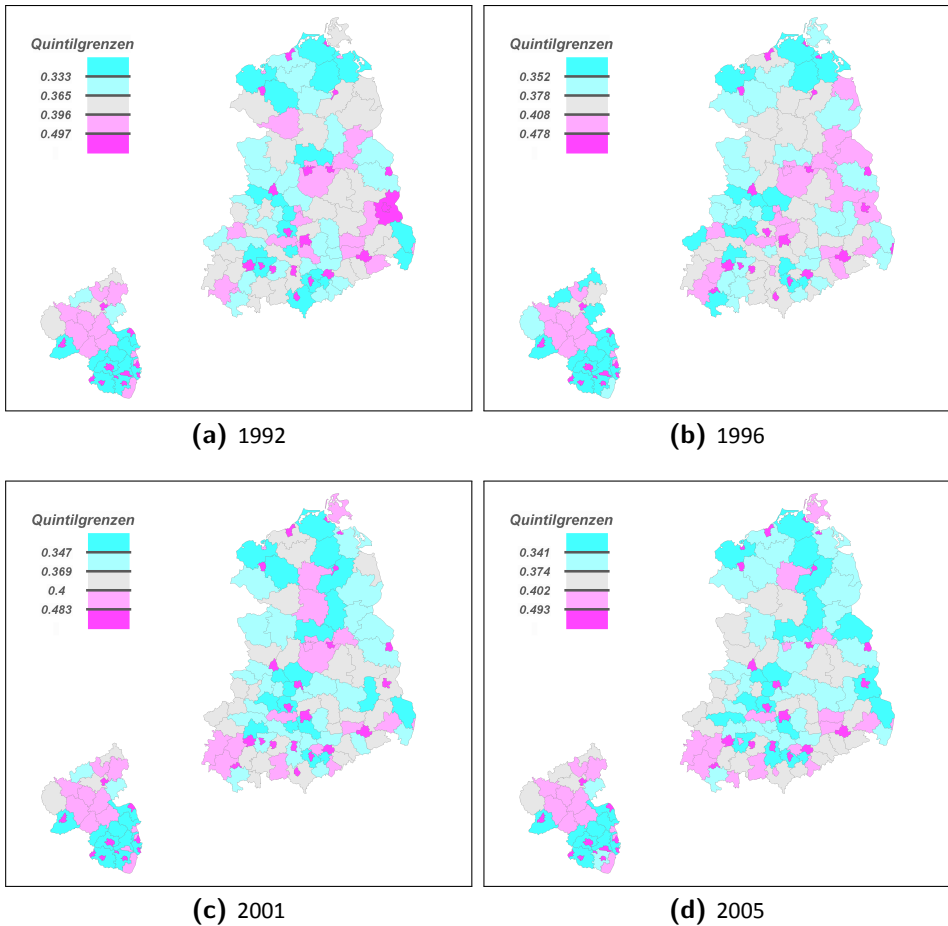


Abb. 6.3.: Brutto-Erwerbstätigenquoten in den Kreisen [Anteil, Quintilkarten]

onsveränderung zum Jahr 2005 verändert die Verhältnisse im Untersuchungsgebiet nicht.

Auffallend ist auch, dass unmittelbar an Altbundesländer angrenzende Kreise unter allen ostdeutschen relativ geringere Arbeitslosenquoten aufweisen. Das trifft vor allem für Randkreise Thüringens und Mecklenburg-Vorpommerns zu; ein erhöhter Anteil an Auspendlern zu Arbeitsplätzen in westdeutschen Kreisen kann hier nur vermutet werden.

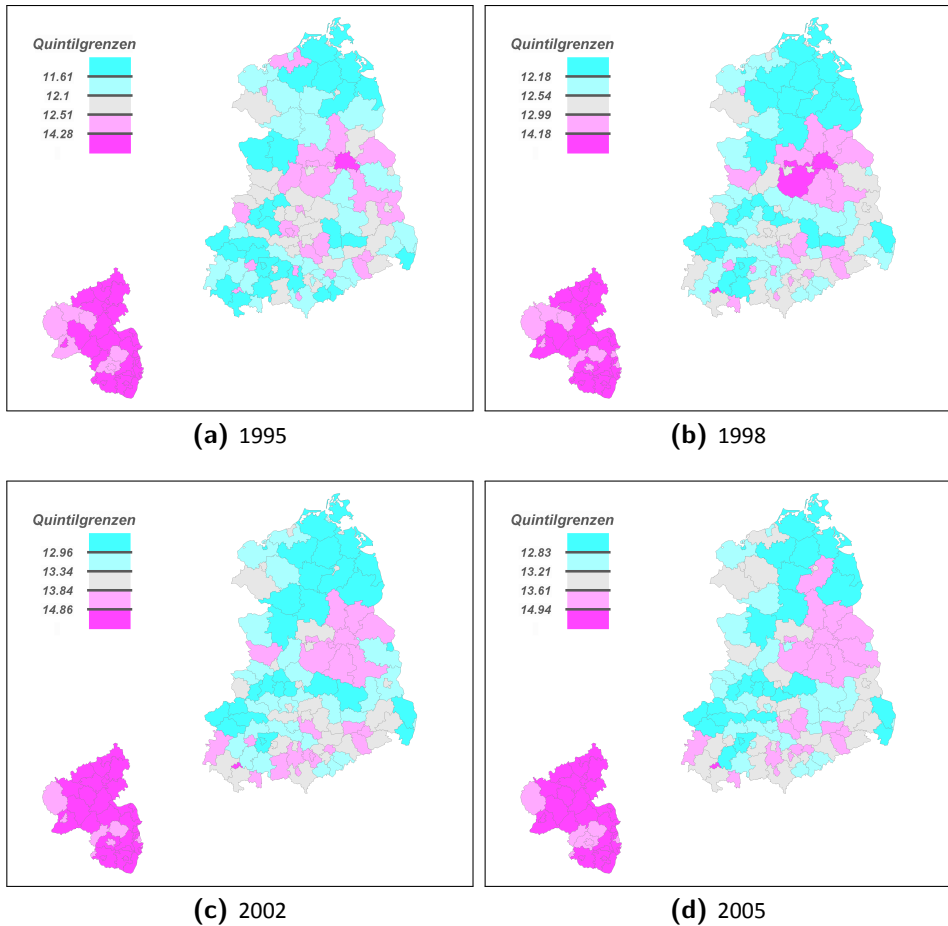


Abb. 6.4.: Verfügbares Haushaltseinkommen je Einwohner [1000€, deflationiert] in den Kreisen (Quintilkarten)

Die mediane Brutto-Erwerbstätigenquote der Kreise des Untersuchungsgebietes schwankt mit Werten zwischen 0,366 (36,6 %, 1993) und 0,392 (39,2 %, 1995 und 1996) eher gering, Kreise mit niedriger Brutto-Erwerbstätigenquote finden sich in Ost wie in West genauso wie solche höherer Brutto-Erwerbstätigenquote (Abb. 6.3). Während die kreisfreien Städte offensichtlich die hohen Brutto-Erwerbstätigenquoten aufweisen, lassen sich die niedrigen Quoten nicht automatisch den Kreisen geringerer Dichte zuordnen.

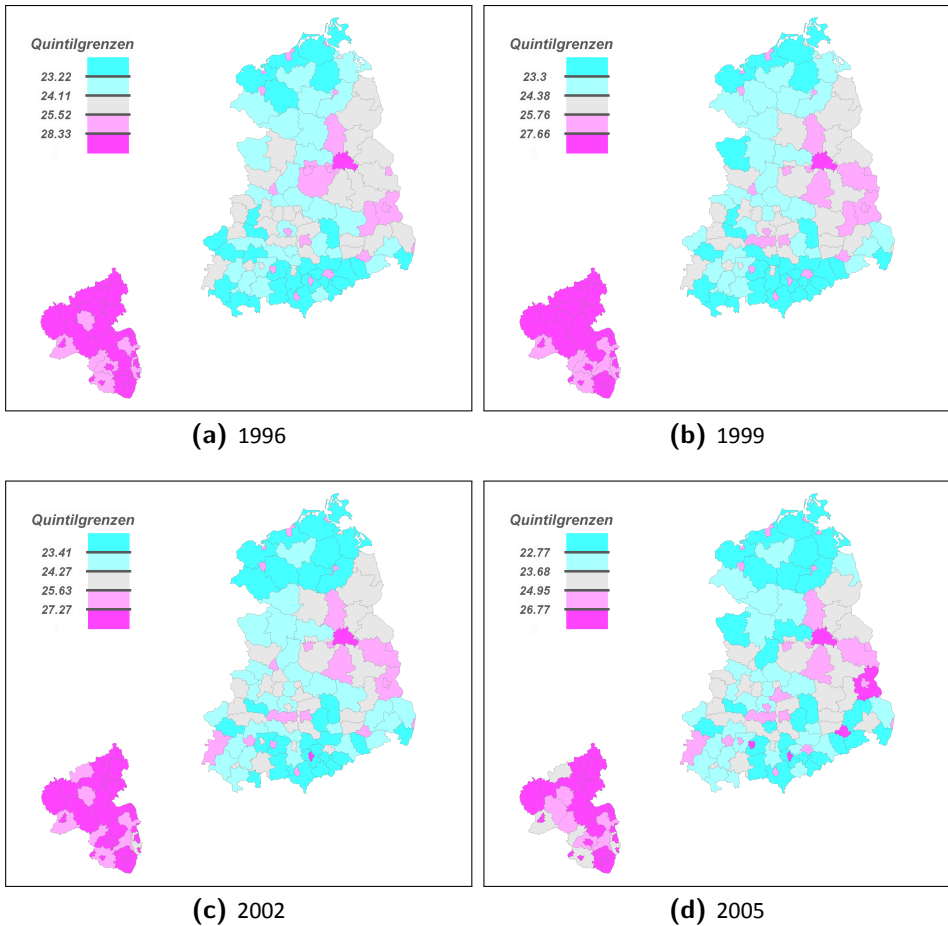


Abb. 6.5.: Arbeitnehmereinkommen je Arbeitnehmer [1000€, deflationiert] in den Kreisen (Quintilkarten)

Ähnliche Brutto-Erwerbstätigenquoten in Ost und West bedeuten hingegen nicht ähnliche verfügbare Haushaltseinkommen (Abb. 6.4). Während Berlin seine Position im obersten Quintilsbereich aufgeben muss, findet sich Suhl in Thüringen seit 1998 als einziger ostdeutscher Kreis dort. Alle Kreise von Rheinland-Pfalz hingegen gehören im Untersuchungszeitraum zu den obersten beiden Quintilsbereichen – eine Annäherung des Wohlstandes findet nicht statt.

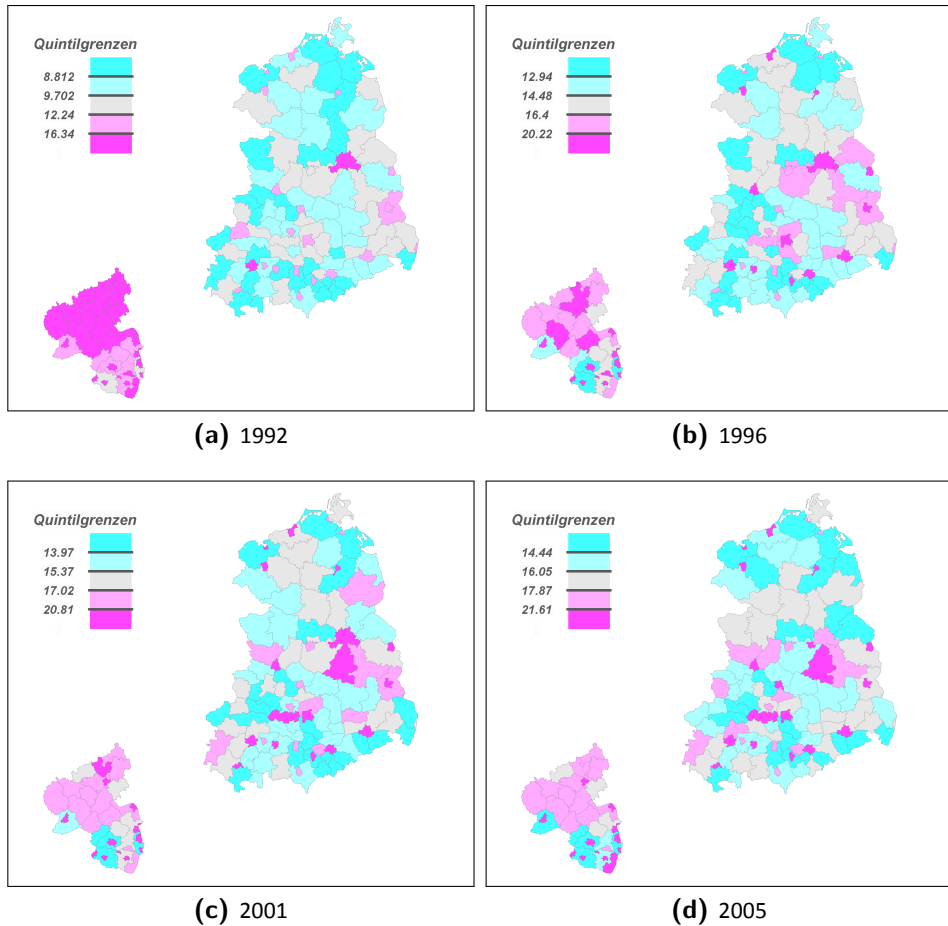


Abb. 6.6.: Brutto-Inlandsprodukt je Einwohner [1000 €, deflationiert] in den Kreisen (Quintilkarten)

Betrachtet man parallel den Verdienst aus angestellten Tätigkeiten über das (mittlere) Arbeitnehmereinkommen¹⁹⁰, so hält Berlin über alle Jahre eine Position im obersten Quintilsbereich. Beginnend mit dem Jahr 2001 gelingt es weiteren ostdeutschen Kreisen sich im obersten zu platzieren, die Kreise der beiden unteren Quintilsbereiche liegen jedoch weiterhin alle in Ostdeutschland (Abb. 6.5).

¹⁹⁰ Inlandskonzept.

Bezüglich des Brutto-Inlandsprodukts je Einwohner ergeben sich zwischen den Kreisen im Untersuchungszeitraum deutliche Verschiebungen. Während 1992 in Rheinland-Pfalz nur zwei Kreise dem mittleren und alle anderen den beiden oberen Quintilsbereichen angehören, finden sich 2005 in allen Quintilsbereichen Kreise aus Rheinland-Pfalz. Das minimale BIP je Einwohner steigt von 7.415 € auf 11.110 € (1996) und verändert sich danach nur noch wenig (zwischen 11.110 € und 11.520 €), der Median wächst von 10.940 € auf 16.700 €, das Maximum liegt 1992 bei 53.440 € und schwankt in den Folgejahren zwischen 49.940 € und 55.510 € (Abb. 6.6). Die Verteilung ist linkssteil und rechtsschief, wie die Verteilung der Arbeitnehmereinkommen je Arbeitnehmer und der Verfügbaren Haushaltseinkommen je Einwohner auch. Für die Modellierung erscheint die Annahme einer lognormalen Verteilung (Abschn. 5.1.5: Verteilungs-Annahmen und -Anpassungen) gerechtfertigt.

Im gesamten Untersuchungszeitraum ähneln die Kreise in Ostdeutschland denen in Rheinland-Pfalz bezüglich der Brutto-Erwerbstätigenquote; die Werte des Brutto-Inlandsprodukts je Einwohner gleichen sich deutlich an. Trotzdem gilt: die wirtschaftliche und soziale Wirklichkeit zwischen ost- und westdeutschen Bundesländern liegt auch 2005 weit auseinander. Ein jeder der Parameter Arbeitslosenquote, Verfügbares Haushaltseinkommen je Einwohner, Arbeitnehmereinkommen je Arbeitnehmer wäre für sich bereits eine guter Indikator für die Trennung zwischen ost- und westdeutschen Bundesländern¹⁹¹.

6.1.3. Grund- und Kostendaten – der Aufwand

Krankenhausanzahl, Anzahl der aufgestellten Betten und durchschnittliche Verweildauer sind nach wie vor beliebte Indikatoren, um schnell über Krankenhäuser zu berichten. Deutlich aussagekräftigere Informationen bieten indikations- und diagnosebezogene Auswertungen oder Darstellungen der empirischen Verteilung einzelner Merkmale.

Tab. 6.1.: Krankenhäuser je Bundesland im Untersuchungsgebiet 1992–2005^x

	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
RP	116	119	119	119	118	118	117	119	119	114	112	103	99	98
BB	66	64	60	60	59	58	55	53	54	53	51	49	47	47
MV	43	41	37	37	37	36	36	34	34	34	36	35	34	34
SN	104	102	97	96	97	98	97	96	93	93	89	88	87	86
ST	71	69	69	66	62	60	59	58	56	55	54	53	51	50
TH	66	65	63	60	57	57	56	53	53	53	49	49	50	46
BE	100	99	102	98	69	69	74	75	77	71	68	70	72	72

^x einschließlich der 3 Bundeswehrkrankenhäuser; siehe auch Tab. 4.2 auf S. 103

Krankenhäuser

Die Anzahl der Krankenhäuser in Deutschland sinkt. Während dieser Prozess für die ostdeutschen Flächenländer mit dem Beginn der 1990er Jahre startet, verringert sich die Krankenhausanzahl in Berlin zwischen 1995 und 1996 abrupt. Im Altbundesland Rheinland-Pfalz beginnt dieser Prozess erst 2000 (Tab. 6.1). Ob es sich bei der Anzahlverringerung vorrangig um Schließung oder Zusammenlegung von Krankenhäusern¹⁹² handelt, ist hier nicht eruierbar.

Ergänzend zu den Anzahlen der Krankenhäuser in den einzelnen Bundesländern ist festzustellen, dass es im Untersuchungszeitraum mit dem Landkreis Ludwigshafen¹⁹³ einen Kreis ohne Krankenhaus gibt. Darüberhinaus gibt es 9 Kreise¹⁹⁴, für die zumindest zeitweise in den Daten der Krankenhausstatistik der Nachweis eines Krankenhauses fehlt. Tab. 6.2 gibt einen Überblick, für wieviele Kreise jeweils wieviele Krankenhäuser registriert sind.

¹⁹¹ repräsentiert durch Rheinland-Pfalz.

¹⁹² z. B. <http://www.eichsfeld-klinikum.de/> » Klinikum » Historie, Zugriff am 11. 10. 2011: *Die Erfolgsgeschichte begann im Jahr 2001. Aus der Fusion des Kreiskrankenhauses Reifenstein mit dem St. Elisabeth-Krankenhaus Worbis entstand das Eichsfeld Klinikum. Ein Jahr später kam das St. Vincenz-Krankenhaus Heiligenstadt als dritter Partner hinzu.* – Das Klinikum hat heute genau die 3 Standorte.

¹⁹³ seit 01. 01. 2004: Rhein-Pfalz-Kreis.

¹⁹⁴ Rheinland-Pfalz 3, Brandenburg 1, Mecklenburg-Vorpommern 1, Sachsen-Anhalt 2 und Thüringen 2.

Tab. 6.2.: Anzahl der Kreise des Untersuchungsgebiets mit Krankenhäusern 1992–2005^x

# ^y	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
0	1	2	2	1	1	1	1	2	2	2	3	3	3	4
1	18	19	21	22	25	27	30	32	33	34	35	36	38	39
2	36	38	41	43	41	39	38	35	36	38	39	41	44	44
3	44	44	43	41	41	41	39	40	38	36	34	33	30	30
4	23	18	17	17	17	17	17	20	19	18	20	20	18	16
5	7	9	6	9	10	10	11	7	8	8	5	4	5	5
6	9	7	7	6	3	3	2	2	3	3	3	6	5	5
≥7	9	10	10	8	9	9	9	9	8	8	8	4	4	4

^x einschließlich der 3 Bundeswehrkrankenhäuser

^y Anzahl Krankenhäuser im Kreis

Unter den Krankenhäusern stellen die Universitätsklinika¹⁹⁵ eine besondere Untergruppe dar. Universitätsklinika *kümmern sich ... um schwerstkranke Patienten. Die Universitätsklinika stehen für Hochleistungsmedizin, Lehre und Forschung.*¹⁹⁶ Sie bilden eine wichtige und besondere Untergruppe, sie gehören zu den besonders großen Krankenhäusern. Unverändert im Untersuchungszeitraum befinden sich je 1 Universitätsklinikum in Rheinland-Pfalz und Thüringen, je 2 in Sachsen, Sachsen-Anhalt und Mecklenburg-Vorpommern, auch wenn letztere wegen fehlender Daten für mehrere Jahre nur begrenzt Teil der Auswertung sein können.¹⁹⁷

Für Berlin finden sich in der Krankenhausstatistik Daten von 4 (1992–1996), über 3 (1997–2000) und 2 (2001–2003) hin zu seit 2004 nur noch einem Universitätsklinikum. Berlins einziges Universitätsklinikum, die *Charité – Universitätsmedizin Berlin*, ist durch Fusion der Universitätsklinika *Charité* und *Virchow* (1997/98) sowie *Benjamin Franklin* 2003 entstanden¹⁹⁸. Diskrepanzen zu den Anzahlen in der Krankenhausstatistik können hier nur angemerkt werden.

¹⁹⁵ Hochschulkliniken im Sinne des Hochschulbauförderungsgesetzes (HBFÜG), Träger ist das jeweilige Bundesland.

¹⁹⁶ http://www.uniklinika.de/vud.php/cat/2/aid/60/title/UEber_uns, Zugriff am 04. 10. 2011.

¹⁹⁷ Aus Mecklenburg-Vorpommern fehlen die Diagnose-Daten 1993 und die Grund- und Kostendaten 1998–2001 jeweils komplett – und somit auch die der zwei dortigen Universitätsklinika.

¹⁹⁸ <http://www.charite.de/charite/organisation/geschichte/>, Zugriff am 04. 10. 2011.

Bundeswehrkrankenhäuser¹⁹⁹ nehmen im Rahmen der Krankenhausstatistik eine Sonderrolle ein, sie berichten nur ihre Leistungen für Zivilpatienten als Grund- und Diagnosedaten, Kostendaten fließen nicht in die Krankenhausstatistik ein. Im Untersuchungsgebiet finden sich über den gesamten Untersuchungszeitraum 3 Bundeswehrkrankenhäuser, je eins in Rheinland-Pfalz, Berlin und Sachsen²⁰⁰.

Plankrankenhäuser sind diejenigen, die in den Landeskrankenhausplan²⁰¹ des jeweiligen Bundeslandes aufgenommen sind, sie bilden die zahlenmäßig größte Gruppe unter den Krankenhäusern. Ihr Anteil sinkt sowohl in Ost- wie in Westdeutschland (Tab. 6.3), wobei die Werte für Rheinland-Pfalz immer unter denen Ostdeutschlands liegen.

Tab. 6.3.: Anteil [%] der Plankrankenhäuser im Untersuchungsgebiet 1992–2005^x

Gebiet	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
RP	87,8	86,4	87,3	88,1	89,7	88,9	88,8	85,6	86,4	86,7	85,6	86,3	85,7	85,6
Osttdt	92,6	92,4	90,0	90,6	95,0	95,5	94,7	94,0	93,3	91,0	91,0	89,4	88,5	87,0
OstFL ^y	97,7	97,6	96,6	97,2	97,1	97,7	97,7	97,3	97,3	96,4	96,4	95,2	94,8	93,5
BE	74,2	74,5	67,7	68,8	85,1	85,1	83,3	82,2	80,0	71,0	68,2	66,2	64,3	62,9

^x ausschließlich der 3 Bundeswehrkrankenhäuser

^y ostdeutsche Flächenländer, d. h. ohne BE

Eine feinere Analyse der Zahlen für Ostdeutschland zeigt, dass der abrupte Anstieg des Anteils der Plankrankenhäuser um 4,4 Prozentpunkte zwischen 1995 und 1996 allein in Berlin seine Begründung findet: eine Steigerung um 16,3 Prozentpunkte steht einer absoluten Verringerung der Krankenhausanzahlen um 29 (29,6 %) gegenüber. Es ist zu vermuten, dass Schließung oder Übernahme in diesem Zeitraum vor allem Nicht-Plankrankenhäuser betrifft.

Die Anteile der Krankenhäuser in den verschiedenen Trägerschaften (Abschn. 3.2.2) verschieben sich überall zuungunsten der öffentlichen Trägerschaft bei sehr unterschiedlicher Ausgangslage 1992 (Tab. 6.4). Während in Rheinland-Pfalz freigemeinnützige und private Träger gleichartig Anteile hinzugewinnen, wächst in Ostdeutschland der 1992 sehr geringe Anteil der privaten

¹⁹⁹ Träger ist der Bund.

²⁰⁰ Das Bundeswehrkrankenhaus Leipzig wird – nach Ende des Untersuchungszeitraums – zum 31. 03. 2007 geschlossen.

²⁰¹ gemäß Krankenhausfinanzierungsgesetz (KHG).

auf ein knappes Drittel. Im frischvereinigten Land Berlin dominieren 1992 die freigemeinnützigen Träger im selben Maß wie in Rheinland-Pfalz, 2005 ist nur noch jedes 18. Berliner Krankenhaus in öffentlicher Trägerschaft.

Tab. 6.4.: Anteil [%] der Krankenhaus-Trägerschaften* im Untersuchungsgebiet 1992–2005^y

Gebiet	1992			2005		
	öff.	freig.	priv.	öff.	freig.	priv.
RP	27,6	57,8	14,7	20,4	61,2	18,4
OstFL ^z	72,9	21,7	5,4	41,8	28,1	30,0
BE	22,4	57,1	20,4	5,6	47,9	46,5

* öffentliche (öff.), freigemeinnützige (freig.) oder private (priv.) Träger

^y einschließlich der 3 Bundeswehrkrankenhäuser

^z ostdeutsche Flächenländer, d. h. ohne BE

Die Veränderungen zwischen den Trägeranteilen verlaufen über die Jahre stetig, nur in Berlin verschieben sie sich zwischen 2000 und 2001 sprunghaft: während der Anteil der öffentlichen Krankenhäuser von 22,4 auf 8,6 % absinkt, wächst der in privater Trägerschaft von 28,9 auf 40,0 %.

566 Krankenhäuser 1992 verringern sich zu 433 im Jahr 2005; Schließungen und Zusammenlegungen stehen Neugründungen gegenüber. Es ist aber zu bezweifeln (Abschn. 4.3.3: Identifikation der Krankenhäuser – Paneldaten; Tab. 4.2), dass 1993 bis 2005 in Berlin 42 Neugründungen²⁰² bei 2005 nachgewiesenen 72 Krankenhäusern ($\hat{=}$ 57 %) gibt, in Rheinland-Pfalz 30 bei 98 ($\hat{=}$ 31 %), in Thüringen 13 bei 46 ($\hat{=}$ 28 %), während Sachsen-Anhalt mit 5 bei 50 ($\hat{=}$ 10 %) registriert ist. Die Eingangsforderung nach lückenloser Identifikation jedes einzelnen Panelsubjekts über alle Panelperioden oder -zeitpunkte scheint für Krankenhäuser weniger gut erfüllt, das Ausmaß auf die folgenden Analysen kann generell nur unzureichend eingeschätzt werden.

Die Krankenhausleistung eines jeden Kreises (Abschn. 5.1.2: Gesamtkrankenhäuseraktivität) hat diese Panel-Identifikationsprobleme nicht.

²⁰² 42 Krankenhaus-Identifikationsnummern mehr als Krankenhäuser 1992.

Betten

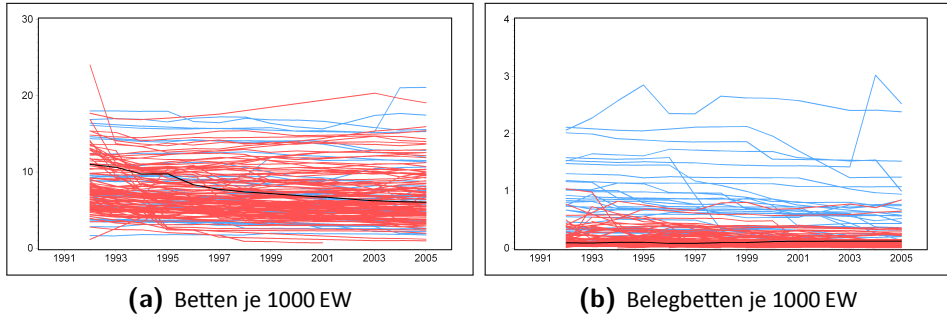


Abb. 6.7.: Bettendichte und Belegbettendichte in den Kreisen
blau: Rheinland-Pfalz (36), rot: ostdeutsche Flächenländer (110), schwarz: Berlin (1)

Die Anzahl bereitgestellter Krankenhausbetten nimmt in jedem Bundesland ab, die Bettendichte ebenso (Abb. 2.3). Betrachtet man diese Entwicklung auf Kreisebene, so wird ersichtlich, dass die Unterschiede zwischen den Kreisen den allgemeinen Trend kaum mehr erkennen lassen (Abb. 6.7 a). Das urbane Agglomerations-Zentrum Berlin nimmt dabei eine mittlere Position ein, Ost und West unterscheiden sich nicht sichtbar.

Belegbetten sind Betten in Krankenhäusern, die durch niedergelassene (krankenhausfremde) Ärzte gegen Entgelt für ihre Patienten genutzt werden. Die Bereitstellung von Belegbetten (Abb. 6.7 b) ist in Rheinland-Pfalz deutlich üblicher als in Ostdeutschland, die Belegbettendichte verändert sich im Untersuchungszeitraum jeweils kaum, eine Angleichung zwischen Ost und West findet nicht statt.

Bereinigte Kosten

Von den mehr als 20 verschiedenen Kosten-Merkmalen der Krankenhausstatistik sind nur die *bereinigten Kosten*²⁰³ unmittelbar longitudinal vergleichbar (Abschn. 4.2.4: Kostendaten). Die mittleren bereinigten Kosten steigen absolut

²⁰³ pflegesatzfähige Kosten.

(ohne Abb.) wie auch auf die Anzahl der Einwohner bezogen. Ost und West unterscheiden sich kaum.

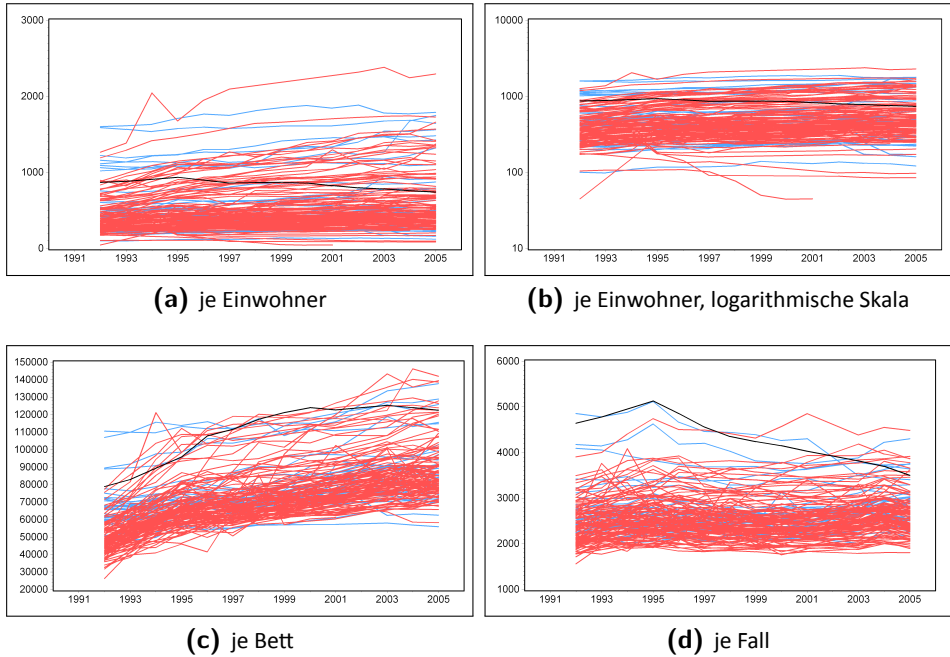


Abb. 6.8.: Relative bereinigte Krankenhaus-Kosten [€, deflationiert] in den Kreisen
 blau: Rheinland-Pfalz (36), rot: ostdeutsche Flächenländer (110), schwarz: Berlin (1)

Im Bezug auf die Einwohnerzahl wird deutlich, dass die bereinigten Kosten leicht steigen, auch wenn Berlin mit seiner Größe diesem Trend widerspricht. Ost und West unterscheiden sich weder in der mittleren Lage noch im Trend merklich, es gibt in beiden Gebieten Kreise mit hohen (Abb. 6.8 a) und niedrigen bereinigten Kosten je Einwohner (b).

Die bereinigten Kosten je Bett (c) steigen im Untersuchungszeitraum stetig an, der Anstieg nimmt in den späteren Jahren etwas ab. Gleichzeitig verringern sich die Unterschiede zwischen Rheinland-Pfalz und Ostdeutschland. Vergleicht man diese Ergebnisse mit der im selben Zeitraum forciert sinkenden Bettendichte (Abb. 2.3) ist zu vermuten, dass ein großer Teil der Erhöhung der bereinigten Kosten je Bett Folge der Veränderungen der Bettenanzahl ist. Genauso wie eine Beziehung der Kosten auf die Bettenanzahl ein jedes Bett als in seiner Wirkung

gleichwertig annimmt, ignoriert die Beziehung auf die Fallanzahl (d) die Unterschiede zwischen unterschiedlich komplexen Behandlungsfällen völlig. Trotzdem ist bemerkenswert, dass die (deflationierten) bereinigten Kosten je Fall wenig Veränderung erfahren und sich in Ost und West in all den Jahren auf dem selben Niveau bewegen.

Die Relation zwischen Personal- und Sachaufwand

Das Personal-Sachkosten-Verhältnis (Abschn. 4.2.4: Kostendaten) definiert sich als Quotient aus den *Personalkosten insgesamt* und den *Sachkosten insgesamt*. Es lässt sich sowohl als Verhältnis als auch – äquivalent – als Anteil der Personalkosten an der Summe aus Personal- und Sachkosten darstellen. Das Personal-Sachkosten-Verhältnis bewegt sich – von Ausnahmen abgesehen – zwischen 5 : 4 (55 % Personalkosten) und 3 : 1 (75 % Personalkosten). Die Kreise von Rheinland-Pfalz haben dabei über den gesamten Zeitraum die höheren Werte bei allgemein leicht sinkendem Trend: der Anteil der Sachkosten an den Gesamtkosten steigt an (Abb. 6.9). Ob eine sparsamere Ausstattung der Krankenhäuser in westdeutschen Bundesländern oder geringere Personalkosten durch weniger Personal bzw. geringere Entlohnung in Ostdeutschland ursächlich dafür verantwortlich sind, kann diesen Daten nicht entnommen werden.

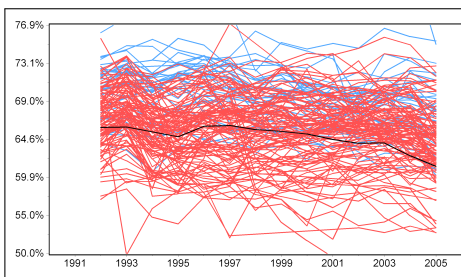


Abb. 6.9.: Personal-Sachkosten-Verhältnis in den Kreisen:
Anteil der Personalkosten an den Gesamtkosten [%; Logit-Skalierung]
blau: Rheinland-Pfalz (36), rot: ostdeutsche Flächenländer (110), schwarz: Berlin (1)

Hinsichtlich des allgemein sinkenden Trends gehen Werblow und Schoffer (2010, Kap. 5.3.4) für ihren Untersuchungszeitraum (2002–2007) von zunehmenden *Outsourcing*-Aktivitäten der Krankenhäuser im Bereich des nichtärztlichen Per-

sonals aus, sicher eine weitere mögliche Ursache für die beschriebenen Veränderungen des Personal-Sachkosten-Verhältnisses.

Personal

Der Anteil der Frauen unter den hauptamtlichen Ärzten in den Krankenhäusern Deutschlands wächst 1991–2000 von 30,4 % auf 33,3 % (Rosenow und Steinberg, 2002). Werblow und Schoffer (2010, Tab. 5-2) berichten, dass der Frauenerwerbsanteil zwischen 2002 und 2007 bei den hauptamtlichen Ärzten von 34 % auf 40 % und beim nichtärztlichen Personal von 80 % auf 81 % ansteigt. Unterschiede zwischen Ost und West werden von weder von Rosenow und Steinberg noch von Werblow und Schoffer betrachtet.

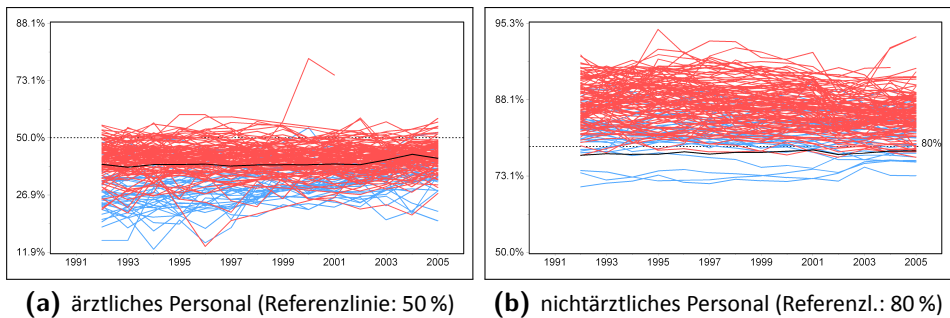


Abb. 6.10.: Frauenanteil am Krankenhauspersonal in den Kreisen* [%]

* blau: Rheinland-Pfalz (36), rot: ostdeutsche Flächenländer (110), schwarz: Berlin (1),
für 1998–2001 fehlen die Daten aus Mecklenburg-Vorpommern

† Logit-Skalierung

Die kreisspezifischen Frauenanteilstwerte unter Krankenhausärzten (Abb. 6.10 a) zeigen die Verschiedenheit zwischen den Kreisen in Ost und West. Während sich der Frauenanteil von (im Mittel) über 40 % im Osten kaum ändert und Berlin immer ein wenig unter dem Mittel aller ostdeutschen Kreise rangiert, nimmt der Frauenanteil in Rheinland-Pfalz stetig zu – mit 2005 weiterhin einem bemerkenswerten Unterschied zu den ostdeutschen Verhältnissen.

Für das nichtärztliche Personal ergibt sich diese Annäherung nur verschwindend gering, der mittlere Frauenanteil in den Kreisen Ostdeutschlands nimmt minimal ab, der in Rheinland-Pfalz verändert sich auf niedrigerem Niveau nicht. Bemerkenswert ist hier, dass Berlin mit unter 80 % für alle Jahre noch unter dem Mittel von Rheinland-Pfalz bleibt.

6.1.4. Diagnose-Statistik (Fälle) – die Leistung

Wichtigstes Ergebnis der krankenhäuslichen Aktivitäten ist die Entlassung geheilter bzw. außerhalb des Krankenhauses rekonvaleszenzfähiger Patienten. Die Fallzahlen spiegeln das nur begrenzt wider. Todesfälle sind ausgewiesen, doch ist die Anzahl der Entlassungen in ein anderes Krankenhaus zur Weiterbehandlung (hier) unbekannt. Die Schwere der Einzelfälle kann die Fallzahl allein ebenfalls nicht widerspiegeln. Ohne Case Mix Index bzw. diagnosis-related group (und DRG-Statistik, Abschn. 4.2.3) können dennoch die langjährig präferiert genutzten Näherungsvariablen (ungewichtete) Fallzahl sowie Verweildauer für einen Überblick über die Leistung der Krankenhäuser der Kreise herangezogen werden.

Fallzahlen

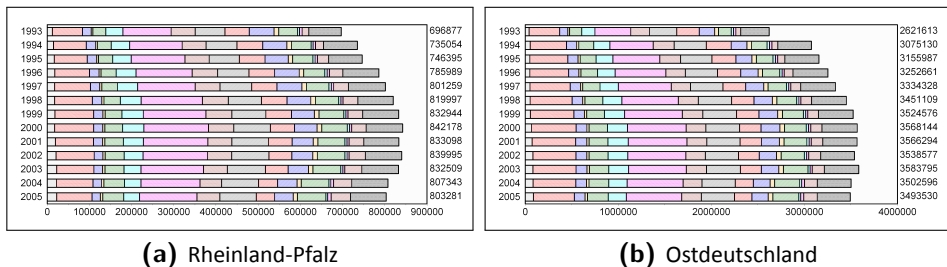


Abb. 6.11.: Fallstatistiken 1993–2005^x – ICD-9-Kapitel I–XVII^y

^x Für 1993 fehlen die Daten aus Mecklenburg-Vorpommern – vermutlich 330.000 Fälle.

^y von links nach rechts: Kapitel I bis Kapitel XVII der ICD-9

Die Anzahl der im Untersuchungsgebiet behandelten Krankenhausfälle liegt für 1993 bei 3,7 Mio., sie erhöht sich bis 2000 stetig, schwankt dann um den Wert

von 4,4 Mio. und verringert sich für 2004 und 2005 auf 4,3 Mio. (Tab. 6.5). Im Mittel entlässt jedes Krankenhaus 1993 6.563 Fälle, bis 2005 steigt diese Zahl auf 9.927. Die empirische Verteilung dieser Fallzahlen je Krankenhaus ist linksteil und rechtsschief²⁰⁴, größere Krankenhäuser entlassen im Jahr weit mehr als 20.000 Fälle. Ost und West unterscheiden sich weder in der Entwicklung der Gesamtzahl der Fälle noch in der Aufteilung der Fälle in die 17 Diagnose-Kapitel der ICD-9 (Abb. 6.11).

Altersverteilungen

Tab. 6.5 zeigt anhand von Perzentilen die Altersverteilung der Fälle über die Jahre 1993–2005. Die Datenausfälle wegen fehlender Altersangaben sind verschwindend gering. Man sieht, dass das mittlere Alter der Fälle erheblich ansteigt – im Median um 10 Jahre, im arithmetischen Mittel um $7\frac{1}{3}$ Jahre – und dass auch das obere Ende der Verteilung (95. und 99. Perzentil) sich hin zu immer größeren Alterswerten verschiebt.

Betrachtet man die Veränderungen der Altersverteilungen der männlichen und weiblichen Fälle über die Jahre (Abb. A.2) so fallen die geringeren Fallzahlen in den Altersstufen der gegen Ende des I. bzw. II. Weltkrieges Geborenen auf, diese verschieben sich erwartungsgemäß mit den Jahren. Das gilt für beide Geschlechter in ähnlichem Ausmaß. Die Masse der multimodalen Verteilung verschiebt sich für beide Geschlechter über die Jahre deutlich hin zu höheren Alterswerten, dies ist auch eine Erklärung für die so extrem erscheinenden Zunahmen von arithmetischem Mittel und Median. Unterschiedlich hingegen sind die Verteilungen für das Kindesalter (Jungen häufiger als Mädchen) und im hohen Alter – die höhere Lebenserwartung²⁰⁵ für Frauen spiegelt sich hier wider. Während sich 1993 eine Häufung der weiblichen Fälle im gebärfähigen Alter zeigt, schwindet diese bis zum Jahr 2005 fast vollständig.

²⁰⁴ Das geometrische Mittel ist hier aussagekräftiger, z. B. gilt für das Jahr 2000: arithmetisches Mittel = 9.078 Fälle, geometrisches Mittel = 4.571 Fälle.

²⁰⁵ Neugeborene, früheres Bundesgebiet, Geburtskohorte 1936: Jungen 67,7 Jahre, Mädchen 76,1 Jahre; 1960: Jungen 75,3 Jahre, Mädchen 82,3 Jahre (Statistisches Bundesamt, 2011, weitere Angaben dort).

Tab. 6.5.: Fallstatistiken 1993–2005 – empirische Altersverteilung:
Anzahl der Fälle^x in den Krankenhäusern des Untersuchungsgebietes,
davon ohne Altersangabe, das Alters-Mittel sowie ausgewählte
Perzentilwerte [Jahre|Monate]

Jahr	Anzahl	ohne Angabe	Alters- Mittel	Alters-Perzentile				
				1.	5.	50.	95.	99.
1993	3.338.898	629	46 11	0 00	3 10	50 01	83 07	89 07
1994	3.826.898	301	47 05	0 01	4 03	50 11	83 10	89 09
1995	3.912.165	116	48 03	0 01	5 00	52 00	84 01	90 00
1996	4.045.624	84	49 12	0 01	5 07	52 11	84 05	90 04
1997	4.138.203	49	49 06	0 01	5 09	53 08	84 07	90 06
1998	4.274.849	62	50 02	0 02	5 09	54 07	85 00	90 11
1999	4.359.276	44	50 11	0 02	6 02	55 07	85 03	91 03
2000	4.412.355	32	51 06	0 02	6 03	56 05	85 06	91 05
2001	4.400.251	10	52 01	0 02	6 03	57 02	85 08	91 10
2002	4.379.841	3	52 08	0 02	6 02	57 11	85 02	92 00
2003	4.417.006	6	53 03	0 02	6 07	58 09	84 08	92 02
2004	4.310.315	17	53 10	0 01	6 02	59 07	84 08	92 04
2005	4.298.398	0	54 03	0 00	6 00	60 02	85 01	92 07

^x Für 1993 fehlen die Daten aus Mecklenburg-Vorpommern – vermutlich 330.000 Fälle.

Die Altersverteilung der Gebärenden in Ost und West (Abb. A.3) verschieben sich im Untersuchungszeitraum zu höheren Werten, für Rheinland-Pfalz wandelt sich eine symmetrische Verteilung mit Modalwert im 30. Lebensjahr zu einer rechtssteilen mit Modalwert im 33., in Ostdeutschland ändert sich die Verteilung von linkssteil zu symmetrisch mit Modalwerten 27. und 29. Lebensjahr; obwohl nicht gleich, nimmt das Alter der im Krankenhaus Gebärenden in beiden Teilen Deutschlands zu. Die Anzahl der Gebärenden sinkt dabei in Rheinland-Pfalz nach Jahren relativer Konstanz mit etwa 37.000 ab 1999 stetig auf 26.901 im Jahr 2005. Anders in Ostdeutschland: von 1994²⁰⁶ mit 97.111 stiegen die Anzahlen auf 123.976 für 2000. Nach rapidem Abfallen sinken die Zahlen dann für 2002 bis 2005 leicht von 108.954 auf 104.434.

Dabei ist mit Fehlern bei den Altersangaben zu rechnen: Abb. A.3 b zeigt die Entbindungen 2003 mit fraglichen Altersangaben für die Gebärende von unter 1 und

²⁰⁶ Für 1993 fehlen die Anzahlen aus Mecklenburg-Vorpommern.

über 70 Jahren. Für Fälle mit einer Hauptdiagnose aus dem ICD-9-Kapitel XV *Bestimmte Affekte mit Ursprung in der Perinatalzeit* ergeben sich – bei etwa 30.000 Fällen pro Jahr – ab dem Jahr 2001 stetig deutliche Veränderungen in den Altersverteilungen. Während unter den 1993–2000 sowohl männlichen wie weiblichen entlassenen Fällen eine bedeutsame Anzahl mit einem Alter von mehr als 3 Jahren – mit für jedes Geschlecht und Jahr einem Maximum bei über 90 Jahren und für Frauen einer Häufung bei 30 Jahren – registriert ist, gibt es 13 Jahre später mit einem weiblich und einen männlichen im jeweils 5. Lebensjahr nur noch zwei Fälle eines Alters von über 3 Jahren (Abb. A.4). Es muss hier offen bleiben, ob diese Diagnosen ab dem Kindesalter nicht mehr im Krankenhaus behandelt werden oder ob in diesen Fällen eine andere Hauptdiagnose – diese dann sicher aus einem anderen Diagnose-Kapitel – gestellt wird.²⁰⁷

Diagnose-Gruppierungen

Für einen besseren medizinisch-epidemiologischen Bezug der Auswertungen sind bestimmte Diagnosen ausgewählt und nach Ähnlichkeit zusammengefasst (Abschn. 4.4.3: Diagnose-Gruppierungen).

Die dabei gewählten Diagnosen decken 53,8% aller berichteten Fälle ab. Sowohl Zoonosen durch Bakterien mit 3.699 Fällen als auch Geschlechtskrankheiten (6.299 Fälle) treten unter den insgesamt 54.114.079 berichteten sehr selten auf; die anderen Diagnose-Gruppierungen umfassen jede mit 52.931 (Entzündliche Krankheiten des Zentralnervensystems) bis 9.066.420 (Krankheiten des Kreislaufsystems) deutlich mehr Fälle.

Verweildauern und Langlieger

Die Verweildauer, die Zeit, die ein vollstationärer Fall zwischen Aufnahme und Entlassung im Krankenhaus verweilt, sinkt in allen deutschen Krankenhäusern

²⁰⁷ Gegen Fehler bei den Altersangaben als alleinige Ursachenannahme sprechen die kontinuierliche Abnahme der Häufigkeit und die Konsistenz der Angaben über die Jahre 1993–2004.

im Mittel von 12,5 Tagen 1993 auf 8,6 Tage 2005 (Statistisches Bundesamt, 2008a)²⁰⁸.

In Übereinstimmung mit Bömermann (2003) ergibt sich die Frage, inwieweit das arithmetischen Mittel das sinnvolle Maß zur Beschreibung von Verweildauern ist. Darüber hinaus ist zu überlegen, zu welchem Anteil diese allgemein als Erfolg gewertete Verringerung auf die administrative Eliminierung²⁰⁹ der Patienten mit potenziell sehr hoher Verweildauer (Abschn. 4.3.3: Verweildauern) zustande kommt²¹⁰.

Als *Langlieger* seien Fälle mit einer Verweildauer von mehr als 6 Wochen (43 oder mehr Tagen) bezeichnet. Deren Anteil sinkt kontinuierlich von 2,94 % 1993 auf 2,74 % im Jahr 2005. Für Frauen liegt der Langlieger-Anteil in jedem Jahr um etwa 0,02 Prozentpunkte über dem der Männer. Für alle Jahre gemeinsam ist unter allen Altersgruppen der Anteil der Langlieger bei den über 90-Jährigen mit 3,06 % am höchsten.²¹¹

Die Langlieger-Anteile in den ICD-9-Kapiteln schwanken zwischen 3,27 % (Kapitel XIII, 1993) und 2,50 % (Kapitel XI, 2005) bei allgemein sinkenden Werten. In allen Jahren ist Kapitel XI (Komplikationen während der Schwangerschaft, der Entbindung oder im Wochenbett²¹²) das mit dem geringsten Langlieger-Anteil. Während 1993 Kapitel XIII (Krankheiten des Skeletts, der Muskeln und des Bindegewebes) den höchsten Langlieger-Anteil aufweist, nimmt dieses Kapitel 13 Jahre später einen mittleren Rang ein; Kapitel V (Psychische Krankheiten) hat 2005 mit 3,12 % den höchsten Anteil an Langliegern.

Weitergehend werden Verweildauern und ihre Verteilungen hier mit Mitteln der Ereigniszeitanalyse (Abschn. 5.1.6) analysiert. Statt für alle Diagnosen gemeinsam

²⁰⁸ 1991 betrug die mittlere Verweildauer noch 14,0 Tage (ibid.).

²⁰⁹ durch Überweisung in den Pflegebereich.

²¹⁰ Dazu wiederum Bömermann (2003): *Der deutliche Rückgang der Verweildauerlänge 1995/96 um 3,3 Tage und 1996/97 um 1,7 Tage ist überwiegend einer Änderung der Grundgesamtheit geschuldet. Bis 1996 gab es in Berlin 29 Krankenhäuser, die speziell chronisch Kranke versorgen. Danach wurden die so genannten Chronikerbetten in Betten in Pflegeeinrichtungen umgewidmet. Bei der Beurteilung der langen Verweildauer in Berliner Krankenhäusern ist ... dieser methodische Effekt in Rechnung zu stellen..*

²¹¹ Für Kleinstkinder bis zu einem Jahr beträgt der Anteil 2,90 %, während er für die Gruppe der 18- bis 30-Jährigen mit 2.64 % am geringsten ist.

²¹² Komplikationslose Entbindungen sind eingeschlossen.

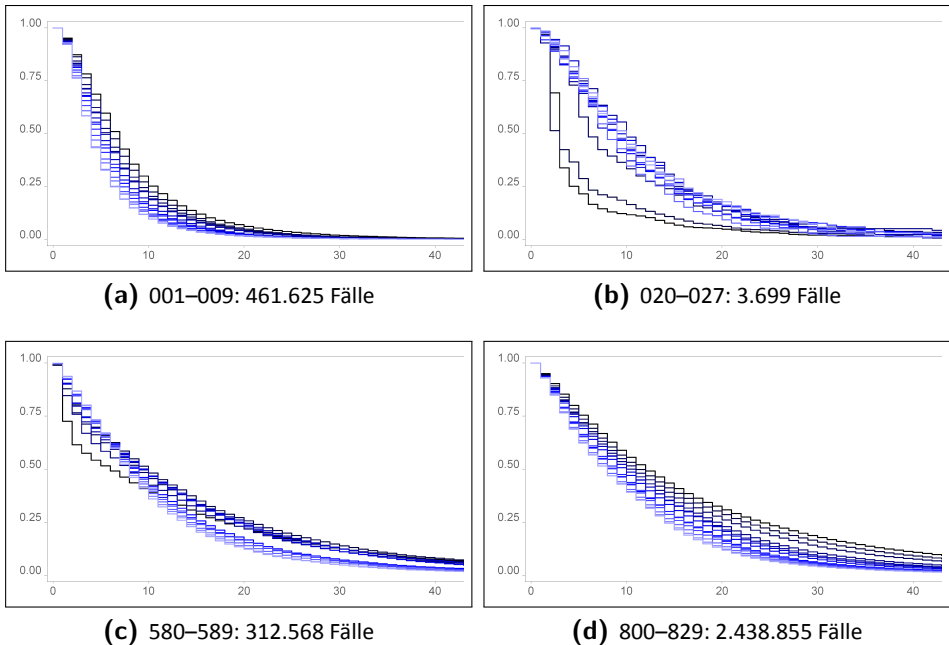


Abb. 6.12.: Ereigniszeitverteilungen zur Verweildauer [d] 1993–2005 ausgewählter Diagnose-Gruppierungen
spätere Jahre in helleren Farbtönen

stehen die Diagnose-Gruppierungen im Mittelpunkt der Analysen. Das höchste in den Gruppierungen beobachtete 75. Perzentil der Verweildauern gilt für 1993 und beträgt 30 Tage (800–829: Frakturen des Schädels, der Wirbelsäule, des Rumpfskelett, der oberen und unteren Extremitäten). Verweildauern sind für alle Werte größer als 42 Tage (6 Wochen, die Langlieger) rechts-zensiert (Abschn. 4.3.3: Verweildauern), die Kaplan-Meier-Schätzungen der jahresweisen Verweildauern werden bis zum Tag 42 dargestellt (Abb. 6.12).

Während sich für die Gruppierungen 001–009 (Infektiöse Krankheiten des Verdauungssystems) bei kontinuierlich steigenden Fallzahlen (1993 ca. 20.000, 2005 ca. 50.000) und 800–829 (Frakturen des Schädels, der Wirbelsäule, des Rumpfskelett, der oberen und unteren Extremitäten) mit jährlich etwa 190.000 Fällen der Trend der stetigen Abnahme der Verweildauern bestätigen lässt, ergibt sich für die (nur) 3.699 Fälle aus 020–027 (Zoonosen durch Bakterien) und die 312.568

aus 580–589 (Nephritis, Nephrotisches Syndrom und Nephrose) ein anderes Bild. Ursächlich verantwortlich für die in den frühen 90er Jahren deutlichen Steigerungen der Verweildauer in diesen Gruppierungen könnte die Tendenz sein, leichtere und somit potenziell verweildauerkürzere Fälle zunehmend ambulant zu behandeln. Auffallend ist, dass die 1993 innerhalb der Gruppierung dominierende Diagnose 025 (Melioidose) mit einem Anteil von über 60 % ab 1996 nur noch in eher geringerer Häufigkeit anzutreffen ist, bei insgesamt sehr unterschiedlichen jährlichen Fallzahlen (1994: 571; 1996 und 1997 je 129; 2000: 430; 2003–2005 je 250–260) dominiert ab 1996 mit jeweils über 80 % die unspezifisch klingende Diagnose 027 (Sonstige Zoonosen durch Bakterien). Eine ähnlich deutliche Verschiebung in der Häufigkeit bestimmter Diagnosen ergibt sich allerdings in der Gruppierung 800–829 nicht.

In Gruppierung 320–326 (Entzündliche Krankheiten des Zentralnervensystems, ohne Abb.) mit 52.931 Fällen ergibt sich ein gemischtes Bild, das 75. Perzentil der Verweildauer sinkt von 27 Tagen 1993 auf 15 Tage 1998 und steigt seitdem auf zuletzt 18 Tage 2005. Für alle anderen der Diagnose-Gruppierungen verringern sich die Verweildauern mit jedem Jahr.

6.2. Ökonometrische Modellierungen

Jenseits der deskriptiven Statistik erfordert die Analyse von Zusammenhängen zwischen mehreren Merkmalen eine Verwendung komplexerer Verfahren der explorativen und konfirmatorischen Statistik. In Abhängigkeit vom Typ der Zielvariablen (Regressanden), der Typen der Einflussgrößen (Regressoren) und des jeweils unterstellten Zusammenhangs fordern die Fragestellungen die Anwendung unterschiedlicher statistisch-ökonometrischer Modelle und Modellklassen.

Betrachtungen zur Auswahl der einzelnen Modelle, ihrer Anpassung an die Daten, zu möglichen Alternativen und zur Interpretation der einzelnen Modellergebnisse erfolgen in diesem Abschnitt (6.2); im folgenden Abschnitt (6.3) werden diese Einzelergebnisse zusammengefasst und in ihren wirtschaftswissenschaftlich-ökonometrischen Kontext gestellt.

6.2.1. Verweildauern und Hazardfunktionen

Mit Ausnahme der Diagnose-Gruppierung 090–099 (Syphilis [Lues] und sonstige Geschlechtskrankheiten) mit jährlich zwischen 407 (1997) und 614 (1995) bei insgesamt 6299 Fällen zeigt der Logrank-Test für jede der Gruppierungen ein auf dem 5 %-Niveau signifikantes Ergebnis: die jahresweisen Ereigniszeitverteilungen unterscheiden sich statistisch – im Lichte der Gruppierungen mit sehr großer Fallzahl ein wenig überraschendes Ergebnis (Abschn. 5.1.4). Wichtiger ist, dass für die Gruppierung 090–099 auch ohne statistische Signifikanz der generelle Trend mit den Jahren abnehmender Verweildauern zutrifft (ohne Abb.).

Modelliert man zusätzlich zu den Ereigniszeitverteilungen für jede Gruppierung die Hazard²¹³-Funktion $\lambda(t)$ in Abhängigkeit vom Berichtsjahr, lassen sich weitere Erkenntnisse über die Verweildauern gewinnen. Die Hazard-Funktion in leicht geglätteter Form (Abschn. 5.1.6) beschreibt die Chance, zur Zeit t entlassen zu werden. Die empirisch ermittelte Weite des Glättungsfensters von 2,2 stellt sich dabei als günstig heraus und liegt allen dargestellten Hazard-Funktionen zugrunde.

Für die 461.625 Fälle der Diagnose-Gruppierung 001–009 (Infektiöse Krankheiten des Verdauungssystems) nimmt mit den Jahren ab 1993 die Chance stetig zu, eher zeitig – bis etwa zum Tag 23 – aus dem Krankenhaus entlassen zu werden, nach dem Tag 23 – jährlich weniger als 5 % der Fälle – verlieren sich die Unterschiede (Abb. 6.13 a). Außerdem zeigt jede Jahreskurve einen Anstieg bis zu einem Maximum (1993 bei 9, 2005 bei 5 Tagen) gefolgt von einem im Wesentlichen monotonen Abfall bis zum Ende der unzensierten Verweildauerwerte am Tag 42.

Sichtbar anders verlaufen die jahresweisen Hazard-Funktionen für die Gruppierung 580–589 (Nephritis, Nephrotisches Syndrom und Nephrose). Deutlicher als bei der Kaplan-Meier-Schätzung der Verweildauer selbst (Abb. 6.12 c) zeigt sich hier (Abb. 6.13 c) ein Bruch zwischen den Jahren 1993–1999 sowie 2000–2005: während in den zeitigeren Jahren die größte Chance entlassen zu werden vor dem Tag 6 liegt, ist dieses Maximum in den späteren Jahren nach dem Tag 9 anzutreffen. Die Annahme, leichtere und somit potenziell verweildauer kürzere Fälle wür-

²¹³ engl: *hazard* $\hat{=}$ Gefahr, Risiko, Wagnis, Zufall.

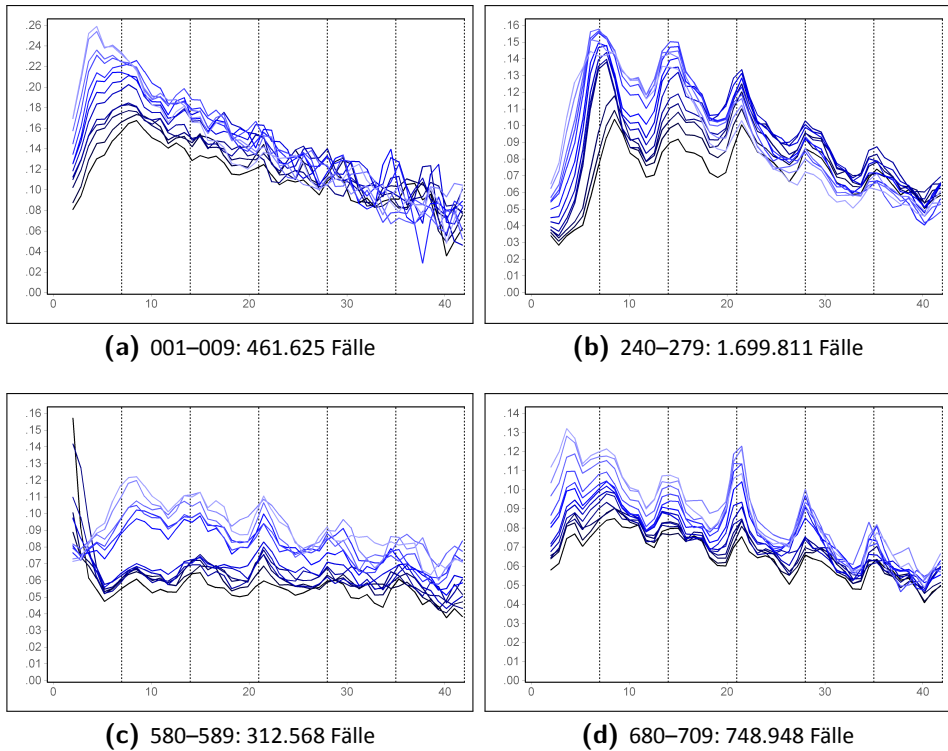


Abb. 6.13.: Hazardfunktionen zur Verweildauer [d] 1993–2005 ausgewählter Diagnose-Gruppierungen
geglättet, spätere Jahre in helleren Farbtönen, Referenzlinien im Wochenabstand

den zunehmend ambulant behandelt, kann diese Veränderungen in den Hazard-Funktionen erklären, den Bruch jedoch erklärt sie nicht.²¹⁴

Für beide Perioden (1993–1999 und 2000–2005) erkennt man jedoch eine Periodizität von 7 Tagen in den (lokalen) Maxima der Hazard-Funktionen. Ausgeprägter noch (Abb. 6.13 b und d) findet sich diese 7-Tage-Periodizität in den 1.699.811 Fällen der Gruppierung 240–279 (Endokrinopathien, Ernährungs- und Stoffwechselkrankheiten sowie Störungen im Immunsystem)²¹⁵ und den 748.948 Fällen der

²¹⁴ Inwieweit eine Veränderung dieser Grundgesamtheit mit dem Wechsel von ICD-9 zu ICD-10 (Abschn. 4.3.2: Leistungsfähigkeit der ICD-Konkordanzlisten) stattgefunden hat und diese ursächlich ist, muss offen bleiben.

²¹⁵ komplett das Diagnose-Kapitel III.

Gruppierung 680–709 (Krankheiten der Haut und des Unterhautzellgewebes)²¹⁶, letztere mit einem zusätzlichen lokalen Maximum am Tag 4 und einem sehr ausgeprägten nach genau 3 Wochen (Tag 21).

Aus medizinischer Sicht scheint das nur schwer nachvollziehbar, erwartet man doch für die jeweilige Grundgesamtheit von Fällen mit ähnlicher Krankheit einen eher stetigen Heilungsprozess. Parallelen zum Wochenrhythmus sind hier jedoch evident und lassen eher einen deutlichen administrativen Einfluss auf die Festsetzung des Entlassungstages für die Fälle vermuten als eine konsistente Umsetzung wirtschaftlicher und medizinisch-epidemiologischer Ziele.

Für die Krankenhäuser in Berlin im Jahr 2000 findet Bömermann (2003, Abb. 1) für alle Behandlungsfälle gemeinsam ein ähnliches Wochen-Muster mit Häufungen an jedem 7. Tag, ohne näher darauf einzugehen.

6.2.2. Die Aufnahme vollstationärer und teilstationärer Fälle

Die Krankenhausstatistik berichtet Fälle zum Zeitpunkt ihrer Entlassung.²¹⁷ Abb. 6.14 hingegen zeigt die monatlichen Aufnahmezahlen der Krankenhäuser in Rheinland-Pfalz bzw. Ostdeutschland für alle Fälle und für die, die davon zu den teilstationären Fällen zu zählen sind. Den Trendberechnungen liegt jeweils eine multiplikative Dekomposition mit saisonaler Komponente von 12 Monaten Länge (SAS Institute Inc., 2004a, Classical Decomposition Operators, S. 812 ff.) zugrunde; die Darstellung erfolgt trotz Fehlens der Daten aus Mecklenburg-Vorpommern für das Jahr 1993.

Der Anteil an teilstationären Fällen (Abschn. 4.2.4: Vollstationär, teilstationär, ambulant oder Stundenfall?) ist nie größer als 5 %, er ist in Ostdeutschland seit 1994 und in Rheinland-Pfalz seit 1998 stetig im Sinken begriffen.

²¹⁶ komplett das Diagnose-Kapitel XII.

²¹⁷ Fälle mit einem Krankenhausaufenthalt über den Jahreswechsel hinweg werden im Folgejahr berichtet. Für Aufnahmen vor Mitte November betrifft dies nur Langlieger und somit für 2005 weniger als 2,75 % der Fälle. Selbst Ende Oktober 2005 aufgenommene Fälle sind Ende Dezember zu mehr als 97 % abgeschlossen und somit im Berichtsjahr 2005 erfasst. Deshalb werden die Zeitreihen hier bis einschließlich Oktober 2005 betrachtet.

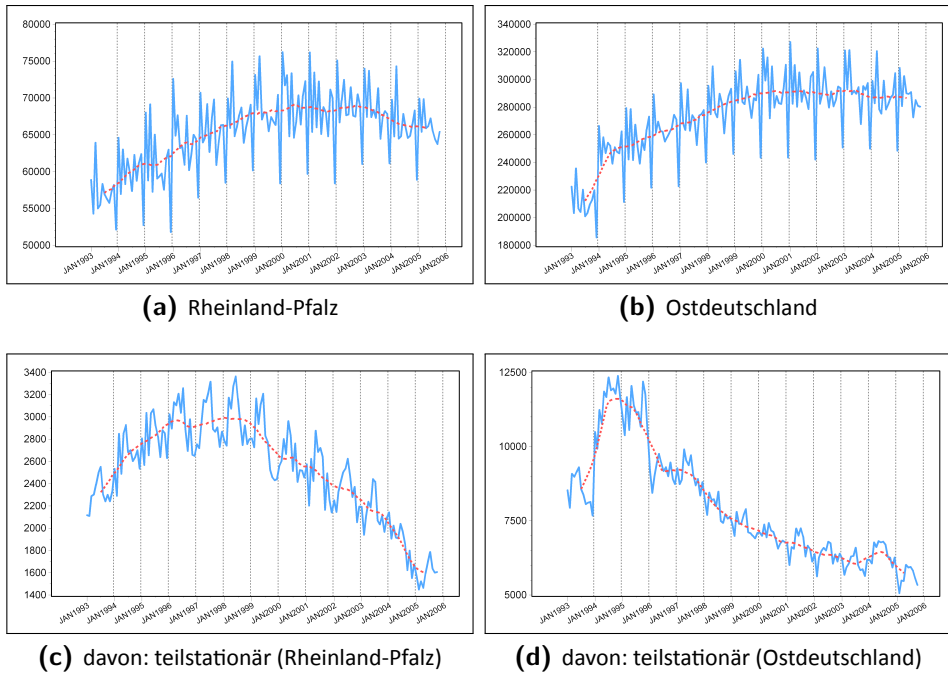


Abb. 6.14.: Zeitreihen monatlicher Aufnahmezahlen Januar 1993 – Oktober 2005*
der Krankenhausfälle mit Trendkomponente, multiplikatives Modell
Referenzlinien an den Jahreswechsell

* Für 1993 entlassene Fälle fehlen die Daten aus Mecklenburg-Vorpommern.

Für alle Fälle spiegelt der Trend die bekannte jährliche Entwicklung der Fallzahlen (Tab. 6.5) wider. Auffallend jedoch ist der jährlich wiederkehrende Abfall der Aufnahmezahlen im Dezember – für Rheinland-Pfalz und Ostdeutschland in ähnlichem Ausmaß. Die Analyse der Saisonkomponente (Abb. A.5) ergibt, dass im Dezember in Rheinland-Pfalz nur knapp 88% und in Ostdeutschland knapp 85% der bei gleichmäßiger Aufnahme über das Gesamtjahr zu erwartenden Zahlen erreicht werden. Kompensiert werden diese durch erhöhte Aufnahmezahlen im jeweils ersten Quartal des Jahres.

6.2.3. Verweildauer-Modelle

Alle Untersuchungen zu Einflussfaktoren auf die Verweildauer beziehen sich auf die vollstationären Fälle unter Ausschluss der vernachlässigbaren Anzahl²¹⁸ derer mit unbekanntem Geschlecht.

Modelle mit gruppierten Altersangaben

Für die Diagnose-Gruppierung 001–009 (Infektiöse Krankheiten des Verdauungssystems) sind die Voraussetzungen für die Anwendung des Cox-Modells (Abschn. 5.1.6 sowie Singer und Willett, 2003, S. 514) glaubhaft erfüllt.²¹⁹ Als exogene Regressoren gehen in den linearen Prädiktor η ein:

- (a) die Binärvariable *ostdt Ist der Fall aus einem Krankenhaus in Ostdeutschland?*,
- (b) das Alter *altgr* des Patienten als Fall bei Aufnahme – ein Faktor mit den Altersstufen [0, 1], [1, 5], [6, 10], [11, 18], [18, 30], [30, 45], [45, 65], [65, 90], [90, ∞], formuliert als 8 Indikatorvariablen $altgr_1, \dots, altgr_8$,
- (c) das Geschlecht *geschl* als Binärvariable,
- (d) die Binärvariable *wohnkreis Kreis des Krankenhauses = Wohnkreis?*,
- (e) das Berichtsjahr als stetige Kovariate *jahr*,
- (f) die (zur Basis 10 logarithmierte) Einwohnerdichte *ewdL* des Kreises des Sitzes des Krankenhauses,
- (g) die Arbeitslosenquote *alqlgt* des Kreises des Sitzes des Krankenhauses (als Logit), und
- (h) die (zur Basis 10 logarithmierte) Brutto-Erwerbstätigenquote *betqL*.

Der (absolutgliedfreie) lineare Prädiktor des Cox-Modells (Gl. 5.3) ergibt sich somit zu

$$\eta = \beta_1 \cdot ostdt + \beta_{21} \cdot altgr_1 + \dots + \beta_{28} \cdot altgr_8 + \beta_3 \cdot geschl \\ + \beta_4 \cdot wohnkreis + \beta_5 \cdot jahr + \beta_6 \cdot ewdL + \beta_7 \cdot alqlgt + \beta_8 \cdot betqL.$$

²¹⁸ 1994: 143 Fälle, in allen anderen Jahren je weniger als 25 Fälle.

²¹⁹ Speziell zeigen die Loglog-Survival-Funktionen für alle Berichtsjahre Parallelität.

Das Modell untersucht den gleichzeitigen Einfluss des Ost-West-Unterschiedes auf Verweildauer unter Einbeziehung demografischer (Alter, Geschlecht), Mobilitäts- (Wohnkreis), Urbanitäts- (Einwohnerdichte), sozialer (Arbeitslosenquote) und erwerbswirtschaftlicher (Brutto-Erwerbstätigenquote) Faktoren.

Wegen der Nichtverfügbarkeit von Angaben zur Arbeitslosigkeit (Abschn. 4.1.4: Bundesagentur für Arbeit – Internetangebot) für die Jahre 1993–1995 werden hier die 392.877 (85,1 %) vollstationären Fälle der Jahre 1996–2005 betrachtet. Einzig für das Geschlecht ergibt sich kein signifikanter Einfluss ($p = 0,941$ als Ablehnungswahrscheinlichkeit für die Wald-Statistik, Abschn. 5.1.5: Modellvergleiche, Modellselektion und Auswahl geeigneter Regressoren); für jeden der anderen Regressoren gilt $p < 0,0001$. Die Frage nach der Bedeutsamkeit des Einflusses einzelner exogener Variabler (und im Falle von Faktoren dem Einfluss einzelner-stufen) lässt sich mittels des Hazard-Verhältnisses beurteilen.

Tab. 6.6.: Diagnose-Gruppierung 001–009:
Verweildauermodellierung der 392.877 vollstationären Fälle 1996–2005

Parameter	Hazard-Verhältnis ^x	95 % Konfidenzschranken	
		untere	obere
ostdt=Ost	0,794	0,783	0,805
altgr=]1, 5]	1,434	1,421	1,449
altgr=]6, 10]	1,872	1,847	1,897
altgr=]11, 18]	1,957	1,933	1,982
altgr=]18, 30]	1,800	1,776	1,824
altgr=]30, 45]	1,264	1,247	1,282
altgr=]45, 65]	0,800	0,790	0,811
altgr=]65, 90]	0,479	0,473	0,484
altgr=]90, ∞]	0,479	0,466	0,493
geschl=weiblich	1,000	0,994	1,007
wohnkreis=ja	0,976	0,969	0,983
jahr	1,080	1,079	1,082
ewdL	1,049	1,041	1,057
alqlgt	0,843	0,831	0,856
betqL	1,151	1,101	1,203

^x Ein Hazard-Verhältnis > 1 entspricht hier einer größeren Chance auf zeitige Entlassung

Tab. 6.6 zeigt mit einem Hazard-Verhältnis von 1,000 für die Frauen (gegenüber den Männern als Referenzkategorie)²²⁰, dass das Geschlecht keinen Einfluss auf die Verweildauer hat.

Für die Kovariate *jahr* bezieht sich das Hazard-Verhältnis von 1,080 auf die Zunahme von einer Einheit in der Kovariaten und reflektiert, dass sich mit steigender Jahreszahl die Chance einer zeitigeren Entlassung erhöht, die Verweildauern sinken, für Ende gegen Anfang des Untersuchungszeitraumes (2005 gegen 1993) bedeutet das ein Hazard-Verhältnis von 2,518. Die 11- bis 18-Jährigen sind diejenigen mit der geringsten Verweildauer, gehört man zu einer Alterskategorie jenseits von 45 Jahren, so ist die erwartete Verweildauer länger als die für Säuglinge – die Referenzkategorie der Variablen *altgr*. Am längsten ist sie für die beiden Kategorien der über 65-Jährigen zu erwarten mit (je) einem Hazard-Verhältnis von $1/0,479 = 2,088$ zur Referenz.

Während für Patienten als Fälle in Krankenhäusern ihres Wohnkreises mit einem Hazard-Verhältnis von $1/0,976 = 1,025$ leicht kürzere Verweildauern vorherzusagen sind als für solche, die außerhalb des Wohnkreises behandelt werden, sind die Verweildauern in Kreisen mit hoher Brutto-Erwerbstätigenquote kürzer als in jenen mit niedrigerer. Obwohl signifikant, scheint diese Verkürzung eher irrelevant: für einen Anstieg vom Minimum der beobachteten Brutto-Erwerbstätigenquoten von 22 % auf das Maximum von 85 % ergibt sich ein Unterschied von 0,59 Einheiten auf der *betqL*-Skala, das Hazard-Verhältnis für diesen äußerst unwahrscheinlichen Anstieg beträgt mit 1,086 nur wenig mehr als das jahresweise für die Kovariate *jahr*. Ähnlich geringe Bedeutung kommt dem Wert 1,049 bei einer Verzehnfachung der Einwohnerdichte zu.

Eine höhere Arbeitslosenquote geht für das Untersuchungsgebiet einher mit höheren Verweildauern, niedrigere somit mit niedrigeren. Eine Senkung der Arbeitslosenquote von 20 % auf 10 % entspricht 0,81 Einheiten und einem Hazard-Verhältnis von $1/0,871 = 1,148$. Für die Fälle Ostdeutschlands ist die erwartete Verweildauer länger als für die Pendants aus Rheinland-Pfalz mit einem Hazard-Verhältnis von $1/0,794 = 1,259$.

Die gewichtete Betrachtung der Hazard-Verhältnisse stellt am ehesten die Bedeutung der einzelnen Variablen gegeneinander dar. Während der Einfluss des Alters

²²⁰ $e^{0,0002418} = 1,000242$.

vorherigen Erwartungen entspricht, erweisen sich unter den eingeschlossenen Variablen nur die Arbeitslosenquote im Kreis des Krankenhauses und die Frage, ob der Kreis zu Ostdeutschland gehört, als merkbar prädiktiv.

Modelle, stratifiziert nach Altersstufen

Alternativ zu Modellen mit Altersstufen als Faktor lässt sich durch eine nach Altersstufen stratifizierte Modellierung der Einfluss des Alters als beachtete Störvariable (Abschn. 5.1.8) beherrschen. Ein solches Modell erkennt das Alter weiterhin als einen wichtigen Einflussfaktor an, es wird jedoch versucht, für diese Störvariable zu korrigieren und diese in der Modellaussage von den interessierenden Einflussfaktoren zu trennen. Das (alters-)stratifizierte Cox-Modell setzt dabei statt der allgemeinen, für alle Subjekte gemeinsamen, Baseline kumulativen Hazard-Funktion $\Lambda_0(t_j)$ (Gl. 5.3) altersgruppenspezifische Baseline kumulative Hazard-Funktionen $\Lambda_g(t_j)$ für die Altersgruppen $g = 1, \dots, 9$ voraus.

$$\Lambda(t_{ij}) = \Lambda_g(t_j) \cdot e^{(\mathbf{x}_{ij}' \beta)} = \Lambda_g(t_j) \cdot e^{(\beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \dots + \beta_p X_{p,ij})}$$

bei ansonsten unverändertem gemeinsamem linearem Prädiktor $\eta_{ij} = \mathbf{x}_{ij}' \beta$.

Eine altersstratifizierte Analyse für die Diagnose-Gruppierung 001–009 (Infektiöse Krankheiten des Verdauungssystems) für die Jahre 1996–2005²²¹ mit ähnlichen Regressoren, aber ohne *ostdt*²²², ergibt wieder ein nichtsignifikantes Testergebnis für das Geschlecht und ähnliche Werte für die Hazard-Verhältnisse der anderen Variablen. Wird statt der Brutto-Erwerbstätigenquote mit dem Brutto-Inlandsprodukt je Einwohner (logarithmiert zur Basis 10) eine Variable mit Bezug zum Wohlstand in einem Kreis, ergibt sich bei einem um 20 % höheren Brutto-Inlandsprodukt je Einwohner²²³ ein Hazard-Verhältnis von 1,044.

Gleichzeitig verändert sich der Einfluss der Arbeitslosenquote: das Hazard-Verhältnis bei einer Senkung der Arbeitslosenquote von 20 % auf 10 % beträgt hier 1,305 (statt 1,148), geringere Arbeitslosenquoten gehen einher mit einer größe-

²²¹ wegen fehlender Jahresdaten zu einzelnen Regressoren (Abschn. 4.1.4).

²²² *ostdt* erweist sich vorab in einem Modell für die Jahre 1996–2005 als wenig prädiktiv.

²²³ 20 % entspricht 2005 etwa einem Aufstieg von der 1. zur 3. oder von der 3. zur 4. Quintilgrenze (Abschn. 5.1.10; Abb. 6.6).

ren Chance einer geringeren Verweildauer. Das ist weniger verwunderlich, denn die Arbeitslosenquote allein ist hier schon ein guter Indikator in allen Jahren für die Unterteilung in ost- und westdeutsche Kreise (Abb. 6.2 und zugehörige Ausführungen).

Für die anderen Diagnose-Gruppierungen ergeben sich ähnliche Ergebnisse. Allen Gruppierungen gemeinsam ist der Einfluss des Berichtsjahres jahr auf die Verweildauer – widergespiegelt durch Hazard-Verhältnisse größer als 1 –, auch wenn die jeweiligen Amplituden sehr unterschiedlich ausfallen. Das Geschlecht spielt als Prädiktor für Verweildauern kaum eine Rolle.

Keins der untersuchten wirtschaftlichen oder sozialen Merkmale als Regressoren zeigt einen über seinen beobachtbaren Wertebereich bedeutsamen Einfluss auf die Hazard-Funktion und somit die Verweildauer. Unterschiede zwischen Ost und West sind marginal, die Hazard-Verhältnisse schwanken zwischen den Gruppierungen um den Wert 1.

6.2.4. Diagnose-Stellung und -Registrierung

Die Hauptdiagnose ist das einzige unmittelbar die Krankheit bzw. den Grund für den Krankenhausaufenthalt beschreibende Merkmal (Abschn. 4.3.3: Diagnosedaten). Veränderungen in den registrierten Hauptdiagnosen über die Zeit werden mit dem Kompositionsabweichungsdiagramm (Abschn. 5.1.13) analysiert.

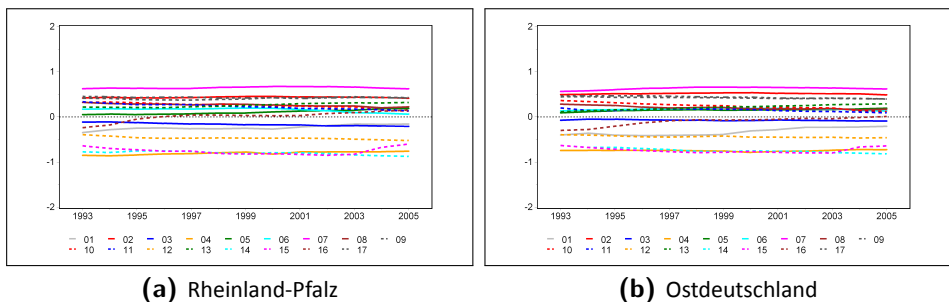


Abb. 6.15.: Diagnosen, ICD-9-Kapitel I–XVII – Kompositionsabweichungsdiagramme

Die Kompositionsabweichungsdiagramme der ICD-9-Kapitel I–XVII für Rheinland-Pfalz und Ostdeutschland (Abb. 6.15) zeigen deutlicher als die gestapelten Balkendiagramme (Abb. 6.11) die nur unwesentlichen Veränderungen in den Anteilen der einzelnen Diagnosekapitel an allen Diagnosen der Jahre 1993–2005. Ost und West unterscheiden sich kaum.

Wie bei der Analyse der Verweildauern (Abschn. 6.1.4: Verweildauern) werden für einen besseren medizinisch-epidemiologischen Bezug nach Ähnlichkeit zusammengefasste Diagnose-Gruppierungen ausgewählt. Für jede dieser Gruppierungen sind die Unterschiede zwischen Ost und West unbedeutend, alle weiteren Kompositionsabweichungsdiagramme beziehen sich auf das gesamte Untersuchungsgebiet.

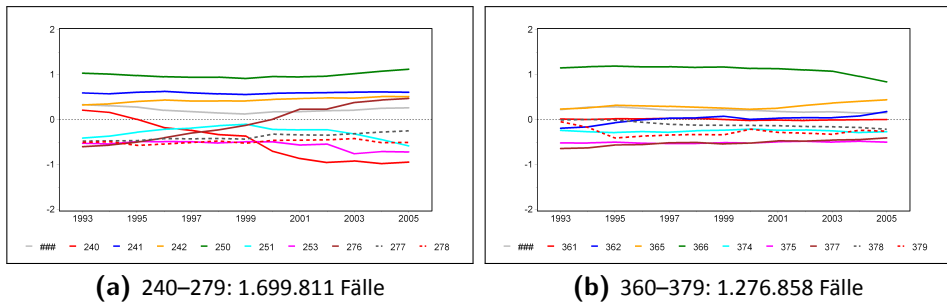


Abb. 6.16.: Diagnosen verschiedener Diagnose-Gruppierungen – Kompositionsabweichungsdiagramme^x

^x jeweils: 9 häufigste Diagnosen und Aggregat (###, per Amalgamation) aller anderen

Von den 20 verschiedenen Diagnosen der Diagnose-Gruppierung 360–379 (Affektionen des Auges und seiner Anhangsgebilde²²⁴) werden die 9 häufigsten in Abb. 6.16 b dargestellt. Die 11 weniger häufigen sind per Amalgamation zusammengefasst, sie haben gemeinsam nur die drittgrößte Häufigkeit²²⁵. Der weitgehend parallele Verlauf der Kurven der Einzelkomponenten zeigt wenig Veränderungen zwischen den Diagnosen dieser Gruppierung an.

Ein wenig anders stellt sich die Situation in der Gruppierung 240–279 (Endokrinoopathien, Ernährungs- und Stoffwechselkrankheiten sowie Störungen im Immun-

²²⁴ das komplette Kapitel VII der ICD-10.

²²⁵ Die häufigsten sind 366: Katarakt, 365: Glaukom.

system²²⁶, Abb. 6.16 a) dar: die 31 selteneren Diagnosen ergeben per Amalgamation wiederum nur die dritthäufigste Komponente, jedoch verändern sich die Häufigkeiten zweier Diagnosen²²⁷ stetig gegenläufig über den gesamten Zeitraum. Eine medizinisch-epidemiologische Begründung ist nicht bekannt.

Für Diagnose-Gruppierungen mit sehr vielen Einzel-Diagnosen ergibt sich bei Beschränkung auf 9 Komponenten, dass die Amalgam-Komponente aus den Komponenten geringster Häufigkeit größer ist als jede der 9 Einzelkomponenten, wie z. B. bei den 30 Einzel-Diagnosen der Gruppierung 680–709 (Krankheiten der Haut und des Unterhautzellgewebes, Abb. 4.3) sichtbar ist. In keiner der anderen betroffenen Gruppierungen 140–208 mit 69, 390–459 mit 70, 460–519 mit 60, 800–829 mit 30 oder 870–897 mit 28 Einzel-Diagnosen ergibt sich eine sichtbare Verschiebung zwischen den Häufigkeiten der Einzel-Diagnosen (ohne Abb.).

Beginnend mit dem Jahr 2002 ergeben sich hingegen gravierende Veränderungen in der Gruppierung 650–669 (Normale Entbindung, Komplikationen im Verlauf der Wehen und der Entbindung, Abb. 6.17) – der Diagnose-Gruppierung Geburtshilfe.

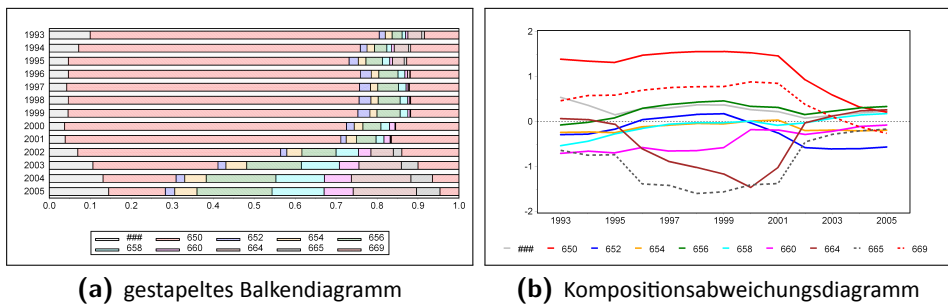


Abb. 6.17.: Diagnosen der -Gruppierung 650–669 (*Normale Entbindung, Komplikationen im Verlauf von Wehen und Entbindung; 1.856.776 Fälle*)^x

^x jeweils: 9 häufigste Diagnosen und Aggregat (###, per Amalgamation) aller anderen

Bei je Jahr zwischen 135.500 (2005) und 163.319 (2000) Geburten zeigt das gestapelte Balkendiagramm (a) die mit dem Jahr 2002 beginnende Abnahme der

²²⁶ das komplette Kapitel III in ICD-9 und ICD-10.

²²⁷ 240: Struma simplex und nicht näher bezeichnete Struma, 276: Störungen des Flüssigkeits-, Elektrolyt- und Säure-/Basen-Gleichgewichts.

Diagnose 650 (Normale Entbindung). Bis zum Jahr 2000 werden jährlich mehr als 68,5 % normale Entbindungen registriert. Nachdem 2001 dieser Anteil leicht auf 67,2 % sinkt, fällt dieser in den Folgejahren über 49,2 %, 30,4 % und 17,8 % auf nurmehr 13,8 % im Jahr 2005. Gleichzeitig sinkt der Anteil der Diagnose 669 (Sonstige Wehen- und Entbindungskomplikationen, anderweitig nicht klassifiziert) mit etwa 4,8 % deutlich unter das langjährige Mittel von knapp 12 %. Das Kompositionsabweichungsdiagramm (b) zeigt neben diesen abfallenden Anteilen deutlich die Diagnosen mit den stärksten Zunahmen im fraglichen Zeitraum: 664 (Verletzung des Dammes und der Vulva während der Entbindung) und 665 (Sonstige Geburtsverletzungen), deren Anteil zuvor – bis zum Jahr 2000 – stetig zurückgegangen ist. Am häufigsten sind für 2005 die Diagnosen 656 (Sonstige fetale und plazentare Störungen, die den Zustand der Mutter beeinträchtigen) mit 18,4 %, 664 mit 15,4 % und 658 (Sonstige Störungen im Zusammenhang mit Amnionhöhle und Eihäuten) während 665 einen Anteil von 7,7 % einnimmt²²⁸.

Eine Reflektion der Tatsache, dass binnen 6 Jahren der Anteil normaler Geburten von mehr als 68 auf weniger als 14 % sinkt und dabei die Verletzungen der Gebärenden unter der Geburt stark zunehmen, kann bisher²²⁹ weder in medizinisch-epidemiologischen Publikationen noch in der Tagespresse gefunden werden. Geht man davon aus, dass eine so deutliche Verschlechterung des Gesundheitszustandes der Gebärenden klinisch bemerkt und in Fachpublikationen intensiv diskutiert würde, müssen andere Erklärungen für diese Veränderungen gefunden werden. Zwischen klinischer Diagnose und Registrierung des Ergebnisses in (u. a.) der Krankenhausstatistik liegt die Kodierung nach ICD-9 bzw. ICD-10. Für die Jahre nach 2001 ist von einer gravierenden Änderung des Kodierverhaltens auszugehen.

Schlussfolgert man konsequenterweise, dass sich die Möglichkeit einer Veränderung im Kodierverhalten nicht auf die Diagnosen 650–669 (Normale Entbindung, Komplikationen im Verlauf der Wehen und der Entbindung) begrenzt²³⁰, so ist es wenig wahrscheinlich, dass sich hinter diesen ICD-9-Mustern noch sinnvoll Bezüge zur lokalen Wirtschaftslage herstellen lassen.

²²⁸ Zum Vergleich: Die 11 Diagnosen geringster Häufigkeit zusammen erreichen 2005 einen Anteil von 14,2 %.

²²⁹ Stand 2011.

²³⁰ z. B. Abb. A.4.

Von weiteren Modellierungen der Daten zur Hauptdiagnose wird hier Abstand genommen.

6.2.5. Wohn- und Behandlungsort

1993 entstammen 69,9 % der Fälle dem jeweiligen Krankenhauskreis²³¹, die Fälle kommen aus der unmittelbaren Umgebung. 12 Jahre später hat sich dieser Anteil auf 64,1 % verringert. Dabei unterscheiden sich Ost und West deutlich: während in Rheinland-Pfalz die Anteile monoton von 56,7 auf 48,5 % um 8,2 Prozentpunkte sinken, sind Ostdeutschland hingegen 1994 69,3 und 2005 67,5 % der Fälle²³² im Kreis des Krankenhauses wohnhaft. 1993 – ohne die Fälle aus Mecklenburg-Vorpommern – beträgt der Anteil sogar 73,6 %.

Eine Betrachtung der Diagnose-Gruppierungen zeigt, dass der Ost-West-Unterschied über die verschiedenen Krankheiten hinweg besteht: für infektiöse Krankheiten des Verdauungssystems (001–009) sind im Untersuchungszeitraum 56,8 % für Rheinland-Pfalz und 78,5 % für Ostdeutschland registriert, während die Zahlen für bösartige Neubildungen (140–208) bei 43,6 und 59,0 % in beiden Gebieten deutlich niedriger liegen.

Modelliert man den Anteil p_{ti} der von innerhalb des Kreises der Krankenhäuser kommenden Fälle für die Berichtsjahre $t \in \{1993, \dots, 2005\}$ und die Krankenhäuser i in einer Panelanalyse mit einer logistischen Regression für wiederholte Messungen, so genügt dies dem verallgemeinerten linearen Modell (Gl. 5.1) mit der Logit-Funktion $\text{logit}(p_{ti}) = \ln\left(\frac{p_{ti}}{1-p_{ti}}\right)$ als Link. Für den linearen Prädiktor $\eta_{ti} = \mathbf{x}_{ti}'\boldsymbol{\beta}$ ist $\boldsymbol{\beta}$ der Vektor von Regressionsparametern zu schätzen. Für die zu modellierenden Anteilswerte p_{ti} gilt $p_{ti} \in [0, 1]$; für die Abhängigkeit der Werte zwischen den Berichtsjahren wird mit AR(1) eine autoregressive Kovarianzstruktur der Ordnung 1 (Gl. 5.2) angenommen. Die Schätzungen erfolgen mit verallgemeinerten Schätzgleichungen (GEE).

Für jede Diagnose-Gruppierung separat untersucht ein sehr sparsames Modell den Einfluss des Indikators für Ostdeutschland und der Zeitvariable `jahr` als fes-

²³¹ Kreis, indem das Krankenhauses seinen Sitz hat; alle Kreise zum Gebietsstand 2005.

²³² fast monoton.

te Effekte sowie einer möglichen Wechselwirkung zwischen beiden. Als linearer Prädiktor wird

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr}$$

angesetzt. Dabei wird offensichtlich, dass die Wechselwirkung sehr gruppierungsspezifisch zu einer guten Modellanpassung beiträgt: während für die Entbindungen (650–669) wohl eine Annäherung der Anteile²³³ p_{ti} über die Jahre stattfindet (mit einem p-Wert der Wald-Statistik von 0,006, Abschn. 5.1.5: Modellvergleiche, Modellselektion und Auswahl geeigneter Regressoren), gilt dieses für die infektiöse Krankheiten des Verdauungssystems (001–009) mitnichten (Wald-p = 0,902).

Obwohl für alle Gruppierungen in Ostdeutschland im Mittel deutlich weniger Fälle von außerhalb des Krankenhauskreises stammen als in Rheinland-Pfalz, ist diese Differenz für die mehr als 5 Mio. Fälle mit bösartigen Neubildungen (140–208) nicht signifikant. Es muss von einer hohen Variabilität zwischen den Anteilen für die 69 verschiedenen Einzel-Diagnosen dieser Gruppierung ausgegangen werden.

Die auch in den Diagnose-Gruppierungen mit den Jahren im Wesentlichen monotone Abnahme der Fälle von innerhalb des Krankenhauskreises lässt eine Modellierung mit linearem zeitlichem Einfluss sinnvoll erscheinen.

Für die Fälle mit Krankheiten der Haut und des Unterhautzellgewebes (680–709) werden die Regressoren Ostdeutschland-Indikator, *jahr*, deren Wechselwirkung und die metrischen Kovariaten Einwohnerdichte als Urbanitätsmaß, BIP je Einwohner als Maß der Wirtschaftskraft (beide logarithmiert, *ewdL* und *bipewL*) und dem Logit der Arbeitslosenquote (*alqlgt*) als Maß der sozialen Lage eingeschlossen:

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} \\ + \beta_3 \cdot \text{ewdL} + \beta_4 \cdot \text{bipewL} + \beta_5 \cdot \text{alqlgt}.$$

Die Modellierung beruht auf den 592.117 Fällen der Jahre 1996–2005²³⁴. Beobachtungseinheit sind wiederum die einzelnen Krankenhäuser mit ihren Jahresangaben, alle Krankenhäuser eines Kreises haben dieselben Werte der Regressoren.

²³³ exakter: Logits der Anteile $\text{logit}(p_{ti}) = \ln\left(\frac{p_{ti}}{1-p_{ti}}\right)$.

²³⁴ Die Arbeitslosenquote steht erst ab 1996 zur Verfügung.

Beurteilt man die Bedeutung der Parameter mit Hilfe der Wald-Statistik, so hat die Arbeitslosenquote den geringsten prädiktiven Wert (Wald- $p = 0,757$) gefolgt von der Wechselwirkung (Wald- $p = 0,162$) und dem Ostdeutschland-Indikator mit Wald- $p = 0,119$. Bedeutsam erscheinen die Einwohnerdichte (Wald- $p = 0,023$), das Jahr (Wald- $p = 0,006$) und das BIP je Einwohner (Wald- $p < 0,0001$). Ausschlaggebend für den nicht so deutlichen Einfluss des Ostdeutschland-Indikators mag die Arbeitslosenquote im Modell sein, denn diese ist selbst ein guter Indikator für die Trennung zwischen Ost und West (Abb. 6.2 und zugehörige Ausführungen). Neben der Veränderung über die Zeit mit in späteren Jahren höheren Anteilen an Fällen von außerhalb zeigt sich, dass eine größere Einwohnerdichte mit einer Erhöhung und ein größeres BIP je Einwohner mit einer Verringerung des Anteils der Fälle von innerhalb des Krankenhauskreises einhergehen: mehr Urbanität geht mit mehr lokalen Fällen einher, eine größere Wirtschaftskraft bringt mehr Fälle von außen.

Bei der Anwendung desselben Modells auf andere Diagnose-Gruppierungen sieht man, dass einzig das BIP je Einwohner in jedem Modell seinen bedeutsamen Einfluss (Wald- $p = 0,0015$ oder kleiner) behält, Kreise größerer Wirtschaftskraft (Abb. 6.6) haben in ihren Krankenhäusern einen höheren Anteil an Fällen von außerhalb. Der prädiktive Wert aller anderen Merkmale für verschiedene Gruppierungen unterschiedlich, auch wenn für fast alle eine Verringerung des Anteils mit der Zeit vorhergesagt wird.

Für die Krankheiten des Kreislaufsystems (390–459) führt die Modellierung zu einem Ergebnis, das bisherigen Erkenntnissen zu widersprechen scheint: der Indikator und die Arbeitslosenquote haben einen bedeutsamen prädiktiven Wert für die Anteilsbetrachtung, laut Indikator ist der Anteil der Fälle von innerhalb des Krankenhauskreises in Ostdeutschland höher als in Rheinland-Pfalz, jedoch ist er in Kreisen höherer Arbeitslosenquoten geringer, was den in Abb. 6.2 gezeigten empirischen Ergebnissen widerspricht. Gleichzeitig spricht das Modell für einen geringen Einfluss der Zeitvariablen *jahr* (Wald- $p = 0,654$) mit hingegen bedeutsamer Zeit-Indikator-Wechselwirkung (Wald- $p = 0,001$), Kollinearitäten zwischen den Regressanden und hohe Variabilität zwischen den Anteilen der einzelnen Krankenhäuser mögen wesentliche Ursachen für diesen Modellierungsmisserfolg sein.

6.2.6. Beschäftigte in den Krankenhäusern

Angaben zu den im Krankenhaus Tätigen finden sich in der Krankenhausstatistik nach verschiedenen Erhebungsmodalitäten: ärztliches und nichtärztliches Personal werden getrennt berichtet, es erfolgt jeweils eine Erfassung nach der Personenanzahl ergänzt um die Angabe der *Vollkräfte im Jahresdurchschnitt insgesamt*. Letztere beschreibt – ähnlich den Personalkosten – den Personalaufwand unabhängig vom Geschlecht. Die Personenanzahl hingegen wird ohne Bezug auf die vertragsmäßige Arbeitszeit in mehreren Kategorien (ärztlich und nichtärztlich, je Station, nach Spezialisierung und für das Krankenhaus insgesamt) und nach Geschlecht erfasst. In diesem Abschn. wird über die in allen Krankenhäusern eines jeden Kreises (Abschn. 5.1.2: Gesamtkrankenhausesaktivität) beschäftigten Personen berichtet.

Zum Vergleich: Die Deutsche Krankenhausgesellschaft berichtet²³⁵ für Deutschland 1993 von 111.608 Krankenhausärzten bei 95.460 Vollbeschäftigten, d. h. einer Personen-Vollbeschäftigten-Rate von 118,2 %. 2003 beträgt diese bei 114.105 Vollbeschäftigten und einer auf 128.853 gestiegenen Anzahl an Ärzten nur noch 114,2 %. Damit sind die Veränderungen für die Ärzte andere als die für nichtärztliche Krankenhausangestellte, deren absolute Anzahlen bei steigenden Raten sinken: 1993 beträgt die Rate 119,2 % (928.858 gegen 779.470), 2003 mit 125,4 % (890.122 gegen 709.834) sind es 6,2 Prozentpunkte mehr.

Ärztliches Personal

Die Entwicklung der männlichen und weiblichen Ärztezahlen lässt sich mit einer Panelanalyse herausarbeiten. Regressand ist hierbei die (logarithmierte) einwohnerbezogene Ärztedichte²³⁶ der Krankenhausärzte eines jeden Kreises – nach Geschlecht getrennt. Es wird angenommen, dass diese Log-Ärztedichte einer Gauss-

²³⁵ eigene Berechnungen auf den Angaben von <http://www.dkgev.de/pdf/856.pdf>, Zugriff am 26. 11. 2011.

²³⁶ bezogen auf je 1000 Einwohner.

Verteilung genügt, Linkfunktion dieser Panelanalyse ist somit die Identität. Regressoren in

$$\begin{aligned} \eta = & \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_3 \cdot \text{geschl} \\ & + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} + \beta_{1,3} \cdot \text{ostdt} \cdot \text{geschl} + \beta_{2,3} \cdot \text{jahr} \cdot \text{geschl} \\ & + \beta_{1,2,3} \cdot \text{ostdt} \cdot \text{jahr} \cdot \text{geschl} \end{aligned}$$

sind der Ostdeutschland-Indikator, die metrische Variable *jahr*, das Geschlecht (*geschl*) und deren Wechselwirkungen. Die Wechselwirkung zweiter Ordnung (Rasch, 1987, S. 716 ff.) zwischen allen drei Faktoren (mit β_{123} als Koeffizienten) ermöglicht dem Modell eine gute Anpassung, falls sich für Ärzte geschlechtsspezifisch unterschiedliche Veränderungen über die Zeit ergeben. Als Kovarianzstruktur zwischen den wiederholten Messungen je Kreis wird mit ARH(1) eine heterogen autoregressive Kovarianzstruktur der Ordnung 1 (Gl. 5.2) mit potenziell unterschiedlichen Varianzen zu den einzelnen Zeitpunkten angenommen. Die Parameter werden mit REML geschätzt.

Die Wald-Statistik (Abschn. 5.1.5: Modellvergleiche, Modellselektion und Auswahl geeigneter Regressoren) spricht mit einem p-Wert von 0,005 dafür, dass auf die Wechselwirkung zweiter Ordnung im Modell nicht verzichtet werden sollte.²³⁷

Die Schätzungen für die heterogen autoregressive Kovarianzstruktur zeigen, dass die Varianz mit den Jahren leicht abnimmt (1992: 0,106, 2005: 0,087). Ein sparsameres Modell mit der Annahme einer (homogen) autoregressiven Kovarianzstruktur kann die Daten fast genauso gut beschreiben.

Die Autokorrelation wird mit 0,985 sehr hoch geschätzt, die Werte eines Kreises ändern sich von einem zum nächsten Jahr eher wenig.

Tab. 6.7 zeigt die Modell-Prädiktionen der Ärztinnen- bzw. Ärztedichten je 1000 Einwohner für ausgewählte Jahre mit 95 %-Konfidenzintervallen. Der modellierte Trend zeigt für Rheinland-Pfalz eine Steigerung der Ärztedichte von 0,791 (1991) auf 1,163 Ärzte je 1000 Einwohner (2005), für Ostdeutschland liegen die mittleren Verhältnisse ähnlich: 0,762 bzw. 1,149 sind die entsprechenden Dichtewerte.

²³⁷ Damit sind alle vorstehenden Summanden – die 3 Wechselwirkungen erster Ordnung sowie jeder Einzelfaktor *ostdt*, *jahr* bzw. *geschl* – für die Modell-Interpretation unabdingbar notwendig.

Tab. 6.7.: Modellierung der Ärzte je 1000 Einwohner, basierend auf den Daten 1992–2005

Gebiet	Geschlecht	Jahr	Prädiktion ^x	95 % KI ^y	Kehrwert ^z
RP	männlich	1991	0,579	[0,448, 0,747]	1728
		1995	0,620	[0,489, 0,786]	1613
		2000	0,675	[0,539, 0,846]	1480
		2005	0,736	[0,586, 0,924]	1359
	weiblich	1991	0,212	[0,164, 0,274]	4717
		1995	0,259	[0,204, 0,328]	3861
		2000	0,333	[0,265, 0,417]	3006
		2005	0,427	[0,340, 0,536]	2340
Ostdt	männlich	1991	0,452	[0,391, 0,521]	2214
		1995	0,502	[0,439, 0,574]	1992
		2000	0,573	[0,505, 0,650]	1745
		2005	0,654	[0,576, 0,743]	1529
	weiblich	1991	0,310	[0,268, 0,357]	3230
		1995	0,354	[0,310, 0,404]	2826
		2000	0,418	[0,369, 0,475]	2390
		2005	0,495	[0,435, 0,562]	2022

^x Krankenhausärzte je 1000 Einwohner

^y 95 % Konfidenzintervall zur Prädiktion [$1/1000 \text{ EW}$]

^z Einwohner je Krankenhausarzt – der reziproke Wert der Prädiktion

Genauso nimmt die Ärztedichte für beide Geschlechter jeweils in Ost und West zu. Deutlich jedoch treten die Unterschiede sowohl zwischen den Geschlechtern als auch zwischen den Rheinland-Pfalz und Ostdeutschland zutage. Dabei sind die Unterschiede zwischen den Dichten an Ärztinnen und derer an ihren männlichen Kollegen in Ostdeutschland viel geringer als in Rheinland-Pfalz; eine merkbare Nivellierung dieser Unterschiede findet jedoch weder in Ost noch West statt. Die Konfidenzintervalle sind relativ weit, der Anteil unerklärter Varianz ist groß.

Die logistische Modellierung des Anteils der weiblichen an allen Krankenhausärzten per Panelanalyse zeigt für die Regressoren Ostdeutschland-Indikator, Zeitvariable *jahr* und deren Wechselwirkungen

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr}$$

Tab. 6.8.: Modellierung des Anteils der Ärztinnen an allen Krankenhausärzten, basierend auf den Daten 1992–2005

Gebiet	Jahr	Prädiktion ^x	95 % KI ^y
RP	1991	25,8	[23,8, 27,9]
	1995	28,5	[26,9, 30,1]
	2000	32,0	[30,6, 33,4]
	2005	35,7	[33,8, 37,7]
Ostdt	1991	40,5	[39,0, 41,9]
	1995	41,0	[40,0, 42,1]
	2000	41,8	[40,9, 42,7]
	2005	42,5	[41,4, 43,7]

^x Anteil [%]

^y 95 % Konfidenzintervall zur Prädiktion [%]

adäquaterweise, dass der Frauenanteil unter den Krankenhausärzten in Ost und West zunimmt, der Eindruck aus Abb. 6.10 a bestätigt sich. Eine Angleichung ist für die unmittelbar dem Untersuchungszeitraum folgenden Jahre nicht zu erwarten.

Abb. 6.18 zeigt die Unterschiede zwischen den Kreisen in Quintilkarten²³⁸. Zur Einhaltung der Anonymisierungsanforderungen (Abschn. 4.1.2) erscheinen die Kreise, in denen weniger als 3 Krankenhäuser ihren Sitz haben, weiß eingefärbt, ihre Werte fließen jedoch in die Quintilberechnungen ein. Sichtbar wird, dass der Anteil solcher Kreise mit den Jahren zunimmt: während dies 1992 für 55 von 147 Kreisen (37,4 %) zutrifft, haben 2005 deutlich mehr als die Hälfte (87, 59,2 %) aller Kreise weniger als drei Krankenhäuser (Tab. 6.2).

Für die darstellbaren Kreise zeigt sich neben dem oben modellierend herausgearbeiteten Ost-West-Unterschied, dass die Kreise mit besonders hohem Ärztenanteil eher in strukturschwachen ländlichen Gegenden liegen und gerade nicht die Urbanitätszentren sind.

²³⁸ Aus Mecklenburg-Vorpommern fehlen die Grund- und Kostendaten 1998–2001 komplett.

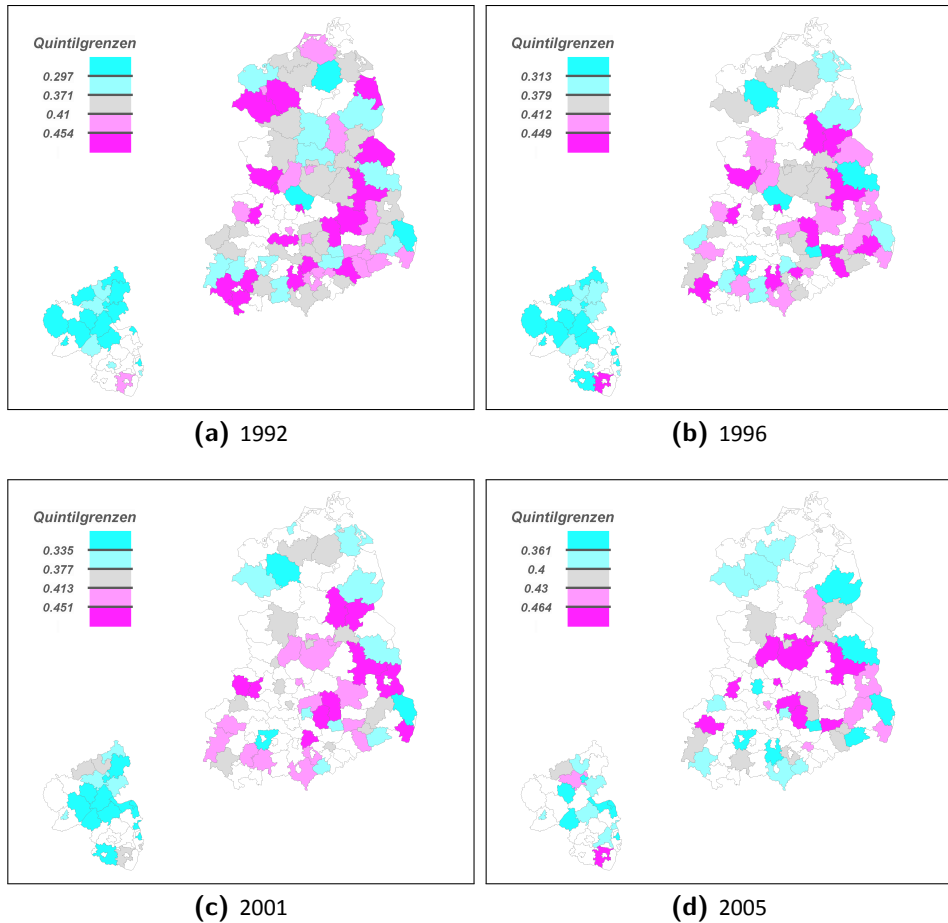


Abb. 6.18.: Krankenhausärztinnen [Anteil, Quintilkarten*]

* Kreise mit weniger als 3 Krankenhäusern sind von der Darstellung ausgeschlossen, sie erscheinen weiß.

Nichtärztliches Personal

Die Anwendung derselben Modelle auf die Angaben zum nichtärztlichen Krankenhauspersonal ergeben (Tab. A.1 und A.2) für die heterogen autoregressive Kovarianzstruktur gleichfalls ein Sinken der Varianz mit den Jahren (1992: 0,093, 2005: 0,079), die Autokorrelation wird mit 0,987 ähnlich hoch geschätzt.

Die Dichte an nichtärztlichen Krankenhaus-Mitarbeitern (je 1000 Einwohner) zeigt für Rheinland-Pfalz einen zunehmenden (1991: 9,04, 2005: 9,38) Trend. Für Ostdeutschland hingegen ergibt sich ein abnehmender Trend (1991: 7,90, 2005: 7,21) trotz der Zunahme bei den männlichen Angestellten (1991: 0,93, 2005: 0,98). Die in Abb. 6.10 b dargestellten Daten werden durch das Modell gut beschrieben.

6.2.7. Bettendichte und Belegungsdichte

Die Angaben zur Bettenanzahl und zu den Belegungstagen gehören zu den Grunddaten in der Krankenhausstatistik, sie dienen der fallunabhängigen Beschreibung der Krankenhäuser. Während die Bettenanzahl das Angebot bzw. das Versorgungspotenzial der Krankenhäuser beschreibt, sind die Belegungstage ein Ausdruck der erbrachten Leistungen; zum Vergleich zwischen Krankenhäusern oder Kreisen sind beide zu standardisieren, z. B. auf die Zahl der Fälle oder der Einwohner für das Angebot bzw. die Zahl der Fälle oder der Betten für die Leistungen.

Zur Beschreibung der kreisweisen kumulativen Krankenhaussituation (Abschn. 5.1.2: Gesamtkrankenhaussituation) werden Bettenanzahl und Belegungstage aggregiert; die Bettendichte ist die Bettenanzahl bezogen auf die Einwohner, die Belegungsdichte die Anzahl der Belegungstage bezogen auf die aufgestellten Betten.

Bettendichte

Die logarithmierte Bettendichte lässt sich gut durch eine Gauss-Verteilung beschreiben; Linkfunktion dieser Panelanalyse ist somit die Identität. Ein sehr sparsames Modell der Bettendichte nutzt den Ostdeutschland-Indikator, die Zeitvariable *jahr* und deren Wechselwirkungen als Regressoren und die Kreise als statistische Einheiten mit wiederholten Beobachtungen:

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr}.$$

Das Modell basiert auf den Daten von 1992 bis 2005. Tests der Regressoren als feste Effekte ergeben für den Indikator einen p-Wert der Wald-Statistik (Abschn. 5.1.5: Modellvergleiche, Modellselektion und Auswahl geeigneter Regressoren) von 0,969, für die Zeitvariable *jahr* von $< 0,001$ und für die Wechselwirkung von 0,123; ein Modell ohne die Wechselwirkung ($\beta_{1,2} \equiv 0$) schiene den Daten gleichfalls gut zu entsprechen.

Tab. 6.9.: Modellierung der Bettendichte, basierend auf den Daten 1992–2005

Gebiet	Jahr	Prädiktion ^x	95 % KI ^y
RP	1995	6,838	[5,761, 8,115]
	2000	6,439	[5,431, 7,633]
	2005	6,063	[5,087, 7,226]
Ostdt	1995	7,075	[6,425, 7,789]
	2000	6,464	[5,875, 7,110]
	2005	5,905	[5,352, 6,515]

^x Bettendichte [$1/1000 \text{ EW}$]

^y 95 % Konfidenzintervall zur Prädiktion [$1/1000 \text{ EW}$]

Tab. 6.9 zeigt die Prädiktionen für ausgewählte Jahre mit Konfidenzintervall, die Modellergebnisse bestätigen den mit Abb. 6.7 a gewonnenen Eindruck: Die Bettendichte sinkt. Jenseits der Sichtbarkeit zeigt das Modell einen gegenüber Rheinland-Pfalz stärkeren Abfall für Ostdeutschland, von deutlich höheren Werten in den neunziger Jahren zu einem deutlich niedrigeren im Jahr 2005.

Ein noch sparsameres Modell mit einzig *jahr* als Regressor ist somit trotz nicht signifikantem Ost-West-Unterschied nicht angebracht, denn die Werte der Prädiktion führten zur Schlussfolgerung derselben (mittleren) Verringerung der Bettendichte für Ost und West.

Im Gegenteil: Erweitert man das Modell um die (logarithmierte) Einwohnerdichte als metrischer Kovariater

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} + \beta_3 \cdot \text{ewdL}$$

und versucht damit, die hohe Variabilität zwischen den Kreisen beider Teilgebiete (Abb. 6.7 a) besser zu erklären, so ergibt sich mit den Prädiktionen der Tab. 6.10 eine deutlich bessere Anpassung an die Daten, der Wert des Bayes'schen Infor-

Tab. 6.10.: Modellierung der Bettendichte in Abhängigkeit der Einwohnerdichte, basierend auf den Daten 1992–2005

Gebiet	EW-Dichte ^x	Jahr	Prädiktion ^y	95 % KI ^z
RP	100	1995	5,657	[4,811, 6,654]
		2005	4,988	[4,218, 5,897]
	1000	1995	8,758	[7,374, 10,40]
		2005	7,720	[6,480, 9,196]
Ostdt	100	1995	6,301	[5,745, 6,912]
		2005	5,325	[4,853, 5,844]
	1000	1995	9,754	[8,461, 11,25]
		2005	8,243	[7,112, 9,554]

^x Einwohnerdichte [$1/km^2$]

^y Bettendichte [$1/1000\text{ EW}$]

^z 95 % Konfidenzintervall zur Prädiktion [$1/1000\text{ EW}$]

mationskriteriums sinkt um 5,9. Ein Teil der Variabilität innerhalb von Rheinland-Pfalz bzw. Ostdeutschlands kann durch Unterschiede in der Einwohnerdichte modelliert werden, urbanere Kreise haben überall eine höhere Bettendichte als die weniger urbanen, im Untersuchungszeitraum ändert sich daran nichts. Dieses nur um einen Regressor bereicherte Modell führt gleichzeitig zu einer den Ergebnissen besser entsprechenden Bewertung der festen Effekte der Regressoren: Kovariate $ewdL$, Wechselwirkung $ostdt \cdot jahr$ und Zeitvariable $jahr$ liefern jede Wald-Statistik-p-Werte kleiner als 0,0001, der Indikator $ostdt$ liegt bei 0,061.

Belegungsdichte

Die mittlere Belegungsdichte wird in Tagen je aufgestelltem Krankenhausbett und Jahr [d/a] angegeben. Modelliert wird sie mit einem verallgemeinerten linearen Modell, wobei für die Belegungstage eine negative Binomialverteilung angenommen wird. Die \log_e -Funktion ist ein passender Link. Der Bezug der Belegungstage auf die Betten wird durch den sogenannten *offset* hergestellt: die \log_e -logarithmierte Anzahl an Betten ($bettL$) wird mit einem festem Koeffizienten $\equiv 1$ Bestandteil des linearen Prädiktors η .

Die Parameter des Paneldaten-Modells werden mit GEE geschätzt; als Kovarianzstruktur zwischen den wiederholten Messungen je Kreis ist mit AR(1) eine autoregressive Kovarianzstruktur der Ordnung 1 (Gl. 5.2) mit gleichen Varianzen zu den einzelnen Zeitpunkten²³⁹ angenommen.

Tab. 6.11.: Prädiktionen der Belegungsdichte [d/a], basierend auf den Daten 1992–2005

Gebiet	1992	1996	2000	2005
RP	310,1	292,9	276,8	257,8
Ostdt	282,9	284,9	287,0	289,6

Ein sehr sparsames Modell

$$\eta = 1 \cdot \text{bettL} + \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr}$$

über die Kreise als statistische Einheiten mit wiederholten Beobachtungen nutzt wiederum den Ostdeutschland-Indikator, die Zeitvariable *jahr* und deren Wechselwirkungen als Regressoren und *bettL* als *offset*. Das Modell basiert auf den Daten von 1992 bis 2005. Tests der Regressoren als feste Effekte ergeben für den Indikator einen p-Wert von 0,001, für Zeitvariable *jahr* und Wechselwirkung von jeweils < 0,001. Tab. 6.11 gibt den Überblick über die Prädiktionsunterschiede zwischen Ostdeutschland und Rheinland-Pfalz.

Erweitert man das Modell um die zeitkonstante Angabe des Regionsgrundtyps (*regtyp*) eines jeden Kreises (Abschn. 3.3.5 sowie Abb. 4.1) in Form zweier Indikatorvariablen (*regtyp₁* und *regtyp₂*) zu

$$\eta = 1 \cdot \text{bettL} + \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} + \beta_{31} \cdot \text{regtyp}_1 + \beta_{32} \cdot \text{regtyp}_2,$$

so verbessert sich die Modellanpassung; der p-Wert der Wald-Statistik (Abschn. 5.1.5: Modellvergleiche, Modellselektion und Auswahl geeigneter

²³⁹ Ähnlich zu den Modellen in Abschn. 6.2.6 verbessert eine allgemeinere ARH(1)-Struktur das Modell kaum.

Regressoren) beträgt für *regtyp* 0,003. Für die Kreise der Agglomerationsräume werden höhere, für die Kreise im ländlichen Raum geringere Belegungsdichten prognostiziert. Die Differenz der Parameterschätzungen zwischen Agglomerations- und Ländlichem Raum beträgt 0,027. Für Rheinland-Pfalz im Jahr 2000 sind die Prädiktionen 281,2 d/a (Agglomerationsraum), 275,8 d/a (Verstädterter Raum) bzw. 273,8 d/a (Ländlicher Raum). Ostdeutschlands Prädiktionen für 2000 sind 291,7 d/a , 286,1 d/a bzw. 284,1 d/a .

Neben dem Regionsgrundtyp wird vom Bundesamt für Bauwesen und Raumordnung jedem Kreis einer von 9 Kreistypen zugeordnet. Ersetzt man im Modell den Regressor Regionsgrundtyp durch 8 Indikatorvariablen für den Kreistyp *krstyp*

$$\begin{aligned} \eta = & 1 \cdot \text{bettL} \\ & + \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} \\ & + \beta_{41} \cdot \text{krstyp}_1 + \dots + \beta_{48} \cdot \text{krstyp}_8, \end{aligned}$$

so beträgt die maximale Differenz der Parameterschätzungen 0,063; besonders hohe Werte weisen die Kernstädte der Agglomerationen und der Verstädterten Räume und besonders niedrige die drei verschiedenen Typen ländlicher Kreise auf, dementsprechend geringer sind die prognostizierten Werte der Belegungsdichte. Die Wald-Statistik liefert für den Kreistyp einen p-Wert von $< 0,0001$.

Der Einschluss von zeitveränderlichen sozio-ökonomischen Merkmalen wie (dem Logit) der Arbeitslosenquote (*alqlgt*), der (logarithmierten) Brutto-Erwerbstätigenquote (*betqL*) oder dem (logarithmierten) BIP je Einwohner (*bipewL*) in das Modell ergibt jeweils für diesen Regressor einen Wald-Statistik-p-Wert von 0,1 oder deutlich darüber: Es kann weder den lokalen wirtschaftlichen Rahmenbedingungen und noch den lokalen sozialen Gegebenheiten ein bedeutsamer Einfluss auf die Belegungsdichte der Krankenhausbetten zugeschrieben werden.

Lediglich der Einschluss der (logarithmierten) Einwohnerdichte (*ewdL*) als Urbanitätsmaß

$$\begin{aligned} \eta = & 1 \cdot \text{bettL} \\ & + \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} \\ & + \beta_{41} \cdot \text{krstyp}_1 + \dots + \beta_{48} \cdot \text{krstyp}_8 + \beta_5 \cdot \text{ewdL} \end{aligned}$$

führt – bei weiterhin p-Werten der Wald-Statistik von $< 0,0001$ für *ostdt*, *jahr* und deren Wechselwirkung – einem relativ kleinen p-Wert von 0,002 für *ewdL* bei gleichzeitiger Erhöhung des p-Wertes für den Kreistyp auf 0,003. Da sich der vom Bundesamt für Bauwesen und Raumordnung bestimmte Kreistyp (Abschn. 3.3.5) zu einem wesentlichen Anteil gleichfalls auf die Einwohnerdichte bezieht, wird damit die bereits dargestellte Abhängigkeit der Belegungsdichte von der Urbanität bestätigt: Die Belegungsdichte ist in ländlicheren Kreisen eher geringer als in urbaneren Kreisen mit höherer Einwohnerdichte.

6.2.8. Kosten der einzelnen Krankenhäuser

Sowohl die bereinigten Kosten²⁴⁰ als auch das Personal-Sachkosten-Verhältnis²⁴¹ (Abb. 6.9²⁴²) lassen sich als ökonometrisch interessante Regressanden in Paneldaten-Regressionen modellieren.²⁴³ Die Krankenhäuser des Untersuchungsgebietes sind hinsichtlich ihrer Größe sehr unterschiedlich. Legt man die Anzahl der abgehenden Fälle je Jahr zugrunde, so beträgt das geometrische Mittel der empirischen Verteilung über alle Berichtsjahre 4572, das 5. Perzentil liegt bei 279 und das 95. bei 25668 Fällen, Maximum und Minimum unterscheiden sich um einen Faktor größer als 100.²⁴⁴ Eine bessere Vergleichbarkeit großer und kleiner Krankenhäuser wird durch Betrachtung der relativen bereinigten Kosten bezogen auf die Fallzahl (Abb. 6.8 d²⁴⁵) hergestellt.

Krankenhäuser bzw. Kreise²⁴⁶ als statistische Einheiten geben potenziell unterschiedliche Einblicke in die Beziehungen zum wirtschaftlichen Umfeld.²⁴⁷ Viele

²⁴⁰ deflationiert (Abschn. 4.4.3: Deflationierung der Kostendaten).

²⁴¹ im Folgenden formuliert als Anteil der Personalkosten an der Summe aus Personal- und Sachkosten.

²⁴² für die kreisbezogene Gesamtkrankenhausaktivität.

²⁴³ Bundeswehrkrankenhäuser berichten keine Kosten, sie sind somit nicht Anteil der Kostenmodellierungen.

²⁴⁴ Diese Überlegungen gelten für die kreisweisen Gesamtkrankenhausaktivitäten gleichfalls.

²⁴⁵ für die kreisbezogene Gesamtkrankenhausaktivität.

²⁴⁶ Abschn. 6.2.9.

²⁴⁷ Während das Panel der Kreise im Untersuchungszeitraum kaum Veränderungen bei den Panelsubjekten aufweist, ändert sich das Panel der Krankenhäuser deutlicher durch Attrition, Zuwachs und Identifikations-Diskontinuitäten seiner Subjekte (Abschn. 4.3.3: Identifikation der Krankenhäuser – Paneldaten).

Kreise umfassen weniger als 3 Krankenhäuser. Deshalb werden hier keine hierarchischen bzw. Mehrebenen-Modellierungen mit den Kreisen als Ebene oberhalb der Krankenhäuser vorgenommen.

Jede der hier (Modelle 1 bis 3) und in Abschn. 6.2.9 (Modelle 4 und 5, mit Kreisen als Panelsubjekte) diskutierten Panel-Regressionen wird getrennt für die zwei interessierenden Regressanden – die relativen bereinigten Kosten sowie das Personal-Sachkosten-Verhältnis – durchgeführt.

Sowohl die \log_{10} -transformierten relativen bereinigten Kosten wie auch das \log_e -transformierte Personal-Sachkosten-Verhältnis²⁴⁸ ($\ln\left(\frac{\text{Personalkosten}}{\text{Sachkosten}}\right)$) lassen sich durch Gauss-Verteilungen adäquat beschreiben. Als Linkfunktion wird somit die Identität gewählt, die Paneldaten-Modelle sind allgemeine lineare Modelle, sie werden mit REML (residual maximum likelihood) geschätzt. Die Abhängigkeitsstruktur der wiederholt gemessenen Panelsubjekte wird als Varianzkomponenten-Modell mit AR(1)- bzw. ARH(1)-Kovarianzstruktur (Gl. 5.2) modelliert.

Modelle 1 und 2 beschreiben die Kostensituation der Krankenhäuser anhand krankenhausspezifischer Parameter, diese Modelle unterstützen die Interpretation in Abschn. 6.3.1: Die Kosten der Krankenhäuser in Transition, während Modell 3 Beziehungen zwischen (lokaler) Wirtschaftsleistung und Krankenhäusern untersucht (Abschn. 6.3.1: Der Einfluss der Wirtschaft auf die Krankenhäuser).

Bereinigte Kosten der Krankenhäuser – Modell 1

Modelliert man die logarithmierten bereinigten Kosten der Jahre 1992–2005 mit den Krankenhäusern als Panelsubjekte linear aus den Regressoren Ostdeutschland-Indikator (*ostdt*), Berichtsjahr (*jahr*), deren Wechselwirkungen sowie der logarithmierten Anzahl der abgehenden Fälle (*fallzL*)

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} + \beta_3 \cdot \text{fallzL},$$

so fällt auf, dass die ARH(1)-Kovarianzstruktur für 1998 und 2001 – das erste bzw. letzte Jahr ohne Angaben aus Mecklenburg-Vorpommern – Varianzen von

²⁴⁸ Das \log_e -transformierte Personal-Sachkosten-Verhältnis entspricht dem Logit von $\frac{\text{Personalkosten}}{\text{Personalkosten} + \text{Sachkosten}}$.

größer als 0,33 schätzt, während diese für jedes der anderen Jahre kleiner als 0,18 sind.²⁴⁹ Die Autokorrelation beträgt 0,961. Das Bayes'sche Informationskriterium hat einen Wert von -9539,7; für die Regressoren ergibt sich eine Bewertung wie in Tab. 6.12 dargestellt. Die Waldsche t-Statistik adressiert die Frage, ob der jeweilige Effekt von 0 verschieden ist.

Tab. 6.12.: Krankenhäuser 1992–2005, Kostenmodellierung #1
Parameterschätzungen (6.089 Beobachtungseinheiten)

Effekt	Schätzung	StdErr ^x	FG ^y	t	Wk ^z
ostdt=West	-0,07349	0,01411	1197	-5,21	<0,0001
jahr	0,004264	0,000953	2342	4,47	<0,0001
jahr·(ostdt=West)	-0,00131	0,002016	2412	-0,65	0,516
fallzL	-0,3028	0,007696	2771	-39,35	<0,0001

^x Standardfehler der Schätzung

^y Freiheitsgrade

^z Ablehnungswahrscheinlichkeit der t-Statistik

Die Wechselwirkung zwischen Berichtsjahr und der Zugehörigkeit des Krankenhauses zu Rheinland-Pfalz oder Ostdeutschland ist hierbei von eher untergeordneter Bedeutung, allen anderen Merkmalen wird ein bedeutender Einfluss auf die Kosten zugeschrieben: die fallbezogenen Kosten steigen mit den Jahren und liegen in Ostdeutschland höher als in Rheinland-Pfalz, sie sind geringer bei Krankenhäusern höherer Fallzahlen.

Quantitativ ergibt sich ein etwas differenzierteres Bild (Tab. 6.13). Die Prädiktionen sprechen für die Jahre 1992–2005 von einer Zunahme der deflationierten Kosten je Fall auf etwa 118 %. Die Unterschiede in den Kosten je Fall zwischen großen und kleinen Krankenhäusern liegen bei 150 % für ein 5.000er Krankenhaus im Vergleich zu einem viermal so großen; diese übersteigen die Ost-West-Unterschiede von 112 % zu 100 % deutlich.

²⁴⁹ Ob diese Varianz-Inhomogenitäten allein auf das Datenausfallmuster zurückzuführen sind, muss hier offen bleiben.

Tab. 6.13.: Krankenhäuser 1992–2005, Kostenmodellierung #1
Prädiktionen (6.089 Beobachtungseinheiten)

Jahr	Gebiet	Fälle	Prädiktion ^x
1992	RP	5.000	2.594
		20.000	1.705
	Ostdt	5.000	2.999
		20.000	1.971
2005	RP	5.000	2.834
		20.000	1.863
	Ostdt	5.000	3.408
		20.000	2.240

^x bereinigte Kosten je Fall [€, deflationiert]. Dabei liegt die obere 95 %-Konfidenzschranke für Ostdeutschland beim 1,04-fachen, für Rheinland-Pfalz beim 1,07-fachen der jeweiligen Prädiktion, die untere demzufolge beim $\frac{1}{1,04}$ - bzw. $\frac{1}{1,07}$ -fachen des Prädiktionwertes.

Bereinigte Kosten der Krankenhäuser – Modell 2

Erweitert man das Modell um die Regressoren Krankenhaustyp ($ktyp$, 3 Indikatorvariablen), Krankenhausträger ($trae$, 2 Indikatorvariablen), deren Wechselwirkung sowie die Kovariaten (logarithmierte) Anzahl vollbeschäftigter Ärzte ($arztzL$) und (logarithmierte) Anzahl Vollbeschäftigter des nichtärztlichen Personals ($naerzL$)

$$\begin{aligned} \eta = & \beta_0 + \beta_1 \cdot ostdt + \beta_2 \cdot jahr + \beta_{1,2} \cdot ostdt \cdot jahr + \beta_3 \cdot fallzL \\ & + \beta_{41} \cdot ktyp_1 + \beta_{42} \cdot ktyp_2 + \beta_{43} \cdot ktyp_3 + \beta_{51} \cdot trae_1 + \beta_{52} \cdot trae_2 \\ & + \beta_{41,51} \cdot ktyp_1 \cdot trae_1 + \dots + \beta_{43,52} \cdot ktyp_3 \cdot trae_2 \\ & + \beta_6 \cdot arztzL + \beta_7 \cdot naerzL, \end{aligned}$$

dann kann wegen des mit -13373,8 deutlich geringeren Wertes des Bayes'schen Informationskriteriums (BIC) auf nahezu gleichbleibenden Beobachtungseinheiten (5.916) von einem deutlich besser angepassten Modell ausgegangen werden (Tab. 6.14).

Wiederum fällt auf, dass die heterogen autoregressive ARH(1)-Kovarianzstruktur für 1998 und 2001 – das erste bzw. letzte Jahr ohne Angaben aus Mecklenburg-

Tab. 6.14.: Krankenhäuser 1992–2005, Kostenmodellierung #2
Parameterschätzungen (5.916 Beobachtungseinheiten)

Effekt (Stufen)	Schätzung	StdErr ^x	FG ^y	F	Wk ^z
ostdt=West	0,008888	0,004820	1, 1802	3,40	0,0654
jahr	0,007035	0,000531	1, 1767	71,47	<0,0001
jahr·(ostdt=West)	-0,00464	0,001084	1, 1650	18,32	<0,0001
ktyp=Universitäts-Klinik	0,04126	0,03610	3, 2856	3,17	0,0234
ktyp=Plankrankenhaus	0,07560	0,01429			
ktyp=KH mit Vers.-Vertrag	0,08103	0,01676			
trae=öff.	-0,01250	0,03605	2, 2542	34,74	<0,0001
trae=freig.	0,3089	0,02604			
(ktyp=UniKl)·(trae=öff.)	0	—	4, 2740	38,70	<0,0001
(ktyp=PlanKH)·(trae=öff.)	-0,01977	0,03609			
(ktyp=PlanKH)·(trae=freig.)	-0,3205	0,02642			
(ktyp=VersV)·(trae=öff.)	-0,01696	0,03879			
(ktyp=VersV)·(trae=freig.)	-0,3322	0,03242			
fallzL	-0,8522	0,006931	1, 3041	15.116	<0,0001
arztzL	0,1781	0,009977	1, 4102	318,7	<0,0001
naerzL	0,6686	0,01084	1, 3577	3.803	<0,0001

^x Standardfehler der Schätzung

^y Zähler-, Nenner-Freiheitsgrade der Wald-F-Statistik

^z Ablehnungswahrscheinlichkeit der Wald-F-Statistik für diesen Effekt

Vorpommern – erheblich größere Werte für die Varianzen schätzen lässt als für jedes der anderen Jahre²⁵⁰. Die Autokorrelation wird mit 0,822 geschätzt.

Mit dem Einbeziehen der die Krankenhäuser differenzierenden Merkmale Krankenhaustyp und Krankenhausträger sowie der Beschäftigtenanzahlen nimmt die Bedeutung des Ost-West-Unterschiedes deutlich ab, der Einfluss des Ostdeutschland-Indikators ist nicht signifikant²⁵¹. Ein Wechselwirkungsterm (ktyp=UniKl)·(trae=öff.) ist nicht schätzbar, jedes der Universitätsklinika ist in öffentlicher Trägerschaft. Obwohl den Beschäftigtenzahlen und den Fallzahlen jeweils ein sehr bedeutsamer Einfluss auf die Modellierung attestiert wird, sind diese Merkmale als wechselseitig hoch korreliert nur begrenzt einzeln betracht- oder beeinflussbar.

²⁵⁰ 1998: 0,224; 2002: 0,159; alle anderen < 0,021.

²⁵¹ auf dem 5 %-Niveau.

Tab. 6.15.: Krankenhäuser 1992–2005, Kostenmodellierung #2
Prädiktionen für Plankrankenhäuser in öffentlicher Trägerschaft
(5.916 Beobachtungseinheiten)

Jahr	Gebiet	Fälle	Ärzte	nä B ^x	Prädiktion ^y
1992	RP	5.000	30	230	2.797
		5.500	30	230	2.579
		5.000	33	230	2.845
		5.000	30	253	2.981
	Ostdt	5.000	30	230	2.621
		5.500	30	230	2.417
		5.000	33	230	2.666
		5.000	30	253	2.794
2005	RP	5.000	30	230	3.005
		5.500	30	230	2.771
		5.000	33	230	3.057
		5.000	30	253	3.203
	Ostdt	5.000	30	230	3.236
		5.500	30	230	2.983
		5.000	33	230	3.291
		5.000	30	253	3.449

^x nichtärztliche Beschäftigte

^y bereinigte Kosten je Fall [€, deflationiert]. Die obere 95 %-Konfidenzschranke liegt jeweils etwa beim 1,11-fachen, die untere beim $\frac{1}{1,11}$ -fachen des Prädiktionwertes.

Dennoch wird ein Überblick über die Veränderungen der bereinigten Kosten bei der Veränderung ausgewählter Regressoren versucht:

Ausgangspunkt ist ein ‚mittleres‘ Krankenhaus, Fallzahl (5.000), Anzahl der Ärzte (30) und Anzahl des nichtärztlichen Personals (230) entsprechen jeweils dem geometrischen Mittel²⁵² der Angaben aller Krankenhäuser in Untersuchungsgebiet und -zeitraum. Jedem dieser Merkmale wird unter Ceteris-paribus-Bedingung ein Zuwachs von 10 % zugestanden, die Auswirkungen auf die Prädiktion der bereinigten Kosten je Fall für ein Plankrankenhaus in öffentlicher Trägerschaft zeigt Tab. 6.15 getrennt für Rheinland-Pfalz und Ostdeutschland. Da Plankrankenhäuser in öffentlicher Trägerschaft einen großen Anteil der Krankenhäuser ausmachen (1992: 59,6 %, 2005: 27,9 %) sind die Konfidenzintervallbreiten mit einer

²⁵² 4.898 Fälle, 30,9 Ärzte bzw. 234,4 Beschäftigte des nichtärztlichen Personals.

oberen 95 %-Konfidenzschranke bei etwa dem 1,11- und einer unteren beim $\frac{1}{1,11}$ -fachen der Prädiktion relativ gering.²⁵³

Die Prädiktionen zeigen, dass²⁵⁴ die bereinigten Kosten je Fall in Ostdeutschland stärker zunehmen als in Rheinland-Pfalz. Der nicht-signifikantem Einfluss des Ostdeutschland-Indikators (*ostdt*, Tab. 6.14) zeigt dieses nicht.

Bereinigte Kosten der Krankenhäuser – Modell 3

Das dritte Modell widmet sich der Frage, inwieweit Wirtschafts- bzw. Sozialparameter der Kreise mit den bereinigten Kosten der lokalen Krankenhäuser assoziiert sind. Die logarithmierten bereinigten Kosten je Fall sind der Regressand in einem Paneldaten-Modell mit den Regressoren aus Modell 2 erweitert um den Kreistyp der siedlungsstrukturellen Gebietstypen des BBR und dessen Wechselwirkung mit Ostdeutschland-Indikator und *jahr*, die (logarithmierte) Einwohnerdichte (*ewdL*), das (logarithmierte) Brutto-Inlandsprodukt je Einwohner (*bipewL*), die (logarithmierte) Brutto-Erwerbstätigenquote (*betqL*) und den Logit der Arbeitslosenquote (*alqlgt*):

$$\begin{aligned} \eta = & \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} + \beta_3 \cdot \text{fallzL} \\ & + \beta_{41} \cdot \text{ktyp}_1 + \beta_{42} \cdot \text{ktyp}_2 + \beta_{43} \cdot \text{ktyp}_3 + \beta_{51} \cdot \text{trae}_1 + \beta_{52} \cdot \text{trae}_2 \\ & + \beta_{41,51} \cdot \text{ktyp}_1 \cdot \text{trae}_1 + \dots + \beta_{43,52} \cdot \text{ktyp}_3 \cdot \text{trae}_2 \\ & + \beta_6 \cdot \text{arztzL} + \beta_7 \cdot \text{naerzL} \\ & + \beta_{81} \cdot \text{krstyp}_1 + \dots + \beta_{88} \cdot \text{krstyp}_8 \\ & + \beta_{1,81} \cdot \text{ostdt} \cdot \text{krstyp}_1 + \dots + \beta_{1,88} \cdot \text{ostdt} \cdot \text{krstyp}_8 \\ & + \beta_{2,81} \cdot \text{jahr} \cdot \text{krstyp}_1 + \dots + \beta_{2,88} \cdot \text{jahr} \cdot \text{krstyp}_8 \\ & + \beta_9 \cdot \text{ewdL} + \beta_A \cdot \text{betqL} + \beta_B \cdot \text{bipewL} + \beta_C \cdot \text{alqlgt}. \end{aligned}$$

Wegen der begrenzteren Verfügbarkeit der Wirtschafts- und Sozialdaten²⁵⁵ beruht dieses Modell auf den Daten der Jahre 1996–2005. Die veränderte Datenba-

²⁵³ Für die jeweils nur gut 2 % Universitätsklinika liegen die Schranken hingegen beim 1,6- bzw. $\frac{1}{1,6}$ -fachen der Prädiktion.

²⁵⁴ zumindest für Plankrankenhäuser in öffentlicher Trägerschaft.

²⁵⁵ Die Arbeitslosenquoten, z. B., sind für die ostdeutschen Kreise erst ab 1996 verfügbar.

sis verhindert den Vergleich der Modelle 2 und 3 mittels des Bayes'schen Informationskriteriums.

Die heterogen autoregressive ARH(1)-Kovarianzstruktur ergibt für 1998 und 2001 – das erste bzw. letzte Jahr ohne Angaben aus Mecklenburg-Vorpommern – wiederum erheblich größere Werte für die Varianzen²⁵⁶, die Autokorrelation selbst wird mit 0,755 geschätzt.

Tab. A.3 im Anhang zeigt die Parameterschätzungen sowie die Ergebnisse der zugehörigen Wald-F-Statistik.

Weder für einen der Wirtschafts- oder Sozialparameter noch für den Kreistyp selbst lässt sich ein wesentlicher Einfluss auf die Prädiktion feststellen. Die bedeutendsten Einflüsse auf die Prädiktion der bereinigten Kosten je Fall sind der Zeitvariablen *jahr*, dem Krankenhaustyp, der Fallanzahl und den Anzahlen an Ärzten sowie an nichtärztlichem Personal zuzusprechen. Die Trägerschaft des Krankenhauses hat über den Wechselwirkungsterm mit dem Krankenhaustyp einen statistisch bedeutsamen Anteil an der Prädiktion, der Ost-West-Unterschied trägt für die Jahre 1996–2005 nicht merklich zur Güte der Prädiktion bei.

Andere Kombinationen von den in dieser Arbeit diskutierten Wirtschafts- und Sozialparametern als Regressoren ergeben kein qualitativ verändertes Bild, alle erscheinen als für die bereinigten Kosten je Fall wenig relevante Einflussfaktoren.

Es sind somit vor allem die Eigenschaften des Krankenhauses wie Größe, Krankenhaustyp und Trägerschaft, die auf die relativen bereinigten Kosten den größten Einfluss haben. Wie bereits an anderer Stelle ausgeführt sind diese Merkmale bzw. Merkmalsgruppen miteinander korreliert, die Eigenschaften bedingen einander zu großen Teilen.

²⁵⁶ 1998: 0,180; 2002: 0,126; alle anderen < 0,016.

Personal-Sachkosten-Verhältnis der Krankenhäuser

Dieselben Modelle²⁵⁷ mit denselben Regressoren und linearen Prädiktoren

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} + \dots$$

werden genutzt, den zweiten konsistent über den Berichtszeitraum verfügbaren Kostenparameter zu modellieren. Regressand ist dabei das \log_e -logarithmierte Personal-Sachkosten-Verhältnis, Linkfunktion ist die Identität. Wiederum unterstützen Modelle 1 und 2 die Betrachtungen zu den Veränderungen der Krankenhäuser selbst (Abschn. 6.3.1: Die Kosten der Krankenhäuser in Transition), während Modell 3 Beziehungen zwischen (lokaler) Wirtschaftsleistung und Krankenhäusern untersucht (Abschn. 6.3.1: Der Einfluss der Wirtschaft auf die Krankenhäuser).

Für alle diese Modelle fällt auf, dass das bei den bereinigten Kosten je Fall beobachtete Phänomen der Inhomogenität der Varianzschätzungen²⁵⁸ zu den Berichtsjahren bei der Modellierung des Personal-Sachkosten-Verhältnisses nicht auftritt.²⁵⁹

Die Parameterschätzungen für Modell 1 ergeben auch für die Wechselwirkung zwischen Berichtsjahr und der Zugehörigkeit des Krankenhauses zu Rheinland-Pfalz oder Ostdeutschland ein signifikant²⁶⁰ von 0 verschiedenes Ergebnis, somit wird allen Regressoren (Ostdeutschland-Indikator (*ostdt*), Berichtsjahr (*jahr*), deren Wechselwirkungen sowie der logarithmierten Anzahl der abgehenden Fälle (*fallzL*)) ein bedeutender Einfluss auf das Personal-Sachkosten-Verhältnis der Krankenhäuser zugeschrieben. Die Varianz innerhalb der Berichtsjahre schwankt zwischen 0,129 und 0,258, die Autokorrelation wird mit 0,935 geschätzt. Das Bayes'sche Informationskriterium hat einen Wert von -4094,0.

Quantitativ (Tab. 6.16) zeugen die Prädiktionen für die Jahre 1992–2005 von einer Abnahme des Personal-Sachkosten-Verhältnisses, d. h. einem steigenden Anteil der Sachkosten an den Gesamtkosten. Die Unterschiede im Personal-Sachkosten-

²⁵⁷ Modelle 1 bis 3 aus vorstehenden Unterabschnitten: Bereinigte Kosten der Krankenhäuser.

²⁵⁸ bei Annahme einer heterogen autoregressiven ARH(1)-Kovarianzstruktur.

²⁵⁹ Die Modellierung mit einer AR(1)-Struktur wäre somit hinreichend.

²⁶⁰ auf dem 5%-Niveau.

Tab. 6.16.: Krankenhäuser 1992–2005, Personal-Sachkosten-Verhältnis #1
Prädiktionen (6.089 Beobachtungseinheiten)

Jahr	Gebiet	Fälle	Prädiktion ^x
1992	RP	5.000	68,7
		20.000	67,7
	Ostdt	5.000	66,6
		20.000	65,6
2005	RP	5.000	67,3
		20.000	66,2
	Ostdt	5.000	62,5
		20.000	61,4

^x Personal-Sachkosten-Verhältnis [%]

Verhältnis zwischen großen und kleinen Krankenhäusern liegen bei 1,0 Prozentpunkten für ein 5.000er Krankenhaus im Vergleich zu einem viermal so großen; die Ost-West-Unterschiede nehmen zwischen 1992 und 2005 von 2,1 auf 4,8 Prozentpunkte vergleichsweise deutlich zu.

In Modell 2 mit den das Krankenhaus beschreibenden Regressoren wird die berichts-jahrinterne Varianz mit Werten zwischen 0,091 und 0,173 relativ homogen geschätzt, für die Autokorrelation ergibt sich 0,907. Das Bayes'sche Informationskriterium hat mit -4375,9 einen gegenüber Modell 1 deutlich verminderten Wert, Modell 2 kann die Daten besser beschreiben. Die Entscheidungen auf Basis der Wald-F-Statistik sprechen den Parametern *ostdt*, *ktyp*, *trae*, *fallzL*, *naerzL* und der Wechselwirkung zwischen Krankenhaustyp und -träger mit Ablehnungswahrscheinlichkeiten von jeweils $< 0,0001$ eine große Bedeutung für Modell und Prädiktionen zu. Während das Jahr einen bedeutsamen Einfluss ($p = 0,017$) hat, könnte im Modell auf den Wechselwirkungsterm zwischen Jahr und Ostdeutschland-Indikator verzichtet werden ($p = 0,582$).

Der Einfluss der (logarithmierten) Anzahl der Ärzte ist als bedeutsam ($p = 0,002$) anzusehen, allerdings überrascht, dass eine Erhöhung der Anzahl der Ärzte zu einer Verringerung des Personal-Sachkosten-Verhältnisses führen soll. Es ist zu vermuten, dass dieser Umstand Ergebnis der hohen Korrelation der Ärzteanzahl mit sowohl der Fallanzahl als auch der Anzahl der nichtärztlichen Beschäftigten ist, die beide ebenfalls Regressoren des Modells 2 sind.

Tab. 6.17.: Krankenhäuser 1992–2005, Personal-Sachkosten-Verhältnis #1
Prädiktionen für Plankrankenhäuser in öffentlicher Trägerschaft
(5.916 Beobachtungseinheiten)

Jahr	Gebiet	Fälle	Ärzte	nä B ^x	Prädiktion ^y
1992	RP	5.000	30	230	70,4
		5.500	30	230	70,1
		5.000	33	230	70,3
		5.000	30	253	70,7
	Ostdt	5.000	30	230	66,6
		5.500	30	230	66,4
		5.000	33	230	66,6
		5.000	30	253	67,0
2005	RP	5.000	30	230	69,5
		5.500	30	230	69,2
		5.000	33	230	69,4
		5.000	30	253	69,8
	Ostdt	5.000	30	230	65,1
		5.500	30	230	64,8
		5.000	33	230	65,0
		5.000	30	253	65,5

^x nichtärztliche Beschäftigte

^y Personal-Sachkosten-Verhältnis [%]

Die Prädiktionen (Tab. 6.17 für die Plankrankenhäuser in öffentlicher Trägerschaft) zeigen, dass für Krankenhäuser mit gleicher Charakteristik im Ost-West-Vergleich eine Differenz von etwa 4 Prozentpunkten im Personal-Sachkosten-Verhältnis mit höheren Werten für Rheinland-Pfalz besteht. Die Differenz nimmt zwischen 1992 und 2005 jedoch nur leicht zu.

Modell 3, das um die Wirtschafts- bzw. Sozialparameter der Kreise erweiterte und auf den Daten der Jahre 1996–2005 beruhende Paneldaten-Modell, ergibt relativ homogene berichts-jahrinterne Varianzschätzungen zwischen 0,107 und 0,174 und eine geschätzte Autokorrelation von 0,911.

Die Parameterschätzungen mit zugehörigen Wald-F-Statistiken zeigen, dass die Einflüsse jeglicher der Wirtschafts- und Sozialparameter²⁶¹ genauso wie die des Kreistyps auf das Modell eher unbedeutend sind. Als bedeutsam für die Prädiktion des Personal-Sachkosten-Verhältnisses stellen sich (wieder nur) die Zeitvariable *jahr*, Krankenhausstyp und -träger nebst Wechselwirkung, die Fallanzahl und die Anzahlen an Ärzten sowie an nichtärztlichem Personal heraus; die Eigenschaften des jeweiligen Krankenhauses sind die wichtigsten Prädiktoren. Allerdings trägt der Ost-West-Unterschied hier merklich ($p = 0,017$) zur Güte der Prädiktion bei: die Personal-Sachkosten-Verhältnisse bleiben auch unter Hinzuziehen der anderen erklärenden Regressoren in ihrem Niveau zwischen Ost und West unterschiedlich.

6.2.9. Krankenhaus-Kosten in den einzelnen Kreisen

Auch auf Kreisniveau stehen die Kosten den Krankenhausleistungen gegenüber, beide lassen sich sinnvollerweise als jeweils kumuliert in ihren Beziehungen zu einander betrachten (Abschn. 5.1.2: Gesamtkrankenhausesaktivität).²⁶² Wiederum stehen die bereinigten Kosten²⁶³ und das Personal-Sachkosten-Verhältnis als Regressanden von Paneldaten-Regressionen im Mittelpunkt des Interesses²⁶⁴; für die bessere Vergleichbarkeit werden die relativen bereinigten Kosten auf die Fallzahl (Abb. 6.8 d) bezogen. Außer dem krankenhaushausfreien Landkreis Ludwigshafen²⁶⁵ ist jeder Kreis Panelsubjekt, das Panel der Kreise mit Krankenhäusern verändert sich im Berichtszeitraum kaum (Tab. 6.2 und zugehörige Ausführungen).

Für die zwei interessierenden Regressanden – die relativen bereinigten Kosten sowie das Personal-Sachkosten-Verhältnis – werden die in Abschn. 6.2.8 beschriebenen Transformationen angewendet, es gelten die dortigen Aussagen zu Verteilungen und gewählter Linkfunktion.

²⁶¹ Wie für die Modelle zur Prädiktion der bereinigten Kosten je Fall gilt auch für das Personal-Sachkosten-Verhältnis, dass andere Kombinationen der in dieser Arbeit diskutierten Wirtschafts- und Sozialparameter als Regressoren kein qualitativ verändertes Bild ergeben.

²⁶² Eine Betrachtung nach Krankenhausstypen und -trägern ist dann gegenstandslos.

²⁶³ deflationiert (Abschn. 4.4.3: Deflationierung der Kostendaten).

²⁶⁴ Bundeswehrkrankenhäuser gehen in diese Betrachtungen nicht ein.

²⁶⁵ seit 01. 01. 2004: Rhein-Pfalz-Kreis.

Kosten in den Kreisen – Modell 4

Die logarithmierten bereinigten Kosten der Jahre 1992–2005 mit den Kreisen als Panelsubjekte werden²⁶⁶ mit den Regressoren Ostdeutschland-Indikator (*ostdt*), Berichtsjahr (*jahr*), deren Wechselwirkungen und der logarithmierten Anzahl der abgehenden Fälle (*fallzL*) und zwei weiteren Regressoren modelliert:

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} \\ + \beta_3 \cdot \text{jahr} \cdot \text{jahr} + \beta_4 \cdot \text{ewdL}.$$

jahr·*jahr* ermöglicht die Betrachtung eines überlinearen Anteils der Zeit (Variable *jahr*), die (logarithmierte) Einwohnerzahl *ewdL* ist eine Näherungsvariable für Urbanität.

Ein Modellierungsversuch mit heterogen autoregressiver ARH(1)-Kovarianzstruktur zeigt sehr homogene berichtsjahterinterne Varianzen, die Modellierung als AR(1)-Kovarianzstruktur ergibt wegen der verringerten Parameteranzahl eine gemäß Bayes'schem Informationskriterium bessere Anpassungsgüte. Die Autokorrelation wird mit 0,967 geschätzt. Der Wald-F-Test zur Frage, ob die Parameter nur unbedeutend von null verschieden sind, führt jeweils zu Ablehnung ($p = 0,022$ oder kleiner).

Ausgehend von den mittleren Anzahlen von Fällen und Einwohnerzahlen²⁶⁷ zeigt Tab. 6.18 Prädiktionen auch je für eine angenommene 10 %ige Zunahme von Fällen bzw. für einen Kreis mit 10 % mehr Einwohnern. Beide Merkmale sind positiv korreliert. Ceteris paribus ergeben mehr Fälle eine Senkung der fallbezogenen bereinigten Kosten während mehr Einwohner zu erhöhten Prädiktionswerten führen.

Allgemein nehmen zwischen 1992 und 2005 die Kosten je Fall zu, dabei sind diese in Rheinland-Pfalz höher als in Ostdeutschland. Die Differenz verringert sich jedoch von 360 € auf 180 €.

²⁶⁶ ähnlich Modell 1.

²⁶⁷ Die geometrischen Mittel über alle Kreise und Berichtsjahre betragen 20.893 Fälle und 114.800 Einwohner.

Tab. 6.18.: Kreise 1992–2005, Kostenmodellierung #4
Prädiktionen (1.813 Beobachtungseinheiten)

Jahr	Gebiet	Fälle	Einwohner	Prädiktion ^x
1992	RP	21.000	120.000	2662
		23.100	120.000	2624
		21.000	132.000	2705
	Ostdt	21.000	120.000	2300
		23.100	120.000	2267
		21.000	132.000	2337
2005	RP	21.000	120.000	2851
		23.100	120.000	2810
		21.000	132.000	2897
	Ostdt	21.000	120.000	2668
		23.100	120.000	2629
		21.000	132.000	2711

^x bereinigte Kosten je Fall [€, deflationiert]

Für die Prädiktion des Personal-Sachkosten-Verhältnisses mit Hilfe desselben Modells zeigt sich, dass weder der Wechselwirkung zwischen Ostdeutschland-Indikator und Berichtsjahr jahr ($p = 0,456$) noch dem quadratischen $\text{jahr} \cdot \text{jahr}$ -Term eine statistische Bedeutung beigemessen wird ($p = 0,117$). Die Annahme einer autoregressiven AR(1)-Kovarianzstruktur führt gegenüber der heterogen autoregressiven ARH(1)-Kovarianzstruktur zu einer marginalen Verschlechterung der Anpassungsgüte gemäß Bayes'schem Informationskriterium.

Ausgewählte Prädiktionen des Modells mit AR(1)-Kovarianzstruktur zeigt Tab. 6.19. Beherrschend ist auch hier der Ost-West-Unterschied mit höheren Personal-Sachkosten-Verhältnis-Werten in Rheinland-Pfalz. Die Differenz sinkt dabei im Untersuchungszeitraum von 4,3 auf 3,7 Prozentpunkte.

Modell 4 unterstützt die Betrachtungen in Abschn. 6.3.1: Die Kosten der Krankenhäuser in Transition.

Tab. 6.19.: Kreise 1992–2005, Personal-Sachkosten-Verhältnis #4
Prädiktionen (1.813 Beobachtungseinheiten)

Jahr	Gebiet	Fälle	Einwohner	Prädiktion ^x
1992	RP	21.000	120.000	70,7
		23.100	120.000	70,4
		21.000	132.000	70,8
	Ostdt	21.000	120.000	66,4
		23.100	120.000	66,1
		21.000	132.000	66,5
2005	RP	21.000	120.000	67,9
		23.100	120.000	67,6
		21.000	132.000	68,0
	Ostdt	21.000	120.000	64,2
		23.100	120.000	63,9
		21.000	132.000	64,3

^x Personal-Sachkosten-Verhältnis [%]

Kosten in den Kreisen – Modell 5

Um den Einfluss von Wirtschaftskraft und sozialer Lage der Kreise auf die kumulativen Krankenhauskosten untersuchen zu können, wird Modell 4 um die Regressoren (logarithmiertes) Brutto-Inlandsprodukt je Einwohner (*bipewL*) und Logit der Arbeitslosenquote (*alqlgt*) erweitert

$$\eta = \beta_0 + \beta_1 \cdot \text{ostdt} + \beta_2 \cdot \text{jahr} + \beta_{1,2} \cdot \text{ostdt} \cdot \text{jahr} \\ + \beta_3 \cdot \text{jahr} \cdot \text{jahr} + \beta_4 \cdot \text{ewdL} + \beta_5 \cdot \text{bipewL} + \beta_6 \cdot \text{alqlgt}.$$

Die Modellierungen²⁶⁸ beruhen wiederum auf den Daten der Jahre 1996–2005, sie unterstützen die Bewertungen im Abschn. 6.3.1: Der Einfluss der Wirtschaft auf die Krankenhäuser.

Es ergibt sich, dass jeweils das Modell mit autoregressiver AR(1)-Kovarianzstruktur eine deutlich höhere Anpassungsgüte besitzt als sein Pendant mit heterogen autoregressiver ARH(1)-Kovarianzstruktur. Die Autokorrelation wird mit 0,966 für

²⁶⁸ ähnlich Modell 3.

die bereinigten Kosten je Fall und mit 0,889 für das Personal-Sachkosten-Verhältnis als Regressanden jeweils sehr hoch geschätzt.

Für das Modell der bereinigten Kosten je Fall erweisen sich gemäß Wald-F-Statistik das Jahr mit quadratischem Einfluss (der $\text{jahr} \cdot \text{jahr}$ -Term), die Fallzahl, die Einwohnerzahl und das Brutto-Inlandsprodukt je Einwohner als für die Prädiktion wichtige Regressoren (jeweils $p < 0,0001$), während dem Ostdeutschland-Indikator ($p = 0,638$), dem Berichtsjahr selbst ($p = 0,908$), deren Wechselwirkungen ($p = 0,148$) und der Arbeitslosenquote ($p = 0,448$) wenig Bedeutung beigemessen wird. Ein linearer Einfluss des Jahres reicht zur guten Modellierung der Daten nicht aus.

Ein wenig anders stellt sich die Modellsituation für das Personal-Sachkosten-Verhältnis dar: Ostdeutschland-Indikator, Berichtsjahr und Fallzahl erscheinen unverzichtbar (jeweils $p < 0,0001$), die Einwohnerzahl ($p = 0,017$), die Ostdeutschland-Indikator-Jahr-Wechselwirkung ($p = 0,022$), das Brutto-Inlandsprodukt je Einwohner ($p = 0,041$) sind die weiteren wichtigen Faktoren. Dem $\text{jahr} \cdot \text{jahr}$ -Term ($p = 0,062$) ist eher geringe und der Arbeitslosenquote ($p = 0,537$) auch hier nahezu keine Bedeutung beizumessen. Bedeutende Ost-West-Unterschiede zeigen sich auf Parameterebene.

Für beide Modelle gilt, dass ein Austausch des Regressors Brutto-Inlandsprodukt je Einwohner durch einen anderen der mit ihm hochkorrelierten kreisspezifischen Wirtschaftsparameter²⁶⁹ qualitativ an den Verhältnissen zwischen den Regressoren nichts ändert.

Vergleicht man diese Modelle mit den Kreisen als Panelsubjekte mit den ihnen ähnlichen krankenhausbezogenen Modellen 3 fällt auf, dass sie sich in Bedeutung von Wirtschaftsparametern als Regressoren unterscheiden, dem Sozialparameter Arbeitslosenquote hingegen kommt in beiden kaum eine Bedeutung für die modellbasierten Prädiktionen zu.

²⁶⁹ z. B. die Brutto-Erwerbstätigenquote.

6.3. Die Modellergebnisse aus ökonomischer Sicht

Deskriptionen (Abschn. 6.1) und Modellierungen (Abschn. 6.2) der Daten der Krankenhausstatistik²⁷⁰ enthüllen viele verschiedene Aspekte der Krankenhauswirklichkeit mittels Kennzahlen, Tabellen, Graphen und Parameterschätzungen. Es bleiben die Fragen nach den Gemeinsamkeiten (und Widersprüchen) dieser Ergebnisse, der Interpretation, des Erkennens von Zusammenhängen, ihrer Stärke und Richtung sowie gemeinsamer Muster (Miller, 2005, *Seven Basic Principles*) – es bleibt das Übersetzen in sozialwissenschaftlich-ökonomische Ergebnisse.

6.3.1. Krankenhäuser und Wirtschaft

Krankenhäuser und Wirtschaft interagieren auf vielfältige Weise. Krankenhäuser – der mit den schwersten Gesundheitsproblemen arbeitende Teil des Gesundheitswesens – ‚produzieren‘ Genesende (und Genesene), deren effiziente Gesundheit die Sozialsysteme entlastet und die u. a. perspektivisch wieder dem Arbeitsmarkt zur Verfügung stehen. Krankenhäuser verbrauchen dabei Ressourcen materieller und personeller Art, die sich in den Kosten widerspiegeln.

Veränderungen der Krankenhäuser

Die Anzahl der Krankenhäuser sinkt in Ost- wie in Westdeutschland, sei es durch Zusammenlegung, Übernahme oder Schließung. Die Analyse der Krankenhausedichte aller 16 Bundesländer (Abb. 2.2) jedoch zeigt, dass die Krankenhausedichte in den Altbundesländern deutlich höher ist als in den Ländern Ostdeutschlands. Das ändert sich im Berichtszeitraum nicht. Die Bettendichten (Abb. 2.3) werden auf vergleichbare Werte reduziert – für Ostdeutschland von im Mittel größeren Werten relativ abrupt, in Westdeutschland gleichmäßiger moderat über den Gesamtzeitraum 1992–2005. Sichtbar wird: die Krankenhäuser in Ostdeutschland sind im Mittel größer als ihre Pendanten in Westdeutschland. Ob es sich dabei um größere Einheiten an jeweils einem Ort oder um administrative Einheiten mit mehreren Standorten handelt, ist in den Daten der Krankenhausstatistik nicht

²⁷⁰ ergänzt um die ausgewählten Wirtschaftsstatistiken.

erfasst. Dies zu wissen wäre aus wirtschaftlicher Sicht jedoch interessant, denn mit den Entfernungen steigen die Kosten für den Transport der Kranken und ihrer den Heilungsprozess unterstützenden Besucher – als Aufenthaltsnebenkosten erscheinen diese nicht in der Krankenhausstatistik.

Politische Vorgaben (Abschn. 3.2.1) zwingen die Krankenhäuser im Berichtszeitraum zu einer Abkehr von budgetbasiertem Mittelzufluss hin zu einer fallpauschalenbasierten Vergütung ihrer Leistungen²⁷¹. Vorgehaltene Kapazitäten werden als Ressourcenbindung wahrgenommen und in ihrem Umfang hinterfragt. Sehr sichtbar wird dies an der Anzahl der aufgestellten Betten, einem leicht ermittel- und interpretierbarem Indikator: Betten werden abgebaut (Abb. 2.3 und 6.7 a).

Modellierungen der Bettendichte zeigen, dass Unterschiede zwischen Ost und West – repräsentiert durch Rheinland-Pfalz – im Vergleich mit den Verringerung der Betten über den gesamten Berichtszeitraum eher unbedeutend sind, beide Teile reagieren gleichartig auf die politischen Forderungen. Für Ostdeutschland geht die Verringerung der Bettendichte einher mit einer leichten Zunahme der Belegungsdichte²⁷², während in Rheinland-Pfalz die Belegungsdichte deutlich abnimmt, so dass Rheinland-Pfalz 2005 die geringste Bettenauslastung aller Bundesländer hat (Statistisches Bundesamt, 2007, S. 246).

Äußerst üblich – national und international – ist die Bewertung der Effizienzsteigerung der Krankenhäuser per mittlerer Verweildauer – allen fachlich-sachlichen Einwendungen zum Trotz. Die Verweildauern werden gesenkt, die Aufenthalte im Krankenhaus werden kürzer. Dies wird zum einen dadurch erreicht, dass der Anteil der Langlieger²⁷³ – auch durch Überführung in Pflegeeinrichtungen – sinkt, zum anderen auch durch kürzere Behandlungszeiten im Krankenhaus. Mit vorliegenden Daten lassen sich die wirtschaftlich drängenderen Fragen nach den Veränderungen der Gesamt-Heilungszeiten und -kosten nicht beantworten.

Der Anteil der Kreise ohne Krankenhaus ist sehr gering.²⁷⁴ Der Anteil der Kreise mit genau einem Krankenhaus nimmt von 12,2 % (1992) auf 26,5 % (2005,

²⁷¹ auch wenn das DRG-basierte Fallpauschalensystem erst am Ende des Berichtszeitraums verbindlich eingeführt wird.

²⁷² der Belegungstage eines Bettes im Verlauf eines Jahres.

²⁷³ Verweildauern von mehr als 6 Wochen, d. h. 43 oder mehr Tagen.

²⁷⁴ genauer: Kreise ohne Sitz eines Krankenhauses im Kreis. Für vier Kreise fehlt 2005 der Nachweis eines Krankenhauses, das wären 2,7 %.

Tab. 6.2) deutlich zu. Inwieweit dieser Zustand neben Wirtschaftlichkeitsüberlegungen auch vom politischen Willen der lokalen Entscheidungsträger beeinflusst wird, im Kreis – z. B. über öffentliche Trägerschaft – mindestens ein ‚eigenes‘ Krankenhaus ansässig zu haben, ist nicht bekannt.

In den Anteilen der Krankenhäuser in öffentlicher, freigemeinnütziger bzw. privater Trägerschaft (Tab. 6.4) ähneln sich 1992 Berlin und Rheinland-Pfalz sehr, während die ostdeutschen Flächenländer sich davon sehr unterscheiden. 2005 hat diese Ähnlichkeit abgenommen, so dass Rheinland-Pfalz, Berlin und die ostdeutschen Flächenländer sehr heterogen hinsichtlich der Krankenhausträger wirken – jedes der drei Gebiete weicht deutlich von gemeinsamen Mittel ab. Über den Gesamtzeitraum 1992–2005 werden im Mittel 42,0 % der Krankenhäuser²⁷⁵ öffentlich, 39,2 % in freigemeinnütziger und 18,8 % in privater Trägerschaft betrieben. Soweit die Anzahlen, kleine Krankenhäuser stehen dabei mit großen auf einer Stufe. Mehr über die unterschiedlichen Krankenhausgrößen²⁷⁶ und die anteiligen Leistungen lässt sich aus den bereinigten Kosten²⁷⁷ ablesen: 62,6 % entfallen auf Häuser in öffentlicher, 26,8 % auf die in freigemeinnütziger und 10,6 % auf die in privater Trägerschaft. Eine gleichartige Betrachtung ergibt, dass die 2,2 % Universitätsklinika unter den Krankenhäusern für 15,0 % der bereinigten Kosten stehen, während den 1,5 % reinen Tages- und Nachtkliniken 0,05 % aller Kosten zuzuordnen sind. Bezüglich der wirtschaftlichen Leistungen und (somit) der Bedeutung der Krankenhäuser für die Patientenversorgung sind kostenbezogene Zahlen eine bessere Beschreibung.

Indikatoren und Gewichtungen, speziell zur Beschreibung der Veränderungen der Krankenhäuser

Die bereinigten Kosten der Einzel-Krankenhäuser wie die des kumulierten Krankenhausbetriebs (Abschn. 5.1.2: Gesamt-Krankenhausaktivität) in den Kreise stei-

²⁷⁵ hier ohne die 3 Bundeswehrkrankenhäuser, die keine Kostendaten berichten.

²⁷⁶ Größenverhältnisse der Krankenhäuser Deutschlands in unterschiedlicher Trägerschaft werden für 2005 auch mittels durchschnittlicher Bettenanzahlen berichtet: 115 Betten je Krankenhaus in privater, 226 in freigemeinnütziger und 364 in öffentlicher Trägerschaft (Statistisches Bundesamt, 2007; vgl. Abschn. 3.2.2).

²⁷⁷ pflegesatzfähige Kosten.

gen stärker als der Verbraucherpreisindex (VPI), die deflationierten Kosten steigen somit auch.

Große Einheiten, seien es Einzel-Krankenhäuser oder die Kumulation aller Krankenhäuser eines Gebietes, haben erwartungsgemäß höhere, kleinere hingegen geringere Kosten; das bestätigt sich mit den Daten der Krankenhausstatistik. In einer Analyse dieser Einheiten widerspiegeln die Kosten zuerst und vor allem die Größe des jeweiligen Hauses. Mittels Ableitung von Beziehungszahlen oder Raten (Rönz und Strohe, 1994, S. 49 f.) versucht man, diesen Größeneinfluss auf die Kosten zu eliminieren. Übliche Raten dabei sind (a) Kosten je Einwohner, (b) Kosten je aufgestelltem Bett und (c) Kosten je Fall (Abb. 6.8).

Alle drei Raten werden häufig als Indikatoren zur Beschreibung von Krankenhäusern oder kumulativen Krankenhauseinheiten herangezogen, bei longitudinalen Vergleichen ist vorab die Frage der Deflationierung zu klären. Die Nennermerkmale werden dabei als Näherungsvariablen für den Aufwand oder die (zu erbringenden) Leistungen der Krankenhäuser angesehen. Die Betrachtungen im Untersuchungszeitraum zeigen jedoch, dass die Bettenanzahl (Abb. 2.3) als Referenz sehr wenig geeignet ist, beschreibt sie doch vor allem ein Vorhalten von Ressourcen, dessen Umfang sich in den Jahren 1992–2005 selbst bei steigenden Einwohner- und Fallzahlen politisch gewollt massiv verringern lässt.

Aus wirtschaftswissenschaftlicher Sicht ergibt sich, dass die Kosten je Einwohner als Rate zwei Konzepte gegeneinander stellt: das Inlandskonzept mit den kreisweisen (kumulierten) Krankenhauskosten und mit den Einwohnern das Inländerkonzept. Kosten werden nicht unmittelbar Leistungen oder Leistungsnutzern gegenübergestellt.

Die beste Beziehung zwischen Aufwand und Leistungen stellen die Kosten je behandelten Fall her, auch wenn dabei jeder Fall als gleichwertig angenommen wird – unabhängig von seiner Schwere oder dem nötigen Aufwand seiner Behandlung.²⁷⁸

²⁷⁸ Erst nach Ende des Untersuchungszeitraumes steht in Deutschland mit den DRG ein der jeweiligen Fallschwere entsprechendes Merkmal zur Verfügung.

Das von Werblow und Schoffer (2010) herangezogene Personal-Sachkosten-Verhältnis²⁷⁹ kann ebenfalls als Indikator zur Beschreibung von Krankenhäusern oder kumulativen Krankeneinheiten herangezogen werden. Dieser Indikator hat neben dem von Werblow und Schoffer postulierten Vorteil der Konsistenz gegenüber dem Brutto-Netto-Brutto-Wechsel der Krankenhauskostenermittlung die Eigenschaft, von der (angenommenen mittleren) Inflation genauso wenig beeinflusst zu werden wie von der Größe des betrachteten Krankenhauses.²⁸⁰

Die Anwendung von Indikatoren in der Deskription and Modellierung von Krankenhäusern oder kumulativen Krankeneinheiten²⁸¹ macht deutlich, dass a priori jeder Einheit als statistischer Einheit (Abschn. 5.1.2) in den Analysen dieselbe Bedeutung zukommt: kleine Krankenhäuser haben dasselbe Gewicht wie große, das Universitätsklinikum steht gleichwertig neben der reinen Tages- oder Nachtambulanz.

Tab. 6.20.: Mögliche Wichtungen dokumentiert anhand der Anteile [%] der Bundesländer Berlin und Rheinland-Pfalz am Untersuchungsgebiet für verschiedene Indikatoren (2005)^z

Gebiet	Krankenhäuser	Einwohner	Betten	Fälle	Bel-Tage ^x	Ärzte	nä B ^y	Kreise
BE	16,6	16,3	15,3	16,1	15,9	21,4	15,7	0,7
RP	22,6	19,5	19,3	18,4	17,9	17,3	22,1	24,5

^z Berechnungen des Autors aus den amtlichen Angaben für 2005 (Statistisches Bundesamt, 2007)

^x Belegungstage

^y nichtärztliche Beschäftigte

Für diese Untersuchung bedeutet das: 2005 berichtet Berlin über 72, Rheinland-Pfalz über 98 Krankenhäuser, das sind 16,6 % bzw. 22,6 % der Krankenhäuser des Untersuchungsgebietes in jenem Jahr. Diese Krankenhausanzahlen stehen u. a. Einwohnerzahlen, Anzahlen aufgestellter Betten, Fallzahlen, Anzahlen der Belegungstage, Ärztezahlen, Anzahlen des nichtärztlichen Personals (bzw. Dienstes, Statistisches Bundesamt, 2007, S. 247) als weitere mögliche Bezugsgrößen gegenüber (Tab. 6.20). Für eine Analyse nach Krankenhäusern gibt Rheinland-Pfalz die

²⁷⁹ dort als Chance (engl.: *odds*) $\text{Personalkosten} : \text{Sachkosten} = \left(\frac{\text{Personalkosten}}{\text{Sachkosten}} \right)$ formuliert, hier äquivalent als Quote $\left(\frac{\text{Personalkosten}}{\text{Personalkosten} + \text{Sachkosten}} \right)$ in % angegeben.

²⁸⁰ Natürlich können sich dennoch Unterschiede im Personal-Sachkosten-Verhältnis zwischen größeren und kleineren Krankenhäusern herausstellen.

²⁸¹ und anderer Wirtschaftseinheiten oder -subjekte.

höchste Anzahl an Subjekten (22,6 %), während dieses Bundesland zu einer ärztebezogenen Analyse nur 17,3, zu einer fallbezogenen hingegen 18,4 % der Information beitragen sollte.

Tab. 6.21.: Mögliche Wichtungen dokumentiert anhand der Anteile [%] an Krankenhäusern bzw. Kreisen aller Bundesländer im Untersuchungsgebiet (2005)

	BE	BB	MV	RP	SN	ST	TH	Σ
Krankenhäuser	16,6	10,9	7,9	22,6	19,9	11,5	10,6	100
Kreise	0,7	12,2	12,2	24,5	19,0	16,3	15,0	100

In die Analyse der kreisweisen kumulativen Krankenseinheiten geht Berlin als ein Kreis unter 147 mit genau demselben Informationsanteil von 0,7 % ein, wie ein Kreis mit nur einem Krankenhaus: Berlins Gewicht in den Analysen ist erheblich gesunken. Die mit dem Perspektivwechsel Krankenhaus zu Kreis einhergehenden Veränderungen sind für Berlin extrem, die anderen Bundesländer verschieben ebenfalls ihre Positionen gegeneinander (Tab. 6.21).

Ausgleich könnte jeweils eine Gewichtung der einzelnen statistischen Einheiten schaffen: nach Bettenzahl, Fällen, Kosten oder – für kreisbezogene Analysen – nach Anzahl der Krankenhäuser, Einwohnerzahl oder dem Wert eines relevanten Wirtschafts- bzw. Sozialparameters. Die Auswahl der Gewichtung hat (potenziell) einen großen Einfluss auf die Ergebnisse, jede Auswahl ist zu begründen.

Die hier gewählte Gewichtung von konstant 1 sowohl für die Ebene der Krankenhäuser als auch die der Kreise gibt den mikroökonomischen Erfassungseinheiten der Krankenhausstatistik bzw. der ergänzenden Wirtschafts- und Sozialstatistiken den Vorrang vor häufigkeits- oder bedeutungsbasierten Gewichten; Differenzen in der Aussage beider Analysegruppen werden in Anerkennung dieser Gewichtungen interpretiert.

Beschäftigte der Krankenhäuser

Im Jahr 2005 sind in den Krankenhäusern im Mittel $1\frac{1}{2}$ -mal so viele Ärzte beschäftigt wie 1992. Demgegenüber sind die Unterschiede zwischen Ost und West bereits 1992 verschwindend gering, sie verringern sich bis 2005 weiter.

Das Geschlechterverhältnis²⁸² unter den Krankenhausärzten verändert sich dabei deutlicher. Der Frauenanteil unter den Krankenhausärzten nimmt in Ostdeutschland um etwa 2 Prozentpunkte von 40,5 % auf 42,5 % zu, die Zunahme in Westdeutschland ist wesentlich stärker bei wesentlich geringerem Ausgangswert. Für Rheinland-Pfalz steigt der Frauenanteil von 25,8 % im Jahr 1992 auf 35,7 % zum Ende des Berichtszeitraums und liegt damit immer noch deutlich unter dem Ostdeutschlands. Für Krankenhausärzte gilt somit ähnliches wie für die beiden deutschen Teilbevölkerungen: Die Erwerbstätigenquoten der Frauen ändern sich zwischen den Fünfjahresperioden 1990–1994 und 2005–2009 von 78,5 %²⁸³ auf 81,6 % für Ostdeutschland und von 64,7 % auf 71,8 % für Westdeutschland (bei deutlich unterschiedlichen Arbeitslosenquoten, Krause et al., 2010, Tab. 2).

Auffällig dabei scheint, dass es sich bei den Kreisen mit den höchsten Frauenanteilen unter den Krankenhausärzten eher um flächenmäßig große, ländliche Kreise handelt (Abb. 6.18²⁸⁴). Diese Beobachtung wird hier nicht weiter verfolgt.

Unter den nichtärztlichen Beschäftigten der Krankenhäuser sind weit geringere Veränderungen des Geschlechterverhältnisses zu beobachten, der Abstand zwischen Ostdeutschland und Rheinland-Pfalz verringert sich minimal durch eine Verringerung des Frauenanteils in Ostdeutschland. Während die Dichte an nicht-ärztlichem Personal von bereits höherem Niveau in Rheinland-Pfalz zunimmt, sinkt sie von niedrigerem in Ostdeutschland weiter ab, der Abstand zwischen Ost und West vergrößert sich. Es kann hier nur vermutet werden, dass die von Werblow und Schoffer (2010) erwähnten *Outsourcing*-Aktivitäten der Krankenhäuser im Bereich des nichtärztlichen Personals eine merkliche Ursache sind.

²⁸² Modell-Prädiktionen.

²⁸³ Wert vom Autor anhand weiterer Angaben der Quelle korrigiert.

²⁸⁴ In der Darstellung unterdrückt sind Kreise mit 2 oder weniger Krankenhäusern. Das Komplement – die Kreise mit 3 oder mehr Krankenhäusern und gleichzeitig häufig die urbaneren – ist von der Informationsunterdrückung ausgenommen.

Die Kosten der Krankenhäuser in Transition

Die fallbezogenen relativen bereinigten Kosten der Krankenhäuser steigen und sind dabei *ceteris paribus* im Mittel in Ostdeutschland höher als in Rheinland-Pfalz. Modell 1²⁸⁵ zeigt aber auch den Einfluss der Krankenhausgröße – angenähert durch die Fallzahl – auf die bereinigten Kosten, größere Krankenhäuser liegen tendenziell unterhalb des Mittels. Die Krankenhäuser Ostdeutschlands sind dabei im Mittel größer als jene in Rheinland-Pfalz.

Eine Modellierung – Modell 2 – mit weiteren, das einzelne Krankenhaus charakterisierenden Merkmalen spricht dafür, dass die ostdeutschen Krankenhäuser 1992 geringere, 2005 jedoch höhere fallbezogene relative bereinigte Kosten als die Pendanten in Rheinland-Pfalz haben. Wichtige Merkmale dabei sind die Trägerschaft, der Krankenhaustyp, die Krankenhausgröße und die Anzahlen an ärztlichen sowie nichtärztlichen Beschäftigten – obwohl hoch korreliert hat jedes dieser Merkmale als Regressor²⁸⁶ einen großen Einfluss auf die Prädiktion der fallbezogenen relativen bereinigten Kosten.

Hinsichtlich der Aufteilung der Kosten in Personal- und Sachkosten ergeben sich für Rheinland-Pfalz immer deutlich höhere Werte im Personal-Sachkosten-Verhältnis als für Ostdeutschland. Über die Gründe geben die Daten keine Auskunft. Vorstellbar ist, dass zum Beginn des Untersuchungszeitraumes das in Ostdeutschland deutlich geringere Lohn- und Gehaltsniveau und der sich über mehrere Jahre erstreckende erhöhte Investitionsbedarf zu dieser Situation bedeutsam beitragen, und dass die ostdeutschen Krankenhäuser in Transition flexibler auf die moderneren Methoden des *Outsourcing* zurückgreifen.²⁸⁷ Ein verstärktes *Outsourcing* von nichtärztlichen Tätigkeiten verschiebt das Personal-Sachkosten-Verhältnis hin zu höheren Sachkosten.

²⁸⁵ Abschn. 6.2.8.

²⁸⁶ ggf. auch in Wechselwirkung mit anderen.

²⁸⁷ Das bedürfte weiterer Analysen.

Der Einfluss der Krankenhäuser auf die Wirtschaft

Zur Sicherstellung ihres Betriebs tätigen Krankenhäuser²⁸⁸ Investitionen, beauftragen Bau- und Reparaturleistungen, kaufen Verbrauchsmaterialien für die medizinische Versorgung und die Pflege der Patienten, binden Personal. Damit beeinflussen die Krankenhäuser lokal den Arbeitsmarkt direkt und zunehmend indirekt durch das Einbinden externer Dienstleister (*Outsourcing*, vgl. Werblow und Schoffer, 2010) zuungunsten eigener Beschäftigter, während die Wertschöpfung bei medizinischen Verbrauchsmaterialien oder größeren Bauinvestitionen zu einem großen Teil auch weniger lokal vermutet werden muss.

Als sicher anzunehmen ist ebenfalls, dass dieser Einfluss von der Größe des Krankenhauses abhängt.

In seiner Arbeit *Zur volkswirtschaftlichen Bedeutung des Gesundheitswesens im Spiegel der amtlichen Statistik* schreibt Richter (2008), dass es erst seit der um 2003 erfolgten Einführung des Begriffs *Gesundheitswirtschaft* als Ziel verstanden wird, den Anteil und die Verflechtungen der Gesundheitswirtschaft amtlich statistisch zu erfassen und ... *auch in die Berechnungen und Schätzungen zur Stellung des Gesundheitswesens im volkswirtschaftlichen Kontext einfließen zu lassen*.²⁸⁹ Krankenhäuser wären mit 27,0 %²⁹⁰ ein wesentlicher Teil.

Für eine Deskription oder mikroökonomische Modellierung des Einflusses der Krankenhäuser auf die Wirtschaft liegen im Rahmen dieses Projekts keine ausreichenden Daten vor.

Der Einfluss der Wirtschaft auf die Krankenhäuser

Während vorliegende Daten einen ursächlichen Einfluss der Krankenhäuser auf die Wirtschaft nicht empirisch untersuchen lassen, können die Daten mikroökonomischen Einblick geben, ob und wie die Krankenhäuser in ihren Kreisen von

²⁸⁸ bzw. deren Träger.

²⁸⁹ Die ersten Veröffentlichungen sind Mackenthun et al. (2003) bzw. Henke et al. (2003) zum Berliner Gesundheitsmarkt.

²⁹⁰ der laufenden Gesundheitsausgaben 2005 in Deutschland (Statistisches Bundesamt, 2007, S. 251).

den wirtschaftlichen Rahmenbedingungen beeinflusst sind. Für diese Analysen werden die Einzelkrankenhäuser kreisweise kumuliert (Abschn. 5.1.2: Gesamt-krankenhausaktivität). Einerseits werden damit Wirtschafts- und Sozialdaten auf Kreisniveau den Krankenhäusern zuordenbar, und andererseits sind die Kreistage die kleinste politisch-administrative Einheit, deren Entscheidungen (im Rahmen des Landeskrankenhausplans²⁹¹) gravierenden Einfluss auf die Krankenhauslandschaft und die Existenzbedingungen der lokalen Krankenhäuser nimmt.²⁹² Das kumulierte Krankenhausangebot steht den Einwohnern, Gästen und Einpendlern zur Verfügung.

Zielparameter der Betrachtungen sind wiederum die bereinigten Kosten je Fall (als Maß der Aufwendungen) und das Personal-Sachkosten-Verhältnis. Wünschenswert für die Einschätzung der Kosten und der Kosteneffizienz der Krankenhäuser in den Kreisen wäre der Bezug zum Schweregrad der jeweiligen Fälle, wie sie z. B. ein DRG-basierter CMI böte.²⁹³ Ohne Schweregrad bietet sich die reine Fallzahl als gute Näherungsvariable an, wenn man davon ausgeht, dass zwischen den betrachteten Gebieten keine Unterschiede in der Verteilung der Fallschwere bestehen. Für einzelne Kreise sehr verschiedener Urbanität scheint diese Annahme wenig sinnvoll; für den Vergleich Ostdeutschlands mit Rheinland-Pfalz als Repräsentanten der westdeutschen Bundesländer ist sie jedoch sinnvoll. Diese Annahme liegt den Betrachtungen zu den bereinigten Kosten je Fall zugrunde.

Paneldaten-Modellierungen²⁹⁴ mit Wirtschafts- und Sozialparametern als zusätzlichen Regressoren lassen sich wegen der Datenverfügbarkeit erst ab dem Jahr 1996 formulieren. Vorgeschaltete Analysen zeigen, dass die verfügbaren Wirtschafts- und Sozialparameter miteinander für den Berichtszeitraum miteinander hoch korreliert sind und jeder einzelne – mit Ausnahme der Brutto-Erwerbstätigenquote – einen guten Ost-West-Separator darstellt (Abb. 6.2–6.6).

Modell 5 zeigt, dass der Einschluss eines der Wirtschafts- und Sozialparameter den unmittelbaren Ostdeutschland-Indikator für das Modell bei den bereinigten Kosten als eher unwichtig erscheinen lässt, für das Personal-Sachkosten-Verhält-

²⁹¹ gemäß Krankenhausfinanzierungsgesetz (KHG).

²⁹² Einige krankenhausspezifische Informationen wie Trägerschaft, Krankenhaustyp sind von sinnvollen Aggregationen ausgenommen.

²⁹³ Solche Analysen werden mit den Daten der DRG-Statistik für die Jahre nach 2006 zunehmend möglich sein.

²⁹⁴ Abschn. 6.2.9.

nis scheint der Einfluss eher zwischen Ostdeutschland-Indikator und Wirtschafts- bzw. Sozialparameter ‚geteilt‘. Der Verzicht auf den Wirtschafts- bzw. Sozialparameter gibt in beiden Fällen dem Ostdeutschland-Indikator seine Bedeutung für ein gutes Modell zurück. Wie anhand der vorgeschalteten Analysen vermutet, können die Wirtschafts- und Sozialparameter weder die Prädiktion der bereinigten Kosten je Fall noch des Personal-Sachkosten-Verhältnisses relevant beeinflussen. Modell 5 für die Kosten zeigt auch, dass das Einbinden einer quadratischen Zeitkomponente²⁹⁵ die Modellanpassung verbessert: die Annahme eines nichtlinearen Trends über die Zeit ist sinnvoll, die Kosten nehmen überlinear zu.

Trotz anderer Gewichtung der Kreise²⁹⁶ und Auslassung der quadratischen Zeitkomponente ergeben sich mit Modell 3²⁹⁷ sehr ähnliche Schlussfolgerungen.

Somit ist der mikroökonomische Einfluss der Wirtschaft auf die Krankenhäuser als eher gering einzuschätzen²⁹⁸, für die Jahre 1996–2005 sind sowohl für das Personal-Sachkosten-Verhältnis als auch die bereinigten Kosten der Ost-West-Unterschied und die Veränderungen über die Berichtsjahre die wichtigsten Erklärenden.

Neben der Tatsache, dass sich die wirtschaftlichen Unterschiede zwischen Ost und West im Berichtszeitraum (noch) nicht nivellieren, ist vermutlich das System der Finanzierung des deutschen Gesundheitswesens (Abschn. 3.1.3) ein entscheidender Faktor. Der Gesundheitsmarkt wird in gesamtgesellschaftlichem Interesse stark reguliert (Abschn. 3.2.6), Bundesgesetze gelten kreisunabhängig im ganzen Land, viele der wichtigsten Finanziierer (von den Allgemeinen Ortskrankenkassen²⁹⁹ über Betriebskrankenkassen und Ersatzkassen bis zu Privatversicherern) sind überregional mit demselben Angebot und Leistungsspektrum – auch und gerade im Umgang mit den Krankenhäusern – tätig.³⁰⁰ Eher ist zu erwarten, dass

²⁹⁵ Quadrat des Merkmals *jahr*.

²⁹⁶ *krankenhausanzahlbezogen* statt konstant 1.

²⁹⁷ Abschn. 6.2.8.

²⁹⁸ Makroökonomisch ist das anders: Die Finanzierung des Gesundheitssystems eines Landes hängt stark von der wirtschaftlichen Leistungskraft und dem erarbeiteten Wohlstand ab.

²⁹⁹ im Berichtszeitraum 16 (bundeslandweise); durch Fusionen aktuell (Stand 01. 01. 2012) 12, gemeinsam vom AOK-Bundesverband vertreten.

³⁰⁰ Auch für die fallpauschalenbasierte Krankenhausvergütung nach *diagnosis-related group* besteht mit der Etappe *Einheitlicher Landesbasisfallwert (LBFW) 2009* das Ziel, bundeseinheitlich Basisfallwerte zu definieren, Abweichungen sollen sich in einem engen Basisfallwertkorridor bewegen dürfen.

sich lokale Unterschiede bei den Krankenhäusern zu einem bedeutenden Teil als Reaktion auf die demografischen Gegebenheiten³⁰¹ und das damit verbundene Krankheitenspektrum verstehen lassen.³⁰²

6.3.2. Wirtschaftlichkeit bei der Behandlung der Patienten

Krankenhäuser behandeln Patienten – verstanden als Fälle –, die Anzahl der Behandlungsfälle ist ein wichtiger Output-Indikator (Abschn. 3.2.4) wie auch die Anzahl der Pflgetage und die daraus abgeleitete Verweildauer der Fälle. Diesem Output steht ein Verbrauch an Ressourcen gegenüber. Wirtschaftlichkeit als das Verhältnis zwischen dem Ertrag und den eingesetzten Mitteln wird hier für die Krankenhäuser über die Beziehungen zwischen Ressourcenverbrauch und Output diskutiert.

Fallzahlen nach Aufnahmemonat

Abb. A.5 zeigt für Rheinland-Pfalz und Ostdeutschland getrennt die jährlich wiederkehrenden Abweichungen in den Patientenaufnahmen pro Monat gegenüber dem allgemeinen langfristigen Trend (Abschn. 6.2.2 und Abb. 6.14). Auffällig ist das jeweilige Dezembertief, wohl verursacht durch eine deutliche Verminderung elektiver³⁰³ Krankenhausaufenthalte, das im Wesentlichen durch die deutlich weniger ausgeprägten Hochs in den drei Monaten des ersten Quartals³⁰⁴ kompensiert wird. Da die Ausstattung der Krankenhäuser den verstärkten Fallzahlen der ersten Quartale genügt, muss davon ausgegangen werden, dass im Dezember ein deutlich geringerer Nutzungsgrad für vorhandene Betten, OP-Säle, Großgeräte anzusetzen ist, gebundene Ressourcen werden weniger genutzt, die sachkostenbezogene Wirtschaftlichkeit ist vermindert.

³⁰¹ vor allem die Altersverteilung, aber auch Geschlechtsverteilung, Umweltgegebenheiten,

³⁰² Für eine solche Analyse wären die Definition von Krankenseinzungsbereichen und demografisch-epidemiologische Daten zu den Einwohnern und Krankenhaussfällen vonnöten.

³⁰³ auswählend; medizinisch: nicht wirklich dringend notwendige und deshalb zeitlich (frei) planbare Eingriffe bzw. Operationen.

³⁰⁴ darunter der Februar mit nur 28 bzw. 29 Tagen.

Einhergehen dürfte das mit einer erhöhten Urlaubsrate um die Weihnachtszeit, Weihnachten und Jahreswechsel dienstfrei für größere Teile des Personals trägt sicher ebenso zu einem guten Betriebsklima bei, wie diese Zeit zu Hause im Familienkreis für viele Genesende die psychosozialen Aspekte des Heilungsprozesses verbessert. Beide Effekte können hier nur vermutet werden, das Bestimmen ihrer Anteile an der Wirtschaftlichkeit des Krankenhausbetriebs und des Gesundheitssystems insgesamt ist als deutlich schwieriger anzusehen.

Wohn- und Behandlungskreis

Friedrich und Beivers (2009) untersuchen in einer Querschnittsanalyse basierend auf allen 6 Millionen Fällen von AOK-Versicherten des Jahres 2006 die räumliche Mobilität von Krankenhauspatienten. Spezielles Augenmerk legen Sie auf Patienten mit einer Hüftendoprothesen-Implantation³⁰⁵. Besonders Elektivpatienten entscheiden sich häufig nicht für das wohnortnächste Krankenhaus sondern für ein Krankenhaus mit einer höheren Anzahl vergleichbarer Operationen, dieses ist oft kleiner, jedoch auf diese Behandlung spezialisiert.³⁰⁶ Mit dem Zugriff auf die Adressen der Krankenhäuser und der Patienten werden die Entfernungen zum nächsten und zum gewählten Krankenhaus sowie die Differenz als Maß für den zusätzlichen Aufwand geschätzt. Gefolgert wird eine bewusste Krankenhauswahl durch Patienten unter Inkaufnahme größerer Entfernungen, wobei anhand der Daten unbekannt bleiben muss, ob die Prioritäten vom (Haus-)Arzt, vom Patienten selbst oder anderen gesetzt werden. Mögliche Ost-West-Unterschiede werden genauso wenig untersucht wie zeitliche Veränderungen.

Die hier vorliegenden Panelanalysen können weder auf die feine Geokodierung von Krankenhaus und Wohnort noch die Einteilung der Fälle in elektive und Notfälle zurückgreifen. Trotzdem zeigt sich auch bei den kreisbasierten Angaben zu Wohnort und Krankenhaussitz, dass zunehmend wohnortfernere Krankenhäuser die Behandlung übernehmen. Neben der von Friedrich und Beivers beschriebenen bewussten Krankenhauswahl unter Umgehung des wohnortnächsten könn-

³⁰⁵ 72.146 Operationen, davon 70,4 % elektiv, 29,6 % Notfallversorgung.

³⁰⁶ Notfallpatienten haben nur geringen Einfluss auf die Wahl des Erstkrankenhauses, doch für die Weiterbehandlung wird – in geringerem Maße – von einer ähnlichen Entscheidungspräferenz berichtet.

ten auch eine zunehmende Spezialisierung (und eine damit einhergehende Verringerung der anderen Behandlungsangebote) der Krankenhäuser oder die Reduktion der Krankenhauszahlen Patienten zu weiteren Wegen zwingen – zwei weitere mögliche Ursache für den Trend. Vorstellbar ist weiterhin, dass Patienten mit bestimmten Diagnosen³⁰⁷ die Anonymität ferner gelegener Krankenhäuser bevorzugen, auch sie nähmen erhöhte Kosten in Kauf. Andererseits könnten Spezialisierung und damit verbundener Erfahrungszuwachs zu geringeren Kosten und somit gesteigerter Wirtschaftlichkeit der Krankenhäuser führen.

So wie Friedrich und Beivers (2009) auf die Unterschiede zwischen verschiedenen Diagnosen hinweisen, zeigen hier vorgelegte diagnosegruppierungsspezifische Analysen ein gemischtes Bild: Sowohl der Anteil der Patienten mit höherer Mobilität (als Fälle mit krankenhausesfernem Wohnkreis) als auch deren Veränderungen im Berichtszeitraum stellen sich für verschiedene Diagnose-Gruppierungen unterschiedlich dar. Friedrich und Beivers folgend ist ohne das Wissen um die Elektivität des Krankenhausfalls für alle diese Analysen ein *omitted variable bias* anzunehmen.

Die Kreise Ostdeutschlands sind mit im Mittel 983 km² flächenmäßig 1,9-mal so groß wie die in Rheinland-Pfalz (512 km²)³⁰⁸. Weitergehende Analysen zur Mobilität der Krankenhauspatienten könnten klären, ob dieses eine wesentliche Ursache dafür ist, dass in allen untersuchten Diagnose-Gruppierungen ein Ost-West-Unterschied bezüglich der die Kreisgrenzen überschreitenden Mobilität der Patienten gegeben ist, sie ist für ostdeutsche Krankenhäuser die geringere.

Verweildauern

Die mittlere Verweildauern – das arithmetische Mittel – ist eine auch international sehr beliebte Kennzahl zum Vergleich von Gesundheitssystemen mit dem Ziel möglichst geringer Werte. Die OECD³⁰⁹ warnt sehr deutlich vor länderspezifischen Definitionsunterschieden besonders bezüglich des Einschlusses von Fällen

³⁰⁷ z. B. Schwangerschaftsabbruch, psychische oder venerische Krankheiten.

³⁰⁸ Statistisches Bundesamt (2007, S. 29) und eigene Berechnungen.

³⁰⁹ http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT und Verweisungen dort, Zugriff am 18. 10. 2010.

mit Verweildauern von 0 oder 1 Tag und der vollständigen Abdeckung der Krankenhauslandschaft.

Verweildauern werden hier aus der wirtschaftlichen Sicht auf die Krankenhäuser analysiert, ein Todesfall beendet den Krankenhausaufenthalt genauso wie eines Genesenden Entlassung oder Überweisung zur Weiterbehandlung, der Aufwand für das Krankenhaus endet. Im Untersuchungszeitraum zeigt sich, dass in beiden Teilen Deutschlands eine bemerkenswerte Verbesserung allein durch die Verringerung des Langliegeranteils möglich ist, vermutlich (auch) durch Verlagerung von Fällen aus den Krankenhäusern in Pflegeeinrichtungen.

Die gesetzlichen Änderungen des Finanzierungssystems des Gesundheitswesens (Abschn. 3.2.1) weg von der Mischfinanzierung mit tagesbezogenen Pflegesätzen hin zu Fallpauschalen verändern das betriebswirtschaftliche Interesse an Pflegetagen bzw. Verweildauern. Das Interesse der Krankenhäuser an kürzeren Verweildauern für gleichartige Fälle wächst.

Die Analysen zeigen jedoch auch bisher nicht diskutierte Muster in den Verweildauerverteilungen. So weisen die Hazardfunktionen (Abb. 6.13) periodische Muster auf, die sich qualitativ auch unter den sich verändernden Finanzierungsbedingungen des Untersuchungszeitraums nicht ändern: es gibt bevorzugte Verweildauern. Dieses Phänomen ist für bestimmten Diagnose-Gruppierungen ausgeprägter als für andere. Lokal erhöhte Chancen entlassen zu werden, bestehen oft im Wochenabstand zum Aufnahmetag. Einfache Erklärungsmuster lassen sich weder medizinisch-epidemiologisch noch betriebswirtschaftlich finden. Inwieweit eine nach dem Wochentag der Aufnahme stratifizierte Analyse weitere Aufklärung bringt, muss hier offen bleiben.

Den Erwartungen entsprechend sind Verweildauern diagnosespezifisch, wie sowohl die Kaplan-Meier-Schätzungen der Verweildauerverteilungen (Abb. 6.12) als auch die geglätteten Hazardfunktionen für jeweils ausgewählte Diagnose-Gruppierungen zeigen. Die Paneldaten-Regressionen für die verschiedenen Diagnose-Gruppierungen lassen jedoch weder einen bemerkenswerten Ost-West-Unterschied noch einen relevanten Einfluss der untersuchten Wirtschafts- oder Sozialparameter erkennen. Wie bereits ausgeführt (Abschn. 6.3.1: Der Einfluss der Wirtschaft auf die Krankenhäuser), ist vermutlich das gesamtnational einheitliche und wenig transitionbetreffene System der Finanzierung des deutschen Gesund-

heitswesens eine wichtige Voraussetzung für den geringen Einfluss der mikroökonomischen Gegebenheiten auf die Verweildauer-Veränderungen.

Der Nachweis behandelter Krankheiten anhand der Kodierung

Die Hauptdiagnose ist das Merkmal in der Diagnosestatistik, das den Grund für den Aufenthalt im Krankenhaus beschreibt, ihre Definition hat sich im Berichtszeitraum nicht geändert, sie ist auch nach der Einführung der DRG-Statistik als in den Krankenhäusern parallel zu erhebender Statistik gültig (Abschn. 4.3.3: Diagnosedaten). Zwischen der Diagnose zum Patienten selbst und dem Eingang der Information in die Krankenhausstatistik ist die Diagnose zu kodieren: aus einer endlichen Anzahl von Schlüsseln ist genau einer auszuwählen, sich leicht unterscheidende Fälle werden gegebenenfalls mit demselben Schlüssel³¹⁰ registriert, nur die Schlüssel stehen einer Analyse zur Verfügung.

Abb. 6.17 (Abschn. 6.2.4) zeigt, wie stark sich im Untersuchungszeitraum die Anteile einzelner Schlüssel für Fälle der Diagnose-Gruppierung *Normale Entbindung, Komplikationen im Verlauf der Wehen und der Entbindung* verändert haben, der Anteil der mit ICD-9-Schlüssel 650 (normale Geburt) kodierten Fälle sinkt von jeweils mehr als 68 % in den Jahren 1993–2000 auf weniger als 14 % im Jahr 2005. Medizinisch bedeutete das einen dramatischen Abfall des Anteils normaler Geburten zugunsten verschiedener Komplikationen (656 – Sonstige fetale und placentare Störungen, die den Zustand der Mutter beeinträchtigen; 658 – Sonstige Störungen im Zusammenhang mit Amnionhöhle und Eihäuten; 665 – Sonstige Geburtsverletzungen). Eine medizinische Fachdiskussion dieser gravierenden Veränderungen ist dem Autor bis heute³¹¹ nicht bekannt geworden. Lediglich im *Krankenhaus-Report 2005. Schwerpunkt Wege zur Integration* widmen sich Heller und Schmidt (2005) unter dem Titel *Wodurch ist die Veränderung der geburtshilflichen Diagnosehäufigkeiten in der Krankenhausdiagnosestatistik zu erklären?* diesem Phänomen für die Jahre 2001–2003³¹². Sie konstatieren: *Die Krankenhausdiagnosestatistik zeigt im Bereich der Geburtshilfe in den vergangenen Jahren drastische*

³¹⁰ zum Wechsel des Kodierungsschemas siehe Abschn. 4.3.2: Leistungsfähigkeit der ICD-Konkordanzlisten.

³¹¹ 17. 01. 2012.

³¹² alle Diagnosen ICD-10-kodiert.

Veränderungen. Es werden zunehmend Diagnosen kodiert, die auf Pathologien hinweisen. Es liegt nahe, dass dieser Sachverhalt am ehesten durch veränderte Abrechnungsbedingungen (DRG-Einführung) bedingt ist. Sie vermuten Upcoding und empfehlen weitergehende Untersuchungen.

Der Trend zur Kodierung von auf Pathologien – krankhafte oder abnorme Befunde – hinweisenden Diagnosen setzt sich 2004 und 2005 verstärkt fort, so dass 2005 nur noch jede 7. (im Krankenhaus erfolgte) Geburt als normal eingestuft wird.

Aus wirtschafts- und sozialwissenschaftlicher Sicht ergeben sich hier weiterführende interessante Fragen:

1. Das vermutete *Upcoding* beginnt mit dem Jahr 2001, die DRG-Einführung für die Krankenhäuser beginnt gestaffelt 2003 bzw. 2004 mit einer Kalibrierungsphase zur Einstellung der Vergütung für die jeweiligen DRG-kodierten Fälle. Die Vergütungen sind dabei gesamtkostenneutral einzustellen. Welche Auswirkungen hat ein *Upcoding* während der Kalibrierung auf zukünftige Vergütungsverhältnisse?
2. 2001 und 2002 sind Jahre ausschließlich der Diskussion (und vehemente Ablehnung durch große Teile der Ärzteschaft) zukünftiger DRG-basierter Prozesse. Ist allein die Präsenz des Themas für eine derart gravierende Verhaltensänderung verantwortlich zu machen?
3. Sind die beschriebenen Kodierveränderungen wirklich nur auf *Upcoding* zurückzuführen, oder ist auch davon auszugehen, dass in Zeiten, in denen mit den medizinischen Diagnosen keine unmittelbaren wirtschaftlichen Folgen verbunden sind, zur Bedienung anderer Erfolgsparameter (z. B. Quote komplikationsloser Geburten) oder aus mangelndem Interesse *Downcoding* die wahren Verhältnissen verzerrt darstellen lässt?

Upcoding ist ein Problem jeder mit wirtschaftlichen Interessen verbundenen Informationsverschlüsselung. Spezifisch für das Gesundheitswesen hat sich auch der Begriff *DRG creep* eingebürgert. In den USA werden diesbezügliche Untersuchungen seit Jahren (auch) vom Gesundheitsministerium (US Department of Health and Human Services) angestrengt, z. B. 1992 *National DRG Validation Study*³¹³ und 1999 *Monitoring the Accuracy of Hospital Coding*³¹⁴ mit der Aussa-

³¹³ <http://oig.hhs.gov/oei/reports/oei-12-89-00190.pdf>, Zugriff am 04. 02. 2011.

³¹⁴ <http://oig.hhs.gov/oei/reports/oei-01-98-00420.pdf>, Zugriff am 04. 02. 2011.

ge: *The DRG system is vulnerable to abuse by providers who wish to increase reimbursement inappropriately through upcoding, particularly so within certain DRGs. Our analysis found noticeable, detectable, and curable upcoding abuses among providers and within specific DRGs.* Als besonders vom *Upcoding* betroffene Diagnose-Gruppierungen werden Erkrankungen der Atemwege, Vergiftungen des Blutes und Gefäßerkrankungen des Gehirns³¹⁵ identifiziert; Geburtshilfe wird nicht erwähnt.

DRG creep-Probleme lassen sich wahrscheinlich aus allen Ländern mit DRG-basierter Krankenhausfinanzierung berichten, diese scheinen systemimmanent und verschwinden nicht mit der Zeit.³¹⁶

In Deutschland befasst sich der Bundestag mit dem Thema *Upcoding* u. a. im *Gutachten 2007 des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen*³¹⁷ im kurz Kapitel 4.2.1.2 *Erwartete Fehlanreize und Gegensteuerung*, wo vor allem auf die dem *Upcoding* entgegenstehenden Gesetze und Prüfmöglichkeiten durch die Krankenkassen bzw. deren Medizinischen Dienst hingewiesen wird.

Das Statistische Bundesamt sieht in seinen jährlichen Qualitätsberichten kein Problem und argumentiert ausgerechnet mit der wirtschaftlichen Bedeutung der Kodierung: *Dadurch, dass die Angabe der Hauptdiagnose für die Krankenhäuser entgeltrelevant ist, kann davon ausgegangen werden, dass die Qualität der Diagnoseangabe hoch ist. Sollten dennoch Unstimmigkeiten auftreten, so können diese in der Datenaufbereitungsphase bei der Plausibilisierung der Angaben berichtigt werden, z.B. wenn Diagnose- und Geschlechtsangabe des Patienten/der Patientin nicht übereinstimmen* (Statistisches Bundesamt, 2006b)³¹⁸.

Sowohl die hier vorgelegten als auch die von Heller und Schmidt (2005) empirisch gewonnenen Ergebnisse widersprechen dieser Aussage wie auch der Aussage von Strausberg (2007): *Befürchtungen über ein flächendeckendes Upcoding nach Ein-*

³¹⁵ u. a. Schlaganfall.

³¹⁶ Ungarn, DRG seit 1993, berichtet 2010: *Physicians and hospitals know approximately which diagnosis and which intervention to code in order to get more money than general.* http://www.knowandpol.eu/fileadmin/KaP/content/Scientific_reports/Orientation3/O3_2.Hungary_health.Version_2.pdf, Zugriff am 04. 02. 2011.

³¹⁷ Deutscher Bundestag, 16. Wahlperiode, Drucksache 16/6339, 07. 09. 2007.

³¹⁸ unverändert auch im Qualitätsbericht für 2011.

führung der DRGs ... haben sich nicht bestätigt. Strausberg argumentiert, dass die Kodierung *plausibel* sei und mit der papierbasierten Krankenakte übereinstimme. Das hieße, dass man eine Diskrepanz zwischen DRG-Kodierung und Diagnosekodierung nach ICD vermutet. Wenn der Verdacht besteht, dass die Diagnosestatistik und die DRG-Statistik – beides amtliche Statistiken der Fälle der Krankenhäuser – mit unterschiedlichen Informationen beschickt werden, böte sich im Rahmen der Qualitätsuntersuchungen ein Abgleich beider an.

Die Einführung der diagnosis-related group als Grundlage fallpauschaler Vergütung der Krankenhäuser und die jährliche Berichterung im Rahmen der DRG-Statistik hat Auswirkungen auf die berichteten Diagnosen in der Krankenhausstatistik. Offensichtlich gilt das für die Fälle der Geburtshilfe der Jahre 2001–2005, ein Ost-West-Unterschied zeigt sich dabei nicht. Wenn auch hier nicht so offensichtlich, muss doch angenommen werden, dass sich der Einfluss nicht auf die Diagnosekodierung der Geburtshilfe beschränkt. Ergebnisse aus anderen Ländern unterstützen diese Sicht.

Neben den wirtschaftlichen Implikationen stellt sich auch die Frage, inwieweit die Diagnosen der Krankenhausstatistik für die qualitativ hochwertige Bearbeitung medizinisch-epidemiologischer Fragestellungen herangezogen werden können.

6.4. Empirische Analysen – Zusammenfassung

Deskription und Exploration der Daten der amtlichen Krankenhausstatistik in Verbindung mit Wirtschafts- und Sozialparametern aus öffentlich zugängigen amtlichen Statistiken geben ein Bild von den Beziehungen zwischen den Krankenhäusern als Teil des Gesundheitswesens und den wirtschaftlichen und sozialen Rahmenbedingungen im Umkreis der Krankenhäuser. Die Jahre des Berichtszeitraums (1992–2005) sind einerseits geprägt von den Veränderungen der Transition in den 1990 wiedererstandenen Bundesländern Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen sowie im wiedervereinten Bundesland Berlin. Andererseits nehmen diese Bundesländer teil an den allgemeinen Veränderungen in Deutschland. Zur Trennung zwischen transitionsbedingten und allgemeinen Veränderungen wird das Bundesland Rheinland-Pfalz als Referenz und Vergleichsgrundlage ebenfalls betrachtet.

In vorliegender Arbeit werden die Ergebnisse eines zweistufigen Verfahrens vorgestellt. Unter den Bedingungen der Arbeit am Gastwissenschaftler-Arbeitsplatz (GWAP) gelingt es, die GWAP-eingeschränkte Version des Registers zur Grundlage der Modellentwicklung zu machen; wesentliche Kollinearitäten werden gefunden, ein erster Ergebniseindruck wird gewonnen, weitergehende Analysen werden vorbereitet.

Das auf dem vollständigen Datensatz basierende Register wird im zweiten Schritt unter den Bedingungen der KDFV (Kontrollierte Datenfernverarbeitung) analysiert, dabei werden die am Gastwissenschaftler-Arbeitsplatz entwickelten Modelle verfeinert und angepasst.

Die ausgewählten Wirtschafts- und Sozialdaten der Kreise zeigen, dass eine Angleichung zwischen Ost und West bis 2005 nur begrenzt stattfindet, besonders Arbeitslosenquoten, Arbeitnehmereinkommen je Arbeitnehmer und verfügbares Haushaltseinkommen je Einwohner können jeweils als guter Indikator einer Ost-West-Trennung der Kreise herangezogen werden, obwohl die Brutto-Erwerbstätigenquote über alle Jahre für Ost und West gemeinsam eher ländliche von urbaneren Kreisen trennt. Lediglich die 1992 noch deutlich Trennung anhand des Brutto-Inlandsprodukts je Einwohner verliert sich etwas.

Weniger deutlich als im allgemein wirtschaftlichen Bereich hat auch im Krankenhaussektor die Transition bis 2005 Unterschiede nicht verschwinden lassen. Das mag erstaunen, wenn man bedenkt, dass das Gesundheitssystem hinsichtlich der Finanzierung unmittelbar auf das sozialversicherungsbasierte Bismarck-System der Altbundesländer mit Ortskrankenkassen, Ersatzkassen und Privatversicherungen umgestellt wird – in den ersten Jahren mit leicht verminderten Beitragssätzen und Kappungsgrenzen. Neben unterschiedlichen Krankenhausgrößen und leichten Differenzen bei den Kosten ist es auch historisch gewachsene soziale Wirklichkeit, die sich in den Krankenhäusern widerspiegelt. Ein höherer Frauenanteil unter den Krankenhausärzten in Ostdeutschland findet seine Entsprechung in einer weiterhin höheren Erwerbstätigenquote unter ostdeutschen Frauen.

Bedeutender als der Einfluss der Transition stellen sich die politisch forcierten Veränderungen für das gesamte Gesundheitswesen als Auslöser gravierender Veränderungen in den Krankenhäusern dar. Die Krankenhausanzahlen sinken wie die Anzahlen aufgestellter Betten, Krankenhäuser in privater Trägerschaft nehmen zu

und in öffentlicher ab, Verweildauern können kontinuierlich verringert werden. Gleichzeitig steigen die Krankenhauskosten je Bett immens – und zeigen dabei, dass einige der in der Diskussion um die ständig steigenden Kosten des Gesundheitswesens verwendeten Kennzahlen nicht die ihnen zugesprochenen Inhalte transportieren.³¹⁹ Wissenschaftliche Reflektionen, Begründungen und Verbesserungsideen können bisher wenig an (auch international) üblichen Kennzahlen ändern.

Gleichzeitig erweisen sich neue, den Gegebenheiten und Veränderungen der Krankenhausstatistik angepasste Kennzahlen als hilfreich für Beschreibung und Analyse. Das Personal-Sachkosten-Verhältnis ermöglicht Einblick in die Kostenaufteilung auch über die Veränderungen der Kostenerhebung nach Brutto- oder Netto-Prinzip hinweg. Ost und West unterscheiden sich, sie unterscheiden sich vielleicht auch wegen unterschiedlich großer Begeisterung für das *Outsourcing*.

Schon der Versuch, die Anteile der Krankenhäuser in jeweiliger Trägerschaft gegenüberzustellen, macht deutlich, dass Gewichtung der einzelnen Beobachtung einen bedeutsamen Beitrag zur Aussage der Modelle wie der in Kennzahlen komprimierten Information hat: 42 % der Krankenhäuser sind in öffentlicher Trägerschaft, 18,8 % in privater³²⁰. Die bereinigten Kosten geben einen anderen Eindruck: 62,6 % für die in öffentlicher und nur 10,6 % für die Krankenhausleistungen in privater Trägerschaft.

Die Diagnosestatistik gibt einen Blick frei auf die Fälle – die Patienten der Krankenhäuser ohne Reidentifikation beim nächsten Aufenthalt –, ihre Verweildauern und ihre Erkrankung – kodiert und registriert als Hauptdiagnose. Vorliegende Daten führen zu der Vermutung, dass in die Kodierungen nicht nur medizinische Überlegungen einfließen. Besonders die zeitliche Koinzidenz mit der Diskussion und Einführung der neuen DRG-Statistik und anstehenden weitgehenden Umstellung der Krankenhausfinanzierung auf das Fallpauschalensystem spricht für einen starken Einfluss wirtschaftlicher Überlegungen, *Upcoding* bzw. *DRG creep* genannt. Damit sind leider erhebliche Zweifel an jeglicher Diagnosestatistik-basierten medizinisch-epidemiologischen Auswertung angebracht.

³¹⁹ Kostensteigerung aber eben auch Bettenabbau führen zu einer Erhöhung dieser Verhältniszahl.

³²⁰ alle Krankenhäuser (außer Bundeswehrkrankenhäuser) über alle Jahre des Berichtszeitraums.

Hier wird, auch unter der Annahme von *Upcoding*, davon ausgegangen, dass intendierte Fehlkodierungen nur äußerst selten die jeweilige Diagnose-Gruppierung verlassen. Mit gruppierungsspezifischen Verweildaueranalysen werden sowohl quantitative Unterschiede wie auch Ähnlichkeiten beim Muster der Hazardfunktionen herausgearbeitet. Aus ökonomischer Sicht sind diese Muster (bisher) nicht sinnvoll erklärbar.

Einheitlich geltende Bundesgesetze und Leistungskataloge der Krankenkassen auf dem stark regulierten Gesundheitsmarkt und die gesamtstaatliche Basis der wichtigsten Financierer sind vermutlich die entscheidenden Gründe, dass der mikroökonomische Einfluss der Wirtschaft im Vergleich mit anderen Parametern³²¹ auf die Krankenhäuser eher gering ist. Der Einfluss der Krankenhäuser auf die lokale Wirtschaft hingegen ist mit vorliegenden Daten nicht analysierbar.

³²¹ Alters- und Geschlechtsverteilungen, damit und mit Umwelteinflüssen korrelierte Krankheiten, Spezialisierungstendenzen der Krankenhäuser selbst.

Kapitel 7.

Antworten und Ausblick

MAN MERKT NIE, WAS SCHON GETAN WURDE;
MAN SIEHT IMMER NUR DAS, WAS NOCH ZU TUN BLEIBT.
Marie Skłodowska Curie³²²

Wissenschaftsprojekte werden mit Fragen begonnen, jedes Projekt wirft in Entwicklung weitere Fragen auf. Neben den zielbezogenen Fragen stehen die Fragen zum Vorgehen und der Durchführung der Einzelschritte. Manche der Fragen müssen, viele werden beantwortet; Erfahrungen sammeln sich an und sind es wert, berichtet zu werden. Und es bleiben Fragen offen – für zukünftige Untersuchungen.

7.1. Arbeiten unter GWAP- und KDFV-Bedingungen

Im Rahmen der Bedingungen des Vertrages zwischen der Universität Potsdam und dem Forschungsdatenzentrum der Statistischen Landesämter (2007) hat sich für die Arbeit des Autors am und mit dem Standort Kamenz des FDZ/L ein zweistufiges Verfahren für die Analysen insgesamt als vorteilhaft erwiesen. Am Gastwissenschaftler-Arbeitsplatz (GWAP) besteht Zugriff auf etwa 80 % des thematisch

³²² Marie Skłodowska Curie (1867–1934), polnisch-französische Wissenschaftlerin, Nobelpreisträgerin 1903 (Physik) und 1911 (Chemie).

definierten GWAP-Zieldatensatzes³²³, für die Kontrollierte Datenfernverarbeitung (KDFV) werden 100 % der zugesicherten Zieldaten³²⁴ analysierbar bereitgestellt.

Der erste Schritt – Arbeit am GWAP – erfordert dabei viel Zeit und hohe Investitionen in die Datenstrukturierung und die rohe Datenexploration gefolgt von einer Definition der als nötig erachteten Datenqualitätsprüfungen. Nicht minder groß ist die Investition in die Entwicklung der nötigen Programme (Skripte) für diese rohe Exploration, die Qualitätsprüfungen und Datenmodifikationen, die Registererstellung sowie die Analyseschritte von Datendeskription bis -modellierung. All das erscheint kompliziert, extrem arbeitsaufwändig und auf den ersten Blick weit vom Ziel entfernt.

Zumindest zeitlich ist es weit vom Ziel entfernt. Es empfiehlt sich, eine inhaltlich gute Dokumentation der Arbeitsschritte und getroffenen Entscheidungen mit Voraussetzungen bzw. Annahmen und Begründungen anzulegen. All dieses amortisiert sich. Weniger amortisieren sich die umfangreichen Arbeiten, die für die Einhaltung der Anonymisierungsforderungen unter GWAP – speziell die Einhaltung der 3-er Regel – erforderlich sind. Eine erfolgreiche Strategie hier ist, alle Skripte per Fallunterscheidung sofort für die Nutzung am GWAP und für KDFV-Rechenläufe vorzusehen. Fallunterscheidungen möglichst zentral an einer Stelle zu verwalten, ist für den eigenen Überblick sehr hilfreich.

Unschätzbarer Vorteil der Tätigkeit am GWAP ist der direkte Zugriff auf Ergebnisse und Fehlermeldungen, die beiden wichtigsten Reaktionen auf einen jeden Skript-Entwicklungsschritt. Die Reaktionszeiten unter KDFV-Bedingungen sind definitions- und prozessgemäß unvergleichlich viel länger. Parallelarbeit am GWAP und über KDFV empfiehlt sich.

Ob die Ergebnisse am GWAP eine gute Vorausschau auf die Ergebnisse mit dem vollen Datensatz unter KDFV sind, lässt sich nicht allgemeingültig beantworten. Bedeutsam ist bereits die Definition der Zufallseinschränkung für den Datensatz der GWAP-Analysen. Für geplante Panelanalysen empfiehlt es sich, die Selektion

³²³ hier: Krankenhausstatistik-Daten der Jahre 1992, 1996, 2000 und 2004 – aus Kostengründen.

³²⁴ hier: Krankenhausstatistik-Daten von 1991–2005.

auf Panelsubjekte zu beziehen und nicht auf die einzelnen Panelsubjektzeitpunkte³²⁵ – Panel-Attrition und Panel-Zuwachs sind nicht von vornherein verzerrt.

Die Anonymisierungsforderungen unter GWAP – speziell die Aggregationen in Anerkennung der 3-er Regel – führen zu veränderten Korrelationsstrukturen, die bezüglich Kollinearität die Wirklichkeit des vollständigen Datensatzes unzureichend abbilden.³²⁶ Dies ist sicher ein Grund, warum speziell Konvergenzprobleme bei der Modellanpassung von GWAP nach KDFV kaum vorhersagbar sind.³²⁷

Bei guter Vorbereitung ist für ein komplexes und umfangreiches Projekt die Arbeit am GWAP vor Ort im Forschungsdatenzentrum der Statistischen Landesämter ein sehr lohnender Schritt, wenn sie als eine Lernphase für beide, den Wissenschaftler und den betreuenden Mitarbeiter des FDZ/L, gestaltet werden kann. Der Gewinn realisiert sich – sicher zu unterschiedlichen Anteilen – für beide Vertragspartner in einer komplikationsarmen KDFV-Phase des Projekts.

7.2. Antworten

Zehn Fragen stehen am Beginn dieses Projektes (Abschn. 2.2), die erarbeiteten Antworten ergeben zusammengefasst einen Überblick zum Stand und den Perspektiven des Projekts:

- i. Der Beitritt ostdeutschen Länder zur Bundesrepublik und die schnelle Übernahme des bundesdeutschen Rechts-, Wirtschafts- und Sozialsystems trennt die Transition der vormaligen DDR von den Transitionen der anderen, selbstbestimmten oder gar sich teilenden Staaten des Ostblocks. Für das Gesundheitssystem (GSS) gilt diese Sondersituation noch ein wenig mehr, besaß doch die DDR als einziger Staat des Ostblocks eine sozialversicherungsbasierte Finanzierung des GSS, die Eingliederung in das sozialversicherungsbasierte Finan-

³²⁵ hier: ein Krankenhaus ist mit seiner kompletten Biografie entweder Teil des GWAP-Datensatzes oder nicht.

³²⁶ Ob und unter welchen Bedingungen diesen Beobachtungen eine Systematik innewohnt, bliebe zu untersuchen.

³²⁷ Positiv für das Arbeiten: Fälle mit KDFV-Konvergenz trotz GWAP-Nichtkonvergenz lassen sich gleichfalls beobachten.

zierungssystem der Bundesrepublik ist ein eher evolutionärer und in kürzester Zeit absolvierter Schritt.

- ii. Im Bereich der Amtlichen Statistik Deutschlands werden flächendeckend keine Daten zu den Patienten (Fällen) oder Einrichtungen der ambulanten Versorgung erhoben, ganz im Unterschied zum stationären Bereich, wo mit der jährlichen Krankenhausstatistik neben jedem Fall die Ausstattung, die Leistungen und die Kosten jedes Krankenhauses berichtet werden.³²⁸ Berichtet wird also nicht über Gesundheit, sondern über die eher schwereren Krankheitsfälle sowie den Aufwand und die Kosten für eine Behandlung in diesen Einrichtungen. Eine Übersicht über die gesellschaftlichen Aufwendungen der Krankheitsbekämpfung insgesamt ist nicht möglich.
- iii. Die Amtliche Statistik hält Daten der Beschreibung von Wirtschafts- und Sozialleistungen auf Kreisniveau vor.³²⁹ Diese sind vielfach öffentlich zugänglich, sie können als Grundlage einer die allgemeine Wirtschaftslage und die Krankenhäuser in Beziehung setzenden Analyse auf Kreisniveau dienen.
- iv. Unaggregierte Daten der Krankenhausstatistik lassen sich nur unter vertraglichen Vereinbarungen für wissenschaftliche Analysen nutzen. Den Qualitätsberichten des StBA zufolge ist die Qualität sehr gut. Es zeigt sich jedoch, dass – bundeslandspezifisch – in vielen Fällen Inkonsistenzen bei der jahresübergreifenden Identifikation von Krankenhäusern wahrscheinlich sind. Medizinisch-epidemiologisch wäre zu bedenken, dass eine Hauptdiagnose viele Fälle nur unzureichend beschreiben kann, und wirtschaftlich, dass keine den Aufwand für den einzelnen Fall beschreibende Information vorhanden ist.³³⁰ Merkmalsdefinitionsveränderungen sind Teil einer ‚lebendigen‘ Statistik, den daraus erwachsenden jahresübergreifenden Kompatibilitätsproblemen wird (zu) wenig Aufmerksamkeit gewidmet. Als besonders problematisch für Wirtschaftlichkeitsanalysen müssen die Brutto-Netto-Brutto-Definitionsänderungen aller Kosten³³¹ angesehen werden, bei denen eine (longitudinale) Vergleichbarkeit ausdrücklich nicht gegeben ist.
- v. Aus den Daten der Krankenhausstatistik lässt sich ein Panel mit den Krankenhäusern als Subjekten erstellen, anhand dessen – wegen der sehr wahrscheinlichen Identifikations-Diskontinuitäten einzelner Krankenhäuser – Panel-Attri-

³²⁸ Vorsorge- und Rehabilitationseinrichtungen berichten ähnlich mit eingeschränktem Datenumfang.

³²⁹ Daten auf Betriebsniveau werden gleichfalls vorgehalten, hier jedoch nicht weiter betrachtet.

³³⁰ Die DRG-Statistik ist nicht Teil der Krankenhausstatistik.

³³¹ Ausnahme: bereinigte Kosten.

tion und Panel-Zuwachs überschätzt werden. Sehr wahrscheinlich ist auch, dass sich die Identifikationsprobleme über die Jahre nur auf die Krankenhäuser selbst und nicht auf den Kreis ihres Sitzes beziehen, ein durch Aggregation entstandenes Panel mit den Kreisen als Subjekten ist von diesen Identifikationsproblemen nicht betroffen. Die hier verwendeten Wirtschafts- und Sozialdaten können als bereits in Panelstruktur bereitgestellt verstanden werden, problematisch sind deren merkmalsweise Nichtverfügbarkeit für die Jahre bis 1995. Die Kreisstruktur (besonders) in Ostdeutschland ändert sich mehrfach, beide Panel müssen dem Rechnung tragen: Die Zuordnung der alten Kreise zu den neuen und die entsprechenden Aggregationen aller betroffenen Merkmale sind möglich, da die neuen Kreise zu überwiegendem Anteil allein durch Fusion der alten entstehen.

- vi. Die Kombination der Krankenhaus- mit den Wirtschaftsstatistiken ermöglicht eingeschränkten Einblick in die mikroökonomischen Interaktionen zwischen Gesundheit bzw. Gesundheitssystem und Wirtschaft. Statt dem GSS ist nur der Bereich der Krankenhäuser analysierbar. Gravierende gesetzliche Eingriffe in das GSS sind nicht in den Daten widerspiegelt. Die Anonymisierungsforderungen unter GWAP- bzw. KDFV-Bedingungen verbieten jedoch bestimmte Analysen.³³²
- vii. Das der Arbeit zugrunde liegende Panel umfasst den Zeitraum 1992–2005. Mit dem Beginn des Jahres 2005 gilt eine neue Definition der Arbeitslosigkeit, die Arbeitslosenquoten ändern sich. Das trifft alle Kreise gleichartig und kann (mit Vorbehalt) in den Analysen berücksichtigt werden. Mit dem Jahr 2006 ändert das Bundesamt für Bauwesen und Raumordnung seine Zuordnung eines Regionsgrundtyps für Trier und Umgebung in Rheinland-Pfalz, folglich ändern sich auch alle Kreistypen in jenem Gebiet. 2007 (Sachsen-Anhalt), 2008 (Sachsen), 2011 (Mecklenburg-Vorpommern) finden wieder Kreisreformen statt, Kreise werden dabei auch in Teilen in neue überführt. Eine Fortsetzung des Panels ist nicht möglich. Wohl denkbar hingegen ist der Versuch einer Neukompilation eines Panels in den neuen Kreisgrenzen unter Nutzung der beschriebenen Aggregationstechniken und Entscheidungen.
- viii. Von einer Analyse der Häufigkeit bestimmter Krankheiten und der Beziehungen zur wirtschaftlichen Lage mit den Daten der Krankenhausstatistik sollte eher Abstand genommen werden, denn es bestehen deutliche Zweifel an der

³³² Paneldaten-Modelle mit Krankenhäusern oder Kreisen als festen Effekten sind wegen der Mindestforderung der Zusammenfassung von je 3 Krankenhäusern nicht anwendbar.

Qualität der als Hauptdiagnose kodierten Information. Mit den medizinisch-epidemiologisch nicht erklärbaren Anteilsverschiebungen zwischen Kodierungen innerhalb bestimmter Diagnose-Gruppierungen ergibt sich nicht nur die Frage nach potenziellem *Upcoding* seit dem Beginn des Übergangs zum DRG-basierten Finanzierungssystem sondern auch die nach intendierten oder auf mangelndem Interesse basierenden Fehlkodierungen in den Jahren davor.

- ix. Für den Untersuchungszeitraum lassen sich mannigfaltige Unterschiede in den mikroökonomischen Wechselwirkungen der Krankenhäuser mit dem wirtschaftlichen und sozialen Umfeld finden, manche davon zeigen Ost-West-Unterschiede und Veränderungen über die Zeit. Vor allem aber zeigen die Analysen, dass die Unterschiede und Veränderungen der Leistungen und des Aufwands der Krankenhäuser kaum von den lokalen wirtschaftlichen Bedingungen beeinflusst sind, den Veränderungen des politischen Willens hin zu einem fallpauschalenbasierten Finanzierungssystem ist der größte Einfluss auf die Entwicklung der Krankenhauslandschaft in den Jahren seit der Wiedervereinigung zuzusprechen. Ost-West-Unterschiede spielen dabei eine untergeordnete Rolle.
- x. Gesetzlich forcierter Veränderungsdruck erzeugt Reaktionen bei den Betroffenen. Rückkopplung an die Politik bedarf des Messens dieser Reaktionen, die dabei verwendeten Maße (Kennzahlen) beeinflussen die Diskussionen erheblich. Der Auswahl aussagekräftiger Kennzahlen und ihrer adäquaten Interpretation sowie der Qualität der zugrunde liegenden Daten muss mehr Aufmerksamkeit geschenkt werden.

7.3. Methoden-Entwicklungen

Die Analysen dieses Projekt basieren auch auf mehreren originären Entwicklungen des Autors. Hervorzuheben sind dabei:

- ▷ die Ähnlichkeitspartitionierung zu vorgegebenen Klumpengrößen für die optimale Zusammenfassung von jeweils (nur) 3 Kreisen anhand des auf mehreren Wirtschaftsparametern basierenden Abstands für die Analysen am GWAP,

- ▷ die Quintil-Choropleth-Karte mit Ausschlussbedingung, deren Notwendigkeit durch die Anonymisierungsforderungen unter KDFV gegeben ist,
- ▷ das Kompositionsabweichungsdiagramm zur Darstellung sich über einer Kovariate verändernder Kompositionen.

Für die Datenaufbereitung durch das FDZ/L vor Übergabe der Daten an den Nutzer sind zwei SAS®-Skripte essentiell für vorliegendes Projekt: (a) ein Konvertierungsprogramm zur Wandlung von ICD-10-kodierten Diagnosen in ICD-9-kodierte und (b) ein AGS-Rekodierungsprogramm zur Korrektur von Kreiskodierungen und deren Transformation auf den Gebietsstand von 2005. Beide Skripte sind Ergebnis intensiver Schnittstellen- und Datenstrukturdiskussionen zwischen dem Autor und dem FDZ/L, Standort Kamenz³³³ und sichern wichtige Voraussetzungen für die Konsistenz der Projekt-Daten.

7.4. Ausblick

Die vorliegenden Analysen sind wiederholbar, und sie sind räumlich erweiterbar. Eine Einbindung der verbleibenden Altbundesländer könnte wegen zwischen 1992 und 2005 fast unveränderter Kreiseinteilungen mit deutlich verringertem Aufwand erfolgen, den Datenqualitätsproblemen³³⁴ wäre jedoch unvermindert Aufmerksamkeit zu widmen; außerdem ist mit dem jahresweisen Komplettdatenausfall eines weiteren Bundeslandes zu rechnen³³⁵.

Ökonometrische Deskriptionen und Modellierungen beruhen auf Annahmen, wie sich hier z. B. bei der Wahl der Gewichtungen zeigt. Veränderungen bezüglich dieser führen mit ein und demselben Datenmaterial zu numerisch veränderten Ergebnissen. Robustheitsuntersuchungen könnten die Bedeutung der unterschiedlichen Gewichtungen klären, eine modifizierte ökonomische Theorie dafür plädieren, dass Krankenhäuser nur über ihre relative Größe (quantifiziert z. B. durch

³³³ Dr. Schoffer war auch hier hilfreicher Partner und übernahm dankenswerterweise die Einbindung.

³³⁴ vermutlich auch hier mit bundeslandspezifischen Anteilen.

³³⁵ <http://www.forschungsdatenzentrum.de/bestand/krankenhaus/index.asp>, Zugriff am 21. 01. 2012.

ihre Kosten) sinnvoll miteinander verglichen werden können. Und nicht zuletzt können hier vorliegende Ergebnisse selbst anders interpretiert werden.

Für dieses Projekt steht die wirtschaftsbezogene Sichtweise im Vordergrund, somit ist der regelmäßige Rückgang der Aufnahmezahlen im Dezember ein Ressourcenauslastungsproblem. Soziologisch, mit der Sicht auf Befindlichkeit der Mitarbeiter, stellt dieser Rückgang vielleicht einen Vorteil dar, und medizinisch einen anderen; beide sind nicht Schwerpunkt dieser Arbeit, beide könnten Schwerpunkt einer anderen Arbeit mit denselben Daten sein.

Anhang A.

A.1. Merkmalsliste Krankenhausstatistik

Satzart / EF-Nr.	Merkmal	Bemerkungen	GWAP
	systemfreie Nr des KH		
	Zuordnungsmerkmal (z.B. BBR-Indikatoren)		
KHGEMU1	Bundesland		
EF2	Satzart [1-6]		
EF3	(je nach Satzart)		
	Satzart 2	030, 120, 150, 190, 220, 250, 260, 310, 350, 360, 390, 510, 540, 570, 630, 690, 770, 800, 820, 830, 870, 900, 930, 990	
	Satzart 3	030, 060, 070, 090, 120, 140, 150, 170, 190, 220, 250, 260, 270, 280, 310, 350, 360, 390, 420, 450, 480, 510, 530, 540, 570, 600, 630, 660, 690, 720, 750, 760, 770, 790, 800, 820, 830, 840, 870, 890, 900, 960, 970, 990, 991, 995, 999	
	Satzart 4	000, 001, 100, 300, 400, 500, 600, 700, 800, 900, 970, 990, 991, 999	
	Satzart 5	120, 150, 190, 220, 250, 260, 310, 350, 360, 390, 510, 540, 570, 630, 690, 693, 800, 820, 830, 870, 900, 930, 990	
1 / 4	KH-Typ		
1 / 5	Träger		
1 / 6-17	Anzahl (Summe) Ausbildungsplätze		
1 / 18	Arzneimittelversorgung		
1 / 19-29	Anzahl (Summe) med-techn Großgeräte		
1 / 44	Bettenkapazität (insgesamt)		<nein>
1 / 45-48	Anteile der Bettenkapazitäten nach ~ HSBauFörderungsgesetz ~ KHFinanzierungsgesetz ~ Vertragsbetten ~ sonst Betten	relative Anteile [auf Hunderstel genau, entsprechend ganzen Prozent]	
1 / 49	Ambul Operationen §115 SGB V		
1 / 50	~ von erm Ärzten	binär	
1 / 86	Entbindungen gesamt		
1 / 92	geborene Kinder insgesamt		
1 / 95	Art der Abrechnung		<nein>

Satzart / EF-Nr.	Merkmal	Bemerkungen	GWAP
2 / 4	Anzahl aufgestellte Betten		
2 / 5	~ Intensiv-/Notfallbetten		
2 / 6	~ Belegbetten		
2 / 7	Berechnungs- & Belegungstage		
2 / 8	~ Tage in Intensivbeh/Notfallüberw		
2 / 9	Patientenzugänge vollstationär		
2 / 10	~ aus anderen KH		
2 / 11	vollstationäre Verlegungen innerhalb KH		
2 / 12	teilstat in vollstat Verlegung in KH		
2 / 13	Entlassungen aus vollstat Beh		
2 / 14	~ Verlegung in andere KH		
2 / 15	~ in stat Reha-Einrichtung		
2 / 16	~ in Pflegeheime		
2 / 17	Verlegung innerh KH: vollstationär		
2 / 18	Verlegung innerh KH: vollstat nach teilstat		
2 / 19	durch Tod		
3 / 4	hauptamtl. Ärzte am 31.12. (Ohne Belegärzte)		
3 / 5	~ männl		
3 / 6	~ weibl		
3 / 7	~ männl Teilzeit- / geringfüg Beschäftigte		
3 / 8	~ weibl Teilzeit- / geringfüg Beschäftigte		
3 / 9	männl leitende Ärzte		geschlechts-
3 / 10	weibl leitende Ärzte		summiert
3 / 11	männl Oberärzte		geschlechts-
3 / 12	weibl Oberärzte		summiert
3 / 13	männl Assistenzärzte		geschlechts-
3 / 14	weibl Assistenzärzte		summiert
3 / 15	(nicht hauptamtl) Belegärzte am 31.12.		
3 / 16	(nicht hauptamtl) von Belegärzten angest Ärzte		
3 / 17	Vollzeitkräfte im Jahresdurchschnitt		
4 / 4	nichtärztl Personal		
4 / 5	~ männl		
4 / 6	~ weibl		
4 / 7	~ männl Teilzeit- / geringfüg Beschäftigte		
4 / 8	~ weibl Teilzeit- / geringfüg Beschäftigte		
4 / 9	Vollzeitkräfte im Jahresdurchschnitt		
5 / 4	vorstationäre Behandlungen	nur KH	
5 / 5	nachstationäre Behandlungen	-- " --	
5 / 6	Tages- und Nachtambulanzplätze	-- " --	
5 / 7	Entlassungen aus teilstat Behandlung	-- " --	
5 / 8	Teilstat Berechnungs-/Belegungstage	-- " --	
6 / 4	~ ärztl Dienst	nur KH	
6 / 5	~ Pflegedienst	-- " --	
6 / 6	~ med-techn Dienst	-- " --	
6 / 7	~ Funktionsdienst	-- " --	
6 / 8	~ klin Hauspersonal	-- " --	
6 / 9	~ Wirtsch & Versorgungsdienst	-- " --	
6 / 10	~ techn Dienst	-- " --	
6 / 11	~ Verwaltungsdienst	-- " --	
6 / 12	~ Sonderdienste	-- " --	
6 / 13	~ sonst Personal	-- " --	
6 / 14	~ nicht zurechenbare	-- " --	
6 / 15	Personalkosten insgesamt	-- " --	

Satzart / EF-Nr.	Merkmal	Bemerkungen	GWAP
6 / 16	~ Lebensmittel	-- "	<nein>
6 / 17	~ med Bedarf	-- "	
6 / 18	~~ Arzneimittel	-- "	
6 / 19	~~ Blut, Blutkons., -plasma	-- "	
6 / 20	~~ Verbandsmittel	-- "	
6 / 21	~~ Verbrauchsmat	-- "	
6 / 22	~~ Narkose & OP-Bedarf	-- "	
6 / 23	~~ Laborbed	-- "	
6 / 24	~~ Implantate	-- "	
6 / 25	~~ Transplantate	-- "	
6 / 26	~ Wasser, Energie, Brennstoffe	-- "	
6 / 27	~ Wirtschaftsbedarf	-- "	
6 / 28	~ wiederbeschaffte Verbrauchsgüter	-- "	<nein>
6 / 29	~ Verwaltungsbedarf	-- "	
6 / 30	~ zentr. Verwaltungsdienst	-- "	
6 / 31	~ zentr. Gemeinschaftsdienst	-- "	
6 / 32	~ pflegesatzfähige Instandhaltung	-- "	
6 / 33	~ sonstige Abgaben	-- "	
6 / 34	~ Versicherungen	-- "	<nein>
6 / 35	~ Sonstiges	-- "	
6 / 36	Sachkosten insgesamt	-- "	
6 / 37	Zinsen u.ä.	-- "	
6 / 38	~ Betriebsmittelkredite	-- "	<nein>
6 / 39	Steuern	-- "	
6 / 40	KH-Kosten insgesamt	-- "	
6 / 41	~ Personal	-- "	
6 / 42	~ Sachaufwand	-- "	<nein>
6 / 43	~ Aufwend aus Ausbildungsstättenumlage	-- "	
6 / 44	Kosten Ausbildungsstätten insgesamt	-- "	
6 / 45	Gesamtkosten	-- "	
6 / 46	~ Ambulanz	-- "	
6 / 47	~ wiss Fo & Lehre	-- "	
6 / 48	~ sonstige	-- "	<nein>
6 / 49	Abzüge insgesamt	-- "	
6 / 50	bereinigte Kosten	-- "	

EF-Nr.	Merkmal	Bemerkungen	GWAP
	systemfreie Nr des KH (wie bei Teilen I+III)		
	Zuordnungsmerkmal (z.B. BBR-Indikatoren)		
	Bundesland		
3	Geschlecht		
4	Geb.-Monat & -Jahr		<nein>
5	Zugangsdatum		monatsgenau
6	Abgangsdatum		<nein>
7	Abgangsart (Sterbefall)		
8	Hauptdiagnose	ICD-10 (ICD-9 Umkodierungen)	
11	Fachabt längster Verweildauer	Schlüsselveränderungen im Erfassungszeitraum	
12	Verweildauer		
13U2	Alter	monatsgenau (bis 99 J 11 M, darüber eine Klasse)	
14	Regionalangabe Pat.-Wohnort (5-stellig = Kreiszuordnung, Ausland)		

Abb. A.1.: Merkmalsliste Krankenhausstatistik zum Projekt FDZ-Nutzung 606-2007 (Forschungsdatenzentrum der Statistischen Landesämter, 2007)

A.2. Programmbausteine

Das Mantelskript (Abschn. 5.2.2) in seinen zwei Varianten unterscheidet sich nur an wenigen Stellen. Beide Versionen bündeln vor allem die für die Arbeit am Gastwissenschaftler-Arbeitsplatz bzw. die unter den Bedingungen der Kontrollierten Datenfernverarbeitung notwendigen Voreinstellungen für das Statistiksystem SAS®.

Kontrollierte Datenfernverarbeitung (KDFV)

```

1  %*** 2009.04.29 created by UHF
   Start aller Auswertungen im batch — KDFV ***;
   %*** 2011.03.29 modified by UHF ***;
6  OPTIONS LS=110 PS=100 NOCENTER NODATE NUMBER PAGENO=1 NOTES BYLINE FIRSTOBS=1 OBS=max
   REPLACE FMTERR STIMER ERRORS=10 MSGLEVEL=i NOSOURCE NOSOURCE2 NOSYMBOLGEN NOMPRINT NOMLOGIC
   LOCALE=en_US;
   ODS RESULTS=off;
   ODS OUTPUT close;
11  ODS LISTING SELECT all;

   %*****;
   %*
   %* Overall Start for KDFV — HF ULBRICH
   %*
16  %*****;

   %GLOBAL stdprnto adt adt_run atm atm_run atm_dif adtm
   uhf_prj uhf_log uhf_lst uhf_mac uhf_sas cmprss odsstd optbefore
21  prsp prspf ostmin psel1 rkpt ttx jrgrko jrdiag selcnd rpl seed1 icd9ks kherrfmt;

   %*** Format für KH in Kontrollprints in logs ***;
   %LET kherrfmt=hodef.;

26  %LET stdprnto= ;
   %*** Kompression (großer, fallbezogener) data sets? ***;
   %LET cmprss=binary;
   %*** Standard-ODS-Auswahl ***;
   %LET odsstd=%str(SELECT all; ODS EXCLUDE EngineHost (PERSIST));
31  ODS &odsstd.;

   %*** Perspektive prsp ***;
   %LET prsp=k;
   %LET prspf=hex2.;
36  %LET ostmin=091x;
   %LET psel1=WHERE=(g&prsp. in (081x 0C1x 0FFx));
   %LET rkpt=typ;

41  %LET adt_run=%sysfunc(date(),ddmmyyp10.);
   %LET atm_run=%sysfunc(time(),time8.);
   %LET atm_dif=%sysfunc(datetime(),13.2);

   %LET uhf_prj
46  =d:\data\231xx Krankenhausstatistik\Nutzungen\606—2007 Uni Potsdam—Strohe\KDFV\Analyse\;

   %LET uhf_sas=&uhf_prj.sas\;
   %LET uhf_mac=&uhf_prj.mac\;

```

```

%LET uhf_log=&uhf_prj.log\;
%LET uhf_lst=&uhf_prj.lst\;
51 %PUT;
%PUT %str() SAS _start_K run startete am &adt_run. um &atm_run. Uhr;
%PUT %nrstr( mit aktueller Auswertungs-Perspektive (&PRSP)) # &prsp..;
%PUT;
56 %*** 'global' macros ***;
%INC "&uhf_sas._start_K.0as";

%LET ttx=_start_K.sas: Beginn (Perspektive %upcase(&prsp.));
TITLES " &adt_run., &atm_run. Uhr.";
61 TITLE6 " ";
%titles;
%titlog;
%PUT _automatic_;
PROC DATASETS LIBRARY=work DETAILS;
66 QUIT;

%stdprnto;
PROC PRINTTO NEW &stdprnto.; RUN;
%PUT;
71 %PUT %str() SAS _start_K run startete am &adt_run. um &atm_run. Uhr;
%PUT %nrstr( mit aktueller Auswertungs-Perspektive (&PRSP)) &prsp..;
%PUT;

%LET adtm=%sysfunc(datetime(),datetime19.);
76 %LET adtm=%substr(&adtm.,1,2).%substr(&adtm.,3,3).%substr(&adtm.,6);

OPTIONS FORMCHAR="|----|+|----+|=|/|<>*";

LIBNAME dat "&uhf_prj.dat";
LIBNAME ori "&uhf_prj.ori";
81 LIBNAME exd "&uhf_prj.exd";
LIBNAME grc "&uhf_prj.grc";

%*** seeds — somewhere needed! ***;
%LET seed1=1774577; %*** dataprep.sas ***;
86 %PUT; %PUT %str() seed1: &seed1.; %PUT;

%postpo(ppd=1.1);

91 PROC CONTENTS VARNUM DATA=sashelp.vlibnam;
RUN;
PROC PRINT WIDTH=minimum UNIFORM OBS=".#" DATA=sashelp.vlibnam;
BY libname NOTSORTED;
ID libname;
96 VAR level readonly path;
RUN;

%*** ACHTUNG: (sicherheitshalber) nicht jr1996—jr2004, denn sonst
Ordnung nach TRANSPOSE (bei nichtbesetzten Variablen) nicht garantiert! ***;
101 %LET jrdiag=jr1993 jr1994 jr1995 jr1996 jr1997 jr1998 jr1999;
%LET jrdiag=&jrdiag. jr2000 jr2001 jr2002 jr2003 jr2004 jr2005;
%LET jrgrko=jr1992 &jrdiag.;

%****~ Einlese- & Korrektur-Prozeduren ~~~~~~*
106 %*~%INC "&uhf_sas._start_K.1as";

%****~ Format-Aufbereitung & Jahresbezug-Fixierung ~~~~~~*

111 %*** UNBEDINGT ausführen! ***; %INC "&uhf_sas._start_K.2as";

%****~ Daten-Überprüfungen ~~~~~~*

%*~%INC "&uhf_sas._start_K.3as";
116 %****~ Template-Einbindung für procs PANEL & GLIMMIX ~~~~~~*

```

```

121 %*** tmlinKM.sas — GLIMMIX & PANEL templates ***;
%LET ttx=tmlinKM.sas: Einbindung/Aktualisierung der GLIMMIX & PANEL templates;
%titles;
%titlog;
%doit (sfn=tmlinKM);

126 %*~*~ Auswertungsschritt—Prozeduren ~~~~~*;
%*** Auswahl und Reihenfolge der ICD—9—Gruppierungen für Auswertungen ***;
%LET icd9ks=%sysfunc(compbl
    01(001<=icd9e<=009) 01(020<=icd9e<=027) 01(090<=icd9e<=099) 02(140<=icd9e<=208) 03 04
131 06(320<=icd9e<=326) 06(360<=icd9e<=379) 06(380<=icd9e<=389) 07 08 09(520<=icd9e<=543)
    10(580<=icd9e<=589) 11(650<=icd9e<=669) 12 17(800<=icd9e<=829) 17(870<=icd9e<=897));

%*** evalua_3.sas — KH—Grunddaten ärztl Personal ***;
%LET ttx=evalua_3.sas: Daten ärztliches Personal (Perspektive %upcase(&prsp.));
%titles;
136 %titlog;
%*—%doit (sfn=evalua_3);

%*** evalua_4.sas — KH—Grunddaten nicht—ärztl Personal ***;
%LET ttx=evalua_4.sas: Daten nicht—ärztliches Personal (Perspektive %upcase(&prsp.));
141 %titles;
%titlog;
%*—%doit (sfn=evalua_4);

%*** evalua_0.sas — Krankenhäuser ***;
146 %LET ttx=evalua_0.sas: Krankenhausdaten (Perspektive %upcase(&prsp.));
%titles;
%titlog;
%doit (sfn=evalua_0);

151 %*** evalua_7.sas — (geopolitisch) kumulierte Krankenhäuser ***;
%LET ttx=evalua_7.sas: Krankenhausversorgungsleistungen (Perspektive %upcase(&prsp.));
%titles;
%titlog;
%*—%doit (sfn=evalua_7);

156 %*** evalua_6.sas — Fallstatistiken, Deskription — Alters—Vg ***;
%LET ttx=evalua_6.sas: Fallstatistiken, Deskription — Alters—Vg (Perspektive %upcase(&prsp.));
%titles;
%titlog;
161 %*—%*—%doit (sfn=evalua_6);

%*** evalua_2.sas — KH—Grunddaten Fachabteilungen ***;
%LET ttx=evalua_2.sas: Fachabteilungsdaten (Perspektive %upcase(&prsp.));
%titles;
166 %titlog;
%*—%doit (sfn=evalua_2);

%*** evalua_1.sas — ICD—Kodierungen ***;
%LET ttx=evalua_1.sas: Diagnosen (Perspektive %upcase(&prsp.));
171 %titles;
%titlog;
%*—%doit (sfn=evalua_1);

%*** evalua_5.sas — Verweildauern ***;
176 %LET ttx=evalua_5.sas: Verweildauern (Perspektive %upcase(&prsp.));
%titles;
%titlog;
%*—%doit (sfn=evalua_5);

181 %*~*~ Log—File—Prüfung & Abschluss ~~~~~*;

%PUT _automatic_;
%postpo(ppd=1.1);

186 ODS SELECT EngineHost;

%*** chklog — Scannen der log—files nach Auffälligkeiten ***;

```

```

%LET ttx=chklogKM.sas: logfiles seit &adt_run. &atm_run.;
%titles;
%titlog;
191 %doit (sfn=chklogKM);

OPTIONS FMTSEARCH=() PAGENO=2147483647;
PROC DATASETS LIBRARY=work DETAILS NOLIST;
196 DELETE tp: / MEMTYPE=data; RUN;
DELETE tp: / MEMTYPE=catalog; RUN;
QUIT;
%LET adt_run=%sysfunc(date(),ddmmyyp10.);
%LET atm_run=%sysfunc(time(),time8.);
201 %LET atm_dif=%sysfunc(putn(%sysfunc(datetime(),13.2)—&atm_dif.,time13.2));
%LET ttx=_start_K.sas (Perspektive %upcase
(&prsp.), nach chklogKM): &adt_run., &atm_run. Uhr, Laufzeit: &atm_dif. h.);

%titles;
206 TITLES " ~ Überblick: library ori";
TITLE6 " ";
%titles;
%titlog;
%libr2l (lib=ori,dest=b);
211 TITLES " ~ Überblick: library dat";
TITLE6 " ";
%titles;
%titlog;
%libr2l (lib=dat,dest=b);
216 TITLES " ~ Überblick: library grc";
TITLE6 " ";
%titles;
%titlog;
%libr2l (lib=grc,dest=b);
221 %titles;
TITLES " ~ Aufräumen: library work, data sets — Überblick";
TITLE6 " ";
%titles;
%titlog;
226 ODS SELECT Directory Members;
PROC DATASETS LIBRARY=work NOLIST;
CONTENTS VARNUM DATA=work._all_; RUN;
QUIT;
231 TITLES " ~ Aufräumen: library work, KILL data sets ~ Überblick";
TITLE6 " ";
%titles;
%titlog;
ODS SELECT Directory Members;
236 PROC DATASETS LIBRARY=work MEMTYPE=data KILL;
CONTENTS VARNUM DATA=work._all_; RUN;
QUIT;
TITLES " ~ Aufräumen: catalog sasmacr";
TITLE6 " ";
241 %titlog;
PROC CATALOG CATALOG=work.sasmacr;
DELETE kxanzl parseL dsappe
titles titlog libr2l cont2l catg2l dsnprn chkpnt postpo extvrs
prcmdl marker newlst efidno mdlchk mdlprn catcnt stdprnto doit / ENTRYTYPE=macro; RUN;
246 CONTENTS; RUN;
QUIT;

%*—LIBNAME dat CLEAR;
%*—LIBNAME ori CLEAR;
251 %*—LIBNAME exd CLEAR;
%*—LIBNAME grc CLEAR;

%PUT;
%PUT %str() SAS _start_K run endete nach &atm_dif. h Laufzeit am &adt_run. um &atm_run. Uhr.;
256 %PUT;

ODS RESULTS=on;

```

```

PROC PRINTTO PRINT=print; RUN;
TITLE; FOOTNOTE; PROC PRINTTO; RUN;
261 %PUT;
%PUT %str() SAS _start_K run endete am &adt_run. um &atm_run. Uhr — nach &atm_dif. h.;
%PUT;
%SYMDEL stdprnto adt adt_run atm atm_run atm_dif adtm
uhf_prj uhf_log uhf_lst uhf_mac uhf_sas cmprss odsstd optbefore
266 prsp prspf ostmin psel1 rkpt ttx jrgrko jrdiag selcnd rpl seed1 cmprss kherrfmt / NOWARN;
OPTIONS LS=110 PS=512 NOCENTER DATE NUMBER PAGENO=1 NOTES BYLINE FIRSTOBS=1 OBS=max
REPLACE FMTERR STIMER ERRORS=10 MSGLEVEL=i NOSOURCE NOSOURCE2 NOSYMBOLGEN NOMPRINT NOMLOGIC;
%*** End: _START_K.SAS *****;

```

In den Zeilen 6 bis 11 werden wesentliche Merkmale der SAS®-Arbeitsumgebung gesetzt, `LOCALE=en_US` sichert dabei, dass SAS in seiner englischsprachigen Originalversion arbeitet. Unverständene Fehlermeldungen lassen sich im Original deutlich besser mit SAS®-Dokumenten oder Einträgen im Internet vergleichen.

Zeile 24 bereitet vor, dass auch im Fall von Fehlermeldungen in der Logdatei eine Ausgabe der Krankenhausnummer unterdrückt wird, `hidef.` ist ein vom Autor definiertes SAS®-Format, das für alle gültigen numerischen Werte konstant eine Zeichenkette (`-*`) ausgibt, Fehlwerte bleiben sichtbar.

`cmprss=binary` in Zeile 28 ermöglicht im Weiteren, große Daten-Dateien komprimiert zwischen- und endabzuspeichern. Unter den Gegebenheiten am Standort Kamenz des FDZ/L erweisen sich die Zeitaufwände für Komprimierung und Dekomprimierung der Diagnosedaten als unvergleichlich viel geringer als der Zugriff auf unkomprimierte Versionen der Daten-Dateien.

Der Block `%*** Perspektive prsp ***;` (Zeilen 33–38) fixiert spezifische Anforderungen der Arbeit unter KDFV (`prsp=k`). Für eine sofort sichtbare Trennung der Ergebnisse der Arbeit am GWAP von denen unter KDFV wird den Klumpen bzw. Kreisen ein jeweils anderer Nummernbereich zugeordnet: den 50 Klumpen am GWAP die Nummern 1 bis 50, den 147 Kreisen unter KDFV hingegen die Nummern 109 bis 255. Die hexadezimale Darstellung erfordert nur 2 Stellen (01–32 bzw. 6D–FF), das erweist sich bei Datenlisten als vorteilhaft. Kreisen bzw. Klumpen aus Rheinland-Pfalz werden jeweils die kleineren Nummern zugeordnet, `ostmin` fixiert die kleinste der Nummern für Ostdeutschland.

In den Zeilen 44–49 wird das Arbeitsverzeichnis deklariert (für die KDFV vom FDZ/L vorgegeben). Unterverzeichnisse ermöglichen die Trennung von spezifisch für diese Arbeit entwickelten SAS®-Steuerungen und -Makros, andernorts entwi-

ckelten und geprüften Makros (Abschn. 5.2.4), der Logdateien und der Ergebniszusammenstellungen (als SAS®-Listings).³³⁶

Der Logdatei-Eintrag der Startzeit erfolgt zweimal: auf den Bildschirm (Zeilen 51–54) und – nach Eröffnung der Logdatei des Rechenlaufes über das Makro %stdprnto (Zeilen 67 und 68) – in die Logdatei³³⁷. Für den entsprechenden Endzeit-Eintrag nebst Angabe der Dauer des Rechenlaufes erfolgt dies in umgekehrter Reihenfolge in den Zeilen 252–254 bzw. 260–262.

Das (potenziell einmalige) Einlesen der Daten, das Generieren von nutzerdefinierten SAS®-Formaten sowie die Suche nach Fehlern in den Daten wird in den Zeilen 104–114 durch das Einbinden von KDFV-spezifischen Subskripts ermöglicht.

Da die SAS-Prozeduren `proc panel` und `proc glimmix` Ergänzungen zur Grundinstallation von SAS/STAT® Version 9.1.3 sind, werden sie von der normalen Softwarewartung am Standort Kamenz des FDZ/L nicht sicher erfasst. Diesem Umstand wird mit jeweiliger Neueinbindung (Zeilen 118–122) der wichtigsten Teile dieser Prozeduren mittel eines vom Autor entwickelten Einzelskripts begegnet.³³⁸

Das Makro %doit organisiert für dabei, dass für das Einzelskript `tmplinKM (.sas)` eine Logdatei – mit Zeitstempel im Dateinamen – und Ergebnisdateien eröffnet und am Ende wieder geschlossen werden. Dieses Makro kommt bei jedem Einzelskript der Auswertung gleichartig zur Anwendung, ein Transparenzgewinn sowohl für den Autor als auch den Betreuenden am FDZ/L. In den Zeilen 124–178 ist die Einbindung der thematischen Einzelskripte (Abschn. 5.2.3) vorgesehen, ein jedes kann dabei separat (hier mit %*-- am Anfang einer Zeile) ausgeblendet werden³³⁹.

³³⁶ Die Bitte des FDZ/L nach zusätzlicher Generierung aller Ergebnisse als rtf-Dateien wird mit einem weiteren Unterverzeichnis `.\rtf` unterstützt.

³³⁷ z. B. `20100203_160756__start_K.log` für den Rechenlauf begonnen am 03.02.2010, 16:07:56 Uhr.

³³⁸ Der Rechenzeitbedarf für dieses Einzelskript ist im Vergleich zur damit gewonnenen Sicherheit der Berechnungen verschwindend gering.

³³⁹ z. B. ist `evalua_3 (.sas)` in Zeile 136 inaktiv.

Als letzte Schritte eines jeden Rechenlaufes erfolgen eine ‚Reinigung‘ von temporären Dateien³⁴⁰ und ein Überblick über die im Rechenlauf generierten Ergebnisdateien.

Gastwissenschaftlerarbeitsplatz (GWAP)

```

5  %*** 2008.01.18 created by UHF
    Start aller Auswertungen im batch — GWAP ***;
6  %*** 2010.08.18 modified by UHF ***;

7  OPTIONS LS=110 PS=100 NOCENTER NODATE NUMBER PAGENO=1 NOTES BYLINE FIRSTOBS=1 OBS=max
    REPLACE FMTERR STIMER ERRORS=10 MSGLEVEL=i NOSOURCE NOSOURCE2 NOSYMBOLGEN NOMPRINT NOMLOGIC
    LOCALE=en_US;
8  ODS RESULTS=off;
9  ODS OUTPUT close;
10 ODS LISTING SELECT all;

11 %*****;
12 %*
13 %* Overall Start for GWAP — HF ULBRICH
14 %*
15 %*
16 %*****;
17
18 :
19 :
20 :

21 %*** Perspektive prsp ***;
22 %LET prsp=1;
23 %LET prspf=z2.;
24 %LET ostmin=13;
25 %LET psel1=WHERE=(g&prsp. in (1 12 13 47 50));
26 %LET rkpt=ptt;

27
28 %LET adt_run=%sysfunc(date(), ddmmyyp10.);
29 %LET atm_run=%sysfunc(time(), time8.);
30 %LET atm_dif=%sysfunc(datetime(), 13.2);

31
32 %LET uhf_prj=h:\gwap\;
33
34 %LET uhf_sas=&uhf_prj.sas\;
35 %LET uhf_mac=&uhf_prj.mac\;
36 %LET uhf_log=&uhf_prj.log\;
37 %LET uhf_lst=&uhf_prj.lst\;
38
39 %PUT;
40 %PUT %str() SAS _start_G run startete am &adt_run. um &atm_run. Uhr;
41 %PUT %nrstr( mit aktueller Auswertungs-Perspektive (&PRSP)) # &prsp..;
42 %PUT;
43 %*** 'global' macros ***;
44 %INC "&uhf_sas._start_G.0as";

45
46 %LET ttx=_start_G.sas: Beginn (Perspektive %upcase(&prsp.));
47 TITLES " _&adt_run., &atm_run. Uhr.";
48 TITLE5 " ";
49 %titles;
50 %titlog;
51 %PUT _automatic_;

```

³⁴⁰ Für einen unmittelbaren erneuten Rechenlauf – das übliche Vorgehen am GWAP – wird die Interaktion mit vorherigen Teilergebnissen verhindert.


```
PROC DATASETS LIBRARY=work DETAILS;
55 QUIT;

%stdprnto;
PROC PRINTTO NEW &stdprnto.; RUN;
%PUT;
60 %PUT %str( SAS _start_G run startete am &adt_run. um &atm_run. Uhr;
%PUT %nrstr( mit aktueller Auswertungs-Perspektive (&PRSP)) &prsp.;
%PUT;

%LET adtm=%sysfunc(datetime(),datetime19.);
65 %LET adtm=%substr(&adtm.,1,2)%substr(&adtm.,3,3)%substr(&adtm.,6);

OPTIONS FORMCHAR="|-----|+|---+|=|/|<|>*";

LIBNAME dat "&uhf_prj.dat";
70 LIBNAME ori "&uhf_prj.ori";
LIBNAME exd "&uhf_prj.exd";
LIBNAME grc "&uhf_prj.grc";
%*-%-LIBNAME library "&uhf_prj.ins";
%*-%-LIBNAME gdevice0 "&uhf_prj.ins";
75 %*-%-LIBNAME gfont0 "&uhf_prj.ins";

%*** seeds — somewhere needed! ***;
%LET seed1=1774577; %*** dataprep.sas ***;
%PUT; %PUT %str( seed1: &seed1.; %PUT;
80 %postpo(ppd=1.1);

%*—PROC OPTIONS HOST; RUN;
PROC CONTENTS VARNUM DATA=sashelp.vlibnam;
85 RUN;
PROC PRINT WIDTH=minimum UNIFORM OBS=".#" DATA=sashelp.vlibnam;
BY libname NOTSORTED;
ID libname;
VAR level readonly path;
90 RUN;

%*** ACHTUNG: (sicherheitshalber) nicht jr1996—jr2004, denn sonst
Ordnung nach TRANSPOSE (bei nichtbesetzten Variablen) nicht garantiert! ***;
%LET jrdiag=jr1996 jr2000 jr2004;
95 %LET jrgrko=jr1992 &jrdiag.;

%*~~~ Einlese- & Korrektur-Prozeduren ~~~~~*;
%*-%INC "&uhf_sas._start_G.1as";
100 %*~~~ Format-Aufbereitung & Jahresbezug-Fixierung ~~~~~*;
%*** UNBEDINGT ausführen! ***; %INC "&uhf_sas._start_G.2as";

%*~~~ Daten-Überprüfungen ~~~~~*;
%*-%INC "&uhf_sas._start_G.3as";

%*~~~ Template-Einbindung für procs PANEL & GLIMMIX ~~~~~*;
110 :
:
:

%PUT;
%PUT %str( SAS _start_G run endete nach &atm_dif. h Laufzeit am &adt_run. um &atm_run. Uhr.;
%PUT;

ODS RESULTS=on;
120 PROC PRINTTO PRINT=print; RUN;
TITLE; FOOTNOTE; PROC PRINTTO; RUN;
%PUT;
%PUT %str( SAS _start_G run endete am &adt_run. um &atm_run. Uhr — nach &atm_dif. h.;
```

```

125 %PUT;
%SYMDEL stdprnto adt adt_run atm atm_run atm_dif adtm
      uhf_prj uhf_log uhf_lst uhf_mac uhf_sas cmprss odsstd optbefore
      prsp prspf ostmin psel1 rkpt ttj jrgrko jrdiag selcnd rpl seed1 cmprss kherrfmt / NOWARN;
130 OPTIONS LS=110 PS=512 NOCENTER DATE NUMBER PAGENO=1 NOTES BYLINE FIRSTOBS=1 OBS=max
      REPLACE FMTERR STIMER ERRORS=10 MSGLEVEL=i NOSOURCE NOSOURCE2 NOSYMBOLGEN NOMPRINT NOMLOGIC;
%*** End: _START_G.SAS *****;

```

Die (gekürzte) Darstellung des Mantelskripts für die Arbeit am GWAP zeigt vor allem die Unterschiede³⁴¹ zum Mantelskript für die KDFV, Auslassungen identischer Passagen sind dabei durch aufeinander folgende Zeilen mit Doppelpunkten gekennzeichnet.

³⁴¹ `prsp=1` (Zeile 24) charakterisiert – aus historischen Gründen – die Version für den GWAP; `ostmin`, die kleinste Nummer eines Klumpens in Ostdeutschland, ist hier 13 (Zeile 26). Das GWAP-spezifische Einlesen der Daten, das Generieren von nutzerdefinierten SAS®-Formaten sowie die Suche nach Fehlern in den Daten wird in den Zeilen 97–107 ermöglicht.

A.3. Tabellen

Ergänzungen zu Abschn. 6.2.6: Nichtärztliches Personal:

Tab. A.1.: Modellierung der nichtärztlichen Krankenhaus-Mitarbeiter je 1000 Einwohner, basierend auf den Daten 1992–2005

Gebiet	Geschlecht	Jahr	Prädiktion ^x	95 % KI ^y	Kehrwert ^z
RP	männlich	1991	1,54	[1,22, 1,94]	650
		1995	1,56	[1,25, 1,95]	640
		2000	1,60	[1,29, 1,97]	627
		2005	1,63	[1,32, 2,02]	614
	weiblich	1991	7,50	[5,94, 9,48]	133
		1995	7,54	[6,05, 9,40]	133
		2000	7,59	[6,15, 9,38]	132
		2005	7,65	[6,17, 9,47]	131
Ostdt	männlich	1991	0,93	[0,81, 1,06]	1079
		1995	0,94	[0,83, 1,07]	1062
		2000	0,96	[0,85, 1,08]	1042
		2005	0,98	[0,87, 1,10]	1022
	weiblich	1991	6,97	[6,11, 7,95]	144
		1995	6,75	[5,96, 7,64]	148
		2000	6,49	[5,76, 7,30]	154
		2005	6,23	[5,53, 7,03]	160

^x nichtärztliche Krankenhaus-Mitarbeiter je 1000 Einwohner

^y 95 % Konfidenzintervall zur Prädiktion

^z Einwohner je nichtärztlichem Krankenhaus-Mitarbeiter – der reziproke Wert der Prädiktion

Tab. A.2.: Modellierung des Anteils der weiblichen an allen nichtärztlichen Krankenhaus-Mitarbeitern, basierend auf den Daten 1992–2005

Gebiet	Jahr	Prädiktion ^x	95 % KI ^y
RP	1991	83,3	[81,8, 84,7]
	1995	83,2	[81,9, 84,3]
	2000	82,9	[81,8, 84,0]
	2005	82,7	[81,4, 84,0]
Ostdt	1991	87,9	[87,2, 88,5]
	1995	87,3	[86,8, 87,9]
	2000	86,6	[86,1, 87,1]
	2005	85,8	[85,2, 86,4]

^x Anteil [%]

^y 95 % Konfidenzintervall zur Prädiktion

Ergänzungen zu Abschn. 6.2.8: Bereinigte Kosten der Krankenhäuser – Modell 3:

Tab. A.3.: Krankenhäuser 1996–2005, Kostenmodellierung #3
Parameterschätzungen (3.904 Beobachtungseinheiten)

Effekt (Stufen)	Schätzung	StdErr ^w	FG ^x	F	Wk ^y
ostdt=West	0,02152	0,01969	1, 2194	0,80	0,3697
jahr	0,005920	0,001700	1, 1657	21,96	<0,0001
jahr·(ostdt=West)	-0,00238	0,001544	1, 1566	2,38	0,1228
ktyp=Universitäts-Klinik	-0,00698	0,04223	3, 1973	19,06	<0,0001
ktyp=Plan-Krankenhaus	0,1755	0,01668			
ktyp=KH mit Vers.-Vertrag	0,1732	0,01862			
trae=öff.	0,1023	0,04290	2, 2085	0,70	0,4950
trae=freig.	-0,03469	0,02856			
(ktyp=PlanKH)·(trae=öff.)	-0,1306	0,04282	3, 2029	3,24	0,0214
(ktyp=PlanKH)·(trae=freig.)	0,01540	0,02905			
(ktyp=VersV)·(trae=öff.)	-0,1248	0,05279			
fallzL	-0,8233	0,007945	1, 2154	10737	<0,0001
arztzL	0,1897	0,01303	1, 2342	211,9	<0,0001
naerzL	0,6355	0,01423	1, 2283	1993	<0,0001
krstyp=1	0,03418	0,01709	8, 1799	0,70	0,6961
krstyp=2	-0,00081	0,02362			
krstyp=3	-0,00584	0,01168			
krstyp=4	-0,00371	0,009354			
krstyp=5	0,02152	0,01586			
krstyp=6	-0,00594	0,01066			
krstyp=7	-0,01094	0,008928			
krstyp=8	-0,00597	0,01060			
(ostdt=West)·(krstyp=1)	-0,03604	0,02891	7, 1684	1,15	0,3267
(ostdt=West)·(krstyp=2)	-0,00433	0,03313			
(ostdt=West)·(krstyp=3)	-0,01927	0,02147			
(ostdt=West)·(krstyp=5)	-0,03924	0,02296			
(ostdt=West)·(krstyp=6)	-0,00182	0,01837			
(ostdt=West)·(krstyp=7)	0,006333	0,02134			
(ostdt=West)·(krstyp=8)	0,006865	0,02345			
jahr·(krstyp=1)	-0,00391	0,002101	8, 1532	1,00	0,4307
jahr·(krstyp=2)	-0,00224	0,004645			
jahr·(krstyp=3)	-0,00062	0,002587			
jahr·(krstyp=4)	-0,00019	0,002359			
jahr·(krstyp=5)	-0,00149	0,002661			
jahr·(krstyp=6)	0,000314	0,002149			
jahr·(krstyp=7)	0,000874	0,002207			
jahr·(krstyp=8)	-0,00044	0,002478			
ewdL	0,01343	0,009946	1, 1729	1,82	0,1772
betqL	-0,06450	0,05532	1, 2038	1,36	0,2438
bipewL	0,04927	0,05477	1, 2264	0,81	0,3684
alqlgt	-0,00401	0,008793	1, 2184	0,21	0,6483

^w Standardfehler der Schätzung^x Zähler-, Nenner-Freiheitsgrade der Wald-F-Statistik^y Ablehnungswahrscheinlichkeit der Wald-F-Statistik für diesen Effekt^z Kodierung gemäß Kreistypen des BBR (3. Spalte in Abb. 3.3)

A.4. Graphen

A.4.1. Altersverteilungen

Ergänzungen zu Abschn. 6.1.4: Altersverteilungen³⁴²:

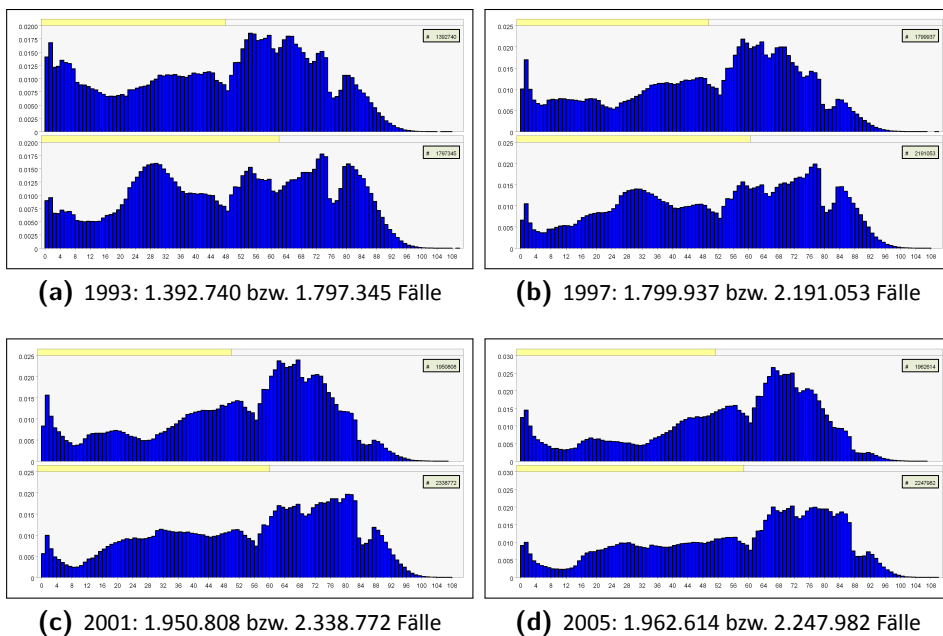
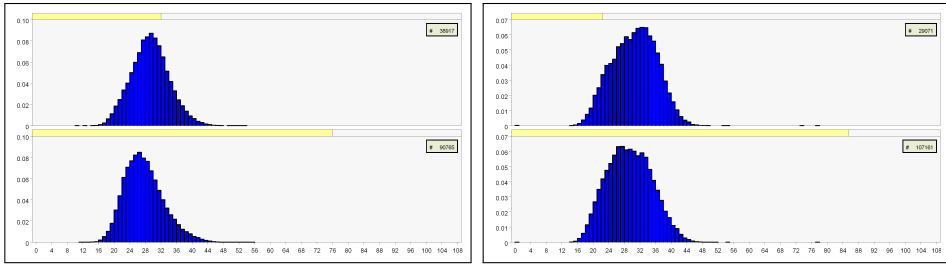


Abb. A.2.: Empirische Geschlechts- und Altersverteilung [Anteile] mit Fallzahl, ICD-9-Kapitel I–XVII: männlich (oberes) und weiblich (unteres Teilpanel)

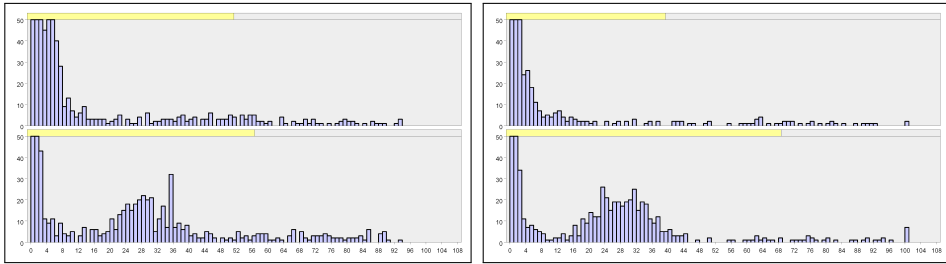
³⁴² Für 1993 fehlen jeweils die Daten aus Mecklenburg-Vorpommern.



(a) 1993: 38.917 bzw. 90.765 Fälle

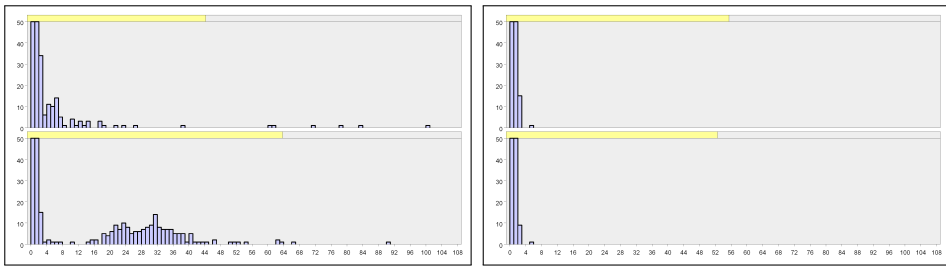
(b) 2003: 29.071 bzw. 107.161 Fälle

Abb. A.3.: Entbindungen (650–669, aus ICD-9-Kapitel XI):
Empirische Altersverteilung der Gebärenden [Anteile] mit Fallzahl,
Rheinland-Pfalz (oberes Teilpanel) und Ostdeutschland (unteres Teilpanel)



(a) 1993

(b) 2000



(c) 2002

(d) 2005

Abb. A.4.: ICD-9-Kapitel XV: bestimmte Affekte mit Ursprung in der Perinatalzeit:
Empirische Altersverteilung [Anzahlen, getrimmt bei 50 Fällen],
männlich (oberes Teilpanel) und weiblich (unteres Teilpanel)

A.4.2. Monatliche Aufnahmezahlen vollstationärer und teilstationärer Fälle

Ergänzungen zu Abschn. 6.2.2 und Abb. 6.14:

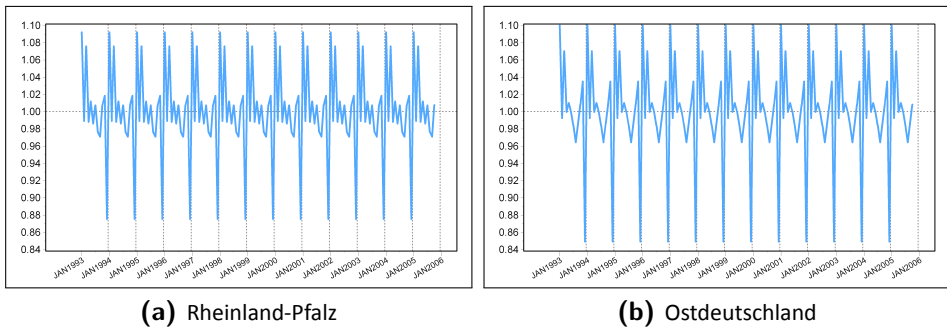


Abb. A.5.: Saisonkomponenten [Anteile] monatlicher Aufnahmezahlen
Januar 1993 – Oktober 2005, multiplikatives Modell
Referenzlinien an den Jahreswechsell

Literatur

- Aitchison, J. (1982). The Statistical Analysis of Compositional Data (with discussion). *Journal of the Royal Statistical Society B (Statistical Methodology)* 44, 139–177.
- Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman & Hall.
- Aitchison, J. (2003). *The Statistical Analysis of Compositional Data (Reprint with additional material)*. Caldwell, NJ: The Blackburn Press.
- Aitchison, J. und J. J. Egozcue (2005). Compositional Data Analysis: Where Are We and Where Should We Be Heading. *Mathematical Geology* 37, 829–850.
- Albrecht, T. und N. Klazinga (2002). Health Manpower Planning in Slovenia: A Policy Analysis of the Changes in Roles of Stakeholders and Methodologies. *Journal of Health Politics, Policy & Law* 27, 1001–1022.
- Albrecht, T., E. Turk, M. Toth, J. Ceglar, S. Marn, R. Pribakovic Brinovec und M. Säfer (2009). *Health Systems in Transition: Slovenia*. Kopenhagen: World Health Organization, on behalf of the European Observatory on Health Systems and Policies.
- Allison, P. D. (2005). *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Institute.
- Amend, M. und B. Lachtrup (1991). *Empirische Ursachenanalyse von Fehlzeiten an einem Unternehmensbeispiel, Sozialökonomische Texte* Band 27. Hamburg: Hochschule für Wirtschaft und Politik.
- Anonym (2005). Vollstationär, teilstationär, ambulant? Zeitliche Abgrenzung von Krankenhausbehandlungen. *Orthopädie & Rheuma* 2005(5), 57.

- von Auer, L. (2007). *Ökonometrie – Eine Einführung* (4. Aufl.). Heidelberg: Springer.
- Backhaus, K., B. Erichson, W. Plinke und R. Weiber (2003). *Multivariate Analysemethoden* (10. Aufl.). Springer.
- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3. Aufl.). Chichester: Wiley.
- Batini, C. und M. Scannapieco (2006). *Data Quality: Concepts, Methodologies And Techniques*. Berlin, Heidelberg, New York: Springer.
- Beck, N. (2005). Multilevel Analyses of Comparative Data: A Comment. *Political Analysis* 13, 457–458.
- Beck, N. (2008). Time-Series Cross-Section Methods. In: J. M. Box-Steffensmeier, H. E. Brady und D. Collier (Hrsg.), *The Oxford Handbook of Political Methodology*, Chapter 20, S. 475–493. New York: Oxford University Press.
- Beck, N. und J. N. Katz (1995). What to do (and not to do) with Time-Series Cross-Section Data. *American Political Science Review* 89, 634–664.
- Belsley, D. A., E. Kuh und R. E. Welsch (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Hoboken, NJ: Wiley Interscience.
- Berk, R. und J. M. MacDonald (2008). Overdispersion and Poisson Regression. *Journal of Quantitative Criminology* 24, 269–284.
- Berrington, A., P. W. F. Smith und P. Sturgis (2006). An Overview of Methods for the Analysis of Panel Data. NCRM Method Review Papers NCRM/007, ERSC National Center for Research Methods, Southampton.
- Blöß, T. (2005). Wirtschaftsfaktor Gesundheitswesen – Die unterschätzte Branche. Altenpflege, Medizintechnik, Patiententourismus und die Versorgung kranker Menschen – mehr als eine Million zusätzliche Jobs könnten im Gesundheitssektor entstehen. *Deutsches Ärzteblatt* 102, A644–A649.
- Böhm, K. (2005). Gesundheitsdaten in Deutschland. *Wirtschaft und Statistik* 2005(8), 828–837.

- Bömermann, H. (2003). Entwicklung und Bewertung der Krankenhausverweildauer auf Makro- und Mikroebene. Ist der arithmetische Mittelwert ein geeignetes Verweildauermaß? *Berliner Statistik Monatschrift* 2003(4), 145–149.
- Bömermann, H. und A. Hauswirth (2002). Anmerkungen zur Methodik und Qualitätsbewertung der amtlichen Diagnosestatistik. *Berliner Statistik Monatschrift* 2002(8), 318–328.
- Bonilla-Chacin, M.-E., E. Murrugarra und M. Temourov (2005). Health care during transition and health systems reform: evidence from the poorest CIS countries. *Social Policy & Administration* 39, 381–408.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Aufl.). Berlin: Springer.
- Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In: R. L. Launer und G. N. Wilkinson (Hrsg.), *Robustness in Statistics*, S. 201–236. New York: Academic Press.
- Breckenkamp, J., C. Wiskow und U. Laaser (2007). Progress on quality management in the German health system – a long and winding road. *Health Research Policy and Systems* 5, paper 7.
- Brewer, C. A. und L. Pickle (2002). Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers* 92, 662–681.
- Breyer, F., P. Zweifel und M. Kifmann (2005). *Gesundheitsökonomik* (5. Aufl.). Springer-Lehrbuch. Berlin, Heidelberg: Springer.
- Brezinski, H. und M. Fritsch (1995). Transformation: The shocking German way. *MOCT-MOST: Economic Policy in Transitional Economies* 5(4), 1–25.
- Brown, H. und R. Prescott (2006). *Applied Mixed Models in Medicine* (2. Aufl.). Statistics in Practice. Chichester: Wiley.
- Bryk, A. S. und S. W. Raudenbush (1992). *Hierarchical Linear Models, Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.
- Buccianti, A. und V. Pawlowsky-Glahn (2005). New Perspectives on Water Chemistry and Compositional Data Analysis. *Mathematical Geology* 37, 703–727.

- Buldyrev, S. V., L. A. Nunes Amaral, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley und M. H. R. Stanley (1997). Scaling Behavior in Economics: II. Modeling of Company Growth. *Journal de Physique I* 7, 635–650.
- Bundesministerium für Justiz (1987). Bundesstatistikgesetz (BStatG) vom 22. Januar 1987 (BGBl. I S. 462), das zuletzt durch Artikel 3 des Gesetzes vom 7. September 2007 (BGBl. I S. 2246) geändert worden ist. http://www.gesetze-im-internet.de/bundesrecht/bstatg_1987/gesamt.pdf (28. 03. 2010).
- Bundesministerium für Justiz (1988). Sozialgesetzbuch (SGB) Fünftes Buch (V) – Gesetzliche Krankenversicherung – (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477), das durch Artikel 2 des Gesetzes vom 22. Dezember 2010 (BGBl. I S. 2309) geändert worden ist. http://bundesrecht.juris.de/bundesrecht/sgb_5/gesamt.pdf (05. 04. 2011).
- Bundesministerium für Justiz (1990a). Berlinförderungsgesetz (BerlinFG) in der Fassung der Bekanntmachung vom 2. Februar 1990 (BGBl. I S. 173), das zuletzt durch Artikel 2 Absatz 5 des Gesetzes vom 5. Dezember 2006 (BGBl. I S. 2748) geändert worden ist. <http://bundesrecht.juris.de/bundesrecht/berlinfg/gesamt.pdf> (28. 03. 2010).
- Bundesministerium für Justiz (1990b). Krankenhausstatistik-Verordnung vom 10. April 1990 (BGBl. I S. 730), die zuletzt durch Artikel 4b des Gesetzes vom 17. März 2009 (BGBl. I S. 534) geändert worden ist. <http://bundesrecht.juris.de/bundesrecht/khstatv/gesamt.pdf> (05. 04. 2011).
- Bundesministerium für Justiz (1991). Krankenhausfinanzierungsgesetz in der Fassung der Bekanntmachung vom 10. April 1991 (BGBl. I S. 886), das zuletzt Artikel 1 des Gesetzes vom 17. März 2009 (BGBl. I S. 534) geändert worden ist. <http://bundesrecht.juris.de/bundesrecht/khg/gesamt.pdf> (05. 04. 2011).
- Bundesministerium für Justiz (1992). Gesundheitsstrukturgesetz vom 21. Dezember 1992 (BGBl. I S. 2266), das zuletzt durch Artikel 205 der Verordnung vom 25. November 2003 (BGBl. I S. 2304) geändert worden ist. <http://bundesrecht.juris.de/bundesrecht/gsg/gesamt.pdf> (29. 05. 2010).
- Bundesministerium für Justiz (1994). Bundespflegesatzverordnung vom 26. September 1994 (BGBl. I S. 2750), die zuletzt durch Artikel 10 des Gesetzes vom 22. Dezember 2010 (BGBl. I S. 2309) geändert worden ist. http://bundesrecht.juris.de/bundesrecht/bpflv_1994/gesamt.pdf (21. 05. 2011).

- Bundesministerium für Justiz (2003a). Bundesdatenschutzgesetz (BDSG) in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), das zuletzt durch Artikel 1 des Gesetzes vom 14. August 2009 (BGBl. I S. 2814) geändert worden ist. http://bundesrecht.juris.de/bundesrecht/bdsg_1990/gesamt.pdf (28.03.2011).
- Bundesministerium für Justiz (2003b). Gesetz zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz – GMG) in der Fassung der Bekanntmachung vom 14. November 2003 (BGBl. I S. 2190), das zuletzt durch Artikel 1 des Gesetzes vom 15. Dezember 2004 (BGBl. I S. 3445) geändert worden ist. <http://www.buzer.de/gesetz/7332/index.htm> (07.06.2011).
- Busse, R. und A. Riesberg (2005). *Gesundheitssysteme im Wandel: Deutschland*. Kopenhagen: WHO Regionalbüro für Europa im Auftrag des Europäischen Observatoriums für Gesundheitssysteme und Gesundheitspolitik.
- Busse, R. und M. Wörz (2009). Ausländische Erfahrung mit ambulanten Leistungen am Krankenhaus. In: J. Klauber, B. Robra und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2008/09. Schwerpunkt Versorgungszentren*, S. 49–59. Stuttgart, New York: Schattauer.
- Campbell, M. (2010). A statistician on a NICE committee. *Significance* 7, 81–84.
- Cleveland, W. S. (1994). *The Elements of Graphing Data* (2. Aufl.). Summit, NJ: Hobart Press.
- Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM* 13, 377–387.
- Codd, E. F. (1982). Relational Database: A Practical Foundation for Productivity. *Communications of the ACM* 25, 391–410.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B (Statistical Methodology)* 34, 187–220.
- Crémieux, P.-Y. und P. Ouellette (2001). Omitted variable bias and hospital costs. *Journal of Health Economics* 20, 271–282.

- Cromley, R. G. und E. K. Cromley (2009). Chloropleth map legend design for visualizing community health disparities. *International Journal of Health Geographics* 8, 52.
- Day, K. und J. Tousignant (2005). Health Spending, Health Outcomes, and Per Capita Income in Canada: A Dynamic Analysis. Working Paper – Document de travail 2005–07, Department of Finance Canada – Ministère des Finances Canada, Ottawa.
- Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics* 30, 109–126.
- DeMaris, A. (2004). *Regression with Social Data: Modeling Continuous and Limited Response Variables*. Hoboken, NJ: Wiley.
- Dietrich, I. und H. G. Strohe (2011a). Die Finanzlage öffentlicher Unternehmen in Deutschland – Statistische Analyse amtlicher Mikrodaten der Jahresabschlüsse. Statistische Diskussionsbeiträge 46, Universität Potsdam, Potsdam.
- Dietrich, I. und H. G. Strohe (2011b). Statistik der öffentlichen Unternehmen in Deutschland – Die Datenbasis. Statistische Diskussionsbeiträge 44, Universität Potsdam, Potsdam.
- Diggle, P. J., K.-Y. Liang und S. L. Zeger (1994). *Analysis of Longitudinal Data*. Oxford Statistical Sciences Series # 13. Oxford: Clarendon Press.
- Domanja, K., G. Philipp und B. Mensch (2007). Krankenhausstatistik: Psychische und Verhaltensstörungen durch Alkohol. *Statistik in Sachsen 2007*(3), 81–86.
- Dranove, D. und R. Lindrooth (2003). Hospital consolidation and costs: another look at the evidence. *Journal of Health Economics* 22, 983–997.
- Egozcue, J. J. und V. Pawlowsky-Glahn (2006). Simplicial geometry for compositional data. In: A. Buccianti, G. Mateu-Figueras und V. Pawlowsky-Glahn (Hrsg.), *Compositional Data Analysis in the Geosciences: From Theory to Praxis*, Geological Society Special Publication Band 264, S. 145–159. London: Geological Society.
- Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras und C. Barceló-Vidal (2003). Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology* 35, 279–300.

- Eisenhart, C. (1947). The Assumptions Underlying the Analysis of Variance. *Biometrics* 3, 1–21.
- European Union (2001). Presseerklärung IP/01/1747 (deutsche Version): Medizinische Versorgung – Kommission schlägt drei gemeinsame EU-Ziele für Gesundheitswesen und Altenpflege vor: Zugang für alle, hohe Qualität und langfristige Finanzierbarkeit. <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/01/1747&language=DE> (05. 05. 2010).
- Evans, I. S. (1977). The selection of class intervals. *Transactions of the Institute of British Geographers – New Series* 2, 98–124.
- Fielding, A. (2004). The role of the Hausman test and whether higher level effects should be treated as random or fixed. *Multilevel Modelling Newsletters* 16(2), 3–8.
- Filzmoser, P., K. Hron und C. Reimann (2009). Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Science of the Total Environment* 407, 6100–6108.
- Forschungsdatenzentrum der Statistischen Landesämter (2007). Nutzung von Mikrodaten der Krankenhausstatistik (FDZ-Nutzung 606-2007, Uni Potsdam – Strohe).
- Frees, E. W. (2004). *Longitudinal and Panel Data*. Cambridge, New York, Victoria: Cambridge University Press.
- Friedrich, J. und A. Beivers (2009). Patientenwege ins Krankenhaus: Räumliche Mobilität bei Elektiv- und Notfalleistungen am Beispiel von Hüftendoprothesen. In: J. Klauber, B. Robra und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2008/09. Schwerpunkt Versorgungszentren*, S. 155–181. Stuttgart, New York: Schattauer.
- Fry, J. M., T. R. L. Fry, K. R. McLaren und T. N. Smith (2001). Modelling zeroes in microdata. *Applied Economics* 33, 383–392.
- Gale, N. und W. C. Halperin (1982). A case for better graphics: the unclassed choropleth map. *The American Statistician* 36, 330–336.
- Gardiner, J. C., Z. Luo und L. A. Roman (2009). Fixed effects, random effects and GEE: What are the differences? *Statistics in Medicine* 28, 221–239.

- von Gaudecker, H.-M. (2006). Differentielle Sterblichkeit in der GRV: Problem-
aufriß und erste Berechnungen. In: *Bericht vom zweiten Workshop des For-
schungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni
2005 in Würzburg, DRV-Schriften* Band 55/2005, S. 242–252. Berlin: Deutsche
Rentenversicherung Bund.
- Gaynor, M. und G. F. Anderson (1995). Uncertain demand, the structure of hospi-
tal costs, and the cost of empty hospital beds. *Journal of Health Economics* 14,
291–317.
- Gelman, A. und P. N. Price (1999). All maps of parameter estimates are misleading.
Statistics in Medicine 18, 3221–3234.
- Gelman, A. und H. Stern (2006). The Difference Between 'Significant' and 'Not
Significant' is not Itself Statistically Significant. *The American Statistician* 60,
328–331.
- Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold.
- Graf, M. (2006). *Swiss Earnings Structure Survey 2002–2004 – Compositional data
in a stratified two-stage sample: Analysis and precision assessment of wage
components*. Swiss Statistics Methodology Report 338-0038. Neuchâtel: Swiss
Federal Statistical Office.
- Graubner, B. (2000). Aktuelle Diagnosen- und Prozedurenklassifikationen. *Zen-
tralblatt für Gynäkologie* 122(12), 611–624.
- Gresenz, C. R., J. Rogowski und J. J. Escarce (2004). Updated Variable-Radius Mea-
sures of Hospital Competition. *Health Services Research* 39, 417–430.
- Grielen, S. J., W. G. W. Boerma und P. P. Groenewegen (2000). Unity or diversity?
Task profiles of general practitioners in Central and Eastern Europe. *European
Journal of Public Health* 10, 249–254.
- Grömping, U. (1996). A note on fitting a marginal model to mixed effects log-
linear regression data via GEE. *Biometrics* 52, 280–285.
- Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The
American Statistician* 60, 19–26.

- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica* 12 (Supplement), 1–118.
- Hall, S., G. E. Mizon und A. Welfe (2000). Modelling economies in transition: an introduction. *Economic Modelling* 17, 339–357.
- Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory* 21, 60–68.
- Haubrock, M. (2005). Finanzierung von Krankenhausleistungen im Wandel – Vom Tagessatz zum leistungsbezogenen Entgelt. In: A. Kerres und B. Seeberger (Hrsg.), *Gesamtlehrbuch Pflegemanagement*, S. 235–268. Heidelberg: Springer Medizin Verlag.
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica* 46, 1251–1271.
- Hedeker, D. (2005). Generalized Linear Mixed Models. In: B. S. Everitt und D. C. Howell (Hrsg.), *Encyclopedia of Statistics in Behavioral Science*, S. 729–738. Chichester: Wiley.
- Hedeker, D. (2008). Multilevel Models for Ordinal and Nominal Variables. In: J. de Leeuw und E. Meijer (Hrsg.), *Handbook of Multilevel Analysis*, S. 237–274. New York: Springer.
- Hellenbrand, W. (2007). Zur Situation wichtiger Infektionskrankheiten in Deutschland: Keuchhusten-Erkrankungen in den neuen Bundesländern seit 2002. *Epidemiologisches Bulletin des Robert-Koch-Instituts [RKI] 2007*, 475–481.
- Heller, G. und S. Schmidt (2005). Wodurch ist die Veränderung der geburtshilflichen Diagnosehäufigkeiten in der Krankenhausdiagnosestatistik zu erklären? In: J. Klauber, B. Robra und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2005. Schwerpunkt Wege zur Integration*, S. 297–300. Stuttgart, New York: Schattauer.
- Henke, K.-D., B. Mackenthun und J. Schreyögg (2003). Gesundheit als Wirtschaftsfaktor. Eine ökonomische Analyse des Berliner Gesundheitsmarktes. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 46, 642–647.
- Herr, A. (2008). Cost and technical efficiency of German hospitals: Does ownership matter? *Health Economics* 17, 1057–1071.

- Herr, A. (2009). *On Hospital Competition: Quality, Efficiency, and Ownership*. Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen.
- Herr, A., H. Schmitz und B. Augurzky (2011). Profit efficiency and ownership of German hospitals. *Health Economics* 20, 0660–0674.
- Himmelreicher, R. K. (2006). Persönliche Kommunikation mit dem Autor.
- Hoover, K. D. (2006). The Methodology of Econometrics. In: T. C. Mills und K. Patterson (Hrsg.), *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, S. 61–87. Basingstoke: Palgrave Macmillan.
- Hron, K. und P. Filzmoser (2008). Robuste Biplots für Kompositionsdaten. In: *Abstracts of the Workshop TU Wien/TU Dresden*, S. 16.
- Hsiao, C. (2007). Panel data analysis – advantages and challenges (mit Diskussion). *Test* 16, 1–57.
- Huntington, S. P. (1991). *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: University of Oklahoma Press.
- Iezzoni, L. I. (1997). Assessing Quality Using Administrative Data. *Annals of Internal Medicine* 127, 666–674.
- Imlach Gunasekara, F., K. Carter und T. Blakely (2008). Glossary for econometrics and epidemiology. *Journal of Epidemiology & Community Health* 62, 858–861.
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2009a). Arbeitspapier Kostenbestimmung, Version 1.0. https://www.iqwig.de/download/Arbeitspapier_Kostenbestimmung_v_1_0.pdf (08. 04. 2011).
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2009b). Arbeitspapier Modellierung, Version 1.0. https://www.iqwig.de/download/Arbeitspapier_Modellierung_v_1_0.pdf (08. 04. 2011).
- Jenks, G. F. und F. C. Caspall (1971). Error on Choroplethic Maps. Definition, Measurement, Reduction. *Annals of the Association of American Geographers* 61, 217–244.
- Kauermann, G. und R. Ortlieb (2004). Temporal pattern in number of staff on sick leave: the effect of downsizing. *Journal of the Royal Statistical Society C (Applied Statistics)* 53, 355–367.

- Kish, L. (1985). Sample surveys versus experiments, controlled observations, censuses, registers, and local studies. *Australian & New Zealand Journal of Statistics* 27, 111–122.
- Klauber, J., M. Geraedts und J. Friedrich (Hrsg.) (2010). *Krankenhaus-Report 2010. Schwerpunkt: Krankenhausversorgung in der Krise?* Stuttgart, New York: Schattauer.
- Kleinsorge, A. (1998). Die Diagnosestatistik im Freistaat Sachsen in den Jahren 1993 bis 1996. *Statistik in Sachsen* 1998(2), 23–34.
- Knödler, R. (2005). Stationäre Krankenhausversorgung und demografische Entwicklung. *Statistisches Monatsheft Baden-Württemberg* 2005(8), 26–29.
- Krapppweis, H., J. Krapppweis und W. Kirch (2000). Pharmaökonomie und Public Health. *Journal of Public Health* 8, 4–13.
- Krause, P., J. Goebel, M. Kroh und G. G. Wagner (2010). 20 Jahre Wiedervereinigung: Wie weit Ost- und Westdeutschland zusammengerückt sind. *DIW Wochenbericht* 44/2010, 2–12.
- Kuchinke, B. A. und A. Wübker (2009). Defizite von öffentlichen Allgemeinkrankenhäusern in Deutschland: Empirische Befunde von 1998 bis 2004 und wirtschaftspolitische Implikationen. *Perspektiven der Wirtschaftspolitik* 10, 290–308.
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge, New York, Victoria: Cambridge University Press.
- Lee, Y. und J. A. Nelder (2004). Conditional and marginal models: Another view. *Statistical Science* 19, 219–238. With discussion.
- Lesaffre, E. und B. D. Marx (1993). Collinearity in generalized linear regression. *Communications in Statistics – Theory and Methods* 22, 1933–1952.
- Maciejewski, M. L., P. Diehr, M. A. Smith und P. L. Hebert (2002). Common Methodological Terms in Health Services Research and Their Synonyms. *Medical Care* 40, 477–484 & 661.
- Maciejewski, M. L., E. M. Weaver und P. L. Hebert (2011). Synonyms in Health Services Research Methodology. *Medical Care Research and Review* 68, 156–176.

- Mackenthun, B., K.-D. Henke und J. Schreyögg (2003). Gesundheitswesen als Wachstumsfaktor (destatis Kolloquium 2003). <http://kolloq.destatis.de/2003/mackenthun.pdf> (17. 04. 2007).
- Maes, H.-J. (2000). World Health Report – Mixtur von harten und weichen Daten. *Deutsches Ärzteblatt* 97, A2289–A2293.
- Mangelsdorf, S. (2007). Transformationsprozess im verarbeitenden Gewerbe in Berlin und Brandenburg in den Jahren 1995 bis 2003. *Zeitschrift für amtliche Statistik Berlin Brandenburg* 2007(1), 42–46.
- Mangelsdorf, S. (2009). *Die Transformation des verarbeitenden Gewerbes in Berlin/Brandenburg unter Berücksichtigung der Exporte*. Dissertation, Universität Potsdam, Potsdam.
- Martín-Fernández, J. A. und S. Thió-Henestrosa (2006). Rounded zeros: some practical aspects for compositional data. In: A. Buccianti, G. Mateu-Figueras und V. Pawlowsky-Glahn (Hrsg.), *Compositional Data Analysis in the Geosciences: From Theory to Praxis, Geological Society Special Publication* Band 264, S. 191–201. London: Geological Society.
- Matthes, N. und A. Wiest (2002). Qualität im Krankenhaus – Überlegungen zu Begriffen und Methoden der Leistungsbewertung. In: M. Arnold, J. Klauber und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2002. Schwerpunkt: Krankenhaus im Wettbewerb*, S. 161–174. Stuttgart, New York: Schattauer.
- Mátyás, L. und P. Sevestre (Hrsg.) (1996). *The Econometrics of Panel Data: Handbook of Theory and Applications* (2. Aufl.). Dordrecht: Kluwer Academic Publishers.
- Mátyás, L. und P. Sevestre (Hrsg.) (2008). *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (3. Aufl.), *Advanced Studies in Theoretical and Applied Econometrics* Band 46. Berlin, Heidelberg: Springer.
- McCulloch, C. E. und S. R. Searle (2001). *Generalized, Linear and Mixed Models*. Wiley Series in Probability and Statistics. New York: Wiley.
- Merkel, W. (1999). *Systemtransformationen: eine Einführung in die Theorie und Empirie der Transformationsforschung*. UTB für Wissenschaft, Bd 2076. Opladen: Leske+Budrich.

- Mertens, A. und A. Haas (2006). Regionale Arbeitslosigkeit und Arbeitsplatzwechsel in Deutschland. Eine Analyse auf Kreisebene. *Jahrbuch für Regionalwissenschaft* 26, 147–169.
- Miller, J. E. (2005). *The Chicago Guide to Writing about Multivariate Analysis*. Chicago: Chicago University Press.
- Monmonier, M. (2005). Lying with maps. *Statistical Science* 20, 215–222.
- Mühlbauer, B. H. (2005). Unternehmenskultur im Krankenhaus zwischen Ethik und Ökonomie. In: A. Kerres und B. Seeberger (Hrsg.), *Gesamtlehrbuch Pflege-management*, S. 315–336. Heidelberg: Springer Medizin Verlag.
- Neubauer, G. und A. Beivers (2010). Zur Situation der stationären Versorgung: Optimierung unter schwierigen Rahmenbedingungen. In: J. Klauber, M. Geraedts und J. Friedrich (Hrsg.), *Krankenhaus-Report 2010. Schwerpunkt: Krankenhausversorgung in der Krise?*, S. 3–12. Stuttgart, New York: Schattauer.
- Nunes Amaral, L. A., S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley und M. H. R. Stanley (1997). Scaling Behavior in Economics: I. Empirical Results for Company Growth. *Journal de Physique I* 7, 621–633.
- Ortlieb, R. (2003). *Betrieblicher Krankenstand als personalpolitische Arena*. Gabler Edition Wissenschaft. Wiesbaden: Deutscher Universitäts-Verlag.
- Ossen, P. (2011). 20 Jahre Landeskrankenhausgesellschaft Thüringen. *Das Krankenhaus 2011*, 171–176.
- Palta, M. und C. L. Seplaki (2002). Causes, Problems, and Benefits of Different Between and Within Effects in the Analysis of Clustered Data. *Health Services and Outcomes Research Methodology* 3, 177–193.
- Pawlowsky-Glahn, V. und J. J. Egozcue (2007). Statistische Analyse von Kompositionsdaten. In: *58. Berg- und Hüttenmännischer Tag 2007 – Tagungsband Geologische Modellierung*, Freiberg, S. 253–260.
- Pawlowsky-Glahn, V., J. J. Egozcue und R. Tolosana-Delgado (2007). Lecture Notes on Compositional Data Analysis. <http://dugi-doc.udg.edu/bitstream/10256/297/1/CoDa-book.pdf> (13. 06. 2008).

- Pearson, K. (1897). On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London* 60, 489–498.
- Pipino, L. L., Y. W. Lee und R. Y. Wang (2002). Data Quality Assessment. *Communications of the ACM* 45(4), 211–218.
- de Pouvourville, G. (2002). Kann Qualität ein Wettbewerbsparameter im Gesundheitswesen sein? In: M. Arnold, J. Klauber und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2002. Schwerpunkt: Krankenhaus im Wettbewerb*, S. 161–174. Stuttgart, New York: Schattauer.
- Radermacher, W. und T. Körner (2006). Fehlende und fehlerhafte Daten in der amtlichen Statistik. Neue Herausforderungen und Lösungsansätze. *Allgemeines Statistisches Archiv* 90, 553–576.
- Rambert, L. (2004). *Beschäftigungsdynamik im Transformationsprozess der brandenburgischen Industrie*. Reihe Quantitative Ökonomie Band 142. Lohmar: Eul-Verlag.
- Rambert, L. und H. G. Strohe (2001). Analyse der Beschäftigungsentwicklung in brandenburgischen Betrieben anhand Paneldaten der amtlichen Statistik. In: U. Schasse und J. Wagner (Hrsg.), *Regionale Wirtschaftsanalysen mit Betriebspaneldaten – Ansätze und Ergebnisse*, S. 23–46. Hannover: Niedersächsisches Institut für Wirtschaftsforschung.
- Ramlau-Hansen, H. (1983). Smoothing Counting Process Intensities by Means of Kernel Functions. *The Annals of Statistics* 11, 453–466.
- Rasch, D. (Hrsg.) (1987). *Biometrisches Wörterbuch*. Berlin: Deutscher Landwirtschaftsverlag.
- Reh, G. (2001). Strukturdaten zur Krankenhausversorgung in Rheinland-Pfalz. *Statistische Monatshefte Rheinland-Pfalz* 2001(2), 27–30.
- Richter, B. (2008). Zur volkswirtschaftlichen Bedeutung des Gesundheitswesens im Spiegel der amtlichen Statistik. *Statistik in Sachsen* 2008(2), 89–101.
- Robert-Koch-Institut (2011). *Daten und Fakten: Ergebnisse der Studie »Gesundheit in Deutschland aktuell 2009«*. Beiträge zur Gesundheitsberichterstattung des Bundes. Robert-Koch-Institut.

- Robra, B., E. Schewart und S. Felder (2002). Perspektiven des Wettbewerbs im Krankenhaussektor. In: M. Arnold, J. Klauber und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2002. Schwerpunkt: Krankenhaus im Wettbewerb*, S. 49–59. Stuttgart, New York: Schattauer.
- Röhmel, J. (2009). *Effizienzmessung im Gesundheitswesen: Eine Analyse der Effizienz ostdeutscher Krankenhäuser*. Hamburg: Verlag Dr. Kovač.
- Rolland, S. (2005a). Krankenhäuser in Deutschland 2003. *Wirtschaft und Statistik 2005*(8), 838–848.
- Rolland, S. (2005b). Statistische Krankenhausdaten: Grund- und Kostendaten der Krankenhäuser 2003. In: J. Klauber, B. Robra und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2005. Schwerpunkt Wege zur Integration*, S. 237–264. Stuttgart, New York: Schattauer.
- Rönz, B. und H. G. Strohe (Hrsg.) (1994). *Lexikon Statistik*. Wiesbaden: Gabler.
- Rosenberg, A. L., M. V. H. Greenfield und J. B. Dimick (2006). Secondary Data Analyses. In: D. F. Penson und J. T. Wei (Hrsg.), *Clinical Research for Surgeons*, S. 175–200. Totowa, NJ: Humana Press.
- Rosenow, C. und A. Steinberg (2002). 10 Jahre bundeseinheitliche Krankenhausstatistik. *Wirtschaft und Statistik 2002*(5), 383–391.
- Rott, W. (2009). *Die Insel. Eine Geschichte West-Berlins. 1948-1990*. München: C. H. Beck.
- Rudzio, K. und W. Uchatius (2010). Arbeiten, bis der Arzt kommt. *Die Zeit 28/2010*, 21–22.
- SAS Institute Inc. (2004a). *SAS/ETS® 9.1 User's Guide*. Cary, NC: SAS Institute.
- SAS Institute Inc. (2004b). *SAS/STAT® 9.1 User's Guide*. Cary, NC: SAS Institute.
- SAS Institute Inc. (2004c). *SAS® 9.1 Macro Language: Reference*. Cary, NC: SAS Institute.
- SAS Institute Inc. (2005). The PANEL Procedure. Technical Report (31. 08. 2005), Cary, NC: SAS Institute.

- SAS Institute Inc. (2006). The GLIMMIX Procedure. Technical Report (30. 06. 2006), Cary, NC: SAS Institute.
- Schneeweiss, S. und O. Sangha (2001). Leistungsvergleiche in der Medizin. Bedarf, Anforderungen und Wege zur Akzeptanz. *Deutsche Medizinische Wochenschrift* 126, 918–924.
- Schoffer, O. (2005). Persönliche Kommunikation mit dem Autor.
- Schoffer, O. (2006). Persönliche Kommunikation mit dem Autor.
- Schoffer, O. (2008). Persönliche Kommunikation mit dem Autor.
- Schoffer, O. (2009). Persönliche Kommunikation mit dem Autor.
- Schulenburg, D. (2005). Was ist vollstationär? *Rheinisches Ärzteblatt* 2005(12), 14.
- Schulte, S. (2005). Statistische Krankenhausdaten: Diagnosedaten der Krankenhauspatienten 2003. In: J. Klauber, B. Robra und H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2005. Schwerpunkt Wege zur Integration*, S. 265–296. Stuttgart, New York: Schattauer.
- Schwierz, C. (2009). *Expansion in Markets with Decreasing Demand – For Profits in the German Hospital Industry*. Ruhr Economic Papers # 106. Bochum, Dortmund, Duisburg, Essen: Rheinisch-Westfälisches Institut für Wirtschaftsforschung.
- Singer, J. D. (1998). Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics* 24, 323–355.
- Singer, J. D. und J. B. Willett (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.
- Skrondal, A. und S. Rabe-Hesketh (2008). Multilevel and Related Models for Longitudinal Data. In: J. de Leeuw und E. Meijer (Hrsg.), *Handbook of Multilevel Analysis*, S. 275–299. New York: Springer.
- Smith, P. C., E. Mossialos und I. Papanicolas (2008). Leistungsmessung zur Verbesserung der Gesundheitssysteme. Erfahrungen, Herausforderungen und Ausichten. Hintergrundpapier, WHO Regionalbüro für Europa, Kopenhagen.

- Sowada, C. (1999). Soziale Reformen in Polen. Zwischen Bewahrung und Neuanfang. Finanzwissenschaftliche Diskussionsbeiträge 23, Universität Potsdam, Potsdam.
- Statistics Canada (2010). Population and dwelling counts. <http://www.statcan.gc.ca/> (10. 03. 2010), Suche nach Kenora District bzw. Prince Edward County.
- Statistikamt Nord (2009). *Statistisches Jahrbuch Hamburg 2009/2010*. Hamburg: Statistisches Amt für Hamburg und Schleswig-Holstein.
- Statistisches Bundesamt (1993–2005c). Grunddaten der Krankenhäuser und Vorsorge- oder Rehabilitationseinrichtungen (Fachserie 12, Reihe 6.1) 1991–2003. <https://www-ec.destatis.de/csp/shop/> (10. 12. 2010).
- Statistisches Bundesamt (2004). Tiefgegliederte Diagnosedaten der Krankenhauspatientinnen und -patienten, 2000 (Excel-Datei). <https://www-ec.destatis.de/csp/shop/> → Krankenhausstatistik → 23131 Diagnosen der Krankenhauspatienten (29. 03. 2009).
- Statistisches Bundesamt (2005a). Diagnosedaten der Patienten und Patientinnen in Krankenhäusern (einschl. Sterbe- und Stundenfälle) (Fachserie 12, Reihe 6.2.1) 2003. <https://www-ec.destatis.de/csp/shop/> (10. 12. 2010).
- Statistisches Bundesamt (2005b). Diagnosedaten der Patienten und Patientinnen in Vorsorge- oder Rehabilitationseinrichtungen (Fachserie 12, Reihe 6.2.2) 2003. <https://www-ec.destatis.de/csp/shop/> (17. 09. 2006).
- Statistisches Bundesamt (2006a). Kostennachweis der Krankenhäuser (Fachserie 12, Reihe 6.3) 2004. <https://www-ec.destatis.de/csp/shop/> (17. 09. 2006).
- Statistisches Bundesamt (2006b). Qualitätsbericht – Diagnosen der Krankenhauspatienten. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/Gesundheitswesen/Krankenhauspatienten,property=file.pdf> (14. 08. 2006).
- Statistisches Bundesamt (2006c). Qualitätsbericht – Grunddaten der Krankenhäuser. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/Gesundheitswesen/Grunddatenkrankenhaeser,property=file.pdf> (14. 08. 2006).

- Statistisches Bundesamt (2006d). Qualitätsbericht – Kostennachweis der Krankenhäuser. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/Gesundheitswesen/Kostennachweiskrankenhaeser,property=file.pdf> (14. 08. 2006).
- Statistisches Bundesamt (2006e). *Statistisches Jahrbuch 2006. Für die Bundesrepublik Deutschland*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2006–2010). Grunddaten der Krankenhäuser (Fachserie 12, Reihe 6.1.1) 2004–2008. <https://www-ec.destatis.de/csp/shop/> (10. 12. 2010).
- Statistisches Bundesamt (2007). *Statistisches Jahrbuch 2007. Für die Bundesrepublik Deutschland*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2008a). Krankenhauslandschaft im Umbruch: Begleitmaterial zur Pressekonferenz am 10. Dezember 2008 in Berlin. http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pk/2008/Krankenhaeuser/begleitheft__krankenhaeuser,property=file.pdf (18. 10. 2011).
- Statistisches Bundesamt (2008b). Qualitätsbericht – Diagnosen der Krankenhauspatienten. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/Gesundheitswesen/Krankenhauspatienten,property=file.pdf> (14. 04. 2008).
- Statistisches Bundesamt (2008c). Qualitätsbericht – Grunddaten der Krankenhäuser. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/Gesundheitswesen/Grunddatenkrankenhaeser,property=file.pdf> (14. 04. 2008).
- Statistisches Bundesamt (2008d). Qualitätsbericht – Kostennachweis der Krankenhäuser. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/Gesundheitswesen/Kostennachweiskrankenhaeser,property=file.pdf> (14. 04. 2008).
- Statistisches Bundesamt (2008e). *Statistisches Jahrbuch 2008. Für die Bundesrepublik Deutschland*. Wiesbaden: Statistisches Bundesamt.

- Statistisches Bundesamt (2011). Generationensterbetafeln für Deutschland: Modellrechnungen für die Geburtsjahrgänge 1896–2009. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/Bevoelkerung/Bevoelkerungsbewegung/Generationssterbetafeln5126101119004,property=file.pdf> (18. 10. 2011).
- Statistisches Landesamt Berlin (2002). Statistischer Bericht: Kennziffern für das Gesundheitswesen Berlin 1991–2001. <http://www.statistik-berlin.de/statistiken/gesundheit/kennziffern.pdf> (12. 04. 2008).
- Statistisches Landesamt Bremen (2009). *Statistisches Jahrbuch 2009*. Bremen: Statistisches Landesamt.
- Stiglitz, J. E., A. Sen und J. Fitoussi (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress (14. 09. 2009). http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf (23. 04. 2011).
- Strausberg, J. (2007). Die Kodierqualität in der stationären Versorgung. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 50, 1039–1046.
- Strohe, H. G. (1993). Weiche Modellierung umweltökonomischer Zusammenhänge. *Allgemeines Statistisches Archiv* 77, 281–310.
- Strohe, H. G. (1999). Zur Statistik der wirtschaftlichen Entwicklung in Ostdeutschland am Beispiel des Landes Brandenburg. In: D. Wagner (Hrsg.), *Bewältigung des ökonomischen Wandels – Entwicklungen der Transformation in Ost und West*, S. 311–331. München: Hampp.
- Strohe, H. G. (2000). The use of Confidential Industrial Microdata of the Brandenburg Official Statistics for Modelling Regional Economics – A Project Report. In: A. Zelias (Hrsg.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*. Kraków: PIW.
- Therneau, T. M. und P. M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Statistics for Biology and Health. New York, Berlin, Heidelberg: Springer.
- Ulbrich, H.-F. (2010). Höherdimensionale Kompositionsdaten – Gedanken zur grafischen Darstellung und Analyse. Statistische Diskussionsbeiträge 43, Universität Potsdam, Potsdam.

- de Veaux, R. D. und D. J. Hand (2005). How to Lie with Bad Data. *Statistical Science* 20, 231–238.
- Vittinghoff, E., D. V. Glidden, S. C. Shiboski und C. E. McCulloch (2005). *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. New York: Springer.
- Wallgren, A. und B. Wallgren (2007). *Register-based Statistics: Administrative Data for Statistical Purposes*. Chichester: Wiley.
- Walter, S. D. (1993). Visual and statistical assessment of spatial clustering in mapped data. *Statistics in Medicine* 12, 1275–1291.
- Wang, R. Y. und D. M. Strong (1996). Beyond accuracy: what data quality means to consumers. *Journal of Management Information Systems* 12(4), 5–34.
- Wein, E. (2002). Datenqualität und Datenprüfungen. *Wirtschaft und Statistik* 2002(2), 110–115.
- Werblow, A. und O. Schoffer (2010). Entwicklung des Personalbestandes in allgemeinen Krankenhäusern Deutschlands 2002–2007. In: J. Klauber, M. Geraedts und J. Friedrich (Hrsg.), *Krankenhaus-Report 2010. Schwerpunkt: Krankenhausversorgung in der Krise?*, S. 71–95. Stuttgart, New York: Schattauer.
- Willich, S. N. und R. Busse (2006). Bedeutung der Gesundheitsökonomie für die medizinische Versorgung. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 49, 1–2.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, London: MIT Press.
- Wooldridge, J. M. (2003). *Introductory Econometrics: A Modern Approach* (2. Aufl.). Mason, OH: Thomson South-Western.
- van der Zee, J. und M. W. Kroneman (2007). Bismarck or Beveridge: a beauty contest between dinosaurs. *BMC Health Services Research* 7, Art. 94.
- Zühlke, S., M. Zwick, S. Scharnhorst und T. Wende (2003). Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. *Wirtschaft und Statistik* 2003(10), 906–911.

Zühlke, S., M. Zwick, S. Scharnhorst und T. Wende (2004). The research data centres of the Federal Statistical Office and the statistical offices of the Länder. *Schmollers Jahrbuch 124*, 567–578.

Potsdamer Schriften zu Statistik und Wirtschaft

herausgegeben von Prof. Dr. em. Hans Gerhard Strohe

ISSN (print) 2192-8061

ISSN (online) 2192-807X

Zuletzt erschienene Ausgaben:

- Band 4** Hannes-Friedrich Ulbrich: Die Krankenhäuser Ostdeutschlands in Transition: Eine registerbasierte Analyse amtlicher Paneldaten | 2012
ISBN 978-3-86956-200-1
URN urn:nbn:de:kobv:517-opus-60408
- Band 3** Irina Dietrich: Öffentliche Unternehmen in Deutschland: Eine Analyse von Mikrodaten der amtlichen Statistik | 2012
ISBN 978-3-86956-182-0
URN urn:nbn:de:kobv:517-opus- 59199
- Band 2** Alexander Mehnert, Andreas Nastansky: Staatsverschuldung und Inflation: Eine empirische Analyse für Deutschland | 2012
ISBN 978-3-86956-181-3
URN urn:nbn:de:kobv:517-opus-59181
- Band 1** Marcus Ruge: Stimmungen und Erwartungen im System der Märkte: Eine Analyse mit DPLS-Modellen | 2011
ISBN 978-3-86956-142-4
URN urn:nbn:de:kobv:517-opus-52900

Vorliegende Arbeit untersucht die ersten eineinhalb Jahrzehnte Transition in Ostdeutschland ökonomisch. Im Mittelpunkt stehen dabei die Interaktionen zwischen der allgemeinen Wirtschaft und den Krankenhäusern sowie – auf Basis der Hauptdiagnose – den in Krankenhäusern behandelten Krankheiten. Verschiedene, in den politischen Diskussionen zur Effizienz der Krankenhäuser übliche Indikatoren werden verglichen und auf ihre Eignung zur adäquaten Beschreibung von Krankenhauskosten und -leistungen geprüft. Durch Einbeziehen der Daten aus Rheinland-Pfalz wird herausgearbeitet, welche der Veränderungen in Ostdeutschland transitionsbezogen sind und ob und wie sich die transitionsbezogenen von eher transitionsunabhängigen Veränderungen trennen lassen. Dazu werden den Daten – jährlich erhobenen Angaben zu einem jeden Krankenhaus bzw. jedem Kreis des Untersuchungsgebietes – Paneldatenmodelle angepasst. Ergänzt um deskriptive Statistiken, Graphen und Choroplethenkarten sind diese Modelle Grundlage einer wirtschaftswissenschaftlichen Interpretation der Veränderungen zwischen 1992 und 2005.

