

Enno Swart Peter Ihle Holger Gothe David Matusiewicz
Herausgeber

Routinedaten im Gesundheitswesen

Handbuch Sekundärdatenanalyse:
Grundlagen, Methoden und Perspektiven

2., vollständig überarbeitete Auflage

HUBER



Swart/Ihle/Gothe/Matusiewicz (Hrsg.)
Routinedaten im Gesundheitswesen

Verlag Hans Huber
Programmbereich Gesundheit

Wissenschaftlicher Beirat:
Ansgar Gerhardus, Bremen
Klaus Hurrelmann, Berlin
Petra Kolip, Bielefeld
Milo Puhan, Zürich
Doris Schaeffer, Bielefeld

HUBER



Enno Swart
Peter Ihle
Holger Gothe
David Matusiewicz

Herausgeber

Routinedaten im Gesundheitswesen

**Handbuch Sekundärdatenanalyse:
Grundlagen, Methoden und Perspektiven**

2., vollständig überarbeitete und erweiterte Auflage

Verlag Hans Huber

Lektorat: Dr. Klaus Reinhardt
Herstellung: Daniel Berger
Bearbeitung: Ulrike Boos, Freiburg
Umschlaggestaltung: Claude Borer, Basel
Druckvorstufe: punktgenau gmbh, Bühl
Druck und buchbinderische Verarbeitung: Hubert & Co., Göttingen
Printed in Germany

Bibliografische Information der Deutschen Nationalbibliothek

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



Anregungen und Zuschriften bitte an:

Verlag Hans Huber
Lektorat Medizin/Gesundheit
Länggass-Strasse 76
CH-3000 Bern 9
Tel: 0041 (0)31 300 4500
verlag@hanshuber.com
www.verlag-hanshuber.com

2. Auflage 2014

© 2005/2014 by Verlag Hans Huber, Hogrefe AG, Bern
(E-Book-ISBN [PDF] 978-3-456-95435-6)
ISBN 978-3-456-85435-9

Inhalt

Vorwort	12
Vorwort der ersten Auflage	
Sekundärdatenanalyse: Aufgaben und Ziele (<i>Enno Swart, Peter Ihle</i>)	16
AGENS – eine schrecklich routinierte Familie	19
(<i>David Matusiewicz, Holger Gothe, Peter Ihle, Enno Swart</i>)	
I. Versorgungsbereiche im Spiegel der Gesetzlichen Krankenversicherung	27
1. Stammdaten und Versichertenhistorien (<i>Thomas G. Grobe, Peter Ihle</i>)	28
1.1 Einleitung	28
1.2 Versichertengruppen	29
1.3 Versichertenbezogene Merkmale	30
1.4 Nutzung für wissenschaftliche Fragestellungen	34
1.5 Auswahl von Bezugspopulationen	35
2. Analysen zur Sterblichkeit (<i>Rembrandt Scholz, Sebastian Sauer, Rolf Müller</i>)	38
2.1 Hintergrund	38
2.2 Untersuchungsgegenstand: Sterberaten	39
2.3 Datengrundlage, methodisches Vorgehen und Selektion	39
2.4 Sterberaten der Deutschen Rentenversicherung und der Gesetzlichen Krankenversicherung versus Bevölkerungsdaten der Human Mortality Database	40
2.5 Ausblick	42
3. Ambulante ärztliche Versorgung (<i>Thomas G. Grobe, Hendrik Dräther</i>)	43
3.1 Hintergrund – Abgrenzung – gesetzliche Grundlagen	44
3.2 Datenerhebungswege – Datenhalter	47
3.3 Datenstruktur – Datenumfang – Kennwerte	48
3.4 Historische Verfügbarkeit – wesentliche Veränderungen	56
3.5 Beispielauswertungen	57
3.6 Resümee: Möglichkeiten und Grenzen	60
4. Dokumentation der Disease-Management-Programme (<i>Dirk Horenkamp-Sonntag, Roland Linder, Dagmar Köppel, Daniel Wildner</i>)	63
4.1 Hintergrund und Historie	64
4.2 Datenentstehung	65
4.3 Datenstruktur	67
4.4 Historische Verfügbarkeit	70
4.5 Evaluation	71
4.6 Daten-Qualität der Dokumentationsdaten	71
4.7 Ausblick: Möglichkeiten und Grenzen	72
5. Arzneimittelverordnungen (<i>Helmut Schröder</i>)	74
5.1 Einführung	74
5.2 Was sind Arzneimittel?	75

5.3	Wie kommen die Arzneimitteldaten zu den Krankenkassen?	76
5.4	Analysen mit Arzneimitteldaten	79
5.5	Kassenartenübergreifende Arzneimitteldaten	82
5.6	Kassenspezifische Arzneimitteldaten	83
5.7	Fazit	85
6.	Heilmittel (<i>Andrea Waltersbacher</i>)	88
6.1	Hintergrund	88
6.2	Wie entstehen die Heilmitteldatenpools?	90
6.3	Heilmitteldatenpools, Kennwerte und Veröffentlichungen von Heilmitteldaten	92
6.4	Ausblick: Möglichkeiten und Grenzen	102
7.	Zahnärztliche Versorgung (<i>Rugzan Hussein, Andrea Hartmann, Thomas Schäfer, Eva-Maria Bitzer</i>)	104
7.1	Hintergrund – Abgrenzung – gesetzliche Grundlagen	105
7.2	Datenerhebungswege – Datenhalter	106
7.3	Ablauf der elektronischen Datenübermittlung	106
7.4	Formulare, die vor der elektronischen Datenübermittlung eingesetzt wurden	108
7.5	Datenstruktur – Datenumfang	112
7.6	Kennwerte	112
7.7	Publikationen und bevölkerungsbezogene Studien in Deutschland – Referenzergebnisse .	116
7.8	Resümee: Möglichkeiten und Grenzen	118
8.	Krankenhausbehandlung (<i>Thomas G. Grobe, Ulrike Nimptsch, Jörg Friedrich</i>)	121
8.1	Hintergrund – Abgrenzung – gesetzliche Grundlagen	122
8.2	Datenerhebungswege – Datenhalter	126
8.3	Datenstruktur – Datenumfang – Kennwerte	128
8.4	Auswertung und Kennwerte	135
8.5	Beispielauswertungen	141
8.6	Historische Verfügbarkeit – wesentliche Veränderungen	143
8.7	Resümee	144
9.	Arbeitsunfähigkeit (<i>Markus Meyer</i>)	146
9.1	Hintergrund und gesetzliche Rahmenbedingungen	146
9.2	Entstehung der Arbeitsunfähigkeitsdaten und Datenhaltung	147
9.3	Datenstruktur: Welche Daten stehen zur Verfügung?	148
9.4	Arbeitsunfähigkeitsdaten in der Gesundheitsberichterstattung und Epidemiologie	149
9.5	Kennwerte	150
9.6	Möglichkeiten und Grenzen	158
9.7	Resümee	159
II. Versorgungsbereiche im Spiegel anderer Sozialversicherungsträger		161
10.	Pflegeleistungen nach Sozialgesetzbuch XI (<i>Rolf Müller, Heinz Rothgang, Rainer Unger</i>)	162
10.1	Hintergrund	163
10.2	Datenentstehung in der Pflegekasse	166
10.3	Daten bei den Pflegekassen	167
10.4	Historische Verfügbarkeit – wesentliche Veränderungen	169
10.5	Amtliche Statistiken zur Pflegeversicherung	170

10.6	Resümee	171
10.7	Ausblick	173
11.	Routinedaten zur Rehabilitation durch die Träger der Sozialversicherung (<i>Anja Bestmann, Ute Polak, Ingmar Hansen</i>)	176
11.1	Hintergrund	176
11.2	Datenentstehung in der Rehabilitation	179
11.3	Datenstruktur, -aufbau und Datenumfang, Besonderheiten	181
11.4	Resümee: Möglichkeiten und Grenzen	186
12.	Rekonstruktion von Erwerbsverläufen (<i>Anja Burghardt, Anita Tisch, Silke Tophoven</i>)	192
12.1	Einleitung	192
12.2	Sekundärdaten der Arbeitsverwaltung	193
12.3	Die Verwendung von erwerbsbiografischen Informationen aus den Sekundärdaten der Bundesagentur für Arbeit und des Instituts für Arbeitsmarkt- und Berufsforschung ..	196
12.4	Die Verknüpfung von administrativen Erwerbsbiografiedaten mit Primärdaten der lidA-(leben in der Arbeit)Kohortenstudie zu Arbeit und Gesundheit	197
12.5	Die Sequenzmusteranalyse als Beispiel der Abbildung erwerbsbiografischer Verläufe.	198
12.6	Zusammenfassung	200
13.	Daten zu Erwerbsminderungsrenten (<i>Ralf K. Himmelreicher</i>)	203
13.1	Hintergrund	203
13.2	Datenentstehung	204
13.3	Datenstruktur	204
13.4	Resümee	210
III.	Ergänzende Routinedaten mit Gesundheitsbezug	213
14.	Gesundheitsbefragungen und Aggregatdaten (<i>Jelena Jaunzeme, Yvonne Marx, Enno Swart, Siegfried Geyer</i>)	214
14.1	Einleitung	215
14.2	Gesundheitsmonitoring des Robert Koch-Instituts	215
14.3	Weitere gesundheitsrelevante Surveys	217
14.4	Bevölkerungsumfragen mit gesundheitsrelevanten Themen	218
14.5	Nutzung von Aggregatdaten	219
14.6	Resümee	221
15.	Möglichkeiten und Grenzen von Befragungsdaten und Daten gesetzlicher Krankenversicherungen (<i>Siegfried Geyer, Jelena Jaunzeme</i>)	223
15.1	Einführung	224
15.2	Befragungsdaten	224
15.3	Daten gesetzlicher Krankenversicherungen	227
15.4	Zusammenfassende Beurteilung	230
16.	Registerdaten (<i>Birga Maier, Holger Gothe, Joachim Kieschke</i>)	234
16.1	Einführung	234
16.2	Was ist ein Register?	234
16.3	Welche Register gibt es?	235
16.4	Epidemiologisches Krebsregister Niedersachsen (EKN) als Beispiel für ein Krankheitsregister	237

16.5	Berliner Herzinfarktregister als Beispiel für ein Qualitätsregister	239
16.6	Welche Vorzüge haben Register?	241
16.7	Welche Grenzen und Probleme weisen Register auf?	241
16.8	Zusammenfassung und Ausblick	243
17.	Daten der Privaten Krankenversicherung (PKV) (<i>Holger Gothe, Anne-Dorothee Köster</i>)	245
17.1	Das System der Privaten Krankenversicherung in Deutschland	245
17.2	Hintergründe für die Datenlage in der PKV	246
17.3	Versorgungsdaten der PKV	247
17.4	Stellenwert der Nutzung von Versorgungsdaten der PKV	251
18.	Leistungserbringerbezogene Merkmale (<i>Ulrike Nimptsch</i>)	254
18.1	Hintergrund	254
18.2	Merkmale zur Abgrenzung von Leistungserbringern	255
18.3	Zuordnung von leistungserbringerbezogenen Merkmalen	258
19.	Routinedaten im Ausland (<i>Holger Gothe</i>)	260
19.1	Hintergrund	260
19.2	Arztbasierte Datensätze	261
19.3	Administrative Datensätze	263
19.4	Fazit	265
IV.	Datenzugang und Datenvalidierung	269
20.	Zugang zu Routinedaten (<i>Ulrike Nimptsch, Anja Bestmann, Michael Erhart, Stefan Dudey, Yvonne Marx, Joachim Saam, Michael Schopen, Helmut Schröder, Enno Swart</i>)	270
20.1	Einleitung	270
20.2	Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder	271
20.3	Reha-Statistik der deutschen Rentenversicherung	275
20.4	Informationssystem Versorgungsdaten (Datentransparenz)	278
20.5	KV-Daten-Stichprobe des Zentralinstituts für die kassenärztliche Versorgung	279
20.6	Versorgungsdaten der Versicherten bei gesetzlichen Krankenversicherungen	283
21.	Datenschutzrechtliche Aspekte bei der Nutzung von Routinedaten (<i>Stefanie March, Angela Rauch, Stefan Bender, Peter Ihle</i>)	291
21.1	Rechtlicher Hintergrund	291
21.2	Organisatorisches Vorgehen	294
21.3	Zwei Beispielstudien unter datenschutzrechtlichen Gesichtspunkten	300
21.4	Fazit	301
22.	Datenhaltung und -analyse aus technischer Sicht (<i>Peter Ihle, David Matusiewicz, Thomas G. Grobe, Gisbert W. Selke</i>)	304
22.1	Einleitung	304
22.2	Datenhaltung	305
22.3	Datenverknüpfung	306
22.4	Datenzugang und Datenübermittlung	306
22.5	Datenvolumen	308
22.6	Software	309
22.7	Fazit	313

23. Prüfung der Datenqualität und Validität von GKV-Routinedaten (<i>Dirk Horenkamp-Sonntag, Roland Linder, Fabian Wenzel, Bettina Gerste, Peter Ihle</i>)	314
23.1 Warum müssen Daten geprüft werden?	315
23.2 Welche Möglichkeiten der Prüfung gibt es?	317
23.3 Anwendungsbeispiele zur Plausibilität	319
23.4 Anwendungsbeispiele zur internen Validierung	322
23.5 Anwendungsbeispiele zur externen Validierung	324
23.6 Prüfstrategien	326
23.7 Fazit zur Prüfung der Daten-Qualität	329
24. Klassifikationssysteme (<i>Anja Bestmann, Carsten Telschow, Michael Stegmann</i>)	331
24.1 Hintergrund	331
24.2 Internationale statistische Klassifikation der Krankheiten und verwandten Gesundheitsprobleme (ICD-10)	332
24.3 Operationen- und Prozedurenschlüssel (OPS)	334
24.4 Einheitlicher Bewertungsmaßstab (EBM)	335
24.5 Anatomisch-therapeutisch-chemisches System (ATC)	338
24.6 Pharmazentralnummer (PZN)	340
24.7 Die Neugestaltung des Tätigkeitsschlüssels	342
24.8 Institutionskennzeichen (IK)	343
25. Datenlinkage (<i>Stefanie March, Christoph Stallmann, Enno Swart</i>)	347
25.1 Allgemeine Anforderungen an das Datenlinkage	348
25.2 Ausgewählte Beispiele des Datenlinkages	350
V. Routinedaten als Informationsquelle für Morbiditätsschätzungen	357
26. Krankheitsereignis: Operationalisierung und Falldefinition (<i>Ingrid Schubert, Ingrid Köster</i>)	358
26.1 Hintergrund	359
26.2 Operationalisierung von Krankheiten in der Versorgungsforschung und Pharmakovigilanz: Stellenwert der Diagnosedaten	360
26.3 Szenarien der Diagnosegenerierung im ambulanten Sektor	361
26.4 Interne Diagnosevalidierung und Falldefinition	363
26.5 Externe Diagnosevalidierung/Plausibilisierung	365
26.6 Fazit	366
27. Bestimmung von Behandlungs- und Erkrankungsperioden in Routinedaten (<i>Sascha Abbas, Peter Ihle</i>)	369
27.1 Hintergrund und Datengrundlage	369
27.2 Behandlungs- versus Erkrankungszeitraum	370
27.3 Prävalenz- und Inzidenzbestimmung – praktische Umsetzung in den Daten	371
27.4 Persistenz der Erkrankung und Diagnosemuster in Routinedaten	374
27.5 Fazit	374
28. Die Population unter Risiko bei Prävalenz- und Inzidenzschätzungen – Nennerkonzepte (<i>Christoph Wagner</i>)	376
28.1 Hintergrund	377
28.2 Besonderheiten von Versichertenstammdaten für eine Nennerdefinition	378
28.3 Zeitliche Verfügbarkeit von Leistungsdaten	381

28.4	Verfahren zur Selektion einer Population unter Risiko	381
28.5	Nennerkonzepte für Daten der Datentransparenzverordnung (§ 303a–e SGB V)	385
28.6	Fazit	386
29.	Fallbeispiele für Prävalenz- und Inzidenzschätzungen (<i>Christoph Wagner, Dirk Horenkamp-Sonntag, Thomas Grobe, Roland Linder</i>)	389
29.1	Einleitung	389
29.2	Neugeborenen-Inzidenz für Mukoviszidose – ein cleverer Spezialfall	391
29.3	Extrapolation der Lebenszeitprävalenz der Aufmerksamkeitsdefizit-/Hyperaktivitäts- störung	393
29.4	Wie umgehen mit der Erkrankungsdunkelziffer bei Osteoporose?	396
29.5	Fazit	397
VI.	Spezielle Methoden, Instrumente und Anwendungen	401
30.	Matching-Verfahren (<i>Sarah Mostardt, Gerald Lux, Helmut Dahl, David Matusiewicz, Janine Biermann</i>)	402
30.1	Hintergrund	402
30.2	Verfügbarkeit von Matching-Variablen in Sekundärdaten	403
30.3	Matching-Verfahren	404
30.4	Herausforderungen und Probleme beim Matching	406
30.5	Fazit und Ausblick	410
31.	Risikoadjustierung und Komorbiditäten (<i>Gerald Lux, Janine Biermann, Helmut Dahl, David Matusiewicz, Sarah Mostardt, Ulrike Nimptsch, Jürgen Wasem, Anke Walendzik</i>)	411
31.1	Hintergrund	411
31.2	Berücksichtigungsfähige Variablen für eine Risikoadjustierung	412
31.3	Komorbiditätsbasierte Risikoindizes	415
31.4	Methodische Herausforderungen der Risikoadjustierung	416
31.5	Anwendungsfelder der Risikoadjustierung	417
31.6	Fazit	421
32.	Bias und Confounding (<i>Raffaella Matteucci Gothe, Barbara Buchberger</i>)	424
32.1	Einführung	424
32.2	Definition und Beispiele von Bias und Confounding	426
32.3	Maßnahmen zur Vermeidung/Adjustierung von Bias und Confounding	428
32.4	Zusammenfassung	432
33.	Routinedaten für kleinräumige Analysen (<i>Janett Powietzka, Enno Swart</i>)	435
33.1	Vorbemerkungen	435
33.2	Regionalisierte Daten der Gesetzlichen Krankenversicherungen	436
33.3	Versorgungsatlas	437
33.4	Raumordnungsdaten des Bundesinstituts für Bau-, Stadt- und Raumforschung	438
33.5	Regionalisierte Daten der amtlichen Statistik	439
33.6	Mikrozensus	440
33.7	Survey-Daten des Robert Koch-Instituts	441
33.8	Das Sozioökonomische Panel	441
33.9	Fazit	442

34. Erfahrungen aus Sozialmedizin und Gesundheitsökonomie (<i>Janine Biermann, Sarah Mostardt, Gerald Lux, Jürgen Wasem, Helmut Dahl, David Matusiewicz</i>)	446
34.1 Hintergrund	446
34.2 Ausgewählte Problemfelder und Erfahrungen im Rahmen von Routinedatenanalysen.	447
34.3 Fazit	458
35. Nutzung von Sekundärdaten in der gesetzlichen Qualitätssicherung (<i>Günther Heller, Björn Borge, Joachim Szecsenyi</i>)	460
35.1 Einführung	460
35.2 Sekundärdaten und Qualitätssicherung?	461
35.3 Gesetzliche Grundlagen	462
35.4 Entwicklung eines sekundärdatenbasierten Qualitätssicherungsverfahrens im Leistungsbereich Dekubitusprophylaxe	463
35.5 Nutzung von stationären Abrechnungsdaten im Rahmen der Veröffentlichung der Ergebnisqualität für Perinatalzentren der Level 1 und Level 2 «Phase B»	467
35.6 Nutzung von Sekundärdaten zur Weiterentwicklung des Leistungsbereichs Cholezystektomie	468
35.7 Resümee und Ausblick	470
36. Vorschlag für eine standardisierte Berichterstattung von Sekundärdatenanalysen (<i>Jochen Schmitt, Enno Swart</i>)	474
36.1 Vorbemerkungen	477
36.2 Zielsetzung und Vorgehen	478
36.3 Bewertung der STROBE-Kriterien und Empfehlungen für deren Ergänzung	478
36.4 Diskussion	480
Anhang	487
Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen (3. Fassung 2012) (Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSM) und der Deutschen Gesellschaft für Epidemiologie (DGEpi))	488
Abkürzungsverzeichnis	498
Glossar	504
Autorenverzeichnis	516
Adressen der Herausgeber	526
Sachregister	527

Vorwort

Vieles ist geschehen, seit im Jahre 2005 das Standardwerk «Routinedaten im Gesundheitswesen» als «Handbuch der Sekundärdatenanalyse» erstmals erschienen ist. So viel, dass die Herausgeber der ersten Auflage es als notwendig erachten, mit dem vorliegenden Buch eine vollständig überarbeitete zweite Auflage zu präsentieren. Gleichwohl haben sich grundlegende Ziele und Aufgaben der Sekundärdatenanalyse nicht geändert, weswegen an dieser Stelle weiterhin ein Blick in das Vorwort der ersten Auflage lohnt, das anschließend folgt.

Was hat sich nun seit der Erstauflage für die Sekundärdatenanalysten geändert? Die markanteste Entwicklung im Gesundheitswesen der vergangenen Dekade, die das Interesse an Sekundärdaten maßgeblich stimuliert haben dürfte, ist wohl die drastisch gestiegene Bedeutung des Nutzenbegriffes in der gesundheitspolitischen Debatte. Aus der Perspektive des Gemeinwohls betrachtet wurden das subjektiv wahrgenommene Bedürfnis und der individuell messbare Therapieerfolg als ehemals dominante Kriterien für die medizinische Entscheidungsfindung vom objektiv(ert)en Bedarf und der kollektiven Nutzenerwartung in den Hintergrund gedrängt. Ein für die Gemeinschaft relevanter Nutzen lässt sich jedoch nur schwerlich nachweisen, wenn die Belege dafür aus quasi in-vitro durchgeführten Therapieexperimenten oder artifiziellen klinischen Studien mit wenigen Probanden extrapoliert werden müssen. Zu hohe Kosten wären in Kauf zu nehmen, zu viel Zeit wäre aufzuwenden, wollte man alle Rahmenbedingungen für einen derartigen Nutznachweis so gestalten, dass sie die Real-life-Situation des Versorgungsalltags abbildeten. Selbst wenn diese Bemühungen gelängen, der Nachweis würde doch nicht in derjenigen Population geführt, für die die Frage nach der Nutzenerreichung beantwortet werden soll. Es liegt auf der Hand, dass Daten, die aus eben dieser Population stammen und die nahezu verzerrungsfrei die tatsächliche Anwendungssituation, das «Roll-out in der rauen Wirklichkeit», widerspiegeln, eine authentische

Urteilsfindung begünstigen und somit inzwischen ein Höchstmaß an Attraktivität besitzen.

Markiert die erste Auflage des Handbuches noch einen Zeitpunkt, zu dem die Routinedaten-Community sich gleichsam in Aufbruchstimmung befand, in der ein begeisterungsfähiges, überwiegend wissenschaftlich orientiertes Publikum erstmals mit Referenzpublikationen wie der Guten Praxis Sekundärdatenanalyse (GPS) in Berührung kam, so flankiert die zweite Auflage nun eine Phase der Proliferation und des Wachstums der Sekundärdatenszene. Seit wenigen Jahren werden viele Player – v. a. diejenigen Institutionen, die selbst Routinedaten halten, insbesondere die Krankenkassen und andere Sozialversicherungsträger – sich des besonderen Wertes ihres Schatzes bewusst und leisten einen willkommenen Beitrag dazu, denselben zu heben und mit Befunden, Analysen und Interpretationen an die interessierte Öffentlichkeit heranzutreten.

Gleichzeitig erkennt die Academia, repräsentiert von den wissenschaftlichen Fachgesellschaften, die Notwendigkeit, die methodisch fundierte, hypothesengeleitete Auseinandersetzung mit diesen ganz besonderen Daten, die auf den ersten Blick gar nicht zum Prozess des traditionellen wissenschaftlichen Arbeitens von der Formulierung der Forschungsfrage über die Hypothesengenerierung und Operationalisierung bis hin zur gezielten Datenerhebung, Analyse und Interpretation passen wollen, zum Gegenstand ihres Interesses zu erklären und sich verstärkt mit der «Sekundärnutzung» bereits vorhandener Datensätze zu befassen. Dies findet seinen Niederschlag in der Gründung zahlreicher Arbeitsgruppen und Arbeitskreise der unterschiedlichsten wissenschaftlichen Disziplinen, von den theoretischen gesundheitswissenschaftlich ausgerichteten Fächern bis hin zu den klinisch-praktischen Disziplinen sowie in der Formulierung von Anleitungen und Methodenpapieren verschiedenster Provenienz. Ihr Geltungsbereich erstreckt sich von der Versorgungsforschung über gesundheitsökonomische Analysen bis hin

zur Verwendung der Daten für Modellierungen im Medical Decision Making und Health Technology Assessment.

Dabei liegt gerade in der Phasenumkehr der altehrwürdigen Regel «erst die Frage, dann die Daten» das besondere Faszinosum der Routinedatenwelt. Mit geeigneten methodischen Mitteln, verbunden mit wissenschaftlicher Kreativität, aus prä-existierenden Datensätzen Erkenntnisse heraus zu präparieren, die sich dem Unkundigen nicht auf Anhieb offenbaren, eine nahezu detektivische Akribie zu entwickeln, um Indizien für wissenschaftliche Annahmen aufzufindig zu machen, die über andere (Daten-)Zugänge nur schwer zu entdecken wären, das übt heutzutage auf eine wachsende Zahl von Forschern starke Anziehungskraft aus. Dabei die gefestigten Wege der Wissenschaft nicht zu verlassen, der Magie der großen Datensätze nicht in naiver Verirrung zu erliegen, kommt mitunter einer immensen Herausforderung gleich. Wie diesen Herausforderungen in den vergangenen Jahren begegnet worden ist, lässt sich unter anderem daran ablesen, dass sich die methodischen Ansätze und das analytische Repertoire, die heute bei der Auswertung von Routinedaten zum Einsatz kommen, im Vergleich zu den vor zehn Jahren praktizierten Vorgehensweisen an Kontur und Präzision, aber auch an Verbindlichkeit hinsichtlich ihrer Anwendung gewonnen haben. Ein Grund mehr, in der vorliegenden zweiten Auflage des Handbuches dieser Entwicklung durch Aktualisierung und Ergänzung der Methodenkapitel Rechnung zu tragen.

Begünstigt wird das zunehmende Interesse an Routinedaten dadurch, dass bereits frühzeitig angelegte Entwicklungstendenzen, wie sie etwa in das am 1. Januar 2004 in Kraft getretene Gesundheitsmodernisierungsgesetz (GMG) Einzug gehalten haben, erst jetzt, nachdem einige Jahre ins Land gegangen sind, ihre volle Wirkungskraft zu entfalten beginnen. Beispielgebend dafür sei die Datentransparenzverordnung (DaTraV) vom 10. September 2012 genannt, mit der das DIMDI die Aufgabe erhielt, das Informationssystem Versorgungsdaten (Datentransparenz) aufzubauen. 2014 sollen nun erstmals aggregierte Versorgungsdaten der gesetzlichen Krankenkassen für

Analysen des Versorgungsgeschehens, z.B. im Rahmen der Versorgungsforschung, über das DaTraV-System zur Verfügung stehen. Es dürfte spannend sein zu beobachten, welche Impulse die angestrebte Transparenz der Datennutzung tatsächlich setzt und ob sich ein neuerlicher dynamischer Schub in Richtung intensiver Nutzung von Routinedaten ergibt.

Ziel der Überarbeitung des vorliegenden Werkes ist es, derartigen Entwicklungen mit fundierten, aktualisierten Informationen und substanziellen Handreichungen zu begegnen. Das Buch richtet sich an Wissenschaftler und Entscheider in allen Bereichen des Gesundheitswesens und der Medizin (Krankenversorgung, Krankenkassen, Selbstverwaltung, Verbände, Medizintechnik- und Pharmaindustrie) sowie an Studierende und Postgraduierte gesundheitsbezogener Studiengänge wie Medizinmanagement, Management im Gesundheitswesen, Health Care Management, Gesundheitsökonomie, Public Health, Krankenhausmanagement und Pflegemanagement. Einsteiger und Umsteiger sowie Führungskräfte aus Medizin und Wirtschaftswissenschaften, Politik, Unternehmensberatungen und anderen Wirtschaftsbranchen werden gleichermaßen angesprochen.

Der Inhalt des vorliegenden Buches wird an dieser Stelle kurz skizziert:

AGENS-Familie. Als eine besondere Zusammenfassung des Buches und als Motivation für dessen Lektüre präsentieren die Herausgeber die «AGENS-Familie», eine fiktive Kurzgeschichte mit Augenzwinkern. An einzelnen Stellen wird auf diejenigen Abschnitte des Buches verwiesen, in denen die weiterführende Behandlung des erwähnten Themas stattfindet. Wir hoffen, dass die «AGENS-Familie» ein Schmunzeln bei den «alten Hasen» hervorruft und Interesse bei «jungen Hasen» weckt. Routinedaten sind viel mehr als «Einsen» und «Nullen».

Teil I: Versorgungsbereiche im Spiegel der Gesetzlichen Krankenversicherung. Im ersten Abschnitt nach der Einleitung geht es um den «Klassiker» – die Gesetzliche Krankenversicherung.

Zunächst werden Stammdaten und Versichertenhistorien der Versicherten beschrieben, da dies meist die erste Anlaufstelle für den «Routinedatenanalysten» ist. Im Folgenden erhält der Leser einen umfassenden Überblick über die Analyse der Routinedaten nach bestimmten Leistungsbereichen der gesetzlichen Krankenversicherung. Hierzu zählen die Krankenhausbehandlung, ambulante ärztliche Versorgung, zahnärztliche Versorgung, Arzneimittel und Heilmittel. Ferner werden in dem Abschnitt bestimmte Analysemöglichkeiten (Arbeitsunfähigkeiten, Disease-Management-Programme und Sterblichkeit) vorgestellt.

Teil II: Versorgungsbereiche im Spiegel anderer Sozialversicherungsträger. Wer das Wort Sozialdaten im Zusammenhang mit Sekundärdatenanalysen hört, denkt oftmals nur an Routinedaten der Gesetzlichen Krankenversicherung. Doch auch andere Sozialversicherungsträger halten Sozialdaten aus verschiedenen Versorgungsbereichen für wissenschaftliche Analysen bereit. Diese werden in einem separaten Abschnitt vorgestellt. In vier Beiträgen werden Daten zum Pflegegeschehen, zur Rehabilitation, zu individuellen Erwerbsverläufen und zu Erwerbsminderungsrenten beschrieben, die von den Pflegekassen, den gesetzlichen Rentenversicherungsträgern und dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit gehalten werden.

Teil III: Ergänzende Routinedaten mit Gesundheitsbezug. Wie bereits erwähnt, stammen Routinedaten nicht nur von den Krankenversicherungen, sie können auch aus Gesundheitsbefragungen, Surveys, Registern oder anderen aggregierten Datensätzen stammen. Diesem Themenfeld, einschließlich der Möglichkeiten und Grenzen solcher Daten, widmet sich der Abschnitt «Ergänzende Routinedaten mit Gesundheitsbezug». Unter den Krankenversicherungsdaten gibt es nicht allein die Daten der Gesetzlichen Krankenversicherung (GKV), sondern auch Daten der Privaten Krankenversicherung (PKV); hierzu findet sich ebenfalls ein Beitrag in diesem Abschnitt. Auch fallen Routinedaten nicht allein in Deutschland an, sondern entstehen ebenso im europäischen

Ausland und in Übersee. Sie werden dort, ähnlich wie bei uns, wissenschaftlich ausgewertet, und die so gewonnenen Erkenntnisse halten Einzug in die Gesundheitsberichterstattung, in die sogenannte «Bedarfsplanung» (besser: Angebots- oder Strukturplanung), in die Qualitätssicherung, die Versorgungsforschung und Versorgungssteuerung usw. Internationale Datenbanken können Referenzcharakter für die deutsche Sekundärdatenforschung haben; sie zu kennen, kann sich daher auch für den deutschen Sekundärdatenforscher als nützlich erweisen.

Teil IV: Datenzugang und Datenvalidierung. Ein weiterer Abschnitt mit sechs Beiträgen thematisiert die Aspekte Datenzugang und Datenvalidierung und damit wichtige Arbeitsschritte bei der Aufbereitung der Sekundär-Rohdaten bis zu einem geprüften und wissenschaftlich nutzbaren Datensatz. Es werden Datenportale jenseits der Sozialversicherungsträger beschrieben, technische Aspekte der Datenaufbereitung und der Datenverwaltung von Routinedaten dargestellt und der wichtige Schritt der Datenvalidierung der ursprünglich zu anderen Zwecken erhobenen Daten wird thematisiert. In diesem Zusammenhang erhalten Klassifikationssysteme in der Medizin eine besondere Bedeutung. Schließlich sind Sekundärdatenanalysen nicht denkbar ohne Beachtung spezifischer Datenschutzbestimmungen und -vorkehrungen. Besondere Relevanz erhalten diese technischen und rechtlichen Aspekte beim zunehmend wichtiger werdenden Linkage verschiedener Datenkörper.

Teil V: Routinedaten als Informationsquelle für Morbiditätsschätzungen. Die vier eng mit einander verzahnten Beiträge dieses Abschnitts beschäftigen sich jeweils aus unterschiedlicher Sicht mit der Bestimmung von Inzidenz und Prävalenz. In den ersten beiden Beiträgen wird aufgezeigt, wie sich erkrankte Fälle in Routinedaten definieren lassen und welche abrechnungstechnischen Rahmenbedingungen hierbei zu beachten sind. Der dritte Beitrag widmet sich dem Thema der Population unter Risiko. Abschließend wird an Beispielen das breite Spektrum des Themas Morbiditätsmessung diskutiert.

Teil VI: Spezielle Methoden, Instrumente und Anwendungen der Sekundärdatenanalyse. Die Beiträge dieses Kapitel zeigen, welche spezifischen Aspekte, Herausforderungen und Chancen sich ergeben, wenn etablierte Instrumente aus unterschiedlichen Fachdisziplinen auf Routinedaten treffen. Abgerundet wird das Kapitel mit der Vision eines Standards bei der Publikation von Ergebnissen und Erkenntnissen aus Routinedatenanalysen.

Zum Schluss möchten wir uns bedanken bei allen, die maßgeblich zu diesem Werk beigetragen haben. Dazu gehören in erster Linie die Kolleginnen und Kollegen aus der AGENS und alle weiteren Autoren, die die einzelnen Kapitel dieses Buches mit Sachverstand und Engagement neben ihrer Hauptarbeit erstellt haben. Weiterhin gebührt Dank Herrn Dr. Reinhardt vom Verlag

Hans Huber, der uns zu dieser Neuauflage ermuntert hat. Und schließlich wurde diese durch finanzielle Zuschüsse der Deutschen Gesellschaft für Epidemiologie (DGEpi) und der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) überhaupt erst ermöglicht.

Wir wünschen eine gewinnbringende Lektüre, die bei der Arbeit mit Routinedaten im Gesundheitswesen unterstützen soll und hoffentlich zum «Licht am Ende des Datentunnels» beitragen wird.

Im Mai 2014

Enno Swart, Magdeburg
Peter Ihle, Köln
Holger Gothe, Dresden und Hall in Tirol
David Matusiewicz, Essen

Vorwort der ersten Auflage

Sekundärdatenanalyse: Aufgaben und Ziele

Enno Swart, Peter Ihle

Die Versorgungsforschung hat in den letzten Jahren erheblich an Bedeutung gewonnen. Augenfalliger Ausdruck dieser Entwicklung ist der seit 2002 jährlich stattfindende Kongress für Versorgungsforschung, der derzeit von rund 30 wissenschaftlichen medizinischen Fachgesellschaften getragen wird. Daran lässt sich die kontinuierlich zunehmende Ausbreitung einer bevölkerungsbezogenen Evaluation des Versorgungsgeschehens ablesen. Auch die öffentliche Förderung der Versorgungsforschung im gemeinsamen Förderprogramm des Bundesministeriums für Gesundheit und Soziale Sicherung und der Spitzenverbände der Gesetzlichen Krankenversicherung kann als Beleg für diese Entwicklung angeführt werden.

Woraus erklärt sich diese Entwicklung? Immer noch stellt sich das tägliche Versorgungsgeschehen in Praxen, Krankenhäusern und anderen Einrichtungen unseres komplexen Gesundheitssystems bezüglich ihrer Kontaktpunkte, diagnostischen und therapeutischen Leistungen oder Ergebnissen häufig als eine Black-Box dar, trotz aller Datenmengen, die routinemäßig im Zuge millionenfacher Versorgungskontakte entstehen. Um unter finanziell enger werdenden Rahmenbedingungen der GKV weiterhin allen Versicherten eine hochwertige medizinische Versorgung anbieten, eine Bewertung der vielfach häufig unreflektiert angewandten Leistungen vornehmen oder spezifischen Versorgungsbedürfnissen einzelner Bevölkerungs- oder Patientengruppen besser gerecht werden zu können, bedarf es vermehrter Anstrengungen für eine höhere Transparenz des Versorgungsgeschehens. Diese beispielhaft aufgezählten Fragestellungen belegen den umfangreichen und vielfältigen Datenbedarf zur Beantwortung versorgungsrelevanter Fragestellungen. Ohne die (Sekundär-)Nutzung bereits vorhandener Informationen wäre dieser Bedarf mit Primärerhebungen kaum zu

vertretbaren Kosten und in absehbarer Zeit zu stillen. Die an vielen Stellen vorhandenen Routinedaten unseres Sozialversicherungssystems bieten sich als kaum erschöpfbare Datenquelle geradezu an, werden aber trotz steigender Tendenz derzeit nur vereinzelt und wenig systematisch genutzt.

Die in diesem Buch vorgestellten Methoden, Probleme und Lösungsansätze sowie Beispiele der Sekundärdatenanalyse sollen auf eine breite Palette nutzbarer Datenquellen, Methoden und Auswertansätze aufmerksam machen, die den genannten Herausforderungen gerecht werden können. Dabei ist die Sekundärdatenanalyse – primär mit Daten der Gesetzlichen Krankenversicherung –, aber zunehmend auch mit Daten anderer Sozialversicherungsträger nichts Neues. Schon vor etwa 25 Jahren gab es erste konzeptionelle Überlegungen zur Nutzung derartiger Daten. Auch eine Arbeitsgruppe der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSM) hatte 1997 ihre Vorstellungen und Erfahrungen zum Umgang mit Gesundheits- und Sozialdaten formuliert. Verbesserte technische Möglichkeiten der Datengewinnung und -verarbeitung, neue gesetzliche Rahmenbedingungen, erweiterte inhaltliche Fragestellungen und ein zunehmendes Interesse an der Nutzung von Prozessdaten gerade auch außerhalb der Wissenschaft legen die Vorlage eines neuen und erweiterten Buches zur Sekundärdatenanalyse mit Gesundheitsdaten nahe.

Sekundärdatenanalyse, also die Nutzung von Daten im Rahmen wissenschaftlicher und/oder praxisrelevanter Untersuchungen ohne direkten Bezug zum primären Erhebungsanlass, wird von den Herausgebern und den Autoren dieses Buches dabei nicht als Zweite-Klasse-Wissenschaft verstanden, sondern als eigenständiges Forschungsfeld, das in einem spezifischen inhaltlichen und strukturellen Kontext spezielle Methoden bereit

hält und gegenüber der Primärdatenforschung gleiche Wertigkeit besitzt. Natürlich muss sich die Sekundärdatenanalyse mit diesem Anspruch auch allgemein anerkannten wissenschaftlichen Standards stellen. Dazu gehört auch die Gute Praxis Sekundärdatenanalyse (GPS), die im Vorfeld dieser Publikation veröffentlicht und zur wissenschaftlichen Diskussion gestellt wurde. An der GPS werden Anknüpfungspunkte zur traditionellen Epidemiologie ebenso erkennbar wie ihre speziellen methodischen und strukturellen Besonderheiten.

Routinedaten der Sozialversicherungsträger bieten sich zur Nutzung aus einer Vielzahl von Gründen an: Sie stehen im Prinzip jederzeit zur Nutzung bereit, sind daher kostengünstig und schnell verfügbar. Wegen ihrer Generierung im Zuge der Leistungserbringung bzw. Kostenerstattung sind sie aktueller als viele andere routinemäßig verfügbare Daten. Weitere herausragende Eigenschaften sind ihr eindeutiger Personen- bzw. Bevölkerungsbezug, ihre Vollständigkeit, die leistungserbringerübergreifende Perspektive oder die Möglichkeit längsschnittlicher Analysen, um nur einige Vorteile zu nennen.

Dass Sekundärdaten – wie jede andere Datenquelle auch – ihre Nachteile haben, soll nicht unterschlagen werden. Häufig geäußerte Kritikpunkte sind etwa das Fehlen klinischer Informationen, unzureichende Validität bezüglich der Klassifikation von Diagnosen und Prozeduren oder Verzerrungen durch die Analyse singulärer Datenkörper einzelner Kostenträger. Dieser Kritik wird in den Einzelbeiträgen durch das Aufzeigen von Lösungsansätzen bei derartigen Problemen begegnet. In dem Maße, wie bezüglich dieser Daten die Vertrautheit, das methodische Verständnis sowie das Instrumentarium wachsen, werden Chancen und Potenziale dieser Datenquellen in realistischer Abschätzung ihrer Grenzen erkennbar.

Das Interesse und das Verständnis für den Umgang mit Sekundärdaten will dieses Buch wecken bzw. fördern. Es richtet sich einerseits an die zunehmend größer werdende Gruppe von Wissenschaftlern im Bereich der Versorgungsforschung und möchte ihnen methodisches Rüstzeug und einen Orientierungsrahmen für die Nutzung dieser

Daten anbieten. Es möchte damit in dem Maße Standards definieren, wie sich allgemeine Prinzipien aus dem Umgang mit Sekundärdaten ableiten lassen. Darüber hinaus soll eine Palette von methodischen Herangehensweisen und Handwerkszeugen dargestellt werden, die in konkreten Nutzungssituationen motivieren und unterstützen sollen.

Andererseits will dieses Buch bewusst die Brücke zwischen Wissenschaft und Praxis schlagen. An dieser datenbezogenen Schnittstelle zwischen Wissenschaft und Praxis scheinen uns die Voraussetzungen für einen Brückenschlag besonders günstig zu sein. Das Buch richtet sich neben Forschern daher auch an Fachleute aus der (Versorgungs-)Praxis und an potenzielle Nutzer von Sekundärdatenanalysen. Wenn in diesem Buch von Versorgungsforschung gesprochen wird, heißt das nicht, dass die Sekundärnutzung von Prozessdaten der Sozialversicherungsträger und anderer Institutionen allein der Wissenschaft vorbehalten ist. Es soll lediglich hervorgehoben werden, dass bei der Nutzung dieser Daten wissenschaftliche Standards und Gepflogenheiten zu beachten sind, was letztendlich der Qualität der Ergebnisse und ihrer Umsetzung in die Praxis zugutekommt.

Die neuen gesetzlichen Regelungen in § 303 des Sozialgesetzbuches V zur Datentransparenz, auf die im einzelnen in diesem Buch noch eingegangen wird, eröffnen Chancen für eine breitere Nutzung von Routinedaten der gesetzlichen Krankenkassen. Allerdings bedürfen die in diesem Paragraphen fixierten Rahmenbedingungen noch der konkreten Ausgestaltung. Die Wissenschaft ist zwar als potenzieller Nutzer der noch im Einzelnen festzulegenden Datenkörper genannt, jedoch nicht als Partner bei der Umsetzung des Gesetzes. Dieses Buch will daher neben den genannten Zielen bei den Selbstverwaltungsorganen die Erkenntnis für die Notwendigkeit der Einbeziehung wissenschaftlicher Expertise und der Beachtung methodischer Aspekte beim Aufbau des Datenkörpers fördern. Das Buch zeigt ebenso, dass datenschutzrechtliche Fragen im Zusammenhang mit der Nutzung von Prozessdaten der Sozialversicherungsträger lösbar sind und kein grundsätzliches (und in der Vergangenheit oft vorgebracht-

tes) Argument gegen die Nutzung dieser Daten darstellen.

Das vorliegende Buch trägt den Untertitel «Handbuch der Sekundärdatenanalyse». Ein Handbuch zeichnet sich üblicherweise durch eine umfassende und stringente Darstellung seines Gegenstandes aus. Bei der Arbeit an diesem Buch haben wir jedoch gemerkt, dass gerade auf Gebieten außerhalb der gesetzlichen Krankenversicherung noch zahlreiche inhaltliche und methodische Fragen bei der Nutzung von Routinedaten der Bearbeitung bedürfen, die hier nur angerissen werden konnten. Ein vollständiger Überblick über Datenkörper, Methoden und Werkzeuge bezüglich der Routinedaten im Gesundheitswesen hätte daher einen erheblich längeren Entstehungszeitraum erfordert.

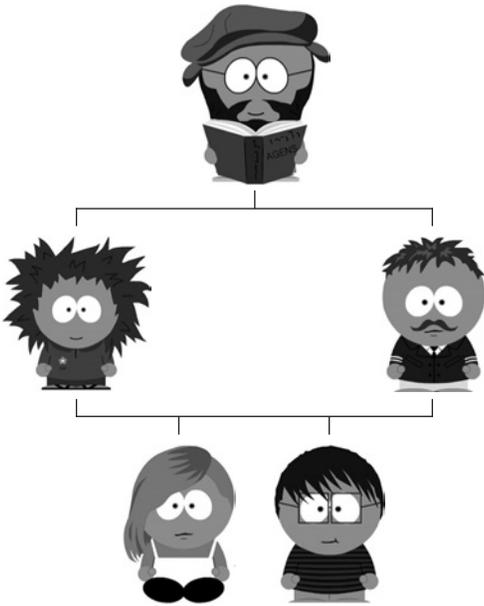
Aus diesem Grund und angesichts eines sich sehr dynamisch entwickelnden Forschungsfeldes, auf dem die Erkenntnisse von heute schnell die überholten Fragestellungen von gestern sind, zugunsten einer besseren Lesbarkeit und höheren Anschaulichkeit und nicht zuletzt zugunsten einer schnellen Veröffentlichung im Rahmen des diesjährigen Kongresses zur Versorgungsforschung haben wir daher den Anspruch eines umfassenden Handbuchs fallen lassen. Entstanden ist dennoch – so meinen wir – ein vielseitiges Werk gut lesbarer, methodisch fundierter und auf dem jahrelangen Erfahrungswissen der Autoren aufbauender Einzelbeiträge für einen großen und breiten Kreis potenzieller Leser. Wir hoffen, mit diesem Buch methodische Orientierung und vielfältige Hilfestellung für Wissenschaft und Praxis leisten zu

können. Daher die Beibehaltung des Begriffs «Handbuch» im Untertitel. Es wird zukünftige Aufgabe aller Akteure auf diesem Gebiet sein, an den offenen und in diesem Buch angesprochenen Fragen der Sekundärdatenanalyse weiter aktiv zu arbeiten.

Die Herausgeber möchten zu guter Letzt allen Personen und Institutionen danken, die zum Gelingen dieses Buches beigetragen haben. In erster Linie sind hier alle Autoren zu nennen, die die Arbeit an diesem Buch zusätzlich zu ihren normalen Aufgaben bereitwillig und engagiert übernommen haben. Viele dieser Autoren kommen aus der AGENS (Arbeitsgruppe zur Erhebung und Nutzung von Sekundärdaten). Diese Arbeitsgruppe der DGSMMP hat durch offene und konstruktive Diskussionen erst die Idee und die Grundlage zu diesem Werk gelegt. Am kontinuierlichen Wachsen dieser Arbeitsgruppe haben wir ein breites Interesse an einem orientierenden und zusammenfassenden Werk zur Sekundärdatenanalyse erkannt und die Motivation zu diesem Gemeinschaftswerk bezogen. Die DGSMMP gibt der AGENS seit Jahren eine wissenschaftliche Heimat und hat konkret durch fachliche und finanzielle Unterstützung geholfen. Dieses Buch ist Dank und gleichzeitig ein kleiner Beitrag zum 100-jährigen Gründungsjubiläum der Vorgängergesellschaft der DGSMMP, der Gesellschaft für soziale Medizin, Hygiene und Medicinalstatistik, das im September 2005 begangen wird. Schließlich sei dem Verlag Hans Huber, speziell Herrn Dr. Reinhardt, für die Offenheit gegenüber diesem Projekt und dessen technische Umsetzung gedankt.

AGENS – eine schrecklich routinierte Familie¹

David Matusiewicz, Holger Gothe, Peter Ihle, Enno Swart



Großvater Anton AGENS ist ein pensionierter Forscher, der nach dem Tod seiner Frau viel Zeit hat. Heute Morgen hat er die gedruckte Form der neuesten Ausgabe des AGENS-Newsletters mit der Post erhalten. Nach dem obligatorischen Spaziergang durch den angrenzenden Forst hat er sich zum 5-Uhr-Tee in sein Arbeitszimmer zurückgezogen und studiert interessiert die Kurzfassungen der neuesten Projektergebnisse und Ankündigungen anstehender Termine.

Die klassischen Rentneraktivitäten wie Kartenspielen und die Pflege eines Schrebergartens sind nichts für ihn. Und Zeit ist nicht das einzige, was er besitzt. Seine Frau hatte ihm überraschend ein kleines Vermögen vererbt, das aus den Verwertungslizenzen eines Patents ihres verstorbenen Bruders stammte, der in den 1920er-Jahren nach Amerika auswanderte, aber Mitte der 1980er-Jahre alleinstehend ohne direkte Nachfahren verstarb. Und so war es eine Fügung des Schicksals, als ihm eines Tages zwei Artikel in die Hände fielen, die sein Le-

ben noch einmal verändern sollten: Ein Artikel berichtete über die Quantified-Self-Bewegung aus San Francisco, deren Anhänger sich das Ziel gesetzt haben, die eigenen Gesundheitsdaten zu erfassen und zu analysieren. Und in einem zweiten Artikel wurde beschrieben, welche Möglichkeiten die Analyse von Routinedaten der Sozialversicherungsträger bieten. Anton hatte plötzlich eine Idee.

Die Familie AGENS war ja schon immer etwas Besonderes. Die Familienmitglieder legen seit langem großen Wert auf Routine in ihrem Alltag. Ungeplante Aktivitäten und Spontanität sind da eher fehl am Platz. Ob es sich um Einkäufe, den Arbeitsalltag oder die Buchung des nächsten Urlaubes dreht: Am liebsten läuft alles nach Plan – so, wie es schon immer war. Es gibt einen standardisierten Einkaufszettel, einen disziplinierten Zeitmanagementplan für die Arbeitswelt, und der Urlaubsort ist schon seit Jahrzehnten der gleiche Campingplatz an der Nordsee. Er ist durch die Familie erprobt, man will ja schließlich kein Risiko eingehen. Aber etwas anderes war Anton auch schon aufgefallen: Jeder dieser Familienmitglieder zeigt Auffälligkeiten in seinem Gesundheits-, besser in seinem Krankheitsverhalten. Aber Anton konnte diese Auffälligkeiten, wie er es als Forscher gewohnt war, an keinem objektiven Maßstab messen, er wusste nicht, wo das normale Mittelmaß lag.

Aber Routinedaten könnten doch geeignet sein, ihm die Antwort zu geben, nach denen er so lange gesucht hatte. In diesem Zusammenhang stellte sich Anton die Frage, wie er einen Zugang zu den Routinedaten der medizinischen Versorgung erhalten könne. Das gestaltete sich schwerer als gedacht.² Die Daten der Krankenversicherung oder

¹ Dies ist eine fiktive Familie, die rein in der Phantasie der Herausgeber existiert. Die angesprochenen Themen enthalten keine Wertigkeit zu einzelnen Aspekten.

² Kap. 20 «Zugang zu Routinedaten»

der anderen Sozialversicherungsträger sind schließlich «sensible» Daten. Und schnell war klar, dass er als Privatperson keine Chancen hatte. Und ihm wurde weiterhin klar, dass nur ein eigenständiges Forschungsinstitut die notwendigen Strukturen vorhalten kann, in dessen Rahmen eine solche Aufgabe gestemmt werden kann. Kurzerhand steckte er die Erbschaft seiner Frau in eine Stiftung und gründete das Routinedaten-Exzellenz-Zentrum AGENS (kurz A-Team; das «A» stand sowohl für Arbeitsgruppe als auch für deren Galionsfigur «Anton»).

Anton hat gute Kontakte zu alten Freunden, die zufällig alle Versorgungsbereiche der Sozialversicherungsträger kennen und dem A-Team auf unterschiedlichen Wegen einen Zugang zu den pseudonymisierten Gesundheitsdaten verschafften. Einige Daten können vom Institutsrechner über einen eigens dafür eingerichteten Zugang analysiert werden, andere werden transportverschlüsselt auf CD zugesendet.³ So erarbeiteten sich die AGENS-Forscher nach und nach etwas über den Datenaufbau und die Datenaufbereitung⁴, indem sie langsam verstanden, wie man Zahlen in den Tabellen in immer größer werdende Matrizen anordnen kann. Dabei fiel ihnen auf, dass gerade das Thema Datenqualität⁵ eine wichtige Rolle spielt. Die Validität von Routinedaten ist immer sorgfältig zu überprüfen, damit gute Ergebnisse erzielt werden, die auch tatsächlich Auskunft über die Gesundheit geben.

Das A-Team nutzte jetzt auch Ordnungssysteme, die auf dem Prinzip einer Klassenbildung beruhen. Das macht AGENS, um mehrere Ausprägungen einer Variablen in einer geringeren Anzahl von Klassen zusammenzufassen und so die Komplexität der Daten zu reduzieren.⁶ In einem großen Daten-Modell sollen schließlich sowohl die verschiedenen Sekundärdatenquellen über geeignete Schlüsselvariablen verknüpft werden, als auch weitere Primärquellen (Befragungen) mit in die Ergebnisse einfließen.⁷

Schließlich wurde Anton von seinen Freunden bei den Sozialversicherungsträgern immer wieder darauf hingewiesen, welch hohen Stellenwert der Datenschutz bei der Nutzung von Routinedaten im Rahmen wissenschaftlicher Forschung besitzt

und welche umfassenden datenschutzrechtlichen Vorgaben zu beachten sind.⁸ Nicht nur die Verwendung einzelner Daten, sondern auch die Verknüpfung unterschiedlicher Datenquellen muss er daher unbedingt datenschutzkonform erledigen. Das bedeutet für das A-Team einen umfassenden organisatorischen und zeitlichen Aufwand, der bei der Planung wissenschaftlicher Untersuchungen berücksichtigt werden muss. Aber Anton ist ja glücklicherweise Rentner und hat diese Zeit. Er entwickelt mit dem A-Team ein ausgeklügeltes Datenschutz- und Datensicherheitskonzept. Ihm ist klar, dass dies ein wichtiger Aspekt ist, denn er möchte schließlich auch, dass Dritte vertrauensvoll mit den Gesundheitsdaten umgehen.

Endlich fängt nun nach umfangreichen Vorarbeiten die eigentliche Arbeit für Anton und sein A-Team an! Zunächst zeichnet er sich einen Ahnenbaum und versucht über alte Dokumente, die Familiendaten und -historien der Vorfahren zusammenzutragen. Das erste Blatt seines Blockes füllen somit die Stammdaten⁹ der Familienmitglieder. Nachdem Anton den Stammbaum bis zum Anfang des 19. Jhd. aufgeschrieben hat, widmet er sich zunächst seiner verstorbenen Frau **Aurelia AGENS**. Zur Analyse der Sterblichkeit¹⁰ nimmt er die Todesursachenbescheinigung, d.h. den Totenschein seiner Frau zur Hand, der vom Standesamt erfasst wurde. Anton lässt durch das A-Team das Alter seiner Frau mit dem Verlauf der Sterblichkeit in den Daten der Renten- und Krankenversicherung vergleichen. Die Kurven erinnern ihn an etwas. Er schaut nach unten auf seine Schuhe und sieht ein Bumerang-Symbol, das dem Kurvenverlauf sehr ähnelt. Nach dem Risiko während und nach der Geburt sinkt also die Sterberate auf einen Minimalwert und im Anschluss steigt der

3 Das geht in der Realität nicht so einfach!

4 Kap. 22 «Datenaufbereitung und Datenaufbau»

5 Kap. 23 «Prüfung der Datenqualität und Validität von GKV-Routinedaten»

6 Kap. 24 «Klassifikationssysteme»

7 Kap. 25 «Datenlinkage»

8 Kap. 21 «Datenschutzrechtliche Aspekte bei der Nutzung von Routinedaten»

9 Kap. 1 «Stammdaten und Versichertenhistorien»

10 Kap. 2 «Analysen zur Sterblichkeit»

Kurvenverlauf wieder kontinuierlicher an, am Anfang allerdings sehr langsam. In jungen Jahren sind es wohl eher Unfälle, Suizide und Morde, die zum Tod führen, während mit zunehmendem Alter offenbar erst akute und dann chronische Krankheiten ausschlaggebend sein mögen. Nachdem Anton lange Zeit die Bumerang-Kurve betrachtet hat, ist er froh, das Aurelia mit ihren 80 Jahren ein «goldenes» Alter erreicht hatte und die Bumerang-Kurve ziemlich weit erklommen hat. Die Risikoaversion der Familie hatte sich da wohl bereits ausgezahlt. Aus den Pflegeleistungen, die seine Frau zuletzt noch bekam, kann er die Pflegestufe, die einzelnen Leistungen und die Zeiträume der Leistungserbringung erkennen. Die Messung von allgemeinen Pflegeinzidenzen und Pflegeprävalenzen gelingt dem A-Team über die Verknüpfung mit den Versicherungszeiten der Krankenkassen. Anders als sie es mit den amtlichen Statistiken zur Pflegebedürftigkeit gekonnt hätten, ist es ihnen gelungen, mit den Routinedaten der Pflegekassen Pflegeverlaufsdarstellungen darzustellen.¹¹ Als er sich die aggregierten Gesundheitsdaten ansieht, die auch seine verstorbene Frau charakterisierten, wird ihm etwas warm ums Herz, und ein leicht abgewandeltes Goethe-Zitat kommt ihm in den Sinn: «Was man tief im Herzen trägt [und in Sekundärdaten über Ewigkeiten schwarz auf weiß besitzt], kann man nicht durch Tod verlieren.»



Anton widmet sich im Anschluss nun dem nächsten Familienmitglied. Da ist zunächst seine Tochter Agata AGENS, die sich im mittleren Alter befindet und regional in Berlin verwurzelt ist. Anton hat schon immer die These vertreten, dass Agata im Sinne Molières eine eingebildete Kranke ist und viele ihrer medizinischen Aktivitäten auf ihre Hypochondrie zurückzuführen sind. Da Agata ihm am Telefon immer wieder von ihren mannigfaltigen Arztbesuchen erzählt hatte, bittet Anton das A-Team, in den Routinedaten nach der Gruppe der High-user zu forschen. Arztbesuche lassen sich aus den Abrechnungsscheinen der ambulant-ärztlichen Versorgung erkennen.¹² Hier

sieht das A-Team, welche Ärzte aufgesucht, welche Diagnosen von diesen gestellt und welche Leistungsziffern (sog. EBM-Ziffern) an welchen Tagen in Rechnung gestellt wurden. Das sind die Daten, die von den ambulant tätigen Vertragsärzten quartalsweise mit den regional für sie zuständigen Kassenärztlichen Vereinigungen abgerechnet und an die jeweils zuständige Krankenkasse weitergeleitet werden. Bei der Gruppe der High-user finden sich Versicherte, die im vergangenen Jahr rund 300 Diagnosen aufwiesen, was nach Recherchen des A-Teams das 10-fache des Durchschnitts darstellt. In diesem Zusammenhang erinnert sich Anton, dass Agata auch in mehreren strukturierten Behandlungs-Programmen (Disease Management Programme)¹³ für chronisch kranke Patienten eingeschrieben ist. Sie zählt also zu einer Gemeinschaft von mehreren Millionen Versicherten, die an mindestens einem Programm für Brustkrebs, Diabetes mellitus, Koronare Herzkrankheit (KHK), Chronisch-obstruktive Lungenerkrankung (COPD) oder Asthma bronchiale teilnehmen. «Be part of it», so lautete schon immer Agatas Devise, und so ist sie gleich in allen Programmen Mitglied. Viel hilft viel, könnte auch Agatas Leitspruch lauten.

Was Anton aber am meisten gegen den Strich geht, ist die Tatsache, dass Agata in jedem Döschen in der Küche ein paar bunte Pillen lagert und diese wie Bonbons konsumiert. Das hat er bei seinem letzten Berlin-Besuch beobachtet. Er staunt nicht schlecht, als er sich den überdurchschnittlichen Verbrauch der Arzneimittel¹⁴ in den Krankenkassendaten grafisch darstellen lässt. Über Pharmazentralnummern (PZN) der Medikamente kann das A-Team diese einzelnen Wirkstoffen zuordnen. Zur Messung des Verbrauches bestimmen sie außerdem Tagesdosen (daily defined doses, DDD). Anton's Team recherchiert den durchschnittlichen Arzneimittelverbrauch je Versichertem und kommt zu dem Schluss, dass es Versi-

11 Kap. 10 «Pflegeleistungen nach SGB XI»

12 Kap. 3 «Ambulante ärztliche Versorgung»

13 Kap. 4 «DMP-Programm»

14 Kap. 5 «Arzneimittelverordnungen»