

Agenda

- Allgemeine Informationsquellen
- Datenzugang
- Aufbereitung/Plausibilisierung

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021





Autor: **Swart, Enno / Ihle, Peter / Gothe, Holger / Matusiewicz, David (Hrsg.)**

Routinedaten im Gesundheitswesen

Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und
Perspektiven

2. vollst. überarb. u. erw. Aufl. 2014. 536 S., 52 Abb., 49 Tab., Gb

Themen: **Methoden der Forschung / Grundlagen und Methoden /**

Info: 536 S.

Erschienen: 7. Oktober 2014

ISBN: 9783456854359

Preis: EUR 39.95 / CHF 53.90

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Swart E et al. Gute Praxis Sekundärdatenanalyse (GPS): ... Gesundheitswesen 2015; 77: 120–126

Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen

Good Practice of Secondary Data Analysis (GPS): Guidelines and Recommendations

3. Fassung; Version 2012/2014*
Third Revision 2012/2014

Autoren E. Swart¹, H. Gothe², S. Geyer³, J. Jaunzema³, B. Maier⁴, T. G. Grobe⁵, P. Ihle⁶

Institute

¹ Institut für Sozialmedizin und Gesundheitsökonomie, Med. Fakultät,
Otto-von-Guericke-Universität Magdeburg, Magdeburg

² UMIT – Private Universität für Gesundheitswissenschaften, Medizinische
Informatik und Technik GmbH, Department of Public Health & Health
Technology Assessment, Hall i.T., Austria; Medizinische Fakultät „Carl Gustav
Carus“ der Technischen Universität Dresden, Lehrstuhl für Gesundheitswis-
senschaften, Dresden

³ Forschungs- und Lehrstuhl Medizinische Soziologie, Medizinische
Hochschule Hannover, Hannover

⁴ Berliner Herzinfarktregister e.V. am Fachgebiet Management im Gesund-
heitswesen, TU Berlin, Berlin

⁵ Institut für angewandte Qualitätsförderung und Forschung im
Gesundheitswesen, AQUA-Institut, Göttingen

⁶ PMV forschungsguppe, Klinik und Poliklinik für Psychiatrie und
Psychotherapie des Kindes- und Jugendalters der Universität zu Köln, Köln

- › Zusammenfassung
- › Abstract
- › Ziele der GPS
- › Zielgruppe der GPS
- › Methodik und Verfahren der Aktualisierung
- › Leitlinie 1: Ethik
- › Leitlinie 2: Forschungsfrage
- › Leitlinie 3: Studienplan
- › Leitlinie 4: Probendatenbanken
- › Leitlinie 5: Qualitätssicherung
- › Leitlinie 6: Datenaufbereitung
- › Leitlinie 7: Datenanalyse
- › Leitlinie 8: Datenschutz
- › Leitlinie 9: Vertragliche Rahmenbedingungen
- › Leitlinie 10: Interpretation und wissenschaftliche Publikation
- › Leitlinie 11: Kommunikation und Public Health
- › Begriffsdefinitionen
- › Literatur

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021





Datengutachten für das
Deutsche Institut für Medizinische Dokumentation
und Information (DIMDI)

Gutachten: Daten für die Versorgungsforschung,
Zugang und Nutzungsmöglichkeiten

AutorInnen

Dr. Ingrid Schubert (PMV)
Peter Ihle (PMV)
Ingrid Köster (PMV)
Dr. Jutta Küpper-Nybelen (PMV)
Melanie Rentzsch (ISMG)
Christoph Stallmann (ISMG)
Dr. Enno Swart (ISMG)
Carolin Winkler (ISMG)

Das Kapitel »Informationssystem Versorgungsdaten« wurde von Dr. Jochen Dreß und Dr. Michael Schopen (DIMDI) erstellt.

Korrespondenzanschriften

Dr. Ingrid Schubert
PMV forschungsguppe an der KJP, Universität zu Köln
Herderstraße 52-54, 50931 Köln
Tel. 0221-4786545, Ingrid.Schubert@uk-koeln.de

Kooperationspartner

Dr. Enno Swart
Institut für Sozialmedizin und Gesundheitsökonomie
(ISMG; Leiter Prof. Dr. B.-P. Robra, M.P.H.)
Med. Fakultät, Otto-von-Guericke Universität Magdeburg
Leipziger Straße 44, 39120 Magdeburg
Tel. 0391-6724306; Enno.Swart@med.ovgu.de

Danksagung

Die Autoren danken Frau Andrea Waltersbacher (WIdO) sowie wie Frau Ulnke Nimptsch (TU Berlin) für unterstützende Hinweise zur Abfassung des Gutachtens.

Vorbemerkung

Zur besseren Lesbarkeit des Textes wird die männliche Form (z. B. Arzt, Patient) verwendet. In diese Bezeichnung sind jedoch Männer und Frauen gleichermaßen eingeschlossen.

Köln, 12. Juli 2014; letztes Update 08. Oktober 2014

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Datenzugang

Wo sind *versichertenbezogene* GKV-Daten verfügbar?

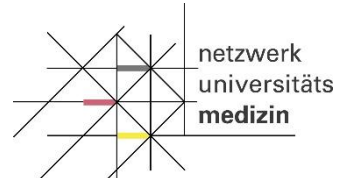
- Krankenkassen
- Wissenschaftliche Institute der Krankenkassen (WIdO, InGef, bifg)
- Apothekenabrechnungszentren
- Kassenärztliche Vereinigungen
- Zentralinstitut der Kassenärztlichen Vereinigungen
- Daten der Datentransparenzverordnung nach SGB V § 303a-f im FDZ Sozialdaten (BfArM)
- §21er Daten bei den FDZ/Statistikämter der Bundes/der Länder
- Forschungseinrichtungen, z. B.
Gepard-Datenbank im BIPS (Bremen)
Versichertenstichprobe Hessen, PMV forschungsguppe (Köln)
- G-BA (Gemeinsame Bundesausschuss ist das höchste Beschlussgremium der gemeinsamen Selbstverwaltung im deutschen Gesundheitswesen)

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021...



Datenzugang

Nutzung auf Antrag

- DaTraV-Daten beim FDZ Sozialdaten im BfArM (ab 1. Jan. 2022)
- Zentralinstitut (Scientific use file, 10% Stichprobe)
- Daten des G-BA

Nutzung vorhandener Forschungsdaten auf Basis projektbezogener Vereinbarung

- Wissenschaftliche Institute der Krankenkassen: WIdO, InGef, bifg
- Forschungseinrichtungen (BIPS, PMV, CoRe-Net)

Datennutzung im Rahmen von auf Basis projektbezogener Vereinbarung auf Basis des §75 SGB X

- Daten einer oder mehrerer Krankenkassen

Datenzugang

Zugang: § 75, SGB X

Übermittlung ist zulässig (...) für (...) Forschung oder Planung im Sozialleistungsbereich durch eine öffentliche Stelle im Rahmen ihrer Aufgaben (...) oder das öffentliche Interesse an der Forschung oder Planung das Geheimhaltungsinteresse des Betroffenen erheblich überwiegt.

Eine Übermittlung ohne Einwilligung des Betroffenen ist nicht zulässig, soweit es zumutbar ist, die Einwilligung des Betroffenen (...) einzuholen oder den Zweck der Forschung oder Planung auf andere Weise zu erreichen.

Die Übermittlung bedarf der vorherigen Genehmigung durch die oberste Bundes oder Landesbehörde (...)

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Datenzugang

- Datenübermittlung und Auswertung am Arbeitsplatz des Forschers
- Auswertung beim Dateneigner (Gastwissenschaftler)
- VPN-Zugang beim Dateneigner
- Zweizeitiges Vorgehen:
 - Übermittlung von Testdaten oder Stichprobe für Skriptgenerierung und Auswertung am Arbeitsplatz des Forschers
 - anschließend Skriptausführung am Volldatensatz bei Dateneigner durch Forscher oder Dateneigner
 - setzt identische Auswertungsumgebungen voraus!!!

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Welche Datengrundlage für welche Studie?

Leitlinie 3: Studienplan



Grundlage einer Sekundärdatenanalyse ist ein detaillierter und verbindlicher Studienplan, in dem die Studiencharakteristika schriftlich festgelegt werden.

Vor der Durchführung einer Sekundärdatenanalyse ist ein Studienplan zu erstellen. Er enthält alle Angaben, die für die Beantwortung der Fragestellung und für die Durchführung der Studie notwendig sind. Dazu gehören folgende Aspekte:

- ▶ Fragestellung und Arbeitshypothesen,
- ▶ Studiendesign,
- ▶ Auswahl der Datenbasis,
- ▶ Größe der angestrebten Studienpopulation mit Begründung,
- ▶ Ein- und Ausschlusskriterien der Beobachtungseinheiten,
- ▶ Auswahl der erforderlichen Merkmale,
- ▶ Auswertungsstrategie einschließlich der statistischen Methoden.

Die folgenden vier Aspekte sind häufig bereits in anderen separaten Dokumenten festgelegt, auf die im Studienplan verwiesen werden kann, sie können jedoch auch Bestandteil des Studienplans selbst sein:

- ▶ Konzept zur Datenbereitstellung und -übermittlung sowie zur Sicherung und Archivierung von Daten, Auswertungsdokumentationen und Ergebnissen (s. Leitlinie 6),
- ▶ Maßnahmen zur Qualitätssicherung (s. Leitlinie 5),
- ▶ Maßnahmen für die Gewährleistung ethischer Prinzipien (s. Leitlinie 1), und des Datenschutzes (s. Leitlinie 8) und
- ▶ Zeitplan mit Festlegung der Verantwortlichkeiten (s. Leitlinie 7).

Bei der Umsetzung dieser Anforderungen sind die Eigenschaften der spezifischen Daten zu beachten.

Empfehlung 3.1 – Studiendesign

Das Studiendesign wird beschrieben und seine Wahl angemessen begründet. Die Eignung der Daten als Basis für Auswertungen im Sinne der wissenschaftlichen Fragestellung ist darzulegen.

Welche Datengrundlage für welche Studie?

- Sind GKV-Routinedaten grundsätzlich zur Beantwortung der Forschungsfrage geeignet?
- Welche Daten/Sektoren werden benötigt?
- Für welchen Zeitraum? Historisch? Aktuell?
- Populationsbezug/Nenner?
- Regionaler Bezug/bundesweite Daten?
- ...

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Herausforderung «Datenzugang»

- Analysen auf der Basis von *versichertenbezogenen sektorübergreifenden* GKV-Routinedaten erfolgen bisher i.d.R. auf den Daten einer Krankenkasse(nart)
- Größter Kritikpunkt ist dabei immer der (vermutete) Krankenkassen(arten)- immanente Bias, der (für einzelne Indikatoren) auch bestehen bleibt, wenn auf wichtige (vorhandene) Einflussfaktoren kontrolliert wird
- Gemeinsame Auswertung mit Poolen der Versichertendaten scheitert (bisher/oft) an der Bereitschaft der Krankenkassen, an datenschutzrechtlichen Vorgaben, Aufwand und Kosten, an fehlenden «Kontakten»
- Daten des FDZ Sozialdaten können zukünftig eine Lösung sein. Aber: keine Pflegedaten, keine Aktualität, Zugang?

Prüfung / Plausibilisierung

Leitlinie 5: Qualitätssicherung

In Sekundärdatenanalysen ist eine begleitende Qualitätssicherung aller relevanten Instrumente und Verfahren sicherzustellen.

Eine interne begleitende Qualitätssicherung ist unabdingbarer Bestandteil jeder Sekundärdatenanalyse.

Sekundärdatenanalysen greifen auf Daten zurück, die primär zu einem anderen Zweck und von anderen Personen erhoben wurden. Auf die primäre Datenerhebung hat der Sekundärdatenutzer keinen Einfluss und kann somit auch nicht die Qualität der Primärdatenerhebung sichern. Aufgrund der Vielfalt potenziell verwendbarer Datenquellen, der unterschiedlichen Verwendungszwecke und der rechtlichen Grundlagen muss besonderer Wert auf die Transparenz der Datenentstehung und -übermittlung gelegt werden. Eine enge Kooperation zwischen Dateneignern und Sekundärdatennutzern ist anzustreben.

Die Qualitätssicherung setzt an verschiedenen Punkten der Datenentstehung und Datenübermittlung an, um valide Auswertungsgrundlagen zu schaffen. Im Einzelnen umfasst die Qualitätssicherung die Durchführung von Plausibilitätskontrollen, Prüfungen der Konsistenz der Daten sowie die Festlegung personeller Zuständigkeiten (s. Empfehlungen 6.4, 6.6, 7.2). Ebenso erstreckt sich die Qualitätssicherung auf die in Leitlinie 6 angesprochene Dokumentation der Datenaufbereitung.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Prüfung / Plausibilisierung

Leitlinie 6: Datenaufbereitung

Für die Erfassung und Haltung aller für eine Studie erforderlichen Daten sowie für die Aufbereitung, Plausibilitätsprüfung, Kodierung und Bereitstellung der Daten ist vorab ein detailliertes Konzept zu erstellen.

Sämtliche der Datenanalyse vorangehenden Prozesse der Datenaufbereitung und -prüfung müssen dokumentiert werden. Die Dokumentation enthält die vollständige und nachvollziehbare Beschreibung der übermittelten Daten und die nachgehenden Transformationen bis zur Generierung eines Auswertedatensatzes. Dies kann auch durch eine abgespeicherte Syntax oder ein Skript der jeweilig eingesetzten Software geschehen.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



- Datenbereitstellung erfolgt i.d.R. auf Basis eines im Vorfeld abgestimmten Datensatzbeschreibung inkl. der für die Selektion notwendigen Aufgreif- und Filterkriterien
- Ein Datenerhebungsprotokoll ist nicht die Regel
- Eine Dokumentation der gelieferten Daten fehlt i.d.R., also auch keine Angabe über die Anzahl der gelieferten Datenzeilen
- Die Lieferung erfolgt i.d.R. als reine Text-Datei, meist mit Trennzeichen (Semikolon/Komma) und damit ohne Formatbeschreibung/Formatdatei
- In der ersten Zeile meist die Spaltennamen, ebenfalls mit Trennzeichen getrennt, zuweilen fehlt auch diese Angabe.

- (GKV-)Routinedaten sind Abrechnungsdaten und dienen in erster Linie administrativen Zwecken
- Die Qualität (Vollständigkeit, Korrektheit, etc.) orientiert sich am Primärzweck
- Vor der Sekundärnutzung erfolgt i.d.R. eine Datenbereitstellung durch Dritte (keine Projektmitarbeiter)
- Bereitstellung erfolgt i.d.R. nicht in geprüften, validierten oder qualitätsgesicherten Arbeitsschritten, sondern individuell/spezifisch für jedes Projekt
- Daten liefernde Institutionen haben jeweils unterschiedliche Datenhaltungssysteme, kein Austausch/Nutzung von Selektionsprogrammen
- I.d.R. erhält der Projektverantwortliche/Projektmanager keinen Einblick in die Selektions- und Aufbereitungsprogramme
- Die Daten liefernden Institutionen plausibilisieren die ausgelesenen Daten i.d.R. nicht

Jahr	Quartal	Anzahl Datensätze				
		r1 Stamm	r2 Amb	r6 Rx	r8fall Krhs	r9sach SoLe
2013	1		299.042	786.494		1.997.627
	2		306.474	820.315		2.151.078
	3		311.852	847.895		2.321.992
	4		306.904	886.701		2.438.014
2014		40.028				
	1		311.429	930.628	39.589	2.473.435
	2		288.213	882.585	37.875	2.349.803
	3		277.083	838.472	36.510	2.262.968
	4		248.532	776.997	35.387	2.114.627
2015		40.789				
	1		230.619	716.343	35.259	1.953.360
	2		203.901	641.626	30.920	1.776.345
	3		186.171	586.966	28.961	1.663.111
	4		161.500	533.690	27.064	1.539.031
2016		39.361				
	1		139.855	457.013	24.508	1.362.421
	2		116.587	392.109	21.842	1.164.744
	3		93.709	311.975	18.299	948.886
	4		67.396	221.153	13.588	680.426

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Gesamt					Versicherte	
r1	r2	r6	r8fall	r9sach	Anzahl	Anteil [%]
X	X	X	X	X	84.209	79,9
X	X	X	X	-	9.041	8,6
X	X	X	-	X	6.557	6,2
X	X	X	-	-	2.786	2,6
X	X	-	X	X	225	0,2
X	X	-	X	-	483	0,5
X	X	-	-	X	96	0,1
X	X	-	-	-	363	0,3
X	-	X	X	X	7	0,0
X	-	X	X	-	15	0,0
X	-	X	-	-	17	0,0
X	-	-	X	X	68	0,1
X	-	-	X	-	613	0,6
X	-	-	-	X	46	0,0
X	-	-	-	-	927	0,9
Gesamt					105.453	100,0

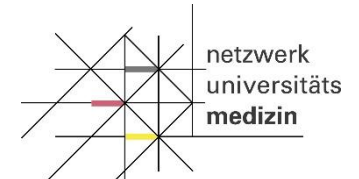
Anzahl Datensätze mit einer Pharmazentralnummer der Länge							
missing	4	5	6	7	8	9	10
			15	77	7		
		4	60	421	12		
		79	1.599	8.479	257		
							885
					2.936		
				55	126		7
					1.297		
14					748		
					4.290		2
					1.859		
	1	17	606	3.557	92	17	103
		6	21	165	17	1	
	1	6	289	1.607	107	22	14
		6	157	734	28	7	22
							2.397
					154		
	9	426	8.077	41.007	1.651	261	19.992
					2.882		145

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



ungültige Datumsangabe
2002-14-00
2002-16-00
2004-16-00
2005-14-00
2005-15-00
2008-14-00
2010-13-00
2010-15-00
2012-13-00
2012-16-00

ungültige Datumsangabe
20111000
20111100
20111200
20120100
20120200
20120300
20120400
20120500
20120600
20120700

ungültige Datumsangabe
2012.04.00
2012.05.00
2012.06.00
2012.07.00
2012.08.00
2012.09.00
2012.10.00
2012.11.00
2012.12.00
2013.01.00

ungültige Datumsangabe
2013-13-19
2013-13-21
2013-13-24
2013-13-26
2013-13-29
2014-14-19
2014-14-21
2015-15-20
2016-16-02
2016-16-25

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021



Fazit

- Prüfung der gelieferten Daten?
Bestandteil der Guten Praxis!
Ein Muss!
- Erhebungskontext beachten
- Basisprüfung weitgehend uniform/automatisiert durchführbar
mit Prüfung der Ergebnisse an bisherigen Erfahrungswerten
- Projektspezifische und Analysebezogene Prüfung
abhängig von Erhebungskontext und Fragestellung
- Ausreichend Zeit für Aufbereitung einplanen

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01KX2021

