

Die Gute Praxis Datenlinkage erweitert die Familie ‚Gute Praxis‘

Enno Swart¹, Stefanie March^{1,2}, Christoph Stallmann¹

¹ Institut für Sozialmedizin und Gesundheitssystemforschung (ISMG)
Medizinische Fakultät, Otto-von-Guericke-Universität Magdeburg

www.med.uni-magdeburg.de

² Hochschule Magdeburg-Stendal, Fachbereich Soziale Arbeit, Gesundheit und Medien



MEDIZINISCHE FAKULTÄT
UNIVERSITÄTSKLINIKUM MAGDEBURG A.Ö.R.

Gliederung

1. **Methodische Herausforderungen**
2. **GEP: Gute Epidemiologische Praxis (2019)**
3. **GPS: Gute Praxis Sekundärdatenanalyse (2015)**
4. **GPD: Gute Praxis Datenlinkage (2019)**
5. **Abrundung: STROSA (2016)**

(alle Dokumente verfügbar über <https://agens.group>)

Hintergrund (I)

- (späte) Etablierung der Epidemiologie in Deutschland seit Anfang der 1990er-Jahre (Gründung der dae 1992 als Vorläuferin der DGEpi)
 - Geringes epidemiologisches Basiswissen und viele ‚Quereinsteiger‘
 - Notwendigkeit der Etablierung epidemiologischer Standards und Handlungsanleitungen
- **Gute Epidemiologische Praxis (GEP)**

Hintergrund (II)


- Zunehmende wissenschaftliche Nutzung von GKV-Routinedaten seit den 1990er-Jahren
- AGAS-Memorandum 1997 zur Public Health Forschung mit Gesundheits- und Sozialdaten (von Ferber et al., Asgard-Verlag)
- Anfänglich Vorbehalte gegen ‚zweitrangige‘ (Versorgungs-) Forschung auf der Basis von Sekundärdaten
- Etablierung von Versorgungsforschung in Deutschland seit 2000
- **Gute Praxis Sekundärdatenanalyse (GPS)**
- In den letzten Jahren steigende Anzahl von Studien die eine Verlinkung von Primär- und Sekundärdaten vorsehen (IidA, NAKO Gesundheitsstudie, Innofonds-Projekte)
- **Gute Praxis Datenlinkage (GPD)**

Gute Epidemiologische Praxis

- **Publiziert 1999**
- **Erste Revision 2005 (Hoffmann et al.; GESU 2005)**
- **Zweite Revision 2017/2018 (publiziert 2019)**

GUIDELINES | [Open Access](#) | Published: 04 March 2019

Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology

[Wolfgang Hoffmann](#), [Ute Latza](#), [Sebastian E. Baumeister](#), [Martin Brünger](#), [Nina Buttmann-Schweiger](#), [Juliane Hardt](#), [Verena Hoffmann](#), [André Karch](#), [Adrian Richter](#), [Carsten Oliver Schmidt](#), [Irene Schmidtmann](#), [Enno Swart](#) & [Neeltje van den Berg](#) 

[European Journal of Epidemiology](#) **34**, 301–317(2019) | [Cite this article](#)

Struktur der GEP

Die GEP gliedert sich in elf Leitlinien (LL; neu sortiert):

- LL 1 – Ethik
- LL 2 – Forschungsfrage
- LL 3 – Studienplan/Operationshandbuch
- LL 4 – Datenschutz
- LL 5 – Probenbanken
- LL 6 – Qualitätssicherung
- LL 7 – Datenhaltung und Dokumentation
- LL 8 – Auswertung epidemiologischer Daten
- LL 9 – Vertraglicher Rahmen
- LL 10 – Interpretation und wissenschaftliche Publikation
- LL 11 – Kommunikation und Public Health

Jede Leitlinie untergliedert sich weiter in mehrere konkrete Empfehlungen zu deren Umsetzung. Erläuterungen zu Leitlinien und Empfehlungen fördern das bessere Verständnis.
Ein besonderer Bezug zu Sekundärdaten(analysen) besteht nicht.

Zielsetzung der Guten Praxis Sekundärdatenanalyse

- **Standard für die Durchführung von Sekundärdatenanalysen**
- **Grundlage für vertragliche Absprachen zwischen Dateneignern und Sekundärdatennutzern**
- **Wirkfeld über GKV-Routinedaten hinaus**
- **Eigenständig, dabei Orientierung an und Bezug zu anderen Leitlinien (insbesondere GEP)**

Entwicklung der Guten Praxis Sekundärdatenanalyse

- Erste Fassung 2004 durch die AGENS beschlossen (GESU 2005)
- 1. Revision durch AGENS und AG Epidemiologische Methoden der DGEpi, 2007 (GESU 2008)
- Akkreditierung als ‚Ausführungsbestimmung‘ der GEP (bei Bewahrung ihrer Eigenständigkeit)
- Abschluss der 2. Revision April 2012, Publikation 2015
- Nächste Revision: ??

Gesundheitswesen 2015; 77(02): 120-126
DOI: 10.1055/s-0034-1396815



Konsensstatement

© Georg Thieme Verlag KG Stuttgart · New York

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

E. Swart, H. Gothe, S. Geyer, J. Jaunzeme, B. Maier, T. G. Grobe, P. Ihle

Struktur der GPS (seit 1. Revision 2008)

orientiert an der früheren Gliederung der GEP

- LL 1 – Ethik (0)
- LL 2 – Forschungsfrage (0)
- LL 3 – Studienplan (6)
- LL 4 – Probenbanken (0)
- LL 5 – Qualitätssicherung (3)
- LL 6 – Datenhaltung und Dokumentation (7)
- LL 7 – Auswertung (3)
- LL 8 – Datenschutz (11)
- LL 9 – vertragliche Rahmenbedingungen (3)
- LL 10 – Interpretation (0)
- LL 11 – Kommunikation und Public Health (3)

ergänzt um konkrete Empfehlungen (Anz.in Klammern) + Glossar

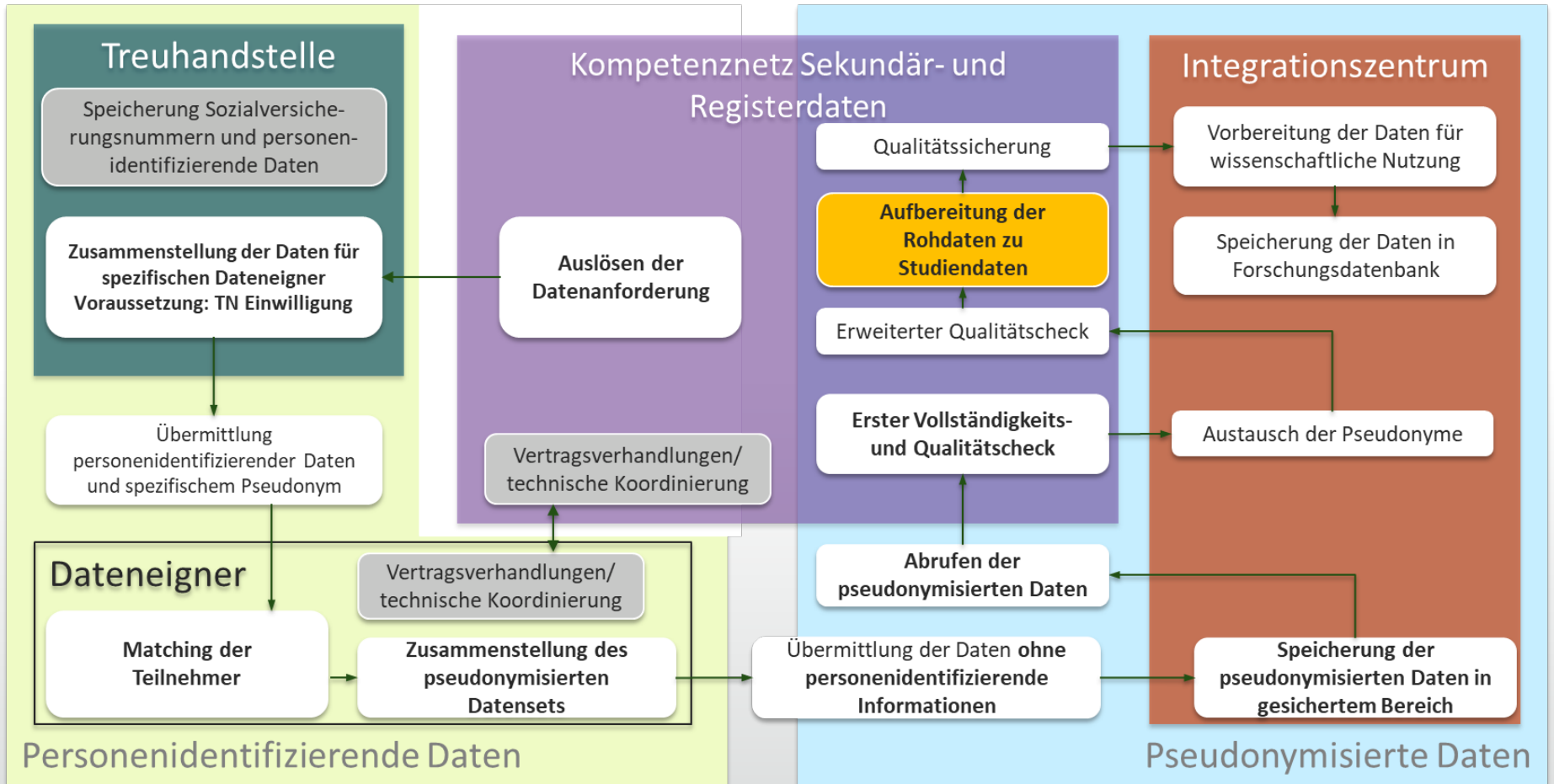
Zielgruppen der Guten Praxis Sekundärdatenanalyse

- **Versorgungsepidemiologen / Versorgungsforscher**
- **Nutzer von versorgungsrelevanten Studien auf der Basis von Sekundärdaten („critical appraisal“)**
- **Dateneigner (als Datennutzer und/oder Studienkonsumenten)**
- **Forschungsförderer (Auswahlkriterium bei Ausschreibungen)**

Gute Praxis Datenlinkage - Ausgangspunkt

- **Datenlinkage en vogue** (z.B. NAKO Gesundheitsstudie, Innofonds)
- **‚best of two worlds‘**
- **hohe (datenschutz-)rechtliche Hürden**
- **damit technisch und organisatorisch anspruchsvoll** (z.B. Hash-Verfahren, Vertrauensstellen, Datenfluss)
- **hoher Bedarf nach Handreichung bei gleichzeitig großer Unsicherheit**
- **Vorbereitung durch Positionspapier**

Zur Veranschaulichung: Datenfluss in der NAKO Gesundheitsstudie



BOX 1 CHECKLISTE DER WICHTIGSTEN FRAGEN BEIM DATENLINKAGE

Frage

- Welche Datenquellen sollen verknüpft werden?
- Kommen personenidentifizierende Merkmale vor?
- Brauche ich eine Einwilligungserklärung für das Datenlinkage?
- Welche Fehler können in Identifikatoren vorkommen und wie können diese minimiert werden?
- Gibt es Datenfelder, die nicht Bestandteil des Identifikators sind, aber auf eine korrekte/fehlerhafte Zuordnung hinweisen können?
- Wie hoch ist die Anzahl verlinkter Datensätze?
- Wie kann die Güte des Linkage-Verfahrens eingeschätzt werden?
- Gibt es strukturelle Unterschiede zwischen verlinkten und nicht verlinkten Datensätzen?
- Wie gut ist die Datenqualität der Identifikatoren?
- Wie groß sind falsch negative bzw. falsch positive Klassifikationen?
- Wie gehe ich mit Zeitaspekten um?

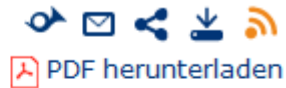
► Tab. 1 Ausgewählte Beispiele für verschiedene Arten des Datenlinkage.

Studientitel	Welche Daten wurden verknüpft?	Welche Arten des Datenlinkage wurden verwendet?	Referenzen
SHARE RV: Verknüpfung von Befragungsdaten des Survey of Health, Ageing and Retirement in Europe (SHARE) mit administrativen Daten der Rentenversicherung	Befragungsdaten mit Rentenversicherungsdaten	deterministisches Linkage, Linkage mit direkten Identifikatoren, Linkage mit Einwilligungserklärung	Czaplicki & Korbmacher (2010) [64] Korbmacher & Czaplicki (2013) [65] http://www.share-project.org/
lidA - Leben in der Arbeit - eine Kohortenstudie zu Gesundheit und Älterwerden in der Arbeit	Befragungsdaten mit Daten der Gesetzlichen Krankenversicherung (GKV) und der Bundesagentur für Arbeit (BA)	deterministisches Linkage, Linkage mit direkten Identifikatoren, Linkage mit Einwilligungserklärung	March et al. (2012) [55] March et al. (2015) [12] www.lidA-studie.de
	Befragungsdaten mit aggregierten Daten der GKV	Linkage mit indirekten Identifikatoren, Linkage ohne Einwilligungserklärung (Antrag nach § 75 SGB X)	
Vergleich unterschiedlicher Linkageverfahren sowie Vollständigkeit klinischer Angaben	Daten der GKV mit Daten eines Krankenhausinformationssystem	Linkage mit direkten und indirekten Identifikatoren	Ohlmeier et al. (2015) [42]
QS-AMI Studie	Daten der GKV mit Daten eines klinischen Registers	deterministisches Linkage, Linkage mit indirekten Identifikatoren	Maier et al. (2015) [11]
ALWA-ADIAB – ALWA-Befragungsdaten verknüpft mit administrativen Daten des IAB	Befragungsdaten mit Daten der BA	deterministisches und probabilistisches Linkage	Antoni et al. (2011) [66] Antoni & Seth (2012) [67]
Validierungsstudie	Daten der GKV mit Daten eines Sterberegisters	probabilistisches Linkage, deterministisches Linkage, exaktes Linkage, fehlertolerantes Linkage, Linkage mit direkten Identifikatoren	Ohlmeier et al. 2016 [58]
Sektorenübergreifende Datenzusammenführung und Evaluation am Beispiel der Schenkelhalsfrakturen	Daten der GKV mit Daten der Externen Qualitätssicherung und Pflegegutachten des Medizinischen Dienstes der Krankenkassen	exaktes Linkage, Linkage mit direkten Identifikatoren, fehlertolerantes Linkage	Ohmann et al. (2005) [68]
Kohortenstudie zur Krebsinzidenz bei Patienten mit Diabetes mellitus Typ 2	Daten aus Disease-Management (DMP)-Programmen (GKV) mit Daten eines Epidemiologischen Krebsregisters	probabilistisches Linkage	Kajüter et al. (2014) [39]

Quelle: March et al.: Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme. GESU 2018



CC BY-NC-ND 4.0 · Gesundheitswesen 2019; 81(08/09):
636-650
DOI: 10.1055/a-0962-9933



PDF herunterladen

Originalarbeit

Eigentümer und Copyright ©Georg Thieme Verlag KG 2018

Gute Praxis Datenlinkage (GPD)

Good Practice Data Linkage

Stefanie March, Silke Andrich, Johannes Drepper, Dirk Horenkamp-Sonntag, Andrea Icks, Peter Ihle, Joachim Kieschke, Bianca Kollhorst, Birga Maier, Ingo Meyer, Gabriele Müller, Christoph Ohlmeier, Dirk Peschke, Adrian Richter, Marie-Luise Rosenbusch, Nadine Scholten, Mandy Schulz, Christoph Stallmann, Enno Swart, Stefanie Wobbe-Ribinski, Antke Wolter, Jan Zeidler, Falk Hoffmann

Mitglieder aus AGENS, DGEpi, DGSMP, DNVF, TMF e.V.)

Gute Praxis Datenlinkage

(March et al.; Das Gesundheitswesen, 2019, open access)

Leitlinie 1: Forschungsziele, Fragestellung, Datenquellen und Ressourcen

Leitlinie 2: Dateninfrastruktur und Datenfluss

Leitlinie 3: Datenschutz

Leitlinie 4: Ethik

Leitlinie 5: Schlüsselvariablen und Linkageverfahren

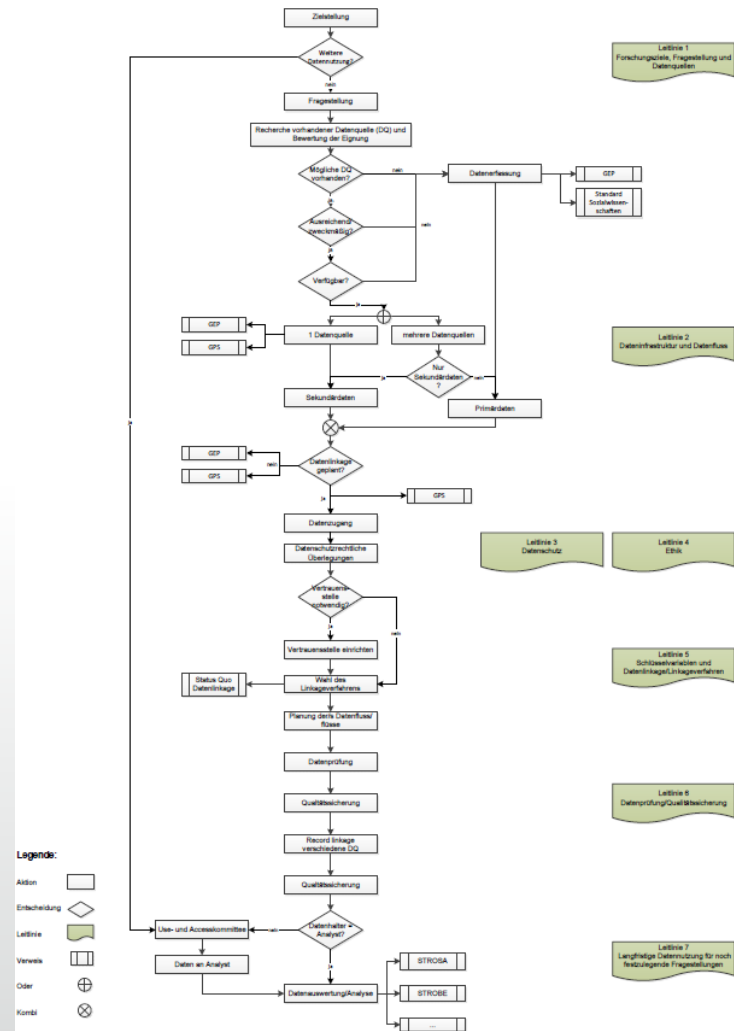
Leitlinie 6: Datenprüfung und Qualitätssicherung

Leitlinie 7: langfristige Datennutzung

Gute Praxis Datenlinkage

Leitlinien orientieren sich am Flussdiagramm eines typischen Verfahrensablaufs (March et al. 2019)

Gute Praxis Datenlinkage – Flussdiagramm (Stand: 01.11.2018)



Gute Praxis Datenlinkage enthält

- **Umfangreiches Glossar**
- **Empfehlungen als Konkretisierung der einzelnen Leitlinien**
- **Umfangreiches Literaturverzeichnis mit Beispielen konkreter Umsetzung**

Nebenzweige der Gute-Praxen-Familie

- **Gute Praxis Gesundheitsberichterstattung** – Leitlinien und Empfehlungen 2.0 - Journal of Health Monitoring S1/2019
- **Gute Kartographische Praxis im Gesundheitswesen** (2015; Überarbeitung in Vorbereitung, AG Health Geography)
- **Gute Praxis Gesundheitsinformation** (Deutsches Netzwerk Evidenzbasierte Medizin e.V.)
- **Gute Praxis Bewertungsportale** (KBV und BÄK)

Berichtsstandard für Sekundärdatenanalyse

- Für nahezu alle Studienarten existieren Berichtsstandards
- Steigende Zahl an wissenschaftliche Publikationen: Wo ist der Weizen hinter all der Spreu?
- Es fehlen konsentierete Kriterien zur systematischen Begutachtung von Studien auf der Basis von Sekundärdaten, die deren spezifischen Rahmenbedingungen berücksichtigen
- Startpunkt 1: Gute Epidemiologische Praxis seit 1999 (Rev. 2004)
- Startpunkt 2: STROBE: STrengthening the Reporting of OBservational Studies in Epidemiology (2007; dt. 2008)

Bestehende Berichtsformate: STROBE

- **ST**rengthening the **R**eporting of **OB**servational Studies in Epidemiology
- 22 Kriterien für Einleitung, Methoden, Ergebnisse und Diskussion, incl. einer unterstützende Checkliste
- STROBE gilt gleichermaßen für alle Typen epidemiologischer Beobachtungsstudien (Kohorten-, Fall-Kontroll- und Querschnittstudien)
- **Aber: STROBE thematisiert nicht die spezifischen Rahmenbedingungen von Sekundärdatenanalysen.**



Standardisierte BerichtsROutine für Sekundärdaten Analysen (STROSA) – ein konsentierter Berichtsstandard für Deutschland, Version 2

A Consensus German Reporting Standard for Secondary Data Analyses, Version 2 (STROSA-Standardisierte BerichtsROutine für SekundärdatenAnalysen)

Autoren

E. Swart¹, E. M. Bitzer², H. Gothe^{3,4}, M. Harling⁵, F. Hoffmann⁶, D. Horenkamp-Sonntag⁷, B. Maier⁸, S. March¹, T. Petzold^{9,10}, R. Röhrig⁶, A. Rommel¹¹, T. Schink¹², C. Wagner^{13,14}, S. Wobbe⁵, J. Schmitt⁹

Institute

Die Institutsangaben sind am Ende des Beitrags gelistet

Gesundheitswesen 2016; 78: e145-e160
DOI <http://dx.doi.org/10.1055/s-0042-108647>

Herzlichen Dank an meine Kolleginnen und Kollegen!!



Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

: enno.swart@med.ovgu.de