

GPS - Gute Praxis Sekundärdatenanalyse: Revision nach grundlegender Überarbeitung

Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS)¹ der
Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP)

und

Arbeitsgruppe Epidemiologische Methoden² der
Deutschen Gesellschaft für Epidemiologie (DGEpi), der
Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie
(GMDS) und der
Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP)

¹ Sprecher: Peter Ihle (Köln) und Dr. Enno Swart (Magdeburg)

² Sprecher: Dr. Thomas Lampert (Berlin) und Dr. Stefanie Klug (Mainz)

Zielsetzung und Verantwortlichkeit

Die Nutzung von Sekundärdaten in gesundheitswissenschaftlichem Kontext hat in den letzten Jahren kontinuierlich zugenommen. Als Sekundärdaten lassen sich in diesem Zusammenhang Routinedaten der gesetzlichen Kranken-, Renten- und Unfallversicherung (Sozialdaten) oder Daten von (bevölkerungsbezogenen) Krankheitsregistern bezeichnen. Möglichkeiten, aber auch Grenzen der Nutzung derartiger Daten, die primär zu anderen Zwecken erhoben wurden, in epidemiologischen Studien und in der (Versorgungs-)Forschung werden intensiv diskutiert. Die zunehmende Bedeutung von Sekundärdaten zeigt ebenfalls der in das Sozialgesetzbuch eingefügte § 303a-f (Datentransparenz), der einen neuen gesetzlichen Rahmen zur Nutzung derartiger Daten aufzeigt.

Die Arbeitsgruppe „Erhebung und Nutzung von Sekundärdaten (AGENS)“ der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) veröffentlichte 2005 in Anlehnung an die seit 2000 vorliegende und von der Arbeitsgruppe „Epidemiologische Methoden“ der Deutschen Gesellschaft für Epidemiologie (DGEpi; vormals DAE), der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) und der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) in Zusammenarbeit mit der Deutschen Region der Internationalen Biometrischen Gesellschaft (DR-IBS) erstellten „Guten Epidemiologischen Praxis“ (GEP) erstmals eine „Gute Praxis Sekundärdatenanalyse“ (GPS).

Ziele der GPS: Es wird damit ein Standard für die Durchführung von Sekundärdatenanalysen formuliert. Gleichzeitig dient die Grundlage für vertragliche Absprachen zwischen Dateneigner (Primärnutzer) und Sekundärnutzer (siehe Glossar).

Zielgruppe der GPS sind neben den Dateneignern sozialmedizinisch und versorgungsepidemiologisch tätige Sekundärnutzer und die Nutzer deren Forschungsergebnisse. Damit sind nicht allein Angehöriger von Hochschuleinrichtungen gemeint, sondern alle Personen, die sich unter wissenschaftlicher Perspektive und unter Anwendung wissenschaftlicher Methoden Sekundärdaten und deren Analyse zuwenden. Wie in den Gesundheitswissenschaften üblich, gehören dazu Angehöriger verschiedener Berufsgruppen, neben Epidemiologen gleichermaßen Mediziner, Statistiker, Sozialwissenschaftler, Gesundheitsökonomien u.v.m.

Methodik / Verfahren der Aktualisierung: Die Resonanz auf die Veröffentlichung der GPS war durchweg positiv. Gleichzeitig wurden in der Zwischenzeit mehrere Verbesserungsvorschläge aus dem Kreis der Dateneigner und Epidemiologen unterbreitet. Diese bezogen sich einerseits auf einzelne Aussagen der GPS, die sich nach Auffassung der Kritiker zu stark an den Sozialdaten der GKV orientierten und daher allgemeiner formuliert werden sollten. Andererseits wurde der Wunsch nach einer stärkeren und formaleren Anbindung an die GEP formuliert, um zu verdeutlichen, dass Planung und Durchführung einer Sekundärdatenanalyse als besondere Studienform der Epidemiologie sich immer auch an der GEP orientieren sollte.

In Reaktion auf diese Rückmeldungen erfolgte die erste Revision der GPS durch eine Querschnittsarbeitsgruppe aus Mitgliedern der AGENS und der AG Epidemiologische Methoden unter Einbeziehung weiterer epidemiologischer Experten. Auf mehreren Arbeitstreffen wurde ein Revisionsentwurf erarbeitet, der anschließend den Mitgliedern der beteiligten Arbeitsgruppen zur Diskussion gestellt wurde. Aus den eingegangenen Über-

arbeitungsvorschlägen wurde durch diese Querschnittsarbeitsgruppe die vorliegende Revision der GPS erstellt.

Die zweite Fassung der Guten Praxis Sekundärdatenanalyse wurde gegenüber der ersten Version umstrukturiert. Die Nummerierung der Leitlinien erfolgt nunmehr in der Reihenfolge der Leitlinien der GEP. Ebenso wurde deren Formulierung wörtlich aus der GEP übernommen. Es wurde mit Ausnahme der Leitlinie 4 (Probendatenbanken) lediglich der allgemeine Ausdruck ‚epidemiologische Studie‘ durch den spezifischen Begriff ‚Sekundärdatenanalyse‘ ersetzt. In den Erläuterungen der Leitlinien und den daraus abgeleiteten Empfehlungen werden in der GPS dann die spezifischen Rahmenbedingungen und Anforderungen einer Sekundärdatenanalyse angesprochen. In einem gegenüber der ersten Fassung erweiterten Glossar werden am Ende der GPS zentrale Begriffe definiert.

Die GPS ist in der vorliegenden Form allgemein gehalten, um Nutzer unterschiedlicher Arten von Sekundärdaten anzusprechen. Konkrete Studienbedingungen und Spezifika bestimmter Daten können es erforderlich machen, von den Empfehlungen der GPS abzuweichen. Ein primäres Ziel der Mitwirkenden an der Formulierung der GPS war es, eine allgemein anwendbare Gute Praxis für Sekundärdatenanalysen zu formulieren.

Die GPS ist daher nicht als verbindlicher Standard zu verstehen, der keinerlei Abweichungen erlaubt. Vielmehr soll sie als Richtschnur bei der Planung, Durchführung und Analyse von Studien auf der Basis von Sekundärdaten dienen. Vor dem Hintergrund der Zielsetzung einer Studie und der verfügbaren Daten soll die Sinnhaftigkeit der Einhaltung der Leitlinien und Empfehlungen geprüft werden. Abweichungen sind dabei jederzeit möglich, bedürfen aber einer nachvollziehbaren Begründung. Es sei ausdrücklich erwähnt, dass die vorliegende Gute Praxis Sekundärdatenanalyse bei aller Orientierung an die GEP aus sich heraus verständlich ist und daher eine eigenständige Leitlinie darstellt. Der Nutzer eines bestehenden Datenkörpers sollte für Dritte nachvollziehbar begründen, ob die geplante Analyse eine Sekundärdatenanalyse im Sinne dieser GPS darstellt oder primär die Leitlinien und Empfehlungen der GEP gelten sollen.

Die GPS wurde in der vorliegenden Fassung nach Abschluss des beschriebenen Revisionsprozesses den Vorständen der vier epidemiologischen Fachgesellschaften DGEpi, DGSMP, GMDS (Gesellschaft für medizinische Dokumentation und Statistik) und IBS-DR (Deutsche Region der Internationalen Biometrischen Gesellschaft) vorgelegt und auf der gemeinsamen Jahrestagung der vier Gesellschaften im September 2007 in Augsburg formell beschlossen.

Die GPS ist gültig bis Ende 2010. Eine Überarbeitung nach diesem Termin kann ohne eine gleichzeitige Revision der GEP unter Federführung der AGENS erfolgen.

Gute Praxis Sekundärdatenanalyse

Version 2 gültig bis 12/2010

Leitlinie 1: Ethik

Sekundärdatenanalysen müssen im Einklang mit ethischen Prinzipien durchgeführt werden und Menschenwürde sowie Menschenrechte respektieren.

Diese Forderung gilt für Sekundärdatenanalysen ebenso wie für Primärdatenerhebungen. Ethische Prinzipien werden in allgemeinen Menschen- und Bürgerrechten formuliert. Ethische Prinzipien sind auch dann einzuhalten, wenn eine rechtliche Verpflichtung hierzu nicht besteht.

Die Empfehlung zur Konsultation einer Ethikkommission kann bei Sekundärdatenanalysen entfallen, wenn alle datenschutzrechtlichen Vorschriften zur Pseudonymisierung personenbezogener Daten erfüllt sind (vgl. auch Leitlinie 8) und keine Verknüpfung mit Primärdaten geplant ist.

Leitlinie 2: Forschungsfrage

Die Planung jeder Sekundärdatenanalyse erfordert explizite und operationalisierbare Fragestellungen, die spezifisch und so präzise wie möglich formuliert sein müssen. Die Auswahl der zu untersuchenden Bevölkerungsgruppen muss im Hinblick auf die Forschungsfrage begründet werden.

Auch in einer Sekundärdatenanalyse muss die Auswahl der zu untersuchenden Populationen im Hinblick auf die Forschungsfrage begründet werden.

Die Forschungsfrage ist unverzichtbarer Ausgangspunkt einer Beurteilung des potentiellen Nutzens einer Sekundärdatenanalyse, und zwar unabhängig von der Art des Datenkörpers. Die explizite Formulierung der Forschungsfrage ist wesentliche Voraussetzung für Planung und Bewertung des Studiendesigns und der Datenextraktion, aber auch des Zeit- und Kostenrahmens der geplanten Untersuchung. Die Forschungsfrage ermöglicht erst die Festlegung der Einzelheiten einer Sekundärdatenanalyse (Auswahl der Untersuchungsgruppe, Wahl des geeigneten Datenkörpers, Bestimmung relevanter Variablen etc.).

In Sekundärdatenanalysen sind konfirmatorische und explorative Analysen voneinander abzugrenzen. Konfirmatorisch zu überprüfende Hypothesen müssen vor Beginn der Analyse formuliert werden.

Leitlinie 3: Studienplan

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

Die Erstellung eines Studienplans vor Beginn einer Sekundärdatenanalyse ist eine wesentliche methodische Voraussetzung für deren Qualität.

Der Studienplan ist eine Zusammenstellung der wichtigsten Angaben, die für die Beantragung und Beurteilung der Studie als Forschungsvorhaben und für ihre Durchführung notwendig sind. Bestandteile des Studienplans im Rahmen einer Sekundärdatenanalyse sollten sein:

- Fragestellung und Arbeitshypothesen,
- Studientyp,
- Datenbasis,

- Studienumfang und dessen Begründung,
- Ein- und Ausschlusskriterien bei der Bestimmung des Datenkörpers,
- Festlegung der geeigneten Datenfelder innerhalb des in Frage kommenden Datenkörpers,
- Konzept zur Datenbereitstellung und –übermittlung sowie zur Archivierung von Roh- und Auswertedatensätzen,
- Auswertungsstrategie einschließlich der statistischen Methoden,
- Maßnahmen zur Qualitätssicherung,
- Maßnahmen für die Gewährleistung des Datenschutzes und ethischer Prinzipien,
- Zeitplan mit Festlegung der Verantwortlichkeiten.

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

Empfehlung 3.1 - Studiendesign

Der Studientyp soll beschrieben und seine Wahl angemessen begründet werden. Die Eignung des Datenkörpers als Basis für Auswertungen im Sinne des Studiendesigns ist zu begründen.

Empfehlung 3.2 – Studienteilnehmer / Datenbasis

Die Sekundärdatenanalyse sollte sich auf eine Studienpopulation beziehen, deren Auswahl auf einer kritischen Beurteilung von Erhebungszweck, Qualität, Reliabilität, Validität der verwendeten Daten und der Verallgemeinerbarkeit der Ergebnisse beruht.

Empfehlung 3.3 – Vermeidung von Verzerrungen, interne Validität

Bereits bei der Planung von Studien auf der Basis von Sekundärdaten soll möglichen Verzerrungen (Bias) der Ergebnisse entgegengewirkt werden, die durch Selektion, Confounding etc. entstehen können. Bei Sekundärdatenanalysen kann das z.B. durch individuelles oder Gruppen-Matchen erfolgen oder aber durch die Berücksichtigung von Informationen, die zur Kontrolle von Störgrößen erforderlich sind.

Empfehlung 3.4 – Repräsentativität, Generalisierbarkeit, externe Validität

Analog zur Minimierung der Nichtteilnahmerate bei Primärdatenanalysen soll bei Sekundärdatenanalysen eine möglichst große Verallgemeinerungsfähigkeit für die untersuchte Grundgesamtheit angestrebt werden.

Empfehlung 3.5 - Variablen

Eine Sekundärdatenanalyse muss die Messgenauigkeit und Vollständigkeit der zu untersuchenden Merkmale und möglicher Störfaktoren in den Primärdaten berücksichtigen. Hierzu gehört eine Beschreibung und Bewertung aller verwendeten Daten(felder) und des Erhebungskontextes.

Empfehlung 3.6 - Studienumfang

Im Studienplan ist eine Begründung des Studienumfangs anzugeben. Insbesondere bei Auswertungen zu seltenen Ereignissen oder bei kleineren Zielpopulationen sollten quantitative Abschätzungen der statistischen Aussagekraft zur Bestimmung der erforderlichen Populationsgrößen durchgeführt werden (Machbarkeitsanalyse).

Empfehlung 3.7 – Operationshandbuch

Ergänzend zum Studienplan sollten in einem Operationshandbuch sämtliche organisatorischen Festlegungen zur Vorbereitung und Durchführung der Sekundärdatenanalyse

sowie deren schrittweise Abarbeitung dokumentiert werden. Dies schließt die Datenbereitstellung beim Dateneigner, die Datenübermittlung an den Sekundärnutzer und die Aufbereitung durch ihn ein.

Empfehlung 3.8 – Ressourcen

Für die Studie sind ausreichende zeitliche und personelle Ressourcen beim Dateneigner und Sekundärnutzer vorzusehen. Dies gilt gleichermaßen für die Datenbereitstellung, Aufarbeitung, Auswertung und Darstellung der Ergebnisse sowie für die erforderliche Kommunikation und Absprachen innerhalb und zwischen den beteiligten Stellen.

Leitlinie 4: Probandenbanken

In vielen epidemiologischen Studien ist die Anlage einer biologischen Probenbank notwendig bzw. sinnvoll. Hierfür und für die aktuelle und vorgesehene zukünftige Nutzung der Proben ist die dokumentierte Einwilligung aller Probanden erforderlich.

Biologische Proben sind gegenwärtig nicht Gegenstand von Sekundärdatenanalysen. Im übrigen sei auf die Leitlinie 4 der GEP mit ihren Erläuterungen und Kommentaren verwiesen.

Leitlinie 5: Qualitätssicherung

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

Eine interne begleitende Qualitätssicherung ist unabdingbarer Bestandteil jeder Sekundärdatenanalyse. Aufgrund der Vielfalt der potenziell einer Sekundärdatenanalyse zugrunde liegenden Datenquellen, ihrer unterschiedlichen ursprünglichen Verwendungszwecke, rechtlichen Grundlagen, Dateneignern sowie Übertragungswege muss besonderer Wert auf die Transparenz bzgl. der Datenentstehung und -übermittlung gelegt werden, um Auswertungsprozesse nachvollziehbar zu gestalten. Die Qualitätssicherung setzt an verschiedenen Punkten der Datenentstehung und -übermittlung an, um valide Auswertungsgrundlagen zu schaffen. Im einzelnen umfasst die Qualitätssicherung die Prüfung der Datenintegrität, Plausibilitätskontrollen und die Festlegung personeller Zuständigkeiten (vgl. Empfehlungen 6.4, 6.6, 7.2). Ebenso erstreckt sich die Qualitätssicherung auf die in Leitlinie 6 angesprochene Dokumentation der Datenaufbereitung.

Empfehlung 5.1 – Prätest

Im Vorfeld der Datenbereitstellung sollte die Möglichkeit einer Bereitstellung von Probedaten mit einer reduzierten Zahl von Beobachtungen erwogen werden.

Empfehlung 5.2 – Anpassung des Studienplans

Ergibt sich während der Durchführung einer Sekundärdatenanalyse die Notwendigkeit, die dort festgelegten Verfahrensweisen zu verändern, so sind diese Änderungen zu begründen und als Ergänzung zum Studienplan zu dokumentieren.

Leitlinie 6: Datenaufbereitung

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

Die Notwendigkeit der Dokumentation erstreckt sich auf sämtliche der Datenanalyse vorgeschalteten Prozesse der Datenaufbereitung und –Prüfung im Zuge der Datenbereitstellung durch den Dateneigner und der Datenübermittlung an den Sekundärnutzer. Die Dokumentation in Form eines Datenaufbereitungsprotokolls sollte die vollständige und nachvollziehbare Beschreibung des Datenhaltungssystems, der übermittelten Daten (Datum der Erhebung, Datum der Lieferung, Anzahl der Datensätze, verwendetes Übertragungsformat, verwendete Kodier- und Referenzlisten, Pseudonymisierungs- und Anonymisierungsschritte etc.) und die nachgehenden Transformationen bis zur Generierung eines Auswertedatensatzes umfassen.

Empfehlung 6.1 - Datenerhebung und -übermittlung

Die Grundlagen und Wege der Daten von ihrer primären Erfassung bis zur Bereitstellung für den Sekundärnutzer sollen dargestellt sein. Dies beinhaltet eine Darstellung des Verwendungszwecks der Datensammlung, der Erfassungsregeln und die Überprüfung der Konstanz von Erfassungsregeln (wie z.B. Diagnose- oder Prozedurenkodierung) und der gesetzlichen Rahmenbedingungen über die berücksichtigte Erhebungsperiode.

Empfehlung 6.2 - Ausgangsdatensatz

Der vom Dateneigner übermittelte Ausgangsdatensatz soll in unveränderter Form über den gesamten Zeitraum der Sekundärdatenanalyse verfügbar sein. Für die Reproduzierbarkeit der Analysen gilt zusätzlich die in Leitlinie 7 genannte Aufbewahrungsfrist.

Empfehlung 6.3 - Datenbeschreibung

Der Umfang und die Struktur der bereitgestellten und verwendeten Daten sollen dokumentiert werden. Hinweise auf fehlende Datensätze und deren Häufigkeit sowie die Feststellung redundanter Datensätze sind gleichfalls zu dokumentieren.

Empfehlung 6.4 - Datenqualität

Die Reliabilität und Validität der verwendeten Daten sollten auf der Basis verfügbarer Informationen überprüft werden. Eine externe Validierung kritischer Merkmale im Rahmen von Primärerhebungen, z. B. für Subpopulationen, ist anzustreben.

Empfehlung 6.5 - Plausibilitätskontrollen

Vor und während der Sekundärdatenanalyse sind Plausibilitätskontrollen durchzuführen. Sie erfolgen prinzipiell auf der Grundlage des Ausgangsdatensatzes. Vor eventuell erforderlichen Korrekturen sollte in der Regel eine Rücksprache mit dem Dateneigner erfolgen. Alle Ergänzungen und Änderungen von Variablenwerten sind vollständig und schriftlich zu dokumentieren.

Empfehlung 6.6 – Operationalisierung - abgeleitete Merkmale

Die Sekundärdatenanalyse umfasst in der Regel die Bildung von Merkmalen und Merkmalkategorien auf der Basis mehrerer Variablen des Ausgangsdatensatzes und deren Ausprägungen. Die Bildung abgeleiteter Merkmale ist nachvollziehbar zu dokumentieren und sollte sich ggf. an vorhandenen Standards orientieren, um eine Vergleichbarkeit der

Ergebnisse zu erleichtern. Ebenso sollte die Bildung neuer Variablen vollständig dokumentiert werden.

Empfehlung 6.7 – Auswertungsdatensatz

Der nach Plausibilitätsprüfung und Datentransformation überarbeitete Datensatz ist als Auswertungsdatensatz zu kennzeichnen und unabhängig vom Ausgangsdatensatz zu speichern und zu sichern. Für die Reproduzierbarkeit der Analysen gilt zusätzlich die in Leitlinie 7 genannte Aufbewahrungsfrist.

Leitlinie 7: Datenanalyse

Die Auswertung von Sekundärdaten soll unter Verwendung adäquater Methoden und ohne unangemessene Verzögerung erfolgen. Die den Ergebnissen zugrunde liegenden Daten sind in vollständig reproduzierbarer Form für mindestens 10 Jahre aufzubewahren.

Um die Überprüfbarkeit der Studienergebnisse zu ermöglichen ist eine konkrete Aufbewahrungsfrist von Ausgangsdaten- und Auswertungsdatensatz sowie der Datenaufbereitungsprotokolle beim Sekundärnutzer vertraglich zu regeln. Maßgeblich für diese Forderung an Sekundärdatenanalysen ist dabei nicht der Entstehungszeitpunkt der Sekundärdaten, sondern der Zeitpunkt der Übermittlung an den Sekundärnutzer.

Die einzelnen Auswertungsschritte sollten nachvollziehbar sein und hinsichtlich möglicher Rückwirkungen auf die Selektion der Studienpopulation, die Operationalisierung der Forschungsfrage, der Reichweite der Ergebnisse, der Entscheidungskriterien bei der Hypothesenprüfung usw. kritisch analysiert werden.

Die Sekundärdatenanalyse bedarf einer an den verfügbaren Daten orientierten Planung der Auswertungsstrategie. Diese muss die Messgenauigkeit und Vollständigkeit (hinsichtlich vorhandener Merkmale sowie hinsichtlich möglicher Störfaktoren und Wechselwirkungen) der Daten berücksichtigen. Die im Rahmen einer Sekundärdatenanalyse zu überprüfenden Hypothesen müssen vor Beginn der Untersuchung formuliert werden, ebenso die bei diesen Prüfungen anzulegenden Entscheidungskriterien.

Empfehlung 7.1 – Analyseplan

Die Auswertung zu den einzelnen Fragestellungen soll nach einem vorab erstellten Analyseplan und auf dem aktuellen Stand des epidemiologischen, statistischen oder methodischen Wissens erfolgen und hinsichtlich der allgemeinen Auswertungsstrategie sowie der einzelnen Auswertungsschritte nachvollziehbar begründet werden. Notwendige Änderungen des ursprünglichen Auswertungsplanes sind zu begründen und zu dokumentieren.

Empfehlung 7.2 – Personelle Verantwortung

Jede für die Datenanalysen verantwortliche Person muss vor Beginn der Arbeit benannt und in die rechtlichen und organisatorischen Voraussetzungen der Datensammlung eingewiesen werden. Dies gilt auch für weitere an den Datenanalysen beteiligte Personen, wenn das Verständnis der Datenerhebung für die sachgerechte Durchführung der Analyse notwendig ist.

Empfehlung 7.3 – Zwischenauswertungen

Wie bei Primärdatenanalysen soll auch bei Sekundärdatenanalysen auf Zwischenauswertungen verzichtet werden. Hiervon ausgenommen sind Zwischenauswertungen im

Studienverlauf, die dem Studienmonitoring dienen und somit Teil der Qualitätssicherung sind.

Empfehlung 7.4 – Prüfung der Ergebnisse

Die Auswertungen der Ergebnisse von Sekundärdatenanalysen sollten vor der Publikation der Gegenprüfung unterzogen werden. Auswertestrategie, die Auswertungen und ihre Resultate sollten durch Dritte reproduzierbar sein. Alle Auswertungen sollten derart dokumentiert werden, dass außenstehende Personen oder Institutionen die Auswertungsstrategie, die eigentlichen Auswertungen und ihre Resultate verstehen und nachvollziehen können. Die den Analysen zugrunde liegenden Daten und Programme sollen anschließend in vollständig reproduzierbarer Form archiviert werden.

Leitlinie 8: Datenschutz

Bei der Planung und Durchführung von Sekundärdatenanalysen ist auf die Einhaltung der geltenden Datenschutzvorschriften zum Schutz der informationellen Selbstbestimmung zu achten.

Die geltenden Datenschutzvorschriften, darunter das Prinzip der Datenvermeidung und der Datensparsamkeit (§ 3a BDSG), sowie ggf. weitere für den verwendeten Datenkörper einschlägige Regelungen müssen berücksichtigt werden. Alle Personen, die im Rahmen eines Forschungsprojektes Umgang mit personenbezogenen Daten haben, müssen über Inhalte, Reichweite und Möglichkeiten der einschlägigen gesetzlichen Bestimmungen informiert sein. Bei der Forschung mit personenbezogenen Daten müssen das Recht des Einzelnen auf informationelle Selbstbestimmung und zugleich das Recht auf Freiheit von Wissenschaft und Forschung beachtet werden.

Empfehlung 8.1 – Zweck der Datenbereitstellung

Der Zweck der Datenbereitstellung (im datenschutzrechtlichen Sinne) ist die Beantwortung der Forschungsfrage (siehe Leitlinie 2) und muss schriftlich fixiert werden.

Empfehlung 8.2 – Pseudonymisieren und Anonymisieren

Von den im Bundesdatenschutzgesetz aufgeführten Möglichkeiten der Pseudonymisierung und Anonymisierung ist Gebrauch zu machen (§ 3a BDSG Datenvermeidung und Datensparsamkeit). Hierbei ist auch die Einbindung eines Datentreuhänders zu prüfen.

Empfehlung 8.3 – De-Pseudonymisierung und Re-Identifikation

In den vertraglichen Rahmenbedingungen ist schriftlich festzulegen, ob und wenn ja, in welchen Fällen eine De-Pseudonymisierung möglich sein soll. Bei der Analyse ist eine unzulässige Re-Identifikation durch geeignete Mittel (technisch und vertraglich) zu verhindern.

Empfehlung 8.4 – Weitergabe personenbezogener Daten an Dritte

Die Weitergabe personenbezogener Daten erfolgt in der Regel nur durch den Dateneigner.

Empfehlung 8.5 – Personenbezogenes Datenlinkage mit externen Datenquellen

Jedes nicht explizit vorgesehene *personenbezogene* Datenlinkage mit externen Datenquellen bedarf der Einhaltung datenschutzrechtlicher Vorschriften.

Empfehlung 8.6 – Verantwortlicher für den Datenschutz

Bei jeder Sekundärdatenanalyse sind die nationalen und internationalen Standards der Datensicherheit und des Datenschutzes einzuhalten. Innerhalb einer Forschungseinheit ist eine für die Datenverarbeitung verantwortliche Person zu benennen, die die Einhaltung dieser Standards zu überwachen hat. Diese Person ist entsprechend ihrer Aufgaben zu qualifizieren.

Empfehlung 8.7 - Löschfristen

Wenn die für die Sekundärdatenanalyse bereitgestellten Daten nach Erreichen des Studienzieles aus datenschutzrechtlichen Gründen zu löschen oder zu anonymisieren sind, muss dies in Einklang mit der nach Empfehlungen 6.2 und 6.7 geforderten Aufbewahrung von Ausgangs- und Auswertungsdatensatz erfolgen. Ebenso muss bei der Festlegung von Löschfristen die nach Leitlinie 7 geforderte Möglichkeit zur Überprüfung der Ergebnisse aus einer Sekundärnutzung mitbedacht werden.

Empfehlung 8.8 – Zusammenarbeit mit Datenschutzbeauftragten

Bereits bei der Planung einer Sekundärdatenanalyse ist von den Beteiligten eine Kontaktaufnahme mit den zuständigen Datenschutzbeauftragten zu prüfen.

Leitlinie 9: Vertragliche Rahmenbedingungen

Die Durchführung einer Sekundärdatenanalyse setzt definierte rechtliche und finanzielle Rahmenbedingungen voraus. Hierzu sind rechtswirksame Vereinbarungen zwischen Auftraggeber und Auftragnehmer sowie zwischen Partnern von Forschungsk Kooperationen anzustreben.

Sekundärdatenanalysen beruhen in der Regel auf einer vertraglichen Übereinkunft zwischen Dateneigner und Sekundärnutzer. Aufgrund der gesetzlichen Auflagen zum Umgang mit Sozialdaten und anderen personenbezogenen Daten sind rechtswirksame Vereinbarungen der Kooperationspartner anzustreben, in denen die rechtlichen und finanziellen Rahmenbedingungen der intendierten Sekundärdatennutzung festgelegt werden. Im übrigen sind bei der Vielfältigkeit der speziellen Konstellationen und möglicher Datenquellen unterschiedliche Vertragsformen möglich.

Empfehlung 9.1 – Vertragsinhalte

Mit dem Dateneigner sollten transparente und realistische Vereinbarungen getroffen werden. Der Vertrag muss die Unabhängigkeit der Forschung, Rechte und Pflichten der Forschung und langfristigen Zugang zu den Daten thematisieren.

Empfehlung 9.2 – Verwendung der Ergebnisse

Die Verwendung der Ergebnisse einer Auftragsforschung für Forschung und Lehre darf nicht verhindert, behindert oder unzumutbar verzögert werden. Die vertragliche Festsetzung einer angemessenen Sperrfrist ist zulässig.

Leitlinie 10: Interpretation

Die Interpretation der Forschungsergebnisse einer Sekundärdatenanalyse ist Aufgabe des Autors/der Autoren einer Publikation. Grundlage jeder Interpretation ist eine kritische Diskussion der Methoden, Daten und Ergebnisse der eigenen Unter-

suchung im Kontext der vorhandenen Evidenz. Alle Publikationen sollten einem externen Review unterworfen werden.

Die Beurteilung der Ergebnisse gehört zu den originären Aufgaben der Sekundärnutzer. Der argumentative Prozess, der einer Interpretation zugrunde liegt, muss in einer schriftlichen Diskussion transparent und nachvollziehbar dargestellt werden. Einschränkungen bzgl. der Übertragbarkeit der Studienergebnisse auf andere Populationen oder andere Zeiträume als die betrachteten sind zu diskutieren.

Leitlinie 11: Kommunikation und Public Health

Sekundärdatenanalysen, deren Anliegen die Umsetzung von Ergebnissen in gesundheitswirksame Maßnahmen ist, sollten die betroffenen Bevölkerungsgruppen angemessen einbeziehen und eine qualifizierte Risikokommunikation mit der interessierten Öffentlichkeit anstreben.

Sekundärdatenanalysen können die Bewertung von Strukturen und Leistungen des Gesundheitssystems oder die Umsetzung und Evaluation gesundheitsrelevanter Maßnahmen zum Gegenstand haben. Dies um so mehr, als sie häufig auf Daten der Versorgungspraxis basieren. Ergibt sich nach dem fachlichen Urteil der Sekundärnutzer aus den Ergebnissen einer Sekundärdatenanalyse die Notwendigkeit von Konsequenzen, sollen diese, beispielsweise in Form einer Empfehlung, explizit formuliert werden können. Dabei müssen sich die Sekundärnutzer auch für eine verständliche Kommunikation mit Nicht-Fachleuten verantwortlich fühlen. Die Sekundärnutzer können darüber hinaus fachlich begründete Empfehlungen an den Dateneigner zur öffentlichen Bereitstellung von Informationen formulieren und an der technischen Umsetzung mitwirken.

Empfehlung 11.1 – Unabhängigkeit des Datennutzers

In Unabhängigkeit ihrer Forschungstätigkeit sollten Sekundärnutzer die Möglichkeit besitzen, sich bzgl. der praktischen Konsequenzen ihrer Analysen für die betroffenen Bevölkerungsgruppen direkt und ohne Abstimmung mit dem Dateneigner zu äußern. Formale Aspekte der direkten Kommunikation der Sekundärnutzer mit der Öffentlichkeit sollten vertraglich geregelt werden.

Empfehlung 11.2 – Transparenz der Methoden

Die in einer Sekundärdatenanalyse verwandten Methoden sollen an geeigneter Stelle veröffentlicht und interessierten Personen und Institutionen im Sinne der Methodentransparenz zugänglich gemacht werden. Die Weitergabe oder Publikation von Auswertungsroutinen, analytischen Verfahren und Berichtsformaten sollte unabhängig vom Dateneigner erfolgen können.

Empfehlung 11.3 – externe Datennutzung

Bei jeder Sekundärdatenanalyse sollte von Dateneigner und Sekundärnutzer geprüft werden, ob und inwieweit der Datensatz der wissenschaftlichen Öffentlichkeit für Forschungsk Kooperationen angeboten wird. Die abschließende Entscheidung verbleibt beim Dateneigner.

Begriffsdefinitionen

Primärdaten sind Daten, die im Rahmen ihres originär vorgesehenen Verwendungszwecks aufbereitet und analysiert werden.

Sekundärdaten sind Daten, die einer Auswertung *über ihren originären, vorrangigen Verwendungszweck hinaus* zugeführt werden. Maßgeblich für die Einstufung als Sekundärdaten sind Unterschiede zwischen dem primären Erhebungsanlass und der nachfolgenden Nutzung. Für die Einstufung ist es unerheblich, ob die weitergehende Nutzung durch den Dateneigner selbst oder durch Dritte erfolgt. Demnach sind beispielsweise Routinedaten einer Krankenkasse nicht nur Sekundärdaten, wenn sie für wissenschaftliche Fragestellungen genutzt werden, sondern z. B. auch dann, wenn sie durch die Krankenkasse für Zwecke der Versorgungsplanung herangezogen werden.

Sekundärdatenanalyse bezeichnet die Nutzung von Sekundärdaten. Die Sekundärdatenanalyse schließt die für die Analyse notwendigen Erhebungs- und Aufbereitungsschritte des Sekundärdatenkörpers ein. Erst durch diese Aufbereitungsschritte sind die Daten für wissenschaftliche Fragestellungen zugänglich. Die begriffliche Abgrenzung zwischen Primär- und Sekundärdaten und damit zwischen Sekundäranalyse und Sekundärdatenanalyse ist a priori und theoretisch nicht immer zweifelsfrei möglich. Es ist daher im Einzelfall die Aufgabe des Nutzers eines bestehenden Datenkörpers für Dritte nachvollziehbar zu begründen, ob die geplante Analyse eine Sekundärdatenanalyse darstellt. Entsprechend dem Ergebnis dieser Bewertung gelten dann primär die Leitlinien und Empfehlungen der Guten Epidemiologischen Praxis (GEP) und/oder der Guten Praxis Sekundärdatenanalyse (GPS).

Dateneigner: Unter diesem Begriff werden im Rahmen der GPS diejenigen Institutionen verstanden, die die Daten (primär) erheben, speichern und nutzen. Dateneigner und **Primärnutzer** sind als Synonym zu verstehen. Der Begriff Dateneigner hebt jedoch zusätzlich hervor, dass der Primärnutzer auch die rechtliche Verfügungsgewalt über die Daten besitzt. Im Bereich der Gesetzlichen Sozialversicherung sind Dateneigner beispielsweise Krankenkassen oder Rentenversicherungsträger, die versichertenbezogene (medizinische) Daten für administrative Aufgaben speichern, ebenso wie (Krebs-)Registerstellen, arbeitsmedizinische Untersuchungsstellen oder epidemiologische Einrichtungen.

Sekundärnutzer: Unter diesem Begriff werden im Rahmen der GPS diejenigen Stellen und Personen verstanden, die Daten von Dateneignern erhalten und zur Bearbeitung von Forschungsfragen unabhängig vom formalen primären Erhebungszweck aufbereiten und auswerten. Er drückt aus, dass es sich meist um andere Personen/Institutionen als die Dateneigner/Primärnutzer handelt.

Personenbezogene Daten: Unter personenbezogenen Daten sind im epidemiologischen Sinne solche Informationen zu verstehen, die einer einzelnen Person als Beobachtungseinheit zugeordnet werden können.

Sozialdaten. Darunter sind Routinedaten der Sozialversicherungsträger (u.a. Krankenversicherung, Rentenversicherung, Unfallversicherung) zu verstehen

Registerdaten sind Daten von krankheitsbezogenen, nicht notwendigerweise bevölkerungsbezogenen Registern.

Pseudonymisierung: *Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren (Bundesdatenschutzgesetz (BDSG) §3 (6a)).* Dabei werden die direkt personenidentifizierenden Daten (z. B. Name, Vorname, Telefonnummer, Sozialversicherungsnummer, Personalausweisnummer) aus den Daten entfernt und durch eindeutige Kennzeichen (z. B. eine Identifikationsnummer) ersetzt. Pseudonymisierte Daten sind weiterhin personenbezogen. Eine Pseudonymisierung ist insbesondere dann notwendig, wenn personenbezogene Daten über ein bekanntes Pseudonym bereits pseudonymisierten Daten zugeordnet werden sollen.

Anonymisierung: Definition nach BDSG §3 (6a) BDSG: *Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.*

Unter den Begriff Anonymisierung fallen Maßnahmen, welche in einem Datenkörper die Identifikation einer Person wirksam verhindern. Dies kann wie bei der Pseudonymisierung das Ersetzen der personenidentifizierenden Daten durch eine Identifikationsnummer sein. Allerdings darf dieses Ersetzen nicht umkehrbar sein. Dies wird beispielsweise erreicht, indem der für die Anonymisierung verwendete Schlüssel oder die Referenzliste vernichtet wird. Häufig werden bestimmte Merkmale mittels Klassifikation (Bildung von Altersgruppen, Kürzen von Postleitzahlen u.a.) bei gleichzeitiger Löschung der unklassifizierten Angaben verändert. Auch innerhalb eines vollständig anonymisierten Datenkörpers können verschiedene Angaben, z. B. im Längsschnitt, auf eine Beobachtungseinheit (ein Patient, ein Versicherter) durch Vergabe einer eindeutigen Kennziffer bezogen werden, aus der aber nicht auf die natürliche Person zurück geschlossen werden kann. Anonymisierte Daten fallen nicht mehr unter die datenschutzrechtlichen Bestimmungen über den Umgang mit personenbezogenen Daten.

An der ersten Version der GPS haben mitgewirkt: Susanne Bisson (Hannover), Wolfgang Bödecker (Essen), Siegfried Geyer (Hannover), Thomas Grobe (Hannover), Werner Hofmann (Hildesheim), Peter Ihle (Köln), Jürgen John (Neuherberg), Hedy Kerek-Bodden (Berlin), Susanne Moebus (Essen), Marina Mosafer (Frankfurt), Rolf Müller (Bremen), Matthias Perleth (Berlin), Burgi Riens (Göttingen), Ingrid Schubert (Köln), Daniel Sieg (Bielefeld), Philipp Storz (Berlin), Enno Swart (Magdeburg)

An der Überarbeitung der GPS haben mitgewirkt: Siegfried Geyer (Hannover), Thomas Grobe (Hannover), Volker Harth (Bochum), Wolfgang Hoffmann (Greifswald), Werner Hofmann (Hildesheim), Peter Ihle (Köln), Alexander Katalinic (Lübeck), Stefanie Klug (Mainz), Thomas Lampert (Berlin), Ingo Langner (Bremen), Ute Latza (Hamburg), Matthias Möhner (Berlin), Andreas Seidler (Berlin), Claudia Spix (Mainz), Enno Swart (Magdeburg)