

Gute Praxis Sekundärdatenanalyse (GPS)

Leitlinien und Empfehlungen

3. Fassung 2012

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

Ziele der GPS

Mit der Guten Praxis Sekundärdatenanalyse (GPS) wird ein Standard für die Durchführung von Sekundärdatenanalysen nach wissenschaftlichen Grundsätzen formuliert.

Die GPS ergänzt die Leitlinien und Empfehlungen für Gute Epidemiologische Praxis 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) in der jeweils aktuellen Fassung (http://dgepi.de/fileadmin/pdf/leitlinien/GEP_mit_Ergaenzung_GPS_Stand_24.02.2009.pdf). Sie konkretisiert die in den GEP formulierten allgemeingültigen Qualitätsanforderungen in der Epidemiologie für den Bereich der Sekundärdatenforschung.

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, begründet von den Empfehlungen der GPS abzuweichen. Die GPS versteht sich als Richtschnur bei der Planung, Durchführung und Analyse von Studien auf der Basis von Sekundärdaten. Gleichzeitig dient sie als Grundlage für vertragliche Absprachen zwischen Dateneignern (Primärnutzer) und Sekundärdatennutzern (s. Glossar am Ende der GPS). Der Nutzer von Daten muss für Dritte nachvollziehbar begründen, ob geplante Untersuchungen eine Sekundärdatenanalyse im Sinne dieser GPS darstellen. Die GPS kann schließlich für Autoren, Reviewer und Leser eine Richtlinie bei wissenschaftlichen Veröffentlichungen darstellen.

Zielgruppe der GPS

Zielgruppe der GPS sind alle Personen, die sich unter wissenschaftlicher Perspektive und unter Anwendung wissenschaftlicher Methoden Sekundärdaten, ihrer Analyse und Interpretation zuwenden. Dies schließt auch die Dateneigner ein. Der Wirkungsbereich dieser Leitlinien bezieht jede Form von Sekundärdaten ein, fokussiert jedoch auf medizinische Sekundärdaten, also typischerweise Routinedaten der gesetzlichen Kranken-, Renten- und Unfallversicherung (Sozialdaten) oder Daten von (bevölkerungsbezogenen) Krankheitsregistern.

Methodik und Verfahren der Aktualisierung

Die Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi) veröffentlichte erstmals 2005 in Anlehnung an die Gute Epidemiologische Praxis (GEP; Hoffmann et al. 2005) eine „Gute Praxis Sekundärdatenanalyse“ (GPS) (Swart et al. 2005). 2008 wurde die erste Fassung der GPS durch Mitglieder der AGENS und der AG Epidemiologische Methoden der DGEpi, DGSMP und GMDS und unter Einbeziehung weiterer Experten evaluiert und umstrukturiert (Swart et al. 2008). Die Formulierung und Nummerierung der Leitlinien wurden dabei an die Gute Epidemiologische Praxis (GEP) angelehnt. Die Aktualität der GPS wird durch die AGENS im Zuge eines regelmäßigen wissenschaftlichen Austauschs geprüft und bei Bedarf revidiert. Nach Überarbeitung durch die AGENS zu Beginn des Jahres 2012 wird nun die 3. Fassung vorgelegt. Die GPS und die GEP stehen auf der Webseite der DGEpi zum Download bereit (www.dgepi.de).

Leitlinie 1: Ethik

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

Ethische Prinzipien werden in allgemeinen Menschen- und Bürgerrechten formuliert und werden im Rahmen der GPS daher nicht weiter konkretisiert.

Die Konsultation einer Ethikkommission ist bei Analysen, die ausschließlich auf Sekundärdaten basieren, nicht erforderlich. Im Einzelfall kann die Konsultation einer Ethikkommission geboten sein, wenn andere Regelwerke dies vorschreiben.

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 Untersuchungspopulation muss im Hinblick auf die Forschungsfrage begründet werden.

Der potentielle Nutzen einer Sekundärdatenanalyse lässt sich nur dann beurteilen, wenn als Ausgangspunkt der Untersuchung eine Forschungsfrage explizit formuliert wird. Dies ist eine wesentliche Voraussetzung für Planung und Bewertung des Studiendesigns und der Datenextraktion, aber auch des Zeit- und Kostenrahmens der geplanten Untersuchung. Erst die Formulierung der Forschungsfrage ermöglicht die Festlegung der Einzelheiten einer Sekundärdatenanalyse (s. Leitlinie 3).

Explorative und hypothesengeleitete Analysen sind bei der Formulierung der Forschungsfrage voneinander abzugrenzen.

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.

Empfehlung 3.2 – Datenbasis / Studienpopulation

Die Studienpopulation wird im Hinblick auf die Fragestellung ausgewählt. Hierbei sind insbesondere Erhebungszweck, Verfügbarkeit, Vollständigkeit, Merkmalsumfang und Merkmalsdifferenzierung als Auswahlkriterien heranzuziehen.

Empfehlung 3.3 – Vermeidung von Verzerrungen

Bereits bei der Planung von Studien soll einer Verzerrung (Bias) von Ergebnissen entgegen gewirkt werden, die unter anderem durch Selektion und Confounding entstehen kann. Bei Sekundärdatenanalysen ist darauf zu achten, dass Merkmale zur Kontrolle von Störgrößen in den Daten enthalten sind und berücksichtigt werden.

Empfehlung 3.4 – Repräsentativität, Generalisierbarkeit

Im Studienplan muss thematisiert werden, ob die Fragestellung auf Generalisierbarkeit und Repräsentativität der erwarteten Ergebnisse und Kernaussagen abzielt.

Empfehlung 3.5 – Variablen und Variablenwerte

Eine Sekundärdatenanalyse muss die Messgenauigkeit und Vollständigkeit der berücksichtigten Merkmale und mögliche Störfaktoren von der primären Informationserfassung bis zur Datenbereitstellung berücksichtigen. Zum Studienplan gehört die Beschreibung aller verwendeten Daten(-felder) bzw. Merkmale mit Aussagen zum Erhebungskontext, zu den Bereitstellungswegen sowie zu den jeweils zulässigen Merkmalsausprägungen. Dabei sind insbesondere auch Veränderungen der Merkmalserfassung im Studienzeitraum darzustellen (z.B. Veränderungen der verwendeten Klassifikationen von Diagnosen oder von Berufstätigkeiten, Veränderungen von Leistungsvergütungen und deren Erfassung etc.).

Empfehlung 3.6 – Studienumfang

Im Studienplan ist eine Begründung des Studienumfangs anzugeben. Zur Bestimmung der erforderlichen Populationsgrößen sollen quantitative Abschätzungen der statistischen Aussagekraft durchgeführt werden.

Leitlinie 4: Probendatenbanken

Immer häufiger werden biologische Probenbanken angelegt. Werden die aus den Proben gewonnenen Informationen (Probendaten) in Sekundäranalysen genutzt, befinden sie sich im Wirkungsbereich der Guten Praxis Sekundärdatenanalyse und unterliegen damit auch den hier formulierten Leitlinien.

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 Sekundärdaten-

nutzer 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.

Empfehlung 5.1 – Testdatensatz

Im Vorfeld der Sekundärdatenanalyse sollte die Möglichkeit der Bereitstellung einer Stichprobe des angestrebten Analysedatensatzes erwogen werden, um die Übermittlung und Verwendbarkeit der Daten, insbesondere hinsichtlich der Datenformate und -struktur, beurteilen zu können.

Empfehlung 5.2 – Anpassung des Studienplans

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

Empfehlung 5.3 – Ressourcen

Für die Qualitätssicherung der Studie und für die erforderlichen Absprachen innerhalb und zwischen den beteiligten Stellen sind ausreichende zeitliche und personelle Ressourcen beim Dateneigner und Sekundärdatennutzer vorzusehen.

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.

Empfehlung 6.1 – Datenerhebung und Datenübermittlung

Der Entstehungszusammenhang und die Weiterverarbeitung der Daten bis zur Übermittlung an den Sekundärdatennutzer müssen nachvollziehbar dokumentiert sein, inkl. des Datums der Lieferung. Dies umfasst eine Darstellung des ursprünglichen Verwendungszwecks der Datensammlung, der Erfassungsregeln und die Überprüfung ihrer zeitlichen Konstanz (wie z.B. bei Diagnose- oder Prozedurenkodierungen).

Empfehlung 6.2 – Ausgangsdatsatz

Der vom Dateneigner übermittelte Ausgangsdatsatz 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

Übertragungsformat, Umfang und Struktur der bereitgestellten und verwendeten Daten werden dokumentiert. Der Umfang fehlender sowie redundanter Daten ist festzuhalten.

Empfehlung 6.4 – Plausibilitätskontrollen

Vor und während der Sekundärdatenanalyse sind Plausibilitätskontrollen durchzuführen. Eventuell erforderliche Korrekturen liegen im Verantwortungsbereich des Sekundärdatennutzers. Dabei wird eine Rücksprache mit dem Dateneigner empfohlen. Alle Ergänzungen und Änderungen sind vollständig und schriftlich zu dokumentieren und zu begründen.

Empfehlung 6.5 – Operationalisierung und abgeleitete Variablen

Die Sekundärdatenanalyse umfasst in der Regel die Bildung von Merkmalen und Merkmalskategorien auf der Basis mehrerer Variablen des Ausgangsdatensatzes und deren Ausprägungen. Die Bildung abgeleiteter Variablen ist nachvollziehbar zu dokumentieren und folgt vorhandenen Standards, um eine Vergleichbarkeit der Ergebnisse zu erleichtern.

Empfehlung 6.6 – Auswertungsdatensatz

Der im Verlauf der Plausibilitätsprüfung und der Datentransformation entstandene Datensatz ist als Auswertungsdatensatz zu kennzeichnen und zu sichern. Alternativ ist die Wiederherstellbarkeit aus dem Ausgangsdatensatz zu gewährleisten. Für die Reproduzierbarkeit der Analysen gilt zusätzlich die in Leitlinie 7 genannte Aufbewahrungsfrist.

Empfehlung 6.7 – Operationshandbuch

Ergänzend zum Studienplan sollen 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 Sekundärdatennutzer und die Aufbereitung durch diesen ein.

Leitlinie 7: Datenanalyse

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

Um die Überprüfbarkeit der Studienergebnisse zu ermöglichen, ist eine Aufbewahrungsfrist von Ausgangs- und Auswertungsdatensatz sowie der Datenaufbereitungsprotokolle beim Sekundärnutzer vertraglich zu regeln.

Die einzelnen Auswertungsschritte müssen nachvollziehbar sein und hinsichtlich möglicher Rückwirkungen auf die Selektion der Studienpopulation, die Operationalisierung der Forschungsfrage, die Reichweite der Ergebnisse und die Entscheidungskriterien bei der Hypothesenprüfung 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 möglicher Störfaktoren und Wechselwirkungen) der Daten berücksichtigen. Die im Rahmen einer Sekundärdatenanalyse zu überprüfenden Hypothesen sollen vor Beginn der Untersuchung formuliert werden, ebenso die für diese Prüfungen anwendbaren 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 und methodischen Wissens

erfolgen. Die Auswertungsstrategie muss nachvollziehbar begründet werden. Änderungen des ursprünglichen Auswertungsplanes sind zu dokumentieren.

Empfehlung 7.2 – Personelle Verantwortung

Jede an den Datenanalysen beteiligte Person muss vor Beginn der Arbeit benannt und in die Arbeit mit den Daten eingewiesen werden.

Empfehlung 7.3 – Prüfung der Ergebnisse

Die Ergebnisse von Sekundärdatenanalysen sollen vor der Publikation einer Gegenprüfung unterzogen werden. Alle Auswertungen sollen so 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 Prozedurdateien 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 rechtliche Verankerung des Datenschutzes findet sich in den Regelungen des Datenschutzrechts auf Bundes- und auf Landesebene (BDSG und DSG der Länder) sowie in allgemeinen gesetzlichen Regelungen, z.B. Sozialgesetzbücher V und X, Infektionsschutzgesetz, Röntgenverordnung, Krebsregistergesetze, Krankenhausgesetze, Gesundheitsdienstgesetze. Die Gesetze auf Landesebene unterscheiden sich erheblich, und auch die Auslegung der Gesetze hängt von dem jeweiligen Landesbeauftragten für den Datenschutz ab. Zu den geltenden Datenschutzvorschriften zählt auch das Prinzip der Datenvermeidung und der Datensparsamkeit (§ 3a BDSG).

Da Forschung mit personenbezogenen Daten immer im Spannungsfeld des Rechts des Einzelnen auf informationelle Selbstbestimmung und des Rechts auf Freiheit der Wissenschaft und Forschung steht, ist der Datenschutz essentieller Bestandteil jeder Sekundärdatenanalyse.

Empfehlung 8.1 – Zweck der Datenbereitstellung

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

Empfehlung 8.2 – Pseudonymisierung und Anonymisierung

Von den im Bundesdatenschutzgesetz aufgeführten Möglichkeiten der Pseudonymisierung und Anonymisierung ist Gebrauch zu machen. Hierbei ist auch die Einbindung eines Datentreuhänders zu prüfen.

Fehler bei der Anonymisierung und Pseudonymisierung können den Rohdatensatz verändern und Auswertungsergebnisse verfälschen. Der Sekundärdatennutzer muss daher den Vorgang der Anonymisierung und Pseudonymisierung durch geeignete Maßnahmen prüfen können, beispielsweise durch stichprobenhafte Prüfung und Sichtung der Ausgangs- und anonymisierten/pseudonymisierten Datensätze oder durch Pseudonymisierung von Testdaten.

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 liegt im Zuständigkeitsbereich des Dateneigners. Eine geplante Weitergabe des Roh- bzw. des aufbereiteten Analysedatensatzes ist mit dem Dateneigner abzustimmen und vertraglich zu regeln.

Empfehlung 8.5 – Personenbezogenes Datenlinkage mit externen Datenquellen

Jedes Datenlinkage mit weiteren Datenquellen bedarf der Einhaltung datenschutzrechtlicher Vorschriften.

Empfehlung 8.6 – 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 Auswertungsdatsatz erfolgen. Ebenso muss bei der Festlegung von Löschfristen die nach Leitlinie 7 geforderte Möglichkeit zur Überprüfung der Ergebnisse aus einer Sekundärdatenanalyse mitbedacht werden.

Empfehlung 8.7 – Zusammenarbeit mit Datenschutzbeauftragten

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

Empfehlung 8.8 – Datensicherheit

Über die hier spezifisch formulierten datenschutzrechtlichen Empfehlungen hinaus sind die allgemeinen Datensicherheitsvorschriften beim Umgang mit personenbezogenen Daten zu beachten (s. beispielsweise IT-Grundschutz des Bundesamtes für Sicherheit in der Informationstechnik). Die Einbindung eines Experten für Datensicherheit wird empfohlen.

Empfehlung 8.9 – Datenschutzkonzept

Alle im Rahmen einer Sekundärdatenanalyse notwendigen datenschutzrechtlichen Punkte sind in einem Datenschutzkonzept schriftlich zu fixieren. Hierzu zählen vor allem die Aspekte der Datenübermittlung, Datenhaltung, die Dauer der Aufbewahrung, sowie die Zugriffsberechtigungen. Das Datenschutzkonzept sollte idealerweise Bestandteil des Vertrages mit dem Dateneigner sein. Bei länger dauernden Studien ist das Datenschutzkonzept ggf. zu aktualisieren, sinnvollerweise unter Hinzuziehung der jeweils zuständigen Datenschutzbeauftragten.

Empfehlung 8.10 – Verantwortlicher für den Datenschutz

Innerhalb einer Forschungseinheit ist eine für den Datenschutz verantwortliche Person zu benennen, die die Einhaltung dieser Standards zu überwachen hat. Diese Person ist entsprechend ihrer Aufgaben zu qualifizieren.

Empfehlung 8.11 – Datenschutzerklärung, Schweigepflicht

Alle Personen, die im Rahmen eines Forschungsprojektes Umgang mit personenbezogenen Daten haben, müssen über Inhalte, Reichweite und Möglichkeiten der einschlägigen gesetzlichen Bestimmungen informiert sein. Sie müssen die datenschutzrechtlichen Bestimmungen anerkennen und schriftlich ihre Schweigepflicht erklären.

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 personenbezogenen Daten sind rechtswirksame Vereinbarungen der Kooperationspartner anzustreben, in denen die rechtlichen und finanziellen Rahmenbedingungen der intendierten Sekundärdatennutzung festgelegt werden. Die Zielsetzung der Sekundärdatenanalysen und deren Inhalte sollten angesprochen werden, ohne die Forschung durch Auflagen einzuschränken. Bei der Vielfalt der speziellen Konstellationen und möglicher Datenquellen sind unterschiedliche Vertragsformen möglich.

Empfehlung 9.1 – Vertragsinhalte

Mit dem Dateneigner sollen transparente und realistische Vereinbarungen getroffen werden. Der Vertrag muss die Unabhängigkeit der Forschung, Rechte und Pflichten der Forschung, incl. der Modalitäten von Publikationen, und den langfristigen Zugang zu den Daten thematisieren. Aspekte des Datenschutzes sollen explizit angesprochen werden.

Empfehlung 9.2 – Verwendung der Ergebnisse

Die wissenschaftliche Publikation sowie die Verwendung von Ergebnissen aus der Auftragsforschung für wissenschaftliche und Unterrichtszwecke darf nicht verhindert, behindert oder unzumutbar verzögert werden. Die vertragliche Festsetzung einer Sperrfrist von maximal sechs Monaten ist zulässig.

Empfehlung 9.3 – Sekundärdatenanalysen für Dritte

Sekundärdaten werden vom Dateneigner in der Regel für a priori definierte Forschungsvorhaben und begrenzte Zeiträume zur Verfügung gestellt (s. Leitlinie 3). Analysen, die über den festgelegten Forschungsrahmen hinausgehen und/oder Anfragen nach Sekundärdatenanalysen für wissenschaftliche Kooperationen, die an den Sekundärdatennutzer herangetragen werden, müssen an den Dateneigner zur Entscheidung weitergeleitet werden. Zu der jeweiligen Anfrage kann der Sekundärdatennutzer gegenüber dem Dateneigner ein Votum abgeben. Analysen dürfen aber erst durchgeführt werden, nachdem der Dateneigner den Sekundärdatennutzer dazu autorisiert hat.

Leitlinie 10: Interpretation und wissenschaftliche Publikation

Die Interpretation der Ergebnisse einer Sekundärdatenanalyse ist Aufgabe des Sekundärdatennutzers. Grundlage jeder Interpretation ist eine kritische Diskussion der Methoden, Daten und Ergebnisse der eigenen Untersuchung im Kontext der vorhandenen Evidenz. Eine wissenschaftliche Publikation in einem Fachmedium ist anzustreben. Publikationen sollen einem externen Review unterzogen werden.

Die Beurteilung der Ergebnisse gehört zu den originären Aufgaben der Sekundärdatennutzer. Der argumentative Prozess, der einer Interpretation zugrunde liegt, muss in einer schriftlichen Diskussion transparent und nachvollziehbar dargestellt werden, zumindest in Form eines Berichtes an den Dateneigner, idealerweise in Form einer wissenschaftlichen Publikation in einem Fachmedium. Einschränkungen bzgl. der Übertragbarkeit der Studienergebnisse auf andere Populationen oder andere Zeiträume als die betrachteten sind zu diskutieren. Die Orientierung an geläufigen Standards zur wissenschaftlichen Publikation von empirischen Studien wird empfohlen. Dabei sollen ggf. Adaptationen an die Rahmenbedingungen der Sekundärdatenanalysen vorgenommen werden.

Leitlinie 11: Kommunikation und Public Health

Ergebnisse aus Sekundärdatenanalysen, die öffentliches Interesse berühren, sollen der Öffentlichkeit in verständlicher Form zugänglich gemacht werden. Dies schließt eine qua-

ifizizierte Risikokommunikation ein.

Sekundärdatenanalysen können die Bewertung von Strukturen und Leistungen des Gesundheitssystems oder die Evaluation der Umsetzung gesundheitswirksamer Maßnahmen zum Gegenstand haben. Dies um so mehr, als solche Analysen häufig auf Daten der Versorgungspraxis basieren. Ergibt sich nach dem fachlichen Urteil der Untersucher aus den Ergebnissen einer Sekundärdatenanalyse die Notwendigkeit von Konsequenzen, so sollen diese, beispielsweise in Form einer Empfehlung, explizit formuliert werden können. Dabei tragen die Sekundärdatennutzer auch Verantwortung für eine verständliche Kommunikation ihrer Forschungsergebnisse an Nicht-Fachleute. Darüber hinaus können sie fachlich begründete Empfehlungen an den Dateneigner zur öffentlichen Bereitstellung von Informationen formulieren und an deren technischen Umsetzung mitwirken.

Empfehlung 11.1 – Unabhängigkeit der Darstellung der Ergebnisse

In Unabhängigkeit ihrer Arbeit sollen Sekundärdatennutzer die Möglichkeit besitzen, sich zu den praktischen Konsequenzen ihrer Analysen für die betroffenen Bevölkerungsgruppen kurzfristig, direkt und ohne Abstimmung mit dem Dateneigner zu äußern. Formale Aspekte der direkten Kommunikation der Sekundärdatennutzer mit der Öffentlichkeit sollen vertraglich festgelegt werden. Dazu gehören die Regelung der vorherigen Information der Dateneigner und das Ausmaß der Abstimmung zwischen Dateneigner und Sekundärdatennutzer.

Empfehlung 11.2 – Transparenz der Methoden

Die in einer Sekundärdatenanalyse verwendeten Methoden sollen an geeigneter Stelle, z. B. in wissenschaftlichen Zeitschriften oder auf Fachtagungen, veröffentlicht und interessierten Personen und Institutionen im Sinne der Methodentransparenz und grundsätzlichen Replizierbarkeit der Analysen zugänglich gemacht werden. Die Weitergabe oder Publikation von Auswertungsroutinen, analytischen Verfahren und Berichtsformaten soll unabhängig vom Dateneigner erfolgen können. Das Recht zur Weitergabe dieser Elemente einer Sekundärdatenanalyse erstreckt sich nicht auf die Weitergabe der Daten an sich.

Empfehlung 11.3 – Externe Datennutzung

Bei jeder Sekundärdatenanalyse soll von Dateneignern und Sekundärdatennutzern geprüft werden, ob und in welcher Form der Auswertungsdatensatz (oder Teile davon) im Sinne der Daten und Methodentransparenz für wissenschaftliche Kooperationen bereit gestellt werden kann, wenn dies für die öffentliche Kommunikation sinnvoll und notwendig ist. Die abschließende Entscheidung verbleibt beim Dateneigner.

Quellen

Hoffmann W, Latza U, Terschüren C. Leitlinien und Empfehlungen zur Sicherung Guter Epidemiologischer Praxis (GEP) – überarbeitete Fassung nach Revision. Das Gesundheitswesen 2005; 67: 217-225

Swart E, Ihle P, Geyer S, Grobe T, Hoffmann W. GPS - Gute Praxis Sekundärdatenanalyse. Gesundheitswesen 2005; 67: 416-421

Swart E, Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten und Arbeitsgruppe Epidemiologische Methoden. GPS - Gute Praxis Sekundärdatenanalyse: Revision nach grundlegender Überarbeitung. Das Gesundheitswesen 2008; 70: 54-60

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 zur Bearbeitung von Forschungsfragen unabhängig vom formalen primären Erhebungszweck aufbereiten und sich unter wissenschaftlicher Perspektive und unter Anwendung wissenschaftlicher Methoden ihrer Analyse und Interpretation zuwenden.

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 wei-

terhin 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.