

Daten der gesetzlichen und privaten Krankenversicherung in der NAKO Gesundheitsstudie

Nutzermanual

VERSION 1.0 / STAND: 31.03.2025

[DOI: 10.4126/FRL01-006510711](https://doi.org/10.4126/FRL01-006510711)

© Die Autoren, NAKO e. V. 2025

Das Manual ist unter der CC BY 4.0 veröffentlicht.

<https://creativecommons.org/licenses/by/4.0/>



Inhaltsverzeichnis

1. Einleitung	4
2. Antragsverfahren	4
3. Auswahl der Krankenversicherungsdaten im Transferhub	6
3.1 Datenstruktur	8
3.1.1 Teilnehmendenpseudonym ID-A.....	8
3.1.2 Struktur der Krankenversicherungsdaten	8
3.2 Inhalte der Krankenversicherungsdaten	8
3.2.1 Sektoren der Krankenversicherungsdaten.....	8
3.2.2 Hinweise zum Sektor „Stammdaten“	9
3.2.3 Weitere Schlüsselvariablen in den Sektoren.....	10
3.2.4 Variablen- und Wertelabels	11
3.2.5 Lebenslange Arztnummer, Betriebsstättennummer und Institutionskennzeichen.....	11
4. Datensatzbeschreibung und Datenbereitstellung	12
4.1 Datensatzbeschreibung	12
4.2 Datenbereitstellung	13
4.3 Record Linkage	13
4.4 Weiterführende Auswertungshinweise	14
4.4.1 Format von Codes	14
4.4.2 Verfügbarkeit bestimmter Variablen und Tabellen	14
4.4.3 Sekundär Diagnosen im Krankenhaus ohne Zuordnung zur Primär diagnose	15
5. Schulungen und Standarddokumente	15
5.1 Schulungen	15
5.2 Standardliteratur	15
6. Autorenliste	16
7. Literaturverzeichnis	17

Tabellenverzeichnis

Tab. 1: Übersicht der Datentabellen Krankenversicherungsdaten im Transferhub	9
Tab. 2: Zusätzliche Schlüsselvariablen nach Tabelle in den Krankenversicherungsdaten	10
Tab. 3: Beispiel Zuordnung von Sekundär Diagnosen in der stationären Versorgung	15

Abbildungsverzeichnis

Abb. 1: Ablauf des Antragsverfahrens für GKV-/PKV-Daten	4
Abb. 2: Startseite des Transferhubs der NAKO	7
Abb. 3: Datenverzeichnis Krankenkassendaten	7
Abb. 4: Beispiel Variablen in Tabelle „Ambulante Versorgung: Diagnosen“	9
Abb. 5: Darstellung der Abhängigkeit von Schlüsselvariablen am Beispiel Arbeitsunfähigkeit	11
Abb. 6: Beispiel Aufbau <i>NAKO data dictionary</i> . Versicherungstyp aggregiert	12
Abb. 7: Variablennamen für das Beispiel ambulante Diagnosen	12

1. Einleitung

Forschungsprojekte können Studiendaten der NAKO grundsätzlich über den Transferhub der Studie (<https://transfer.nako.de/transfer/index>) im Rahmen eines Use & Access Verfahrens beantragen. Auf der Startseite des Transferhubs werden hierfür die notwendigen Informationen zur Antragsberechtigung und zum Antragsverfahren bereitgestellt.

Das vorliegende Dokument dient der Beschreibung der Beantragung der in der NAKO Gesundheitsstudie (NAKO) verfügbaren Routinedaten von gesetzlichen und privaten Krankenversicherungen (GKV-/PKV-Daten) sowie ergänzend der Erläuterung der grundlegenden Struktur und Inhalte der entsprechenden Daten. Zudem werden Hinweise zur Notwendigkeit der Beantragung von sogenannten Schlüsselvariablen sowie zu Besonderheiten der Daten gegeben.

Das Dokument stellt keine Grundlagenliteratur zur Arbeit mit und Analyse von GKV-/PKV-Daten dar. Hinweise zu entsprechenden Kursen und zu Grundlagenliteratur finden Sie in Kapitel 5.

2. Antragsverfahren

Das nachfolgend beschriebene Verfahren (*Abb. 1*) dient der Orientierung für Wissenschaftler*innen, die über das Use & Access Verfahren ausschließlich GKV-/PKV-Daten oder diese zusätzlich zu weiteren Daten der NAKO beantragen möchten. Vor der Antragstellung wird empfohlen, die Expertengruppe Sekundär- und Registerdaten/Mortalitäts-Follow-up zu konsultieren.

Prüfung der Antragsberechtigung

Vor der Entwicklung einer Forschungsfrage / eines Forschungsprojekts sollten interessierte Wissenschaftler*innen prüfen, ob sie entsprechend der Nutzungsordnung der NAKO antragsberechtigt sind. Die jeweils aktuelle Nutzungsordnung sowie ergänzende Dokumente sind auf dem Datenportal der NAKO hinterlegt: <https://www.nako.de/transferhub> (deutsch und englisch).

Wenn Sie Fragen bezüglich der Antragsberechtigung oder allgemein zum Transferhub haben, treten Sie bitte mit der Transferstelle der NAKO in Kontakt: transfer@nako.de

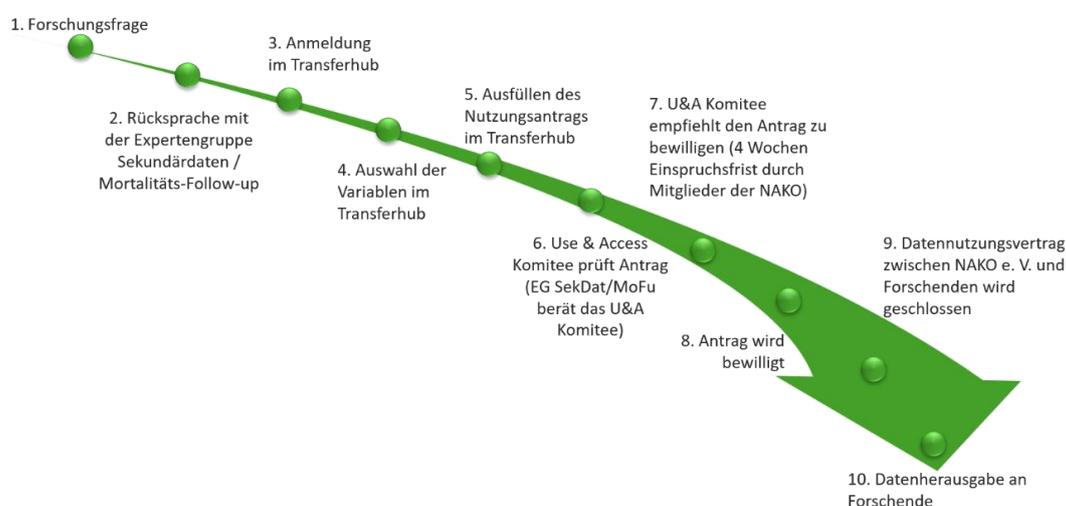


Abb. 1: Ablauf des Antragsverfahrens für GKV-/PKV-Daten

Im Folgenden werden die zehn Schritte des Antragsverfahrens näher erläutert.

1. Forschungsfrage

Die Forschungsfrage sollte klar definiert und die Nutzung der GKV-/PKV-Daten im Rahmen des Forschungsdesigns/ der Methodik klar beschrieben und begründet sein. Dies beinhaltet auch, möglichst alle relevanten Sektoren und ggf. erforderlichen Variablen (einschließlich Hilfs- und Schlüsselvariablen) festzulegen. Zudem Datennutzungsanträge nicht auf eine Datenart (Primär- oder Sekundärdaten) beschränkt. Alle im Transferhub verfügbaren Datensektoren können beantragt werden, sofern sie für die Beantwortung einer Fragestellung erforderlich sind. Die Forschungsfrage bildet die Grundlage für die Antragstellung und muss im Datenportal der NAKO hinterlegt werden. Eine Zusammenfassung wird ebenfalls gefordert. Diese sollte den etablierten wissenschaftlichen Kriterien gerecht werden und entsprechend in Hintergrund/Ziel/Methodik/erwartete Schlussfolgerungen gegliedert sein.

2. Rücksprache mit der Expertengruppe Sekundär- und Registerdaten/Mortalitäts-Follow-up

Dieser Schritt ist optional und dient vor der Einreichung eines Datennutzungsantrags der Konkretisierung der zu beantragenden Daten. Hierdurch können zeitintensive Rückfragerunden zwischen Use & Access Komitee, der Expertengruppe Sekundärdaten/Mortalitäts-Follow-up und den Antragsteller*innen nach Einreichung des Nutzungsantrags vermieden und somit das Antragsverfahren beschleunigt werden (siehe auch Schritt 6). Für die Rücksprache nehmen Sie bitte Kontakt mit der Expertengruppe Sekundärdaten/Mortalitäts-Follow-up der NAKO auf. Als zentrale Kontaktadresse verwenden Sie bitte: christoph.stallmann@med.ovgu.de

Sie erhalten dann einen Termin, an dem Sie Ihr Forschungsvorhaben in einem Kurzvortrag (ca. 5-Minuten) vorstellen können. Die Expertengruppe Sekundärdaten/Mortalitäts-Follow-up beantwortet Ihnen Fragen und gibt Ihnen Rückmeldung zu Ihrem Vorhaben.

3. Anmeldung im Transferhub (Datenportal) der NAKO

Registrieren Sie sich und melden Sie sich im Transferhub der NAKO an: <https://www.nako.de/transferhub>

4. Auswahl der Variablen im Transferhub

Nach erfolgreicher Anmeldung im Transferhub können Sie im ersten Schritt die für Ihren Forschungsantrag erforderlichen Variablen auswählen. Die GKV-/PKV-Daten befinden sich im Verzeichnis „Sekundär- und Registerdaten“ → „Krankenkassendaten“. Die Daten sind nach Versorgungssektoren aufgelistet. Speichern Sie Ihre Auswahl als entsprechenden Datensatz.

Ihrer Forschungsfrage entsprechend ist es auch möglich, Daten aus anderen Bereichen der NAKO mit auszuwählen.

5. Ausfüllen des Nutzungsantrags

Nach Auswahl und Speicherung der erforderlichen Variablen können Sie den Datennutzungsantrag ausfüllen. Zwischenstände können gespeichert und zu einem späteren Zeitpunkt wieder bearbeitet werden. Korrekturen in der Datenauswahl sind ebenfalls möglich. Wenn das Antragsformular mit allen erforderlichen Informationen ausgefüllt ist, kann der Antrag eingereicht werden.

6. Use & Access Komitee prüft den Antrag

Das Use & Access Komitee der NAKO prüft jeden eingereichten Antrag hinsichtlich der Einhaltung der Nutzungsordnung der NAKO. Bei Forschungsanträgen, die zusätzlich oder ausschließlich GKV-/PKV-Da-

ten beinhalten, wird vor der Beschlussfassung zusätzlich die Expertengruppe Sekundärdaten/Mortalitäts-Follow-up konsultiert. Die Expertengruppe verfasst eine Stellungnahme. Falls die Expertengruppe entsprechend Schritt 2. bereits in den Prozess der Antragstellung eingebunden war, erfolgt eine vereinfachte Stellungnahme mit Hinweis auf die Vorab-Konsultation. Wurde vorab kein Kontakt zur Expertengruppe aufgenommen, kann es vonseiten der Expertengruppe zu Auflagen für die Datennutzung und/oder zu der Bitte um Rücksprache mit den Antragstellenden kommen. Die rechtzeitige Konsultation der Expertengruppe wird daher ausdrücklich empfohlen.

7. Use & Access Komitee empfiehlt die Bewilligung des Antrags

Nach Prüfung des Antrags durch das Use & Access Komitee und unter Berücksichtigung der Stellungnahme der Expertengruppe Sekundärdaten/Mortalitäts-Follow-up wird der Antrag nach positivem Entscheid zur Annahme empfohlen. Ab diesem Tag gilt für die an der NAKO beteiligten Institutionen eine Einspruchsfrist von vier Wochen.

8. Bewilligung des Antrags

Sofern kein Einspruch erfolgt ist oder die Einsprüche bearbeitet wurden, wird der Antrag i. d. R. nach den vier Wochen bewilligt. Die Antragsteller*innen werden per E-Mail informiert.

9. Vertrag zur Datennutzung

Zwischen dem NAKO e. V. und den Wissenschaftler*innen wird ein Datennutzungsvertrag geschlossen. Dieser bildet die rechtliche Grundlage für die Datenübermittlung und -nutzung. Die Vertragsvorlage wird dem/der hauptverantwortlichen Antragsteller*in durch die Transferstelle der NAKO zur Verfügung gestellt.

10. Datenherausgabe

Ist der Datennutzungsvertrag von allen Parteien unterzeichnet, stellt die Transferstelle die für das Forschungsvorhaben erforderlichen Daten zusammen. Die Antragsteller*innen werden nach erfolgter Zusammenstellung benachrichtigt. Die Daten können dann aus dem Transferhub heruntergeladen werden. Zusätzlich wird den Forscher*innen ein data dictionary zur Verfügung gestellt, in dem alle im Antrag sowie Datensatz enthaltenen Daten inkl. der Ausprägungen einzelner Variablen beschrieben sind.

3. Auswahl der Krankenversicherungsdaten im Transferhub

Die Auswahl von Daten für Forschungsprojekte erfolgt über den Transferhub der NAKO Gesundheitsstudie (Abb. 2). Der Transferhub kann über die Adresse <https://www.nako.de/transferhub> erreicht werden. Dieses Datenportal enthält das Datenverzeichnis und eine Funktion zur Stellung eines Datennutzungsantrags. Zudem können Zusammenfassungen bereits genehmigter Forschungsprojekte eingesehen werden.

Für die Datenauswahl und Antragstellung ist eine Registrierung aller Wissenschaftler*innen, die an einem Antrag beteiligt sind, erforderlich (siehe oben Schritt 3).

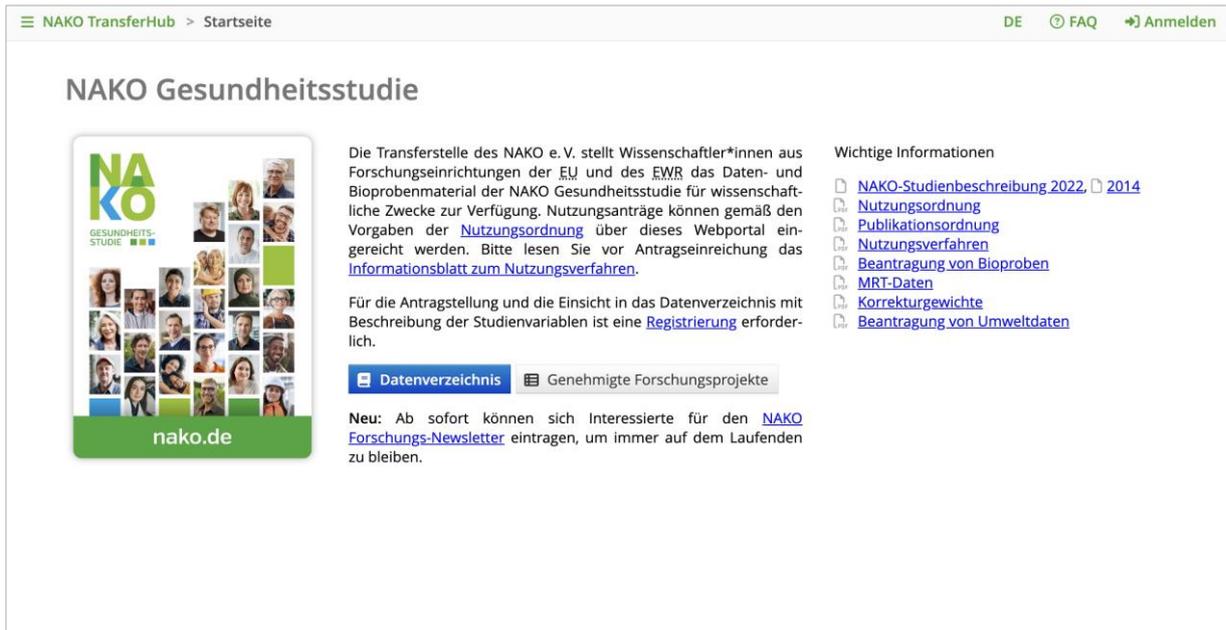


Abb. 2: Startseite des Transferhubs der NAKO

Die Krankenversicherungsdaten befinden sich nach Leistungssektoren (z. B. vertragsärztliche oder stationäre Versorgung) aufgelistet im Verzeichnis „Sekundär- und Registerdaten“ und dann „Krankenkassendaten“ (siehe *Abb. 3*). Es handelt sich dabei um eine gemeinsame Datenbank für GKV-/PKV-Daten. Die Unterscheidung/Selektion beider Versicherungsarten kann anhand einer hierfür vorgesehenen Variable erfolgen (siehe Kapitel 3.2.2). Werden Daten zusammengestellt, befinden sich in den Tabellen immer Daten aller bis zum Zeitpunkt der Zusammenstellung in der Datenbank befindlichen Daten.

Projektbezogene Datenauswahl

Die Auswahl von einzelnen Variablen hat stets geleitet durch die Fragestellung zu erfolgen. Eine Auswahl aller Krankenversicherungsdaten, die nicht durch die Forschungsfrage begründet werden, ist daher nicht zweckmäßig und führt zu einer Ablehnung bzw. Auflage zur Überarbeitung eines Forschungsantrags.

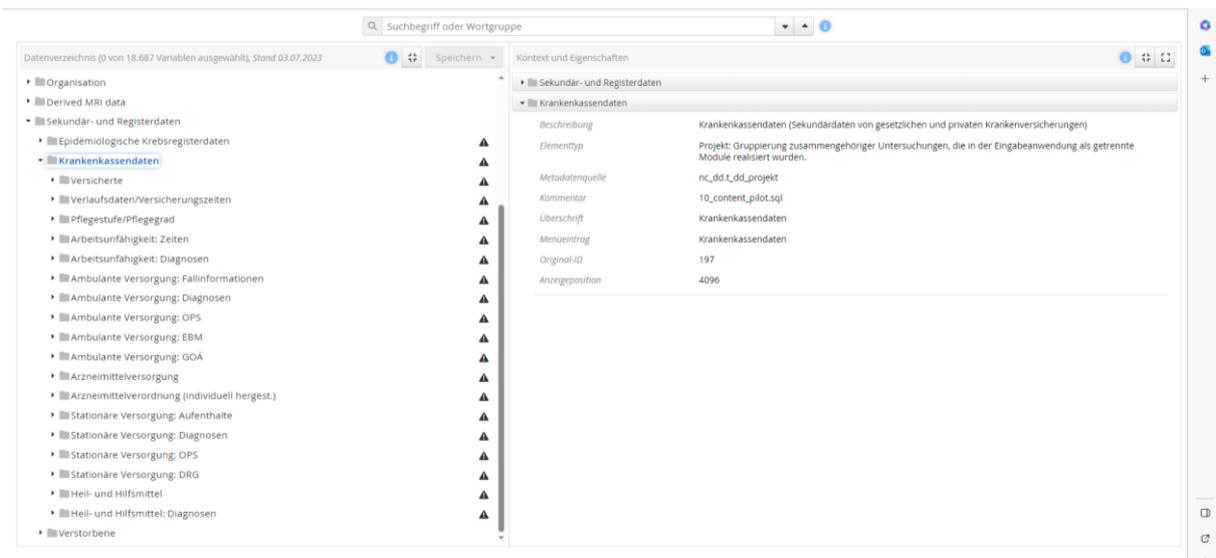


Abb. 3: Datenverzeichnis Krankenkassendaten im Transferhub der NAKO

3.1 Datenstruktur

Zur Orientierung wird für die Auswahl aller relevanten Variablen im Nachfolgenden ein Überblick über die Datenstruktur sowie -inhalte gegeben.

Alle die Krankenversicherungsdaten betreffenden Tabellen werden an Forschende im csv-Format übergeben.

3.1.1 Teilnehmendenpseudonym ID-A

Für jedes über den Transferhub der NAKO beantragte Forschungsvorhaben wird jedem in den Daten abgebildeten NAKO-Teilnehmenden ein projektspezifisches neues Pseudonym „ID-A“ zugewiesen. Eine Verknüpfung mit Daten anderer Forschungsvorhaben ist daher ausgeschlossen. Unabhängig davon, welche Variablen für ein Vorhaben ausgewählt werden, ist dieses Pseudonym in allen übergebenen Tabellen eineindeutig einem NAKO-Teilnehmenden zugeordnet. Die ID-A wird automatisch für jede Tabelle mitgeliefert und kann nicht selektiert/deselektiert werden. Durch die ID-A können unterschiedliche Inhalte der Primärdaten der NAKO (bspw. Selbstangaben zu Erkrankungen) mit anderen angebotenen Daten (bspw. den GKV-/PKV-Daten) verknüpft werden.

3.1.2 Struktur der Krankenversicherungsdaten

Den Krankenversicherungsdaten liegt ein relationales Datenbankmodell mit derzeit 18 Tabellen zugrunde (Gesamtübersicht siehe

Anhang 1). Die Tabellen sind untereinander anhand des enthaltenen projektbezogenen Pseudonyms **ID-A** auf Ebene der Teilnehmenden miteinander verknüpfbar (siehe Kapitel 3.2). Weitere Schlüsselvariablen können notwendig sein (siehe Unterkapitel 3.2.3). Die Struktur der Ordner im Transferhub entspricht dabei den 18 Tabellen der relationalen Datenbank.

3.2 Inhalte der Krankenversicherungsdaten

3.2.1 Sektoren der Krankenversicherungsdaten

Die enthaltenen Versorgungssektoren und Variablen orientieren sich an den in der Versorgungsrealität etablierten sowie für die Sekundärdatenforschung gemeinhin bekannten Strukturen (Beispiele in Klammern):

- *Stammdaten* (Versicherungszeiten, Versicherungstyp)
- *Ambulante ärztliche Versorgung* (Haus- und andere Fachärzte, ambulante Leistungen, Diagnosen)
- *Stationäre Versorgung* (Krankenhaus; auch ambulante Leistungen der KH, Diagnosen)
- *Arzneimittelversorgung* (bei Apotheken eingelöste Rezepte)
- *Pflege* (Pflegrade/-stufen, Pflegezeiten)

In Teilstichprobe zusätzlich:

- *Heil- und Hilfsmittel* (Physiotherapie, Ergotherapie, Sehhilfen, Rollator etc.)
- *Arbeitsunfähigkeit* (Krankheitstage und Grund der AU)
- *Arzneimittelversorgung: Bestandteile individuell hergestellte Lösungen*
- *Stationäre Versorgung: DRG*

Aufgrund der NAKO-Datenbankstruktur sowie der Granularität der hinterlegten Datentabellen lassen sich die in *Abb. 3*, linke Seite hinterlegten Datenordner für die Krankenversicherungsdaten wie folgt einordnen:

Tab. 1: Übersicht der Datentabellen Krankenversicherungsdaten im Transferhub

Ordner	Sektor	Unterer/Untere/Tabelle
Krankenkassendaten	Stammdaten	Versicherte
		Verlaufsdaten/Versicherungszeiten
	Pflege	Pflegestufe/Pflegegrad
	Arbeitsunfähigkeit	Arbeitsunfähigkeit: Zeiten
		Arbeitsunfähigkeit: Diagnosen
	Ambulante ärztliche Versorgung	Ambulante Versorgung: Fallinformationen
		Ambulante Versorgung: Diagnosen
		Ambulante Versorgung: OPS
		Ambulante Versorgung: EBM
		Ambulante Versorgung: GOÄ
	Arzneimittel	Arzneimittelversorgung
		Arzneimittelverordnung (individuell hergest.)
	Stationäre Versorgung (Krankenhaus)	Stationäre Versorgung: Aufenthalte
		Stationäre Versorgung: Diagnosen
		Stationäre Versorgung: OPS
		Stationäre Versorgung: DRG
	Heil- und Hilfsmittel	Heil- und Hilfsmittel
		Heil- und Hilfsmittel: Diagnosen

Die Auflistung der in den Tabellen enthaltenen Variablen ist in *Abb. 4* auf der linken Seite zu sehen. Auf der rechten Seite werden zu den einzelnen Variablen Detailinformationen wie enthaltene Werte und dazugehörige Ausprägungen (Bedeutung) angezeigt. Falls zu den hinterlegten Werten externe Informationen zur Interpretation herangezogen werden müssen, sind bei der Beschreibung entsprechende Hinweise dazu vermerkt.

The screenshot shows a software interface for data management. On the left, a tree view lists various data categories and variables, with 'Diagnosesicherheit' highlighted in pink. On the right, a detailed view of the 'Diagnosesicherheit' variable is shown, including its description, values, and metadata.

Anzeigetext	Wert	Bedeutung
Diagnosesicherheit	1	Ausschluss
Diagnosesicherheit	2	gesichert
Diagnosesicherheit	3	Verdacht auf
Diagnosesicherheit	4	Zustand nach
Diagnosesicherheit	-9	keine Angabe

Additional metadata shown includes: Skalentyp: nominal; Skala: kv_ambdiagsure; Eindeutiger Schlüssel: kvad_amb_diag_sure; Einschränkungen: eingeschränkte Nutzbarkeit - Eingeschränkte Beantragbarkeit wegen zeitlich befristeter oder unbefristeter Nutzungsvorbehalte Dritter (z. B. Daten aus Kooperations- oder Level 3-Projekten und Daten, die von Dritten außerhalb der NAKO erhoben wurden); Beantragung möglich? Nur in Ausnahmefällen oder unter Auflagen; Zusätzliche Begründung nötig? ja.

Abb. 4: Beispiel Variablen in Tabelle „Ambulante Versorgung: Diagnosen“

3.2.2 Hinweise zum Sektor „Stammdaten“

Basisdaten zur Demografie sind nicht in den GKV-/PKV-Daten enthalten. Alter und Geschlecht sowie weitere, nicht regulär in Abrechnungsdaten enthaltene Informationen (Primärdaten wie bspw. aktuelle Berufsangaben und Bildung zum Zeitpunkt der Datenerhebung), können zusätzlich im Transferhub ausgewählt werden unter Basisdaten/Demographie/Basisdaten/Allgemein:

- Soziodemografie: basierend auf Selbstangaben der Teilnehmenden
- Alter zum Zeitpunkt der Erstuntersuchung (Hinweis: Variable „Untersuchungsdatum“ mit anfordern)

Weiterhin sollte die Variable „Versicherungstyp“ immer mit ausgewählt werden. In der NAKO sind Routinedaten von sowohl gesetzlich als auch privat krankenversicherten Teilnehmenden verfügbar. Für die Zuordnung eines Teilnehmenden zur gesetzlichen oder privaten Krankenversicherung muss die Variable „Versicherungstyp“ daher immer mit ausgewählt werden. Die Art der Krankenversicherung hat Einfluss auf die gesundheitliche Versorgung und beinhaltet zudem indirekt weitere Informationen zu den Personen, weshalb der Versicherungstyp bei der Analyse und Interpretation der Daten unbedingt berücksichtigt werden muss. Dies betrifft insbesondere den Bereich der ambulanten ärztlichen Versorgung. Den erbrachten Leistungen liegen hier unterschiedliche Abrechnungssysteme zugrunde (GKV: Einheitlicher Bewertungsmaßstab – EBM; PKV: Gebührenordnung der Ärzte – GOÄ; siehe Kapitel 4.1).

3.2.3 Weitere Schlüsselvariablen in den Sektoren

Weitere Schlüsselvariablen, die zusammen mit der ID-A in den GKV-/PKV-Daten eine Eindeutigkeit herstellen, bestehen innerhalb der verschiedenen Sektoren. Sie sind der folgenden Tabelle zu entnehmen.

Tab. 2: Zusätzliche Schlüsselvariablen nach Tabelle in den Krankenversicherungsdaten

Übergeordnete Tabelle	Unterordner/Tabelle	Weitere Schlüsselvariable
Arzneimittelversorgung	Arzneimittelverordnungen (individuell hergest.)	Verordnungs-ID
		Zähler der applikationsfertigen Einheit (Charge)
Heil- und Hilfsmittel	Heil und Hilfsmittel: Diagnosen	Verordnungs-ID
Arbeitsunfähigkeit: Zeiten	Arbeitsunfähigkeit: Diagnosen	Arbeitsunfähigkeits-ID
Ambulante Versorgung: Fallinformationen	Ambulante Versorgung: Diagnosen	Fall-ID
	Ambulante Versorgung: OPS	Fall-ID
	Ambulante Versorgung: EBM	Fall-ID
	Ambulante Versorgung: GOÄ	Fall-ID
Stationäre Versorgung: Aufenthalte	Stationäre Versorgung: Diagnosen	Fall-ID
		Laufende Nummer
	Stationäre Versorgung: OPS	Fall-ID
		Laufende Nummer
	Stationäre Versorgung: DRG	Fall-ID
		Laufende Nummer

Beispiel zur Verwendung von mehreren Schlüsselvariablen:

Werden für Teilnehmende Analysen hinsichtlich von Arbeitsunfähigkeit geplant, ist zu entscheiden, ob lediglich Zeiten der Arbeitsunfähigkeit betrachtet werden sollen. In diesem Fall ist die Auswahl der Variablen aus der Tabelle „Arbeitsunfähigkeit: Zeiten“ ausreichend. Falls zudem die mit der Arbeitsunfähigkeit verbundenen Erkrankungen (in Form von Diagnosen nach ICD-10) betrachtet werden sollen, ist die entsprechende Variable aus der Tabelle „Arbeitsunfähigkeit: Diagnosen“ mit auszuwählen. Damit ein eindeutiger Bezug einer Arbeitsunfähigkeitsdiagnose zu dem dazugehörigen AU-Zeitraum hergestellt werden kann, muss zusätzlich die Variable „**AU-ID**“ innerhalb beider Tabellen selektiert werden (siehe Abb. 5).

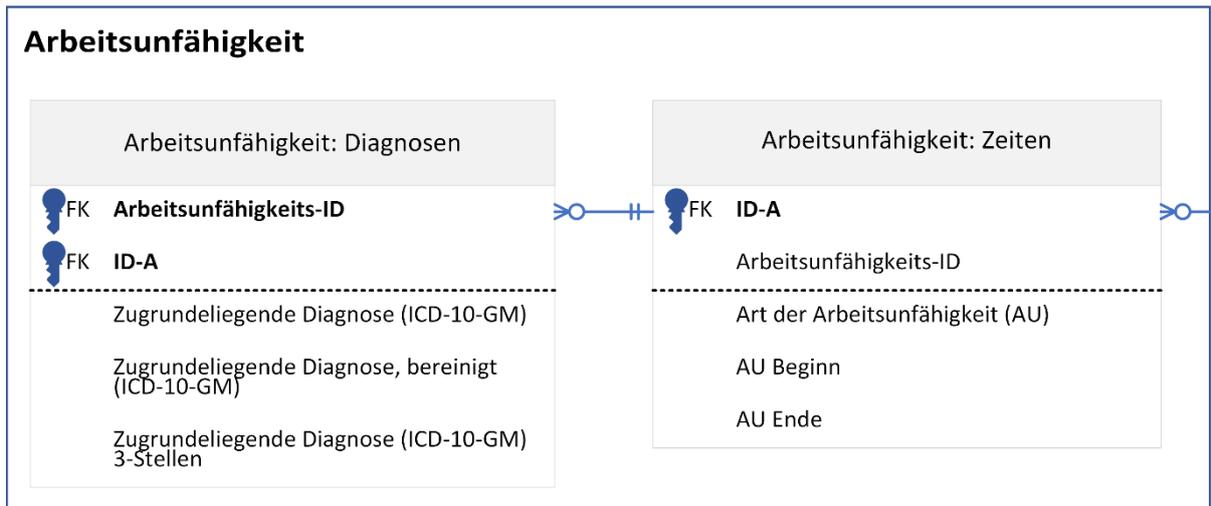


Abb. 5: Darstellung der Abhängigkeit von Schlüsselvariablen am Beispiel Arbeitsunfähigkeit

3.2.4 Variablen- und Wertelabels

Für jede bereitgestellte Tabelle werden die jeweiligen Namen der beinhalteten Variablen in der ersten Zeile der Datentabelle mitgeliefert. Die Labels der Variablenausprägungen (Wertelabels) sind nicht im Datensatz enthalten. Diese können bereits während der Datenselektion über den Transferhub für jede Variable nachvollzogen werden. Zudem wird bei Datenherausgaben jeweils auch ein Auszug des NAKO Data Dictionary mit den für das Vorhaben selektierten Variablen mit zur Verfügung gestellt. In diesem sind die Variablenausprägungen ebenfalls enthalten.

Ausprägungen der in Variablen enthaltenen Werte



Bei der Bereitstellung von Daten enthalten die einzelnen Tabellen keine Wertelabels. Die Ausprägungen müssen, je nach verwendetem Analysetool von den Forscher*innen den einzelnen Werten zugeordnet werden.

Zur Hilfestellung werden auf Anfrage vom Kompetenznetz Syntaxdateien für die Programme SPSS und SAS zur Verfügung gestellt.

3.2.5 Lebenslange Arztnummer, Betriebsstättennummer und Institutionskennzeichen

Die Lebenslange Arztnummer (LANR), die Betriebsstättennummer (BSNR) und das Institutionskennzeichen (IK) werden ausschließlich pseudonymisiert zur Verfügung gestellt (in Vorbereitung). Davon betroffene Sektoren sind die ambulante Versorgung, stationäre Versorgung, Arzneimittelversorgung sowie Heil- und Hilfsmittel.

Der aus der LANR des/der behandelnden Arztes/Ärztin hervorgehende Fachgruppenschlüssel ist als eigenständige Variable „Facharztqualifikation“ verfügbar und kann bei der Antragsstellung selektiert werden.

Die Region, abgeleitet aus den IK der Leistungserbringer (Krankenhäuser, Apotheken) ist ebenfalls als eigenständige Variable verfügbar.

Die Region der Kassenärztlichen Vereinigung des/der behandelnden Arztes/Ärztin, abgeleitet aus den BSNR der Leistungserbringer ist ebenfalls als eigenständige Variable im Bereich der Arzneimittelversorgung verfügbar.

4. Datensatzbeschreibung und Datenbereitstellung

4.1 Datensatzbeschreibung

Die Datensatzbeschreibung (data dictionary) aller in der NAKO verfügbaren Krankenversicherungsdaten ist für Forschende im Transferhub in deutscher und englischer Sprache hinterlegt. Bei der Bereitstellung der Daten für Projekte durch die Transferstelle wird zudem ein projektbezogenes data dictionary mit bereitgestellt.

Das data dictionary der verfügbaren Primärdaten wird aktuell als Excel-Tabelle (*Excel-Auszug aus Datenverzeichnis*) über die Startseite des Transferhubs zur Verfügung gestellt. Hierfür ist keine Anmeldung im Transferhub erforderlich. Projekte, die Krankenversicherungsdaten mit beantrag haben, erhalten bei der Datenübergabe ein erweitertes data dictionary mit den Inhalten der ausgewählten Daten inkl. der Krankenversicherungsdaten (siehe Beispiel in *Abb. 6*).

In den Datensätzen und ihren Beschreibungen sind die Namen gesetzlicher Krankenkassen und von privaten Krankenversicherungen nicht enthalten. Eine Unterscheidung gesetzlich und privat krankenversicherter Personen ist durch die Variable *kvper_shi_phi_agg* möglich.

Reiter „Variablen“										
Untersuchung	ID	Name	Abschnitt	Frage	Label	Einheit	Belegungsliste	#Belegungen	DB-Datentyp	#Plausi
Verlaufdaten/Versicherungszeiten	12802	kvper_shi_phi_agg			Versicherungstyp aggregiert		kv_instypagg		2 TINYINT(1)	0

Reiter „Codierung“									
	Belegungsliste	Wert			Label				
; value	kv_instypagg	1 =			#NAME?	"	PKV	"	
		2 =			#NAME?	"	GKV	"	

Abb. 6: Beispiel Aufbau NAKO data dictionary. Versicherungstyp aggregiert.

Die Variablenamen orientieren sich an den Datenbankkonventionen der NAKO:

- Alle Tabellen und Variablen der Krankenversicherungsdaten haben das Präfix KV_
- Bereits abgeleiteten Variablen ist an zweiter Stelle zusätzlich ein *_a_* vorangestellt
- Beispiel (siehe *Abb. 7*):
 - Tabelle: Ambulante Diagnosen **KV_AMB_DIAG**
 - Standardvariable: Ambulante Diagnose (ICD-10) *kvad_amb_diag*
 - Abgeleitete Variable: Diagnose 3-Steller (ICD-10) *kvad_a_amb_diag_3*

Untersuchung	ID	Name	Abschnitt	Frage	Label	Einheit	Belegungsliste	#Belegungen	DB-Datentyp	#Plausi
Ambulante Versorgung: Diagnosen	12845	kvad_amb_diag			Diagnose (ICD-10_GM), Großbuchstaben ohne Leerzeichen				VARCHAR(12)	0
Ambulante Versorgung: Diagnosen	12847	kvad_a_amb_diag_3			Diagnose (ICD-10-GM), 3 Stellen				VARCHAR(3)	0

Abb. 7: Variablenamen für das Beispiel ambulante Diagnosen

Die Auswertung von Abrechnungsziffern bzw. in Anspruch genommener (ambulanter) Leistungen ist sowohl für GKV- als auch für PKV-Versicherte möglich. Hierbei sind die unterschiedlichen Vergütungssysteme und Klassifikationen zu berücksichtigen. Entsprechend sind die benötigten Variablen aus den Tabellen der ambulanten Versorgung OPS, EBM und GOÄ zu selektieren. Für die stationäre Versorgung sind die erbrachten Leistungen in der DRG- und OPS-Tabelle unabhängig von der Versicherungszugehörigkeit enthalten.

Weitere Informationen und Inhalte der Leistungsziffern finden Sie unter folgenden Links:

- EBM (GKV): <http://www.kbv.de/html/ebm.php> [1]
- OPS: https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/_node.html [2]
- GOÄ (PKV): <https://abrechnungsstelle.com/goae/ziffern/> [3]
- DRG: <https://www.g-drg.de/> [4]

Bei der Zuspielung ergänzender Informationen, wie bspw. der den EBM- oder GOÄ-Ziffern hinterlegten Leistungen, ist die Dynamik der Codierungen zu beachten. Im Zeitverlauf kann es zu einem Wegfall oder der Neuvergabe von Codierungen und damit zu einem Bedeutungswechsel kommen. Entsprechend müssen bei der Analyse die jeweiligen Datumsvariablen (bspw. Abgabe-, Entlassungs-, Abrechnungs-, Leistungsdatum) berücksichtigt werden.

Arzneimitteldaten: ATC, Medikamentenname etc. müssen durch die Datennutzer*innen selbst ergänzt werden – bspw. über den WIdO Arzneimittelindex [5] oder die ABDA-Datenbank [6]. Die zeitliche Abhängigkeit und damit Dynamik von hinterlegten Werten/Informationen zu den Pharmazentralnummern (PZN) muss beachtet werden.

Weitere Informationen und Anträge:

- GKV-Arzneimittelindex [5]
- ABDA-Datenbank [6]
- Referenzdatenbank für Fertigarzneimittel am BfArM [7]

4.2 Datenbereitstellung

Die beantragten Daten werden in Tabellen als .csv-Dateien über den Transferhub bereitgestellt. Dabei wird den Datennutzer*innen einmalig der zum Zeitpunkt der Datenübermittlung aktuelle Datenstand übermittelt. Es erfolgt keine automatische Ergänzung oder Korrektur der Daten bei Vorhandensein neuer Zeitabschnitte oder bei Nachkorrekturen. Für dahin gehende Änderungen müsste ein erneuter Antrag gestellt werden.

Es werden ausschließlich die im Transferhub verfügbaren relationalen Daten aus den jeweiligen Tabellen/Sektoren zur Verfügung gestellt. Die Ableitung von projektspezifisch benötigten Indikatoren und weitere Bearbeitungsschritte sind von den Datennutzer*innen selbst durchzuführen.

Falls eine erneute Datenbereitstellung mit einem erweiterten Zeitabschnitt beantragt wird, ist zu beachten, dass durch Neulieferungen die Daten fortgeschrieben werden. Das bedeutet, dass

- eine Übermittlung neuer Zeitabschnitte erfolgt und
- Nachkorrekturen bei bereits bestehenden Daten möglich sind.

Hinsichtlich der Variablen, welche für die Bearbeitung einer Forschungsfrage erforderlich sind, sollte vor der Einreichung eines Datennutzungsantrags Klarheit bestehen. Eine Nachselektion von Variablen ist möglich, allerdings bedarf dies der erneuten Genehmigung des Vorstandes des NAKO e.V. und ist entsprechend zeitaufwendig.

Der bereits gelieferte Datensatz und ein möglicher aktualisierter Datensatz sind daher ggf. inhaltlich auch in Bezug auf die Daten der ersten Lieferung unterschiedlich und nicht nur aufgrund der zusätzlichen Zeitabschnitte. Das Teilnehmendenpseudonym ID-A ändert sich bei einer Datenaktualisierung nicht.

4.3 Record Linkage

Datennutzungsanträge in der NAKO sind nicht auf eine Datenart (Primär- oder Sekundärdaten) beschränkt. Alle im Transferhub verfügbaren Datensektoren können beantragt werden, sofern sie für die Beantwortung einer Fragestellung erforderlich sind.

Über die projektspezifischen Teilnehmendenpseudonyme (ID-A) ist ein exaktes Linkage zwischen Primär- und Sekundärdaten möglich.

Für die Krankenversicherungsdaten müssen weitere Schlüsselvariablen zur Herstellung der Eindeutigkeit berücksichtigt werden (siehe Kapitel 3.2.3)

4.4 Weiterführende Auswertungshinweise

4.4.1 Format von Codes

Bei der Analyse von unbereinigten ICD-Codes muss beachtet werden, dass diese teilweise mit trennenden Zeichen und Sonderzeichen aus der Kreuz-Stern-Klassifikation [8] übermittelt werden. Daher müssen die entsprechenden Variablen vor der Analyse und ggf. dem Zuspielen ergänzender Informationen vereinheitlicht werden.

Bei EBM-Codes ist zu beachten, dass hier mitunter an 6. Stelle ein Buchstabe codiert sein kann (bspw. zum Zweck der Kennzeichnung der Seitenlokalisierung oder der Impfreihefolge). Dies ist bei der Analyse und ggf. dem Zuspielen ergänzender Informationen zu beachten.

4.4.2 Verfügbarkeit bestimmter Variablen und Tabellen

1. Sektion Verlaufsdaten: Die Kennzeichnung besonderer Personenkreise wie Bezieher von Arbeitslosengeld wird nicht von allen GKVn (in Zukunft verfügbar) bzw. PKVn übermittelt.
2. Sektion ambulante ärztliche Versorgung – Behandlungsfälle: Die Information, ob es sich bei der Behandlung um eine ambulante Behandlung/Operation, eine stationäre Leistung oder eine Reha-/Anschlussbehandlung handelt, wird nicht von allen GKVn/PKVn übermittelt.
3. Sektion ambulante ärztliche Versorgung – Operationen und Prozeduren: Für ambulante Operationen und Prozeduren (OPS) liegt die Variable des OPS-Datums für die Daten der GKV-Versicherten nicht vor.
4. Sektion ambulante ärztliche Versorgung – Behandlungsfälle sowie Medikamentenverordnungen: Leistungserbringerbezogene Informationen (z. B. BSNR, LANR) werden nicht von allen PKVn übermittelt.
5. Einzelbestandteile von individuell hergestellten Arzneimitteln:
 - a. Diese Daten werden nicht von allen GKVn/PKVn übermittelt.
 - b. Beispiele für Themen, bei denen die Daten relevant sind:
 - i. Zytostatikazubereitungen
 - ii. Parenterale Antibiotikatherapie
 - iii. Parenterale Schmerzmitteltherapie
6. Heil- und Hilfsmitteldaten werden nicht von allen GKVn/PKVn übermittelt.
7. Arbeitsunfähigkeitsdaten werden nicht von allen GKVn/PKVn übermittelt.
8. Pflegedaten werden nicht von allen GKVn/PKVn übermittelt.
9. Ambulante Behandlungen im Krankenhaus (bspw. durch Belegärzte) werden nicht von allen GKVn/PKVn übermittelt.

Neben von GKVn/PKVn vollständig nicht übermittelten Variablen und Tabellen können auch innerhalb von Datenlieferungen bestimmte Ausprägungen nicht übermittelt werden. Dadurch hervorgerufene Implikationen in Hinblick auf das Forschungsvorhaben können ggf. im Rahmen der Konsultation der Expertengruppe Sekundär- und Registerdaten/Mortalitäts-Follow-up (siehe Kapitel 2) vertieft werden.

4.4.3 Sekundär Diagnosen im Krankenhaus ohne Zuordnung zur Primär diagnose

Zu bestimmten ICD-10 codierten Primär diagnosen kann eine ergänzende Sekundär diagnose codiert werden, über das sogenannte Kreuz-Stern-Klassifikationssystem. Ein Beispiel wäre die Codierung einer Polyneuropathie (ICD-10-GM G63.2*) bei Primär diagnose Diabetes mellitus Typ 2 mit neurologischen Komplikationen (ICD-10-GM E11.4[†]).

Nur wenn Primär- und Sekundär diagnose gemeinsam in einer Tabellenzeile übermittelt werden, besteht eine korrekte Zuordnung und Darstellung des Krankheitsgeschehens.

Von einigen Kassen werden die Primär- und Sekundär diagnose in unterschiedlichen Zeilen und dafür mit Kennzeichnung für Primär diagnose bzw. Sekundär diagnose übermittelt. In diesen Fällen ist es unklar, zu welcher Primär diagnose die jeweilige Sekundär diagnose gehört. Diese Sekundär diagnosen sind dann mit Diagnosetyp = 25 gekennzeichnet, wobei die Diagnose in der Spalte der Primär diagnose nicht zwingend die tatsächliche Primär diagnose darstellt.

Die nachfolgende Tabelle enthält für Person 1 eine korrekte Übermittlung der Sekundär diagnose und für Person 2 eine, bei der die Sekundär diagnose nicht zugeordnet werden konnte.

Tab. 3: Beispiel Zuordnung von Sekundär diagnosen in der stationären Versorgung

ID-A	kvhd_hosp_case_id	kvhd_diag_prime	kvhd_diag_sec	kvhd_hosp_diag_type
1	2222222	E11.4 [†]	G63.2*	23
1	2222222	F32.0		20
1	2222222	J44.01		20
2	1234567	E12.4		20
2	1234567	F32.0	G63.2	25
2	1234567	J44.01		20

5. Schulungen und Standard dokumente

5.1 Schulungen

Für die Analyse von Sekundär daten gibt es in Deutschland Einführungs- und Vertiefungskurse. Diese werden u. a. angeboten von der AGENS – Arbeitsgruppe Erhebung und Nutzung von Sekundär daten und dem Deutschen Netzwerk für Versorgungsforschung e. V. (DNVF).

Die AGENS bietet mehrmals jährlich Schools mit verschiedener Ausrichtung im Rahmen der Routine datenanalyse an. Die aktuellen Termine finden sie auf der Homepage der AGENS: <https://agens.group>
Das DNVF bietet jährlich eine Spring School an. Aktuelle Termine finden sie auf der Webseite des Netzwerks: <https://www.dnvf.de>

5.2 Standard literatur

Bei der Arbeit mit Routedaten sind bestimmte Standards zu beachten. Die nachfolgende Tabelle liefert einen Überblick über die wichtigsten Standards, erhebt aber keinen Anspruch auf Vollständigkeit. Weitere fachspezifische Standards können ebenfalls relevant sein (z. B. für Untersuchungen zur Anwendung von Arzneimitteln).

- Routedaten im Gesundheitswesen: Handbuch Sekundär datenanalyse: Grundlagen, Methoden und Perspektiven [9]
- Prozessorientierter Leitfaden für die Analyse und Nutzung von Routedaten der Gesetzlichen Krankenversicherung [10]

- Gute Praxis Sekundärdatenanalyse (GPS) [11]
- Gute Epidemiologische Praxis (GEP) [12, 13]
- Gute Praxis Datenlinkage (GPD) [14, 15]
- STROSA - STandardized Reporting Of Secondary data Analyses [16]
- STROBE - Strengthening the Reporting of Observational Studies in Epidemiology [17]
- Vorgaben und Technische Anlagen des elektronischen Datenaustausches der GKVn [18]

6. Autorenliste

Christoph Stallmann¹, Berit Brandes², Ludwig Goldhahn¹, Thore Wulf², Achim Reineke², Enno Swart¹, Wolfgang Ahrens², Iris Pigeot², Wiebke Schäfer²

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

² Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen

7. Literaturverzeichnis

- [1] Kassenärztliche Bundesvereinigung (KBV). KBV - EBM: Kassenärztliche Bundesvereinigung (KBV) (16.03.2025). Im Internet: <https://www.kbv.de/html/ebm.php>; Stand: 16.03.2025
- [2] Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). OPS: Operationen- und Prozedurenschlüssel (OPS) (16.03.2025). Im Internet: https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/_node.html; Stand: 16.03.2025
- [3] abrechnungsstelle.com. Online GOÄ-Ziffern 2025 (Gebührenordnung für Ärzte) (16.02.2025). Im Internet: <https://abrechnungsstelle.com/goae/ziffern/>; Stand: 16.03.2025
- [4] InEK GmbH. Website des deutschen DRG-Systems (16.03.2025). Im Internet: <https://www.g-drg.de/>; Stand: 16.03.2025
- [5] Wissenschaftliches Institut der AOK (WIdO). GKV-Arzneimittelindex (16.03.2025). Im Internet: <https://www.wido.de/forschung-projekte/arszneimittel/gkv-arszneimittelindex/>; Stand: 16.03.2025
- [6] Avoxa – Mediengruppe Deutscher Apotheker GmbH. ABDADatenbank² - ABDA (05.09.2024). Im Internet: <https://abdata.de/produkte/abdadatenbank2/>; Stand: 16.03.2025
- [7] Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). Referenzdatenbank für Fertigarzneimittel gemäß § 31b SGB V (16.03.2025). Im Internet: https://www.bfarm.de/DE/Arzneimittel/Arzneimittelinformationen/Referenzdatenbank/_node.html; Stand: 16.03.2025
- [8] Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). Kategorie und Kode in der ICD-10-GM (16.03.2025). Im Internet: <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/Systematik/kodestruktur.html>; Stand: 16.03.2025
- [9] Swart E, Ihle P, Gothe H, Matusiewicz D, Hrsg. Routinedaten im Gesundheitswesen: Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. 2. Aufl. Bern: Huber; 2014
- [10] Neubauer S, Zeidler J, Lange A, Schulenburg J-M von der. Prozessorientierter Leitfaden für die Analyse und Nutzung von Routinedaten der Gesetzlichen Krankenversicherung. Baden-Baden: Nomos; 2017
- [11] Swart E, Gothe H, Geyer S et al. Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen. Gesundheitswesen 2015; 77: 120–126; DOI: 10.1055/s-0034-1396815
- [12] Hoffmann W, Latza U, Baumeister SE et al. Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology. Eur J Epidemiol 2019; 34: 301–317; DOI: 10.1007/s10654-019-00500-x
- [13] Hoffmann W, Latza U, Terschüren C. Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP) -- überarbeitete Fassung nach Evaluation. Gesundheitswesen 2005; 67: 217–225; DOI: 10.1055/s-2004-813850
- [14] March S, Andrich S, Drepper J et al. Good Practice Data Linkage (GPD): A Translation of the German Version. Int J Environ Res Public Health 2020; 17; DOI: 10.3390/ijerph17217852
- [15] March S, Andrich S, Drepper J et al. Gute Praxis Datenlinkage (GPD). Gesundheitswesen 2019; 81: 636–650; DOI: 10.1055/a-0962-9933
- [16] Swart E, Schmitt J. STandardized Reporting Of Secondary data Analyses (STROSA)—Vorschlag für ein Berichtsformat für Sekundärdatenanalysen. Z Evid Fortbild Qual Gesundhwes 2014; 108: 511–516; DOI: 10.1016/j.zefq.2014.08.022
- [17] Vandembroucke JP, Elm E von, Altman DG et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Ann Intern Med 2007; 147: W163-94; DOI: 10.7326/0003-4819-147-8-200710160-00010-w1
- [18] GKV-Spitzenverband. GKV-Datenaustausch (16.03.2025). Im Internet: <https://www.gkv-datenaustausch.de/>; Stand: 16.03.2025

Anhang 1: Datenbankmodell der Krankenversicherungsdaten

