

Record Linkage-Use Cases

Einführung und Übersicht

Timm Intemann



Workshop zum White Paper
"Verbesserung des Record
Linkage in Deutschland"

24. Mai 2022, Berlin



Leibniz-Institut
für Präventionsforschung und
Epidemiologie – BIPS



Use Cases in der Gesundheitsforschung

- **Verschiede Anwendungsfälle** der Verknüpfung von personenbezogenen Gesundheitsdaten auf Individualebene
- **Überblick** über praktische Umsetzung von Record Linkage-Projekten
- Einblick in Record Linkage-Projekte aus unterschiedlichen Bereichen der Gesundheitsforschung – kein Anspruch auf Vollständigkeit
- Use Cases unterscheiden sich bei **Datenquellen** und **durchführenden Einrichtungen**

- **Kurzvorträge** von ~6 Minuten plus ~5 Minuten Fragen und Diskussion
 - Hintergrund und Ziele
 - Datenquellen und Datenfluss
 - Identifier und Record Linkage-Methode
 - Herausforderungen

Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt (Timm Intemann)</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum

Weitere Use Cases des White Papers

- Linkage am Beispiel der molekularen Sars-CoV-2 Surveillance (DESH)
- Use Case Treuhandstelle: Die Unabhängige Treuhandstelle der Universitätsmedizin Greifswald
- Künftiger Use Case A: Standortübergreifendes Record Linkage in der Medizininformatik-Initiative
- Künftiger Use Case B: Standortübergreifendes Record Linkage im Netzwerk Universitätsmedizin (NUM-CODEX)

Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum

Use Case: Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf

Timm Intemann, Christoph Stallmann, Stefan
Rach, Iris Pigeot und Wolfgang Ahrens

Workshop zum White Paper
"Verbesserung des Record
Linkage in Deutschland"

24. Mai 2022, Berlin



Leibniz-Institut
für Präventionsforschung und
Epidemiologie – BIPS



Hintergrund und Ziele



Studie zum Erkrankungsverlauf bei Personen mit einer COVID-19-Erkrankung bzw. einem positiven Test auf SARS-CoV-2

- Determinanten schwerer Erkrankungsverläufe
- Demonstration: Krankenkassendaten zur Untersuchung von Erkrankungsverläufen nützlich

<https://www.bips-institut.de/covid-19.html#9176>

Datenquelle: Krankenkassendaten

- Krankenkassen sammeln individuelle Daten zu Abrechnungszwecken
 - Demografische Daten
 - Arzneimittelverordnungen, ambulante/stationäre Leistungen, Diagnosen
- Vorteile gegenüber Befragungsdaten, da z.B. keine Erinnerungs- oder Wissenslücken

➤ Datenschatz

- **Allerdings:** Datenschatz verteilt sich aktuell auf 97 gesetzliche und 46 private Krankenversicherungen (GKV/PKV)

Pigeot & Ahrens (2008) Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. *Pharmacoepidemiology and Drug Safety* 17, 215-223

Stallmann et al (2015) Individuelle Datenverknüpfung von Primärdaten mit Sekundär- und Registerdaten in Kohortenstudien: Potenziale und Verfahrensvorschläge. *Gesundheitswesen* 77, e37-e42.

Ahrens et al (2020) Erforschung von Erkrankungen in der NAKO Gesundheitsstudie. Die wichtigsten gesundheitlichen Endpunkte und ihre Erfassung *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 63, 376-384

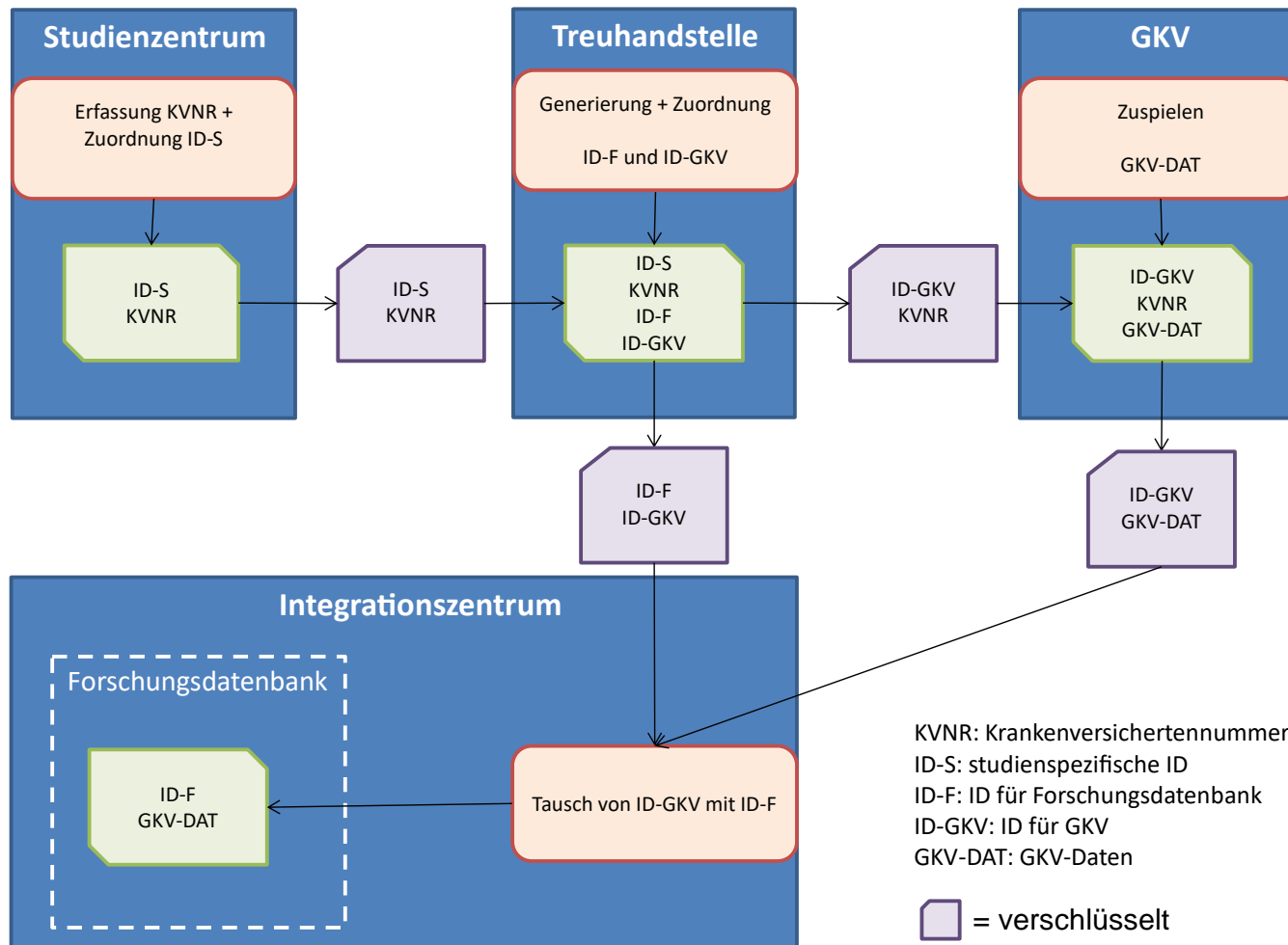
https://gkv-spitzenverband.de/krankenversicherung/kv_grundprinzipien/alle_gesetzlichen_krankenkassen/alle_gesetzlichen_krankenkassen.jsp, Zugriff: 20.05.2022

<https://portal.mvp.bafin.de/database/InstInfo/>, Zugriff: 20.05.2022, Suchkategorie: Krankenversicherer (VA)

Datenquelle: epidemiologische Primärdaten

- Kontaktaufnahme über Gesundheitsamt Bremen
- 1908 Personen in Studie eingewilligt
- Befragungsdaten u.a. zu
 - Soziodemografie
 - Vorerkrankungen und Medikamenten
 - Lebensstil
- 1617 Personen in Verknüpfung mit Krankenkassendaten eingewilligt und Krankenversicherungsnummer (KVNR) übermittelt

Möglicher Datenfluss analog zur NAKO-Gesundheitsstudie



Identifizierung und Record Linkage-Methode

- Zusätzliche Erhebung der KVNR stellt bei Primärdatenerhebungen keine Hürde dar
- Mit KVNR steht **direkter Identifikator** zur Verfügung, der relativ einfach für exaktes Record Linkage verwendet werden kann

Herausforderungen

- Einholung der Einwilligungserklärungen
- Einwilligungsbereitschaft in der Bevölkerung sehr hoch
 - CoVerlauf: **85%**
 - NAKO-Gesundheitsstudie: **94%**
- Organisatorischer Aufwand für die Kooperation mit **101** Krankenkassen (lt. Selbstangabe)
 - Einschluss: Kassen mit mindestens 20 Probanden und Probandinnen:
 - **Reduktion des Stichprobenfangs auf 1352 (71%) und 14 Kassen**
- Zweigliedriges Krankenversicherungssystem in Deutschland
 - GKV- und PKV-Daten fallen unter verschiedene Datenschutzregeln
 - Ausschluss der PKV
 - **Reduktion des Stichprobenfangs auf 1325 (69%)**

Schipf et al (2020) Die Basiserhebung der NAKO Gesundheitsstudie: Teilnahme an den Untersuchungsmodulen, Qualitätssicherung und Nutzung von Sekundärdaten. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 63, 254-266

Stallmann (2018) Die Nutzung von Sekundärdaten in epidemiologischen Kohortenstudien: ein Zugewinn für die Gesundheitsforschung? Otto-von-Guericke-Universität Magdeburg

Vielen Dank für die
Aufmerksamkeit!



Kontakt:

Timm Intemann

intemann@leibniz-bips.de

**Leibniz-Institut für
Präventionsforschung und
Epidemiologie - BIPS**

www.nfdi4health.de



Leibniz-Institut
für Präventionsforschung und
Epidemiologie – BIPS

Gefördert durch



Deutsche
Forschungsgemeinschaft



Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum

Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum

Use Case: Linkage von Krankenkassen- und Krebsregisterdaten am Beispiel des Projekts DFG-Linkage

Timm Intemann, Bianca Kollhorst, Klaus
Kraywinkel, Wolfgang Ahrens und Iris Pigeot

Workshop zum White Paper
"Verbesserung des Record
Linkage in Deutschland"

24. Mai 2022, Berlin



Leibniz-Institut
für Präventionsforschung und
Epidemiologie – BIPS



Hintergrund und Ziele

- DFG-Linkage: „*Evaluierung eines indirekten Linkage-Ansatzes anhand einer Beispielstudie zum Risiko einer Krebsneuerkrankung und der Krebsmortalität bei Patienten und Patientinnen mit Typ-2-Diabetes unter Behandlung mit verschiedenen Antidiabetika*“
- Für zwei Krebsentitäten: Schilddrüsen- und Darmkrebs
- **Ziel:** Evaluierung eines Linkage-Ansatzes basierend auf **indirekten** Identifikatoren im Vergleich zu einem Ansatz mit **direkten** Identifikatoren

Pigeot et al (2021) Verknüpfung von Abrechnungsdaten gesetzlicher Krankenkassen mit Daten epidemiologischer Krebsregister: Möglichkeiten und Limitationen.

Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 65, 615-623

Pigeot et al (2021) Nutzung von Sekundärdaten für die pharmakoepidemiologische Forschung - machen wir das Beste draus! Gesundheitswesen 83, S69-S76

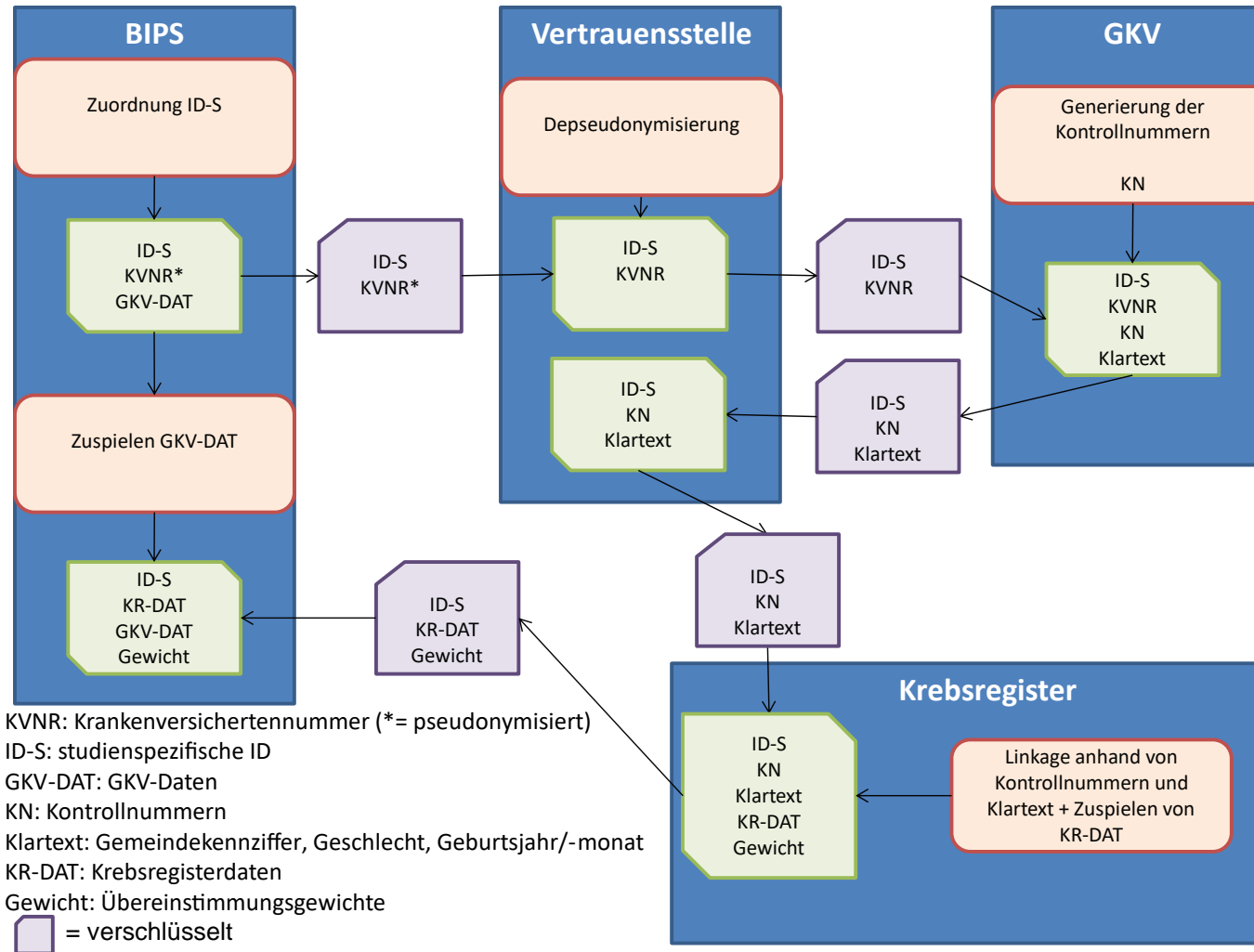
Datenquellen

- Vier epidemiologische Krebsregister
 - Tumorstadium
 - Valide Tumordiagnose
- Pharmakoepidemiologische Forschungsdatenbank (GePaRD):
Abrechnungsdaten von ~25 Mio. Versicherten (TK, DAK, hkk, AOK Bremen)
 - Diabetesmedikation



Pigeot & Ahrens (2008) Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. *Pharmacoepidemiology and Drug Safety* 17, 215-223

Datenfluss



Pigeot et al (2021) Nutzung von Sekundärdaten für die pharmakoepidemiologische Forschung - machen wir das Beste draus! Gesundheitswesen 83, S69-S76

Identifizierung und Record Linkage-Methode

Probabilistisches Linkage basierend auf *direkten* Identifikatoren

- Etabliertes Kontrollnummernlinkage
- Kontrollnummern abgeleitet aus u.a. *Vorname, Nachname, Tagesangabe des Geburtsdatums*

Deterministisches Linkage basierend auf *indirekten* Identifikatoren

- Datenschutzrechtlich weniger aufwändig, da keine Vertrauensstelle notwendig
- Identifizierung: *Geburtsjahr, Geschlecht, Gemeindekennziffer, Krebsart und Diagnosedatum*
- Falls keine eindeutige Zuordnung: *Genauigkeit der ICD-Kodierung, stationäre Diagnose, Differenz des Diagnosedatums*
- Sensitivität des Linkage: 72% (Darmkrebs) und 67% (Schilddrüsenkrebs)
- Analyseergebnis: Risiko für Darmkrebs in Abhängigkeit von Diabetesmedikation um 10 Prozentpunkte unterschätzt

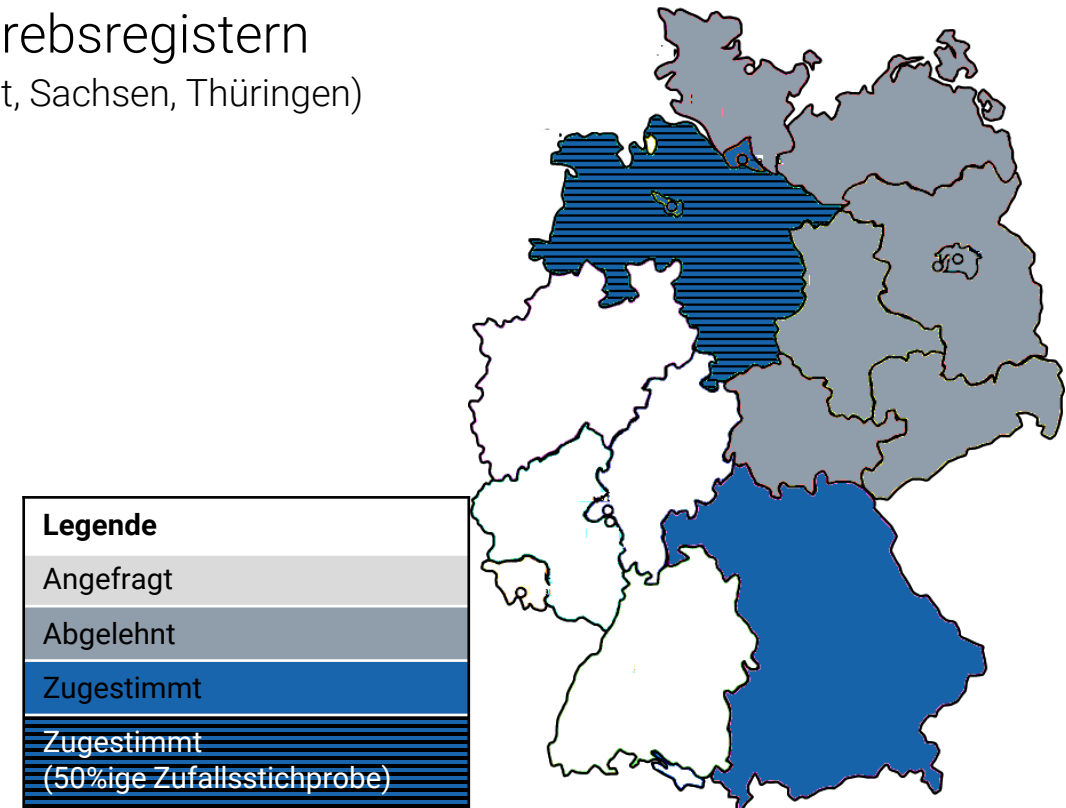
Hentschel & Katalinic (2008) Das Manual der epidemiologischen Krebsregistrierung, Zuckschwerdt, München

Pigeot et al (2021) Verknüpfung von Abrechnungsdaten gesetzlicher Krankenkassen mit Daten epidemiologischer Krebsregister: Möglichkeiten und Limitationen.

Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 65, 615-623

Herausforderungen

- **Problem:** keine Verlinkung über Zentrum für Krebsregisterdaten (ZfKD) aus rechtlichen Gründen möglich
- **Konsequenz:** Anfrage bei sechs epidemiologischen Krebsregistern
 - GKR (Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen-Anhalt, Sachsen, Thüringen)
 - Bayern
 - Bremen
 - Hamburg
 - Niedersachsen
 - Schleswig-Holstein
- **Unterschiede:**
 - Art der Antragstellung
 - Einzureichende Unterlagen
 - Zuständige Stelle
 - Geäußerte datenschutzrelevante Bedenken
 - Erfolgte Entscheidung



Pigeot et al (2021) Verknüpfung von Abrechnungsdaten gesetzlicher Krankenkassen mit Daten epidemiologischer Krebsregister: Möglichkeiten und Limitationen. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 65, 615-623

Herausforderungen

- Kontrollnummernlinkage für Krebsregistrierung entwickelt, dafür nötige Merkmale nicht vollständig bzw. nicht in gleicher Form in Versicherungsdaten enthalten
 - Zukünftig: für GKV-Versicherte erleichtert, wenn KVNR genutzt werden dürfte
- Studien dieser Art zeit- und kostenintensiv: weitere Anträge und Genehmigungen
 - Zustimmung der Kassen
 - §75, SGB X
- **Flächendeckende bundesweite Analysen dieser Art erscheinen momentan unrealistisch**
- **Gesetz zur Zusammenführung von Krebsregisterdaten:** mögliche Verbesserung
 - ZfKD und andere aufgefordert „*Konzept zur Schaffung einer Plattform, die eine bundesweite anlassbezogene Datenzusammenführung und Analyse der Krebsregisterdaten aus den Ländern sowie eine **Verknüpfung von Krebsregisterdaten mit anderen Daten ermöglicht***“

Pigeot et al (2021a) Verknüpfung von Abrechnungsdaten gesetzlicher Krankenkassen mit Daten epidemiologischer Krebsregister: Möglichkeiten und Limitationen. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 65, 615-623

Gurung-Schönfeld & Kraywinkel (2021) Krebsregistrierung heute: zwischen Epidemiologie, Qualitätssicherung und Forschung. Epidemiologisches Bulletin 4, 3-9

Vielen Dank für die
Aufmerksamkeit!



Kontakt:

Timm Intemann

intemann@leibniz-bips.de

**Leibniz-Institut für
Präventionsforschung und
Epidemiologie - BIPS**

www.nfdi4health.de



Leibniz-Institut
für Präventionsforschung und
Epidemiologie – BIPS

Gefördert durch



Deutsche
Forschungsgemeinschaft



Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum

Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum

Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum

Use Case	Vortragende	Organisation
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie CoVerlauf	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage von Primärdaten und Krankenkassendaten am Beispiel der Studie SHIP	<i>Carsten Oliver Schmidt</i>	Institut für Community Medicine, Universitätsmedizin Greifswald
Linkage von Krankenkassen- und Krebsregisterdaten (Projekt DFG-Linkage)	<i>Timm Intemann</i>	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS
Linkage der Landeskrebsregister am Beispiel von Nordrhein-Westfalen und Niedersachsen	<i>Sebastian Bartholomäus</i>	Landeskrebsregister NRW
Linkage von Daten aus Routinedaten und -proben und klinischen Erhebungen innerhalb eines rechtlichen Trägers	<i>Ulrich Sax</i>	Institut für Medizinische Informatik, Universitätsmedizin Göttingen
Standortübergreifendes Record Linkage im Deutschen Konsortium für Translationale Krebsforschung (DKTK)	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum
Linkage von klinischen Routine- und Studiendaten über mehrere rechtliche Träger hinweg am Beispiel des nationalen Netzwerks Genomische Medizin Lungenkrebs	<i>Jori Kern</i>	Deutsches Krebsforschungszentrum