Nutzung von Sekundärdaten für die pharmakoepidemiologische Forschung

Lassen sich Limitationen der Daten durch moderne statistische Verfahren oder Datenlinkage ausgleichen?

Iris Pigeot, Vanessa Didelez, Dirk Enders, Bianca Kollhorst

Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS

Herbstworkshop 2016, Berlin

Inhalt

- Pharmakoepidemiologische Forschung
- Die Forschungsdatenbank GePaRD
- Statistische Ansätze zur Berücksichtigung von Confounding
- Bisherige Linkage-Erfahrungen
- Rechtliche Grundlage

Pharmakoepidemiologische Forschung

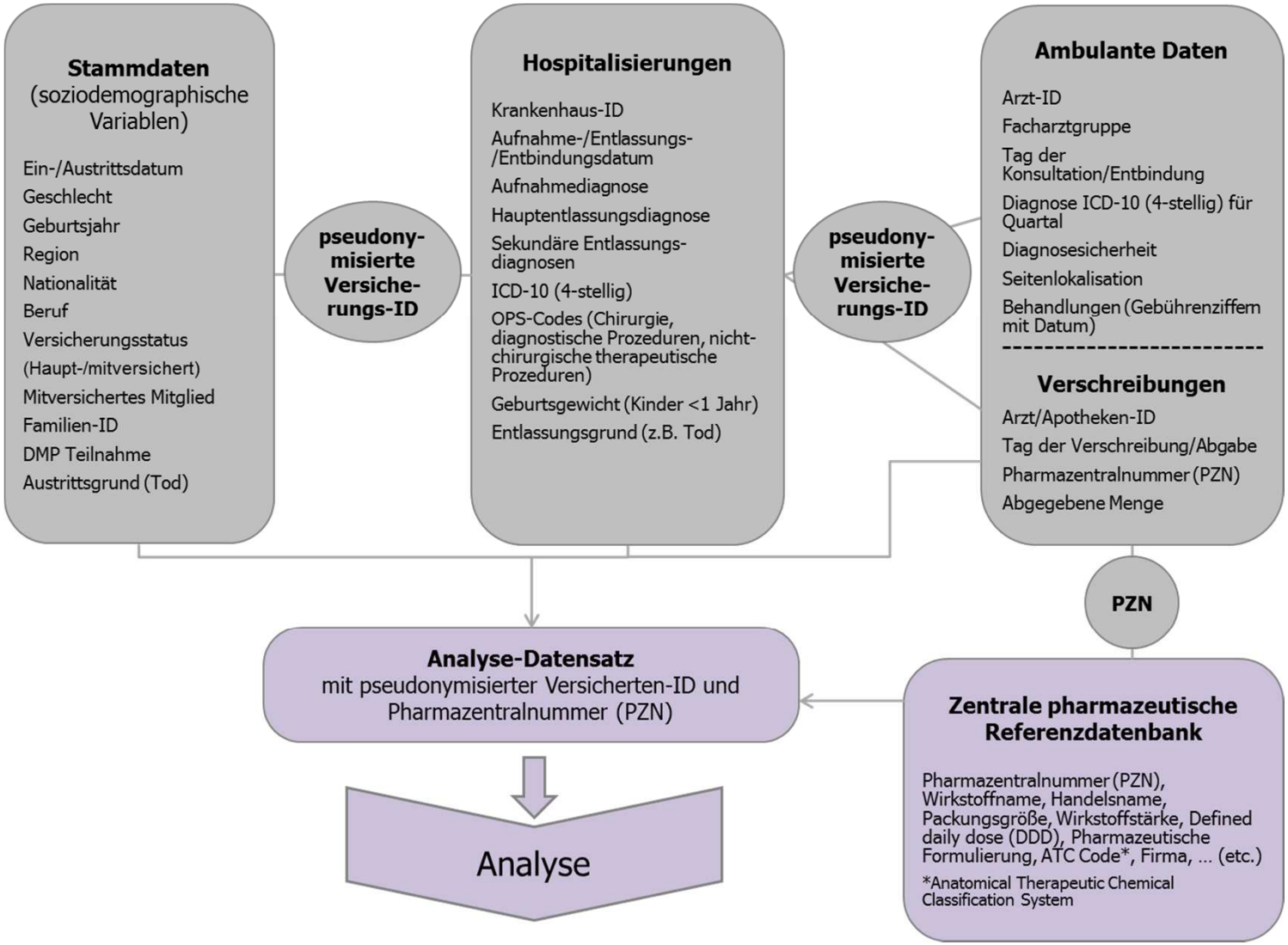
- Untersucht Anwendung, Wirkungen und Risiken von Arzneimitteln nach Marktzulassung
- **Limitationen der klinischen Studien vor Zulassung**
  - Zu geringe Größe (i.d.R. 3000-5000 Patienten)
  - Zu geringe Studiendauer
  - Vulnerable Patientengruppen häufig ausgeschlossen
- **Risiken von nicht bestimmungsgemäßem Gebrauch nach Marktzulassung**
  - Nicht zugelassene Indikation, Unter- oder Überdosierung
  - Nichteinhaltung von Kontraindikationen
  - Verbotene Begleitmedikation

Die Forschungsdatenbank GePaRD

- Aufbau und Pflege **seit 2004**
- Basiert auf Abrechnungsdaten 4 gesetzlicher Krankenversicherungen von **> 20 Mio. Versicherten**
- Enthält Daten von ungefähr **17% der deutschen Allgemeinbevölkerung** und deckt alle geographischen Regionen der Bundesrepublik ab
- Enthält demographische Angaben, Informationen zu Krankenhausaufenthalten, ambulanten Arztbesuchen und Arzneimittelverschreibungen

Die Forschungsdatenbank GePaRD

Enthaltene Informationen



Die Forschungsdatenbank GePaRD

- Monitoring der Anwendung von Arzneimitteln und Impfstoffen nach Zulassung
- Untersuchung von
  - Arzneimittel- und Impfstoffanwendung, Fehlversorgung bei bestimmten Indikationen
  - (Seltenen) Arzneimittel- und Impfstoffrisiken
  - Arzneimittel-/Impfstoffrisiken mit einer langen Latenzzeit
  - Arzneimittelinteraktionen

Die Forschungsdatenbank GePaRD

- Studien vergleichsweise schnell und kostengünstig durchführbar
- Retrospektive Kohortenstudien und Fall-Kontroll-Studien möglich
- Große Patientenzahl und viele Datenjahre
- Untersuchung seltener, schwerwiegender und sich spät manifestierender Ereignisse möglich
- Berücksichtigung besonderer Patientengruppen
  (z.B. Alte, Pflegebedürftige, Schwerkranke, Verstorbene)
- Spezifische Angaben zu erhaltenen Leistungen, Diagnosen und Arzneimitteln
- Keine studienbedingten Verzerrungen
  (Erinnerungsfehler etc.)
- Hohe Repräsentativität
  (keine Einwilligung nötig, kaum Loss-to-Follow-up)

Die Forschungsdatenbank GePaRD

Daten wurden zu anderem Zweck erhoben:

Großes Wissen über Möglichkeiten und Limitationen unerlässlich!!

Pharmacoepidemiology and Drug Safety *Explore this journal >*

Original Report

Linkage of mother–baby pairs in the German Pharmacoepidemiological Research Database

Edeltraut Garbe, Marc Suling, Sebastian Kloss, Christina Lindemann, Ulrike Schmid

First published: 11 November 2010 Full publication history

Original Report

Evaluation of pregnancy outcome records in the German Pharmacoepidemiological Research Database (GePaRD)†

RT Mikolajczyk, AA Kraut, E Garbe

First published: 3 June 2013 Full publication history

DOI: 10.1002/pds.3467 View/save citation

Behr S, Kollhorst B, Dippel F-W, Garbe E. **Identification of patients with diabetes mellitus type 2 in insurance claims data - Are problems caused by ambiguous ICD-10 codes?** 28th International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE), 23-26 August 2012, Barcelona, Spain. (Abstract published in: Pharmacoepidemiology & Drug Safety. 2012;21(Suppl.3):396)

Die Forschungsdatenbank GePaRD

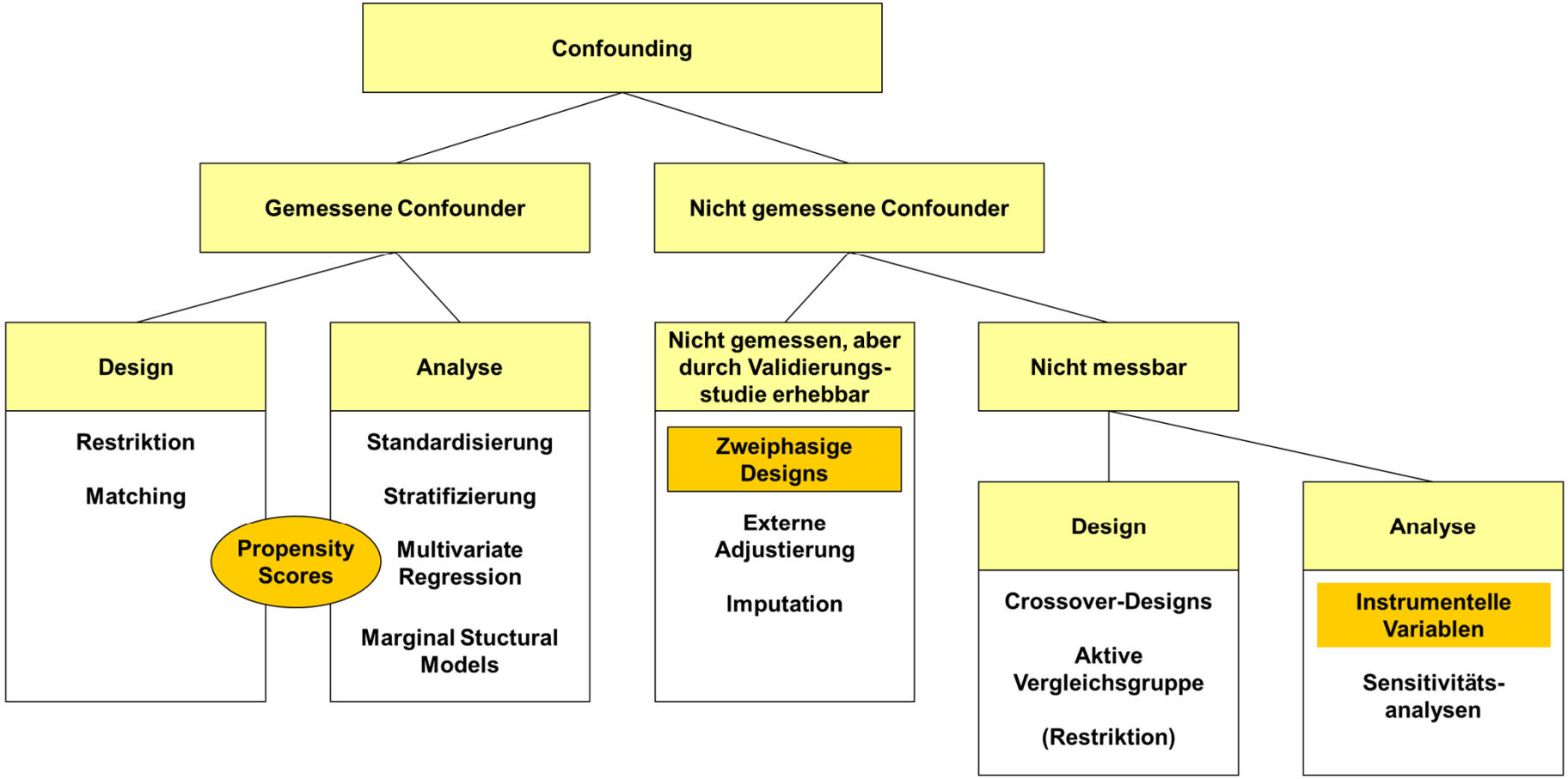
- Potential für Über-, Unter- und Fehlkodierungen durch Verwendung von Abrechnungsdaten
- Fehlende Zeiträume vor Einrichtung der Datenbank
- Fehlende Arzneimittel-Angaben
  (z.B. tatsächliche Arzneimitteleinnahme, Tagesdosis, Therapiedauer, OTC-Präparate)
- Fehlende Laborparameter
- Fehlende Angaben zu Lebensstilfaktoren
  (z.B. Bewegung, Gewicht, Rauchen)

➢ Unmeasured Confounding

Mögliche Lösungsansätze

- **Statistische Ansätze**
  - Propensity Scores
  - Instrumental Variables
  - Two-Phase Designs
  - ….
- **Datenanreicherung durch Linkage mit anderen Datenquellen**
  - Primärdaten
  - Registerdaten
  - Sekundärdaten
  - …

Statistische Ansätze



High-dimensional Propensity Scores I

- **Propensity Score (PS):** W-keit für Behandlung, bedingt auf **manuell** definierten Kovariablen ($\boldsymbol{X}_m$)
- **Problem:** Viel Information der Datenbank bleibt ungenutzt

➢ **High-dimensional PS (HDPS) Algorithmus:** Erweiterung des PS um **automatisch** selektierte Kovariablen ($\boldsymbol{X}_a$) mit größtem Potential zur Confounderkontrolle (B)

High-dimensional Propensity Scores II

HDPS-Algorithmus:

1. Bildung von Kovariablen zu häufigsten Codes in Datenbank für jeweilige Studienpopulation
2. Berechnung:

$$B = \frac{P_{C1}(RR_{CD}-1)+1}{P_{C0}(RR_{CD}-1)+1}, \text{falls } RR_{CD} > 1 \; ; \; \frac{P_{C1}\left(\frac{1}{RR_{CD}}-1\right)+1}{P_{C0}\left(\frac{1}{RR_{CD}}-1\right)+1} \text{ sonst} \quad \text{(Bross 1966)}$$

P_{C1}; P_{C0}= Prävalenz unter Behandelten; Nicht-Beh.; RR_{CD}= rel. Risiko für Outcome

3. Sortierung nach B, $\boldsymbol{X}_a$= top n Kovariablen (oft mehrere hundert)
4. Anpassung eines logistischen Modells: $\text{logit}(HDPS) = \beta_0 + \boldsymbol{\beta}'_m \boldsymbol{X}_m + \boldsymbol{\beta}'_a \boldsymbol{X}_a$
5. Nutzung des HDPS analog zum PS (z.B. Matching, Stratifizierung)

High-dimensional Propensity Scores III

- **Ziel:**
  Vergleich des Mortalitätsrisikos zwischen ambulant und stationär durchgeführten perkutanen koronaren Interventionen (PCIs) zur Behandlung der koronaren Herzkrankheit (KHK)

- **Design:**
  Kohortenstudie basierend auf Daten von 3 Krankenkassen (2004-2009)

- **Confounding durch Krankheitsschwere:**
  Sektor der PCI abhängig von allg. Gesundheitszustand des Patienten und Schwere der KHK

- **Statistische Analysen zur Kontrolle des Confounding:**
  - Primäre Analyse: Multivariable Cox-Regression, adjustiert für kardiovaskuläre Komorbiditäten und Prädiktoren von Krankenhausmortalität
  - Sensitivitätsanalysen basierend auf PS
    (Matching, Inverse-Probability-of-Treatment-Weighting)

Enders D, Ohlmeier C, Garbe E. The potential of high-dimensional propensity scores in health services research: An exemplary study on the quality of care for elective percutaneous coronary interventions. Health Services Research 2016 (im Druck)

High-dimensional Propensity Scores IV

- **Ergebnis:**
  Signifikant reduzierte Mortalität um 40% für ambulante PCIs in allen Analysen
- **Kontraintuitiv:**
  Leicht erhöhte Mortalität erwartet (Standardisierung und bessere Überwachung im Krankenhaus)

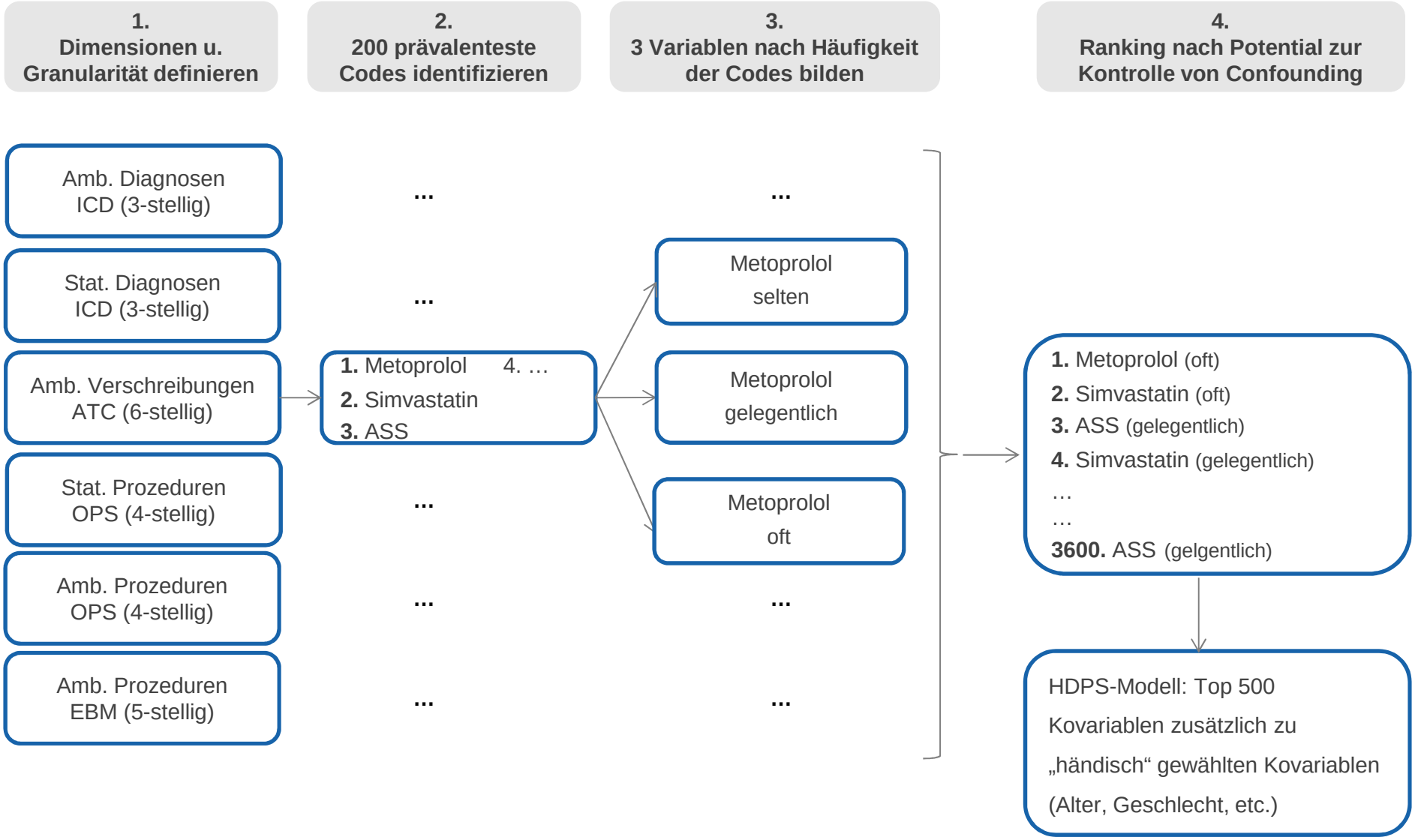
➢ **Möglicherweise verursacht durch ungemessenes Confounding:**
Schwere der KHK nicht erfasst in GePaRD

➢ **Lösungsansatz: High-dimensional PS (HDPS)**

- Automatische Selektion von hunderten empirischen Confoundern aus GePaRD
- „Hoffnung“: Gesamtheit an empirischen Confoundern ist Proxy für Schwere der KHK

High-dimensional Propensity Scores V

Anwendungsbeispiel: HDPS-Algorithmus



High-dimensional Propensity Scores VI

- Erneute Analyse der Kohorte basierend auf HDPS
  (Matching, Inverse-Probability-of-Treatment-Weighting)
- **Ergebnis:**
  Erhöhte Mortalität um 20% für ambulante PCIs in HDPS-Analysen (nicht signifikant)

➢ Plausible Ergebnisse der HDPS-Analysen

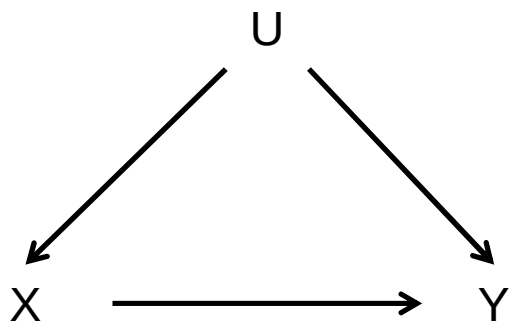
➢ Mit Vorsicht zu interpretieren:

- Generalisierbarkeit eingeschränkt: Gesündere und wesentlich kleinere Studienpopulation als in der ursprünglichen Kohorte
- Ungemessenes Confounding trotz umfangreicher Adjustierung nie auszuschließen

Instrumentelle Variablen I

Interesse:

Effekt einer Exposition X auf den Gesundheitsstatus Y



Problem:

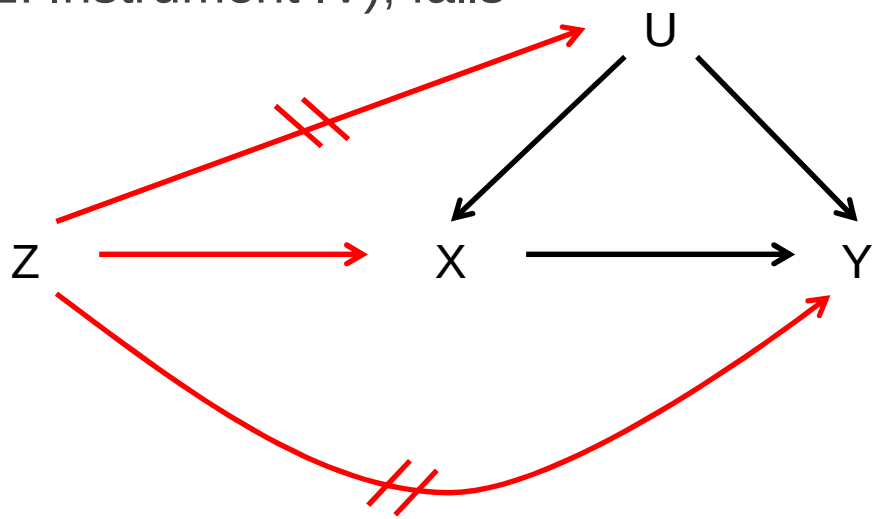
Exposition und Gesundheitsstatus abhängig von weiteren, insb. **unbeobachteten** Einflussfaktoren U

➢ **Confounding** → verzerrte Schätzung des Expositionseffekts

Instrumentelle Variablen II

Idee: Finde beobachtbare Variable Z, die zufällig Exposition voraus-sagt, aber keinen direkten Einfluss auf Gesundheitsstatus hat

- **Z instrumentelle Variable** (kurz: Instrument IV), falls
  1. $Z \not\perp\!\!\!\perp X$
  2. $Z \perp\!\!\!\perp U$
     Independence Assumption
  3. $Z \perp\!\!\!\perp Y \mid (X,U)$
     Exclusion Restriction



$A \perp\!\!\!\perp B \mid C$: A bedingt unabhängig von B, gegeben C

Instrumentelle Variablen III

Beachte:

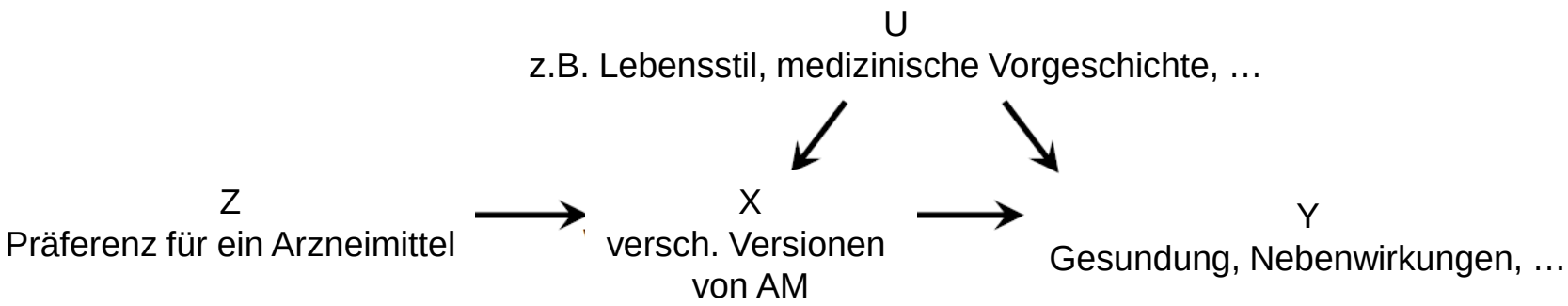
Obige Annahmen nicht unkritisch, aber notwendig für konsistente Schätzung:

1. Benötigt wird IV, das starke Assoziation mit X aufweist:
   falls **Z-X Assoziation schwach** (schwaches IV) im Verhältnis zum Stichprobenumfang: IV-Schätzer basierend auf (Z, X, Y) typischerweise verzerrt in Richtung naïvem Schätzer (Stock et al. 2002)

2. und 3. können nicht direkt empirisch überprüft werden
   $\rightarrow$ **Hintergrundwissen !**

Instrumentelle Variablen IV

Beliebtes IV: Arztpräferenz



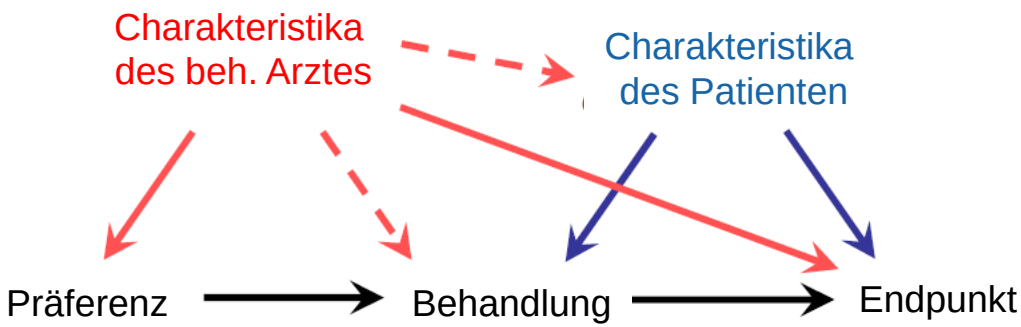
Beispiel:

- Behandlung X = Coxibe vs. NSAIDs (Brookhart et al. 2006)
- „Confounding by Indication“ großes Problem
- Vorschlag: Arztpräferenz für bestimmtes Arzneimittel als IV

Instrumentelle Variablen V

Größtes Problem:

IV Arzt-Präferenz nicht durch Confounding aufgrund Patienten-Charakteristika beeinflusst, aber durch Confounding aufgrund von Charakteristika des behandelnden Arztes



Hier:

- $Z \not\perp\!\!\!\perp Y \mid (X,U)$ und
- möglicherweise $Z \not\perp\!\!\!\perp U$ (Z=Präferenz, U=Patientencharakteristika)

Lösung: stratifiziere oder adjustiere nach Charakteristika des beh. Arztes

Instrumentelle Variablen VI

Anwendungsbeispiel: Arztpräferenz IV für aktuelle Coxibe-Verschreibung?

- **3 Definitionen, um die Arztpräferenz zu schätzen:**
  - Behandlung, die dem vorherigen Patienten verschrieben wurde
  - Anteil aller vorheriger Patienten, denen COX-2 Hemmer verschrieben wurden
  - Indikator für Anzahl der COX-2 Hemmer in vorherigen 7 Verschreibungen (nur Subkohorte 3)

- **Gesamte Kohorte + 3 Subkohorten:**
  - Nur Patienten mit rheumatoider Arthritis oder Osteoarthritis
  - Nur Patienten, die von Allgemeinmedizinern behandelt wurden
  - Nur Patienten von Ärzten, die mind. 20 Patienten in der gesamten Kohorte behandelt haben

Instrumentelle Variablen VII

Zu 1 (starkes IV?):

Stärke der Assoziation zwischen Behandlung und Instrument mittels

- partieller *F*-Statistik
- partieller quadrierter Korrelation r^2 und
- geschätztem Effekt des IVs auf Behandlungswahrscheinlichkeit

Zu 2 (Independence Assumption?):

Balancierung der gemessenen Kovariablen mittels partieller *F*-Statistik

Zu 3 (Exclusion Restriction?):

Assoziation zwischen IV und gastroprotektiven Medikamenten mittels logistischer Regression

Instrumentelle Variablen VIII

- Anteil aller vorheriger Patienten, denen COX-2 Hemmer verschrieben wurden:
  - Moderate Stärke der Assoziation zwischen Behandlung und IV
  - Balancierung der gemessenen Kovariablen (Independence Assumption)
  - Assoziation zwischen IV und gastroprotektiven Medikamenten in der gesamten Kohorte und Subkohorte 3 (Exclusion Restriction)

➢ **Valides Instrument in einem Gesundheitssystem nicht zwangsläufig übertragbar auf andere Gesundheitssysteme**

Instrumentelle Variablen VIV

	Konventionelle Analyse RD (95% KI)	IV Analyse RD (95% KI)
Gesamte Kohorte	-0,01 (-0,15; 0,12)	-0,02 (-0,80; 0,76)
Subkohorte 1	0,08 (-0,17; 0,32)	0,16 (-1,95; 2,27)
Subkohorte 2	0,00 (-0,18; 0,18)	0,05 (-1,05; 1,15)
Subkohorte 3	0,08 (-0,16; 0,32)	-0,10 (-1,24; 1,04)

RD: Risikodifferenz; KI: Konfidenzintervall

- IV Schätzer immer weitere KIs als konventionelle Schätzer, bedingt durch moderate Assoziation zwischen Behandlung und Instrument
- Möglicherweise keine Beeinträchtigung der konventionellen Analyse durch ungemessenes Confounding

Instrumentelle Variablen VIV

	Konventionelle Analyse RD (95% KI)	IV Analyse RD (95% KI)
Gesamte Kohorte	-0,01 (-0,15; 0,12)	-0,02 (-0,80; 0,76)
Subkohorte 1	0,08 (-0,17; 0,32)	0,16 (-1,95; 2,27)
Subkohorte 2	0,00 (-0,18; 0,18)	0,05 (-1,05; 1,15)
Subkohorte 3	0,08 (-0,16; 0,32)	-0,10 (-1,24; 1,04)

RD: Risikodifferenz: KI: Konfidenzintervall

ELSEVIER

CrossMark

Journal of Clinical Epidemiology 69 (2016) 96–106

Journal of Clinical Epidemiology

The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs

Bianca Kollhorst[a,*], Michal Abrahamowicz[b,c], Iris Pigeot[a,d]

[a]*Leibniz Institute for Prevention Research and Epidemiology—BIPS, Department Biometry and Data Management, Achterstr. 30, Bremen 28359, Germany*
[b]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal H3A 1A2, Québec, Canada*
[c]*Division of Clinical Epidemiology, McGill University Health Centre Royal Victoria Hospital, 687 Pine Avenue West, Montreal H3A 1A1, Québec, Canada*
[d]*Faculty of Mathematics and Computer Science, University of Bremen, Bibliothekstraße 1, Bremen 28359, Germany*

Accepted 24 August 2015; Published online 2 September 2015

Two-Phase Designs I

1. **Studiendesign:** Fall-Kontroll-Studie
2. **Zwei Datenquellen:**
   - Daten für alle Patienten (Phase 1): Fall-Kontroll-Status, Exposition, Confounder
   - Zusätzliche Confounder für Teil der Patienten (Phase 2)
3. **Stratifizierung aus Exposition und Auswahl an Confoundern in Phase 1**
   - Teilnahme an Phase 2 hängt nur von Stratifizierung und Fall-Kontroll-Status ab (Missing-at-Random)
   - Keine leeren Strata in Phase 1 und Phase 2 für Fälle bzw. Kontrollen
4. **Analysemethoden:**
   Weighted Likelihood, Pseudo Likelihood, **Maximum Likelihood**

Two-Phase Designs II

1. Phase

Exposition	Krankheit +	Krankheit -
+	N_{11}	N_{01}
-	N_{12}	N_{02}

Balanciertes Design

$n_{11}=n_{01}=n_{12}=n_{02}$

2. Phase

Exposition	Krankheit +	Krankheit -
+	n_{11}	n_{01}
-	n_{12}	n_{02}

Confounder	C_1	C_2	C_3

z.B. BMI, Rauchen

Exposition						
+	n_{111}	n_{011}	n_{112}	n_{012}	n_{113}	n_{013}
-	n_{121}	n_{021}	n_{122}	n_{022}	n_{123}	n_{023}

Two-Phase Designs III

- **Ziel:**

  Vergleich Therapie Sulfonylharnstoffe vs. DPP4-Inhibitoren hinsichtlich des kombinierten Risikos aus Myokardinfarkt, Schlaganfall und Herzinsuffizienz in Typ 2-Diabetikern mit 1st-Line Therapie Metformin

- **Studiendesign:**

  Eingebettete Zwei-Phasen Fall-Kontroll-Studie

- **Datenquellen:**
  - Phase 1: Pharmakoepidemiologische Forschungsdatenbank (GePaRD)
  - Phase 2: Disease Management Programm Diabetes mellitus Typ 2

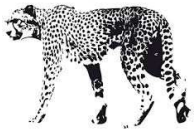
Two-Phase Designs V

- **Fälle:**
  Patienten mit Myokardinfarkt, Schlaganfall oder Herzinsuffizienz innerhalb des Follow-up
- **Kontrollen:** gematcht im Verhältnis 1:10
- **Exposition bei Kohorteneintritt:**
  - Current Exposure Sulfonylharnstoffe
  - Current Exposure DPP4-Inhibitoren

➢ **Teilnahme an Phase 2:**
Untersuchung im DMP vor Kohorteneintritt bis 28 Tage danach

Two-Phase Designs VI

Phase 1

GePaRD 

- Alter, Geschlecht
- Kardiovaskuläre Komorbiditäten
- Schwere des Diabetes
- Sonstige Komorbiditäten
- Komedikationen

Phase 2

- Rauchverhalten
- BMI
- HbA1c-Wert
- Schulungen
- Diabetes-bezogene Überweisungen

Two-Phase Designs VII

1. Phase 1 Analyse

- Logistische Regression mit allen Patienten
- Nur Confounder aus GePaRD

2. Complete-Case Analyse

- Gewichtete logistische Regression mit Patienten aus Phase 2
- Confounder aus GePaRD und DMP
- Gewicht=Inverse der Selektionswahrscheinlichkeit für Teilnahme an Phase 2 (geschätzt)

3. 2-Phasen Analyse

- 2-Phasen logistische Regression
- Confounder aus GePaRD und DMP
- Stratifizierung nach Exposition, Alter & Geschlecht
- Maximum Likelihood Methode

Two-Phase Designs VIII

	Odds Ratio (95%-Konfidenzintervall)		
	Phase 1 Analyse	**Complete-Case Analyse**	**2-Phasen Analyse**
DPP4-Inhibitoren vs. Sulfonylharnstoffe	0.81 (0.60–1.09)	0.76 (0.52–1.12)	0.83 (0.61–1.13)
Geschlecht (Referenz: weiblich)	1.43 (1.05–1.96)	1.52 (1.01–2.29)	1.37 (0.99–1.89)
Alter (Referenz: <55 Jahre)			
55- < 65 Jahre	1.60 (1.08–2.39)	1.78 (1.05–3.04)	1.62 (1.08–2.45)
>= 65 Jahre	2.72 (1.87–3.97)	2.78 (1.65–4.68)	3.01 (1.98–4.60)
Vorgeschichte ischämische Herzkrankheit (Referenz: nein)	1.91 (1.43–2.55)	1.76 (1.18–2.63)	1.78 (1.22–2.59)
Vorgeschichte Schlaganfall (Referenz: nein)	3.19 (2.00–5.10)	2.81 (1.38–5.71)	2.64 (1.31–5.29)
Vorgeschichte Herzinsuffizienz (Referenz: nein)	2.29 (1.58–3.31)	2.09 (1.27–3.44)	2.18 (1.32–3.58)
HbA1c >=7.5% (Referenz: nein)	-	1.23 (0.86–1.75)	1.22 (0.86–1.72)
BMI (Referenz: >=30)	-		
<25	-	1.17 (0.66–2.09)	1.17 (0.66–2.09)
25-<30	-	1.09 (0.76–1.56)	1.09 (0.76–1.56)
Raucher (Referenz: nein)	-	1.38 (0.83–2.30)	1.38 (0.83–2.30)

Two-Phase Designs IX

- Odds Ratio für Behandlung nur unwesentlich durch Informationen des DMP beeinflusst
- Mögliche Erklärungen:
  - Adäquate Berücksichtigung wesentlicher Confounder durch Information in GePaRD
  - Geringe Qualität der DMP-Daten
- 2-Phasen Analyse im Vergleich zu Complete-Case Analyse präziser

Journal of Diabetes and Its Complications 30 (2016) 1339–1346

Contents lists available at ScienceDirect

Journal of Diabetes and Its Complications

journal homepage: WWW.JDCJOURNAL.COM

Comparative risk for cardiovascular diseases of dipeptidyl peptidase-4 inhibitors vs. sulfonylureas in combination with metformin: Results of a two-phase study

Dirk Enders [a], Bianca Kollhorst [a], Susanne Engel [b], Roland Linder [b], Frank Verheyen [b], Iris Pigeot [a,*]

[a] Leibniz Institute for Prevention Research and Epidemiology—BIPS, Bremen, Germany
[b] Scientific Institute of TK for Benefit and Efficiency in Health Care, Hamburg, Germany

Bisherige Linkage-Erfahrungen I

- Daten eines Krankenhausinformationssystems (KIS)
- Evaluierung eines indirekten Linkage-Ansatzes im Vergleich zu direktem Linkage (hohe datenschutzrechtliche und damit große zeitliche/finanzielle Anforderungen) als Goldstandard
- Übereinstimmung von Aufnahme- und Entlassungsdiagnose sowie zusätzliche Diagnosen

➢ Spezifität durchgängig hoch

➢ Sensitivität abhängig vom Grad der Übereinstimmung der Diagnosen

- 30%-70%: 86,7%
- 100%: 41,7%

➢ Viele fehlende Angaben in KIS

➢ Aufwand gerechtfertigt?

Gesundheitswesen 2015; 77(02): e8-e14
DOI: 10.1055/s-0034-1395644

Originalarbeit

© Georg Thieme Verlag KG Stuttgart · New York

Verknüpfung von Routinedaten der Gesetzlichen Krankenversicherung mit Daten eines Krankenhausinformationssystems: Machbar, aber auch „nützlich"?

Linkage of Statutory Health Insurance Data with those of a Hospital Information System: Feasible, but also "Useful"?

C. Ohlmeier[1], F. Hoffmann[2, 2], K. Giersiepen[2, 2], H. Rothgang[2, 2], R. Mikolajczyk[2, 4], H.-J. Appelrat[5], A. Elsässer[6], E. Garbe[2, 2]

Bisherige Linkage-Erfahrungen II

- Daten des Bremer Mortalitätsindex (BreMI)
- Probabilistisches Linkage auf Basis personenbezogener Daten (Vorname, 2. Vorname, Nachname, Geburtsname, Geschlecht, Geburtsdatum, Adresse des letzten Wohnorts)
- Deterministisches Rekord-Linkage anhand von pseudonymisierter Versichertennr. nicht möglich, da nicht vorhanden im BreMI

➢ Linkage möglich für 83,7% aller Todesfälle aus GePaRD

➢ Todesdatum in 97,1% korrekt

➢ Keine Unterschiede bzgl. Geschlecht, Alter oder Versichertenstatus

Pharmacoepidemiology and Drug Safety Explore this journal >

Original Report

Validating mortality in the German Pharmacoepidemiological Research Database (GePaRD) against a mortality registry

Christoph Ohlmeier, Ingo Langner, Edeltraut Garbe, Oliver Riedel

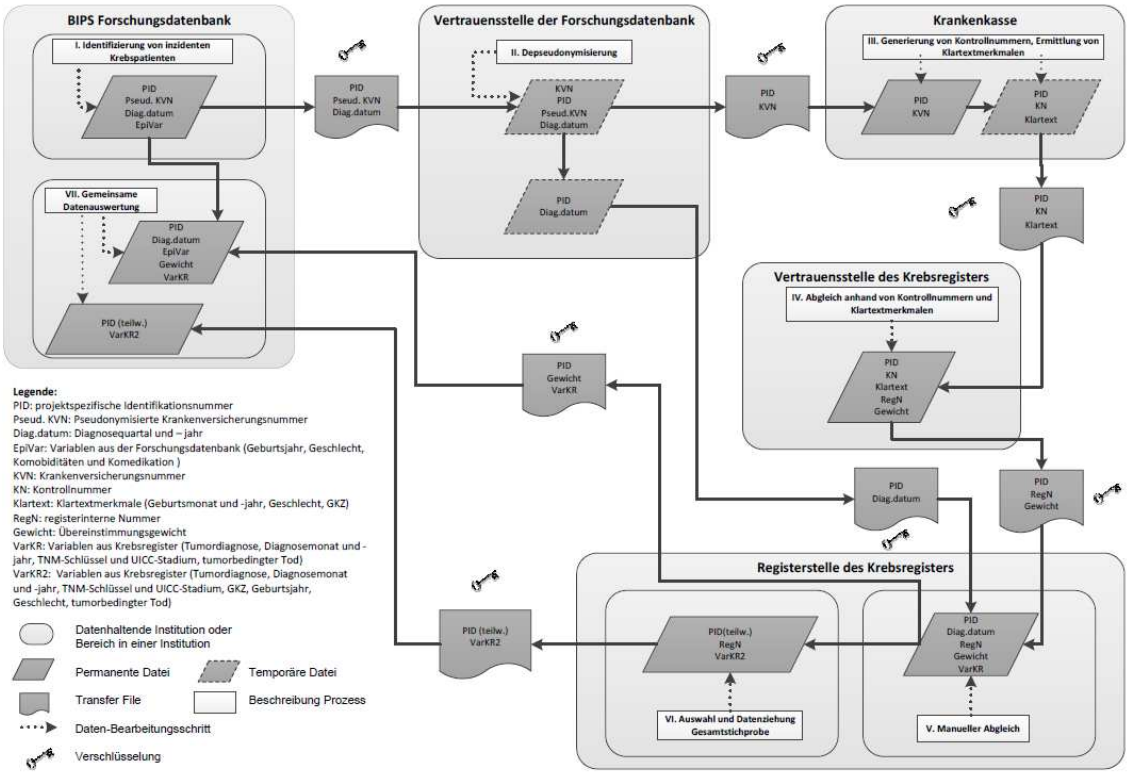
First published: 7 April 2016 Full publication history

Bisherige Linkage-Erfahrungen III

- Daten epidemiologischer Krebsregister
- DFG-Projekt: Beginn 01.12.2016

➢ **Ziel:** Evaluierung eines indirekten Linkage-Ansatzes im Vergleich zu direktem Linkage → kanzerogenes Potential von Antidiabetika

➢ Hohe datenschutzrechtliche und damit große zeitliche/finanzielle Anforderungen



Bisherige Linkage-Erfahrungen IV

- Befragungsdaten (Einverständniserklärung erforderlich)
- Aufwendiges Datenschutzkonzept
- Two-Phase Design, d.h. Teilstichprobe mit Befragungsdaten

➢ Geringe Response: 15.4%
(4000 angeschrieben → 3280 Briefe zugestellt → 505 Interviews durchgeführt)

Pharmacoepidemiology and Drug Safety *Explore this journal >*

Original Report

Does additional confounder information alter the estimated risk of bleeding associated with phenprocoumon use—results of a two-phase study

Sigrid Behr, Walter Schill, Iris Pigeot

First published: 2 February 2012 Full publication history

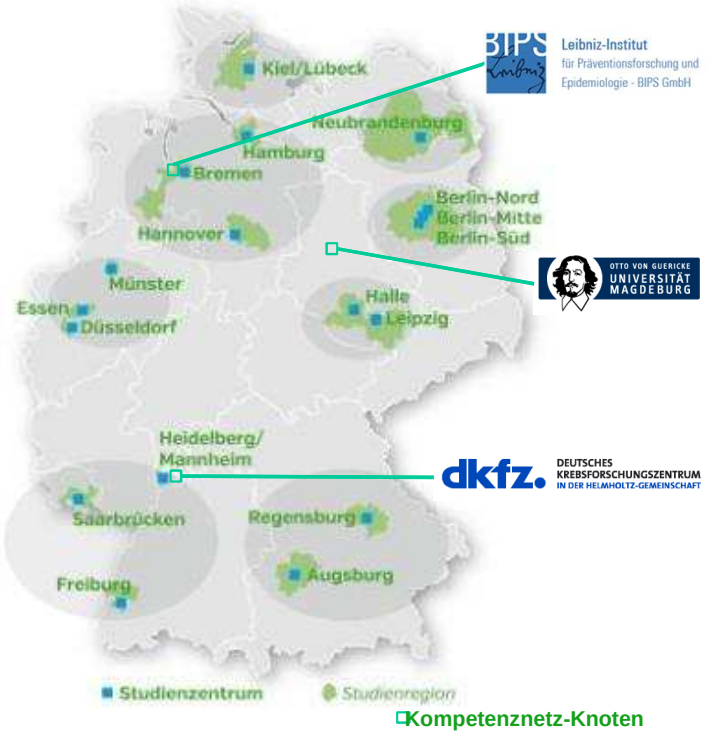
Bisherige Linkage-Erfahrungen V

- NAKO Gesundheitsstudie: prospektive Kohortenstudie (2013-2022; geplant für weitere 20-30+ Jahre) mit 200.000 Personen, 20-69 Jahre, 18 Studienzentren
- Kompetenznetz Sekundär- und Registerdaten

➢ **Aufgabe:** Beschaffung und Aufbereitung von GKV-Daten, Beratung von Nutzern

➢ Sehr hohe datenschutzrechtliche Anforderungen

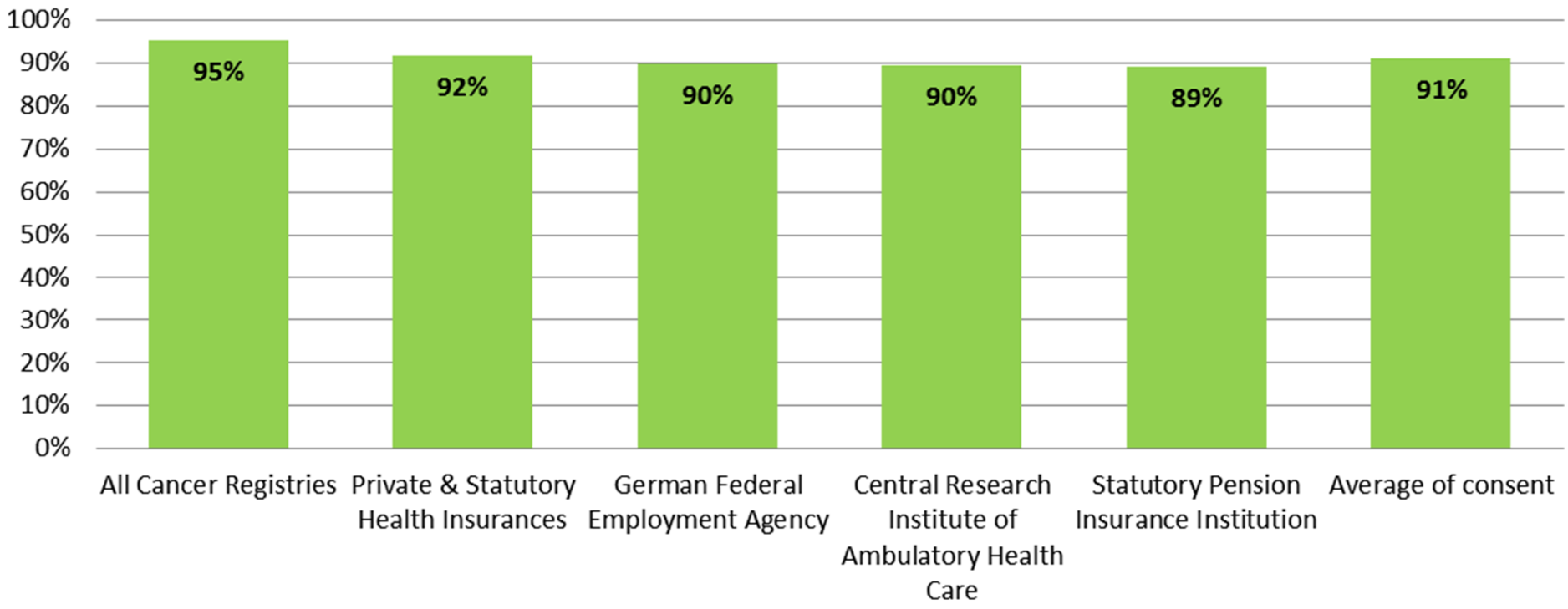
➢ Sobald Daten vorliegen, niedrigschwellige Nutzung für Forschung möglich



NAKO Gesundheitsstudie

Einverständnis zur Verknüpfung

- n=39.503 von 01.03.2015 bis 31.12.2015



NAKO Gesundheitsstudie

- Wichtige Informationen ohne zusätzliche zeitliche Belastung der Teilnehmer (z.B. Berufsbiographie)
- Keine Erinnerungslücken und hohes Maß an Spezifität, die bei Befragungen nicht erreicht werden kann (z.B. Medikamenteneinnahme)
- Erfassung vergangener und zukünftiger Ereignisse (z.B. Erkrankungen), auch wenn der Teilnehmer (z.B. aus gesundheitlichen Gründen) nicht ins Studienzentrum kommen kann
- Einzigartige Datenbasis durch Kombination verschiedener Datenquellen
- Validierung/Ergänzung der Primärdaten
- Vermeidung von Fehlklassifikation
- Unterstützung des Follow-up: Kontinuierliche detaillierte Informationen zu Gesundheit, Leistungsinanspruchnahme, Rentenverlauf und Berufsbiografie

Diskussion

Rechtliche Grundlage: § 75 SGB X

(1) Eine Übermittlung von Sozialdaten ist zulässig, soweit sie erforderlich ist für ein bestimmtes Vorhaben

1. der wissenschaftlichen Forschung im Sozialleistungsbereich oder der wissenschaftlichen Arbeitsmarkt- und Berufsforschung oder
2. der Planung im Sozialleistungsbereich durch eine öffentliche Stelle im Rahmen ihrer Aufgaben

und schutzwürdige Interessen des Betroffenen nicht beeinträchtigt werden oder das öffentliche Interesse an der Forschung oder Planung das Geheimhaltungsinteresse des Betroffenen erheblich überwiegt.

Folgen einer **restriktiven Auslegung:**

Separate Datenziehung für jede einzelne Forschungsfrage

Diskussion

- **Zweckbindung**
- **Zeitliche Befristung**

Damit:

- Verhinderung des Aufbaus langfristiger Forschungsdatenbanken
- Fehlende medizinische Vorgeschichte

⇨ Kein Arzneimittelmonitoring

⇨ Keine schnelle Reaktion auf vermeintliche Risiken

⇨ Keine Untersuchung des Off-label Use bei Kindern

⇨ ….

Aus dem Tätigkeitsbericht 2013/14 der BfDI:
„Ich empfehle dem Gesetzgeber, eine Änderung des § 75
herbeizuführen, die sowohl die Interessen der Wissensch
Rechte der betroffenen Bürgerinnen und Bürger angemes

Vielen Dank für Ihre Aufmerksamkeit!

www.leibniz-bips.de

Kontakt

Prof. Dr. Iris Pigeot

Leibniz-Institut für Präventionsforschung
und Epidemiologie – BIPS
Achterstraße 30
28359 Bremen

pigeot@leibniz-bips.de