

Medizinische Fakultät Carl Gustav Carus
Zentrum für Evidenzbasierte Gesundheitsversorgung

Gesundheitsberichterstattung mit GKV Routinedaten – Möglichkeiten, Grenzen und Herausforderungen

Falko Tesch, Anja Zscheppang, Jochen Schmitt
Hamburg, 24. April 2024

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Interessenkonflikte

Falko Tesch und Anja Zscheppang sind angestellt im Projekt RESILIENT vom Bundesministerium für Gesundheit im Förderschwerpunkt „Strukturelle Stärkung und Weiterentwicklung des Öffentlichen Gesundheitsdienstes (ÖGD)“.

Jochen Schmitt erhielt Forschungsmittel durch den Gemeinsamen Bundesausschuss, das Bundesministerium für Gesundheit, das Bundesministerium für Bildung und Forschung, die Europäische Union, das Bundesland Sachsen, Novartis, Sanofi, ALK und Pfizer. Er gibt Zahlungen von Sanofi, Lilly und ALK für Beratung an.

Jochen Schmitt ist Mitglied im Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen und in der Pflege (SVR) sowie Stellvertretender Vorsitzender des Deutschen Netzwerks Versorgungsforschung e.V.

Hintergrund

Für die Gesundheitsberichterstattung (GBE) ist es unabdingbar, Daten aus verschiedenen Datenquellen heranzuziehen, wie:

- **Daten des Öffentlichen Gesundheitsdienstes (ÖGD)**
- **Daten des Umweltamtes**
- **Daten von Umfragen**
-

Sekundär- bzw. Routinedaten sind „**prozessproduzierte, umfangreiche Informationssammlungen, die im Rahmen der Verwaltung, Leistungserbringung bzw. Kostenerstattung** (z. B. bei der Gesetzlichen Kranken-, Renten- oder Unfallversicherung sowie in Arzt- oder Krankenhausinformationssystemen) anfallen und elektronisch erfasst sind“ (Hoffmann & Glaeske, 2017).

Informationen zu medizinischen Diagnosen und Behandlungen fallen an bei:

- **Arztpraxen**
- **Apotheken- sowie Heil- und Hilfsmittelrechenzentren**
- **Kassenärztlichen Vereinigungen**
- **Kliniken**
- **Krankenkassen**

Routinedaten der gesetzlichen Krankenversicherung I

Mit der Einführung der digitalen Datenerfassung im Gesundheitssektor wurden nach und nach Daten für die Forschung möglich. Seit 1985 wird etwa der Arzneiverordnungs-Report veröffentlicht. Seit 2004 liegen ambulante Diagnosen bei Krankenkassen vor.

- **Stammdaten (§ 28a Abs.3 SGB V)**
Alter, Geschlecht, Nationalität, Schul- und berufliche Bildung, Erwerbsstatus, berufliche Tätigkeit (seit 2011: 1.256 versch. Tätigkeiten), Wohnort, Versicherungsart, Versicherungszeiten
- **Ambulante Diagnosen und Leistungen (§ 295 SGB V)**
Jahr, Quartal, Fachgruppe, Diagnosen und -sicherheit, EBM-Ziffern, Vertragsarzt, Scheinart (Original/Überweisung), OPS-Codes für ambulante Operationen, Kosten
- **Arzneimittel-Verordnungen (§ 300 Abs. 1 SGB V)**
Jahr, Quartal, Arztbezogene Informationen, Anzahl Verordnungen, Abgabepreis, Präparat, Wirkstoffgruppe, Zuzahlungen, Abgabedatum
- **Heil- und Hilfsmittel (§ 302 SGB V)**
z.B. Physio- und Ergotherapie, Logopädie, Fußpflege, Ernährungsberatung (seit 2018), Rollstuhl, Gehhilfen, Inkontinenzprodukte

Routinedaten der gesetzlichen Krankenversicherung II

- **Stationäre Aufenthalte (§ 301 SGB V)**
Institutskennzeichen, Fachabteilung, Aufnahme- und Entlassdatum, Haupt- und Nebendiagnosen, Verlegungen, Prozeduren, DRGs, Entbindungen
- **Ambulante Operationen (§115b SGB V)**
- **Arbeitsunfähigkeiten (§ 5 EFZG)**
Anfangs- und Enddatum, Diagnosen
- **Medizinische Rehabilitation sofern nicht Renten/Unfallversicherung zuständig (§ 40 SGB V)**
- **Disease Management Programme (§ 137f SGB V teilweise Labordaten)**
- **Psychiatrische Institutsambulanzen (§ 118 SGB V)**
- **Spezialambulanzen (§ 116 SGB V, § 117 SGB V, § 118a SGB V, § 119 SGB V)**
- **Pflegedaten (§ 15 SGB XI)**
- **Prävention von Krankheiten (§ 20 SGB V)**

Routinedaten der gesetzlichen Krankenversicherung III

Vorteile

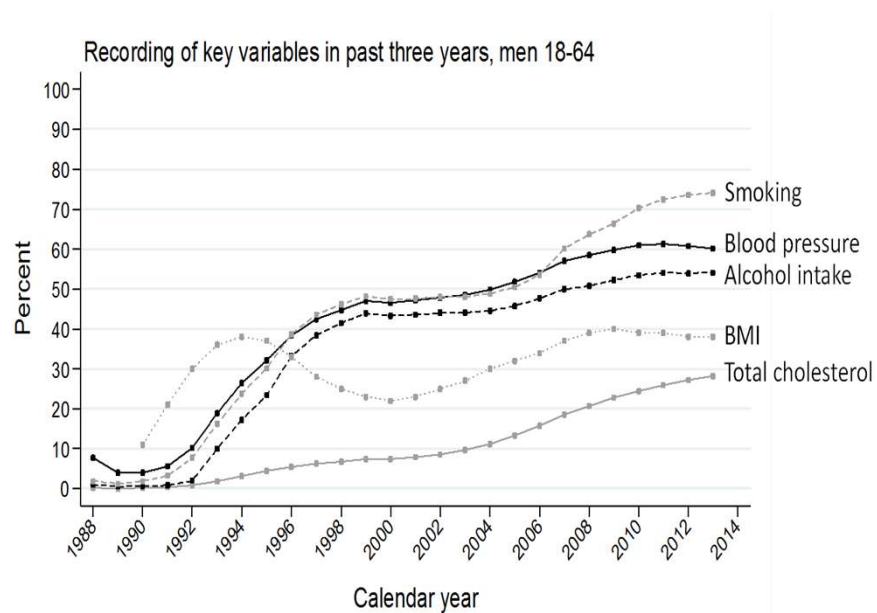
- Personenbezogen
- Sektorenübergreifend (ambulant, stationär)
- Ohne Verzerrungen („unbiased“)
- Kein Erinnerungsbias
- Kein Interviewerbias
- Keine Selektion durch Teilnahmeverweigerung oder fehlenden Zugang (Schwerkranke, Hochbetagte)
- Langzeitbeobachtung möglich
- Keine fehlenden Daten (außer Versicherungswechsel)
- Kostengünstig

Nachteile

- Studienpopulation z.T. nicht auf Gesamtbevölkerung übertragbar u.a. keine Landwirte/Privatversicherte
- Keine Angaben über Krankheiten ohne Arztkontakt
- Keine Angaben über Selbstmedikation/Gesundheitsverhalten
- Keine Privatrezepte/verschreibungsfreie Medikamente
- Potenzieller Misklassifikationsbias
- Daten sind abhängig von politischen und abrechnungstechnischen Regularien
- Kaum klinische Daten (BMI, Labor, Pathologie, Histologie) und Gesundheitsverhalten (ePA könnte ab 2025 diese Lücke füllen)

Routinedaten der gesetzlichen Krankenversicherung IV

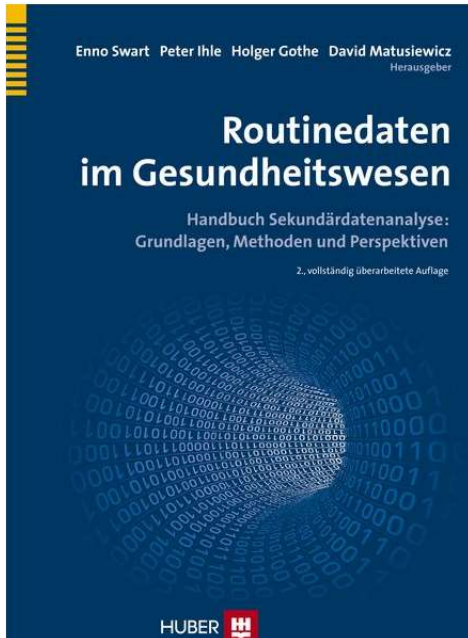
Die geplante Einführung einer elektronischen Gesundheitskarte ab 2025 könnte bei den Kassen nach den Diagnose-/Therapiedaten auch Labor-, Bild- und Verhaltensdaten ergänzen. Aktuell befindet sich ein Forschungsdatenzentrum Gesundheit im Aufbau, was mit einem Teil der Daten der gesetzlichen Krankenkassen gespeist wird (Regionalbezug Postleitzahl).



In Großbritannien werden ca. 11 Millionen Menschen vom Hausärztenetz des Clinical Practice Research Datalink (CPRD) abgedeckt, wofür teilweise Angaben zu BMI, Raucherstatus, Alkoholkonsum, Bluthochdruck und Cholesterin vorliegen (Herrett, 2015).

In Südkorea liegen von 12 der 50 Millionen Einwohner*innen neben den obigen auch Angaben zu Nüchternblutglucose, Nierenfunktion (Glomeruläre Filtrationsrate) und körperlicher Aktivität vor, die nicht älter als 2 Jahre sind (Lee, 2022).

Leitlinien und Berichtsstandards



Standardisierte BerichtsRoutine für Sekundärdaten Analysen (STROSA) – ein konsentierter Berichtsstandard für Deutschland, Version 2

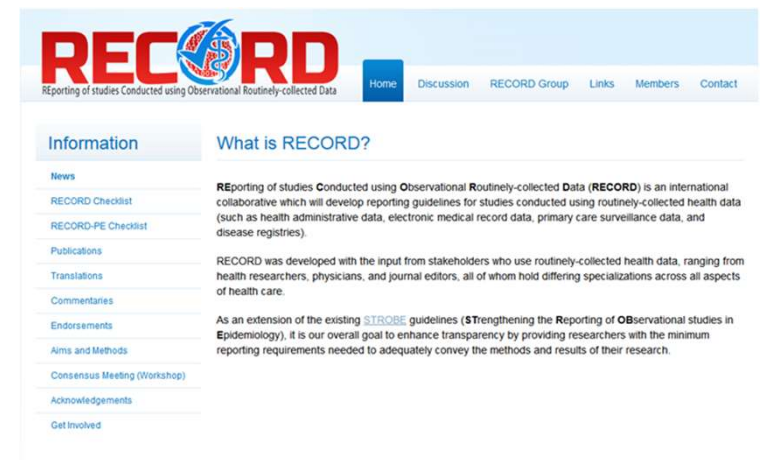
A Consensus German Reporting Standard for Secondary Data Analyses, Version 2 (STROSA-Standardisierte BerichtsRoutine für SekundärdatenAnalysen)

Gute Praxis Sekundärdatenanalyse (GPS)

Leitlinien und Empfehlungen

3. Fassung 2012, geringfügig modifiziert 2014

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



Validierung von GKV Routinedaten I

Die Datensammlung erfolgt für einen anderen Anlass als für wissenschaftliche Untersuchungen. Der Forschende hat keinen Einfluss auf die Datenerhebung und Qualität der Dokumentation, was eine Datenvalidierung erforderlich macht (Hoffmann, 2008).

Zu beachten ist:

- Stimmt eine Angabe aus den Routinedaten mit der Realität überein?
- Prüfungen der internen Konsistenz und Validität der Sekundärdaten
- Angabe ambulanter Diagnosen bisher nur quartalsgenau, damit keine Reihenfolge des Auftretens der Erkrankungen in kurzer Frist erkennbar
- Oftmals unspezifische Kodierungen „n.n. bez“ oder J45.9 (Asthma unspezifisch – unklar, ob allergische oder nichtallergische Ursache)
- Mehrdeutige ICD10-Codierungen gerade bei Seltenen Erkrankungen (Betroffen weniger als 1 von 10.000)
- Vielfalt der Codierungsmöglichkeiten (z.B. Symptom: Schwindel mit 42 verschiedenen ICD10-Codes)
- Erkrankungen weitgehend ohne Therapiemöglichkeiten in der GKV, wie etwa Adipositas oder Rauchen, werden stark untererfasst

Validierung von GKV Routinedaten II

Interne Konsistenzprüfung

Suche nach auffälligen Mustern innerhalb der Daten einer Krankenkasse, z.B.

- Konsistente Fallzahlen im Zeitverlauf
- Wiederkehrende Muster
- Abgleich von Verordnungs- und Abgabedatum von Rezepten
- Abgleich der Daten mit anderen Krankenkassen bzw. vergleichbaren Datenquellen

Interne Diagnosevalidierung

Abgrenzung zwischen sicheren und unsicheren Diagnosen sowie Schweregraden einer Erkrankung durch Nutzung von weiteren verfügbaren Informationen

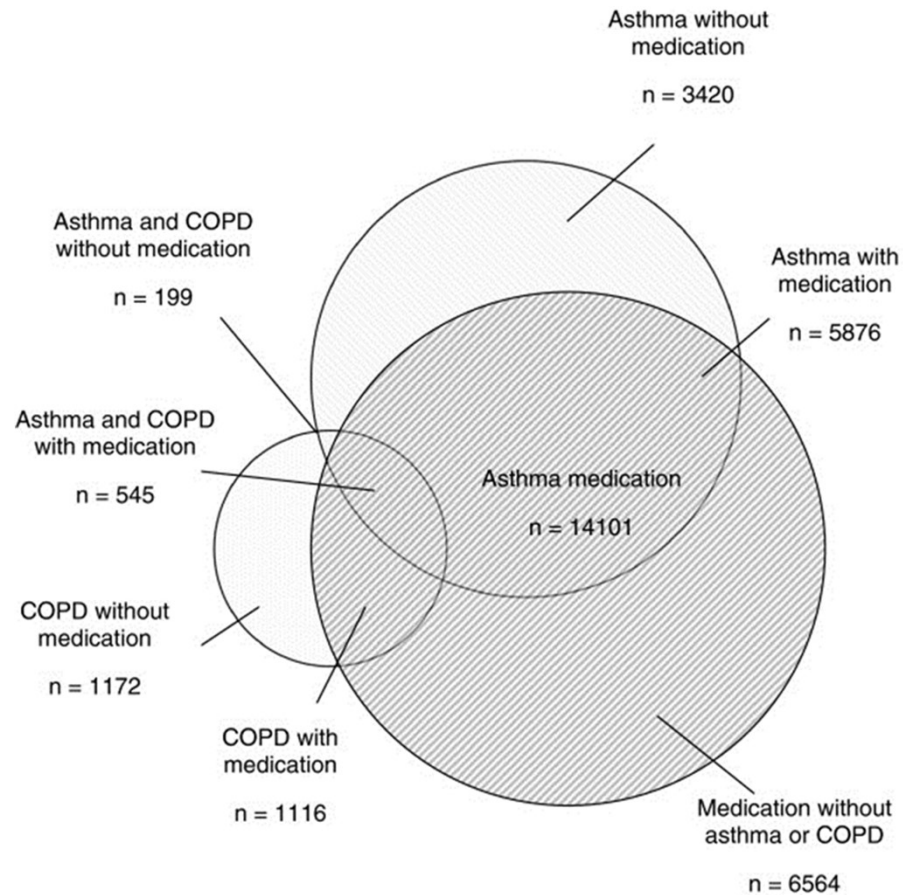
- Prozeduren, Medikamente, wiederholte Behandlungen, Fachärzte, stationäre Aufenthalte, Tod, ...

Externe Validierung

Validierung einer Teilstichprobe gegenüber einem externen Goldstandard, z.B. Ärzte- und/oder Patientenbefragung, Patientenakten, Laborberichte, ...

Validierung von GKV Routinedaten III

Wer hat Asthma?



Weidinger P, Nilsson JL, Lindblad U. Medication prescribing for asthma and COPD: a register-based cross-sectional study in Swedish primary care. *BMC family practice*. 2014;15:54.

Umgang mit seltenen Diagnosen I

Datenstabilität & Modifiable Areal Unit Problem (MAUP)

Tabelle 1
Datenstabilität

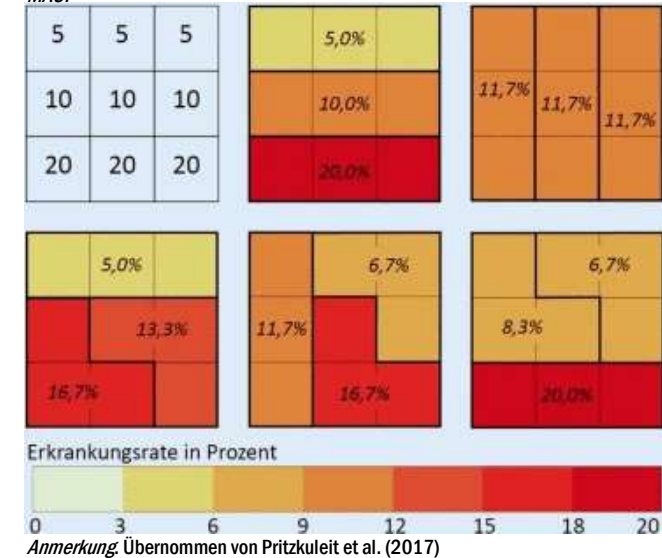
	Gemeinde A	Gemeinde B
Einwohner	1.700.000	1700
Erkrankungen pro Jahr	4000	4
Erkrankungsrate	235,3 pro 100.000	
+ 1 Erkrankung	235,4 pro 100.000 (+0,03 %)	294,1 pro 100.000 (+25 %)

Anmerkung: Übernommen von Pritzkeleit et al. (2015)



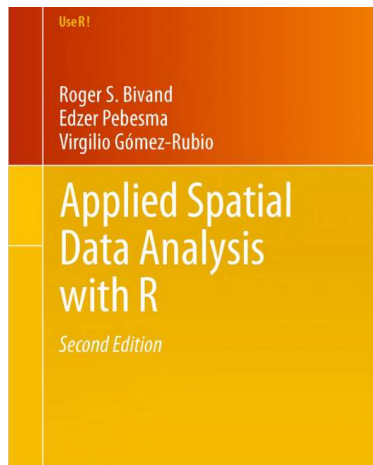
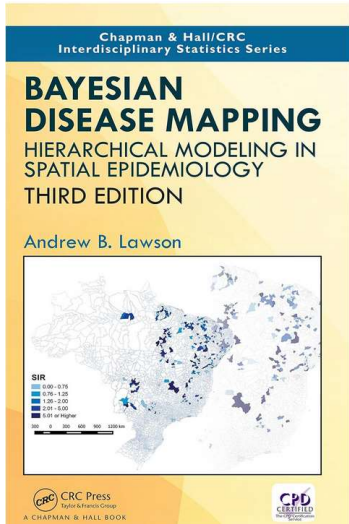
„Je geringer die unter Risiko stehende Bevölkerung und je seltener die Erkrankung, desto stärker neigen Inzidenz- bzw. Mortalitätsraten zu Extremraten.“ (Pritzkeleit et al., 2017)

Abbildung 1
MAUP



Umgang mit seltenen Diagnosen II

Zeitproblem und das „Disease Mapping“



- Bei kleinen räumlichen Ebenen und/oder seltenen Ereignissen kann man nicht 100 Jahre warten bis hinreichend Ereignisse aufgetreten sind, weil sich inzwischen auch die dahinterliegenden Risikofaktoren geändert haben.
- **Das Problem wird über zwei Wege gelöst:**
 - Die Inzidenz für das gesamte Untersuchungsgebiet kann relativ genau bestimmt werden und dies als „prior“ verwendet werden → Je weniger Fälle pro Einheit auftreten, desto mehr „drücken“ diese im Modell die Fallzahl zu dem, was für das gesamte Gebiet erwartet worden wäre.
 - Umliegende Einheiten sind sich oft ähnlicher als weit entfernte Einheiten.
- Annahme dahinter ist, dass die Rate für eine Einheit systematisch von den umliegenden Einheiten abhängt und zufällig um den Mittelwert der gesamten Untersuchungsregion normalverteilt streut.
- Beides wird in einem Bayesianischen CAR Modell oder auch BYM Modell berücksichtigt. Geringe Fallzahlen können trotzdem zu einer zu starken Glättung führen, weswegen auch 5er Jahreszeiträume vereinzelt zur Anwendung kommen könnten.

Herausforderungen

1. Datenhalter

- Eigeninteresse an den Ergebnissen der Forschung oft nicht gegeben

2. Regulatorische Hürden

- Datenschützer der Datenhalter überzeugen (Datenaggregation beim Datenhalter oder beim Datennutzer)
- Aufsichtsbehörde Bundesamt für Soziale Sicherung (BAS) oder Landessozialministerium überzeugen

3. Geographische Ebene

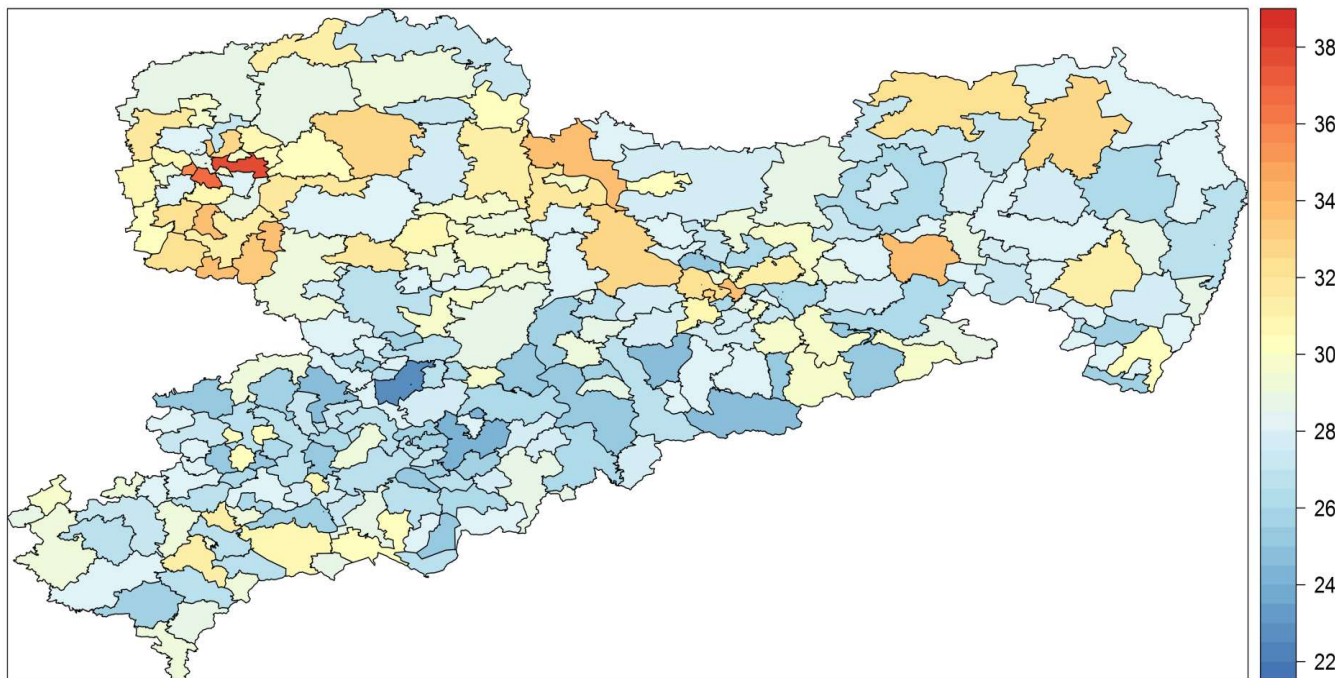
- Hinreichend hohe Zahl an GKV-Versicherten für die jeweiligen Merkmale erforderlich
- Postleitzahlen sind auf kleinteiliger Ebene weniger geeignet, aber oft unterste Ebene im Data Warehouse
- Adressdaten müssen auf möglichst (sozial)homogene Gebiete aggregiert werden

4. Zeitverzug

- Prüfung ambulanter Abrechnungen bei KV und Krankenkasse kann Zeitverzug von 7 bis 9 Monaten bedeuten
- Weiter zurückliegende Daten könnten bei Genehmigung des Vorhabens aufgrund von Löschfristen nicht mehr verfügbar sein
- Datenlieferung kann Zeitplan für Forschungsprojekte gefährden

Krebserkrankungen und Umweltfaktoren in Sachsen I

Mund- und Rachenkrebs Inzidenz (2010-14)

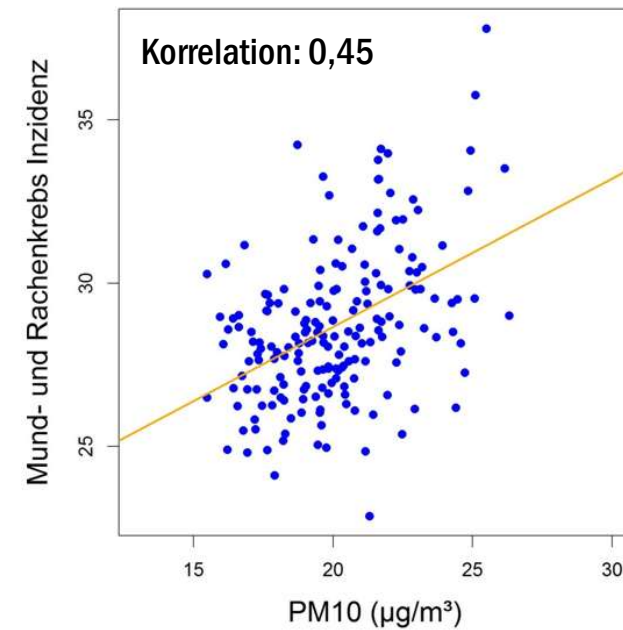
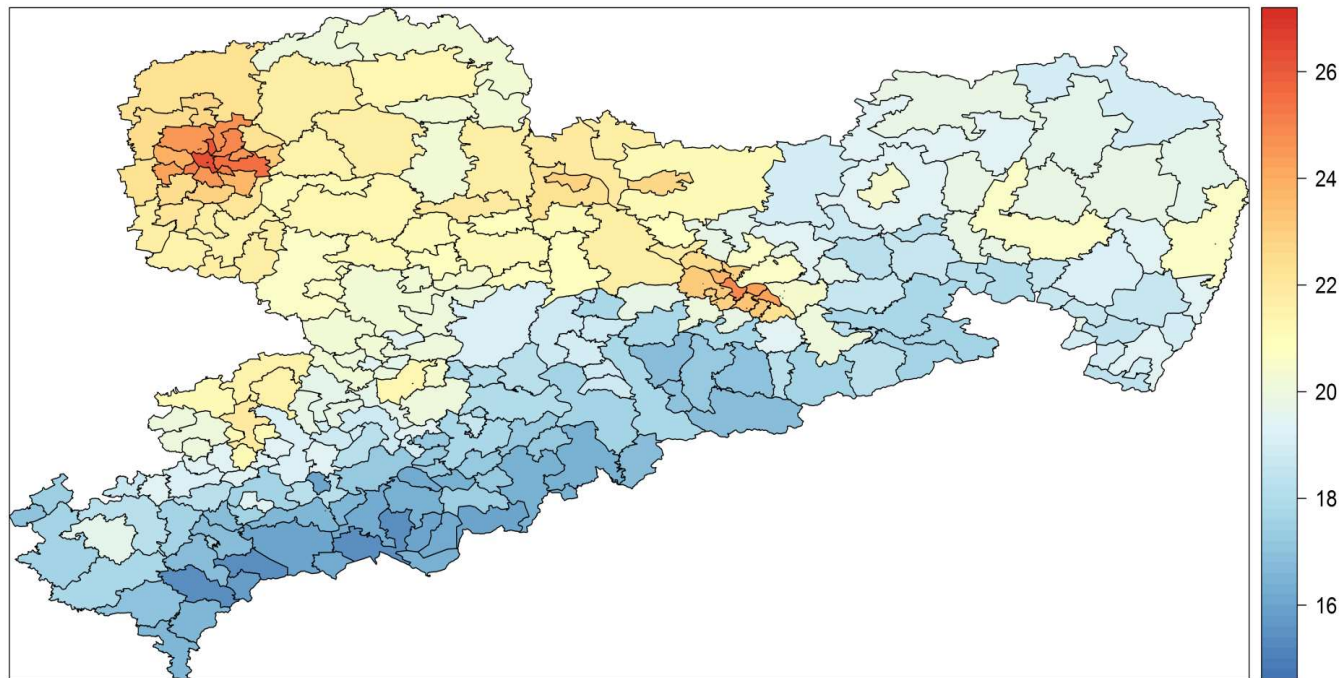


Darstellung der Inzidenz für Krebserkrankungen in Sachsen aufgrund von Daten der AOK PLUS für den Zeitraum 2008 bis 2014 (Basisperiode 2008/09) in 186 vierstelligen Postleitzahlbezirken.

Übertragung der Umweltdaten vom Small Area Health Statistics Unit <http://www.sahsu.org/content/data-download> auf die gleiche geographische Ebene

Krebserkrankungen und Umweltfaktoren in Sachsen II

Feinstaubbelastung 2007- PM10

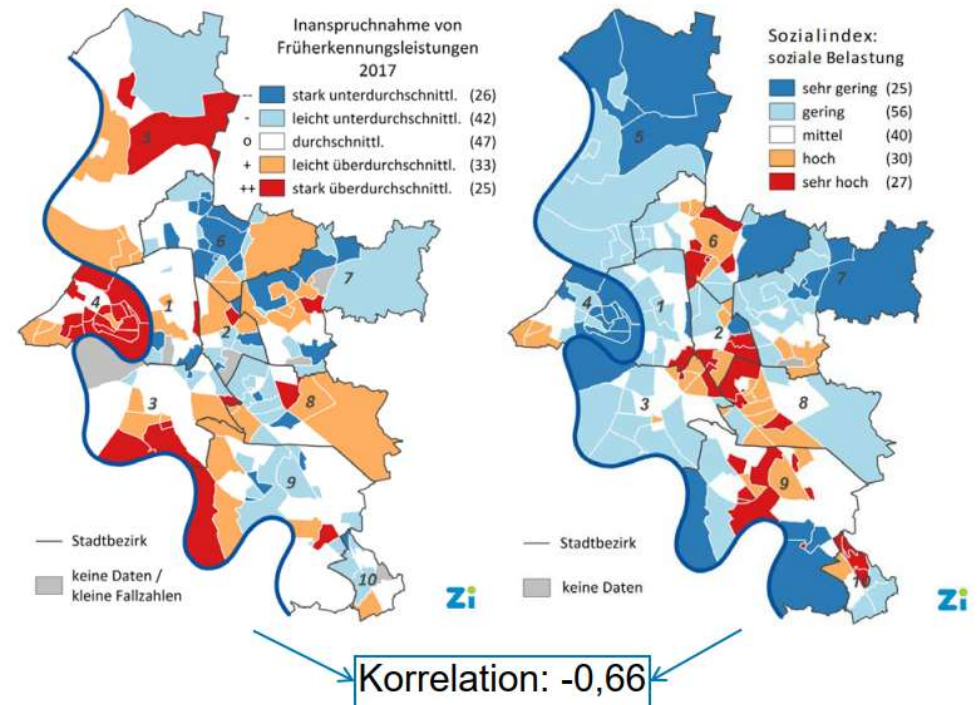


Diagnosen und Sozialdaten in Düsseldorf I

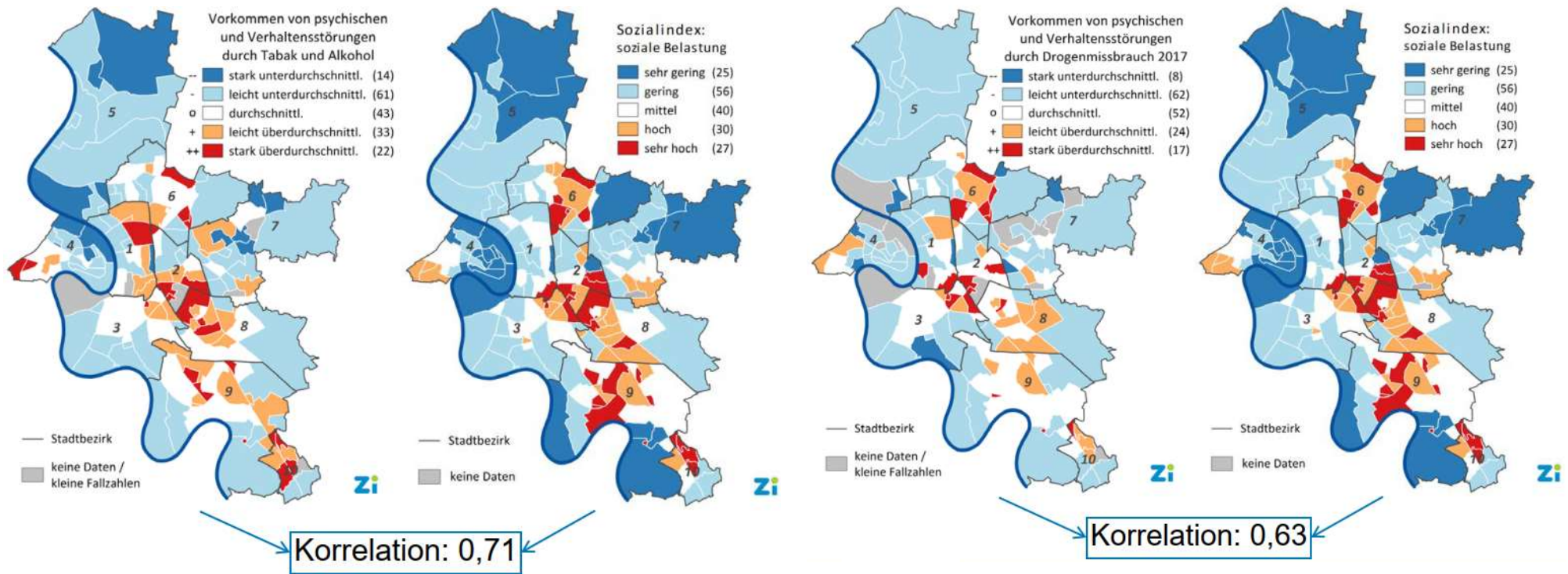
1997 wurden vom Jugendamt und Amt für Statistik und Wahlen 179 Sozialräume als sozial homogene Gebiete in Düsseldorf definiert.

Für diese Gebiete wurde der Index für soziale Belastung und Diagnosedaten der Kassenärztlichen Vereinigungen für GKV Versicherte für 2017 verbunden.

Der Sozialstatus kann einen starken Einfluss auf die Gesundheit haben.

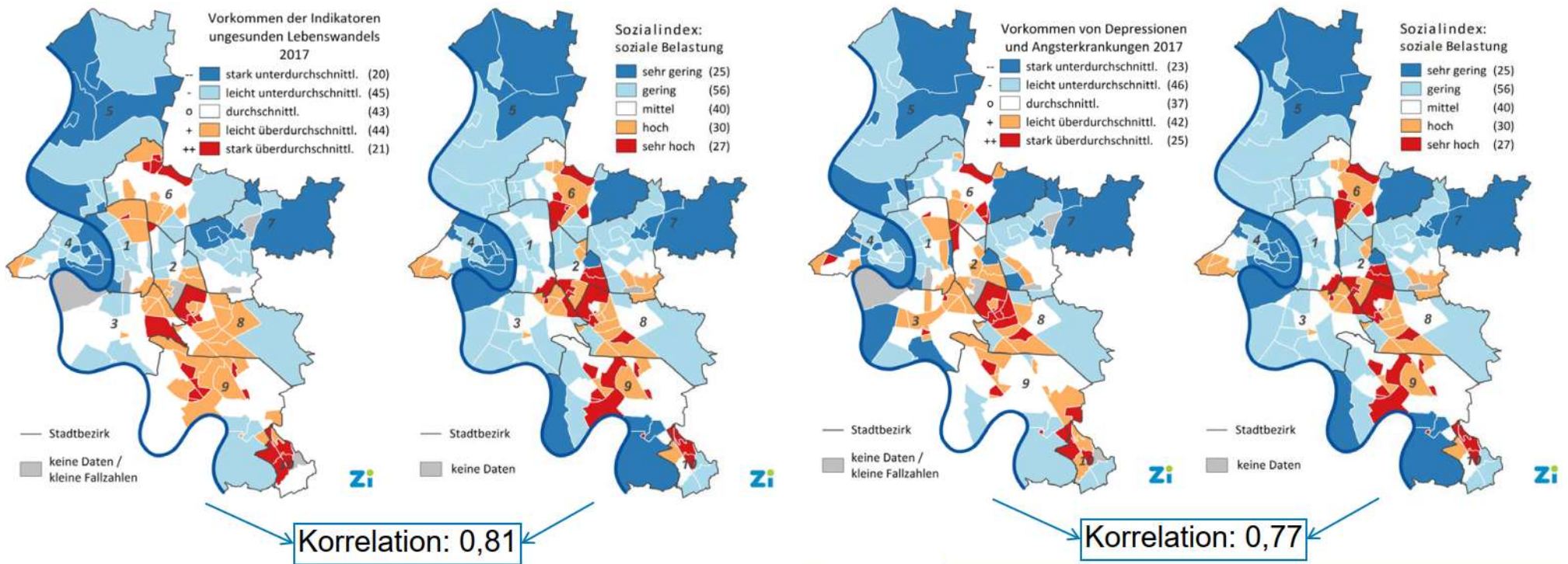


Diagnosen und Sozialdaten in Düsseldorf II



Diagnosen für Drogenkonsum sind dort häufiger, wo die „Soziale Belastung“ höher ist.

Diagnosen und Sozialdaten in Düsseldorf III



Diagnosen für somatische und psychische Erkrankungen sind dort häufiger, wo die „Soziale Belastung“ höher ist.

Fazit: Gesundheitsberichterstattung mit GKV Routinedaten

Möglichkeiten:

- GKV Routinedaten bieten umfangreiche Gesundheitsinformationen und sind auch auf tiefer regionaler Ebene abbildbar
- Daten sind nicht abhängig vom Erinnerungsvermögen oder der Teilnahmebereitschaft/ Teilnahmefähigkeit der Versicherten
- Perspektivisch können auch Laborwerte, Bildbefunde und Gesundheitsverhalten Teil des Datenkörpers werden (Einführung elektronische Patientenakte)

Grenzen:

- Keine Angaben über Krankheiten ohne Arztkontakt des Versicherten sowie Selbstmedikation
- Daten sind abhängig von der Organisation des Gesundheitssystems, etwa Untererfassung von Erkrankungen ohne bzw. mit eingeschränkten Therapieoptionen

Herausforderungen:

- Datenvalidierung ist ein essentieller Bestandteil des Forschungsprozesses
- Zugang zu GKV Routinedaten stellt trotz Bemühungen des BMG (Innovationsfonds und FDZ Gesundheit) größte Hürde dar, da dies auf der Freiwilligkeit der GKV Datenhalter und der Zustimmung der Aufsichtsbehörden basiert.

Referenzen

- Benchimol E. et. al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PloS Med. Oct 6 2015;12(10)
- Datzmann T et. al. Outdoor air pollution, green space, and cancer incidence in Saxony: a semi-individual cohort study. BMC Public Health. 2018 Jun 8;18(1):715. doi: 10.1186/s12889-018-5615-2. PMID: 29884153; PMCID: PMC5994126.
- Herrett, E. et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD), International Journal of Epidemiology, Volume 44, Issue 3, June 2015, Pages 827–836
- Hering, R. Kleinräumige kommunale Gesundheitsberichterstattung in Düsseldorf. Vortrag auf der Jahrestagung der Arbeitsgemeinschaft Nord-West im VDSt in Mettmann / 26.11.2019
- Hoffmann, F.; Glaeske, G. (2017): Analyse von Routinedaten. In: Pfaff, H.; Neugebauer, E. A. M.; Glaeske, G.; Schrappe, M. (Hrsg.): Lehrbuch Versorgungsforschung, 2. Aufl. Stuttgart: Schattauer: 122–127
- Hoffmann F. et al. (2008) Validierung von Sekundärdaten – Grenzen und Möglichkeiten. Bundesgesundheitsbl – Gesundheitsforsch – Gesundheitsschutz. 51:19.1118-1126
- Pritzkeleit, R. Die kartografische Darstellung regionaler Unterschiede in der Morbidität Möglichkeiten der Datenanalyse am Beispiel des kleinräumigen Krebsatlasses Schleswig-Holstein. Bundesgesundheitsbl 2017 · 60:1319–1327
- Lee SW, Lee J, Moon SY, et al. Physical activity and the risk of SARS-CoV-2 infection, severe COVID-19 illness and COVID-19 related mortality in South Korea: a nationwide cohort study. Br J Sports Med. 2022;56:901-912. [PMID: 34301715] doi: 10.1136/bjsports-2021-104203
- Swart E. et al. (2015) Gute Praxis Sekundärdatenanalyse (GPS) – Leitlinien und Empfehlungen. Gesundheitswesen 77: 120-126.

Medizinische Fakultät Carl Gustav Carus
Zentrum für Evidenzbasierte Gesundheitsversorgung

Vielen Dank für Ihre Aufmerksamkeit

Gefördert durch:



Bundesministerium
für Gesundheit

aufgrund eines Beschlusses
des Deutschen Bundestages