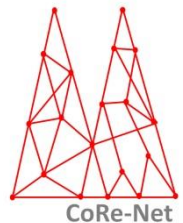




Aufbau und Entstehung unserer regionalen GKV-Datenbank in CoRe-Net

Dr. Nadine Scholten CoRe-Net Koordinatorin

Aufbau einer einzigartigen Datenbank für Köln



GKV-Routinedaten

4 Krankenkassen stellen Daten für den Aufbau der regionalen Datenbank aus Versicherten mit Wohnort in Köln (Regierungsbezirk Köln) zur Verfügung:

- AOK Rheinland/Hamburg (ca. 340.000 / ca. 1.54)
- BARMER (ca. 116.000 / ca. 510.000)
- DAK Gesundheit (ca. 74.000 / ca. 320.000)
- Pronova BKK (? / ca. 660.00)

Zeitraum: 2012 - 2017

Somit stehen in der Datenbank die GKV-Daten von über 50 Prozent der Kölner Bevölkerung zur Verfügung

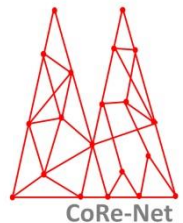
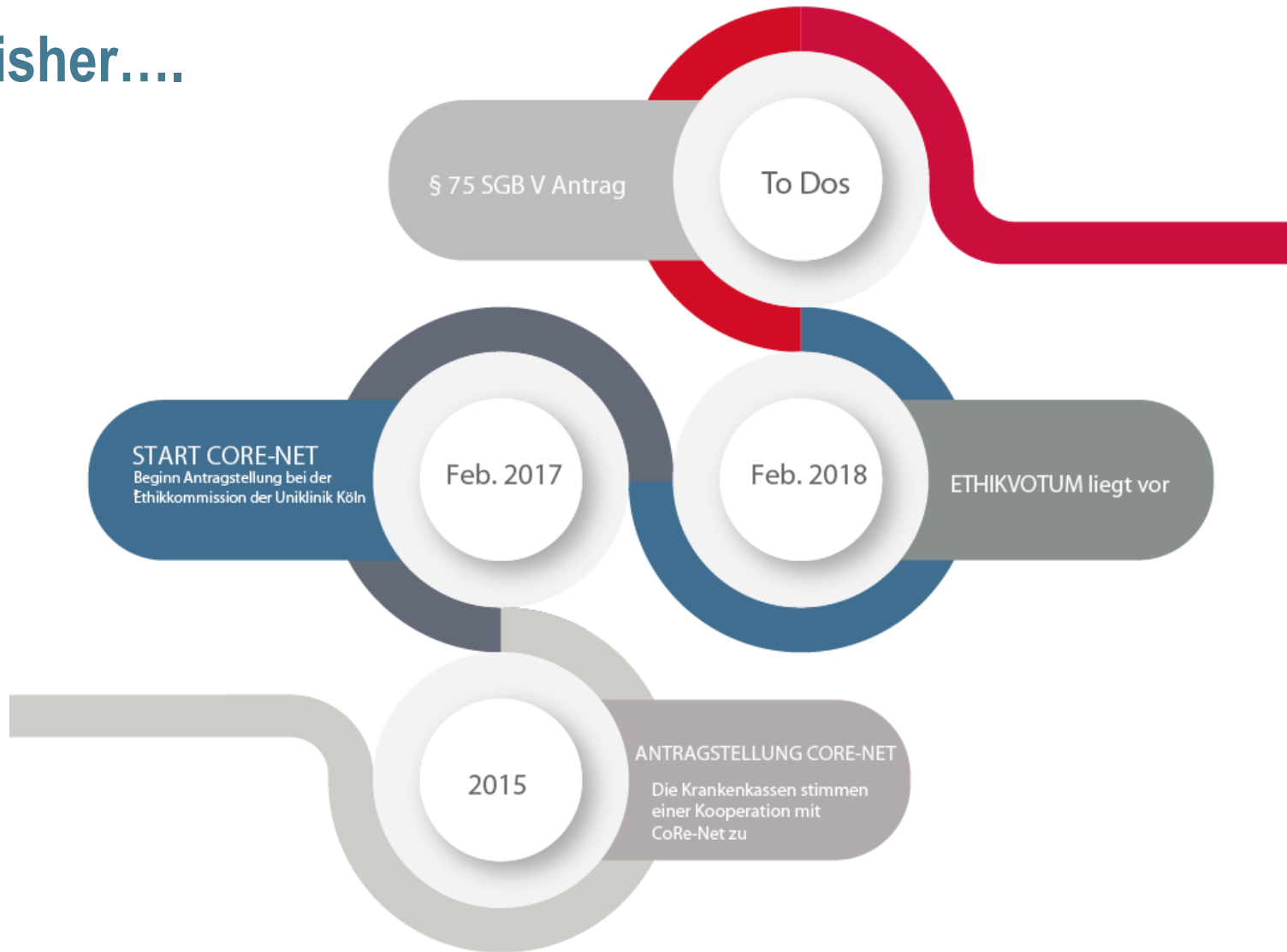


BARMER

DAK
Gesundheit

pronova BKK
Partner für Ihre Gesundheit

Bisher....



Rechtsgrundlage: § 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. Eine Übermittlung ohne Einwilligung des Betroffenen ist nicht zulässig, soweit es zumutbar ist, die Einwilligung des Betroffenen nach § 67b einzuholen. Angaben über den Familien- und Vornamen, die Anschrift, die Telefonnummer sowie die für die Einleitung eines Vorhabens nach Satz 1 zwingend erforderlichen Strukturmerkmale des Betroffenen können für Befragungen auch ohne Einwilligungen übermittelt werden.

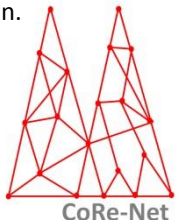
(2) Die Übermittlung bedarf der vorherigen Genehmigung durch die oberste Bundes- oder Landesbehörde, die für den Bereich, aus dem die Daten herrühren, zuständig ist. Die oberste Bundesbehörde kann das Genehmigungsverfahren bei Anträgen von Versicherungsträgern nach § 1 Absatz 1 Satz 1 des Vierten Buches auf das Bundesversicherungsamt übertragen. Die Genehmigung darf im Hinblick auf die Wahrung des Sozialgeheimnisses nur versagt werden, wenn die Voraussetzungen des Absatzes 1 nicht vorliegen. Sie muss

1. den Dritten, an den die Daten übermittelt werden,
2. die Art der zu übermittelnden Sozialdaten und den Kreis der Betroffenen,
3. die wissenschaftliche Forschung oder die Planung, zu der die übermittelten Sozialdaten verwendet werden dürfen, und
4. den Tag, bis zu dem die übermittelten Sozialdaten aufbewahrt werden dürfen,

genau bezeichnen und steht auch ohne besonderen Hinweis unter dem Vorbehalt der nachträglichen Aufnahme, Änderung oder Ergänzung einer Auflage.

(3) Wird die Übermittlung von Daten an nicht-öffentliche Stellen genehmigt, hat die genehmigende Stelle durch Auflagen sicherzustellen, dass die der Genehmigung durch Absatz 1 gesetzten Grenzen beachtet und die Daten nur für den Übermittlungszweck gespeichert, verändert oder genutzt werden.

(4) Ist der Dritte, an den Daten übermittelt werden, eine nicht öffentliche Stelle, gilt § 38 des Bundesdatenschutzgesetzes mit der Maßgabe, dass die Kontrolle auch erfolgen kann, wenn die Daten nicht automatisiert oder nicht in nicht automatisierten Dateien verarbeitet oder genutzt werden.

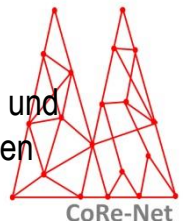


Datenschutz:

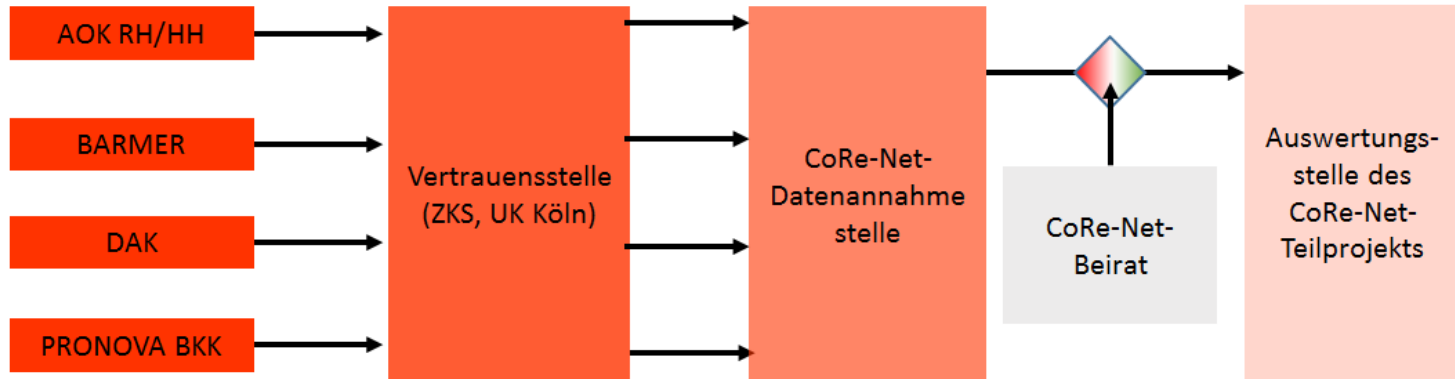
3. Der Aufbau der Routinedatenbank erfolgt auf Basis versichertenbezogener Daten, die im Lauf der Übermittlung von den Krankenkassen an CoRe-Net durch mehrere kaskadierende Bearbeitungsschritte für die CoRe-Net-Subprojekte anonymisiert werden.

1. Bereits in den Krankenkassen werden direkt personenidentifizierende Angaben (Name, Vorname, etc.) gelöscht, das Geburtsdatum wird nur jahresgenau geliefert und ein krankassenspezifisches Pseudonym generiert.
2. In der Pseudonymisierungsstelle werden die krankenkassenspezifischen Pseudonyme durch ein Einweghashverfahren (SHA-3) unter Verwendung eines über die Laufzeit des CoRe-Net-Projekts eindeutigen Salts zweitpseudonymisiert und an die CoRe-Net-Vertrauensstelle, wieder kryptographisch verschlüsselt, übermittelt. (Durchlaufstation mit Funktion der Zweitpseudonymisierung.)
3. Nach Übermittlung der Daten an die CoRe-Net-Vertrauensstelle wird aus dem Hashpseudonym ein CoRe-Net spezifischen Studienidentifikator (Dritt pseudonym) erzeugt, der über die Laufzeit von CoRe-Net identisch bleibt.
4. Nach Erstlieferung bzw. nach Verlinkung der aktuellen Datenlieferung mit dem bisherigen Datenstand erfolgt in der Vertrauensstelle eine Prüfung auf Anonymität der Daten. Bei Nachlieferung wird insbesondere geprüft, ob durch Hinzunahme der Angaben aus diesen Nachlieferungen die Anonymität weiterhin sichergestellt ist bzw. die Reidentifikation einzelner Personen nicht bzw. nur mit unverhältnismäßig hohem Aufwand möglich ist.
5. Die Bereitstellung der GKV-Routinedaten erfolgt projektspezifisch: Filterung auf Spaltenebene (nur benötigte Variablen), auf Zeilenebene (nur definierte Zielpopulation, nur notwendige Angaben, z.B. ICDs) und auf Profilebene (nur benötigte Sektoren), Aggregation von Variablen und Vergabe eines projektspezifischen Studienidentifikators.

Zunahme des Grads der Anonymität



Datenfluss im Rahmen von CoRe-Net



Die Daten werden für die Übermittlung an die Vertrauensstelle transportverschlüsselt.

Die **Vertrauensstelle** pseudonymisiert die von den Krankenkassen gelieferten Daten mit einem Salt-Wert und Einweg-Hashverfahren

Die Vertrauensstelle generiert hierfür einen über die Zeit konstanten Salt-Wert.

Es werden keine Sozialdaten gespeichert, sondern lediglich der Salt-Wert (Tresor und Backuptresor).

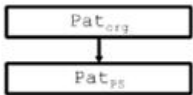
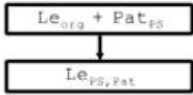
Die **Datenannahmestelle** erhält pro Lieferung nur den von den Krankenkassen jeweils neu übermittelten Datenbestand. Durch das über die Zeit gleichbleibende Zweitpseudonym können die neu gelieferten Daten mit dem bisherigen Datenbestand (faktisch anonym) verlinkt werden.

Die (teilprojektbezogene) **Auswertungsstelle** erhält einen für die Beantwortung der Frage zugeschnittenen Subsample der GKV-Daten für einen zeitlich begrenzten Zeitraum. Die Daten werden zeitlich befristet per kontrolliertem VPN-Zugang zur Verfügung gestellt. Löschrufen und Archivierung der Auswertungsdaten werden vertraglich festgeschrieben.

Einschub: Versichertenbezogene Anonymisierung der Leistungserbringer

Trennung von Personendaten und Leistungsdaten zur Erhöhung der Datensicherheit

„Für eine Versichertenbezogene Sekundärdatenanalyse beispielsweise zu Art und Umfang von Mehrfachinanspruchnahme (so genanntes Doctor-hopping) muss die Datenbasis die Fragestellung zulassen, ob ein Versicherter über die Zeit immer denselben Arzt oder unterschiedliche Ärzte in Anspruch nimmt, nicht jedoch, ob zwei Verschiedene Versicherte zum selben Arzt gehen.“

	Patientenbezogene Datenbasis
Beispiel für Fragestellung	Wie viele Ärzte sucht ein Patient (wg. einer Erkrankung) auf?
Pseudonymisierung des Patienten-kennzeichens	 <p>Konventionelle Pseudonymisierung des Patienten-kennzeichens (Pat_{org}) mit Generierung eines Patienten-pseudonyms (Pat_{PS}).</p> <p>Ein Patient erhält immer dasselbe Pseudonym.</p>
Pseudonymisierung des Leistungserbringer-kennzeichens	 <p>Nach Verkettung von Leistungserbringerkennzeichen (Le_{org}) und bereits pseudonymisiertem Patienten-kennzeichens (Pat_{PS}) wird ein patientenbezogenes Leistungserbringer-pseudonym generiert ($Le_{PS,Pat}$).</p> <p>Ein Leistungserbringer erhält in Abhängigkeit von dem behandelten Patienten <i>mehrere</i> Pseudonyme.</p>

S6 FB I, AG ERHEBUNG UND NUTZUNG VON SEKUNDÄRDATEN (AGENS)

» Die Vertrauensstelle im Rahmen der Sekundärdatenforschung – Lösungsansätze zum Problem der Datenkonzentration

P. Ihle¹, J. Krappweis², I. Schubert¹
¹ Forschungsgruppe Primärmedizinische Versorgung, Medizinische Einrichtungen der Universität zu Köln
² Institut für Klinische Pharmakologie, Technische Universität Dresden

Gesundheitswesen 2001; 63 Sonderheft 1: S6–S12
 © Georg Thieme Verlag Stuttgart · New York
 ISSN 0949-7013

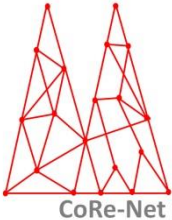
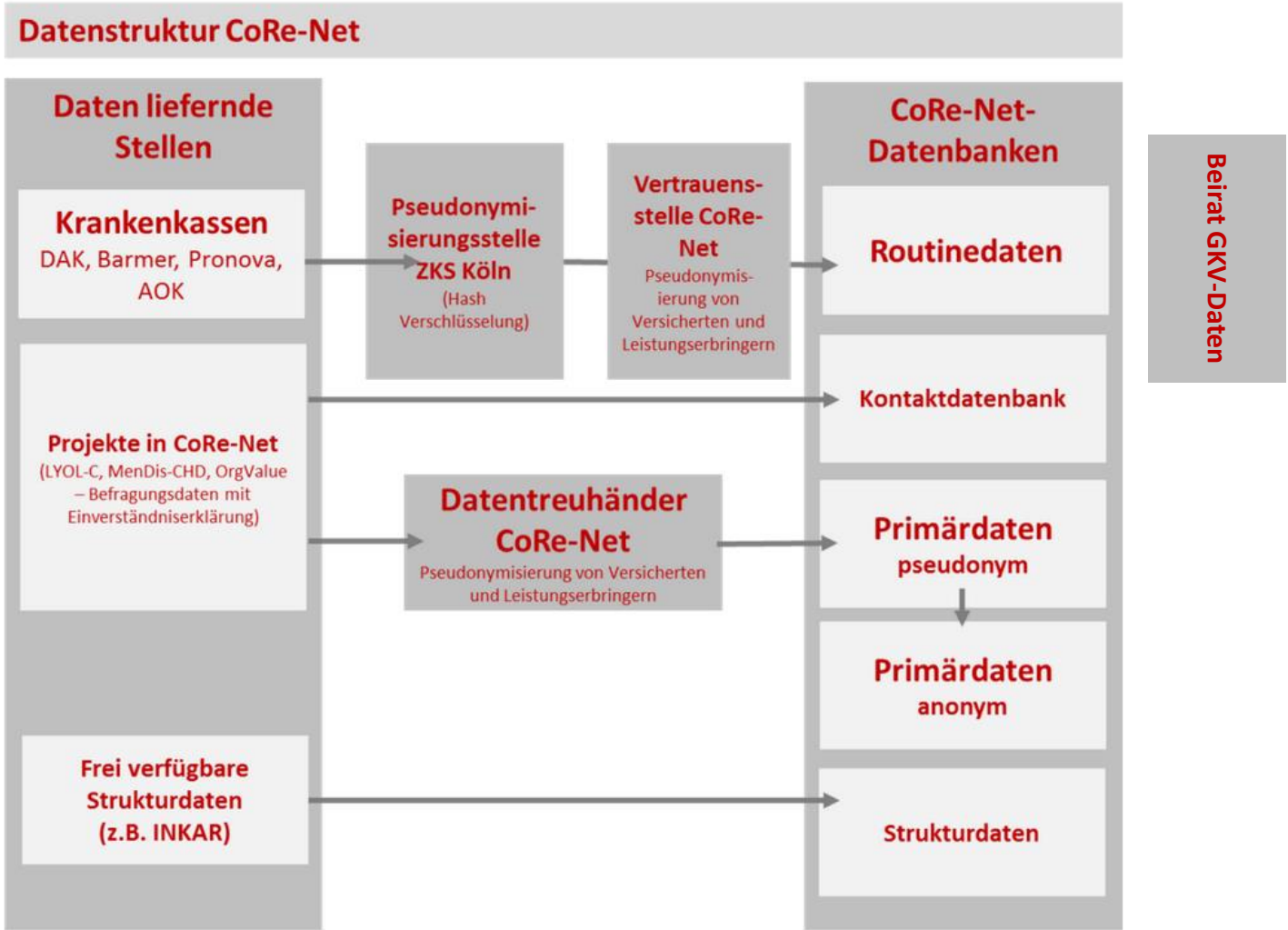
Regelung des Datenzugangs für weitere Projekte

Voraussetzung: positives Ethikvotum und Vorliegen eines ein-stimmigen Votums aller beteiligten Krankenkassen.

Ablauf:

1. Ziehen eines Subsamples, welches lediglich die zur Beantwortung der Forschungsfrage notwendigen Daten enthält (Stichwort: Datensparsamkeit). Filterung auf Spaltenebene (nur benötigte Variablen), auf Zeilenebene (nur definierte Zielpopulation, nur notwendige Angaben, z.B. ICDs) und auf Profilebene (nur benötigte Sektoren)
2. Die Studienidentifikatoren werden durch ein neues subprojektspezifisches Viertpseudonym ersetzt.
3. Die forschungsfragespezifischen Daten werden durch die Vertrauensstelle weiter vergrößert/aggregiert, dass eine Re-Identifikation einzelner Personen ausgeschlossen ist.
4. Das Subsample wird den Subprojekten in der Regel per VPN-Zugang zur Verfügung gestellt.
5. Eine physische Übermittlung der Daten erfolgt nur in begründeten Ausnahmefällen. Der Zugang ist für die Laufzeit des Subprojekts zeitlich beschränkt.

Datenbank





Vielen Dank für Ihre Aufmerksamkeit

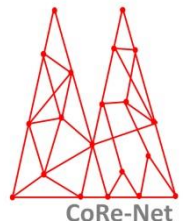
Dr. Nadine Scholten und
IMVR

nadine.scholten@uk-koeln.de

Peter Ihle
PMV forschungsgruppe

peter.ihle@uk-koeln.de

   www.core-net.koeln



CoRe-Net