

# INDEED – Record Linkage (RL)

RL von Notaufnahmeroutinedaten / Daten  
der kassenärztlichen Vereinigung im G-BA  
Projekt INDEED

Gefördert vom G-BA unter der Nummer 01VSF16044

## Ziel des Projekts und RL-Szenario



Inanspruchnahme und sektorenübergreifende Versorgungsmuster von Patienten in Notfallversorgungsstrukturen in Deutschland

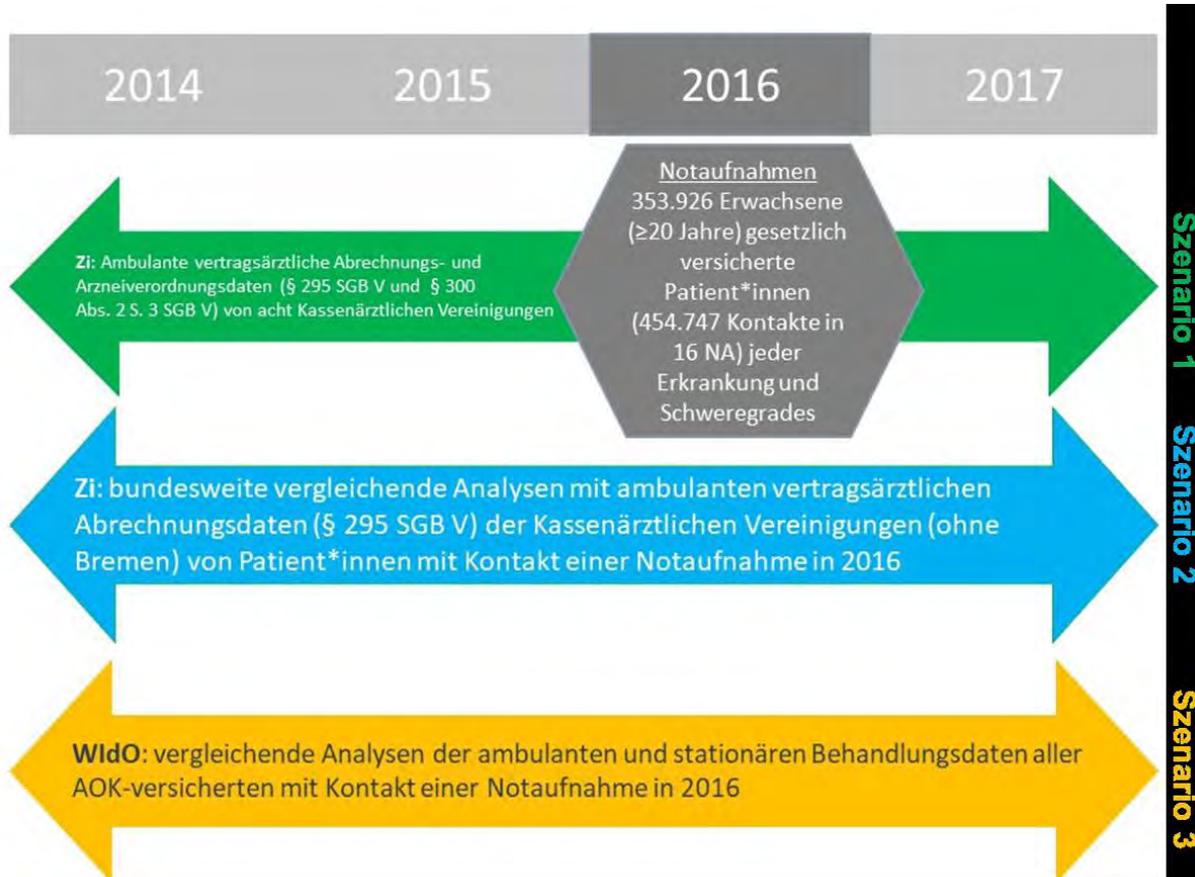
**Ziel:** überregionale, sektorenübergreifende und interdisziplinäre Versorgungsforschung im Bereich Notfall- und Akutmedizin zu ermöglichen

Drei Analyseszenarien:

- > nur Szenario 1 wird vorgestellt
- > In 2 & 3 keine Datenverknüpfung

**Partner:** Charité Berlin, Zi - Zentralinstitut für die kassenärztliche Versorgung, TMF - Technologie- und Methodenplattform für die vernetzte medizinische Forschung, Wido Wissenschaftliches Institut der AOK, Universitätsklinikum Magdeburg, TU Berlin

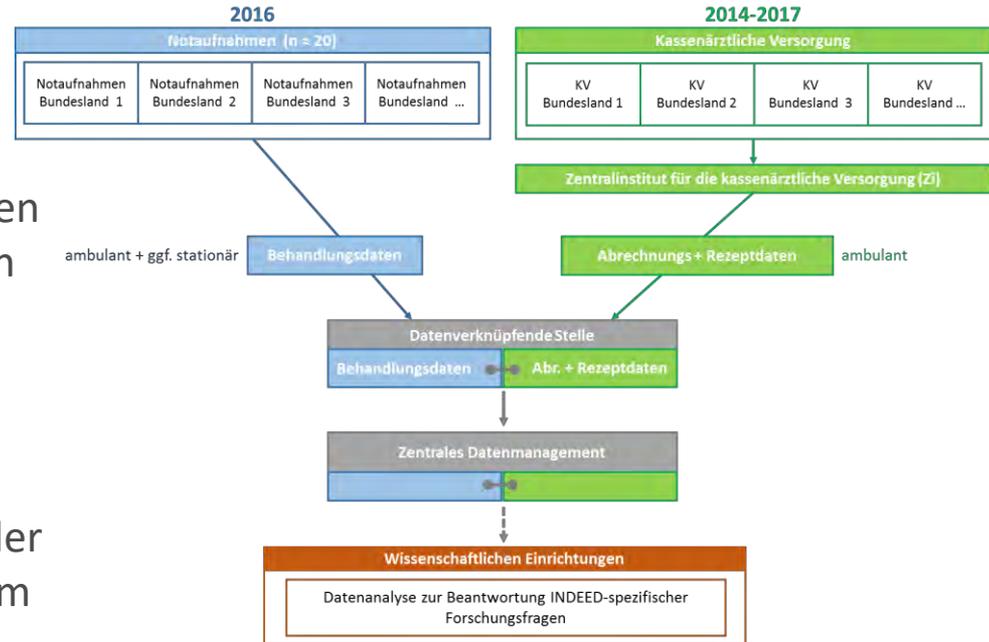
Szenarienübersicht (Folie auf GMDS nicht präsentiert)



## Abbildung „Studiendesign und Datenquellen der INDEED-Studie“

Entnommen dem öffentlichen Projektergebnisbericht:  
<https://innovationsfonds.g-ba.de/beschluesse/indeed-inanspruchnahme-und-sektoreneuebergreifende-versorgungsmuster-von-patienten-in-notfallversorgungsstrukturen-in-deutschland.157>

**Szenario 1:** für das Kalenderjahr 2016 wurden die routinemäßig erhobenen **Krankenhausnutzdaten zur Notaufnahmebehandlung** eines Patienten und zum ggf. anschließenden stationären Krankenhausaufenthalt mit den ambulanten **Abrechnungsdaten (Leistungs- und Diagnosedaten) und Arzneimittelverordnungsdaten** der kassenärztlichen Vereinigungen (KVen) der jeweiligen Bundesländer für den Zeitraum 2014 - 2017 verknüpft werden



Ergebnisbericht Anlage 2\_INDEED\_01VSF16044\_DS-Konzept\_v0.5\_gekuerzt.pdf

Verfügbar unter: <https://versorgungsforschung.charite.de/forschung/ressourcen/indeed/>

A. Fischer-Rosinsky, A. Slagman, R. King, T. Reinhold, L. Schenk, F. Greiner, D. von Stillfried, G. Zimmermann, C. Lüpkes, C. Günster, N. Baier, C. Henschke, S. Roll, T. Keil, and M. Möckel, INDEED—Utilization and Cross-Sectoral Patterns of Care for Patients Admitted to Emergency Departments in Germany: A Retrospective Cohort Study. *Frontiers in Public Health* 9 (2021).

Es wurden Daten aus 16 Notaufnahmen (NA) und 8 Kassenärztlichen Vereinigungen (KV) schlussendlich verarbeitet.

Die Daten wurden:

- > bei den Datenhaltern direkt pseudonymisiert (1. Stufe)
- > an die Vertrauensstelle OFFIS geleitet, gelinkt und mit neuem Pseudonym ausgestattet (2. Stufe)
- > an das zentrale Datenmanagement geleitet → die Auswertedaten vereinigt und bereinigt und mit neuen Auswerte-IDs
- > für verschiedene Auswertende/Fragestellungen aufbereitet



## Für die Notaufnahmedaten:

Auf Basis des Bundesdatenschutzgesetzes (BDSG), Landeskrankenhausgesetzes (speziell Berlin) LKGH Bln, da die Daten für die datenverarbeitenden Stelle als faktisch anonym gelten.

Auf eine **Einwilligung wurde verzichtet**, da das retrospektive Studiendesign sowohl für die Dauer bis zur Bereitstellung der Daten und für die Behandlung positiver als reine prospektive Studie ist.

Als **Offenbarungsbefugnis** (der ärztlichen Schweigepflicht (vgl. § 9 MBO-Ä). Die unbefugte Offenbarung von Patientendaten an Dritte stellt eine Straftat nach § 203 StGB) erlaubt gestattet § 25 Abs. 3 LKHG Bln die Übermittlung von Patientendaten an einrichtungsübergreifende Forschungsvorhaben. Die Vorschrift stellt damit eine **bereichsspezifische Rechtsgrundlage** zur Weitergabe von Patientendaten dar.

Dazu dann noch Abstimmung und Genehmigung der behördlichen Datenschutzbeauftragten (bDSB)

## Für die KV Daten:

Die Übermittlung der vertragsärztlichen Abrechnungsdaten der beteiligten KVen zu Forschungszwecken bedarf gemäß § 75 Abs. 2 SGB X der „vorherigen Genehmigung durch die oberste Bundes- oder Landesbehörde, die für den Bereich, aus dem die Daten herrühren, zuständig ist“. Entsprechende Genehmigungen werden durch die jeweiligen KVen in Zusammenarbeit mit dem Zi bei den Aufsichtsministerien eingeholt werden.

## Ergebnis:

Im Projekt durften Patienten auf gesetzlicher Grundlage ohne Einwilligung eingeschlossen und Ihre Daten gelinkt werden, was „daher umfangreiche Maßnahmen zur Pseudonymisierung auf allen Verarbeitungsebenen enthält, die zugleich mit der im Projektverlauf in Kraft getreten DSGVO kompatibel sind“  
[Öffentlicher Ergebnisbericht]

**Ausgangspunkt waren die Patienten in deutschen Notaufnahmen, direkt aus Daten des Notaufnahmeinformationssystems (EDIS)**

**Identifikatoren geplant (Hashes):**

- > Lebenslang eindeutige eGK Nummer [KVNr]
- > Name, Vorname, Geburtstag (NVG)

**Prüffelder (Text):**

- > Geschlecht, Geburtsjahr

**Identifikator zum Linkage innerhalb der EDIS:**

- > Fallnummer (Fall-Nr.)

**Datenkörper NA (EDIS)**

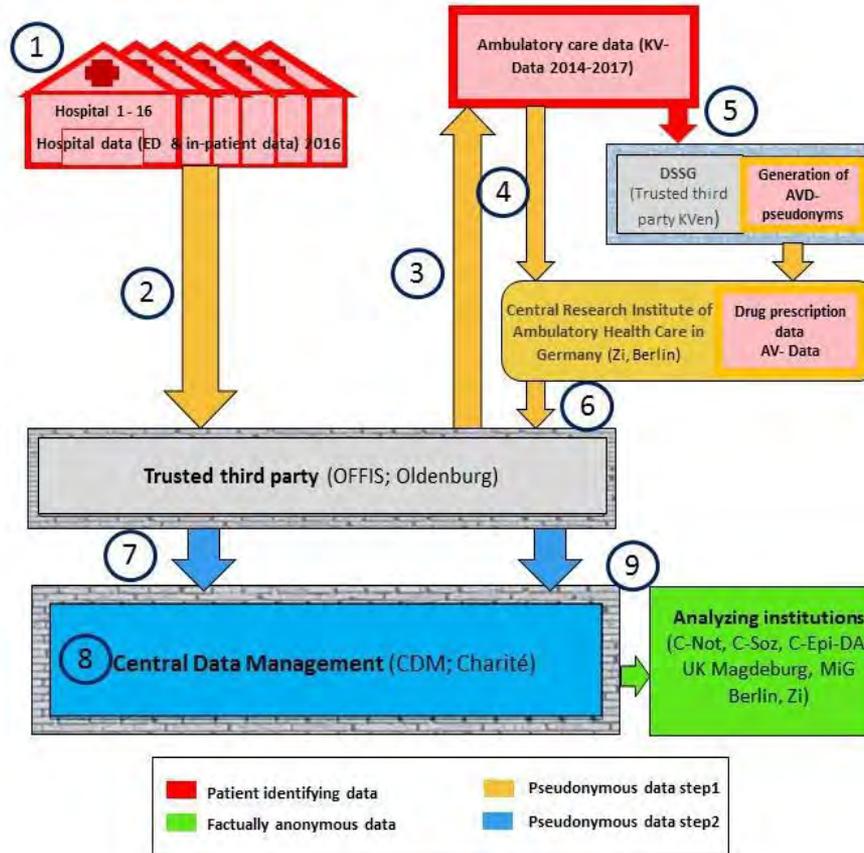
- > IDAT+MDAT in einfachen CSV-Dateien, teilweise nur über Fall-Nr. gelinkt

**KV+AV Daten (Aus KVen):**

- > IDAT+MDAT in spezifizierten CSV Dateien

# Schematische Übersicht der Datenflüsse

## Szenario 1



<https://www.frontiersin.org/articles/10.3389/fpubh.2021.616857/full>

Front. Public Health, 16 April 2021  
Sec. Disaster and Emergency Medicine  
Volume 9 - 2021 | INDEED—Utilization  
and Cross-Sectoral Patterns of Care for  
Patients Admitted to Emergency  
Departments in Germany: Rationale and  
Study Design

Je CSV: Zufällige Zeilen-ID (ZID) für spätere Zuordnung hinzufügen

Standardisierung (angelehnt an Unicon) / Prüfung eGK und NVG

Ableiten des Geburtsjahrs, Prüfen ob das Geburtsdatum das Einschlusskriterien ( $\geq 20$  Jahre) erfüllt

Zufügen eines Salt-Werts, Hashen des eGK und NVG zu P-eGK, P-NVG (Pseudonym Stufe 1)

Aufteilen der CSV in IDAT und MDAT und blockweise vertauschen der Zeilen und asymmetrisch zielgruppenorientiert verschlüsseln

Vertrauensstelle: IDAT weiterverarbeiten

1. (NA) Hashes ersetzen durch fortlaufende Integer (I-P-eGK und I-P-NVG), P-eGK und P-NVG prüfen bzw. mittels FallNr ergänzen
2. Whitelist für KV/AV bilden => nur eGK und NVG im NA-Datensatz dürfen auch bei KVen ausgeleitet werden / IPNr erzeugen (Pseudonym Stufe 2)
3. Whitelists über ZI an KVen
4. KV-Daten erhalten: Zuordnen zu bekannten P-eGK, P-NVG und IPNr
5. Zuordnungslisten (IPNr, ZIDs) ans ZDM

Bzgl. Datenschutz/Datensicherheit:  
(technisch) hervorragend aber viel zu  
Rechenintensiv (asymetrische  
Verschlüsselung je zufallssortierter Zeile  
im Speicher)

eGK ist (wenn verfügbar) für Record  
Linkage geeignet: In NA/EDIS liegt diese  
nicht immer vor, es gibt Karten-Sharing  
und manuelle Fehl-Eingaben. => eGK  
war in INDEED nicht sicher valide!

Es gab generische Platzhalter für eGK/  
NVG und Fall-Nr.

**Aufruf: Software(anforderungen) nicht  
mit „fancy Datenschutzideen“  
verkomplizieren!**

**Lösungsansatz für generische Platzhalter  
in INDEED:**

Fall-Nr. mit  $\geq 3$  verschiedenen P-eGK  
wurden als „generisch“ verworfen sowie  
P-eGK mit  $> 70$  verschiedenen Fall-Nr.

Nutzung von Whitelists bei Zi/KV war guter Ansatz, aber es gab zwei Probleme:

Eine NA hatte keine eGKs geliefert.  
NVG-Filter funktionierte in KV nicht,  
(Ursache unklar: Entweder KV Daten-  
oder Softwareproblem)

eGK/NVG Filter auf HASHes  
funktioniert nicht bei „schlechten“  
Daten (fehlende Werte, Platzhalter  
Werte)

**NA IDAT enthielten Platzhalter z.B.**

**Max Mustermann, „“**

**123456789 oder 00000000...**

**Mittels Salt und Hash wurden daraus  
gültige P-eGK / P-NVG in der KV-  
Whitelist**

**Auch KVen nutzten Platzhalter....**

- Der Zeitplan lief komplett aus dem Ruder: Die Daten für das RL sollten am 31. Mai 2018 vorliegen, RL 31.12.2018 abgeschlossen sein => Wurde 15.08.2020
- Absprachen zum Datenfluss wurden im Projekt massiv geändert / Personalwechsel

**Datenschutzvoten dauerten lange;  
Abstimmung zum Datenfluss mit den  
KVen / Zi waren langwierig**

**Schlussendlich: Die Verlinkung war laut zentralem Datenmanagement gut: Basierend auf deren Prüfungen erfolgte eine über 99% korrekte (=widerspruchsfreie) Zuordnung**

Der nun verfügbare Auswertedatensatz ist sehr wertvoll für die Gesundheitsforschung

→ aktuell [Stand 09/2023] laufen intensiv Auswertungen „kostenneutral“ (2 ½ Jahre nach Förderende!)

Datenprojekte sollten besser in zwei separate Projektförderphasen geteilt werden:

- 1) Datenaquise mit Genehmigungsverfahren und
  - 2) Datenbereitstellung und Auswertung
- => Verhindert eine Verschiebung der Fristen zu Ungunsten der Auswertung

Zu den Daten: Es gibt nichts, was es nicht gibt!