

Record Linkage in der Gesundheitsforschung – Bedarf an übergreifend eindeutigen Identifiern

WS15

Workshop „Record Linkage von unterschiedlichen
Datenarten“ | GMDS 2023 Heilbronn | 20.09.23

Dr. Knut Kaulke

TMF e.V.

Eigenschaften eines Unique Identifiers

- ▶ **Eindeutig und universell** sein (jede Person hat genau einen Identifier).
- ▶ **Keine Offenlegung persönlicher Daten** (d.h. keine persönlichen Daten enthalten)
- ▶ **Unveränderlich**
- ▶ **Verifizierbar**, um Eingabefehler zu vermeiden.

Hillestad et al. (2008)

Anwendungsfälle – Herausforderungen beim Record Linkage bezüglich Identifikatoren

- ▶ 2.1 Krebsregisterdaten und NAKO-Daten
 - ▶ Abgleich allein nach KVNR noch nicht ausreichend, da KVNR nicht für alle Patienten gegeben und in Einzelfällen fehlerhaft sein kann.

- ▶ 2.2 Notaufnahmedaten (Führend), Daten der KVen (für Personen in den Notaufnahmedaten):
 - ▶ Verfügbarkeit und Fehlerhaftigkeit/Uneindeutigkeit der eGK Nummer (KVNR)

- ▶ 3.1 Epidemiologische Primärdaten, Krankenkassendaten (GePaRD-Datenbank), Krebsregisterdaten
 - ▶ Linkage mit Kontrollnummern für Daten der Krankenversicherung nur teilweise passend
 - ▶ KVNR nicht bzw. mittlerweile nur teilweise verfügbar
 - ▶ zusätzliche Verwendung von KVNR zur Plausibilitätsprüfung nötig

- ▶ 3.2 Daten des Kinderkrebsregisters (Führend), Daten der Krankenkassen (für Personen im DKKR):
 - ▶ Einweg-Hashes aus eGK (KVNR) nicht von allen beteiligten Krankenkassen möglich

Herausforderungen bedingt durch das Fehlen eines übergreifenden Unique Identifiers

- ▶ **Eingeschränkte Nutzbarkeit der Krankenversicherungsnummer (KVNR):** Nicht universell anwendbar

- ▶ Qualität der Datenverknüpfung durch Verwendung von fehlerhaften und indirekten Identifikatoren beeinträchtigt. Diese Problematik tritt insbesondere in folgenden Szenarien auf:
 - ▶ manuelle Eingabe von Identifikatoren
 - ▶ geringe Überschneidung der vorhandenen Identifikatoren in den zu verknüpfenden Datenquellen
 - ▶ Verwendung von sehr „unpräzisen“ indirekten Identifikatoren, wie etwa Geschlecht oder Gemeindekennziffer
 - ▶ niedrige Datenqualität der Identifikatoren
 - ▶ mangelnde Standardisierung der Namensgebung, insbesondere im Kontext der phonetischen Diversität in der deutschen Bevölkerung (z.B. Kölnische Phonetik noch zeitgemäß?)
 - ▶ Verknüpfung von umfangreichen Datenquellen → Fehlerraten können bei größeren Datenmengen jedoch deutlich steigen

Herausforderungen bedingt durch das Fehlen eines übergreifenden Unique Identifiers

- ▶ **Manuelle Nachbearbeitung:** Aufwendig und zeitraubend
- ▶ **Hoher Anteil von auszuschließenden Verknüpfungen:** Führt zu Datenverlust und verzerrten Forschungsergebnissen

Aktuelle Optionen in Deutschland: SVN, KVNR, Steuer-ID



- ▶ **SVN und KVNR:** Vorteile und Nachteile in Bezug auf Universalität, Datenabdeckung und Historie.
- ▶ **Steuer-ID:** Fast universelle Abdeckung, aber bereichsfremd.
 - ▶ Aber Erweiterung der Nutzung ist im Registermodernisierungsgesetz 2021 bereits vorgesehen (zur Verbesserung, Entbürokratisierung und nutzerfreundlichen Digitalisierung von Verwaltungsleistungen)
 - Dann Zugriff von 51 verschiedene Behörden, Datenbanken und Verzeichnisse (vom Melderegister im Einwohnermeldeamt, Personenstandsregister im Standesamt und Personalausweis- und Passregister über Fahrzeugregister, Luftfahrerdatei, nationalem Waffenregister und Ausländerbehörde bis hin zu Stammsatzdateien der Renten- und Sozialversicherungen und dem Versichertenverzeichnis der Krankenkassen)

Ein neuer Ansatz: Bereichsspezifisches Forschungspseudonym

- ▶ **Bereichsspezifische Gesundheits-ID:** Ableitung einer bereichsspezifische Gesundheits-ID (auch als „Forschungspseudonym“ bezeichnet) für das Record Linkage von primären Identifikatoren → Zuordnung durch zentrale Stelle (Vertrauensstelle) verwaltet
- ▶ **Zentrale Identitätsverwalter** = Instanzen für die Vergabe und Verwaltung von solchen Identifikatoren
 - ▶ zentralen Identitätsverwalter müssen mit den Strukturen der sekundären Datennutzung kompatibel sein.
- ▶ **Implementierung:** Notwendigkeit der Integration in Dokumentations- und Abrechnungsprozesse.

Ein neuer Ansatz: Bereichsspezifisches Forschungspseudonym

- ▶ **Vereinheitlichung von Regeln** für das Record Linkage und die Pseudonymisierung sinnvoll, da enge Verzahnung der Vorgänge (↑ Linkage-Qualität und Datenschutz)
- ▶ **exaktes Record Linkage unter Verwendung eines Unique Identifiers** (d. h. des dazugehörigen umgeschlüsselten Forschungspseudonym bzw. der Gesundheits-ID)

NFDI4Health White Paper: Verbesserung des Record Linkage für die Gesundheitsforschung in Deutschland

- ▶ Für mehr Details:



<https://doi.org/10.4126/FRL01-006461895>

Vielen Dank für Ihre Aufmerksamkeit!

Weitere Informationen unter www.tmf-ev.de

Kontakt

Geschäftsstelle TMF e.V.
Charlottenstraße 42/Dorotheenstraße
10117 Berlin

+49 (30) 22 00 24 70
info@tmf-ev.de
www.tmf-ev.de | [@TMF_eV](https://www.instagram.com/TMF_eV)