

Einführung in das Themengebiet

WS15

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

Dr. Knut Kaulke

TMF e.V.

TMF-Workshops rund um Record Linkage bei GMDS-Jahrestagungen

- 1. Vergleich elektronischer Treuhänderdienste in der medizinischen Forschung** (THS Greifswald, GECKO (HS Heilbronn), PMV forschungsgruppe (Köln), Schütze Consulting AG)
 - ▶ 63. Jahrestagung der GMDS | Osnabrück | 5. Sep 2018
- 2. Standortübergreifendes Record Linkage mit und ohne Treuhandservice** (THS Greifswald, HiGHmed, DKFZ, Universität Leipzig)
 - ▶ 64. GMDS-Jahrestagung | Dortmund | 08. - 11. September 2019
- 3. Record Linkage von unterschiedlichen Datenarten**
 - ▶ 68. GMDS-Jahrestagung | Heilbronn | 17. - 21. September 2023

Programm

- 8:15** **TOP 1 Einführung in das Themengebiet** *Kaulke*
- TOP 2 Vorträge zum Record Linkage unterschiedlicher Datenarten Teil 1**
- 8:25** ▶ TOP 2.1 Krebsregisterdaten und NAKO-Daten *Hartz*
- 8:45** ▶ TOP 2.2 Notaufnahmeroutinedaten und Daten der kassenärztlichen Vereinigung *Lüpkes*
- 09:05** **Kaffeepause**
- TOP 3 Vorträge zum Record Linkage unterschiedlicher Datenarten Teil 2**
- 09:15** ▶ TOP 3.1 Epidemiologische Primärdaten und Krebsregisterdaten *Intemann*
- 09:35** ▶ TOP 3.2 Kinderkrebsregisterdaten und Routinedaten der GKV *Ihle*

Programm

TOP 4 Übergreifende Themen

09:55 ▶ TOP 4.1 Rechtliche Voraussetzungen und Herausforderungen *Steiner*

10:10 ▶ TOP 4.2 Bedarf an übergreifend eindeutigen Identifiern *Kaulke*

TOP 5 Diskussion & Zusammenfassung

10:25 ▶ TOP 5.1 Diskussion, Zusammenfassung und Handlungsbedarf *Drepper, Alle*

10:45 Ende

- ▶ zunehmende elektronische Erfassung von Gesundheitsdaten
→ hoher Informationsgehalt der Sekundärdaten für Forschung
- ▶ Einschränkungen der benötigten Daten für Gesundheitsforschung (Primärdaten und/oder Sekundärdaten)
 - ▶ für bestimmte Zwecke oder Forschungsfragen
 - ▶ Fehlen von relevanten Informationen zur Beantwortung der Fragestellung
- ▶ Verknüpfung dieser verschiedenen Daten(arten) ermöglicht Informationsgewinn, ohne neue Datenerhebung

Verfahren und Arten des Record Linkage

Exaktes Linkage

- ▶ nur bei exakter Übereinstimmung eines eindeutigen Verknüpfungsschlüssels (z. B. Sozialversicherungsnummer) oder mehrerer Linkage-Variablen zusammen (Match–Merge Linkage)
- ▶ bei Fehlern oder unterschiedlichen Schreibweisen → keine Zusammenführung

Fehlertolerantes Linkage

- ▶ Wenn kein exaktes Datenlinkage möglich
- ▶ Arten: regelbasiertes, distanzbasiertes und probabilistisches Linkage

Verfahren und Arten des Record Linkage

Regelbasiertes Linkage

- ▶ Regeln definieren, welche Identifikatoren komplett und welche nur teilweise übereinstimmen müssen

Probabilistisches Linkage

- ▶ Übereinstimmung mancher Identifikatoren mehr Aussagekraft hinsichtlich der Zusammengehörigkeit als andere; z.B. Wahrscheinlichkeitsregeln bei Nachnamenübereinstimmung und Geschlechtsübereinstimmung

Distanzbasiertes Linkage

- ▶ Berechnung der Ähnlichkeit der Merkmalsausprägungen einzelner Identifikatoren durch String-Metriken (Stringähnlichkeitsfunktionen)

Verfahren und Arten des Record Linkage

Linkage mit direkt und indirekten Identifikatoren

- ▶ Direkt = z.B. über Sozialversicherungsnummer
- ▶ Indirekt = Bestimmung von Linkage-Variablen, die in Kombination zum Linkage führen

Linkage mit Klartextangaben und mit verschlüsselten Identifikatoren

- ▶ Einsatz von Kontrollnummern oder Bloomfiltern bei Identifikatoren, die nicht im Klartext stehen
 - ▶ Fehlertolerantes Linkage auf verschlüsselten Daten möglich (PPRL)

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)