



Record Linkage

Kees Kleihues - van Tol

8. Qualitätskonferenz 2020

- 11 Entitäten
- 28 Register
- 2.053.734 Datensätze



8. Qualitätskonferenz 2020

Neu: Einheitliche Datenmatrix für alle Entitäten

Zusammenführung in Zentraler Datenbank

- Für alle Entitäten
- Für alle Register

Vorteile:

- Zentrale Bereinigung
- Zentrale Plausibilitätskontrolle
- Zentrale Dublettenkontrolle

Bereinigung:

calc_ Felder

Alle Datum-Felder gültiges Datum im Format TT.MM.JJJJ

pTNM_T

- pT1 -> 1
- T1 -> 1
- 1 -> 1

Datenbank mit c. 2000 Regeln für Bereinigung

2.053.734 Datensätze:

Geburtsdatum:

- 1.234.124 Monatsgenau
- 819.610 Tagesgenau

Diagnosedatum:

- 650.488 Monatsgenau
- 1.403.246 Tagesgenau

Personenhash

- 1.307.428

Patienten Hash:

Hash aus Geburtsdatum, Geschlecht und ersten 5 Buchstaben des
Vornamens und ersten 5 Buchstaben des Nachnamens mit Großbuchstaben.
`md5(ddmmyyyyGVVVVNNNN)`

Kontrollgruppe:

- Kolorektales Karzinom (C18-C21)
- Register mit Hash

Dubletten:

- Gleiches Diagnosejahr
- Gleicher Hash

1. Kontrolle auf Dubletten: 652 aus 270.041

Ausschluss von Registern ohne Dubletten (Hash nicht gleich erzeugt?)

2. Kontrolle auf Dubletten 652 aus 174.568

652 Dubletten:

- 652 Geburtsdatum gleich (Monatsgenau)
- 528 Diagnosedatum gleich (Monatsgenau)
- 555 Histologie gleich
- 562 Histologie gleich (ersten 2 Ziffern)
- 273 PLZ gleich
- 279 PLZ gleich (ersten 2 Ziffern)
- 327 PLZ gleich (erste Ziffer)

Geburtsdatum gleich (Monatsgenau)
Diagnosedatum gleich (Monatsgenau)

+Histologie gleich
451 true positive
67.913 false positive

+PLZ gleich
250 true positive
2.750 false positive

+Histologie gleich
+PLZ gleich
241 true positive
1344 false positive

Mit Hash letzten 2 Symbole:
Geburtsdatum gleich (Monatsgenau)
Diagnosedatum gleich (Monatsgenau)

+Histologie gleich
451 true positive
67.913->256 false positive

+PLZ gleich
250 true positive
2.750->12 false positive

+Histologie gleich
+PLZ gleich
241 true positive
1.344->6 false positive

Mit Hash letzten 4 Symbole:
Geburtsdatum gleich (Monatsgenau)
Diagnosedatum gleich (Monatsgenau)

+Histologie gleich
451 true positive
67.913->6 false positive

+PLZ gleich
250 true positive
2.750->0 false positive

+Histologie gleich
+PLZ gleich
241 true positive
1.344->0 false positive

Geburtsdatum gleich (Monatsgenau)
Diagnosedatum gleich (Monatsgenau)
528 true positive
128.695 false positive

+Hash 2 Symbole
528 true positive
489 false positive

+Hash 4 Symbole
528 true positive
7 false positive



Vielen Dank