International Journal of Population Data Science

Journal Website: www.ijpds.org

Comparing Record Linkage methods for real-world perinatal and neonatal data without unique identifiers

Schnell, R¹* and Borgs, C¹

¹University of Duisburg-Essen

Background

Data on newborns is regularly linked for epidemiological research. However, hospital data often suffers from incomplete data. We report on a linkage of two population-covering administrative health databases containing neonatal and perinatal data without unique personal identifiers and with incomplete information in standard patient identifiers.

Goal

To study the effects of a policy-induced change from linking a national database without standard patient identifiers to a privacy-preserving Record Linkage method, we compare the linkage system in use to clear-text and privacy-preserving Record Linkage techniques. We expected large proportions of missing identifiers since they are not needed for clinical practice. Therefore, we expected missing links caused by missing identifiers. To study the impact of these missing identifiers on these successful links, we compared several linkage methods. Furthermore, we study the variations of linkage success between hospitals.

Methods

Perinatal and neonatal data from population-covering real-world administrative databases was linked using several variants of state of the art methods, including Privacy-preserving Record Linkage (PPRL) techniques such as multiple match keys and Bloom filter methods.

Results

We report on the variation of linkage results between the hospitals and give possible explanations for the differences.

The resulting linkage success is reported for each method. The impact of incomplete data on linkage success for each method is documented.

Finally, we report on the relative performance of the modified techniques compared to standard linkage procedures used in practice.

Conclusion

Implementing a record linkage system based on identifiers not required for clinical practice caused a large number of missing identifiers. Since this information is essential for successful clear-text and private linkage methods, emphasizing the need for documenting patient identifiers, especially in cases where auxiliary information (such as stable addresses, date of birth or health insurance numbers) are missing, is of central importance for implementing a privacy-preserving Record Linkage system.

*Corresponding Author:

Email Address: rainer.schnell@uni-due.de (R Schnell)