Proposed abstract submission for

- *ISPOR Orlando 2020*

Improving Accuracy of Record Linkage using Graph Structures: Relevance for Health Outcomes Research?

Niek IJzerman[1,2], Pauline Lin[1], Maarten IJzerman[3,4], Uwe Aickelin[1]

1. University of Melbourne, School of Information Sciences, Parkville, Australia
2. University of Amsterdam, Faculty of Science, Amsterdam, the Netherlands
3. University of Melbourne, Centre for Health Policy, Parkville, Australia
4. Peter MacCallum Cancer Centre, Parkville, Australia

Objectives

The use of administrative and claims data for health outcomes research has greatly increased in recent years. However, appropriate use for health services research requires multiple data sources to be linked, known as record linkage. Multiple methods for record linkage have been developed and tested, mostly using probabilistic (PRL) and deterministic linkage (DRL). Both DRL and PRL use attribute (e.g. name, year, ID) information to link records. This simulation study evaluates if record linkage can be improved by combining attribute with structural information that is being obtained by representing records as nodes in network graphs.

Methods

The simulation study is performed using Python software with Jellyfish library add-on. Data used to simulate different scenarios was obtained from two publicly available set of academic publications with 2,294 and 2,616 records obtained from the Leipzig Database Group. Simulated scenarios were based on the change in accuracy following including attribute information on top of the graphical information from the REGAL algorithm. % accuracy was calculated as the ratio between the true positive linkages to the total number of linkages.

Results

By using structural information solely, the % of accurate linkages ranged from 8% to 39%. By adding attribute information on top of structural information, the % correct linkages ranged from 56% to 63%, partly attributed to improvements due chance. Revising and adjusting the network alignment algorithm could account for more accurate linkages.

Conclusion

It is concluded that adding attribute information on top of structural information within the REGAL framework has been shown to positively influence accuracy of record linkage. There is, however, room to further improve the alignment process, in particular to improve the accuracy for node pairs that have differences in connectedness across graphs. While the number of correct linkages remains relative after using other linkage algorithms, the implications for health outcomes research will be discussed.

Key words: Information systems, record linkage, graphical networks, health services research

Words: 300

Author/s:
IJzerman, N; Lin, P; IJzerman, M; Aickelin, U

Title:
IMPROVING ACCURACY OF RECORD LINKAGE USING GRAPH STRUCTURES: RELEVANCE FOR HEALTH OUTCOMES RESEARCH?

Date:
2020-05-01

Citation:
IJzerman, N., Lin, P., IJzerman, M. & Aickelin, U. (2020). IMPROVING ACCURACY OF RECORD LINKAGE USING GRAPH STRUCTURES: RELEVANCE FOR HEALTH OUTCOMES RESEARCH?. VALUE IN HEALTH, 23, pp.S323-S323. ELSEVIER SCIENCE INC. https://doi.org/10.1016/j.jval.2020.04.1206.

Persistent Link:
http://hdl.handle.net/11343/258841

File Description:
Accepted version