International Journal of Population Data Science

Journal Website: www.ijpds.org

How do you measure up? Methods to assess linkage quality

Ferrante, Anna^{1*}, Boyd, James¹, Randall, Sean¹, Brown, Adrian¹, and Semmens, James¹

¹The University of Western Australia

Objective

Record linkage is a powerful technique which transforms discrete episode data into longitudinal person-based records. These records enable the construction and analysis of complex pathways of health and disease progression, and service use. Achieving high linkage quality is essential for ensuring the quality and integrity of research based on linked data. The methods used to assess linkage quality will depend on the volume and characteristics of the datasets involved, the processes used for linkage and the additional information available for quality assessment. This paper proposes and evaluates two methods to routinely assess linkage quality.

Approach

Linkage units currently use a range of methods to measure, monitor and improve linkage quality; however, no common approach or standards exist. There is an urgent need to develop “best practices” in evaluating, reporting and benchmarking linkage quality. In assessing linkage quality, of primary interest is in knowing the number of true matches and non-matches identified as links and non-links. Any misclassification of matches within these groups introduces linkage errors. We present efforts to develop sharable methods to measure linkage quality in Australia. This includes a sampling-based method to estimate both precision (accuracy) and recall (sensitivity) following record linkage and a benchmarking method - a transparent and transportable methodology to benchmark the quality of linkages across different operational environments.

Results

The sampling-based method achieved estimates of linkage quality that were very close to actual linkage quality metrics. This method presents as a feasible means of accurately estimating matching quality and refining linkages in population level linkage studies.

The benchmarking method provides a systematic approach to estimating linkage quality with a set of open and shareable datasets and a set of well-defined, established performance metrics. The method provides an opportunity to benchmark the linkage quality of different record linkage operations. Both methods have the potential to assess the inter-rater reliability of clerical reviews.

Conclusion

Both methods produce reliable estimates of linkage quality enabling the exchange of information within and between linkage communities. It is important that researchers can assess risk in studies using record linkage techniques. Understanding the impact of linkage quality on research outputs highlights a need for standard methods to routinely measure linkage quality. These two methods provide a good start to the quality process, but it is important to identify standards and good practices in all parts of the linkage process (pre-processing, standardising activities, linkage, grouping and extracting).

*Corresponding Author:

Email Address: a.ferrante@curtin.edu.au (A. Ferrante)