International Journal of Population Data Science

Journal Website: www.ijpds.org

Assessing the impact of different grouping methods: time to rethink and regroup?

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

¹Curtin University

Objective

The grouping of record-pairs to determine which administrative records belong to the same individual is an important process in record linkage. A variety of grouping methods are used but the relative benefits of each are unknown. We evaluate a number of grouping methods against the traditional merge based clustering approach using large scale administrative data.

linkage quality at far lower cost than current methods for improving linkage quality. By improving linkage quality at low cost, this resource can be further utilised by health researchers.

Approach

The grouping of record-pairs to determine which administrative records belong to the same individual is an important process in record linkage. A variety of grouping methods are used but the relative benefits of each are unknown. We evaluate a number of grouping methods against the traditional merge based clustering approach using large scale administrative data.

Results

Results suggest alternate grouping methods will yield large benefits in linkage quality, especially when the quality of the underlying repository is high. Stepwise grouping methods were clearly superior for one-to-one linkage. There appeared little difference in linkage quality between many-to-one grouping approaches. The most appropriate techniques for ongoing linkage depended on the quality of the population spine and the underlying dataset.

Conclusion

These results demonstrate the large effect that the choice of grouping strategy can have on overall linkage quality. Ongoing linkages to high quality population spines provide large improvements in linkage quality compared to merge based linkages. Procuring or developing such a population spine will provide high

*Corresponding Author:

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