International Journal of Population Data Science

Journal Website: www.ijpds.org

The impact of administrative dataset characteristics on clerical review: A Tasmanian case study

Wiggins, Nadine^{1*} and Stokes, Brian¹

¹Menzies Institute for Medical Research, University of Tasmania

Objectives

The Tasmanian Data Linkage Unit (TDLU) routinely collects detailed statistical data specific to time spent preparing datasets for linkage, performing clerical review and quality assurance for all datasets added to its Master Linkage Map (MLM). This information is used for a range of functions including resource planning, performance monitoring, preparation of data linkage quotes, data quality reporting and identifying potential process improvements.

Approach

Detailed metadata for each dataset imported to the MLM is collected to determine likely linkage quality and to estimate clerical review resource requirements. Attributes including the time period covered, data completeness, data quality and accuracy, the population characteristics and an understanding of how the data was collected are documented. In addition to this metadata, clerical reviewer staff record the number of record pairs and groups reviewed within confidence ranges and the time taken to complete review.

Results

The data collected provides strong evidence that administrative datasets with more problematic characteristics take longer to prepare, clerically review and quality assure. Tasmanian public hospital emergency presentations and admitted patient datasets, which are high in quality and completeness, have required less clerical review (8% and 6% respectively of total records) compared with perinatal baby (30% of total records clerically reviewed and significant additional time spent on quality assurance) and a small cohort of ambulance patients where 78% of all records were manually reviewed. The ambulance and perinatal baby datasets contained incomplete name, address and hospital

identifier fields, and spelling and data errors were evident in the ambulance dataset as a result of the data capture method (over the phone with critical time constraints).

Conclusion

The underlying dataset quality and completeness impacts to a significant degree the time taken to clerically review a dataset, and in turn the quality of links made in a MLM. Additionally, each new dataset added to the MLM introduces an additional layer of complexity necessitating additional processing time due to the number of record comparisons made. Having an agreed set of formal processes to prepare administrative data prior to data linkage, and understanding how source data is collected and processed, is crucial to obtaining high levels of quality within the MLM.

*Corresponding Author:

Email Address: nadine.wiggins@utas.edu.au (N. Wiggins)