International Journal of Population Data Science

Journal Website: www.ijpds.org

Real-time linkage: when is near enough good enough?

Irvine, Katrina^{1*}, Williamson, Jennifer², and Pye, Victoria¹

¹Centre for Health Record Linkage, New South Wales Health

²New South Wales Ministry of Health

Objectives

Timely data on the health and vital status of study participants is important for cohort management and patient recruitment processes in addition to traditional observational research. We have investigated the potential to modify data sourcing and linkage processes so that data is linked daily, in near real-time. Such modifications will improve the currency of linked data for end users but impact on the quality of the data for recent periods. This work aims to provide a conceptual framework for understanding drivers of quality and empirical evidence on the type and extent of quality impacts associated with linking death and recent hospitalisation data in near real-time for a large population.

Approach

We compared different linkage strategies including real-time linkage to study the type and extent of quality impacts observed using a large population cohort linked to jurisdictional level hospitalisation and death data.

Results

Changes in data quality arise from several sources: incomplete enumeration of the most recently occurring events (known as delayed notifications), differences in the availability of up-to-date personal information to assist in data linkage, and the adoption of faster probabilistic linkage techniques that skip some computational and/or manual steps (e.g. clustering and clerical review) that are performed in conventional gold-standard probabilistic linkage processes.

Compared to a gold standard process faster linkage techniques with reduced personal information produced quality metrics of precision, recall and F measure in excess of 98.0 for linkage of a population cohort with death registrations. While precision

was maintained at 98.0 for linkage to hospitalisation data recall was reduced to 91.1.

Delayed notifications have a pronounced effect on quality. If incomplete enumeration is factored into quality metrics, the effective recall for the last quarter of data drops to 0.82 and 0.33 for deaths and most recent hospitalisation respectively.

Conclusion

High quality and current information about health and vital status can be obtained through data linkage, but there are trade-offs between timeliness and ascertainment. Delayed notifications in jurisdictional level administrative data systems appear more problematic than subtle changes in linkage technique. The quality impacts of near-real time linkage vary significantly by data collection and so its usefulness is highly dependent on its desired purpose and quality requirements.

*Corresponding Author:

Email Address: katie.irvine@moh.health.nsw.gov.au (K. Irvine)