International Journal of Population Data Science

Journal Website: www.ijpds.org

Understanding the coverage of Statistics New Zealand's Integrated Data Infrastructure

Minshall, Gareth^{1*}, Gomez, Antony¹, and Christine, Bycroft¹

¹Statistics New Zealand

Objectives

Statistics New Zealand's Integrated Data Infrastructure (IDI) combines information from a range of government agencies (such as tax, health and education data) in order to provide the insights government needs to improve social and economic outcomes for New Zealanders. New Zealand has no national population register or unique identifier used in common across these multiple data sources, and probabilistic linkages are a feature of the IDI. A challenge for researchers is to understand the impact of linkage errors and coverage issues present in the linked data, and to develop the rules necessary to define their target population. We outline the statistical infrastructure Statistics New Zealand is developing to help researchers navigate these issues.

Approach

A method has been developed to identify NZ residents at a given time from the much larger number of individuals present in the IDI. Census data linked to the IDI offers insight into the coverage of key population groups and the quality of the attribute information held in the IDI (e.g. location and ethnicity). We are assessing ways that Statistics New Zealand could use these findings to assist researchers in forming their population of interest and assess the potential for bias.

Results

The derived administrative resident population is compared with the official population figures and patterns of under- and over-coverage are identified at an aggregate, and individual level. Some coverage discrepancies may be improved through reducing linkage errors. Comparison with census data reveals some significant quality issues with location and ethnicity variables in administrative collections. Work is underway to improve meth-

ods for combining information from multiple sources of varying quality.

Conclusion

Identifying NZ residents at a given time, and quantifying errors in administrative data sources will assist researchers ability to recognise and adjust for these errors in their analysis. Simply quantifying (often for the first time) the limitations of administrative sources also provides impetus to improving the collection of these variables at source.

*Corresponding Author:

Email Address: gareth.minshall@stats.govt.nz (G. Minshall)