International Journal of Population Data Science

Journal Website: www.ijpds.org

Data Integration for Assessing Census Coverage

Cloutier, Sylvain^{1*}

¹Statistique Canada

Objective

Statistics Canada initiated the Canadian Statistical Demographic Database (CSDD) research project to determine if and how administrative data could be used to support the Canadian Census Program. The project's goal is to create a census spine from administrative data sources. The CSDD's current scope is limited to basic information (name, sex, birth date and usual place of residence) for all Canadians.

Method

Two 2011 CSDD prototypes were built using and linking hundreds of administrative files obtained mainly from other federal departments. Extensive pre-processing activities must take place prior to linkage to remove duplicates and standardize file variables. Given that Canadians do not possess a single unique identifier, administrative files were linked using record linkage methods; key matching variables were identified, validated and used to perform the linkage. This work led to the development of auxiliary files, which serve specific purposes related to the CSDD development. They also provide useful linkage keys to other Statistics Canada statistical programs.

Results

The outcome of the CSDD is determined by comparing it to two references. First, comparisons were done at the aggregate level (Canadian, provincial and sub-provincial levels) by contrasting the results with Demography Division's official population estimates for the 2011 Census. The CSDD was also compared with the 2011 Census of Population's Response Database (RDB), which allows for analysis at the micro (record) level. The RDB contains non-imputed data on name, sex, birth date and usual place of residence as provided by individual census respondents. Comparisons with the RDB have allowed us to address the question, "Does the CSDD put the right person at the same address

as the 2011 Census does?"

Results are promising. At the aggregate level, the CSDD compares well with the demographic estimates for the 2011 Census at the national, provincial/territorial and some urban area levels. At the micro level, the CSDD contains more individuals than the RDB. Improvements are needed with regards to its ability to place persons accurately in rural areas due to the lack of good residential addresses in administrative data files. Initial results led to the planning of new CSDD prototypes, this time for 2016, in line with the 2016 Census of Population.

Conclusion

The presentation will give an overview of the methods and principles behind the construction of the CSDD. Basic analytical results will present areas of strength and weakness. Lessons learned and upcoming challenges along with their proposed solutions will complete the presentation.

*Corresponding Author:

Email Address: sylvain.cloutier@canada.ca (S. Cloutier)