International Journal of Population Data Science

Journal Website: www.ijpds.org

Understanding variations in the trajectories of academic achievement and mental health service utilization of foreign-born adolescents in British Columbia, Canada: A population-based cohort study

Trudeau, Richard¹

¹Statistics Canada

Objectives

The Social Data Linkage Environment (SDLE) at Statistics Canada promotes the innovative use of existing administrative and survey data to address important research questions and inform socio-economic policy through record linkage. It expands the potential of data integration across multiple domains, such as health, justice, education and income, through the creation of linked analytical data files without the need to collect additional data from Canadians.

Approach

At the core of the SDLE is a Derived Record Depository (DRD), essentially a national dynamic relational data base containing only basic personal identifiers. The DRD is created by linking selected Statistics Canada source index files for the purpose of producing a list of unique individuals. These files are brought into the environment, processed and linked only once to the DRD. Each individual in the DRD is assigned an SDLE identifier. Some of the source index files used to build the DRD include tax records, vital statistics registration records (births and deaths), and immigrant data. Updates to these data files are linked to the DRD on an ongoing basis. Only basic personal identifiers are stored in the DRD. Examples of personal identifiers stored in the DRD include surnames, given names, date of birth, sex, insurance numbers, parents' names, marital status, addresses (including postal codes), telephone numbers, immigration date, emigration date and date of death. The paired SDLE identifiers and source index file record IDs resulting from the record linkage are stored in a Key Registry. To reduce the risk of privacy intrusiveness and to minimize the risk of disclosure, source files are separated into source index files and source data files. Employees performing the record linkages in SDLE have access to only the basic personal identifiers needed for link-

age. Employees who build the analytical files for research have access only to the data stripped of personal identifiers.

Results

The SDLE is a highly secure environment that facilitates the creation of linked population data files for social analysis. It is not a large integrated data base.

Conclusions

The SDLE program facilitates pan-Canadian social and economic statistical research. It is a record linkage environment that: increases the relevance of existing surveys without collecting new data; substantially increases the use of administrative data; generates new information without additional data collection; maintains the highest privacy and data security standards; and promotes a standardized approach to record linkage processes and methods.

*Corresponding Author:

Email Address: richard.trudeau@canada.ca (R. Trudeau)