

# Technical Documentation

## BC Demographic Survey: DIP Data Linkage Rates

### Table of Contents

Purpose .....	2
Project Overview .....	2
Methodology .....	2
Overview .....	2
Caveats .....	3
Software .....	3
GitLab .....	3
References .....	4

**PROJECT:** Evaluating the BC Demographic Survey Data in the DIP

**PREPARED BY:**

Amelia Lowery, Julie Hawkins & Lindsay Fredrick  
Data Science Partnerships Program  
BC Stats BC Data Systems  
Ministry of Citizens' Services

**DATE:** 2024-05-13

## Purpose

This document summarizes the analytical methods, caveats, and technical software used to generate the linkage summaries provided in the dashboard [BC Demographic Survey: DIP Data Linkage Rates](#) and made available on the [BC Data Catalog](#)

## Project Overview

In 2023, BC Stats conducted the BC Demographic Survey. More than 200,000 people responded to the voluntary survey, providing information about many aspects of their identity (such as race, ethnicity, ancestry, gender and many others).

The Data Innovation Program (DIP) securely links and de-identifies data from multiple ministries, organizations or agencies in a secure platform. Many DIP datasets also contain demographic related information. However, the number of datasets with demographic information is limited, and are only partially complete pictures. The BC Demographic Survey aims to improve our understanding of how people with varying backgrounds interact with public services by broadening the scope of available demographic information to each data (and therefore service) provider.

The analysis presented in this dashboard used the secure platform to access available datasets from the DIP, and linked these datasets, where possible, to the data from the BC Demographic Survey. Overall linkage rates were investigated. The methods, tools, and caveats associated with the dashboard are explored more fully below.

## Methodology

### Overview

To compare the BC Demographic Survey to each individual DIP dataset, a list of unique StudyIDs (which represent unique individuals as determined by PopData) was created for each individual dataset. The BC Demographic Survey list was compared to the individual DIP dataset, to determine:

- The number of individuals in each DIP dataset.
- The number of individuals within a DIP dataset that have a survey record.

## Caveats

**This dashboard and data source serves as a guide only. Researchers are ultimately responsible for determining the feasibility and reliability of the variables themselves prior to using the BC Demographic Survey variable information for their project.**

Here we list some important distinctions to keep in mind when browsing the dashboard and related datasets:

- This dashboard was produced in 2024. Linkage rates will change overtime.
- Not every DIP dataset has a StudyID.
- Some DIP datasets contain multiple StudyIDs.
- All counts provided here are unweighted. DIP researchers are responsible for providing their own weights where necessary.
- To mitigate the privacy risks of re-identification and residual disclosure, some small counts (and associated percentages) have been masked.
- Rounding has been performed on all percentages.

## Software

This analysis is implemented in the R and python programming languages Van Rossum and Drake Jr (1995). The code used to generate this analysis was reviewed by three data scientists. Key tools used to complete this work include the Apache Arrow project (Richardson et al. 2021), the tidyverse (Wickham et al. 2019) and the internal SAE package dippy (Fredrick 2023).

## GitLab

All code is stored under the [git version control](#) system and shared inside the secure environment in these GitLab repos:

- Creation of Demographic Survey Variables:  
<https://projectsc.popdata.bc.ca/shares/arda-demographic-survey>
- Creation of Linkage Statistics:  
<https://projectsc.popdata.bc.ca/shares/linkage-litmus-test>

## References

Fredrick, Lindsay. 2023. *Dippy: Provide Functions to Efficiently Import SRE Data*.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Richardson, Neal, Ian Cook, Jonathan Keane, Romain François, Jeroen Ooms, and Apache Arrow. 2021. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.

Van Rossum, Guido, and Fred L Drake Jr. 1995. *Python Tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.