

Identifying the New Zealand resident population in the IDI

The choice of a study population in any given project in the IDI will depend on the research question and the study design. Many studies will require a resident population, such as that estimated on census night or [published by SNZ](#). We summarise some options for identifying a residential population in the IDI. Every method has limitations and an in-depth understanding of the data will support better interpretation of study results.

Estimated Residential Population

Datalab applications can now request access to an IDI table for the estimated New Zealand resident population (the IDI-ERP). The IDI-ERP is estimated at 30 June for the years 2007-2014. (Future iterations may include estimates for other dates as well such as 31 December for additional flexibility.) Anonymised individual records are selected when there is evidence that an individual has been active in health, ACC, taxation and education datasets in the year leading up to the date of the estimate, and when the record is linked in the IDI spine. Individuals are censored if they died before the reference date or if they had moved overseas (that is they were overseas for at least 6 out of 12 months spanning the six months either side of the reference date). For children under five years old, having a NZ birth registration or a visa approval (excluding visitor or transit visas) before the reference date is sufficient for inclusion in the population. The IDI-ERP is 2% larger than the official population estimate. Further work is needed to understand the causes of undercoverage or overcoverage in some population groups (eg overcoverage in young men in the IDI-ERP compared to the census). For example, [a VHIN Catalyst project](#) used a version of the IDI-ERP for identifying a denominator population in cardiovascular disease research.

More information on the methods is available from SNZ:

Webpage: [Experimental population estimates from linked administrative data: methods and results](#)

Census Transformation, Statistics New Zealand. Experimental population estimates from linked administrative data 2017 release. Retrieved from <http://innovation.stats.govt.nz/assets/experimental-population-estimates-from-linked-admin-data-2017-release.pdf>

Gibb S, Bycroft C, Matheson-Dunning N (2016). Identifying the New Zealand resident population in the Integrated Data Infrastructure (IDI). Retrieved from <http://archive.stats.govt.nz/methods/research-papers/topss/identifying-nz-resident-pop-in-idi/methods.aspx>

Gibb S & Shrosbree E (2014). Evaluating the potential of linked data sources for population estimates: The Integrated Data Infrastructure as an example. <http://archive.stats.govt.nz/methods/research-papers/topss/evaluating-potential-linked-data-sources.aspx>

Bespoke Residential Estimated Population

There will be projects where you require a population denominator for a particular date outside of the available IDI-ERPs that all relate (currently) to the 30 June. On the SNZ Wiki and Meetadata there is SAS code that can be adapted to identify the IDI-ERP for any given date. For example, the VHIN Healthier Lives Earthquake project used a bespoke IDI-ERP to estimate a resident population at the start of the Canterbury Earthquake sequence in 2010 (See [this page](#) and [code here](#)).

Code is also available on the IDI Wiki, at Meetadata and at these links;

- http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-resources/examples-idi-code.aspx
- <https://github.com/StatisticsNZ/IDI/blob/IDI-Code/VHIN.zip>

By Andrea Teng, June Atkinson

Version: Original 10 March 2017