IDI Spine

What is the spine?

For any analysis of the collection of datasets in the IDI, it is important to understand the function of the IDI spine. All datasets in the IDI are linked through the spine. For example microdata on health and education are each linked to the spine, and thus health and education records can only be linked together for individuals that have a record in the spine. The current spine is called the 'prototype spine' and is a list of individuals made by combining individuals with tax (from 1999), births (from 1920) or visa information for specific visa types (from 1997). Visa types exclude short-term visitors but include people with visas that allow people to reside, work, or study in New Zealand. The spine aims to record everyone who has ever been a resident in New Zealand (the target population). The following papers provide more information about how the spine was created.

More information about how the spine was created:

- Black, A (2016) *The IDI prototype spine's creation and coverage.* 2016 (Statistics New Zealand Working Paper No 16–03). Retrieved from: http://archive.stats.govt.nz/methods/research-papers/working-papers-original/idi-prototype-spine.aspx

- Statistics New Zealand (2015) *Quality and the Integrated Data Infrastructure prototype spine.* 2015. Retrieved from: http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-resources/quality-and-idi-proto-spine.aspx
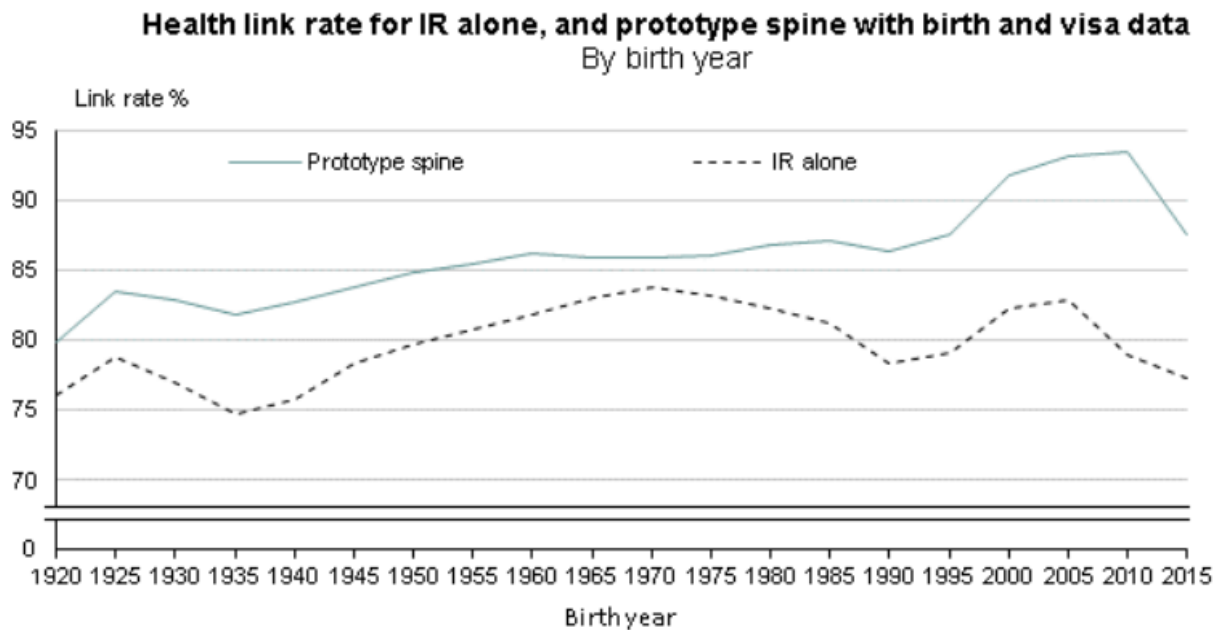
How can I use the spine?

The majority of research projects in the IDI are likely to be restricted to records that are in the IDI spine. For example the current methods for producing an estimated residential population (ERP) from IDI restricts the population to people in the spine. This is because if people are not in the spine they cannot be connected to border movements information, and the border movements information is essential for determining how much time people have spent in New Zealand. If the ERP did not use border movements, the ERP would end up including (for example) people who came to NZ for a two-week holiday and visited hospital for a broken leg.

The personal_detail table contains a variable called 'snz_spine_ind' that indicates whether an individual is in the spine or not.

In the wiki, there is information about link rates that show what proportion of agency datasets (eg health datasets) were linked to the spine in each IDI refresh (eg Figure 1). People that are not linked to the spine are those with no records in the tax, visa or birth administrative data used to construct the spine. For the health-spine link, an individual may be present in health data but not in the spine for several reasons including: they were a tourist who needed health care; they are a resident of New Zealand but are not included in the spine (the spine has a small amount of under coverage) see Black (2016); they were a false negative link. Further work is required to understand known areas of over- and under-coverage of the spine, and its potential impact on selection bias if particular groups for example are undercounted.

Figure 1: Health link rate to the Prototype Spine, from Black (2015)

Health link rate for IR alone, and prototype spine with birth and visa data
By birth year

Source: Statistics New Zealand

More information about linking methodology that might be useful:

- Statistics New Zealand (2014). *Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure project.* Retrieved from: http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-resources/linking-methodology-statsnz-idi.aspx

By Dr Andrea Teng and Dr Sheree Gibb

Version: Original 4 May 2017