

Ethnicity and the IDI

The post aims to improve understanding about the challenges of using ethnicity data in national data collections such as the IDI.

Ethnicity is a measure of cultural affiliation and is different from concepts of ancestry, nationality or citizenship. It is a widely used measure of cultural identity in New Zealand and is important in monitoring the well-being of ethnic groups as well as the impact of policy and services on wellbeing, health and the health sector. The persistence of stark differences in fundamental population health measures, such as a seven year gap in life expectancy between Māori and non-Māori reinforce the need for understanding, and addressing the underlying drivers of health inequalities in New Zealand.

Our aims here are to:

- describe the ethnicity information currently available in the IDI
- discuss some of the challenges in using ethnicity data in the IDI and make some suggestions for improvement.

This post focusses on the context and quality of ethnicity data in the IDI and how it can be used for research. A large body of research has been devoted to understanding the concept of ethnicity how it is related to colonisation, and structural, interpersonal and internalised racism. A conceptual understanding of ethnicity based on the literature is essential for robust study design and interpretation of analytical results for any study that examines ethnic differences within a population. This post does not review this literature, but researchers intending to use ethnicity data are encouraged to read [Cormack \(2010\)](#).

How is ethnicity measured in New Zealand?

Statistics NZ have produced a statistical standard for ethnicity that specifies how ethnicity information should be collected and aggregated. It was developed to ensure that ethnicity is collected consistently across all survey and administrative data collections within the official statistical system. The standard specifies a range of guidelines for collecting ethnicity, including:

- ethnicity should be self-identified
- individuals should have the opportunity to specify multiple ethnicities (there should be space for at least three, but preferably six)
- the prioritising of ethnic responses to one per individual should be discontinued

(For the full standard, see [Statistical Standard for Ethnicity](#))

The statistical standard recommends using multiple response measures of ethnicity, rather than the prioritised ethnic identification measures traditionally used in health research. More than 11% of the population identified with multiple Level 1 ethnic groups in the 2013 census, and the levels of multiple ethnic identification are even higher for children and young adults (Statistics NZ, 2014). Therefore, multiple response ethnicity provides a better reflection of the ethnic identity of New Zealand population than prioritised ethnicity (Goodyear 2009). The ethnicity prioritisation process doesn't necessarily align with the respondent's perception of their ethnicity and is therefore inconsistent with the conceptual foundations of ethnic identity (Kukutai, StatsNZ 2008). Furthermore, the application of the prioritisation process can result in under-representation of ethnic groups lower on the prioritisation, especially Pacific (Statistics NZ 2006) resulting in significant under-counts and age-based selection bias. Therefore, multiple response measures are recommended for description and analysis of ethnic-specific outcomes.

Ethnicity can be recorded and output at different levels of detail. The standard classification of ethnicity is a hierarchical classification of four levels. Level 1 of the classification has six categories: European; Maori; Pacific; Asian; MELAA (Middle Eastern, Latin American and African); and Other. Level 2 has 21 categories, Level 3 has 36 categories and Level 4 has 233 categories (see [Statistical Standard for Ethnicity](#)).

Where can I find ethnicity information in IDI?

In the IDI ethnicity data is available from a range of datasets including: 2013 census; health; education; ACC; and births. These collections vary in how ethnicity information is collected, the level of detail and quality of the data, and the context in which data were provided.

Ethnicity information from different IDI collections is combined using the source ranked ethnicity method, which reports an ethnicity profile for each person by drawing on several IDI data sources. Data sources containing ethnicity information are ranked according to quality, and each individual is assigned the ethnic profile from the highest ranked data source available. The table records ethnicity in total response format, so an individual can have more than one ethnic group recorded. Ethnicity is recorded at Level 1 groupings (European, Maori, Pacific, Asian, MELAA, Other). The ranking order for the sources has been selected on the basis of how well the ethnicity information in each agency's dataset agrees with census data (see figure below for some examples). Census ethnicity data is given the highest priority, followed by birth data and then the Ministry of Health data (the full ranking can be found in the IDI wiki under the 'central tables' heading).

This source of ethnicity information has several strengths:

- it simplifies the use of ethnicity data in IDI by providing a single ethnic profile for each individual
- it is recommended by Stats NZ because it has the best agreement rates with the census. For example for Māori/non-Māori ethnicity, 95% of people in the census and health datasets have the same ethnicity recorded.

However, it also has some limitations:

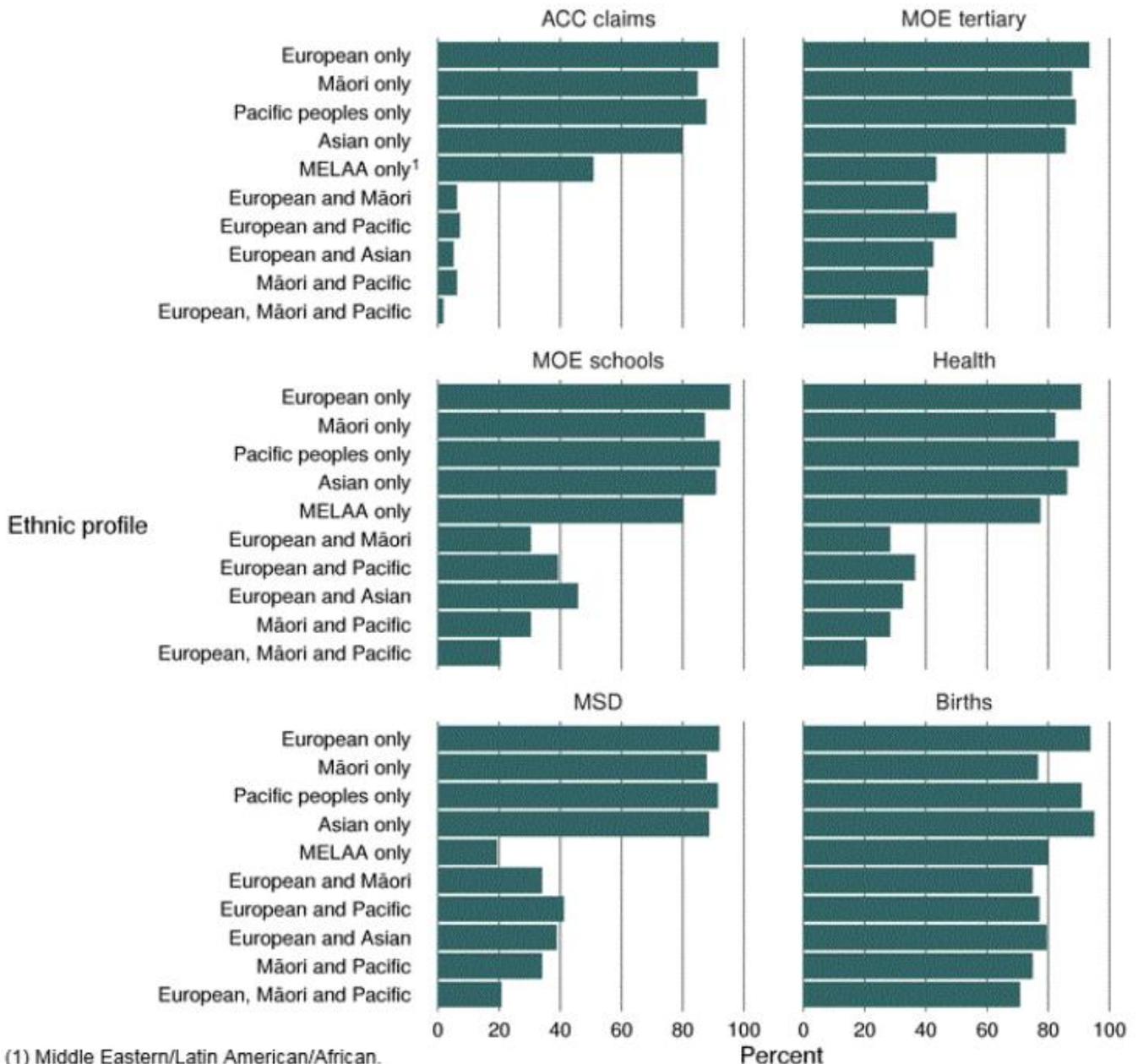
- ethnicity is assigned from a single source, so it may underestimate the number of people with multiple ethnicities
- ethnicity information is provided at Level 1 only so for example it cannot be used to identify different Asian or Pacific sub-group ethnicities (eg Indian, Samoan). Level 2 ethnicity information is available in some datasets (eg Census 2013, health) but this does not flow through to the ethnicity variables in the personal detail table.
- the ranking of sources is based on how well they match the 2013 census, which may not suit all purposes. For example, researchers may want to give higher priority to sources that match the context in which they are working (eg health for a health context). Alternatively, they may wish to prioritise ethnicity information that was collected recently (for many individuals ethnicity may have changed since being collected on their birth record, which by definition cannot be based on self-identification).

Changes to ethnicity recording over time in IDI

Prior to the July 2018 refresh of the IDI, source ranked ethnicity information was only available in a separate source-ranked ethnicity table. From the July 2018 refresh onwards the source ranked ethnicity variables were moved to the 'personal_detail' table, and the source_ranked_ethnicity table was retired.

Ethnicity variables are available in the personal_detail table prior to the July 2018 refresh, however these are not source ranked ethnicities. Instead an 'ever recorded' rule has been used to assign ethnicity. So, the ethnic profile for an individual contains all ethnicities ever recorded for that individual across all IDI data sources. Recording ethnicity in this way results in an over-representation of all ethnic groups. For this reason, use of ethnicity from the personal detail table prior to the July 2018 refresh is not recommended.

Percent agreement with census ethnic profiles, by administrative source



(1) Middle Eastern/Latin American/African.

Source: Statistics New Zealand

Source: Reid et al, 2016.

Challenges when working with ethnicity data in IDI

There are several challenges when working with ethnicity data in the IDI.

A given individual may have different ethnicities recorded in different datasets. This can occur for several reasons:

- First, individuals are likely to give a different response to an ethnicity question depending on the context, the way they are asked and for what reason they are being asked. For example if an individual is arrested, rushed to the emergency department, applying for ACC, or filling out a census form, there are many different factors that may influence what ethnicity you self-identify with and how it is recorded. In many cases ethnicity data may not be self-identified at all. Improving the quality of ethnicity data has been a focus of significant effort across the health sector, but there are still limitations. For example: if the recorded ethnicity is based upon whether a patient is asked, or a receptionist makes a judgement, or a family member fills in the form. Consequently, an individual's ethnicity can easily be recorded differently between datasets, even within the health sector.
- Second, ethnic identity can change over time (ethnic mobility) and much of the ethnicity data in the IDI is not currently time stamped, unless it was collected on a particular date. For example the 2013 census data for example was collected on a fixed date (5 March 2013), and ethnicity from birth records can be tied to the birth date. The lack of time-stamping makes it very difficult to examine ethnic mobility over time.
- Third, some datasets do not fully allow for multiple ethnic responses. Data on multiple ethnicities is required for total ethnicity response or combination output ethnicity analyses as recommended by Stats NZ. The report by Stats NZ comparing the quality of combination ethnicity responses between admin datasets and the 2013 census reveals the poor performance of many of these datasets (including health and ACC) to identify someone with more than one ethnicity compared to the census ([Reid et al, 2016](#)).
- As with any data, there may be errors in recording.

Ethnicity analysis – what could be improved?

How an analysis accounts for ethnicity will depend on the study question and context, however there are some tools that researchers may find helpful and improvements that could be implemented.

The datasets used to define ethnicity can be selected in a way that to improve the precision of the timing of ethnicity data collection. For example using to the 2013 census alone gives a fixed date but also increases the missing ethnicity data through incomplete census-spine linkage and census undercount (which varies by age, ethnicity and gender). If the study is interested in someone's ethnicity at a historical point in time, then note that the source-ranked ethnicity table includes the most recent measures of ethnicity from each of the contributing datasets (and you may consider using an earlier refresh to obtain ethnicity data).

Improvements are required to increase the quality of ethnicity data from administrative sources. The national study on primary care utilisation demonstrated that updating ethnicity information by asking the patient to complete the census ethnicity question is a simple and quick way to improve the quality of the ethnicity data (Crengle et al 2005). Time-stamping of updates to the ethnicity field would be

an easy way to identify the applicability and reliability of ethnicity information from an individual source. In some cases (eg health) these timestamps are collected by the source agency but are not currently included in the IDI. If we are to realise the value of the IDI to understand the determinants of different outcomes within New Zealand's population, a key area for improvement is the standardisation of questions and categories used for ethnicity recording across agencies and data sources.

The strength of the IDI is that the infrastructure provides a means of linking de-identified information at an individual level across multiple data sources. While linkage rates are very high, they are not uniform and vary by both dataset combination as well as individual demographic characteristics. A greater understanding of how incomplete linkage, linkage bias and incomplete coverage of the estimated residential population impact is important to inform our understanding of the limitations of IDI based investigations of differences in outcomes within the NZ population. Source-ranked ethnicity is not corrected for linking errors or measurement errors in particular data sources. (See VHIN posts about the [spine](#) and the [estimated resident population](#)).

Summary

The IDI contains ethnicity data from multiple sources and is a valuable tool for analyses that include ethnicity. However there is some distance to go until we have a fuller understanding about the quality of ethnicity data in the IDI. In the meantime it is important that researchers understand and acknowledge the limitations of ethnicity data in IDI, and make an effort to understand the broader issues surrounding ethnicity in New Zealand. Addressing the underlying drivers of health inequalities in New Zealand is a government responsibility and Treaty of Waitangi obligation.

For more information on ethnicity measures using administrative data please see;

A collection of Statistics NZ papers about ethnicity is available at: http://archive.stats.govt.nz/browse_for_stats/population/census_counts/review-measurement-of-ethnicity/papers.aspx

Ethnicity categories from the data.source_ranked_ethnicity table in the IDI;

snz_ethnicity_grp1_nbr is European,
snz_ethnicity_grp2_nbr is Maori,
snz_ethnicity_grp3_nbr is Pacific,
snz_ethnicity_grp4_nbr is Asian,
snz_ethnicity_grp5_nbr is MELAA,

snz_ethnicity_grp6_nbr is Other

[PDF](#) available here (for a printable version)

References

Bycroft, C, Reid, G, McNally, J, Gleisner, F (2016). *Identifying Māori populations using administrative data: A comparison with the census*. Available from <http://archive.stats.govt.nz/methods/research-papers/topss/maori-info-admin-data.aspx>

Crengle S, Lay-Yee R, Davis P, Pearson J. 2005. A Comparison of Māori and Non-Māori Patient Visits to Doctors: The National Primary Medical Care Survey (NatMedCa): 2001/02. Report 6. Wellington: Ministry of Health.

Cormack D (2010). The practice and politics of counting: ethnicity data in official statistics in Aotearoa/New Zealand. Wellington: Te Rōpū Rangahau Hauora a Eru Pōmare. Cormack, D, & McLeod, M (2010). Available from <http://www.otago.ac.nz/wellington/otago600095.pdf>

Cormack D & McLeod M. (2010). Improving and maintaining quality in ethnicity data collections in the health and disability sector. Te Rōpū Rangahau Hauora a Eru Pōmare: Wellington. Available from <http://natlib.govt.nz/records/22945073?search%5Bpath%5D=items&search%5Btext%5D=Improving+and+maintaining+quality+in+ethnicity+data+collections+in+the+health+and+disability+sector>

Goodyear, RK (Statistics New Zealand) (2009). *The differences within, diversity in age structure between and within ethnic groups*. Wellington: Statistics New Zealand. http://archive.stats.govt.nz/browse_for_stats/population/census_counts/review-measurement-of-ethnicity/papers.aspx#technical

Kukutai, Tahu (Tahatū Consulting), Statistics New Zealand (2008). Ethnic Self-prioritisation of Dual and Multi-ethnic Youth in New Zealand, Wellington: Statistics New Zealand.
http://www.stats.govt.nz/browse_for_stats/population/census_counts/review-measurement-of-ethnicity/papers.aspx#technical

Reid, G, Bycroft, C, Gleisner, F (2016). Comparison of ethnicity information in administrative data and the census. Available from <http://archive.stats.govt.nz/methods/research-papers/topss/comp-ethnic-admin-data-census.aspx>

Statistics New Zealand (2005). Ethnicity New Zealand Standard Classification. Available from <http://archive.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx>

Statistics New Zealand (2006). The Impact of Prioritisation on the Interpretation of Ethnicity Data. http://archive.stats.govt.nz/browse_for_stats/population/census_counts/review-measurement-of-ethnicity/papers.aspx#technical

Statistics New Zealand (2014). 2013 Census QuickStats about culture and identity. Available from <http://archive.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture-identity.aspx>

Original 4/9/17, by Andrea Teng, Sheree Gibb, Andrew Sporle

Last updated 23 January 2020