

Experimental ethnic population estimates from linked administrative data



Census Transformation

Stats NZ

New Zealand Government



Crown copyright ©

See Copyright and terms of use for our copyright, attribution, and liability statements.

Disclaimer

The results in this paper are not official statistics. They were created for research purposes using the Integrated Data Infrastructure (IDI) managed by Stats NZ.

Access to the anonymised data used in this study was provided by Stats NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation. The results in this paper were confidentialised to protect these groups from identification.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI.

See <u>Privacy impact assessment for the Integrated Data Infrastructure</u> for more information.

Citation

Stats NZ (2018). *Experimental ethnic population estimates from linked administrative data*. Retrieved from www.stats.govt.nz.

ISBN 978-1-98-852867-0 (online)

Published in April 2018 by

Stats NZ Tatauranga Aotearoa Wellington, New Zealand

Contact

Stats NZ Information Centre: <u>info@stats.govt.nz</u> Phone toll-free 0508 525 525 Phone international +64 4 931 4600

www.stats.govt.nz

Contents

Background
Census transformation in New Zealand5
About this report5
Introduction
Aim and scope7
Explaining ethnicity
Classification of ethnicity in New Zealand8
Outputs
Challenges in measuring ethnicity9
Data sources11
New Zealand Census of Population and Dwellings11
Estimated resident population11
Integrated Data Infrastructure12
Reviewing the ranking method14
Comparing admin sources with the 2013 Census14
Source ranking versions
Chosen ranking version compared with 2013 Census16
Level 1 ethnicity
Level 2 ethnicity
Level 3 and level 4 ethnicity21
IDI-ERP compared with the ERP22
Ethnic distribution changes over time
Age structure differs across ethnic groups24
Ethnic distribution varies across geographical areas27
Area case studies
Discussion
Conclusion
References
Appendix 1: Quality of ethnicity information in the IDI
Appendix 2: Comparison of ranking versions

List of tables and figures

Tables

1. Distribution of ethnic sources, ranking 1 and ranking 3	16
2. Total responses, combined admin data and 2013 Census, level 1 ethnic groups	17
3. Ethnicity in admin sources for individuals coded as Fijian Indian in 2013 Census	21
4. ERP and IDI-ERP for level 1 ethnic groups, at 30 June 2006 and 2013	22
5. Ethnic distribution in linked census–IDI-ERP and unlinked populations	27
A1. Coverage of ethnicity information, by admin source	
A2. Ratio of total responses in admin sources to 2013 Census, level 1 ethnic groups	
A3. Ratio of total responses in admin sources to 2013 Census, level 2 ethnic groups	
A4. Options for prioritised ranking of ethnicity sources	
A5. Ratio of total responses in combined admin data to 2013 Census, level 1 ethnic groups	
A6. Ratio of total responses in combined admin data to 2013 Census, level 2 ethnic groups	
A7. Total responses, combined admin data and 2013 Census, level 2 ethnic groups	

Figures

1. Mean number of level 1 ethnicities, by data source and single year of age14
2. Ratio of total responses in combined admin data to 2013 Census, by single year of age and level 1
ethnic group18
3. Ratio of total responses in combined admin data to 2013 Census, by level 2 ethnic group
4. Ratio of total responses in combined admin data to 2013 Census, by five-year age group and level
2 ethnic group
5. Ratio of total responses in combined admin data to 2013 Census, by five-year age group and level
2 ethnic group
6. Change in population share, by level 1 ethnic group, 30 June 2006–1323
7. Annual population change, by level 1 ethnic group, year ended 30 June 2007–16
8a. Comparison between ERP and IDI-ERP, Māori ethnic group, at 30 June 2006–1624
8b. Annual growth rate in ERP and IDI-ERP, Māori ethnic group, year ended 30 June 2007–1624
9. Age pyramids for level 1 ethnic groups in ERP and IDI-ERP, at 30 June 201325
10. Percentage difference between ERP and IDI-ERP, by five-year age group and level 1 ethnic group,
at 30 June 2013
11. Comparison of ethnic share in ERP and IDI-ERP, by TALB and level 1 ethnic group28
12. Comparison of ERP and IDI-ERP – Howick local board area, by level 1 ethnic group
13. Comparison of ERP and IDI-ERP – Manurewa local board area, by level 1 ethnic group
14. Comparison of ERP and IDI-ERP – Porirua city, by level 1 ethnic group
15. Comparison of ERP and IDI-ERP – Selwyn district, by level 1 ethnic group

Background

Census transformation in New Zealand

In March 2012 the New Zealand Government agreed to a Census Transformation Strategy. This strategy has two strands:

- a short-to-medium-term focus on modernising the current census model and making it more efficient
- a longer-term focus on exploring the feasibility of a census based on administrative data (Stats NZ; 2012, 2014a).

Investigations into the long-term direction for census are focused on understanding future census information requirements and the ability of administrative (admin) data to meet those requirements.

<u>Census transformation – a promising future</u> (a 2015 Stats NZ Cabinet paper) recommended that Stats NZ work actively towards a future census based primarily on Government's admin data, supported by redeveloping its household surveys.

<u>Census transformation in New Zealand</u> has more information.

About this report

Ethnicity is a core demographic variable describing the New Zealand population. Ethnicity data collected in the census is widely used and is the basis for official population statistics by ethnicity. In an admin-based census, ethnicity is an essential variable we would need to obtain from admin sources.

Through 2016 and 2017 we have released experimental series of national and subnational population estimates, derived from linked admin data in Stats NZ's Integrated Data Infrastructure (IDI). This release extends the series to include ethnicity for the first time.

We update the findings of Reid, Bycroft, and Gleisner (2016) by comparing ethnicity information from admin sources with that from the 2013 Census, extending the analysis to include level 2 ethnic groups. We discuss options for combining various admin data sources.

We use the selected method to generate experimental population estimates by ethnic group for 2006–16, which we assess against the estimated resident population (ERP).

This published experimental data series includes estimates at 30 June 2006–16 for:

- level 1 and level 2 ethnic groups by five-year age group and sex
- level 1 and level 2 ethnic groups by territorial authority and Auckland local board (TALB) area, five-year age group, and sex.

We invite your feedback on any of the methods and results covered in this report, to support our future development work.

Use this form to submit your feedback.

Note that these experimental ethnic population estimates are not official statistics. Rather, they are published as output from research into a different methodology than that currently used to produce official population estimates.

Introduction

Ethnicity is the principal measure of cultural identity in New Zealand, and is used across the Official Statistics System. Collecting ethnicity information in the Census of Population and Dwellings is a legislative requirement under the Statistics Act 1975; the census is an important source of ethnicity data for small areas and small ethnic groups. The census also provides the basis for official population estimates and projections of major ethnic groups.

To meet Treaty of Waitangi commitments, official estimates of the Māori ethnic population are an essential requirement for future censuses (Gleisner, Downey, & McNally, 2015). Ethnicity data is also used with other characteristics of the population – to inform resource allocation, policy development, and research.

The ability to produce ethnicity data from admin sources is therefore a key consideration when determining if a census based on admin data is feasible. By using ethnicity information from linked admin data we may also be able to improve the current production of official ethnic population estimates that are based on the five-yearly census.

Reid, Bycroft, and Gleisner (2016) compared ethnicity data from the 2013 Census with the ethnicity information collected by admin sources that is available in the IDI. The study used data from the May 2015 IDI refresh and was limited to level 1 ethnic groups. They found that nearly everyone had ethnicity recorded in at least one admin source, but that consistency with census responses varied considerably depending on the collection. They addressed the question of how to best combine ethnicity from multiple admin sources.

The method used to combine these multiple sources has a major impact on the result. The 'everreported' approach, which combines all ethnic groups reported in any source, inflated the counts of all ethnic groups, particularly those for Māori and Pacific peoples, and is not recommended.

Ranking admin sources (based on their agreement with the census), and using the ethnicity information from the highest ranked source available for each person, brings the admin data closer to the census. However, this approach undercounted most ethnic groups. Reid et al examined three ranking orders and chose one as having the best overall performance.

Aim and scope

Building on this previous work, we investigate the potential for admin data to provide ethnicity information comparable to the current census and the official ethnic population estimates. We review the ranking methods used by Reid et al (2016), and also examine the consistency of level 2 ethnicity information. We compare the resulting estimates against the official ERP.

This report accompanies the release of an experimental data series of ethnic population estimates produced from linked admin data – see <u>experimental population estimates from linked</u> <u>administrative data</u> on the Stats NZ experimental initiatives page for more detail.

Our aim is to update customers about our progress in producing these admin-based estimates, and in understanding their quality. We also hope to receive input from our customers.

Note: this series has been produced solely from the linked admin data. The estimates do not incorporate any additional statistical modelling.

Explaining ethnicity

According to the statistical standard for ethnicity (Stats NZ, 2005), ethnicity is:

the ethnic group or groups that people identify with or feel they belong to. Ethnicity is a measure of cultural affiliation, as opposed to race, ancestry, nationality or citizenship. Ethnicity is self-perceived and people can belong to more than one ethnic group.

Classification of ethnicity in New Zealand

The 2005 New Zealand standard classification of ethnicity is a hierarchical classification of four levels. Level 1 has six categories and is used solely for output, not for collection. Apart from Māori, level 1 categories are ethnic groups (a collection of related ethnicities), not ethnicities as such.

Ethnicity level 1 categories:

- 1 European
- 2 Māori
- 3 Pacific peoples
- 4 Asian
- 5 Middle Eastern/Latin American/African (MELAA)
- 6 Other ethnicity
- 9 Residual categories

Level 2 has 21 categories, which include larger ethnicities that are within the level 1 groups, such as New Zealand European, Samoan, and Indian. Level 3 has 36 categories and level 4 has 233 categories (excluding residual categories). Individual ethnicities are aggregated into progressively broader ethnic groupings from level 3 up to level 1, according to geographical location or origin, or cultural similarities.

In the standard classification 'New Zealander' is a separate ethnic group at level 4, which aggregates to 'Other ethnicity' at level 1. This approach was used in the 2013 Census. For time series purposes, the standard suggests that New Zealander responses can be combined with the New Zealand European group. For example, the official estimated resident population (ERP) combines European and Other ethnicity. This is largely to minimise the effect of New Zealander responses being inconsistent across censuses, and being inconsistent when compared with other data sources.

Current recommended practice in the health sector codes 'New Zealander' separately, but aggregates these responses to the New Zealand European category for output purposes (Cormack & McLeod, 2010; Ministry of Health, 2017). Other admin sources may not collect New Zealander as a separate response.

For comparability with the ERP and admin sources, we include any New Zealander responses coded at level 4 in the European category.

Outputs

The presence of multiple ethnicities for the same person needs to be considered when reporting ethnic results. In the 2013 Census, 10 percent of individuals identified with two or more level 1 ethnic groups.

There are two standard outputs for ethnicity data:

- In 'total response' data, individuals are counted in all their reported ethnic groups. For example, a person who reported their ethnicities as Samoan and Māori, is counted once in the Pacific peoples category and once in the Māori category for level 1 outputs. The number of grouped total responses tends to be greater than the total population.
- 'Single and combination' data counts people in mutually exclusive categories. People reporting two or more ethnic groups are counted once in the relevant 'combination' group. This means the total number of responses equals the total number of people who stated their ethnicity. In the above example, this person is counted once in the 'Māori and Pacific' combination at level 1.

This report primarily focuses on total response counts.

Challenges in measuring ethnicity

The official New Zealand Standard and Classification for ethnicity is available and used in the census, others Stats NZ surveys, and some admin sources. Reid et al (2016) compared the consistency of the main admin sources in the IDI with the statistical standard for ethnicity. Most admin sources capture the same concept of ethnicity, and aim for self-identification where possible. Despite this consistency, collecting ethnicity information is challenging for several reasons. We outline the challenges here.

Question format affects responses

The format of a question influences responses. The question used for ethnicity differs widely across admin data sources, and also differs within each admin source – depending on the collection mode or the form used. Birth registrations are a national collection that uses the same question format as the census, and shows the greatest consistency with the census.

Ethnic group can change with context

A person may give a different response depending on the question's context. For example, when filling in a self-administered form a person may respond differently from when an interviewer asks about their ethnic group. The social or cultural setting may also affect the ethnicity response reported.

Similarly, not all responses are self-identified. Ethnicity in births data is reported by parents, while some census forms are not completed by the individual themself. Other admin collections could also be a mixture of self-identified and third-party responses.

We see evidence of contextual effects when we compare admin sources against the census. The high degree of consistency between birth registrations and the census may also be partly because both are formal and neutral contexts, where defining ones identity is seen as important.

Ethnic group changes over time (inter-ethnic mobility)

The ethnic group or groups that someone identifies with may change over time. We need to allow for inter-ethnic mobility in longitudinal surveys and admin databases. Inter-ethnic mobility also affects how we integrate different datasets – the same person may have given different ethnic group responses at different times, or in different collections. Generally, each data source will measure ethnicity at a moment in time. Ideally, collections should allow for time referencing of ethnicity responses, to help assess inter-ethnic mobility over time.

The census measures ethnicity at a single point in time (census day), and official population estimates also measure ethnicity at a given reference date. Birth registrations record ethnicity for a single event in a person's life. Other sources may combine responses that have been recorded over time.

People can have multiple ethnicities

People may identify with more than one ethnic group so it's essential to be able to collect multiple ethnic groups for each individual. The census collects up to six ethnic group responses per person. Most, but not all, admin sources record multiple responses. On the whole, at least three ethnic groups are recorded, the standard's minimum requirement. Reid et al (2016) found that, with the exception of birth registrations, fewer people report multiple ethnicities in admin sources than they do in the census.

Data sources

This chapter describes the data used for this research: the 2013 Census, the official estimated resident population by ethnic group, and the relevant admin sources in the IDI.

New Zealand Census of Population and Dwellings

The Census of Population and Dwellings is the official count of people and dwellings in New Zealand. It provides a snapshot of our society at a point in time and tells the story of social and economic change in New Zealand. The census aims to count everyone who is in New Zealand on census night. Overseas visitors are included in the census, while New Zealand residents who are not in New Zealand on census night are not included. The census forms the basis for official statistics of ethnicity and provides the best indication of an individual's ethnicity on census day.

Ethnicity is a 'foremost' variable in the census, which means Stats NZ manages it to produce information of the highest quality. The non-response rate for ethnicity for usual residents who returned a 2013 Census form was 0.7 percent. With substitute forms included, the non-response for ethnicity was 5.4 percent.

In the 2013 Census, net undercount varied by ethnic group. The highest undercount was for Māori (6.1 percent), followed by Pacific peoples (4.8 percent), Asian (3.0 percent), and European (1.9 percent) (Stats NZ, 2014b).

Estimated resident population

The estimated resident population (ERP) of New Zealand is an estimate of all people who usually live in New Zealand at a given date (see <u>Standard for population terms</u>). We use the ERP to assess the plausibility of our admin estimates at the aggregate level.

The current methodology for producing the ERP relies on a periodic full-enumeration census.

The ERP at a given date is derived by updating the census usually resident population count for estimates of:

- net census undercount (as estimated by the Post-enumeration Survey)
- residents temporarily overseas on census night
- natural increase (births less deaths) between census night and the given date
- net migration (arrivals less departures) between census night and the given date (Stats NZ, 2014c).

Additionally, ethnicity is imputed for individuals who don't respond to the census ethnicity question. The adjustment for net census undercount accounts for the ethnic differentials in census response, and includes large increases in the ERP for the young adult Māori and Pacific populations. Good ethnicity data is available for natural increase from birth and death registrations. However no direct ethnicity data is available for migration, so ethnic breakdowns for residents temporarily overseas and for net migration are estimated based on other characteristics.

We publish official estimates for the total national population by age and sex each quarter, and subnational populations are published annually.

Estimates for the level 1 ethnic groups, including by subnational area, are published at 30 June of each census year. Only official estimates of the national Māori population are updated between censuses; we publish these each year.

Stats NZ also produces national projections of future populations for the level 1 ethnic populations, and for the Samoan, Chinese, and Indian ethnicities.

Integrated Data Infrastructure

Stats NZ developed the Integrated Data Infrastructure (IDI) as an environment in which multiple data sources are linked in a systematic and secure way. It was developed to produce official statistics and to allow Stats NZ staff and external researchers to conduct policy evaluation and research on people's transitions and outcomes. The IDI contains de-identified admin and survey datasets, linked at the individual level.

We use the IDI as a test environment for examining the potential of linked admin data sources to produce ethnic population estimates.

The basic structure of the IDI is a central 'spine' to which a series of data sources are linked. The target population for the spine is all individuals who have ever been residents of New Zealand. Black (2016) provides more information on the IDI spine's formation.

Other data sources are linked to the IDI spine (see Stats NZ, 2014d for a description of the linking process). The linked datasets cover a wide range of subjects and include: employer and employee job and earnings information based on Inland Revenue data; health information, including GP enrolment and hospital visits, from the Ministry of Health; education data from Ministry of Education; benefit dynamics data from the Ministry of Social Development; migration movements data from the Ministry of Business, Innovation and Employment; and data from Stats NZ's Household Labour Force Survey and 2013 Census.

The IDI continues to change as new datasets are added.

<u>Data in the IDI</u> explains the current information. We used the June 2017 IDI refresh for all analysis presented in this paper.

Ethnicity information in the IDI

Ethnicity information is available in the IDI, from many sources. We focus on ethnicity information from five agencies identified as being the most reliable:

- Department of Internal Affairs (DIA) ethnicity of the child as reported on the birth registration
- Ministry of Education (MOE) ethnicity as reported on schools and tertiary enrolment data
- Ministry of Health (MOH) ethnicity as recorded in the National Health Index system, a unified national person list compiled by the MOH
- Ministry of Social Development (MSD) ethnicity as reported by individuals obtaining working-age benefits
- Accident Compensation Corporation (ACC) ethnicity reported for an individual who has made an ACC claim.

For a more detailed explanation of these sources see Reid et al (2016) and Cormack & MacLeod (2010).

The IDI also includes ethnicity information from the 2013 Census, and other admin and survey data sources not covered in this report.

Time referencing varies across sources. Some sources, such as DIA births and the census, include ethnicity at a single date, while others include multiple pieces of information relating to each time ethnicity was reported. The remainder contain only a single record for each individual with no time referencing.

Where necessary, the IDI applies business rules to standardise the ethnic codes received from each agency – to align with level 1 of the 2005 ethnicity standard. The IDI has two tables that include ethnicity from a range of sources. The personal details table applies the 'ever-recorded' method, and so includes an ethnicity whenever it is recorded by any agency. A second table uses the source rankings suggested by Reid et al (2016) – 15 sources are included, with the 2013 Census given the highest priority, followed by DIA, MOH, and MOE.

Identifying the New Zealand resident population in the IDI

Stats NZ (2017) describes a method used to identify individuals in the IDI spine who are resident in New Zealand at a given point in time, with the resulting population referred to as the IDI-ERP.

The IDI-ERP is the basis for our admin ethnic population estimates. While the IDI-ERP often shows close agreement with official ERP figures, we still find marked differences for some age groups and local areas. These differences tend to flow through to the ethnic estimates as well.

Linking the 2013 Census to the IDI-ERP

The census provides the best indication of an individual's ethnicity on census day. We therefore treat the 2013 Census ethnicity as a 'gold standard' for assessing the quality of the admin ethnicity. Of course, the census itself is not perfect, and there may be genuine reasons for differences between the census and admin data, such as the time or context of collection. By using the June 2017 IDI refresh, there will also be information in the admin data that was not available at the time of the census.

The 2013 Census is linked to the IDI spine, which allows us to make individual-level comparisons with the ethnicity information in the admin data. We restrict the population for comparison to those individuals who: are in the IDI-ERP at 5 March 2013, have been linked to the 2013 Census, and have at least one ethnicity recorded in both the census and admin data.

The total population for comparison is 3,763,641 (89 percent of the census usually resident population count or 84 percent of the IDI-ERP).

Reviewing the ranking method

The source ranking method for combining the multiple sources, and recommended by Reid et al (2016), uses the ethnicity information from the highest ranked source available for each person, regardless of when that information was collected. We reviewed the ranking order proposed by Reid et al, to highlight any changes in the admin data sources by the June 2017 IDI refresh and to take account of any new information from the level 2 comparisons.

Comparing admin sources with the 2013 Census

Following the approach of Reid et al, we used the 2013 Census linked to the IDI-ERP to assess the consistency of ethnic responses in admin data with that from the census. Close agreement provides strong support for good measurement in both sources, but where we observe differences, it is difficult to determine which is the correct response. As outlined above, there are several reasons why an individual might record different ethnicity responses in the census and the IDI, and not all indicate errors in either source.

We first compared results for each admin source against the 2013 Census. Results are consistent with Reid et al. We found that coverage and accuracy of ethnic information differs across the sources (see tables A1 and A2 in appendix 1).

DIA births show very close agreement with the census for level 1 ethnic groups, but are largely available only for people born in New Zealand from 1998 onwards. Other admin data sources tend to record fewer ethnic groups per person than the census, resulting in some undercount for most ethnic groups when compared with the census. Extending the analysis to breakdown by age, figure 1 shows that younger ages are much more affected by the admin data having lower rates of people with multiple ethnic groups in all sources except DIA births.

Mean number of level 1 ethnicities



Figure 1

Source: Stats NZ

The increasing ethnic diversity of New Zealand's population, especially pronounced among children, is a well-recognised phenomenon (Khawaja et al, 2007). Good collection practices and the need to determine appropriate analysis for people who report multiple ethnic groups are critical in

interpreting ethnicity responses in the Growing Up in New Zealand study (Atatoa Carr et al, 2017). It is important to recognise the high quality of DIA birth registrations when combining data from multiple admin sources.

We found that the number of ethnic responses per person in the MOE tertiary data can vary considerably, based on how ethnicity is determined. Many individuals have multiple records for each year of enrolment – ethnicity is reported independently for each enrolment. We explored two different ways of using these records. Firstly, we aggregated all ethnicities recorded for a given individual – as done in our previous analysis. The aggregated tertiary enrolments have a much higher rate of multiple level 1 ethnic groups than the 2013 Census, or any other source. In contrast, using the ethnicity recorded on only one enrolment record results in fewer ethnic groups for each person, slightly below the 2013 Census.

There is higher consistency for level 1 ethnic groups than for level 2 across all data sources; DIA births continues to show the highest consistency with the census for level 2 ethnicities (table A3).

There is no clear evidence that two additional years of data has made any material difference to our earlier results, despite the presence of new ethnic data collected after the 2013 Census.

Appendix A: Quality of IDI ethnicity information by source has more detail about our findings.

Source ranking versions

We assessed six different versions of ranking the source agency data. (See table A4 in appendix 2 for complete details). Three versions are very similar to those used in Reid et al's analysis, with the others based on our updated findings. Each version assigns DIA birth registrations the highest priority and ACC the lowest priority. The key differences tend to come from the relative placement of MOH and MOE tertiary sources.

Table 1 uses two examples to show the effect of changing the ranking order. Ranking 1 places MOH data above MOE schools and MOE tertiary enrolments, respectively; ranking 3 prioritises tertiary enrolments. In each case DIA provides ethnicity information for 18 percent of the IDI-ERP. For ranking 1, almost all the remaining population (74 percent overall) is given MOH ethnicity, with only 3 percent from MOE tertiary. In ranking 3, 47 percent of the population is assigned ethnicity from MOE tertiary, with another 31 percent coming from MOH. In both approaches less than 3 percent of individuals use MOE schools, MSD, or ACC ethnicity.

In ranking 1, MOH is the most common source for all ages 15 years and above. In contrast, MOE tertiary becomes the most used source for ages 15–60 in ranking 3, and MOH continues to be the main source for ages over 60 years.

Distribution of ethnic sources									
Ranking 1 and ranking 3									
Banking	Ranking	1	Ranking 3	3					
order	Source	% of IDI-ERP using source	Source	% of IDI-ERP using source					
1	DIA births	18.3	DIA births	18.3					
2	МОН	73.8	MOE tertiary	46.9					
3	MOE schools	0.6	МОН	30.5					
4	MOE tertiary	3.2	MOE schools	0.2					
5	MSD	1.2	MSD	1.2					
6	ACC	1.8	ACC	1.8					
	No ethnicity information	1.2	No ethnicity information	1.2					
Symbol: . Source: St	Symbol: not applicable Source: Stats NZ								

Table 1

The versions that rank MOH higher tend to undercount Māori by 10 percent when compared with the census, reflecting the effect of relatively low counts of Māori in MOH data. Versions ranking MOE tertiary higher tend to have similar counts as the census for European, Māori, and Asian ethnic groups, but slightly higher counts than the census for Pacific and MELAA (table A5, appendix 2). Level 2 results are also generally closer to the census for rankings where MOE tertiary is given priority over MOH (table A6).

Appendix B: Comparison of ranking versions has more detail for our results.

We chose ranking 3, prioritising tertiary education data over MOH and using aggregated tertiary data, as the preferred method. We considered it to be the most consistent overall with the census – at both level 1 and level 2. It also retains the main features of Reid et al's preferred ordering.

However, we note that other ranking orders could be equally valid, depending on any specific requirements. Each admin source has certain limitations, and these will not be fully removed with any ranking order.

One key limitation of using a ranking method is that it makes no allowance for inter-ethnic mobility. In particular, using our suggested ranking order, anyone with ethnicity information on their birth record will continue to be given this ethnicity. This may be a reasonable approach at present, when ethnicity from birth records is of high quality and is still only available for children, but it may not be desirable to use birth registration ethnicity as the priority source into adult life. However, other sources are not reporting the same degree of multiple ethnicity for this younger cohort, so attempting to account for inter-ethnic mobility by using ethnicity from these other sources at later times may introduce spurious volatility.

Throughout the rest of this report, we focus on results using ranking 3. We also exclude level 1 'Other ethnicity', due to the high level of disagreement between the admin data and the census.

Chosen ranking version compared with 2013 Census

In this section we compare ranking 3, our preferred method for combining ethnicity from several admin sources, with the 2013 Census – for level 1 and level 2 ethnicities. We use the linked census–

IDI-ERP dataset, and only the individuals with at least one ethnicity in each source (3,763,641 in total).

The linked census–IDI-ERP data does not give us counts for all the resident population. However, the census is our best reference for examining the consistency of known ethnicity responses currently used for official statistics, and the admin-based ethnic estimates. Later we describe results for the admin ethnic estimates for the full resident population.

We use the ratio of total responses for each ethnic group to compare the 2013 Census with the preferred ranking 3. A ratio close to 1 indicates we would expect similar results in the census and the admin source when estimating the same ethnic population, although the same individuals are not necessarily included. A value greater than 1 would indicate more people in the admin data than in the census are reporting they are that particular ethnicity. A ratio less than 1 would suggest the admin data is not capturing everyone who reported being that ethnicity in the census.

Level 1 ethnicity

Table 2 shows the ratio of total responses in the combined admin data to the 2013 Census for each level 1 ethnic group.

The numbers of total responses for European and Māori in the admin data are very close to the census, with ratios of 0.99. Slightly fewer people are identifying with Asian ethnicity than in the census (0.97), while both Pacific and MELAA ethnic groups are overestimated in the admin data compared with the census, with ratios of 1.06 and 1.08, respectively.

Total responses, combined admin data and 2013 Census							
Level 1 ethnic groups							
Ethnic groupCombined admin data (ranking 3)2013 CensusRatio of combined admin to 2013 Census							
European	2,846,037	2,872,380	0.99				
Māori	553,881	559,230	0.99				
Pacific	290,451	272,970	1.06				
Asian	412,131	424,761	0.97				
MELAA 45,786 42,447 1.08							
Source: Stats NZ	Source: Stats NZ						

Table 2

Figure 2 shows the ratio of total responses broken down by single year of age. Compared with the census:

- a very similar number of individuals identify with the European ethnic group in the admin data across most ages
- slightly fewer individuals identify with the Asian ethnic group in the admin data, consistently across all ages
- more individuals identify with Māori ethnicity in the admin data for ages 19–35, but considerably fewer at older ages
- more individuals identify with Pacific ethnicity in the admin data across all ages the peak differences for both Māori and Pacific correspond to the ages where the majority of individuals are assigned MOE tertiary ethnicity

• more individuals identify with MELAA ethnicity in the admin data, particularly for ages 9–21, mostly due to the MOE schools data.



Figure 2

Level 2 ethnicity

Level 2 of the ethnicity classification includes larger ethnicities separately, such as Samoan or Indian, while the remaining smaller ethnicities are grouped (eg as 'Other Pacific' or African). Figure 3 shows the ratio of total responses for level 2 ethnicities. Table A7 has the corresponding numbers.

Most single ethnicities are well-represented in the admin data. Their ratios are between 0.96 and 0.99, showing a small undercount compared with the census. Indian is slightly lower at 0.94. The exceptions are New Zealand European, which is undercounted in the combined admin data (ratio of 0.82), and Fijian, which is significantly overcounted (ratio of 2.59).

The Other European, Other Pacific peoples, and Other Asian categories are all overcounted in the admin data with ratios of 1.21, 1.59, and 1.22, respectively. Two MELAA subgroups, Middle Eastern and Latin American, are undercounted (ratios of 0.84 and 0.91, respectively), while African is overcounted (ratio of 1.60).

The 'not further defined' (nfd) categories are effectively missing data at level 2 (although there is enough information to code these to a level 1 ethnic group). The large ratios (table A7) reflect higher levels of missing level 2 categories in the admin sources than in the census.



Figure 3



Source: Stats NZ

Figures 4 and 5 show the ratio of total responses for selected level 2 ethnicities by five-year age groups. Figure 4 shows examples of level 2 single ethnicities that perform well, with overall total response ratios just under 1, and patterns that are fairly consistent for all age groups. The ratio for all these ethnic groups drops at ages 15–19, which coincides with the youngest ages without ethnicity information from DIA birth records in 2013. This dip is expected to disappear over time if birth registrations continue to be treated as the highest priority source.

Figure 4



Figure 5 highlights much larger differences for some other level 2 ethnicities. New Zealand European is underestimated for ages over 25, largely driven by under-reporting in MOE tertiary data.

Correspondingly high numbers in the European nfd category suggest these responses are being incorrectly coded (see table A3). Of particular note is that both Fijian and African ethnicities are overestimated in the admin data relative to the census.



Figure 5

The issue with Fijian ethnicity is not isolated to a single source. Each admin source has more people identifying as Fijian than the census does, with three times as many responses in MOE tertiary and twice as many in MOH (table A3). This discrepancy appears to result from differences in the coding of 'Fijian Indian' responses. The New Zealand Statistical Standard for Ethnicity (2005) codes 'Fijian Indian' as 43112, which aggregates to Indian at level 2 and Asian at level 1. A recent review of ethnicity coding in the health sector found that Fijian Indian responses were sometimes coded separately to both 'Fijian' and 'Indian' (Ministry of Health, 2017).

Table 3 shows the ethnicity in each admin source for individuals who recorded Fijian Indian ethnicity, but not Fijian ethnicity, in the 2013 Census. There is evidence across each source that some individuals are being coded differently in the admin data. DIA births and MSD data appear to be largely consistent with the census, with most individuals coded as Indian only. However, more than two-thirds of individuals identifying as Fijian Indian in the census were coded as Fijian in the MOE tertiary data, and more than half (54 percent) as Fijian in the combined admin ethnicity.

Some of these differences will be genuine cases, where an individual has reported their ethnicity differently in different collections. However, there is strong evidence that the admin sources are misclassifying Fijian Indian responses in some cases, which contributes to the overestimate of people identifying as Fijian and the undercount of Indian.

Ethnicity in admin sources for individuals coded as Fijian Indian ¹ in 2013 Census									
		Admin source							
Ethnicity in admin data	DIA births	МОН	MOE schools	MOE tertiary	MSD	ACC	Combined admin data		
Indian only	897	5,325	1,902	1,296	2,418	3,789	4,293		
Fijian and Indian	291	2,589	423	2,340	174	1,044	3,717		
Fijian only	33	1,608	654	1,206	183	1,953	1,881		
Neither Fijian or Indian	51	438	198	315	177	993	474		
Total linked responses	Total linked responses 1,272 9,963 3,174 5,154 2,952 7,782 10,368								
1. Excludes individuals also coded as Fijian Source: Stats NZ									

Table 3

The African ethnic group is similarly over-reported in all sources except birth registrations (and MSD, which records no African ethnicity) (see table A3). One possible reason for this is that 'South African European' responses are miscoded to African, rather than Other European. Some residual responses may also be being incorrectly coded as African. The other MELAA groups, 'Middle Eastern' and 'Latin American', are both under-reported at level 2.

Both the African and Indian examples also affect the level 1 ethnic counts, resulting in more people being identified as MELAA and Pacific, respectively. Without either problem, we would expect both observed ratios to be closer to 1.

These examples are useful to highlight the potential difficulties in using admin data. Inevitably there will be differences in the treatment of certain responses that affect estimates produced solely from admin data. It is therefore crucial that we continue to understand any problems, and to work across government to ensure consistency in the collection and reporting of ethnicity.

Level 3 and level 4 ethnicity

Several rapidly growing ethnicities such as Filipino and Korean are not represented at level 2 of the ethnicity standard; the many smaller migrant and refugee communities will only be identifiable and counted through level 4 ethnicities. The census and DIA birth registrations collect ethnicity responses at the most detailed level 4, and aggregate these responses to progressively higher groupings. ACC also collects level 4 ethnicity while other admin agencies collect ethnicity at levels 2 or 3 (Reid et al, 2016).

Using the same ranking method as above, we would currently obtain level 3 ethnicity for 67 percent of the IDI-ERP (18 percent from DIA births, 47 percent from MOE tertiary, and 2 percent from ACC), and level 4 ethnicity for 20 percent of the IDI-ERP (18 percent from DIA births and 2 percent from ACC). While not investigated here, DIA also records ethnicity for the parents of registered births. Including these parents would provide detailed level 4 ethnicity for some of the adult population.

Altering the ranking order to prioritise sources with more detailed ethnicity information (in particular ACC) would result in more people having level 3 and 4 ethnicity, although this would also potentially result in lower quality estimates for level 1 and level 2 ethnicity.

IDI-ERP compared with the ERP

Using our chosen ranking version, we produced ethnicity estimates for the IDI-ERP at 30 June 2006– 16. In this section we analyse the key characteristics of these estimates by making comparisons with the official ERP.

In particular, we highlight how:

- ethnic distribution changes over time
- age structure differs across ethnic groups
- ethnic distribution varies across geographic areas.

For these comparisons we use the official ERP by level 1 ethnic group for 2006 and 2013. We also use the annual Māori ERP, which is available for all years between 2006 and 2016.

Ethnic distribution changes over time

Official figures show that New Zealand is rapidly transitioning to a more multi-ethnic society. New Zealand experiences high migration flows, with over one-quarter of the 2013 population born overseas; this high migration contributes to a growing ethnic diversity. The Asian population has expanded markedly since 2000.

These same broad patterns are seen in the admin ethnic estimates. Between 2006 and 2016, the IDI-ERP increased for all the main level 1 ethnic groups (table 4).

ERP and IDI-ERP for level 1 ethnic groups										
At 30 June 2006 and 2013										
Etheric encour		El	RP			IDI-	ERP			
Ethnic group	At 30 Jun	ie 2006	At 30 Jur	ie 2013	At 30 Jur	ne 2006	At 30 Jur	ie 2013		
	Count	Percent	Count	Percent	Count	Percent ¹	Count	Percent ¹		
European	3,213,300	76.8	3,312,100	74.6	3,137,847	75.4	3,235,446	72.9		
Māori	624,300	14.9	692,300	15.6	683,529	16.4	732,282	16.5		
Pacific	301,600	7.2	344,400	7.8	322,281	7.7	379,125	8.5		
Asian	404,400	9.7	541,300	12.2	363,903	8.7	505,563	11.4		
MELAA 38,600 0.9 53,100 1.2 34,896 0.8 56,736 1.3										
1. Excludes people with missing ethnicity										
Source: Stats NZ	Ζ									

Table 4

The relative proportions of people identifying with each ethnic group have been changing. We see a decreasing proportion of the population identifying as European, and increases in other groups.

Figure 6 shows the change in population shares between 2006 and 2013 for both the ERP and IDI-ERP. Both measures have very similar patterns, with the proportion with European ethnicity declining around 2.2 percentage points, the Asian ethnicity increasing around 2.5 percentage points, and smaller relative increases for each of the Māori, Pacific, and MELAA ethnic groups. The main difference is that the IDI-ERP shows evidence of slightly lower growth for Māori and higher for Pacific ethnicity.



Figure 6

An advantage of the admin series is that ethnic data is available every year, not just for census years. Figure 7 shows annual population change in the IDI-ERP for each level 1 ethnic group for 2006–16. The drop in population growth in 2012 for all ethnicities, except Asian, reflects low net migration that year. Rising growth after then, to 2015 and 2016, also reflects strong levels of international migration.



Figure 7

Māori ethnicity

Official estimates of the Māori population are published annually, which provides a useful comparison for the IDI-ERP.

Figures 8a and 8b compare the annual Māori population for the ERP and the IDI-ERP. Both show continued population growth. The IDI-ERP Māori population is higher than the ERP, although the gap closed over the 10-year period. There is evidence of more varied annual change within the IDI-ERP measure, ranging from 0.4 percent in 2012 to 1.7 percent in 2016. By comparison, the lowest annual growth rate in the ERP was 1.2 percent.

Each method has limitations that could explain these differences. The IDI-ERP makes no attempt to measure inter-ethnic mobility, so population change for the Māori population is restricted to people who enter or leave the overall population. The ERP estimation method relies on assumptions about patterns of Māori migration, since ethnicity is not recorded on arrival and departure cards. These assumptions may not fully capture year-to-year variation.

Māori population estimates has more information on Māori migration.

It seems plausible that the IDI-ERP could be more effective at identifying changes in the pattern of growth over time. This indicates there could be ways to improve the methods used to produce the official Māori estimates by making more use of the available admin data.



Figures 8a and 8b

Age structure differs across ethnic groups

Official figures show the age structures of ethnic populations differ markedly, with the Māori and Pacific populations having a much younger age profile than the older European population. The Asian population is somewhere between. Figure 9 presents age pyramids for each level 1 ethnic group, at 30 June 2013 and comparing the ERP and the IDI-ERP.



Figure 9 Age pyramids for level 1 ethnic groups in ERP and IDI-ERP, at 30 June 2013

Across all ethnic groups, both the ERP and IDI-ERP appear very similar. The most noticeable difference is for MELAA, with a larger grouping aged 15–24 in the IDI-ERP that is not apparent in the ERP. This difference is likely to be due to MELAA ethnicity being overstated in tertiary enrolments. Otherwise, the admin data appears to be effectively measuring genuine differences in age structure.

We also looked more closely at the differences between the ERP and IDI-ERP at 30 June 2013, as shown in figure 10.



Figure 10

Source: Stats NZ

Most trends are similar to those for the linked census-IDI population in figure 2. The European population is slightly lower in the IDI-ERP than in the ERP, while Pacific ethnicity is higher. We see the same pattern for people aged under 25 with MELAA ethnicity in the linked population.

However, some patterns are different from those in the linked census-IDI population. Both the Māori and Pacific populations are higher in the IDI-ERP relative to the ERP, while the Asian population is lower, especially for ages 35 and older. This suggests there are underlying differences in the remaining unlinked part of these populations.

Table 5 shows the distribution of ethnicity for people in the linked census–IDI-ERP population, and for the remaining people who were in the ERP and IDI-ERP, respectively, but were not linked.

We see that for those not in the linked data, the proportion identifying with either Māori or Pacific ethnicity is higher in the IDI-ERP than the ERP, while the proportions identifying with European or Asian is lower. This result is unexpected – all else being equal, we would expect to see little difference between the linked and unlinked groups.

Ethnic distribution in linked census–IDI-ERP and unlinked populations							
Ethnic group	In linked census-	IDI-ERP population	Not in linked census-IDI-ERP population				
	2013 Census	Combined admin data	ERP	IDI-ERP			
European	76.3	75.6	64.8	58.7			
Māori	14.9	14.7	19.6	26.9			
Pacific	7.3	7.7	10.5	13.4			
Asian	11.3	10.9	17.2	14.1			
MELAA 1.1 1.2 1.6 1.6							
Source: Stats NZ							

Table 5

It's possible the unlinked individuals are reporting ethnicity in the admin data differently to the linked population. However, all individuals are selected in the IDI-ERP using the same method, and we would generally expect the admin data to be reasonably consistent.

In contrast, the ERP relies on multiple adjustments in which ethnicity is imputed or otherwise estimated. The ERP includes imputation when ethnicity is not supplied in the census (mainly for substitute records), for residents temporarily overseas, and adjustments for net census undercount. Some of these adjustments may be underestimating the number of individuals with Māori and Pacific ethnicity, and overestimating the number with European or Asian ethnicity.

Ethnicity is not the only variable that could be affected. For example, we have previously observed differences in the IDI-ERP by age and sex, which could be occurring for similar reasons.

Ethnic distribution varies across geographical areas

The distribution of ethnic groups varies considerably across territorial authority and Auckland local board areas (TALBs), which is clearly reflected in the IDI-ERP.

Figure 11 compares the ethnic share for each TALB area. The x-axis shows the percentage of that area's ERP identifying with the given ethnicity, and the y-axis shows the percentage of the IDI-ERP identifying with the same ethnicity. Encouragingly, there is a high level of consistency for all ethnic groups and across all areas. All TALB areas with a high proportion of a specified ethnicity in the ERP have similar results for the IDI-ERP.

Figure 11



Comparison of ethnic share in ERP and IDI-ERP, by TALB and level 1 ethnic group

Generally, the IDI-ERP has slightly higher proportions of most areas identifying as Māori and Pacific, and slightly lower proportions identifying as European and Asian, in line with the national patterns.

We also see that the ethnic distribution varies across areas. Most areas have a majority of individuals identifying as European, but in two areas (Mangere-Otahuhu and Otara-Papatoetoe) this proportion is only 20 percent. In these two areas more than 50 percent of the population identify with Pacific ethnicity; the majority of remaining areas are below 5 percent.

There is also considerable variation in the number of people identifying as Māori or Asian across TALB areas, which is observed in both the ERP and the IDI-ERP.

Area case studies

In this section we examine several TALB areas in more detail to show their differing ethnic distributions.

Howick local board area

Howick local board area was one of the fastest growing TALB areas between 2006 and 2016. Figure 12 shows that almost all this growth comes from individuals who identify as Asian. Between 2009 and 2016, Howick's Asian population grew by 21,200, most of the area's 24,700 overall growth.

Figure 12



The European population increased in Howick during the first three years, although this may be largely due to improvements in the admin data rather than genuine population change. It then remains relatively constant through to 2016, in line with the 2013 ERP.

The Māori, Pacific, and MELAA ethnic groups each made up less than 7 percent of Howick's population in 2016. However, each population has been growing at a faster rate than Howick's overall population.

Manurewa local board area

Manurewa local board area has one of the most-even distributions at level 1 ethnicity. In the IDI-ERP at 30 June 2016, at least 21 percent of the population identified with each of the European, Māori, Pacific, and Asian ethnic groups.

Figure 13 shows considerable change in the area's ethnic composition between 2006 and 2016. The European population decreased by 1,500 (4.8 percent), while the remaining ethnic groups each grew by at least 15 percent.

Asian was the fastest growing ethnicity, more than doubling over the 10-year period. The Pacific population grew by 13,500 to overtake European as the most-commonly identified group. Most of the Pacific population's growth was for the Samoan ethnicity, while the change in Asian ethnicity was driven by growth in people identifying as Chinese.

We see similar trends in the ERP between 2006 and 2013, although European was still the largest ethnic group in 2013. Manurewa's Asian and European populations are slightly higher in the IDI-ERP for both periods. The ERP suggests comparatively lower growth for the Māori and Pacific populations than our admin estimates, with larger differences for the two populations in 2013.



Figure 13

Porirua city

Not all areas have a rapidly changing ethnic composition. Figure 14 shows the distribution of level 1 ethnic groups remained relatively stable in Porirua city between 2006 and 2016. Each ethnic group increased in size during that period, and results for the ERP and the IDI-ERP are very close in the census years.



Figure 14

Selwyn district

Selwyn district was the fastest growing TALB area in both the ERP and IDI-ERP between 2006 and 2016. Figure 15 shows that more than 90 percent of the population identified as European.

Unlike in Howick, most growth in Selwyn was due to an increase in the European population, with growth of 28,500 (136 percent). All other ethnic groups also more than doubled between 2006 and 2016, though from much lower bases.

Selwyn's European population is lower in the IDI-ERP than the ERP, particularly for 2006. This highlights one difficulty of using admin data to produce population estimates by area. In a high growth area such as Selwyn, there tend to be delays in reporting address changes, which can lead to underestimating these populations.



Figure 15

Discussion

The ability to produce ethnic estimates for the New Zealand population, and by subnational area, is a key requirement of any census model. This report discusses our progress in producing these estimates using linked administrative data in the IDI.

Official ethnic statistics are based on one survey source (the census) collected at one point-in-time. Statistical adjustments are used to produce the official ethnic ERP. In contrast, the ethnicity we have derived from several admin sources is an amalgam of responses obtained in different contexts at different times. The way in which we combine these different sources affects the results. We used consistency with the 2013 Census as a measure for ranking admin sources, but other ranking orders could reasonably be chosen.

The overall results reflect our expected patterns of increasing ethnic diversity – as minority ethnic group populations grow at a faster rate than the European group. The age structure and regional distribution of ethnicity are consistently in line with official measures. The possibility of producing more frequent ethnic estimates is a clear advantage of this approach.

Results for level 2 ethnicity are also encouraging. They can provide more frequent and detailed information about particular populations for which official estimates are not currently available. They also usefully highlight potential problems in how ethnicity is coded in admin sources, such as for the Fijian-Indian and New Zealand European populations. However, estimates for smaller ethnicities included at level 4 of the classification are only available from DIA and ACC.

When we look at the results in more detail we see several anomalies. Consistency of the admin IDI-ERP with the ERP varies by ethnic group. The European and Asian IDI-ERPs are slightly lower than the ERP; the Māori IDI-ERP is higher than the ERP, despite good consistency between the linked admin data and the census at the individual level.

Pacific peoples are somewhat overestimated in the admin data according to our comparisons with census responses, but even more so in the Pacific IDI-ERP. These results raise questions about the reliability of ethnicity information for the unlinked group – in both the admin data and in estimating ethnicity for New Zealand residents lacking an ethnicity response in the census.

More broadly, we have key questions that relate to the nature of the ethnicity we wish to measure in official statistics. Some admin data sources have limitations in how they collect ethnicity. While some improvement is evident over time, many limitations may be inherent in the way information is collected. In particular, admin sources (birth registrations are a key exception) under-represent people in multiple ethnic groups when compared with the census. Combining multiple records for an individual within a given source overcomes this to an extent, but inevitably results in some overstating instead. It also accentuates the effects of coding or reporting errors.

Ethnicity is not a fixed concept, and can change over time. Our current approach does not allow measurement of inter-ethnic mobility – individuals will be assigned ethnicity from the highest ranked source, regardless of when that information was provided. The lack of consistent time referencing in some admin data sources means it will be difficult to amend the approach to allow it to take account of ethnic mobility.

We also need to ask how much the consequences of admin collection contexts should influence official ethnicity statistics. The census is a relatively neutral context, but this may not be true of all admin collections.

The trade-offs between these factors will influence how we prioritise particular data sources or methods. Understanding the requirements of data users will help us further focus our approach.

We will investigate a statistical modelling approach based on a validation survey as a way to overcome the limitations of admin data.

Conclusion

The population estimates for ethnic groups that are produced from admin data give a good picture of New Zealand's ethnic make-up, with certain exceptions that we can trace to a particular collection or coding practice. The broad trends are consistent with other measures. The admin estimates also have advantages over official estimates, in frequency and level of detail. By working with agencies, we can expect to see improvements in admin collections as problems are recognised and addressed.

Improvements to our methods for selecting and locating the admin-based resident population will flow through to improvements in ethnic estimates. However, we'll still need to clearly explain why and how the results may differ from current census-based outputs.

Of more concern is the sustainability of the ranking method as a way to combine ethnicity from multiple admin sources, especially as it does not allow for changing self-perception of ethnicity over time. The main limitation of all admin sources (except birth registrations) is the under-reporting of multiple ethnic groups, which is mainly an issue for younger age cohorts. At present we have good results because we use birth registrations as the priority source for people born in New Zealand since 1998. A key question is how we allow for genuine changes to ethnic identification over time and how to do this without introducing spurious volatility.

References

Atatoa Carr, P, Bandara, D, Berry, S, Kingi, T, Grant, C, & Morton, S (2017). <u>Ethnic identification</u> <u>complexity across generations: Evidence from Growing Up in New Zealand</u>. *New Zealand Population Review 43* [PDF, 181p], **35–61**.

Black, A (2016) <u>The IDI prototype spine's creation and coverage</u>. (Statistics New Zealand Working Paper No 16-03). Retrieved from www.stats.govt.nz.

Cormack, D, & McLeod, M (2010). <u>Improving and maintaining quality of ethnicity data collections in</u> <u>health and disability sector</u>. Wellington: Te Rōpū Rangahau Hauora a Eru Pōmare.

Gleisner, F, Downey, A, & McNally, J (2015). <u>Enduring census information requirements for and about Māori</u>. Retrieved from www.stats.govt.nz.

Khawaja, M, Boddington, B, & Didham, R (2007). <u>Growing ethnic diversity in New Zealand and its</u> <u>implications for measuring differentials in fertility and mortality</u>. [PDF, 48p]. Wellington: Statistics New Zealand. Retrieved from www.stats.govt.nz.

Ministry of Health (2017). <u>HISO 10001:2017 Ethnicity Data Protocols</u>. Wellington: Ministry of Health.

Reid, G, Bycroft, C, & Gleisner, F (2016) <u>Comparison of ethnicity information in administrative data</u> <u>and the census</u>. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2005). Statistical standard for ethnicity. Retrieved from www.stats.govt.nz.

Stats NZ (2012). <u>Transforming the New Zealand Census of Population and Dwellings: Issues, options,</u> <u>and strategy</u>. Retrieved from www.stats.govt.nz.

Stats NZ (2014a). <u>An overview of progress on the potential use of administrative data for census</u> <u>information in New Zealand: Census Transformation programme</u>. Retrieved from www.stats.govt.nz.

Stats NZ (2014b). <u>Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration</u> <u>Survey</u>. Retrieved from www.stats.govt.nz.

Stats NZ (2014c). <u>Estimated resident population 2013: Data sources and methods</u>. Retrieved from www.stats.govt.nz.

Stats NZ (2014d). <u>Linking methodology used by Statistics New Zealand in the Integrated Data</u> <u>Infrastructure project</u>. Retrieved from www.stats.govt.nz.

Stats NZ (2017). <u>Experimental population estimates from linked administrative data: 2017 release</u>. Retrieved from www.stats.govt.nz.

Appendix 1: Quality of ethnicity information in the IDI

This section provides a brief quality assessment of ethnicity information from admin sources. It updates Reid et al (2016) using data from the June 2017 IDI refresh. We compare ethnicity information from individual admin data sources with the 2013 Census by using the linked census–IDI-ERP dataset.

Coverage of ethnicity information in the IDI

Table A1 shows the coverage of ethnicity information from several admin sources. The first column indicates that more than 90 percent of the IDI-ERP has ethnicity information from MOH. DIA births and both MOE datasets all have lower coverage overall, but each of these sources has very high coverage for specific age groups. Compared with the findings of Reid et al (2016), coverage had increased slightly for most sources in 2017, as more individuals enter the relevant populations from 2015 onwards. Overall, almost 99 percent of the IDI-ERP has ethnicity information from at least one source.

Table A1

Coverage of ethnicity information							
By admin source							
Data source	Percent of census-night IDI-ERP with ethnicity information	Percent of census-night IDI-ERP missing ethnicity					
DIA births	18.3	0.1					
МОН	91.8	4.7					
MOE schools	29.4	0.0					
MOE tertiary	48.1	0.5					
MSD	33.7	1.2					
ACC	80.3	3.7					
Any source	98.5						
Symbol: not applicable							
Source: Stats NZ							

Ratio of total responses, level 1 ethnicity

Table A2 shows the ratio of total responses in the admin data to the census – for each level 1 ethnic group. We only made comparisons for individuals with at least one ethnicity in both sources.

A ratio close to 1 indicates we'd expect similar results in the census and the admin source when estimating the same ethnic population, although the same individuals are not necessarily included. A value greater than 1 would indicate more people in the admin source than in the census are reporting they are that particular ethnicity. A ratio less than 1 would suggest the admin data is not capturing everyone who reported being that ethnicity in the census.

Ratio of total responses in admin sources to 2013 Census										
Level 1 ethnic groups										
	Admin source									
Ethnic group	DIA births MOH MOE MOE tertiary MOE tertiary MSD ACC (aggregated) (latest)									
European	1.00	0.97	0.91	0.99	0.96	0.94	0.93			
Māori	1.00	0.84	0.90	1.03	0.95	0.96	0.81			
Pacific peoples	1.01	0.95	0.90	1.10	1.01	0.94	0.89			
Asian	1.05	0.94	0.92	0.96	0.91	0.92	0.85			
MELAA	1.11	1.12	1.41	1.02	0.76	0.18	0.87			
Other ethnicity	Other ethnicity 1.81 4.58 14.03 120.11 69.33 112.79 127.19									
Source: Stats NZ										

Table A2

DIA births have the most similar results to the census. Aggregated MOE tertiary results are closer to the census than when we take only the latest enrolment, although Pacific peoples are 10 percent higher than in the census. Among the other sources, most ethnic groups tend to be underestimated, which is consistent with fewer ethnicities being recorded in the admin data. Of note is the low ratio (0.84) for Māori in MOH.

The pattern of results is consistent with Reid et al (2016). However, many of the ratios in table A2 are slightly higher than the corresponding tables Reid et al presented. Those previous comparisons included the small number of individuals with no ethnicity recorded in the admin data, which resulted in lower counts overall. Our differences do not imply changes in the admin data.

Ratio of total responses, level 2 ethnicity

Table A3 shows the ratio of total responses in the admin data to the census, for each level 2 ethnic group. Births again appear to be the best performing, with most ethnic groups having ratios close to 1. The aggregated tertiary enrolments and MSD also do reasonably well.

The larger ethnicities are fairly well reported; New Zealand European, Samoan, Tongan, Chinese, and Indian are relatively consistent across most sources. However, the New Zealand European population is significantly underestimated by MOE tertiary enrolments, with many of these individuals instead being included in the 'European nfd' category.

The 'not further defined' and 'other' categories tend to have larger differences. This could indicate difficulties in coding certain ethnicities, leading to more individuals being included in the residual groupings.

Table A3

Ratio of total responses in admin sources to 2013 Census

Level 2 ethnic groups

Level 2 ethnic groups							
				Admin source	!		
Ethnic group	DIA births	МОН	MOE schools	MOE tertiary (aggregated)	MOE tertiary (latest)	MSD	ACC
European nfd	1.69	3.19	4.85	81.08	42.75	17.88	6.61
New Zealand European	0.98	0.93	0.89	0.68	0.63	0.92	0.93
Other European	2.79	1.19	0.67	0.86	0.68	0.02	0.73
Māori	1.00	0.84	0.90	1.03	0.95	0.96	0.81
Pacific peoples nfd	0.00	4.73	0.00	0.00	0.00	0.00	9.06
Samoan	1.01	0.87	0.87	1.00	0.93	0.92	0.78
Cook Islands Maori	1.00	0.73	0.68	0.99	0.86	0.87	0.79
Tongan	1.00	0.87	0.88	1.01	0.93	0.92	0.79
Niuean	1.00	0.71	0.69	0.99	0.85	0.82	0.59
Tokelauan	0.92	0.74	0.75	1.09	0.91	0.94	0.56
Fijian	1.28	2.14	1.46	3.13	2.57	1.07	1.83
Other Pacific peoples	1.13	1.08	1.25	1.93	1.22	3.32	1.14
Asian nfd	1.09	5.20	6.78	15.06	5.82	37.27	9.74
Southeast Asian	1.11	0.55	0.90	0.80	0.71	0.30	0.43
Chinese	1.06	0.88	0.85	0.96	0.91	0.92	0.81
Indian	1.02	0.94	0.90	0.92	0.86	0.94	0.86
Other Asian	1.07	1.28	0.76	1.19	0.99	0.20	1.04
Middle Eastern	1.11	1.05	1.12	0.64	0.53	0.31	0.79
Latin American	1.30	0.88	0.81	0.78	0.68	0.00	0.65
African	0.97	1.43	2.09	1.82	1.45	0.00	1.21
Other ethnicity	1.81	4.58	14.03	120.11	69.33	112.79	127.19
Note: nfd not further define	ed						
Source: Stats NZ							

Appendix 2: Comparison of ranking versions

Options for prioritised ranking of ethnic sources									
Ranking order	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5	Ranking 6			
1	DIA	DIA	DIA	DIA	DIA	DIA			
2	МОН	МОН	MOE tertiary (aggregated)	MOE tertiary (latest)	MOE tertiary (aggregated)	MOE tertiary (aggregated)			
3	MOE schools	MOE tertiary (aggregated)	МОН	МОН	MOE schools	MSD			
4	MOE tertiary	MOE schools	MOE schools	MOE schools	МОН	мон			
5	MSD	MSD	MSD	MSD	MSD	MOE schools			
6	ACC	ACC	ACC	ACC	ACC	ACC			
Source: State	s NZ								

Table A4

Table A5

Ratio of total responses in combined admin data to 2013 Census							
Level 1 ethnic groups							
Ethnic group	Ranking version						
	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5	Ranking 6	
European	0.98	0.98	0.99	0.97	0.99	0.99	
Māori	0.90	0.90	0.99	0.95	0.99	0.99	
Pacific	1.01	1.01	1.06	1.03	1.06	1.06	
Asian	0.96	0.96	0.97	0.95	0.97	0.97	
MELAA	1.12	1.12	1.08	0.99	1.13	1.00	
Other ethnicity	11.31	11.69	65.50	36.71	66.47	75.64	
Source: Stats NZ							

Ratio of total responses in combined admin data to 2013 Census							
Level 2 ethnic groups							
Ethnic group	Ranking version						
	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5	Ranking 6	
European nfd	5.87	5.88	37.72	20.68	37.74	38.58	
New Zealand European	0.93	0.93	0.82	0.79	0.82	0.82	
Other European	1.31	1.31	1.21	1.13	1.18	1.12	
Māori	0.90	0.90	0.99	0.95	0.99	0.99	
Pacific peoples nfd	3.37	3.37	1.28	1.28	0.92	0.85	
Samoan	0.94	0.94	0.99	0.96	0.99	0.98	
Cook Islands Maori	0.87	0.87	0.96	0.91	0.96	0.97	
Tongan	0.95	0.95	0.99	0.96	0.99	0.99	
Niuean	0.86	0.86	0.96	0.90	0.96	0.96	
Tokelauan	0.84	0.84	0.97	0.90	0.97	0.99	
Fijian	2.12	2.13	2.59	2.33	2.55	2.45	
Other Pacific peoples	1.23	1.24	1.52	1.21	1.54	1.74	
Asian nfd	5.46	5.39	8.69	4.76	8.58	10.31	
Southeast Asian	0.64	0.64	0.74	0.70	0.79	0.71	
Chinese	0.91	0.91	0.96	0.93	0.96	0.96	
Indian	0.95	0.95	0.94	0.92	0.94	0.94	
Other Asian	1.22	1.23	1.22	1.11	1.17	1.14	
Middle Eastern	1.03	1.03	0.84	0.78	0.84	0.74	
Latin American	0.93	0.94	0.91	0.86	0.90	0.87	
African	1.41	1.41	1.60	1.41	1.77	1.53	
Other ethnicity	11.31	11.69	65.50	36.71	66.47	75.64	
Note: nfd not further defined							
Source: Stats NZ							

Table A6

Table /	47
---------	----

Total responses, combined admin data and 2013 Census						
Level 2 ethnic groups						
Ethnic group	Combined admin data (ranking 3) 2013 Census		Ratio of combined admin data to 2013 Census			
European nfd	906,720	24,036	37.72			
New Zealand European	2,177,904	2,652,918	0.82			
Other European	286,098	236,814	1.21			
Māori	553,881	559,230	0.99			
Pacific peoples nfd	1,155	900	1.28			
Samoan	131,796	133,773	0.99			
Cook Islands Maori	55,305	57,336	0.96			
Tongan	54,537	55,092	0.99			
Niuean	21,153	22,101	0.96			
Tokelauan	6,480	6,678	0.97			
Fijian	34,713	13,428	2.59			
Other Pacific peoples	14,718	9,684	1.52			
Asian nfd	36,717	4,227	8.69			
Southeast Asian	52,896	71,400	0.74			
Chinese	144,186	150,375	0.96			
Indian	135,423	143,520	0.94			
Other Asian	73,122	60,060	1.22			
Middle Eastern	15,510	18,486	0.84			
Latin American	10,491	11,583	0.91			
African	19,998	12,465	1.60			
Note: nfd not further defined						
Source: Stats NZ						