Comparison of ethnicity information in administrative data and the census

Census Transformation

Giles Reid, Christine Bycroft, and Fran Gleisner

Liability
While all care and diligence has been used in processing, analysing, and extracting data and information in this publication, Statistics New Zealand gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Disclaimer
The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand.

The opinions, findings, recommendations, and conclusions expressed in this paper are those of the author(s), not Statistics NZ.

Access to the anonymised data used in this study was provided by Statistics 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 have been 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. Further detail can be found in the Integrated Data Infrastructure extension: Privacy impact assessment (6th ed) available from www.stats.govt.nz.

Note: All figures presented in this paper have been rounded to the nearest hundred to protect confidentiality.

Citation
Reid, G, Bycroft, C, Gleisner, F (2016). *Comparison of ethnicity information in administrative data and the census.* Retrieved from www.stats.govt.nz

ISBN 978-0-908350-46-9 (online)

Published in June year by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact
Statistics New Zealand Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4600
www.stats.govt.nz

Contents

List of tables and figures

List of tables

List of figures

1 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 focus in the short-to-medium term on modernising the current census model and making it more efficient
- a longer-term focus on investigating alternative ways of producing small-area population and social and economic statistics. This includes the possibility of changing the census frequency to every 10 years, and exploring the feasibility of a census based on administrative data (Statistics New Zealand, 2012).

The next census in 2018 will be significantly modernised, including an online completion target of 70 percent and re-use of administrative data to support collection and processing.

Continuing to meet critical information needs must underpin decisions on the future of census. Investigations into the long-term direction for census are focused on developing an understanding of future census information requirements, and the ability of administrative sources to meet those requirements.

Read more about Census Transformation in New Zealand.

About this paper

Ethnicity is a core demographic variable for describing the New Zealand population. Ethnicity data collected by the census is widely used and is the basis for official population statistics by ethnicity. In an administrative-based census, ethnicity would have to be obtained from administrative sources. This paper compares ethnicity data from the 2013 Census with the ethnicity information collected by administrative sources currently available in Statistics NZ’s Integrated Data Infrastructure.

We describe how the main administrative sources collect ethnicity data, and compare this against the formal ethnicity standard: the statistical standard for ethnicity. We show rates of agreement between administrative sources and the census using the 2013 Census data linked to the administrative sources. Finally, we discuss the problem of combining ethnicity information from multiple administrative sources and evaluate some alternatives.

This paper is one of a series of investigations as part of the Census Transformation programme. The programme explores the potential for administrative data sources to provide census-type information.

2 Introduction

What is ethnicity?

According to the statistical standard for ethnicity (Statistics NZ, 2005a), ethnicity relates to the ethnic group or groups that people identify with or feel they belong to, and is a measure of cultural affiliation. Ethnicity is self-perceived and a person can belong to more than one ethnic group.

Why is ethnicity information important?

The ability to produce ethnicity from administrative sources is a key consideration when determining the feasibility of a census based on administrative data. Ethnicity is the principal measure of cultural identity in New Zealand, and is used across the Official Statistics System. The collection of ethnicity information in the Census of Population and Dwellings is a legislative requirement under the Statistics Act 1975, and the census is an important source of ethnicity data for small areas and small ethnic groups. The census provides the basis for official population estimates of major ethnic groups, and for population projections for Pacific Island, Māori, and non-Māori populations.

Ethnicity data are widely used with other characteristics of the population to inform resource allocation, policy development, and research. The major uses of census ethnicity information and ethnic group population estimates and projections are:

- to monitor the changing ethnic diversity of New Zealand's population at the national, regional, and local levels so that services can be appropriately targeted
- to provide denominators for calculating rates by ethnicity for topics such as fertility, mortality, morbidity, and crime
- to derive measures for monitoring the well-being of ethnic groups, particularly in the health sector (eg morbidity rates, immunisation rates)
- to monitor the demographic, social, and economic progress of ethnic groups
- to assist in the planning of services directed at the special needs of ethnic groups in areas such as education, housing, health, and social welfare
- to evaluate the impact of government policies on the economic and social well-being of ethnic groups
- to assist in the allocation of funds from government agencies to ethnic groups.

What are the challenges of measuring ethnicity?

Official New Zealand standards and classifications are available, and these are used by Statistics NZ surveys and the census, and by some administrative sources. Despite this, collection of ethnicity information is challenging for several reasons.

Ethnic group changes with context

A person may give a different response depending on the context. For example, when filling in a self-administered form a person may respond differently to when asked his/her ethnic group by an interviewer. The social or cultural setting may also affect the ethnicity response reported.

Ethnic group changes over time (ethnic mobility)

The ethnic group or groups that someone identifies with may change over time. Longitudinal surveys and administrative databases must allow for ethnic mobility. Ethnic mobility also affects the integration of different datasets, as the same person may have given different answers in different collections. Rather than using both datasets' responses, the decision on what to use must be made case by case.

Multiple ethnicity

People may identify with more than one ethnic group, so provision to collect multiple ethnic groups for each individual is essential. The statistical standard recommends the collection of up to six ethnic group responses per person. The ability to collect three responses is the minimum requirement to meet the standard.

Previous comparisons

Previous comparisons of ethnicity between census and administrative data have been undertaken by Blakely, Atkinson, and Fawcett (2008) and Tan, Blakely, and Atkinson (2010), who compared the 2001 Census with mortality data from 2001–04 and 2004–06. Using a total responses measure of ethnicity (a count of the number of people in each ethnic group, regardless of multiple response) they found generally close agreement between the census and mortality data. However, they found that fewer people had multiple ethnic groups recorded in mortality records than in census data, resulting in greater sole Māori counts on the mortality data.

Statistics NZ (2005b, 2014c) compared birth and death registrations for children who died before their fifth birthday. Both studies found fewer multiple ethnic responses recorded on death registrations compared with birth registrations.

O'Byrne, Bycroft, and Gibb (2014) provides a summary of early findings about the potential use of administrative data sources for census information. However, this first assessment was based on metadata and intended to be indicative only. No analysis of the data itself was undertaken. This paper builds on the finding from O'Byrne et al that ethnicity was 'likely' to be satisfied by administrative data and suggested several potential administrative sources of ethnicity information.

Aims and scope

This paper describes preliminary analysis of ethnicity information from linked administrative sources available in Statistics NZ's Integrated Data Infrastructure (IDI).

The overall aim is to investigate the potential of producing estimates of ethnic populations from linked administrative sources, using Statistics NZ's Integrated Data Infrastructure (IDI) as a test environment.

Three research questions guide this work:

- What ethnicity information is available from linked administrative data?
- What is the quality of ethnicity information in the IDI?
- What would be required to improve the potential for producing estimates of ethnic population from linked administrative sources?

This paper provides reference information about the statistical concepts and about administrative data sources relevant to ethnicity, and presents findings from analysis comparing information in the census with administrative sources.

We investigate the potential for administrative data to provide ethnicity information comparable to the current census and the official series of ethnic population estimates, with the latest 2013 Census as the base reference.

We include level 1 of the standard classification of ethnicity and do not extend to more detailed ethnicity breakdowns. Level 1 is commonly used in reporting and for public policy. We limited the administrative sources investigated to those available in the IDI in May 2015.

3 Classification of ethnicity in New Zealand

Statistical standard for ethnicity

A statistical classification is a way to group a set of related categories in a meaningful, systematic, and standard format. New Zealand statistical standards and classifications are designed for use across official statistics collections, both for Statistics NZ and other agencies. We compare the collection of ethnicity in the 2013 Census and the relevant administrative data sources with the official ethnicity standard and classification.

The statistical standard for ethnicity defines ethnicity as follows:

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.

An ethnic group is made up of people who have some or all of the following characteristics:

- a common proper name
- one or more elements of common culture which need not be specified, but may include religion, customs, or language
- unique community of interests, feelings, and actions
- a shared sense of common origins or ancestry
- a common geographic origin.

Classification of ethnicity

The 2005 New Zealand standard classification of ethnicity is a hierarchical classification of four levels. Level 1 of the classification has six categories and is used solely for output, not for collection. Apart from Māori, level 1 categories are ethnic groups, 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 the larger ethnicities within the level 1 groups – for example New Zealand European, Samoan, Indian, and African. 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.

When collecting ethnicity information, people need to be able to state their specific ethnicities without being forced to identify themselves in a more general category grouping of ethnicities. The standard is designed so that detailed ethnic group information can be collected and responses can be classified to specific ethnic group categories at a detailed level of the classification. Where it is not possible to collect data at level 4 of the classification, for instance in administrative data collections where written responses are not able to be coded, ethnic group information should be collected at level 2 of the classification, which is less detailed.

Output

The presence of multiple ethnicities for the same person needs to be taken into account 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:

- Total response data, in which individuals are counted in all of their reported ethnic groups. A person who reported their ethnicities as, for example, English, Irish, and Māori, will be counted once in the European category and once in the Māori category for level 1 outputs. The number of grouped total responses will be greater than the total population, as individuals can provide more than one response.
- Single and combination data, which counts people in mutually exclusive categories. People reporting two or more ethnic groups are counted once in the relevant 'combination' group. This means that the total number of responses equals the total number of people who stated their ethnicity. In the above example, this person would be counted once in the ‘European and Māori’ combination at level 1.

Total response data, grouped total response data, and single/combination data are considered the best means of outputting ethnicity data (Kukutai & Statistics New Zealand, 2008).

We use both total response classifications and single/combination classifications in this paper.

A further type of output, largely discontinued following the 2004 Review of the Measurement of Ethnicity (Statistics New Zealand, 2004) but still used by some administrative collections, is called prioritisation of ethnicity. This ensures that the total number of responses equals the total population. In doing so, prioritisation conceals diversity within and overlapping between ethnic groups by eliminating multiple ethnicities from data (Statistics New Zealand, 2006). This systematic prioritisation of the data gives highest priority to Māori – meaning, for example, an individual who might self-identify as both Pacific and Māori would be counted as Māori.

4 Data sources

This chapter describes the data sources used in this investigation: the New Zealand Census of Population and Dwellings, the relevant administrative sources in the IDI, and the linked Census-IDI data.

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. Census has a wide range of uses within and outside government. The latest census was held in March 2013.

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.

For this investigation we are only interested in New Zealand residents, not those visiting New Zealand temporarily on census night.

In the 2013 Census the net undercount varied by ethnic group. The highest undercount was for Māori (6.1 percent), followed by Pacific (4.8 percent), Asian (3.0 percent), and European (1.9 percent) (Statistics New Zealand 2014a).

Ethnicity information in the census

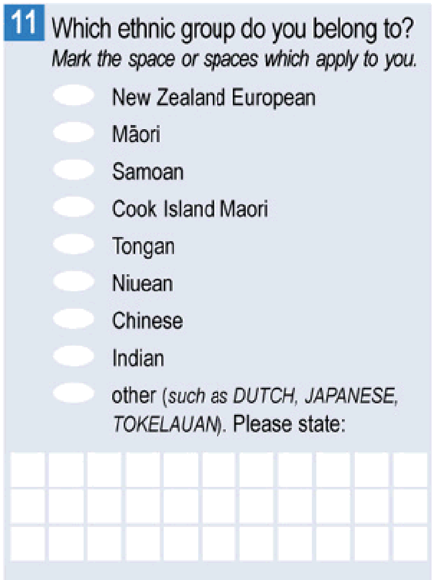
The census uses the statistical standard and classification for ethnicity described above.

Ethnicity is a 'foremost' variable in the census, which means that it is managed to produce information of highest quality. The non-response rate for ethnicity for usual residents who returned a form in the 2013 Census was 0.7 percent. If the substitute forms created to account for people who did not fill out a form are included, the non-response to ethnicity was 5.4 percent.

Figure 1 shows the ethnicity question for the 2013 Census. Up to six responses per person are recorded.

Figure 1

The ethnicity question in the 2013 Census



11 Which ethnic group do you belong to?

Mark the space or spaces which apply to you.

○ New Zealand European

○ Māori

○ Samoan

○ Cook Island Maori

○ Tongan

○ Niuean

○ Chinese

○ Indian

○ other (*such as DUTCH, JAPANESE, TOKELAUAN*). Please state:

Integrated Data Infrastructure (IDI)

Statistics NZ developed the IDI as an environment in which to link multiple data sources in a systematic and secure way. It was developed to produce official statistics outputs and to allow Statistics NZ staff and external researchers to conduct policy evaluation and research on people's transitions and outcomes. The IDI contains administrative and survey datasets, linked at the individual level. The IDI continues to change as new datasets are added. This section describes the structure and content of the IDI in May 2015.

The structure of the IDI is shown in figure 2, and can be described as a central 'spine' to which a series of data collections are linked. The target population for the spine is all individuals who have ever been residents of New Zealand.

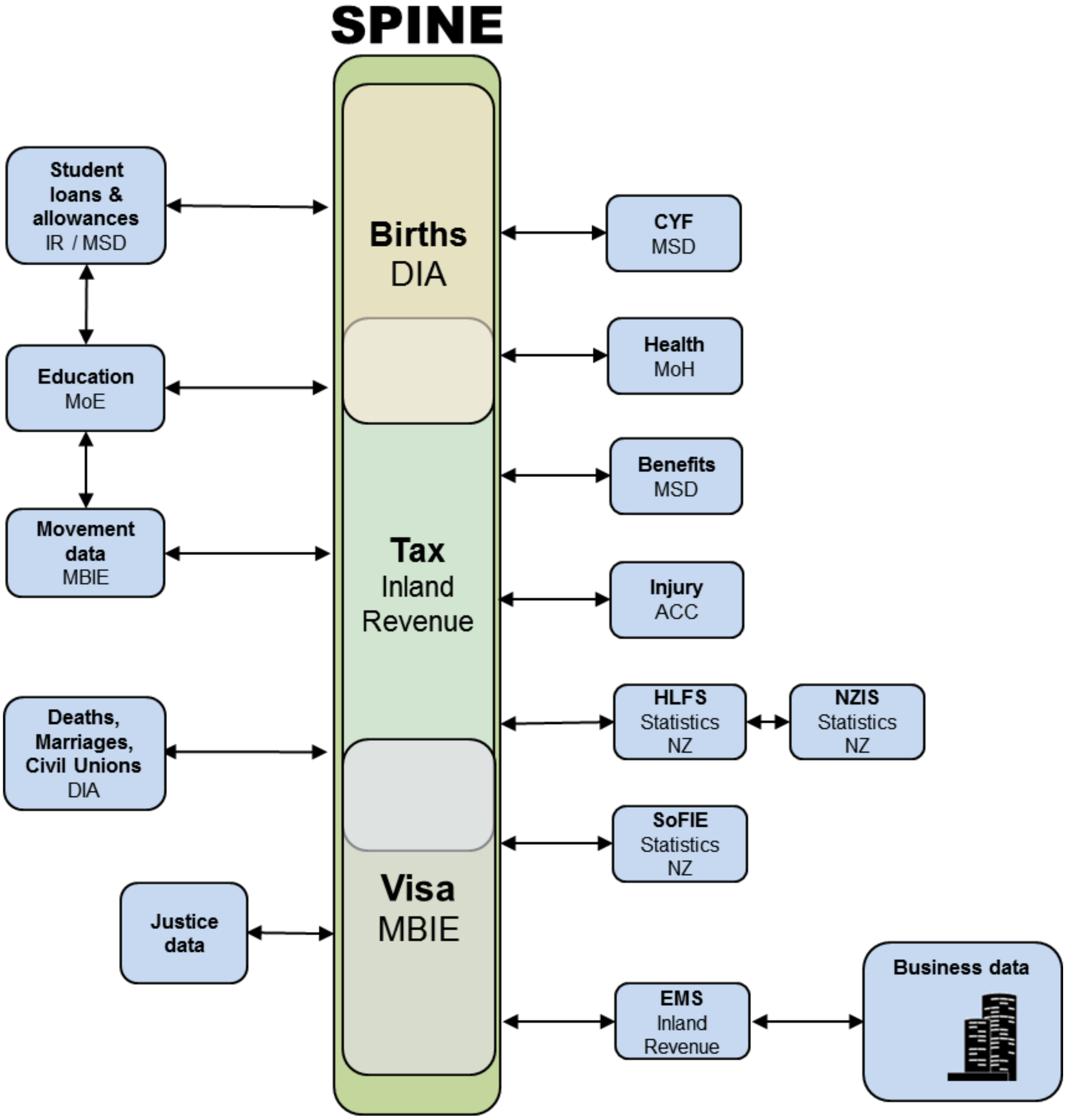
Three data sources are linked together probabilistically to create the spine:

- a list of all IRD numbers that have been issued by Inland Revenue (IR)
- a list of all births registered in New Zealand since 1920
- a list of all visas granted to migrants from 1997 (excluding visitor and transit visas).

Other data sources are linked to the IDI spine, and cover a wide range of subject areas. Statistics New Zealand, 2014b describes the linking methodology. Priority is placed on obtaining a high precision rate, ie minimising creating erroneous links, with the trade-off that more correct links may be missed. In practice, linkages are designed so that under 2 percent of links made are erroneous.

The IDI also contains summary tables that provide core information about individuals (age, sex, ethnicity, geographic location) summarised from across the available data sources.

Figure 2
Structure of the Integrated Data Infrastructure (IDI) in May 2015



Ethnicity information in the IDI

Ethnicity information in the IDI is contained within data collections from several government agencies. The dataset descriptions below are primarily based on Cormack (2010) and Cormack & McLeod (2010), which provide a thorough background to official collections of ethnicity.

Accident Compensation Corporation (ACC)

ACC is a Crown entity set up to deliver New Zealand's personal no-fault injury insurance scheme as set out in the Accident Compensation Act 2001.

Ethnicity data collected by ACC is used to produce injury statistics and to monitor access to the services provided by ACC. It has been collected since 1997, although at that time only one ethnicity was collected, and this collection was only done for a limited number of claims. Since 2001, ACC records up to three ethnicities, using level 2 of the standard classification of ethnicity from the 1996 ethnic standard. The question asked on the form is not standard, and includes a tick box option for ‘I’d prefer not to say’, which approximately 7 percent of respondents tick (Cormack & McLeod, 2010).

The ACC data within the IDI at May 2015 includes only claims made for work-related injuries.

Ministry of Education (MoE)

MoE collects information on ethnicity from providers of early childhood, primary, secondary, and tertiary education. This information is generally collected on enrolment forms and is used to produce a range of information and statistics (eg student participation for different ethnicities).

MoE uses Statistics NZ's definition of ethnicity, and since 2007, has recorded ethnicity as a numeric code using level 3 of the standard classification of ethnicity from the 2005 Ethnic Standard. All enrolment forms should allow students to identify with up to three ethnic groups; however, the Ministry requires some data providers to report a student as being in one ethnic group only. MoE uses Statistics NZ’s prioritisation method outlined in the 1996 ethnic standard to decide which ethnic group to use when a student identifies with more than one ethnic group.

The ethnicity question(s) on enrolment forms can differ between providers. Although the ministry provides guidelines on their website, it is likely that questions are not consistent with each other or with the census question.

Ministry of Health (MoH)

Information about ethnicity has been collected for several years in the health sector, with varying degrees of standardisation and completeness. Several key collections hold ethnicity, including the National Health Index (NHI), and several registries/databases such as the New Zealand Cancer Registry (NZCR), and the National Minimum Dataset (NMD). Ethnicity information is usually collected during contact with a health service or health provider, which can affect the quality and completeness of ethnicity information in the key collections and databases.

Since 1996, MoH has aligned its collection of ethnicity with Statistics NZ’s approach with the key collections holding at least one ethnicity for each individual (mandatory, ‘principal’ ethnicity), and having the ability to hold up to three ethnicities. Before 1996, only one was recorded. The introduction of the *Ethnicity Data Protocols for the Health and Disability Sector* in 2004 was a significant development for the health sector. The protocols provided guidance for standardising data collection and outputs across the health and disability sector.

Since 2008, MoH has aligned the health sector with the Statistics NZ standard classification of ethnicity from the 2005 ethnic standard, including the use of consistent level 1 codes. Ethnicity data in the NHI collection is recorded at level 2 of the Statistics NZ standard classification of ethnicity from the 2005 ethnic standard, and up to three ethnic groups are recorded per individual.

The MoH data in the IDI includes several different tables holding ethnicity information. In this study we used the combined NHI dataset, which is a unified national person list compiled by MoH. Te Rōpū Rangahau Hauora a Eru Pōmare, based at University of Otago, Wellington, has published a series of reports and discussion papers about ethnicity data in New Zealand with a particular focus on the health sector (see Publications of Te Rōpū Rangahau Hauroa a Eru Pōmare). These give a good summary of different ethnicity collections in the health sector and the quality of that information.

Ministry of Social Development (MSD)

MSD collects ethnicity information for individuals obtaining Work and Income services (benefits). Ethnicity can be collected on application forms, or through other interactions with Work and Income (eg in person, online, or through call centres). However, it is not a compulsory field because it is not related to entitlement or eligibility for assistance. Ethnicity information is needed, however, to understand how access to benefits and social welfare is related to disability, access to health care, and health outcomes for Māori.

Ethnicity data has been collected since 1991 by MSD; however, they have used several different systems and classifications. Information in the IDI is available since 1993. Since the late 1990s, the collection of ethnicity information has been more consistent. This improvement was mainly due to the introduction of the SOLO system by Work and Income. This system is used to record information about job seekers and the provision of employment services and allows for individuals to identify with up to three ethnic groups at level 3 of the classification.

The different collections across MSD vary in their adherence to the statistical standard. For example, the question used to collect ethnicity on application forms for financial assistance (benefits) varies – both in the question and the categories used for responses. The voluntary nature of the question and the variability of questions are likely to affect the quality of the ethnicity data collected by MSD.

Department of Internal Affairs (DIA)

The Department of Internal Affairs is responsible for birth registrations, and records go back as far as the 19th century. Until 1962, separate registers were kept for Māori births. The Māori birth register included tribe, residence, and iwi details completed by the parents; however, for the most part these fields have not been digitised.

Between 1962 and September 1995, information was collected on “the degree of Māori or Pacific Island blood and the tribe or island of the newborn's mother and father” (Statistics NZ, 2015). Parents who were not of Māori or Pacific Island descent were not asked to provide any ethnicity information. A new birth registration form was introduced in September 1995. It included an ethnicity question consistent with the concept of ethnic self-identification. The registration form includes ethnicity questions for the mother, father, and child. This form has since been updated to align with the 2005 ethnicity standard.

Since 1998, birth registrations have been recorded digitally and considerable effort has been put into response rates and data quality.

Death registrations also contain ethnicity information, but they were not part of this study.

Statistics NZ survey collections

Some of Statistics NZ’s household survey collections are also included in the IDI. Ethnicity information in these surveys uses the ethnicity standard and is typically very good quality, but the number of people covered by these surveys is relatively small compared with the administrative sources. For this reason, they were not investigated as individual sources in this paper.

Personal details table

Within the IDI, business rules are applied to standardise the ethnic codes received from each agency.

The six level 1 categories from each selected administrative source are summarised in a ‘personal details table’ for each individual. Ministry of Justice data is excluded due to quality concerns. Some individuals do not have any ethnicity recorded in the IDI – for example, if they have not interacted with an agency that collects ethnicity.

As a result of this process, an individual's ethnicity information in the personal details table is a combination of the original responses given to separate agencies, coded to level 1 of the 2005 ethnic standard. An ethnic group is included wherever it is captured by any agency, at any point in time (ever-recorded). It is not possible to directly identify the source(s) of ethnicity, or the date it was captured, in the personal details table, but ethnicity responses for each dataset can be examined individually.

Linking the census to the IDI

To enable individual-level comparisons between the ethnicity information in the IDI and the ethnicity information in the census, the census has been linked to the IDI at the individual level. This link was created by Census Transformation in May 2015. The linking was done to better understand the coverage and quality of census information in the IDI, and the linked data was only available to approved Statistics NZ staff working on the Census Transformation programme.

The census was linked to the May 2015 version of the IDI spine. Linking was completed in Quality Stage using probabilistic matching techniques. The variables used in the linkage process were full name, date of birth, sex, meshblock of usual residence, and country of birth.

Overall, 3,920,364 (or 92 percent of) census usual residents were linked to the IDI. Of most interest for this paper, 95 percent of census records for New Zealand residents in households where forms were returned were linked to the IDI. The match rate was much better for individuals who had used electronic forms (98 percent linked) compared with paper forms (93 percent linked). There were around 250,000 individuals in the census who provided an ethnicity but for whom a link could not be found in the IDI.

The links in this dataset have an estimated false positive rate of less than 1 percent (a false positive is when an incorrect link has been made between two different individuals).

5 Methods for examining the quality of ethnicity information

Understanding causes of error in ethnicity data

The concepts of coverage error and measurement error provide a framework for assessing the accuracy of data sources (Zhang, 2011).

Coverage describes the relationship between the ideal target population and the actual set of people present in a dataset. For the census ethnicity variable, the population of interest is all New Zealand residents. Aggregate-level comparisons are most useful in providing insight into differences in coverage.

Measurement errors cause a recorded response to differ from its true value. If these errors are not random they may result in a systematic bias. Measurement error may occur when administrative definitions, concepts, or questions do not align well with the statistical concept being measured. Measurement errors in both the census and administrative data may also be due to errors within the respective collection and processing systems, and may result in missing or incorrect information. The individual-level comparisons can inform our understanding of measurement error.

The ability to integrate information with other sources through linking the same units also affects accuracy. Linkage errors are of two types: links may be missed (eg if a person's name is recorded differently on different files); or two different people may be wrongly linked (eg if their names and dates of birth are very similar). Linkage errors may reduce the coverage of an administrative source (no information is available if links are not made when they should be), or they may introduce measurement error if the wrong people are linked together.

Evaluating the quality of administrative sources of ethnicity

This investigation uses the following methods to evaluate the quality of the ethnicity information in IDI.

Comparison of concepts and definitions

The concepts and definitions of ethnicity used in the IDI and its individual data collections are compared to the statistical standard for ethnicity. Ideally the concepts and definitions should be consistent across collections and consistent with the standard.

Comparison of aggregate counts

Aggregate comparisons are used to examine the coverage of the administrative sources, and to compare total responses for each administrative source with the census. Analysis is restricted to those individuals in the linked census-IDI dataset.

Comparison of individual-level information

The ethnicities recorded for an individual in the IDI are compared against those recorded for the same individual in census. These comparisons can only be made for the group of people who had records in the IDI and the census which were linked together, and for whom an ethnicity was recorded in both the administrative source in the IDI and the census.

Close agreement of responses in administrative data and the census is a strong suggestion that the measurement in both sources is good. However, when responses are different, it is harder to determine which is likely to be the correct response. There are several reasons why an individual might record different ethnicity responses in the census and the IDI, and not all indicate errors in one source. People can identify with different ethnic groups over time, or in different contexts. Because questions on different administrative forms can be slightly different, this may prompt different responses from a person, which are all correct from their point of view.

While erroneous linkages are kept to a minimum, linkage errors could explain a small proportion of cases where ethnicity information is found to be different between the census and the administrative sources in the IDI. Apart from birth registrations, which form part of the spine, two linkages are involved in the comparison of census ethnicity and ethnicity in administrative sources: the linkage between the census and the IDI spine, and between administrative sources and the IDI spine.

Treatment of ‘New Zealander’ response

For comparability with the estimated resident population and administrative sources, the ‘New Zealander’ response has been included in the ‘European’ category in this investigation.

In the standard classification ‘New Zealander’ is coded to ‘Other ethnicity’, and this approach is used in the 2013 Census. However, the official estimated resident population series codes the ‘New Zealander’ response to ‘European’.

On the whole, administrative sources in the IDI do not have ‘New Zealander’ as a response. Current usual practice in the health sector is to code ‘New Zealander’ to the ‘European’ category (Cormack & McLeod, 2010). This is likely to be similar across other administrative collections.

6 Results

The results are divided into four categories:

1. consistency with the standard for ethnicity
2. coverage and missing data
3. comparison of aggregate counts
4. comparison of individual-level records between the census and administrative sources.

Consistency with the standard for ethnicity

The census collection of ethnicity is consistent with the statistical standards described in section 3. Most administrative sources also capture the same concept of ethnicity as cultural affiliation, and aim for self-identification where possible. Most, but not all, sources record multiple responses. On the whole, up to three ethnic groups are recorded, the minimum requirement of the standard.

The 2005 standard classification appears to be used by most administrative data sources, with all the larger sources capturing responses between level 4 and level 2. Some administrative sources use prioritised ethnicity for reporting – a practice that is not consistent with the standard.

The question asked for ethnicity differs widely across administrative data sources, and often differs within each administrative source depending on the mode of collection or the form used.

All collections, both census and administrative, rely on the respondents' understanding of the concept of ethnicity. This may depend on the context, and may vary over time.

Because a person may change the ethnicity or ethnicities they identify with over time, the time reference of a source collection is important. The census measures ethnicity at a single point in time (census day), and official population estimates also measure ethnicity at a given reference date. In contrast, many administrative sources reflect information collected from individuals at different points in time. Birth registrations record ethnicity for a single event in a person's life, while other sources may capture ethnicity at multiple times depending on contact with the agency. What is recorded may represent the latest value, or an agency may accumulate ethnicities over time. Administrative systems do not always report the date at which ethnicity was collected.

These differences in the timing of data collection introduce conceptual differences between the measurement of ethnicity used currently in the census and official population estimates, and any measures of ethnicity produced from administrative sources.

Table 1 summarises the collection of ethnicity information across the administrative sources and the census.

Table 1

Key elements of ethnicity in the statistical standard as applied in the census and for administrative sources in the IDI

Consistency with statistical standard	Census	DIA births (from 1998)	ACC	Ministry of Education		Ministry of Health	MSD
			Claims	Schools	Tertiary	NHI	Benefits
Self-identified cultural affiliation?	Y	Y	Y	Y	Y	Y	Y
Is multiple response recorded?	Y	Y	Y	Y	Y	Y	N
Statistics NZ 2005 ethnic classification?	Y	Y	Y	Y	Y	Y	N
What level of classification is used?	Level 4	Level 4	Level 4	Level 2	Level 3	Level 2	Level 2 equivalent
Is the question consistent?	Y	Y	N	N	N	Y	N

Key:
Y = consistent with statistical standard
N = inconsistent with statistical standard

Coverage and missing data

This section examines the availability of ethnic data in each administrative source and for all sources combined, by age and sex, and by ethnic group.

The census includes all age groups. The administrative sources in the IDI currently provide information about different age groups depending on the source. Individuals of all ages access the New Zealand health care system and have the potential to be captured in the Ministry of Health data. The MSD benefits source has low coverage, because only a small proportion of adults receive working-age benefits. The education system provides information about children from five years of age. Births data covers all children born in New Zealand since 1995, although we restrict our analysis to records since the digitisation in 1998.

Table 2 shows the coverage of each source, with the base population being the total linked Census-IDI population. The low coverage rates for education and births data are partly because data is only available for recent years.

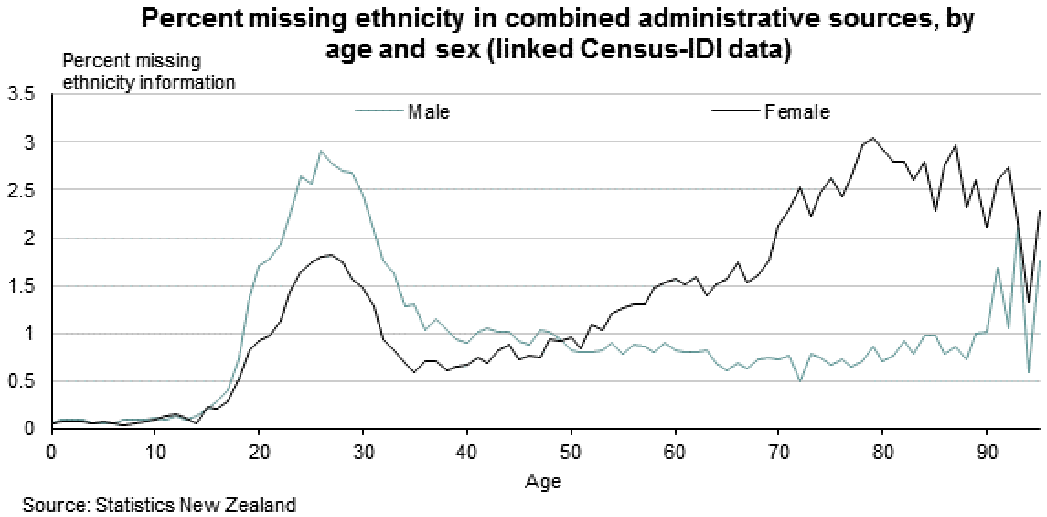
Table 2

Coverage of ethnicity information in each main administrative source (linked Census-IDI data)

Data source	**% of linked IDI-Census population with ethnicity information**
Birth registrations	18
Ministry of Education (schools)	27
Ministry of Education (tertiary)	44
Ministry of Social Development	32
ACC (claims)	47
Health (NHI)	98

Overall, 99 percent of individuals in the linked Census-IDI data have at least one ethnic code recorded in the IDI personal details table. However, the proportion of individuals with ethnicity information recorded varies with age. Figure 3 shows the percentage of individuals in the linked Census-IDI data who have no ethnicity recorded in any administrative source, by age and sex.

Figure 3



Percent missing ethnicity in combined administrative sources, by age and sex (linked Census-IDI data)

Source: Statistics New Zealand

Some ethnicity information is available for more than 97 percent of individuals in the linked Census-IDI at every year of age. There are very few children missing ethnicity data because nearly all the children in the IDI spine come from recent birth records, which have very low rates of non-response. Education records are also good quality for recent years. From around age 19 to 35 there is an increase in missing ethnicity. There is a gradual increase in missing ethnicity from age 50 up for linked females, which is not seen in linked males.

Table 3 shows the number and percentage of respondents for each major ethnic group from the census that have no ethnicity information in the IDI.

Table 3

Percent missing ethnicity in combined administrative sources, by ethnic group, (linked Census-IDI data)		
Ethnic group	**Responses in the census, but missing in the IDI**	
	Number	**Percent**
European	20,400	0.69
Māori	2,800	0.50
Pacific peoples	2,400	0.85
Asian	12,100	2.65
MELAA	1,200	2.61
Note: MELAA = Middle Eastern/Latin American/African		

Overall, a very high proportion of individuals across all ethnic groups have ethnicity information from some administrative data in the IDI. The level of missing ethnicity data varies somewhat across ethnic groups. Māori, European, and Pacific peoples have the lowest rates of missing ethnicity, all under 1 percent. The Asian and MELAA (Middle Eastern/Latin American/African) groups have higher levels of missing data, at around 4 percent.

Comparing aggregate counts

We first compare total responses output for level 1 ethnic groups in the census and for each administrative source. In table 4 we show the ratio of total responses in the administrative source to the census. We only use records from the particular administrative source that have been linked to the census. A ratio close to 1 shows that similar results could be expected between census and the administrative source when estimating ethnicity for the same population group.

Table 4

Comparison of total response counts between census and each administrative source						
Ethnic group	**Administrative source to census ratio for total responses to ethnic group**					
	Births	**Health**	**Tertiary**	**Schools**	**MSD**	**ACC**
European	0.98	0.90	0.93	0.91	0.90	0.90
Māori	0.99	0.79	0.92	0.89	0.93	0.74
Pacific peoples	1.01	0.91	0.98	0.89	0.91	0.87
Asian	1.04	0.89	0.88	0.92	0.89	0.79
MELAA	1.03	1.07	0.65	1.42	0.18	0.74
Other	27.27	3.67	73.39	14.06	212.33	126.56
Note: MELAA = Middle Eastern/Latin American/African						

In most cases ratios are less than one, indicating fewer people with each ethnic group in the administrative sources than in the linked census records. In all datasets, the ‘Other’ group is much more prevalent in the administrative data. This is most likely due to inconsistencies in coding. The ratios for MELAA ethnicities are also poor in several data sources, which again may indicate a problem with coding.

Birth registrations have ratios closest to 1, indicating a good agreement at an aggregate level. The other datasets all have at least one ethnicity that is problematic, though there does not appear to be a consistent pattern across ethnic groups.

Comparing individual-level responses

Because the 2013 Census records have been linked to the IDI, we can compare census responses to those recorded in administrative sources for each individual. This analysis uses an individual’s ethnicity in the census as a benchmark. What we are analysing is disagreements between census and administrative responses, which can generally be referred to as measurement errors.

In some cases the disagreements may not result from mistakes or misreporting. For example, an individual may not have identified as Māori when they filled out the forms for educational enrolment in 2002, but by the time of the 2013 Census, had come to identify as Māori.

Comparing ethnic group total responses

We first compare ethnicity using total responses for level 1 ethnic groups. We want to know, for each source and each ethnic group, how closely the ethnicity responses agree with census responses.

Table 5 shows one example of an administrative source compared with the census. Any individual with an ethnicity response in both sources must have either a ‘yes’ or a ‘no’ response in each source. If the various errors and conceptual difficulties described above did not exist, we would expect everyone to have either a ‘yes’ in both sources or a ‘no’ in both sources.

Table 5

Comparison of Māori ethnic response in census and Ministry of Health data

		Census Māori ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Health Māori ethnicity	Yes	411,800	11	28,900	1	440,700	12
	No	149,400	4	3,203,200	84	3,352,600	88
Total		561,200	15	3,232,100	85	3,793,300	100

Table 5 shows that while 95 percent of people have the same responses in the two datasets, around 5 percent have different responses. This difference in classification has a greater impact on the smaller ethnic group. The under-reporting of Māori in health data compared with the census seen in the aggregate comparisons in table 4 is largely because over one-quarter of people with Māori ethnicity according to the census do not have Māori ethnicity in the Ministry of Health data. A smaller group do have Māori ethnicity in the health data, but not in the census. We can carry out the same type of analysis for each administrative source and ethnic group.

Each panel in figure 4 shows the percentage agreement between a source in the IDI and the census responses linked to that source.

The percentages in the left-hand panels are calculated as:

$$100\% \times \frac{\text{Number of linked people answering 'yes' in both sources}}{\text{Total number of linked people with a 'yes' response in census}}$$

And on the right-hand side:

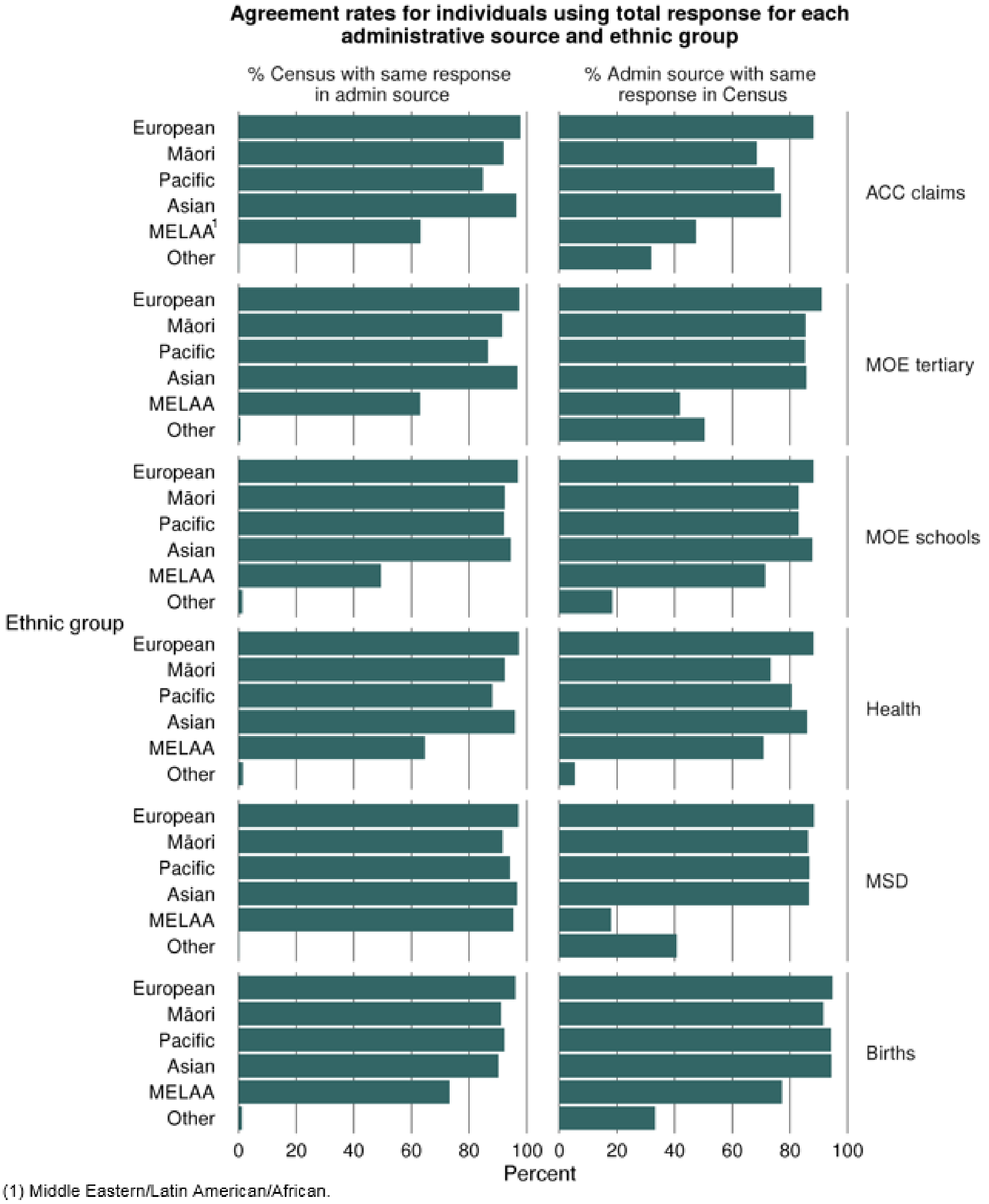
$$100\% \times \frac{\text{Number of linked people answering 'yes' in both sources}}{\text{Total number of linked people with a 'yes' response in the administrative source}}$$

These two measures together provide agreement rates for administrative records compared with the census. A low percentage in the left-hand panel indicates a source that has failed to identify members of an ethnic group according to the census. A low percentage in the right-hand panel indicates a source in which a large proportion of people are incorrectly flagged as belonging to a certain ethnic group. The appendix includes tables for all combinations of source agency by ethnic group.

European and Asian ethnic groups show higher consistency than other level 1 ethnic groups. There is also a general trend for the percentages on the right-hand side to be higher than those on the left, typically over 90 percent. In other words, people with a given ethnic group in the IDI are likely to have it in the census as well. On the left-hand side, some datasets such as ACC and Health assign Māori ethnicity to less than 80 percent of census Māori.

One notable feature of the results is that the ‘Other’ ethnic group has very low rates of agreement. Only a very small number of the people flagged in the administrative data as having ‘Other’ ethnicity have ‘Other’ ethnicity in the census. There are further problems in certain datasets, such as ACC where almost everyone with MELAA ethnicity also has ‘Other’ ethnicity, making the MELAA ethnic profile results inconsistent with census. Because of these problems we have excluded the ‘Other’ ethnic group from the remaining analysis in this paper.

Figure 4



(1) Middle Eastern/Latin American/African.

Source: Statistics New Zealand

Comparing combinations of responses

The results in the previous section considered total responses to each ethnicity independently. As people may belong to two or more ethnic groups, we now compare different combinations of ethnicities for a given individual. We use the term ‘ethnic profile’ to refer to these combinations of responses, for example ‘European and Māori’ or ‘Asian only’. These profiles are mutually exclusive (so an individual can only be in one of the categories).

Figure 5 shows the percentage of people who have the same ethnic profile as recorded in the census, for each administrative source. The denominator used to calculate the percentages is the number of people who have each ethnic profile in the census. Only census people linked to the dataset under consideration are counted in this figure. People with missing ethnicity data are also excluded from the counts.

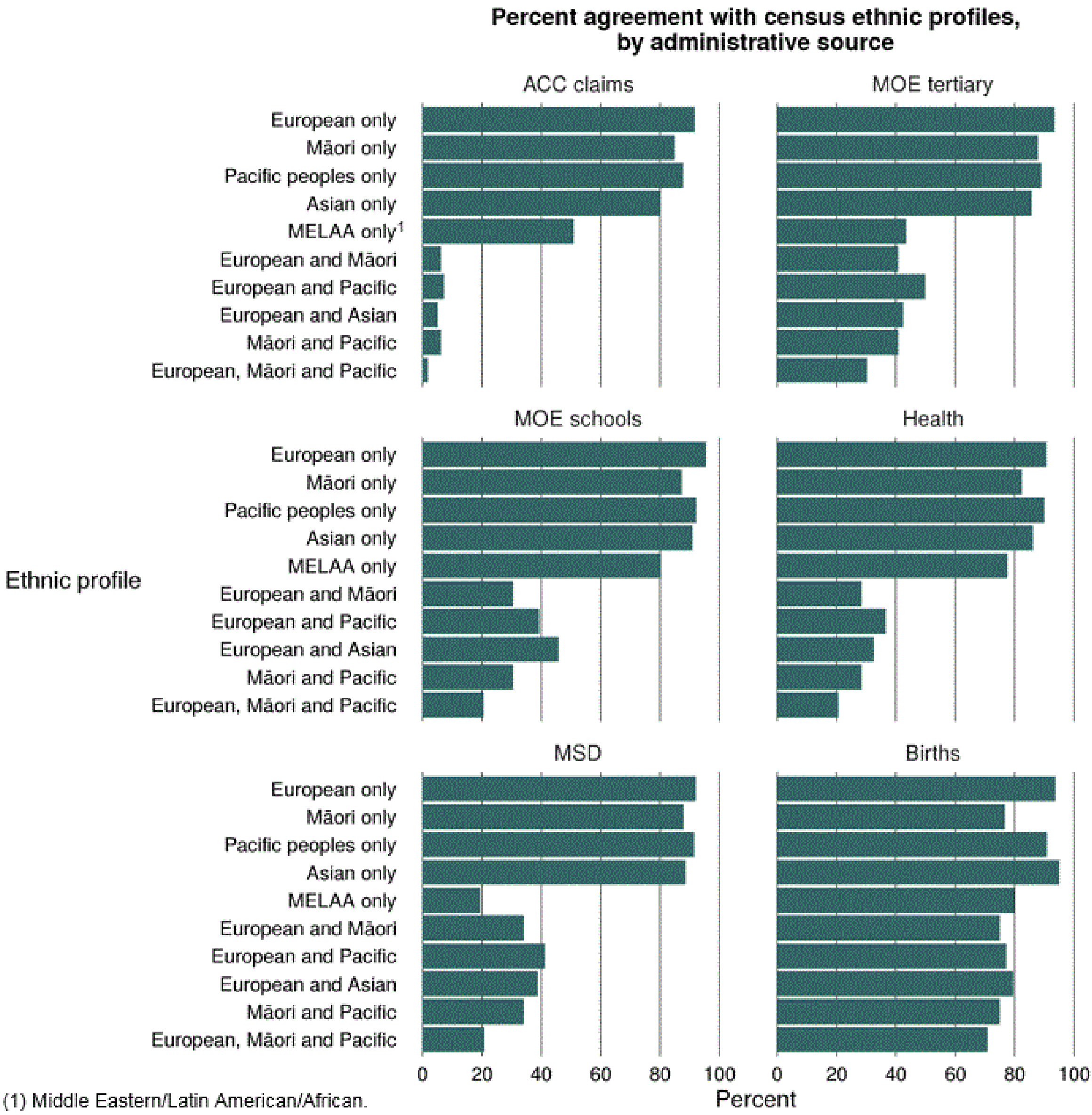
The main feature of figure 5 is the much higher agreement rates for single ethnic profiles, than for those with combinations of two or more ethnicities. This is true for all data sources except birth registrations. Lower rates of people with multiple ethnicities in administrative sources appears to be the main reason for under-reporting of ethnic groups compared with the census.

Some of this disagreement may be because results of older coding practices are still included in the IDI data used here. Birth registrations show the highest consistency across all profiles, which is probably a result of the standardisation and quality controls introduced since the late 1990s.

The different ways of comparing ethnicity data for individuals shown in the tables above represent a spectrum of strictness of comparison. Asking an administrative source to provide an exact ethnic profile (figure 5) for an individual is a more difficult test than asking how many people have the same level 1 ethnic group as the census (figure 4).

We also need to remember that ethnicity in administrative data is reported by people at different points in time, compared with the single reference date of the census. We would expect those who strongly consider themselves to belong to only one ethnic group to be more consistent across sources than those with multiple ethnicities whose responses may be more affected by differences in time or context.

Figure 5



(1) Middle Eastern/Latin American/African.

Source: Statistics New Zealand

7 How should we derive ethnicity from linked individual responses?

We have several sources of a person's ethnicity, each with different coverage patterns and different quality issues. An individual may have ethnicity recorded in one or up to six of the main sources in any combination. We need to determine the 'best' ethnic response for an individual. We have compared two basic ways of solving this problem.

Ever-recorded ethnicity

The method used in the IDI personal details table up to 2015 is to take a 'yes' on any individual source for each ethnic group to be a 'yes' in the final ethnic profile, regardless of what is recorded in other sources.

However, this method is likely to result in too many people being counted as members of some ethnic groups. This is because every 'yes' response, whether from any mistakes at source, linking errors in the IDI, or changes over time in the person's self-identification, will be elevated to the person's final ethnic profile.

Up to 10 percent of individuals have more than one ethnicity recorded in the census and in individual administrative sources, but approximately 20 percent of individuals have more than one ethnicity in the personal details table.

Source ranking

Alternatively, we could define a ranking of the administrative sources and use the highest-ranked information available for each individual. This method would be expected to reduce some of the overcounting problems seen in the 'ever-recorded' approach, but would not be able to correct for linking errors and measurement errors in particular sources.

Using the earlier results for individual sources, we created three different source rankings to use for comparison and to investigate how sensitive the results are to the ranking chosen. These example rankings are shown in table 6.

We put births first in each ranking, as it was the best source on all measures. It covers approximately 18 percent of linked census people. Different rankings of the education and health data were tested because they each have good quality but different patterns of performance for different ethnic profiles. Depending on the ordering, health was the source for 30 percent to 80 percent of records. The education sources only accounted for a few percent of people when they were ranked lower, but for up to 14 percent for schools and up to 44 percent for tertiary enrolments when ranked above health.

Table 6

The three source rankings used to decide from which source to take individual's ethnicity response

	Sources		
Rank	Ranking 1	Ranking 2	Ranking 3
1	Births	Births	Births
2	MOH	MOE schools	MOE tertiary
3	MOE schools	MOH	MOE schools
4	MOE tertiary	MOE tertiary	MOH
5	ACC claims	ACC claims	ACC claims
6	MSD benefits	MSD benefits	MSD benefits

Comparison of the methods

We can use the same comparisons as we did for individual sources to evaluate different methods for combining ethnicity from multiple sources. Table 7 is the same as table 4 except it uses the results of the combination rules described in the previous section.

Table 7

Comparison of total response counts between census and different methods of combining sources

Ethnic group	Combined administrative sources to census ratio for total responses to ethnic group			
	'Ever-recorded'	Ranking 1	Ranking 2	Ranking 3
European	1.06	0.91	0.91	0.93
Māori	1.20	0.86	0.88	0.93
Pacific peoples	1.42	0.97	0.97	0.99
Asian	1.04	0.89	0.90	0.90
MELAA	9.79	1.07	1.14	0.96

Note: MELAA = Middle Eastern/Latin American/African

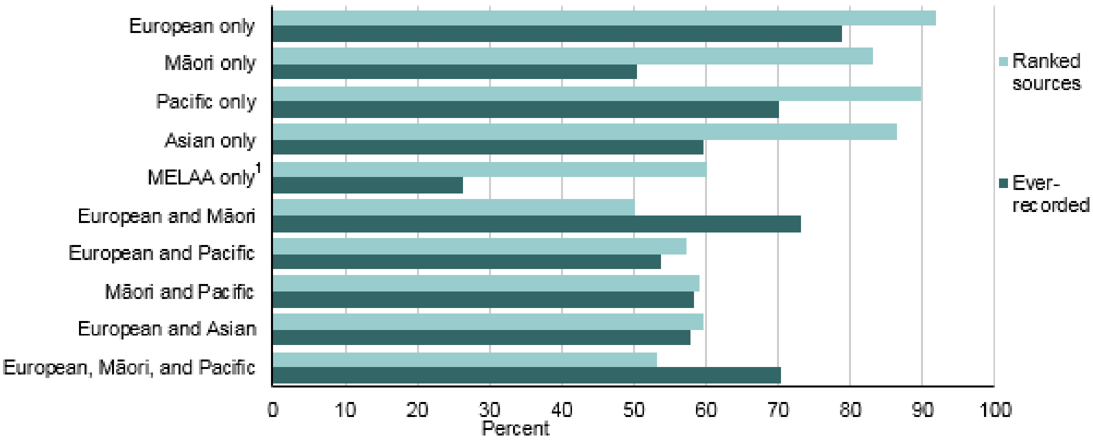
The pattern of this table demonstrates limitations of the 'ever-reported ethnicity' method used in the IDI personal details table. The 'ever-recorded' method overestimates the number of people in all ethnic groups, and for the Māori, Pacific peoples, and especially MELAA groups, the overestimate is very large.

In contrast, using the ranked source method results in an underestimate of the number of people in each ethnic group. That is, people tend to be missing ethnicities which they have in the census. The different rankings chosen do produce different results, but the general pattern is much the same. Ranking 3 has the best overall performance for all ethnic groups, and Māori in particular.

In figure 6 we have compared the ethnicity aggregation methods using the same ethnic profile comparison as we used in figure 5 for the individual sources. The same definitions of ethnic profile agreement were used for these graphs as for the ones in the earlier section. We used ranking 3 in the comparison because it performed the best in the total response ratio comparison in table 7.

Figure 6

Percent agreement with census ethnic profiles, for two methods of combining sources



(1) Middle Eastern/Latin American/African.

Source: Statistics New Zealand

The ranked source method results in higher rates of agreement than the ever-recorded ethnicity method for all of the single ethnic group profiles. The agreement rates for multiple ethnicity profiles are mixed – 'European and Māori' show closer agreement with census using the ever-recorded method, while the other two-way ethnic combinations are slightly better in the ranked data.

Other methods

We have compared two simple but contrasting methods of combining multiple data sources. The results show that the choice of method can have a major impact on the estimates for different ethnic groups. Figure 6 and table 7 show clearly that the ranked sources method is better than the ever-recorded method. They also show, however, that the ranked sources method is still far from perfect, especially for people with more than one ethnicity in the census. Although we will always be ultimately limited by the accuracy of the administrative source data, complex rulesets or statistical models could improve on the simple methods presented above.

One possibility would be to introduce a more complex set of rules, which could be based on a majority vote idea (eg if 3 out of 4 sources say a person has an ethnicity, then assign them that ethnicity) or on a combined rank/vote system (eg the ethnicity reported in a high-ranked source might be overruled if three 'low-quality' sources disagree). Removing older data may mean results align more closely with census results.

Statistical models and machine learning methods can effectively create rulesets that are more complex and optimised according to some statistical measure. An example is latent class analysis, which has been used to analyse questionnaire responses where multiple questions try to measure the same concept but appear to be unreliable or inconsistent (Biemer, 2011). Machine learning methods such as classification trees, which are designed for similar problems, could also be investigated.

8 Discussion

This paper examined the quality of ethnicity information in administrative sources by comparing collection practices with the statistical standard, and by comparing the consistency of ethnic reporting in administrative sources available in the IDI with the ethnic groups that individuals reported in the 2013 Census.

Summary of main findings

Most administrative sources in the IDI use the official New Zealand statistical standard for ethnicity and aim to capture the same concept of ethnicity of cultural affiliation. The key aspects of this concept are for an individual's ethnicity to be self-identified, and the ability to report multiple ethnicities.

In practice, the collection of ethnicity by government agencies is not entirely in line with the statistical standard for ethnicity. The most common discrepancy is that the question asked differs from the standard. Some agencies also have non-standard ways of treating multiple responses.

The census measures ethnicity at a single point in time (census day), and official population estimates also measure ethnicity at a given reference date. In contrast, many administrative sources reflect information collected from individuals at different points in time.

While the coverage of administrative sources varied, almost all people (99 percent) linked to the census had ethnicity information recorded in at least one data source in the IDI. Ministry of Health data had the highest coverage (98 percent), while coverage of other agencies is more limited because they come into contact with a limited part of the population (for example students, those on working age benefits, or ACC). For the New Zealand–born, ethnicity is available from birth registrations since 1995.

Consistency with census responses varies considerably by agency and by ethnic group. Birth registrations show the highest agreement with census, with ratios of total responses close to 1. Other agencies typically produce lower counts compared with the census for all the main ethnic groups.

A marked difference is seen between people who report single or multiple ethnic groups. For those reporting a single European, Māori, Pacific, or Asian ethnic group in the census, between 80 percent and 96 percent also have the same ethnic group in the administrative sources. However, apart from birth registrations, consistency is much lower for those reporting two or three ethnic groups in the census (less than half agree). The most likely cause of the lower numbers of total responses in the main ethnic groups appears to be fewer people with multiple ethnicities in the administrative sources.

Because of the differing coverage and quality of the data available, in practice responses from multiple sources must be combined in some way. The 'ever-recorded' method combines all ethnic groups reported in any source. This inflates the counts of the main ethnic groups, and in particular results in many people with Māori and Pacific ethnicity in administrative data who do not have those ethnicities in the census. Ranking sources on the basis of their agreement rates with census, and using the data from the highest-ranked source for each individual, brings the administrative data closer to the census. While the ranking method does not overcome all the limitations of the source data, results for multiple ethnic groups are better than for any of the sources used alone, apart from birth registrations. A key conclusion is that the method used for combining ethnicity data from multiple sources has a major impact on the results.

Further considerations

Measuring ethnicity is inherently challenging – partly because it is not a fixed concept. People may change how they identify themselves over time, or may identify themselves differently in different environments. Official population statistics on ethnicity are based on responses to the ethnicity question that people give in the census. The census is a single collection, in a relatively neutral context, and does not directly affect the individual. In contrast, the context in which ethnicity is collected in administrative sources varies considerably, and may influence how people respond.

The time reference period is a conceptual difference between official population statistics on ethnicity and that measured by administrative sources. The census measure refers to a single point in time, while administrative sources collect ethnicity at different times depending on the contact a person has with the agency.

The statistical standard for ethnicity encourages using a standardised concept, definition, collection, coding method, and output to promote data consistency and comparability in all official statistics. One of the main areas for improvement for administrative sources is in reducing the variety of questions used. Time-stamping of updates to the ethnicity field also needs to be made available with the administrative data. Efforts to improve ethnicity data already in place (for example by the Ministry of Health) are likely to improve data quality beyond what is evident in this comparison of 2013 data.

If administrative data is to be used to produce official ethnic population statistics in place of the census, then we face additional challenges that arise from using a combination of administrative sources. No single source covers the entire New Zealand–resident population, and an individual may have ethnicity recorded in one or up to six of the main sources in any combination. Linkage errors also introduce incorrect ethnicities.

Better methods of producing ethnicity from a combination of sources are needed. This may include rule-based approaches to remove older or spurious data, and using statistical models combined with independent sample surveys, which could be used to calibrate administrative responses.

Consideration could also be given to rationalising the collection of ethnicities to fewer agencies – thus reducing burden on the public, and allowing resources to be more focussed on quality.

References

Biemer, PP. (2011). *Latent class analysis of survey error*. New Jersey: John Wiley & Sons, Inc.

Blakely, T, Atkinson, J, Fawcett, J (2008). Ethnic counts on mortality and census data (mostly) agree for 2001–2004: New Zealand census-mortality study update. *The New Zealand Medical Journal*, *121*(1281), 58–62; ISSN 1175 8716. Available from www.nzma.org.nz.

Cormack, D (2010). The politics and practice of counting: Ethnicity in official statistics in Aotearoa/New Zealand. Te Rōpū Rangahau Hauora a Eru Pōmare: Wellington.

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.

Crothers, C, Woodley, A, & Davies, T (2007). Utilisation of official statistics in the Auckland region. Official Statistics Research Series. Retrieved from www.stats.govt.nz.

Kukutai, T (Tahatū Consulting), Statistics New Zealand (2008). Ethnic self-prioritisation of dual and multi-ethnic youth in New Zealand. Retrieved from www.stats.govt.nz.

O'Byrne, E, Bycroft, C, & Gibb, S (2014). An initial investigation into the potential for administrative data to provide census long-form information: Census Transformation programme. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2004). Report of the Review of the Measurement of Ethnicity, June 2004. Retrieved from www.stats.govt.nz.

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

Statistics New Zealand (2005b). Infant deaths: Demographic characteristics, ethnic reporting. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2006). The impact of prioritisation on the interpretation of ethnicity data. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2012). Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2014a). Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey. Retrieved from www.stats.govt.nz

Statistics New Zealand (2014b). Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure project. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2014c). Comparing ethnicity on death registrations with ethnicity on birth registrations. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2015). IDI Data Dictionary: Life event data (July 2015 edition). Available from www.stats.govt.nz.

Tan, L, Blakely, T, Atkinson, J (2010). Ethnic counts on mortality and census data 2001–06: New Zealand census-mortality study update. *The New Zealand Medical Journal*, *123*(1320), 37–44; ISSN 1175 8716. Retrieved from www.nzma.org.nz.

Zhang, L-C (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica, 66,* 41–63. doi: 10.1111/j.1467-9574.2011.00508.x

Appendix: Total response comparisons by ethnic group and data source

The tables below compare individual responses in each administrative data source with census by level 1 ethnic grouping. The population for each table is the number of individuals with ethnicity information in both the census and the administrative source.

See appendix tables for data sources:

1 Ministry of Social Development and census

2 Accident Compensation Corporation and census

3 Tertiary education and census

4 Schools and census

5 Births and census

6 Health and census

Appendix table 1

Ministry of Social Development and census
Total response comparisons by ethnic group and data source

		Census European ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
MSD European ethnicity	Yes	784,600	64	17,400	1	802,000	65
	No	102,500	8	328,800	27	431,400	35
	Total	887,100	72	346,300	28	1,233,400	100
		Census Māori ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
MSD Māori ethnicity	Yes	213,900	17	16,900	1	230,900	19
	No	34,000	3	968,500	79	1,002,500	81
	Total	247,900	20	985,400	80	1,233,400	100
		Census Pacific ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
MSD Pacific ethnicity	Yes	94,100	8	4,700	0	98,800	8
	No	14,300	1	1,120,400	91	1,134,600	92
	Total	108,400	9	1,125,000	91	1,233,400	100

Appendix table 1, continued

		Census Asian ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
MSD Asian ethnicity	Yes	94,100	8	4,700	0	98,800	8
	No	14,300	1	1,120,400	91	1,134,600	92
	Total	108,400	9	1,125,000	91	1,233,400	100

		Census MELAA ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
MSD MELAA ethnicity	Yes	2,600	0	100	0	2,700	0
	No	11,900	1	1,218,800	99	1,230,700	100
	Total	14,500	1	1,218,900	99	1,233,400	100

Note: MELAA = Middle Eastern/Latin American/African

		Census Other ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
MSD Other ethnicity	Yes	200	0	92,800	8	93,000	8
	No	300	0	1,140,100	92	1,140,400	92
	Total	400	0	1,233,000	100	1,233,400	100

Appendix table 2

Accident Compensation Corporation and census
Total response comparisons by ethnic group and data source

		Census European ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
ACC European ethnicity	Yes	1,308,700	72	19,600	1	1,328,300	73
	No	173,600	10	307,800	17	481,400	27
	Total	1,482,300	82	327,400	18	1,809,800	100

		Census Māori ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
ACC Māori ethnicity	Yes	155,400	9	11,900	1	167,300	9
	No	71,600	4	1,570,900	87	1,642,500	91
	Total	227,000	13	1,582,700	87	1,809,800	100

		Census Pacific ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
ACC Pacific ethnicity	Yes	66,700	4	11,100	1	77,800	4
	No	22,700	1	1,709,200	94	1,731,900	96
	Total	89,500	5	1,720,300	95	1,809,800	100

		Census Asian ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
ACC Asian ethnicity	Yes	105,200	6	3,000	0	108,200	6
	No	31,400	2	1,670,200	92	1,701,600	94
	Total	136,600	8	1,673,200	92	1,809,800	100

		Census MELAA ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
ACC MELAA ethnicity	Yes	6,400	0	3,500	0	9,900	1
	No	7,100	0	1,792,800	99	1,799,800	99
	Total	13,500	1	1,796,300	99	1,809,800	100

Note: MELAA = Middle Eastern/Latin American/African

		Census Other ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
ACC Other ethnicity	Yes	200	0	89,000	5	89,200	5
	No	500	0	1,720,100	95	1,720,500	95
	Total	700	0	1,809,000	100	1,809,800	100

Appendix table 3

Tertiary education and census
Total response comparisons by ethnic group and data source

		Census European ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Tertiary European ethnicity	Yes	1,156,200	68	22,700	1	1,179,000	69
	No	113,600	7	408,300	24	521,900	31
	Total	1,269,800	75	431,000	25	1,700,900	100

		Census Māori ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Tertiary Māori ethnicity	Yes	223,400	13	18,200	1	241,700	14
	No	38,000	2	1,421,200	84	1,459,200	86
	Total	261,400	15	1,439,400	85	1,700,900	100

		Census Pacific ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Tertiary Pacific ethnicity	Yes	86,900	5	12,300	1	99,200	6
	No	14,900	1	1,586,800	93	1,601,700	94
	Total	101,800	6	1,599,100	94	1,700,900	100

		Census Asian ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Tertiary Asian ethnicity	Yes	177,300	10	4,300	0	181,600	11
	No	29,400	2	1,489,900	88	1,519,200	89
	Total	206,700	12	1,494,200	88	1,700,900	100

		Census MELAA ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Tertiary MELAA ethnicity	Yes	8,200	0	4,500	0	12,700	1
	No	11,300	1	1,676,800	99	1,688,100	99
	Total	19,500	1	1,681,300	99	1,700,900	100

Note: MELAA = Middle Eastern/Latin American/African

		Census Other ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Tertiary Other ethnicity	Yes	400	0	54,200	3	54,600	3
	No	400	0	1,645,900	97	1,646,300	97
	Total	700	0	1,700,100	100	1,700,900	100

Appendix table 4

Schools and census
Comparison of ethnicity information in administrative data and the census

		Census European ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Schools European ethnicity	Yes	669,900	64	17,500	2	687,500	66
	No	89,600	9	267,500	26	357,000	34
	Total	759,500	73	285,000	27	1,044,500	100

		Census Māori ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Schools Māori ethnicity	Yes	188,800	18	13,500	1	202,300	19
	No	38,600	4	803,700	77	842,200	81
	Total	227,300	22	817,200	78	1,044,500	100

		Census Pacific ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Schools Pacific ethnicity	Yes	98,300	9	7,100	1	105,400	10
	No	20,200	2	918,800	88	939,100	90
	Total	118,500	11	926,000	89	1,044,500	100

		Census Asian ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Schools Asian ethnicity	Yes	110,600	11	5,400	1	116,000	11
	No	15,300	1	913,200	87	928,500	89
	Total	125,900	12	918,600	88	1,044,500	100

		Census MELAA ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Schools MELAA ethnicity	Yes	9,900	1	9,700	1	19,600	2
	No	4,000	0	1,021,000	98	1,024,900	98
	Total	13,800	1	1,030,700	99	1,044,500	100

Note: MELAA = Middle Eastern/Latin American/African

		Census Other ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Schools Other ethnicity	Yes	100	0	7,000	1	7,100	1
	No	400	0	1,037,000	99	1,037,400	99
	Total	500	0	1,044,000	100	1,044,500	100

Appendix table 5

Births and census
Comparison of ethnicity information in administrative data and the census

		Census European ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Births European ethnicity	Yes	510,600	71	17,900	2	528,500	74
	No	27,800	4	161,400	22	189,200	26
	Total	538,400	75	179,300	25	717,700	100

		Census Māori ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Births Māori ethnicity	Yes	172,900	24	14,800	2	187,600	26
	No	15,800	2	514,200	72	530,000	74
	Total	188,700	26	529,000	74	717,700	100

		Census Pacific ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Births Pacific ethnicity	Yes	86,000	12	6,100	1	92,100	13
	No	5,300	1	620,300	86	625,600	87
	Total	91,300	13	626,300	87	717,700	100

		Census Asian ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Births Asian ethnicity	Yes	65,100	9	6,300	1	71,400	10
	No	3,900	1	642,400	90	646,300	90
	Total	68,900	10	648,700	90	717,700	100

		Census MELAA ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Births MELAA ethnicity	Yes	5,900	1	2,000	0	7,900	1
	No	1,700	0	708,000	99	709,700	99
	Total	7,700	1	710,000	99	717,700	100

Note: MELAA = Middle Eastern/Latin American/African

		Census Other ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Births Other ethnicity	Yes	100	0	6,500	1	6,600	1
	No	200	0	710,900	99	711,000	99
	Total	200	0	717,400	100	717,700	100

Appendix table 6

Health and census
Comparison of ethnicity information in administrative data and the census

		Census European ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Health European ethnicity	Yes	2,560,700	68	53,700	1	2,614,300	69
	No	342,700	9	836,200	22	1,178,900	31
	Total	2,903,400	77	889,900	23	3,793,300	100

		Census Māori ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Health Māori ethnicity	Yes	411,800	11	28,900	1	440,700	12
	No	149,400	4	3,203,200	84	3,352,600	88
	Total	561,200	15	3,232,100	85	3,793,300	100

		Census Pacific ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Health Pacific ethnicity	Yes	221,500	6	27,100	1	248,600	7
	No	53,000	1	3,491,700	92	3,544,700	93
	Total	274,500	7	3,518,800	93	3,793,300	100

		Census Asian ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Health Asian ethnicity	Yes	363,400	10	11,900	0	375,300	10
	No	59,000	2	3,358,900	89	3,418,000	90
	Total	422,500	11	3,370,800	89	3,793,300	100

		Census MELAA ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Health MELAA ethnicity	Yes	29,800	1	15,300	0	45,100	1
	No	12,300	0	3,735,900	98	3,748,200	99
	Total	42,100	1	3,751,200	99	3,793,300	100

Note: MELAA = Middle Eastern/Latin American/African

		Census Other ethnicity					
		Yes		No		Total	
		Number	%	Number	%	Number	%
Health Other ethnicity	Yes	100	0	6,000	0	6,100	0
	No	1,600	0	3,785,600	100	3,787,200	100
	Total	1,700	0	3,791,600	100	3,793,300	100