# Identifying Māori populations using administrative data: A comparison with the census

Census Transformation

Christine Bycroft, Giles Reid, Justine McNally, 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.

# Contents

# List of tables and figures

# 1   Background

## Census Transformation in New Zealand

In March 2012 the New Zealand Government agreed to a Census Transformation strategy, which 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

Previous work by the Census Transformation programme identified information requirements for and about Māori that a census must provide, regardless of how the census is conducted in the future. Essential census requirements that are specific to Māori are: Māori ethnicity, Māori descent, iwi, and te reo Māori (the Māori language).

In this paper we investigate the potential for administrative sources to provide this information. For three of these variables (Māori ethnicity, Māori descent, and iwi) we compare data from the 2013 Census with information from the administrative data sources available in Statistics NZ's Integrated Data Infrastructure (IDI).

For each variable, we describe the main features of each administrative data source and compare the concepts and definitions used with the relevant statistical standard. We describe the coverage of each data source. We also use the 2013 Census data linked to the administrative source to compare consistency between census responses and the values recorded in the administrative sources.

Some administrative sources provide very good information for and about Māori. However, the lack of completeness and lower quality of other sources means that administrative data cannot, at present, replace the essential information needs that the current survey-based census provides. There is some promise that with improvements to data collection, Māori ethnicity could be provided through linked administrative data sources. For Māori descent, better coverage for adults is needed. The feasibility of collecting high-quality iwi information from government agencies or Māori organisations remains uncertain and will require government to work in partnership with iwi. Language proficiency is unlikely to be suitable for collection through administrative sources.

# 2 Introduction

Enduring census information requirements for and about Māori, based on an understanding of partnership through the Treaty of Waitangi and legal obligations, are established in Gleisner, Downey, & McNally (2015).

They identify four variables as essential census information requirements specific to Māori: Māori descent, Māori ethnicity, iwi, and te reo Māori (the Māori language).

**Māori ethnic group** and **Māori descent** are the primary identifiers of belonging as Māori, and are therefore essential for census to produce any Māori information at all. Both are legal requirements: under the Statistics Act 1975 (ethnicity) and Electoral Act 1993 (Māori descent).

**Iwi** (Māori tribal groupings) is also a core identifier for Māori and fundamental to Treaty settlements, both to support the settlement process and to monitor post-settlement outcomes.

**Te reo Māori** is clearly of paramount importance to Māori, and the Crown's commitment as a Treaty partner is reflected in the Māori Language Act 1987. Information is necessary to monitor the health of te reo and is used by both government and Māori bodies, such as iwi and Te Taura Whiri i te Reo Māori. The census is the only source permitting analysis of spoken language for the entire population that allows detailed breakdowns (eg te reo speakers by iwi and age groups).

The population identifiers and basic demographics provided by the census serve as the population reference point for other data sources. The census also provides a sampling frame for surveys that are targeted to Māori ([Te Kupenga](#) 2013 for example).

Any future census model based on administrative sources must also be able to provide this essential information. As Māori ethnicity and Māori descent are population identifiers, an administrative-based census must be able to source these variables from administrative data for all New Zealand residents. If a large-scale sample survey forms a component of a future census then iwi and te reo may be collected either through the survey, from administrative data if possible, or a combination of both.

## Aims and scope

This paper summarises the availability and quality of administrative data sources for these critical information needs for and about Māori. We first provide reference information about the statistical concepts and about the relevant administrative data sources. We then describe preliminary analysis of Māori ethnicity, descent, and iwi information, where this was available from the linked administrative sources within Statistics NZ's Integrated Data Infrastructure (IDI). Rates of agreement between administrative sources and the census are found by comparing individual responses from 2013 Census data linked to the IDI.

Analysis has been limited to administrative sources available in the IDI in May 2015. In addition, we include aggregate comparisons against the electoral roll for Māori descent. Analysis is only for national-level results for each variable. There are no breakdowns for age, sex, or by region. While we discuss iwi registers, no analysis was possible as this data is not available in the IDI. Similarly, there is no individual-level analysis of electoral roll data, because it is not available in the IDI. We found no administrative sources suitable for analysing language.

The results for Māori ethnicity are summarised from Reid et al (2016), who compare census level 1 ethnic groups with ethnicity data from IDI administrative sources.

# 3   Methods

## A framework for assessing accuracy

The concepts of coverage error and measurement error provide a framework for assessing the accuracy of data sources (Zhang, 2012).

Coverage describes the relationship between the ideal target population and the actual set of people present in a dataset.  For each variable discussed here, the target population of interest is New Zealand residents of all ages.

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 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

We now describe the methods used to evaluate the quality of the information in administrative sources.

A brief description of each data source provides basic information on population coverage. The concepts and definitions used in the administrative data collections are compared with the relevant national standard and related classification. Ideally concepts and definitions should be consistent across collections and consistent with the standard.

For each variable we then compare the administrative data with 2013 Census data. A linked Census-IDI dataset (described below) provides the basis for all the data analysis. A high linkage rate provides a good basis for this comparison. No adjustment is made for any remaining bias due to differential linkage rates.

Results summarise coverage for the administrative source compared with the New Zealand resident population. We gain insights into measurement error from aggregate and individual-level analysis.

For the aggregate analysis we compare the total responses in the administrative source to total responses in the census.  As coverage of each data source varies considerably, we restrict the comparisons to responses in each administrative data source that were linked to usual residents in the census.

Even close results at the aggregate level may be a result of classification differences balancing out. The linked dataset allows us to compare the values recorded for an individual in the administrative sources against those recorded for the same individual in census. The individual-level analysis helps to show what is driving the results seen in the aggregate comparisons. The individual-level comparisons are made for the group of people who had records in the IDI and the census, their records were linked together, and a value was recorded for that variable.

# 4 Classifications and related standards

A statistical classification is a way to group a set of related categories in a meaningful, systematic, and standard format. A statistical standard provides a comprehensive set of guidelines for surveys and administrative sources collecting information on a particular topic (Statistics NZ, nd).

See [Classifications and related standards.](#)

New Zealand statistical classifications and standards are designed for use across official statistics collections, both for Statistics NZ and other agencies. Statistical standards allow us to collect reliable statistics using consistent procedures. If we follow these standards, we can integrate data over time and across different data sources.

## Māori ethnicity

The New Zealand 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.

The 2005 standard classification of ethnicity is a hierarchical classification of four levels. 'Māori' is a single ethnicity appearing at all levels of the classification – level 1 has six categories, including Māori.

The presence of multiple ethnicities for the same person means there are two standard outputs for ethnicity data.

- **Total response output** shows the total counts of all responses given for each ethnic group. The number of total responses will be greater than the total population as individuals can appear in more than one ethnic group.
- **Single/combination output** places each person into only one category depending on the combination of ethnic groups reported; for example Māori only, or Māori and European.

The distinction between reporting Māori as the sole ethnicity, and people also identifying as belonging to other ethnic groups is important for understanding Māori ethnicity.

## Māori descent

'Māori descent' is based on a genealogical or biological concept, rather than on cultural affiliation (as ethnicity is). The statistical standard defines Māori descent as: "A person has Māori descent if they are of the Māori race of New Zealand; this includes any descendant of such a person."

The classification for Māori descent has two main categories: 'Māori descent' and 'No Māori descent', plus residual categories ('don't know', 'refused to answer' and so on).

# Iwi

The statistical standard for iwi is also based on genealogical or biological concepts. The standard definition is:

> The Iwi today is the focal economic and political unit of the traditional Māori descent and kinship based hierarchy of:
>
> Waka (founding canoe)
> |
> Iwi (tribe)
> |
> Hapū (sub-tribe)
> |
> Whānau (family).
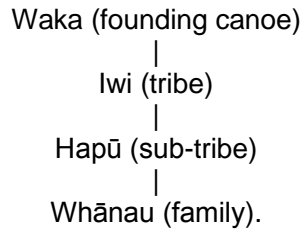
Thus membership of an iwi first requires a person to be of Māori descent. As with ethnicity, the iwi standard states that multiple response should be expected when collecting iwi information. Since many people do not know their iwi, the standard advises that a 'don't know' tick box option should be provided.

The iwi classification used by Statistics NZ included 128 iwi categories in 2015. The classification is prepared for statistical purposes only and is not intended to be a definitive list of all iwi. The criteria for including an iwi in the classification are complex and include historical, economic, and political considerations.

# Language

The standard classification of language is a hierarchical classification of three levels. Māori appears at Level 3 of the classification, within Central Pacific languages.

There is no standard concept or definition for collecting information on language proficiency. Census provides information on spoken language. The Te Kupenga survey obtains more detail on respondents' ability to speak, listen, read, and write in te reo Māori.

Figure 1 illustrates the relationships between the populations identified by these four variables. Information about te reo is collected for all New Zealand residents, and is not limited to those of Māori ethnicity or descent.

**Figure 1**
**Relationship between Māori ethnicity, Māori descent, and iwi affiliation**
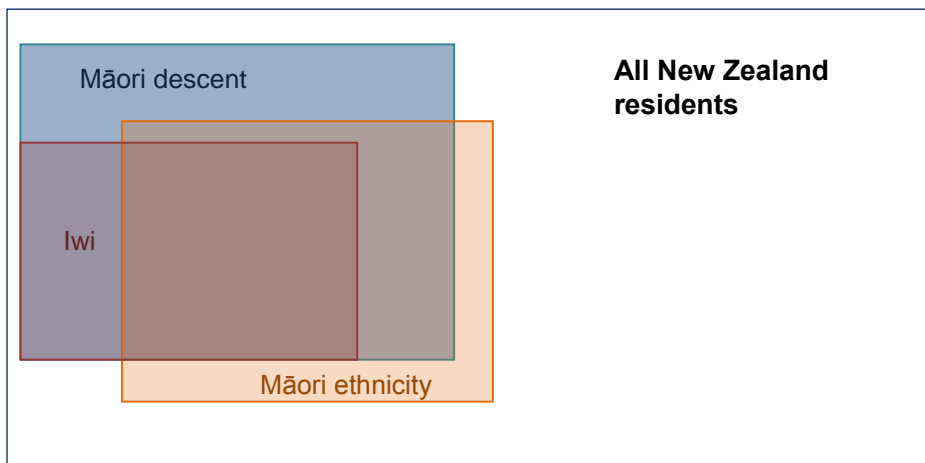
# 5   Data sources

This section describes the data sources used in this investigation: the New Zealand Census of Population and Dwellings (the census), the main administrative sources for the four variables, how these administrative sources are brought together in Statistics NZ's Integrated Data Infrastructure (IDI), and the linked Census-IDI dataset.

## New Zealand Census of Population and Dwellings

The census 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 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 on census night.

### Census coverage and missing data

The 2013 Census usual resident population count is 4,242,048 people. The census count includes 4.8 percent (203,052) substitute records (Statistics NZ, 2014a). A substitute is a census record created (imputed) where there is sufficient evidence received during the collection process that a person exists, or a dwelling was occupied, but we obtained no corresponding census form. As such, substitutes are part of census non-response. While the census imputes values for age and sex, there is no imputation in published census outputs for the variables considered in this paper

Coverage in the census is measured by the Post-Enumeration Survey (PES) (Statistics NZ, 2014b). Net census undercount for the 2013 Census was estimated at 2.4 percent. Younger adults aged 15–29 years (4.8 percent) had a higher percentage undercount than other age groups. Net undercount also varies by ethnicity, with the percentage undercount for Māori (6.1 percent) and Pacific peoples (4.8 percent), with young age structures, being higher than for Asian (3.0 percent) and European (1.9 percent) ethnic groups.

### The 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 (Statistics NZ Standard for population terms). New Zealand's ERP is derived by adjusting the census usually resident population count for net census undercount (as estimated by the PES) and the estimated number of residents temporarily overseas on census night. To obtain the ERP at a given date after census night, updates are made for natural increase (births less deaths) and net migration (arrivals less departures) between census night and the given date. The official ERP series provides the best measure of who is living in New Zealand at a given date.

### Variables in the census

The census uses the statistical standard and classification for the four variables considered here. We summarise the main census results and the relationships between ethnicity, descent, and iwi.

See QuickStats about Māori (Statistics NZ, 2013) for further detail about census results.

**Ethnicity**

In the 2013 Census, 598,605 people usually living in New Zealand identified with the Māori ethnic group. Almost half these people (278,196 or 46.5 percent) identified Māori as their only ethnicity.

Ethnicity is a 'foremost' variable in the census, which means it is managed to produce information of highest quality. The non-response rate for ethnicity for those who return a census form is low (0.7 percent in the 2013 Census). However, the overall non-response to ethnicity, including substitute forms, is 5.4 percent. The census reports Māori as making up 14.9 percent of the population, while the ERP (which adjusts for non-response) reports 15.6 percent.

**Māori descent**

The census Māori descent question asks respondents: "Are you descended from a Māori (that is, did you have a Māori birth parent, grandparent or great-grandparent, etc.)?"

In 2013, 16 percent (668,724 people) answered 'yes' to the Māori descent question and 72 percent answered 'no'; 2 percent answered 'don't know', and 10 percent did not respond.

Māori descent and Māori ethnicity are closely related concepts, but census results demonstrate that people do respond differently. Of the 2013 Census respondents who said they were of Māori descent, 16 percent (107,391) said they were not ethnically Māori. A smaller group (4,212 people) identified as being of Māori ethnicity while stating they had no Māori descent.

**Iwi affiliation**

Only those who selected 'Yes' to Māori descent are asked to provide information about their iwi. The question is: "Do you know the name(s) of your iwi (tribe or tribes)?" Respondents are able to state up to five iwi or rohe (region).

Iwi information collected in the census is subject to some processing and quality issues. The census question on iwi required a written-in response. A team of specialist process operators were employed to ensure responses were coded as accurately as possible.

While 80 percent of those with Māori descent provided at least one valid iwi in the 2013 Census, 17 percent said that they did not know their iwi, and a further 3 percent of responses could not be coded. At the same time, 14,000 people gave a valid iwi in the 2013 Census but did not respond to the Māori descent question. These were not included in either the descent or iwi census counts.

**Te reo Māori**

Information about people's ability to speak te reo Māori is collected from the general 'languages spoken' question in the Census. The 2013 Census asked respondents to identify the languages in which they could "have a conversation about a lot of everyday things". Māori is a tick box response option.

The non-response rate for the 2013 language question was 6.3 percent, of which most were substitute records.

# Administrative sources

Several government agencies and Māori organisations collect information about the four variables. Table 1 summarises these sources by the variables available. The remainder of this section describes key features of the sources available for each variable.

**Table 1**
**Administrative sources for the four variables**

| Source | Māori ethnicity | Māori descent | Iwi | Te reo |
|---|---|---|---|---|
| Department of Internal Affairs (Births and Deaths) | ✓ | ✓ | .. | .. |
| Ministry of Health | ✓ | .. | .. | .. |
| Ministry of Education | ✓ | .. | ✓ | ✓ |
| Ministry of Social Development | ✓ | .. | .. | .. |
| Accident Compensation Corporation | ✓ | .. | .. | .. |
| Electoral Commission | .. | ✓ | .. | .. |
| Iwi registers | .. | ✓ | ✓ | ✓ |
| Tūhono Trust | .. | ✓ | ✓ | .. |
| Symbol: ..not available | | | | |

## Administrative sources for ethnicity

The main government agencies that collect ethnicity are: Department of Internal Affairs (DIA) Birth and Death registrations, Ministry of Health and health service providers, the Ministry of Education, and the Ministry of Social Development. Accident Compensation Commission data is also available.

Reid et al (2016) describe the ethnicity information collected by each of these agencies, primarily based on Cormack (2010) and Cormack & McLeod (2010). Most agencies apply the standard concepts of cultural affiliation, self-identification (where possible), and allow people to belong to more than one ethnic group.

Government agencies vary in the collection mode and questionnaire used to collect ethnicity information. Some forms, such as the birth registration form used since 1995, align very closely with the statistical standard, including having a nearly identical question to the census. Some other agencies use a question that is conceptually in line with the standard but differs in wording or presentation.

Response coding differs. Some agencies code to a higher (less detailed) level than the full level 4 classification, and others deal differently with multiple responses. Older data is often limited, and not consistent with the current standard.

The level of quality controls in place also varies. DIA works closely with Statistics NZ, which processes the data for publication of official statistics and closely monitors quality. Other agencies have few external checks on their data collection.

## Administrative sources for Māori descent

Relatively few government administrative sources collect information on Māori descent. Only birth and death registrations and electoral enrolments ask about Māori descent directly. Birth and death registrations are available in the IDI. Individual-level data for electoral enrolments are not available to Statistics NZ, due to restrictions in the Electoral Act 1993, although aggregate tables can be compared with census results.

DIA is responsible for birth registrations, and records go back to the 19th century. Until 1962, the agency kept separate registers 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. While this definition is not consistent with the measurement of ethnicity, it is consistent with a measure of descent for Māori.

DIA introduced a new birth registration form in September 1995 that included an ethnic question consistent with the concept of ethnic self-identification. In addition the form included a question on Māori descent. The registration form includes ethnicity and Māori descent questions for the mother, father, and child. Since 1998, birth and death records have been recorded digitally.

Under the Electoral Act 1993, all people eligible to vote are required to enrol with the Electoral Commission. As part of registering to vote, you must answer the question "Are you a New Zealand Māori or a descendant of a New Zealand Māori?" Tick boxes for 'yes' and 'no' are provided (Electoral Commission, 2015). Responses to this question are used to determine the number of Māori electorates. Only people answering 'yes' to this question are eligible for the Māori electoral roll and to vote in the Māori electorates. The collection processes and quality of the data collected are generally good since it is crucial for running the electoral system. The major limitation of the electoral roll for this investigation is that only people 18 years or older are eligible to vote.

As a result of a concerted cross-government effort in the 1990s, Māori descent information from birth and death registrations, electoral enrolment data, and the census, is well-standardised. These sources each have the advantage of being single, centralised collections across the entire country that are all handled by a single agency. Birth and death registrations and electoral enrolment are also important legal processes.

## Administrative sources for iwi

The main government source of iwi information is the Ministry of Education (MoE). Iwi information is now collected from all sectors of the education system (Education Counts).

See Iwi data: collection and use for more details.

Tertiary providers have been required to provide iwi affiliation of all first-year Māori students from 2002, although many providers have also provided comprehensive information on Māori students who had first enrolled in previous years.

All School Roll Returns have included iwi from 2007, and systematic collection of iwi for early childhood providers began in 2014. Coverage for iwi is therefore limited to younger age groups. Information is collected during the enrolment process at each early childhood centre, school, or tertiary institution.

The MoE provides guidelines for collecting iwi information on enrolment forms. Iwi affiliation is based on self-identification, and forms should allow for up to three iwi. Statistics NZ's iwi classification is used, and the iwi codes are organised into regional groupings for reporting. However, the wording of the question, and response options, varies widely across different schools and tertiary institutions.

Other government agencies have collected some iwi information, including through the Student Loans and Allowances scheme, by the Ministry of Health, by DIA in birth registrations, and by the Department of Corrections. However, due to the limited amount of data, these sources are not considered further.

Māori organisations also collect iwi information, and by default they also collect Māori descent. Most iwi have established their own registers of enrolled members – either as a precursor to, or a condition of, Treaty of Waitangi settlements. Unlike the census and most government agency data sources, iwi membership is not based on self-identification but on acknowledgement of whakapapa (genealogy), endorsed by the iwi or hapu

kaumatua (elders) (Walling, Small-Rodriguez, & Kukutai, 2009). The registration process depends on the iwi's own protocol and its position in the settlement process.

Iwi registers vary in completeness and quality, depending on the success of Māori- or iwi-driven initiatives and, to some extent, the iwi's position in the Treaty settlement process. Walling et al (2009) note the main sources of error on the Waikato-Tainui register are duplicate records, invalid applications, and deceased members being retained. Registers will only include those who register as members of the iwi, which is likely to be a subset of the iwi-affiliated population provided in the census. These issues are likely to result in some under-coverage and some over-coverage in iwi registers compared with the census.

No data for iwi registers was available in the IDI for this investigation.

[The Tūhono Trust] is an important pan-tribal iwi organisation. Through a 2003 amendment to the Electoral Act, the trust has a legislated role as Kaitiaki (guardian) of the iwi affiliation of Māori who are registered to vote. Secure systems allow Tūhono to facilitate sharing of information about Māori registered to vote with their affiliated iwi. Other pan-tribal organisations, such as urban Māori authorities and Māori business entities, also have an interest in the iwi affiliation of their members.

## Administrative sources for te reo

Government data relating to te reo Māori is limited. MoE data about enrolment in kura kaupapa Māori (Māori-medium schools) or te reo courses may provide some information about its uptake, but studying a language is not the same as being able to have a conversation in that language. MoE data would not capture people who learned te reo at home or overseas, or who completed their education before the early 1990s.

MoE information is also available about te reo teachers. Since 2014, MoE has also collected information on 'language(s) spoken at home' from early childhood establishments. Some iwi have collected measures of te reo proficiency. While these sources might provide an indication of language proficiency, their coverage of the population is low.

The limited population coverage of these data sources, and the conceptual differences between them and the census, make replacing census Māori language information with administrative data unlikely at present. We do not consider te reo further in this paper.

# Integrated Data Infrastructure

Statistics NZ developed the Integrated Data Infrastructure (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 3 (appendix). The structure 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 issued by Inland Revenue; a list of all births registered in New Zealand since 1920; and a list of all visas granted to migrants from 1997 (excluding visitor and transit visas). Other datasets are linked to the IDI spine and include a wide range of subject areas.

Statistics NZ (2014c) 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 contains summary tables that provide core information about individuals (age, sex, ethnicity, and geographic location) summarised from across the available data sources.

For ethnicity, IDI business rules are applied to standardise the ethnic codes received from each agency. Ethnic information for each individual is combined in the IDI Personal Details table. In the process applied in 2015, ethnicity in the Personal Details table is a combination of the original responses given to separate agencies, coded to level one of the 2005 Ethnic Standard. An ethnic group is recorded wherever it is captured by any agency, at any point in time (ever-recorded).

Of the administrative sources described above, most were available in the May 2015 IDI. The exceptions are electoral roll data, Māori-owned sources such as iwi registers, and data from pan-iwi organisations.

# Linking the census to the IDI

The 2013 Census has been linked to the IDI spine to enable comparisons between the information available for an individual in the administrative sources, and the responses provided by the same person in the census. This linked Census-IDI dataset 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; the linked data was only available to approved Statistics NZ staff working on the Census Transformation programme.

Linking was completed in Quality Stage using probabilistic matching techniques. The variables full name, date of birth, sex, meshblock of usual residence, and country of birth were used in the linkage process. Overall, 3,920,364 census records were linked to the IDI (92.4 percent of the census count). Of most interest for this paper, 95.4 percent of census records for New Zealand residents in households where forms were returned (non-substitute households) were linked to the IDI. The linkage rate was better for individuals who had used e-forms (98 percent linked) than for paper forms (93 percent linked). The links in this dataset have an estimated false positive rate of less than 1 percent (where an incorrect link is made between two different individuals).

# 6 Results

Results are presented for each of the four variables in turn. We first assess the coverage of the data sources. For those in the linked Census-IDI dataset, we compare aggregate results, and then show a cross-classification of individual-level responses.

## Māori ethnicity

### Coverage

While the coverage of administrative sources varies depending on the agency, nearly everyone (98.6 percent) in the linked Census-IDI data has an ethnic code recorded from at least one source in the IDI. While there is some variation by age, sex, and ethnic group, ethnicity information is available for more than 95 percent at every single year of age, and for each main ethnic group.

### Aggregate comparison

Table 2 shows the total responses for the Māori ethnic group in the administrative sources and the census. While almost 600,000 people reported being of Māori ethnicity in the census, table 2 includes only records available in the particular administrative source and that were linked to usual residents in the census. For example, of the birth registration records linked to the census, 188,700 are recorded as Māori in the census, and 187,600 have Māori ethnicity in the births data. The final column shows these total responses expressed as a ratio of the administrative source to the census (0.99 in this case) and provides a measure of the consistency of each source compared with the census.

Birth registrations have a ratio of almost 1, indicating very good agreement with the census at an aggregate level. All other single sources have a ratio less than 1, indicating fewer people are stating they are Māori than in the census.

**Table 2**

**Total responses for Māori ethnic group for the census and administrative sources (linked Census-IDI source data only)**

2013

| Administrative source | Total Māori ethnicity responses | | Ratio Admin:Census |
|---|---|---|---|
| | Census | Administrative source | |
| DIA Birth registrations | 188,700 | 187,600 | 0.99 |
| Health NHI | 561,200 | 440,700 | 0.79 |
| MoE Tertiary enrolments | 261,400 | 241,700 | 0.92 |
| MoE School enrolments | 227,300 | 202,300 | 0.89 |
| MSD | 247,900 | 230,900 | 0.93 |
| ACC | 227,000 | 167,300 | 0.74 |
| Ever-recorded, all sources | 569,600 | 678,400 | 1.19 |
| Ranked sources | 566,700 | 522,500 | 0.92 |

To provide ethnicity information for the whole population, we must combine the information from these data sources. The bottom two rows in table 2 show results from applying two methods of combining ethnicity from different administrative sources. The

'ever-recorded' ethnicity assigns a person to the Māori ethnic group if they are recorded as Māori in any contributing source, at any time. This is the method used in the IDI Personal Details table in 2015. The total responses ratio of 1.19 shows that the ever-recorded method inflates the Māori ethnicity count compared with the census by nearly 20 percent.

Another approach is to choose only one source of ethnicity for each person, but to choose the highest quality source available. In the ranking used here, birth registrations have the highest ranking, followed by Ministry of Education, Ministry of Health, and lastly Accident Compensation Corporation data. The ratio of 0.92 for this ranking method is closer to 1 than the ever-recorded ratio; however, Māori are still being undercounted compared with the census. Ranking the sources does provide almost full coverage, but it cannot fully make up for the lower number of Māori responses in many sources.

Similar patterns are seen for other ethnic groups when these methods are applied.

## Individual-level comparison

The linked Census-IDI data allows us to compare responses at an individual level. Much of the discrepancy between administrative sources and the census is due to the lower reporting of multiple ethnicities in administrative data. Figure 1 shows agreement of the administrative sources with the ethnic profile reported in the census, for single Māori ethnicity or Māori in combination with another ethnic group. Reporting of sole Māori is reasonably consistent across all administrative sources with 80 to 90 percent of those reporting sole Māori in the census also reporting sole Māori in the administrative sources. Only birth registrations achieves a similar consistency for Māori in combination with another ethnic group. Other administrative sources have less than 50 percent reporting Māori in combination with another ethnicity when they do in the census. Similar patterns are found for other level 1 ethnic groups reported singly or in combination.

**Figure 2**
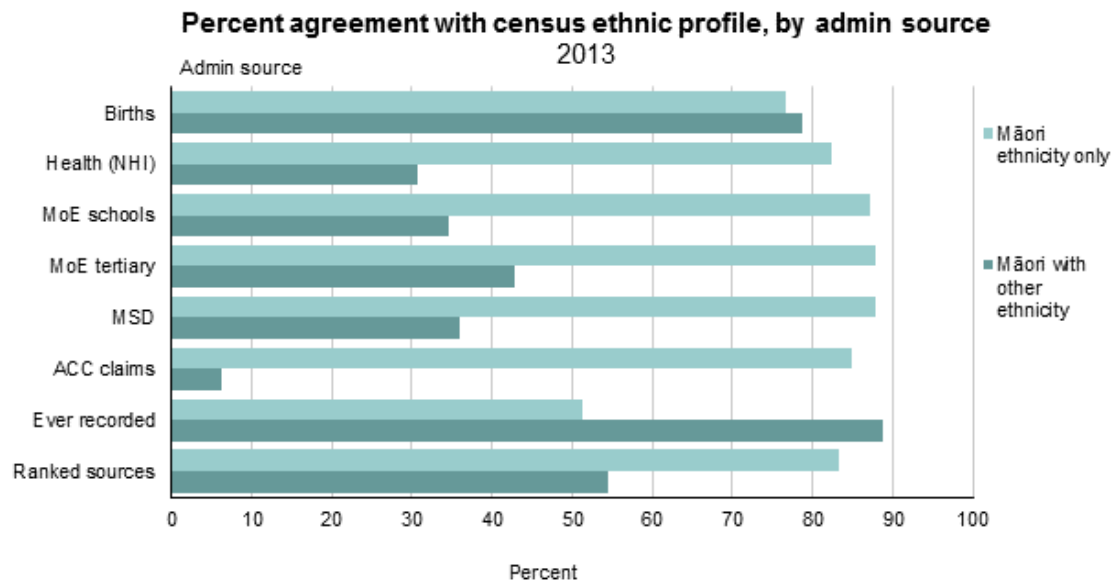


The two methods that bring together all the data sources show opposite results. The ever-recorded method has only 50 percent agreement with the census for sole Māori, but almost 90 percent agreement for Māori combined with another group. In contrast, the ranking method retains the high agreement for sole Māori seen in the individual sources, but still has lower agreement than the census for the combined Māori ethnic group.

Reid and Gleisner (2016) provide more detailed results and for all level 1 ethnic groups.
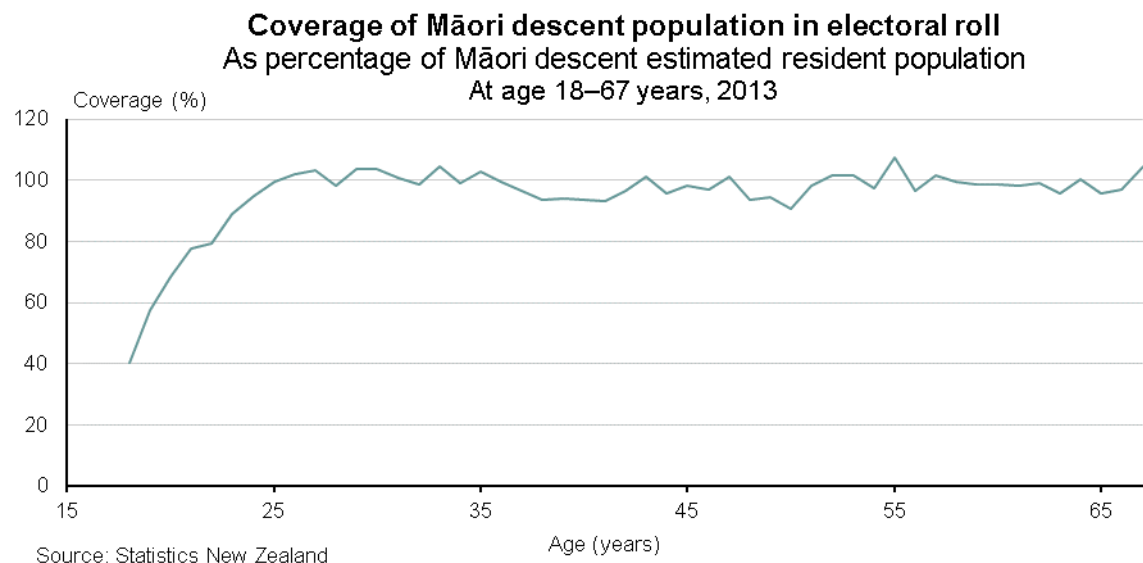
# Māori descent

## Coverage

Māori descent information is available from two government sources, electoral enrolments and birth registrations, which provide complementary coverage of adults and children, respectively.

Adults from age 18 years are entitled to vote. The electoral roll's Māori descent population in 2013 is 93 percent of the Māori descent estimated resident population (ERP) aged 18 years and over. Coverage is lower for younger adults, but from about age 25, electoral numbers hover around 100 percent of the ERP (figure 3). These aggregate results may conceal additional over-coverage and under-coverage compared with the ERP.

**Figure 3**

Coverage of Māori descent population in electoral roll
As percentage of Māori descent estimated resident population
At age 18–67 years, 2013



Source: Statistics New Zealand

Birth registrations provide Māori descent for babies born in New Zealand since 1995. Completion of the Māori descent question is high, with about 0.5 percent missing responses and a further 2.9 percent coded 'don't know' or 'unknown'. Children of Māori descent not included are those born overseas to Māori parents who return to live in New Zealand. Of people stating they were of Māori descent in the 2013 Census, 97.9 percent were born in New Zealand and 2.1 percent were born overseas. Birth registrations therefore provide close to full coverage of those with Māori descent born since 1995.

Since 1995 birth registrations have also recorded Māori descent for the parents of these babies, which provides some coverage of the adult population. However, there is no information on Māori descent for children born overseas. While we might assume most are not of Māori descent, we cannot confirm this directly.

Birth registrations before 1995 collected information on the 'degree of Māori blood'. While unsuitable as a measure of ethnicity, this may have potential as a partial source of descent information. We do not investigate this data here.

Over time, the combination of birth registrations since 1995 and electoral enrolment data is likely to provide high coverage of Māori descent for people living in New Zealand. However, change to the Electoral Act will be required to make full use of the electoral roll in a linked-data environment.

## Aggregate comparison

For this analysis we use birth registrations data from 1998, when digitisation was introduced. Registrations from 5 March 1998 are used to coincide with children aged 0 to 14 years in the 2013 Census. Since we use the linked Census-IDI data to compare the individual responses, linkage rates may affect the analysis. The first row of table 3 shows linkage results for all census children 0–14 years, but includes substitute records that cannot be linked to the IDI and have missing values for Māori descent.

Of the 764,600 children aged 0–14 years in the census with non-missing values for Māori descent, 96 percent were linked to the IDI spine and 87 percent (662,000) to their birth registration. Most of those not linked to the New Zealand birth registrations are born overseas, and respond as 'Māori descent = No'. Just 3 percent stated they were born overseas and with 'Māori descent = Yes'.

For these children linked to their birth registration, 450,000 mothers and 421,700 fathers are present in the registration data. Parents may have more than one child registered, so we expect to see more children than parents. For parents to be included in this analysis, they must first be linked to the IDI spine. Lower linkage rates in the IDI between birth parents and the IDI spine reduce the effective coverage of the parent information in birth registration records, particularly for mothers. Of all the birth parents, 52 percent of mothers and 72 percent of fathers were linked back to their census record.

We note that this final percentage is calculated using different units: 'number of individuals' (linked to the census) are divided by the 'number of records' (of parents in the birth registrations). The processes used in the IDI mean that mothers and fathers could be present more than once in the birth registrations where they have multiple children, but the parent records were not recognised as belonging to the same individual. Parents linked to the spine and then to the census can be assumed to be unique individuals.

**Table 3**
**Linkage rates and missing data for census records linked to birth registrations**
For children, mothers, and fathers, 2013

| | Base number | Linked to IDI spine | | In census-births linked data | |
|---|---|---|---|---|---|
| | | Number | Percent of base | Number | Percent of base |
| Census 0–14 years | 865,600 | 795,300 | 92 | 715,600 | 83 |
| Census 0–14 years Māori descent Yes/ No | 764,600 | 735,900 | 96 | 662,000 | 87 |
| Mothers of linked births | 450,000 | 252,500 | 56 | 236,000 | 52 |
| Fathers of linked births | 421,700 | 357,400 | 85 | 304,500 | 72 |

Using the records linked between census and birth registrations as we did for ethnicity, we compare total responses for 'Māori descent = Yes' between census and birth registrations. Table 4 shows that at an aggregate level birth registrations are very similar to the census – the ratio of 0.97 for children shows births are slightly under-reporting those with Māori descent when compared with the census. Results are consistent across children, mothers, and fathers.

**Table 4**

**Total responses for 'Māori descent=Yes' for census and birth registration records (linked data only)**

2013

|  | Census | Birth registrations | Ratio Births:Census |
|---|---|---|---|
| Children | 205,000 | 197,900 | 0.97 |
| Mothers | 50,900 | 48,600 | 0.95 |
| Fathers | 46,700 | 45,500 | 0.97 |

## Individual-level comparison

We compare Māori descent responses for birth registrations with 2013 Census responses for the same individuals, now also excluding records with missing responses in the birth registrations. Tables 5a, 5b, and 5c show the cross-classification of 'yes' and 'no' Māori descent responses, and the percent of the total in each cell.

For the children, 96 percent of individuals provide the same 'yes', or 'no' Māori descent response for births and census. The births-to-census ratio of slightly less than 1 for Māori descent = 'yes' is a result of 'no' responses in the births data where the census response was 'yes', being about twice the number that were misclassified in the opposite direction. A similar pattern is seen for mothers and fathers.

**Table 5a**

**Comparison of Māori descent response for children in birth records and census**

Counts and percentage of total (linked data only), 2013

|  |  | Births: children Māori descent response | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | Yes | | No | | Total | |
|  |  | Number | % | Number | % | Number | % |
| **Census Māori descent response** | Yes | 177,900 | 28 | 17,400 | 3 | 195,300 | 30 |
|  | No | 8,000 | 1 | 439,800 | 68 | 447,800 | 70 |
|  | Total | 185,800 | 29 | 457,300 | 71 | 643,100 | 100 |

**Table 5b**

**Comparison of Māori descent response for mothers in birth records and census**

Counts and percentage of total (linked data only), 2013

|  |  | Births: mothers Māori descent response | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | Yes | | No | | Total | |
|  |  | Number | % | Number | % | Number | % |
| **Census Māori descent response** | Yes | 45,500 | 21 | 3,200 | 1 | 48,700 | 22 |
|  | No | 1,100 | 1 | 169,200 | 77 | 170,300 | 78 |
|  | Total | 46,600 | 21 | 172,400 | 79 | 219,000 | 100 |

**Table 5c**

**Comparison of Māori descent response for fathers in birth records and census**
Counts and percentage of total (linked data only), 2013

| | | Births: fathers Māori descent response | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes | | No | | Total | |
| | | Number | % | Number | % | Number | % |
| **Census Māori descent response** | Yes | 41,000 | 15 | 3,600 | 1 | 44,600 | 16 |
| | No | 1,700 | 1 | 231,800 | 83 | 233,600 | 84 |
| | Total | 42,800 | 15 | 235,400 | 85 | 278,200 | 100 |

# Iwi affiliation

The starting point for iwi results is the New Zealand residents who answered 'yes' to Māori descent in the 2013 Census, and were linked to MoE data through either the school roll returns (schools) or tertiary qualification enrolments (tertiary).

## Coverage

Coverage of the MoE information is restricted to younger ages because students are usually children or younger adults, and because of the fairly recent start to data collection. The schools data used in this analysis is mainly for those aged 5 to 25 years in 2013; tertiary data are mainly for people aged 18 years and over, with most individuals being under 40 years.

The census obtains valid iwi responses from 80 percent of those indicating they are of Māori descent. The MoE response rates for iwi are lower, with 68 percent of tertiary data providing a valid iwi when they identified as Māori descent in the census, and just 34 percent for schools (table 6).

**Table 6**

**Valid iwi responses in the census and linked Ministry of Education data**
2013

| Base population group | Age range covered | Māori descent in Census | Valid iwi responses | |
|---|---|---|---|---|
| | | | Number | Percent |
| Census usual residents | All ages | 668,700 | 535,900 | 80 |
| Census linked to IDI | All ages | 640,000 | 513,600 | 80 |
| MoE schools linked to census | 5–25 years | 253,100 | 86,700 | 34 |
| MoE tertiary linked to census | 18 years + | 291,100 | 198,500 | 68 |

The combination of coverage limitations and missing data mean that combining MoE data from either source provides valid iwi responses for 42 percent of the 2013 Census Māori descent population.

## Aggregate comparison

As with ethnicity, people provide multiple responses for iwi. Table 7 shows the ratio of total responses for an iwi in the schools and tertiary education sources compared with the census.

Because the classification has 128 iwi, we present results for the 12 largest iwi only. The first column shows the total responses from the full 2013 Census dataset for comparison. The 'Linked Census iwi responses' columns show the total number of responses received for each iwi in the census, restricted to individuals who linked to the education datasets. The schools and tertiary iwi responses columns are the number of responses for that iwi in the education data linked to the census. The ratio is the education total divided by the linked census total. As before, a ratio close to 1 indicates high consistency between the census and education data.

Note: the education datasets in the IDI can contain multiple records for an individual, corresponding to distinct enrolments across different institutions or years. Iwi information in each record can vary. For this paper, we use the record containing the most iwi responses, and have not attempted to reconcile or combine responses from different records.

**Table 7**

**Total responses for 12 largest iwi**

Census and Ministry of Education (linked data only)

| Iwi name | 2013 Census | Schools | | | Tertiary | | |
|---|---|---|---|---|---|---|---|
| | Total iwi responses | Linked Census iwi responses | Iwi responses | Ratio admin: census | Linked Census iwi responses | Iwi responses | Ratio admin: census |
| Ngāpuhi | 125,600 | 49,900 | 22,300 | 0.45 | 53,800 | 46,600 | 0.87 |
| Ngāti Porou | 71,000 | 28,300 | 12,600 | 0.44 | 32,500 | 29,800 | 0.92 |
| Ngāi Tahu | 54,800 | 20,900 | 6,300 | 0.30 | 24,000 | 16,300 | 0.68 |
| Waikato | 40,100 | 16,000 | 1,700 | 0.11 | 18,000 | 5,900 | 0.32 |
| Ngāti Tuwharetoa | 35,900 | 14,400 | 5,200 | 0.36 | 15,900 | 12,400 | 0.78 |
| Ngāti Maniapoto | 35,400 | 13,800 | 4,700 | 0.34 | 15,900 | 11,500 | 0.72 |
| Tuhoe | 34,900 | 14,400 | 7,000 | 0.49 | 15,200 | 13,900 | 0.91 |
| Ngāti Kahungunu ki Te Wairoa | 21,100 | 8,200 | 800 | 0.10 | 10,000 | 3,600 | 0.36 |
| Te Arawa | 19,700 | 8,000 | 5,000 | 0.63 | 8,900 | 11,100 | 1.24 |
| Ngāti Kahungunu, reg unspecified | 18,300 | 7,500 | 7,000 | 0.94 | 7,900 | 14,100 | 1.78 |
| Te Rawara | 16,500 | 6,200 | 1,900 | 0.30 | 7,800 | 5,400 | 0.69 |
| Ngāti Awa | 16,200 | 6,300 | 2,800 | 0.45 | 7,600 | 6,600 | 0.87 |

All but two of the ratios are less than 1, indicating that the education data consistently undercounts iwi affiliation when compared with the census. Ratios for schools data, being mainly less than 0.5, are lower than tertiary institutions – where most are between 0.7 and 0.9. There is also considerable variation among iwi. Tuhoe, Ngāti Porou, Ngāpuhi, and Ngāti Awa show ratios closest to 1 in both schools and tertiary data, while Waikato has the lowest ratio in both sources. Differences in people's responses and coding practices are likely to be affecting results for the two Ngāti Kahungunu categories.

We note that MoE reports their results on iwi regional groupings, rather than individual iwi. This may improve some results, particularly where there are coding difficulties.

## Individual-level comparison

For people who provide an iwi in the education data, we can also look at how consistent the reported iwi affiliation is with the iwi recorded for the same person in the census. Results in figure 3 are for the 12 largest iwi. Of individuals who reported a specific iwi in the census, 20 to 40 percent reported belonging to the same iwi in the schools data. The tertiary results are better, with 30 to 60 percent belonging to the same iwi as stated in the census. The same four iwi perform best in this individual-level analysis, and Waikato again shows the lowest consistency between census and MoE data.

**Figure 4**

### Percent agreement with census iwi – schools and tertiary data sources
#### 2013
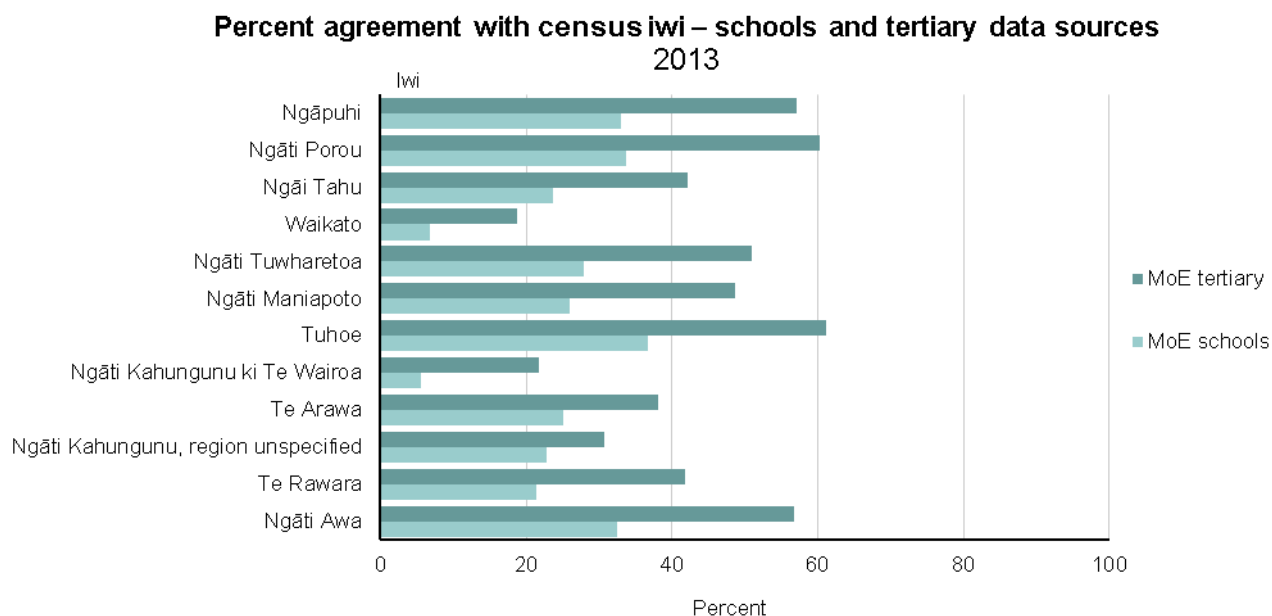


Source: Statistics New Zealand

# Summary

Table 8 summarises results by coverage for the agency, and quality of the data collected for each variable. The high, medium, or low quality ratings are somewhat subjective, but are an attempt to provide a reasonable assessment of the findings reported here.

**Table 8**
**Summary of quality measures for administrative sources for four essential variables**

| Source | Coverage of source | Quality of administrative responses | | | |
|---|---|---|---|---|---|
| | | **Māori ethnicity** | **Māori descent** | **Iwi** | **Te reo** |
| DIA Registered births | • 0–14 years<br>• All NZ born | HIGH | HIGH | … | … |
| Ministry of Health | • All ages<br>• Most NZ residents | MEDIUM | … | … | … |
| MoE schools | • 5–25 years<br>• Most NZ residents | MEDIUM | … | LOW | LOW |
| MoE tertiary | • 18 to under 40<br>• Students | MEDIUM | … | MEDIUM | LOW |
| Ministry of Social Development | • 15 years +<br>• Working-age benefit receipt | MEDIUM | … | … | … |
| ACC | • 15 years +<br>• Injury claimants | LOW | … | … | … |
| Symbol: … not applicable | | | | | |

# 7  Discussion

Four variables have been identified by the Census Transformation programme as essential census information requirements specific to Māori: Māori descent, Māori ethnicity, iwi, and te reo Māori (the Māori language). This paper summarises the availability and quality of administrative data sources for these critical census information needs. While the potential for a future census based on administrative sources provides the context for this investigation, the findings are relevant for other uses of the administrative data.

These four variables are collected by a number of government agencies. We have examined the statistical properties of government administrative data sources by considering three aspects: the concepts underlying the data collection, coverage, and measurement error.

The relevance of the data in the census context is determined by how close the concepts, definitions, and questions are to the statistical concept we wish to understand. Coverage tells us how much of the population we are able to obtain information for. Measurement error looks at the accuracy of the data that is collected.

## Consistency with statistical standards

We considered consistency with the concepts, definitions, and guidelines of the statistical standards, and the use of the standard classifications for each variable. For ethnicity, descent, and iwi, we found that government agencies do now largely collect data for these variables in way that is consistent with the key concepts of the standards. These developments are relatively recent and coincide with the development of the current standards from the mid-1990s. Older data is often not fit for purpose, and it is important that data collected before certain dates can be removed for analysis.

However, we found marked variation in the questions used. Some agencies adhere closely to the question guidance given in the standard, and the question used by the census, while other agencies use a wide variety of form types and questions.

## Coverage

The census provides information for people of all ages. Agencies collect information from people who interact with their particular service. Only the Ministry of Health achieves high coverage of New Zealand residents across all ages; the coverage of other agencies is limited by the nature of their service and also by the year from which data is available. Using linked data (as we have in the IDI) means that coverage gaps in one data source may be filled by other sources. Only ethnicity currently shows almost full coverage from the available administrative sources. Implications of using a system of linked data include the need to be able to identify unique individuals in the source data and for linkages to be accurate.

When combining data from different sources, better methods need to be developed to deal with inconsistent or conflicting data from different sources.

## Measurement error

Measurement error is considered by comparing the 2013 Census with the administrative sources, based on the linked Census-IDI dataset. We compare the aggregate results that would be obtained from the administrative sources with those from the census. We also compare the values recorded for an individual in the administrative sources against those recorded for the same individual in the census. This analysis produces a number of measures of consistency between the administrative data and the census. Strong consistency across all these measures suggests that data in both sources are accurate. Where there are differences, conclusions are less obvious. In some cases causes may

be clear – for example: where multiple responses for ethnicity are not retained, response patterns are very different; the use of different questions seems likely to be a significant cause of different responses; large amounts of missing data suggest underlying problems with data collection.

However, no two sources will ever provide exactly the same responses. There is an underlying variation in people's responses over time, and response depends on context. This is a feature of ethnicity data collection but is also seen in what might be considered more stable attributes such as descent. Inevitably, some errors are introduced during subsequent processing for both surveys and administrative sources, including data linkages. Of all the data sources examined, DIA's birth registrations data since 1998 shows the highest consistency with the census. The level of agreement seen here may be as high as can be expected anywhere across the system.

## Māori organisations collection of statistical data

Statistics NZ will need to work in close partnership with iwi and other Māori organisations towards improving iwi information collected within government, and to establish a potential role for iwi registers in contributing to statistical information for and about Māori. McNally and Gleisner (2015) provide more detail on these issues.

## Government collection of statistical data

There is a question about what kind of environment government agencies need to operate within in order to collect good statistical information. Registration of a birth is a legal requirement, and DIA has a centralised system of data collection. DIA collects ethnicity and descent for statistical purposes and has worked closely with Statistics NZ to ensure the statistical standard is followed and the quality of data is maintained over time. The registration questions are almost identical with the census questions.

In contrast, the challenges of gathering information from a large and widely dispersed collection system are revealed in the high levels of missing data and low consistency of some data sources compared with the census. This is despite providing good guidelines, and use of the standard classifications.

A more coordinated approach by government might rationalise data collection, concentrating resources to achieve high quality in a small number of agencies and allow core demographic information to be shared between agencies. Collection of data about indigenous populations will only succeed where there is a genuine partnership between Māori organisations and government agencies.

## Conclusion

While some administrative sources provide very good information for and about Māori, the lack of completeness and lower quality of other sources means that administrative data cannot at present replace these essential information needs provided by the current survey-based census. There is some promise that ethnicity and Māori descent could be provided in future censuses through linked administrative data sources, though for iwi information this is more uncertain. Key requirements are:

- improvements to the quality of ethnicity data collected by government

- a source of Māori descent information for adults (possibly through access to electoral roll data)

- government to work in partnership with iwi to develop iwi information sources.

Te reo proficiency is not suitable for collection through administrative sources. Therefore te reo information will require continued survey collection. It seems likely that information about iwi will also need to be obtained through surveys for some time to come.

# 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 and the results in this paper are 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 Privacy impact assessment for the Integrated Data Infrastructure (available from www.stats.govt.nz) for more details.

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

# References

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.

Education Counts (nd). Iwi data: Collection and use. Retrieved 29 April 2016 from www.educationcounts.govt.nz.

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

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

Statistics NZ (nd). Classifications and related standards. Retrieved 29 April 2016 from www.stats.govt.nz.

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

Statistics New Zealand (2013). Census QuickStats about Māori. Wellington, New Zealand. Retrieved from www.stats.govt.nz.

Statistics New Zealand (2014a). Understanding substitution and imputation in the 2013 Census. Retrieved from www.stats.govt.nz.

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

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

Tūhono (nd) Tūhono Trust. Retrieved from www.tuhono.net.

Walling, J, Small-Rodriguez, D, Kukutai, T (2009). Tallying tribes: Waikato-Tainui in the census and iwi register. Social Policy Journal of New Zealand 36. Available from www.academia.edu.

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: The IDI

**Appendix figure 1**
**Structure of the IDI, at May 2015**