2018 Census: How we are planning to adjust for missing information

2018 Census: How we are planning to adjust for missing information provides an introduction to the statistical terms and processes used to adjust for missing information, or 'non-response' in the census.

It updates an earlier paper *2018 Census: Technical support information* published in June 2018.

Summary of key points

Stats NZ is making use of data from alternative sources to adjust for missing information from the 2018 Census by including:

- real information from administrative sources
- real information from the 2013 Census.

Statistical methods can be applied when alternative sources are not available:

- unit imputation adds to the census count
- item imputation fills in missing census characteristics.

A post-enumeration survey will provide an independent check on the accuracy of the census count.

Introduction

The census is a count of the population and the dwellings in New Zealand. Every five years Stats NZ asks everyone in the country to provide information about themselves and their household. This information is used to help make informed decisions by government, iwi, businesses, and community groups across the country.

In 2018, we introduced a new model for collecting the information, with a focus on online participation and paper available on request, followed by postal reminders and household visits for people who hadn't taken part.

After census day on 6 March, we deployed a range of planned follow-up activities to encourage the people who hadn't taken part to provide us with their completed forms – either online, or by paper. We also extended the field operation by two weeks to help reach more people.

Because participation in the 2018 Census was lower than planned, we need more time than originally anticipated to draw on other information sources and new methods to finalise a fit-for-purpose dataset.

When Stats NZ produces any statistical information, we use a range of statistical methods to adjust for limitations in the data collected. This includes adjustments for missing data to improve the quality of the published information. While the census aims to collect information directly from all people in New Zealand on census night, it is usual for some people not to fill in census forms, or submit forms with some questions unanswered. Not responding to census and surveys nearly always disproportionately affects some groups more than others. If we were to only use the information from people who completed a question, this would mean counts were lower than in reality and proportions biased towards categories that tend to get responses at higher rates.

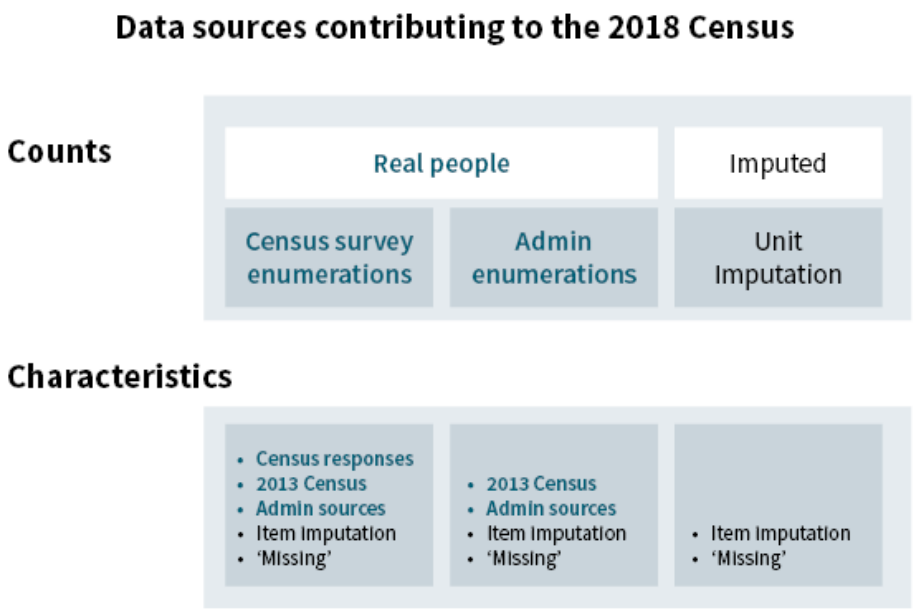
Stats NZ, in common with other countries, uses statistical processes to compensate for missing data. These processes are designed to reduce the bias caused by differences between the people who do respond and those who do not.

For the first time, Stats NZ is making use of data from alternative sources for the 2018 Census. The 2018 Census will use this data to add people to the census count. This is known as an 'admin enumeration'. We are using data from administrative data sources and the 2013 Census to fill in gaps in census responses. When we use the 2013 Census and admin sources, we are obtaining information from real people, but provided at a different time or in a different context.

Statistical methods called 'imputation' can be applied when the alternative sources are not available. We have also extended our previous use of statistical imputation methods to fill in missing data where possible.

The sources contributing to the 2018 Census file are shown in figure 1. The census *count* is made up of people who responded to the census, plus people added through admin enumerations and imputation. The *characteristics* of those people will come from the best available source – census responses, the previous census in 2013, admin data, imputation – or will be left as missing. The details vary depending on the variable and the source of the count. These terms are described in the following sections.

Figure 1



Alternative sources for census information

We will use the 2013 Census and other government data about real people, such as births, tax, health, and education records, to compensate for the information we are missing.

Much of this information is held in Stats NZ's Integrated Data Infrastructure (IDI). The IDI is a large research database that contains a number of datasets about people and households from government agencies, Stats NZ surveys, and non-government organisations. It also holds data from the 2013 Census.

Stats NZ has undertaken a privacy impact assessment (PIA) to use admin data in the IDI to supplement the 2018 Census data (see Privacy impact assessment for using admin data in 2018 Census data). This PIA assesses the risks identified against privacy principles and outlines how we will mitigate these risks.

Including people from administrative data – admin enumerations

We will use information from administrative sources – admin enumerations – so that some of the people who have not responded to the census can be included.

In 2018 we had more households than in the past where no one responded to the census. In previous censuses we used imputation to improve the census count. This time we are using government data as well as imputation to help us fill in some of the information gaps, rather than relying solely on imputation.

Research from our Census Transformation team puts us in a good position to develop appropriate models using administrative sources. Our work over the last four years has shown that we have good coverage of the New Zealand population from the wide range of government data we hold. This is an opportunity to make greater use of administrative data in the 2018 Census than anticipated and fits with the long-term census transformation strategy.

The work involves assessing how many people to bring into the census data file from administrative data sets, and exactly how we will do this. Initial analysis suggests these admin enumerations should improve the representation of people in the census data set.

We are now testing these methods and working through the implications of trading-off timeliness and effectiveness for two main statistical options. We'll make decisions over the next few months on the best approach to produce census data.

In conducting this work, we are drawing on the support of our colleagues in other countries who are developing similar methods for their own censuses, and our external data quality panel. The work needs careful and time-intensive evaluation to ensure we produce a useful census data set.

Filling out census characteristics from 2013 Census and administrative data

In 2018, we will use real information from the 2013 Census and administrative sources to fill in missing characteristics where we can. We had already planned to use these alternative sources for incomplete census responses where we can link to admin records (see 2018 Census – a modernised, digital-first census). An advantage of admin enumerations is that they are already linked to the IDI, and we can use real information for census characteristics as well.

We use 2013 Census and administrative sources to fill in census characteristics for:

- people who have responded to the census, but have not answered some questions or the answer is not able to be used (item non-response)
- admin enumerations when we also use alternative sources to fill out characteristics of those people.

The 2013 Census is a good source of information for variables that do not tend to change much over time, for example, country of birth, religion and number of bedrooms.

Admin data sources vary in quality and suitability for statistical use. Admin data will be used to fill gaps in the census when we are confident that it will improve the quality of the census file. We are in a good position to adopt this approach as we have been investigating future census models that would make much greater use of admin data for the census.

Imputation

Admin sources and the 2013 Census will not provide all the missing census information. Statistical methods called 'imputation' can be applied when the alternative sources are not available. As with alternative information sources, we use imputation methods:

- to add to the census count – unit imputation
- to fill in missing census characteristics – item imputation.

While previous censuses have included both types of imputation, we have made some changes for 2018.

Unit imputation

Unit imputation helps improve the coverage of the census. Unit imputations are individual census records created where there is sufficient evidence that a person exists, but we have not received a

census response for them, and we have not included them from administrative data. In previous censuses we have used 'donor imputation' to impute members of fully non-responding households, choosing a responding household from the same neighbourhood to be the 'donor' (see Understanding substitution and imputation in the 2013 Census).

In 2018, this approach will be used for 'fully non-responding households' where we have evidence that the dwelling was occupied on census night, but no census information was received from the household, and we have not created an admin household. Individual records will be created for members of the imputed household.

A more complex methodology is also being explored whereby people not counted in the census could be imputed, without the need to know which dwelling they belong to.

The rate of unit imputation has increased steadily over time, from 2.9 percent of the total number of people counted in the 2001 Census, to 3.3 percent in 2006 and 4.8 percent in 2013 (see Post-enumeration Survey: 2013 and A Report on the 2006 Post-enumeration Survey). This rate refers to the total number of individual records created for partial and fully non-responding households and is the proportion of the census count not completed by respondents. We do not yet know the extent of unit imputation in the 2018 Census.

In a change from previous censuses, people who are listed on a household form are treated as responses, even when no individual form has been received. There is no unit imputation for 'partially responding households' in 2018.

The number of people listed as belonging to a household, but where no individual form was received, increased in the 2018 Census. The household listing includes age, sex, and place of residence so we have good information to count these people as a response, even though we do not have an individual form for them. In some cases, we will be able to use information from the 2013 Census and administrative data to fill out missing variables.

Item imputation

Since at least the 1996 Census, we have imputed four key census variables: sex, age, usual residence meshblock, and work and labour force status. The methods in previous censuses used information provided by census respondents and known variable distribution patterns (see Understanding substitution and imputation in the 2013 Census). For all other questions that were unanswered, we coded the respective variables to 'not stated'.

In 2018, we will use item imputation when we do not have good data from the 2013 Census or admin sources. We use a different methodology from previous censuses for item imputation that fills in the missing variables by 'borrowing' information from similar people or similar households that have responded. The item imputation is implemented in the CANCEIS software, developed by Statistics Canada, and is used by a number of other countries in their censuses (see Editing the 2011 Census with CANCEIS and options considered for 2016 [PDF, 9p]).

In the 2018 Census, item imputation will be used for a wider range of variables than in the past (see 2018 Census – a modernised, digital-first census). However, there are still some variables where we do not have good alternative sources and we will not be able to impute, for example the housing quality variables that were introduced in the 2018 Census.

Uncertainty due to donor imputation

The use of the previous census and administrative data to adjust for missing data helps to counteract the impact of census non-response, but cannot fill all the gaps. Donor imputation is an established statistical method that relies on the assumption that we are able to find 'similar' households, so that filling in the gaps produces a representative distribution in the finalised census data. This assumption, known as 'missing at random', is harder to defend in areas where there are high levels of non-response, and for variables where we cannot obtain information either from the previous census (because the information is likely to have changed) or from administrative sources (because data is not available).

For the first time in a New Zealand census, in 2018 we will measure the uncertainty in the census counts due to household donor imputation. This will allow us to assess our confidence in the unit imputation and provide a measure of uncertainty for users of census data.

Measuring coverage

Given the strategic significance of the census data and its diverse applications, Stats NZ, like other national statistical organisations, makes concerted efforts to ensure we understand the quality of the census data we produce. Along with evaluating our methods for imputation, another important set of data we need is the results of our 2018 Post-enumeration Survey, which is underway now.

A post-enumeration survey is an independent check on the accuracy of the census count and aims to provide information on the completeness of census coverage (see Post-enumeration Survey: 2013). The survey provides a gauge of how many New Zealand residents were missed, or counted more than once, in the census.

Census coverage relates the number of people who were counted in the census to the number who should have been counted. It is usually expressed as a percentage of what should have been the complete count (for example, the expected number of usual residents in New Zealand on census night).

The full picture, just like in every other census, won't be available for some time, but we will start releasing data from the census as soon as we can.

Census Transformation in New Zealand

Stats NZ's Census Transformation programme is investigating the feasibility of a future census based on administrative data (see Census transformation in New Zealand). The research into the availability and quality of administrative data for census information has formed the basis of the use of administrative data in the 2018 Census for item-non-response. (See Census transformation – research papers, The potential for linked administrative data to provide household and family information, and Experimental population estimates from linked administrative data.)

Note

2018 Census: Technical support information was published on 1 June 2018. It provided a brief introduction to the statistical terms and processes used to measure and adjust for missing information. Since then we have revised the methods we will use, and this earlier document is no longer applicable.

Citation
Stats NZ (2018). *2018 Census: How we are planning to adjust for missing information*. Retrieved from www.stats.govt.nz.

ISBN 978-1-98-858304-4

Published in December 2018 by
Stats NZ Tatauranga Aotearoa
Wellington, New Zealand

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

www.stats.govt.nz