2018 Census: Potential impacts of revised methodology

This paper presents the context to help customers understand the revised methodology for producing the 2018 Census dataset. It outlines the potential data quality impacts and the approaches we will use to reduce these impacts.

This paper is intended to prompt discussion and support our commitment to engaging with customers as we analyse and process 2018 Census data.

Summary

The 2018 Census of Population and Dwellings differed from previous censuses. This year we used a digital-first approach, which encouraged respondents to complete the census online.

Parts of the census operation were successful. Interim figures showed we surpassed our online participation target of 70 percent. However, some aspects were more challenging than expected. While we are yet unable to produce a definitive response rate, the number of individual responses we received is lower than expected. Interim figures show that we have full or partial information for at least 90 percent of individuals, compared with 94.5 percent for the 2013 Census.

When we produce statistical information, we use a range of statistical methods to adjust for limitations in the data collected and to understand the quality of the results produced. To compensate for the missing data, we will be using revised methodology that involves using other sources of information and different imputation approaches to achieve the highest-quality dataset possible.

Note that because of ongoing work on the revised methodology, we are now looking towards a first release of census data in March 2019.

Background to 2018 Census

The census is the official count of New Zealand's population and dwellings. Every five years, Stats NZ asks everyone in the country to provide information about themselves and their dwellings. The data from the census is used by government, iwi, businesses, and community groups across the country to make informed decisions. Census data enables accurate population projections and estimates. It underpins decisions about policy development, funding allocation, and service provision, and determines how billions of dollars of government funding is spent. The census counts are also key to determining the general and Māori electorate boundaries, and calculates the number of general and Māori electorates.

The main advantage of the census over other surveys is that census data is available at a neighbourhood level and provides detailed characteristics of small population groups.

For the 2018 Census we used a new model for collecting the information. We focused on online participation followed up with postal reminders and household visits for those who had not taken part on census day.

After census day on 6 March, we began follow-up activities to encourage those who had not taken part to complete their forms – either online or by paper form. We also extended the field operation by two weeks to boost the uptake, completing nearly one million visits within a seven-week follow-up period.

One of the goals of the 2018 Census was to improve data quality while modernising. The objective was to ensure accuracy of national counts and reduce variation in subnational response rates (see 2018 Census strategy). Our interim calculations show that we have not reached our target coverage rates of 94 percent or higher.[1] However, our interim calculations show full or partial information for about 90 percent of individuals. The interim response rate varies across subpopulations and small geographic areas. To be able to provide good-quality information for all subpopulations and small geographic areas, we need to have very high coverage and response rates across **all** of New Zealand.

Potential impacts

This section presents the potential impacts of the revised methodology.

It is normal for some people to not fill in census forms, or for forms to have some unanswered questions. New Zealand, like other countries, uses a statistical process called 'imputation' to improve the quality of census data. Imputation involves inserting a value when a respondent has not provided a valid response.

Given the interim position of individual response rates for the 2018 Census, we are looking at expanding our imputation approach. We are investigating how we can impute households, and cases of item non-response. Both item and unit imputation will improve data coverage and, occasionally, data quality, but not for all census variables. If we do not impute, there will be large amounts of missing data that will affect the overall quality of the dataset.

We are in the early stages of analysing census responses. Looking at the potential impacts of the revised methodology will help us to better understand what we are missing and where the most significant gaps are.

[1] At a national level, the 1996 Census achieved net coverage of 98.4 percent and a response rate of 95.6 percent. The 2001 Census achieved net coverage of 97.8 percent and a response rate of 95 percent. The 2006 Census achieved net coverage of 98 percent and a response rate of 94.8 percent. The 2013 Census followed this downward trend, achieving net coverage of 97.6 percent and a response rate of 92.9 percent. These national rates may be higher or significantly lower for subpopulations and small population areas.

Extending our imputation approach may help fill those gaps and mitigate potential impacts on the data. By doing this we will be addressing:

- **Small area (SA) counts.** Response rates vary significantly in many small areas for the 2018 Census (down to SA1 and SA2) (see Statistical standard for geographic areas 2018). Some areas will be improved through imputation while others will still have data quality issues.
- **Time series.** Consistency is an important aspect of the census and comparisons over time will be affected.
- **Ability to cross-tabulate attribute data for people and dwellings in small geographies**. The key variable of ethnicity for example, is likely to be affected by a lower response but can be improved by imputation.
- **Variables or geographic areas that may not be publishable due to low quality**. Explanations about data quality and other measures of accuracy may be required to help customers use the data effectively.

More extensive metadata will also be required to:

- communicate the confidence in outputs
- support explanations about the time series.

Other impacts

This section outlines other impacts that may occur regardless of whether we impute or not.

- Lower response rates could have downstream effects on population estimates and projections – at this stage we are uncertain what these may be. This may impact on delivery schedules and data quality.
- Some key data areas that make the census unique, including family and household structure, could be affected by lower response rates. Due to their unique nature, it may be difficult to find alternative data sources that can be used for imputation. In addition, we do not have a methodology for repatriating absent household members back to their usual residence. This issue will be compounded when there is a need to cross-tabulate with other 2018 Census attribute data.
- Some SA2s and rural settlements with small populations and potentially in 'hard to reach/enumerate' areas (eg business SA2s, remote land and water SA2s) will be affected. Imputation of administrative data might not be a good approach for these areas due to a lack of address information.
- There is no administrative data available for new content, such as housing quality and activity limitations. Donor imputation will go some way towards improving this data.
- No administrative data sources are available for some variables including iwi, occupation, household, and families. Donor imputation will help towards improving this data.

Addressing the impacts through imputation

This section describes the imputation methods we will use and how we plan to address and reduce their potential impacts on the data.

Unit imputation

Unit imputation is used to add to the census count where we have sufficient evidence from the collection process that a person exists or a dwelling was occupied, but we did not receive a corresponding form.

Unit imputation helps improve the coverage of the census (see measuring coverage for more information on census coverage). Unit imputation (formerly known as substitute records) are individual census records created where there is sufficient evidence that a person exists, but we have not received an individual census form for them.

This can occur in two situations:

- partially responding household – an individual record is created in an occupied dwelling where information on the number of census night occupants has been received from the household
- fully non-responding household – the dwelling was occupied on census night, but no census information has been received from the household. Individual records are created for all the members of the household.

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. 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.

Partial-response dwellings have increased in the 2018 Census. However, we still have a good basis for unit imputation to contribute to the census count. For a partial-response dwelling, we have a dwelling form or household summary page which lists the people at the dwelling on census night. This provides us with the number of people in the household despite not receiving an individual form from all of them.

In some cases, we will use information from the 2013 Census and administrative data to populate missing variables.

Tables 1 and 2 show the sources we will use for imputing selected individual and household census variables.

Table 1 – Imputation sources for individual variables

Information	Source for imputation	Administrative data source
Usual residence five years ago	Census	2013 Census
Country of birth	Census and administrative datasets	Department of Internal Affairs; Ministry of Business, Innovation and Employment (Department of Labour)
Years since arrival in New Zealand	Census and administrative datasets	Department of Internal Affairs; Ministry of Business, Innovation and Employment (Department of Labour)

Information	Source for imputation	Administrative data source
Ethnicity	Census and administrative datasets	Department of Internal Affairs; Ministry of Business, Innovation and Employment (Department of Labour)
Language	Census	2013 Census
Māori descent	Census and administrative datasets	Department of Internal Affairs
Religion	Census	2013 Census
Regular smoker	Census	2013 Census
Ever smoked	Census	2013 Census
Study participation	Ministry of Education	Ministry of Education
Highest secondary school qualification	Census and administrative datasets	Ministry of Education
Post-school qualification	Census and administrative datasets	Ministry of Education
Source of income	Inland Revenue	Inland Revenue
Annual income	Inland Revenue	Inland Revenue
Employer	Inland Revenue	Inland Revenue

Table 2 – Imputation sources for household variables

Information	Source for imputation	Administrative data source
Dwelling type	Census and administrative datasets	Tenancy bonds
Room count	Census	2013 Census
Bedroom count	Census and administrative datasets	Tenancy bonds
Tenure of household	Census and administrative datasets	Tenancy bonds
Sector of landlord	Tenancy bonds	Tenancy bonds
Rent amount	Tenancy bonds	Tenancy bonds
Rent period	Tenancy bonds	Tenancy bonds

In 2018, there are more households where no one has responded to the census than previous censuses. In previous censuses, we used donor imputation to impute members of fully non-responding households, choosing 'donors' from a responding household in the same neighbourhood. This time more work is required in our use of government data to help fill some information gaps for those households, rather than relying solely on donor imputation.

It is more difficult to use administrative data to reliably derive whole households. However, we are well-positioned to develop appropriate models based on work carried out by our Census

Transformation programme. This is an opportunity to make greater use of administrative data in the 2018 Census than anticipated, and aligns with the long-term Census Transformation strategy.

For the remainder of fully non-responding households where we cannot use the administrative data to predict the members of the household, we will continue to use donor imputation as we have done in previous censuses.

Item imputation

Item imputation fills in values where a census form has been provided, but some questions are unanswered. A number of imputation methods can be used to ensure the final census data reflects the population structure and its characteristics as closely as possible.

Since the 2001 Census, we have imputed four key census variables – sex, age, usual residence meshblock, and work and labour force status. Item imputation methods in previous censuses used information provided by census respondents and known variable distribution patterns. For all other unanswered questions, we coded the respective variables to 'not stated'.

For the 2018 Census, we will use item imputation for a wider range of variables. For the first time, we will also use data from the 2013 Census and select government administrative sources. If the data from these sources is inadequate, the imputation will fill in the missing variables by 'borrowing' information from similar people and dwellings that have responded. This imputation method is known as 'donor imputation'.

Donor imputation

The use of previous census data and administrative sources for imputation mitigates the impact of census non-response, but cannot fill all the gaps. The second step is to use donor imputation. Donor imputation relies on the assumption that we are able to find similar households, allowing the filled gaps to produce a representative distribution in the data. This assumption, known as 'missing at random', is harder to carry out in areas where there are high levels of non-response. It is also difficult for variables where we cannot use information from the previous census, as the information may have changed, or from administrative sources.

Part of our census evaluation involves developing methods to assess our confidence in donor imputation.

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, an important information source is the 2018 Post-enumeration survey (PES). The 2018 PES is underway.

The PES, an independent check on the accuracy of coverage in the census, aims to provide information on the completeness of census coverage. It gauges 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 (eg the expected number of New Zealand 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 are confident that we have produced the highest-quality dataset possible.

Census transformation

Stats NZ's Census Transformation programme is investigating the feasibility of a future census based on administrative data. 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 related information for links to census transformation papers.

Much of the administrative data we have investigated comes from the Integrated Data Infrastructure (IDI). The IDI is a large research database. It holds microdata about people and households. The data is about life events like education, income, benefits, migration, justice, and health. Information is gathered from government agencies, Stats NZ surveys, and non-government organisations (NGOs). The data is linked together, or integrated, to form the IDI. The 2013 Census is linked into the IDI.

Related information

Here are links to related pages.

Understanding substitution and imputation in the 2013 Census

2018 Census – a modernised, digital-first census

2018 Census data quality management strategy

Post-enumeration Survey: 2013

A report on the 2006 Post-enumeration Survey

Census Transformation programme

Census transformation – research papers

The potential for linked administrative data to provide household and family information

Experimental population estimates from linked administrative data

Integrated Data Infrastructure

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

Citation
Stats NZ (2018). *2018 Census: Potential impacts of revised methodology*. Retrieved from www.stats.govt.nz.

ISBN 978-1-98-852880-9

Published in July 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