Stats NZ
Tatauranga Aotearoa

2018 Census: Changes and how they might affect data

New Zealand Government

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

Citation
Stats NZ (2018). *2018 Census: Changes and how they might affect the data*. Retrieved from www.stats.govt.nz.

ISBN 978-1-98-852873-1 (online)

Published in May 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

Contents

List of tables and figures

List of tables

List of figures

Purpose and summary

2018 Census: Changes and how they might affect the data describes the changes made to content and operational processes of the 2018 Census of Population and Dwelling to better reflect the needs of customers, and how these changes might affect the comparability of the data over time.

The changes relate to the classifications, imputation, and questionnaire design; the way the data is processed, variable by variable; and the predicted impacts on the data. Where relevant, we include changes due to actual changes in the population ('real-world changes'). The new model we employed in the 2018 Census to better reflect the needs of customers has resulted in changes that may impact on the comparability of the data over time. We outline the anticipated impacts this report.

Summary

The 2018 New Zealand Census is a change census, meaning that for the first time since 2001, the variables collected in the census have undergone significant review. With the change model, and a general update of systems, there are considerations to be made for data quality. We have made every effort to ensure data quality remains high, and the trust our data users and members of the public place in the census is preserved. There are instances where changes may have an impact on the quality of data, although in many cases these changes will lead to positive outcomes.

We outline possible implications of the changes in content and process of 2018 Census for each of the variables. While we've tried to limit the impacts on comparability with previous data there are instances where this was not possible, given the changed approach for 2018 and need to refresh and update content. Until the data is analysed it is impossible to fully understand the impacts of the changes, therefore our assessments are based on what we knew leading into the 2018 Census operation. Following the analysis of the 2018 Census data, we will publish metadata and further information about any impacts of changes on 2018 Census data.

Table 1 summarises the predicted level of change for each variable, with priority 1 variables first. Appendix 1 summaries the likelihood of eight types of change for each variable.

Table 1

Predicted level of real-world change and data collection change for 2018 variable		
Variable	**Level of real-world change (predicted)**	**Level of data collection change (predicted)**
Age/Date of Birth	Moderate	Minor
Sex	Moderate	Minor
Usual residence	Moderate	Minor
Census night address	Moderate	Minor
Ethnicity	Moderate	Minor
Māori descent	Minor	Moderate
Iwi affiliation	Moderate	Moderate
Birthplace	Moderate	Minor / Moderate
Usual residence one year ago	Not applicable	Not applicable
Legally registered relationship status	Minor	Minor
Partnership status in current relationship	Minor	Minor
Household composition, family type, extended family type and child dependency status	Minor	Major
Main means of travel to education and Educational institution address	Not applicable	Not applicable
Highest Secondary School Qualification	Minor	Minor
Post-school qualification	Minor	Moderate / Major
Field of study	Minor	Moderate
Highest qualification (derived variable)	Minor	Moderate / Major
Sources of personal income	Minor	Moderate
Total personal income	Moderate	Moderate
Work and labour force status	Minor / Moderate	Minor / Moderate

Hours worked in employment per week	Minor	Minor
Status in employment	Minor / Moderate	Minor
Main means of travel to work	Minor	Minor / Moderate
Occupied dwelling type	Moderate	Moderate
Tenure of household	Minor	Minor
Years at usual residence	Minor	Minor
Years since arrival in New Zealand	Moderate	Moderate
Languages spoken	Moderate	Minor
Religious affiliation	Moderate	Major
Study participation	Minor	Major
Disability / activity limitations	Minor / Moderate	Major
Cigarette smoking behaviour	Moderate	Minor
Number of children born	Minor	Minor
Individual home ownership	Minor	Minor
Occupation	Minor / Moderate	Minor / Moderate
Industry	Minor / Moderate	Minor / Moderate
Sector of ownership	Minor / Moderate	Minor
Unpaid activities	Minor	Minor
Number of rooms and bedrooms	Minor	Moderate
Main types of heating	Moderate / Major	Major
Access to telecommunication systems	Major	Minor
Number of motor vehicles	Minor	Minor/Moderate
Housing quality	Not applicable	Not applicable

Changes to the 2018 Census

The aim of the New Zealand census, which has been occurring since 1851, is to provide a high quality count of the New Zealand population and dwellings every five years. The census is unique in that it is able to provide information on small populations and groupings that is not achievable in sample surveys.

It is important that the census reflects the data needs of New Zealand at the time. Some of these needs do not change, while others are emerging. This mix is reflected in the change model that the 2018 Census operated under and informs the strategic aims of the 2018 Census.

These strategic aims are to:

- improve data quality while modernising
- reduce the cost of collection operations
- contribute to organisation capability
- increase the use of administrative data
- make digital engagement easy
- adopt a test-drive development
- deliver customer driven products and services.

See 2018 Census strategy for more information about the strategy.

The strategic aims of the census fit within the larger organisational roles of Stats NZ., which include ensuring we continue to provide independent and trusted data and that we enable New Zealand's decision makers. It is also important for Stats NZ, and therefore the census, to innovate to drive value for our customers while ensuring that we are trusted stewards for New Zealand's data now and into the future.

2018 changes

As well as making changes to the wording and design of the questionnaires, we made some substantial operational and process changes to the 2018 Census. We updated processes, methodologies, systems, and tools across the census programme to ensure the census is as cost-effective as possible while ensuring high quality data for our customers.

See:

- Respondent interaction
- Concept and questionnaire changes
- Collection methodology
- Classification changes
- Imputation
- Document layout
- Prioritisation of variables.

Respondent interaction

A key element of the change model is the difference in how respondents interacted with the 2018 Census. Most respondents received a letter with an access code which they can then use to complete the census online. We expect the move to online will result in data quality improvements.

We sent access codes to households listed on an address register, which had been constructed then physically checked by Stats NZ. We mailed out the codes to about 80 percent of dwellings, while the remaining 20 percent were 'list-leave', where a field officer visited the dwelling to deliver either an access code or paper forms with an access code. While online was the primary mode of completion, people who do not receive paper forms through the 'list-leave' process were able to request them.

This paper describes some of the possible data quality implications of the change model by variable, as well as some other changes that may impact on data quality.

See 2018 Census – a modernised, digital-first census for more about the change model for 2018 Census.

Concept and questionnaire changes

The changes we made to the concepts of questions and to the design and layout of the questionnaires were a result of changes to the data needs of our customers, or where a question required changes due to respondent difficulty in the 2013 Census or during the testing phase of the 2018 Census.

See 2018 Census report on final content for a summary of these changes.

Collection methodology

Changes to the collection methodology may have some impact, positive or negative, on data quality. We made changes to ensure the census meets the needs of respondents and customers in 2018, as well as reducing the cost of running the census.

In previous censuses, staff visited every dwelling in New Zealand to deliver and collect paper census forms as well as delivering internet access codes in 2006 and 2013. In 2018, we moved to the digital-first approach, where most dwellings received an internet access code on the mail. We employed fewer field than in previous censuses. Our aim was to enable more respondents to complete the census without the prompting of field staff, and to target our resources to respondents who had not completed their census. From previous censuses, we know certain groups are less likely to complete their census. In order to encourage these groups to participate, we had targeted response strategies, a comprehensive communications and marketing campaign, and ran focused community engagement activities. The aim was to help mitigate some of the challenges in achieving our expected response and coverage rates as a result of the new collection model and a general trend down in response rates.

The uptake in online responses to the census will result in a significant improvement on the quality of the data we receive. On the whole data, collected using the internet collection system (ICS) is of better quality than that of paper. The online form used automatic routing and as-you type lists which helped respondents answer the necessary questions with their correct information easily. As well as these features, which help respondents, we used radio buttons and check-boxes and mandatory questions to ensure the data collected is of high quality. Where this is applicable, the possible

implications of these changes have been discussed, see the online optimisation sections in Changes and their effect on each variable.

Changes to the processing system have the potential to impact on how the data looks compared to earlier censuses. However, we carefully considered all changes to ensure that, where possible, any we made will improve the quality of the data received.

Classification changes

The classifications for a number of variables, for example religion and iwi, have been reviewed and updated to ensure our classifications are moving with society's changes. The impacts of the changes to classifications are discussed in relevant sections of Changes by variable.

Changes to the statistical geographies allow us to report more detailed information at small geographies for our customers to use while ensuring that we maintain the privacy of respondents. See Statistical standard for geographic areas 2018 (SSGA18).

Imputation

Imputation is discussed at length with relevant variables. Two key types of imputation are used: donor and administrative data imputation. We intend to use imputation for a number of variables, although not all variables, where there is sufficient evidence that imputed data will be of acceptable quality. However, a final decision about imputation will be made once we've received all the data.

Donor imputation finds similar respondents with a response to the variable in question. The processing system then finds the closest match to the respondent with missing or unidentifiable data and imputes the donor respondent's response. This method of imputation uses the CANCEIS (Canadian census edit and imputation system) developed by Statistics Canada and used in their census. The system is based on the Nearest-neighbour Imputation Methodology (NIM) and was first used to perform edit and imputation on census data.

Administrative data imputation uses information provided by the respondent available in other datasets (eg health or education datasets) and imputes a response based on this information.

Changes and their effect on each variable

Here are possible implications for each of the 2018 Census variables of the changes we've made to various aspects of the census.

The variables in this chapter are ordered by priority.

Prioritisation of variables

We applied one of three levels of priority to each variable, based on the 2018 Census data quality management strategy.

Priority 1 variables include those that are the highest priority in terms of quality, time, and resources across all phases of the census. These variables are core census output variables and contribute to population and dwelling counts as well as provide the necessary data for electoral purposes. On the online census form, priority 1 variables are mandatory questions, , while priority 2 and 3 are optional.

Priority 2 variables contribute to data that is of high interest to the public and is closely linked to the main purpose of the census.

Priority 3 variables are important to certain groups and result in information that can only be collected from are a census.

As well as assigning a priority level, we gave each variable a change rating (minor, moderate, or major) for real-world change and for the data collection changes (including questionnaire/concept changes and processing changes).

Note: the changes in this chapter are indicative, based on what we knew before analysing the 2018 Census data.

Age / Date of birth

Moderate real-world change, minor data collection change
Priority 1 variable

There are few significant changes from 2013 to the way age is being processed, but no changes to the content collected. Age is needed for population estimates, and as such is a high priority variable and where necessary it will be imputed. Imputation was also used in 2013 and a similar approach is being used in 2018.

We will take age data from the dwelling form if it is missing from the individual form, but available on the dwelling form. Failing this, we will use donor imputation. An edit is planned to ensure that if any other information is provided an imputed age cannot be possible.

Imputation

Imputation will not be required for forms completed online as date of birth is a mandatory field. However, field officers will not be checking paper forms for completion so some non-response items on paper forms will need to be corrected with imputation.

Collection methodology

A new collection model was implemented in 2018, with a digital-first approach, with specific targeted strategies to engage hard-to-reach populations. We do not yet know the impact of the changes to the collection methodology on different age groups – there may be some impact on the data as a result.

Real-world changes

It is expected that the shifts we observe in the age make-up of the New Zealand population in 2018 will be as a result of real-world changes. It is expected that the proportion of 15–39-year-olds and 65+ year-olds will rise due to an increase in migration and an ageing population, respectively. We expect to see fewer 0–14-year-olds as the birth rate continues to be low, and a lower proportion of those aged 40–64 as there are fewer migrants in this age group compared to the 15–39-year-old population.

Sex

Priority 1 variable
Moderate real-world change, minor data collection change

Like age, sex is a high priority variable and has remained largely unchanged from 2013 in order to maintain a reliable time-series for analysis. There have been no content or processing changes in the sex question from 2013. The imputation approach used in 2013 will be employed for the 2018 census, however due to the attention the sex variable has attracted in the lead up to the 2018 census, there may be an increased use of the imputation method.

Responding to intersex

As there is no intersex or indeterminate option in 2018, intersex respondents who contact Stats NZ will be advised that they can request a paper form and tick both male and female. These responses will kept to facilitate future analysis of these responses, however, they will be imputed to male or female — the same process that occurred in 2013.

Imputation

In general, where the sex variable is missing, imputation for sex will is planned to have a similar approach for age. Sex will be taken from the dwelling form if information is available. If information from the dwelling form is unavailable, sex will be derived from name, if possible. Failing these two approaches, donor imputation will be used.

It is predicted that changes to the sex make-up of the 2018 population will largely be driven by real-world events. However, we are expecting to see minimal change from the 2013 data.

Usual residence

Priority 1 variable
Moderate real-world change, minor data collection changes

Usual residence is a priority 1 variable, with minimal change in content since the 2013 Census. The main changes are to do with the statistical geographies used. There are also some changes to how

respondents will experience the online form. These changes should improve the quality of responses we receive from those using the online form, and with 70 percent of the population expected to use the online form, the overall quality should improve. A similar approach to imputation for missing items will be used as was applied in the 2013 Census.

New geographies

A new statistical geography, the Statistical standard for geographic areas 2018 (SSGA18), has been developed, and replaces the 1992 standard which was published as the New Zealand Standard Areas Classification 1992 (NZSAC92). Meshblocks remain as the lowest level of geographical area for census data outputs including for usual residence outputs.

SSGA18 has new statistical geographies – statistical area 1 (SA1) and statistical area 2 (SA2) and urban and rural areas built up from meshblocks. The new statistical area 1 has no equivalent in the NZSAC92 classification, SA2 statistical areas replace area units, and there is a new urban and rural area classification. SSGA18 will be used for publishing 2018 Census geographical outputs.

Geographical output data from the 2006 and 2013 censuses will be rebased on the new geographies so comparisons across time of geographical outputs derived from the usual residence variable shouldn't cause issues with data quality.

Online optimisation

The online form has undergone significant redevelopment in the lead up to the 2018 census with the aim of ensuring a good respondent experience and utilising available technologies to improve data quality. Respondents were asked to confirm their census night address linked to their access code. Once they confirmed the address, they could refer to this address when answering the usual residence question. If the address is different from the one where they were on census night they were able to start typing the address and then select from an as-you-type list. We are expecting these development changes will improve the quality of responses for responses collected online.

Imputation

In 2013, we only used imputation when geocoding of usual residence on the paper or internet form to the Census Dwelling Frame failed. The same method is planned to be employed in 2018, and we expect geocoding success rates to be between 95 and 98 percent. In cases where a valid address is not given, and geocoding fails, imputation will be used based on the approach taken in 2018. An address from the dwelling form will be imputed if possible, failing this donor imputation will be used.

Real-world changes

We expect very little of the change in the data in 2018 to arise from changes to census processes or methodology. We expect there will be some real-world changes, for example growing populations of usual residents in Auckland, Waitemata, Selwyn, and Queenstown-Lakes, accompanied by declines in Grey, Wairoa, and Buller districts.

Census night address

Priority 1 variable
Moderate real-world change, minor data collection change

As well as asking about usual residence address, the census asked about what address respondents were at on census night. This question has not changed since the 2013 Census. While the concept has not changed, there have been changes to how the information is collected in the online census form. As we expected, the majority of respondents to complete their census online, we expect some mode effect on the data. However, we expect the online uptake to result in higher quality data.

Online optimisation

Respondents who completed their census online were asked to confirm the address they are presented with (which is linked to their unique access code) is their census night address. If the address is incorrect they are able to enter the correct address. An as-you-type address field is enabled to help respondents to enter the correct address, if required.

Geographical output data from the 2006 and 2013 censuses will be rebased on the new geographies so comparisons across time of geographical outputs derived from the census night address variable shouldn't cause issues with data quality.

Ethnicity

Priority 1 variable
Moderate real-world change, minor data collection change

Ethnicity is a high priority variable due to its importance for many of our customers and data users. The ethnicity concept or question has not changed since the 2013 Census, although with the modernised census approach there have been changes to the collection methodology and processing which may have an impact on the data series.

Online optimisation

In particular, the design of the online form is expected to improve the quality of data we receive from respondents. As ethnicity is a priority 1 variable, respondents using the online form were required to answer the question before they could submit their census form. Additionally, there was as-you-type functionality for typing in an ethnicity. We anticipate that this will result in fewer non-identifiable response and more detailed ethnic group responses.

The paper form had no such checks, and field collectors were not checking paper forms for completeness. This may lead to a decrease in the quality of responses on the paper forms. However, as we expect a 70 percent online uptake, the decrease in quality from paper responses should be mitigated by the improvements in the online collection.

Imputation

Imputation is planned in 2018 with the aim of improving data quality. The imputation method in 2018 will use a combination of administrative data and donor imputation to assign one or more ethnicities to a missing response. Where possible, missing ethnicities will be found using an administrative dataset (IDI). As the IDI is likely to only record high level or level 1 ethnic groupings,

where the IDI produces a level 1 ethnicity response, donor imputation will only consider donors with the same high level ethnicities.

It is expected that this imputation method will produce higher quality, and more detailed ethnicity data than in previous censuses.

Collection methodology

Some significant changes to the collection methodology in 2018 may have an impact on the data. In response to the changing collection methodologies, we are carrying out targeted response strategies to encourage groups who have traditionally had high levels of non-response in the census to complete their census forms. These targeted groups include Māori and Pacific populations.

Targeted strategies for the 2018 Census were designed to reduce the differences in response rates between Māori and Pacific populations and the general population. A successful targeted strategies campaign is likely to result in an increase in the proportion of respondents who are of Māori and/or Pacific ethnicity. If this happens, it is likely that the census data will be more in line with Stats NZ population projections and the Post Enumeration Survey (PES) results.

Real-world changes

Aside from an expected increase in the proportion of the population reporting that they are Māori or Pacific ethnicity, we expect the proportion of the European population will decline slightly from 2013 and an increase in both Asian and Middle-Eastern, Latin American and Africa (MELAA) populations. These changes are expected to come about as a result of real-world changes.

Māori descent

Priority 1 variable
Minor real-world change, moderate data collection change

We expect a significant increase in data quality for the Māori descent variable. The Māori descent variable typically has a very high non-response rate. In the 2013, non-response for this variable was about 10 percent.

Māori descent question was made a mandatory question for online respondents; respondents were not be able to progress until they answered the question. There were no such checks on the paper form and the extensive test of the census form in Whanganui in April 2017, demonstrated a moderate non-response on the paper form, although at 2 percent it was much lower than the 2013 Census.

The only change to the Māori descent question was in the ordering of the response categories. There are three possible responses to the question about Māori descent: ‘yes’, ‘don’t know’, and ‘no’. In 2013, the order of options was ‘yes’, ‘no’, and ‘don’t know’. In 2018, the ‘no’ and ‘don’t know’ options were switched. The reason for this change was primarily a decision related to the routing associated with this question. The change in routing should have no impact on the data for this variable.

Online optimisation

As we expect approximately 70 percent of respondents to complete the census online, where Māori descent is a mandatory question, we expect to see some substantial changes to the data. We expect the proportion of 'not stated' to decline. It is difficult to estimate, but based on the proportion of non-responses on the paper forms in 2013, the rate of non-response could reasonably be something like three to four percent.

We expect that as a result of not allowing respondents to skip the Māori descent question on the primary mode (the online form), we will see an increase in the proportion of the population answering 'no' to the Māori descent question. In the 2013 Census, 95 percent of respondents who did not answer the Māori descent question also gave no iwi information. It is supposed that respondents who are not of Māori descent feel that the question is not applicable to them and skip this and the iwi question. As a result of the increase in 'no' responses to the question we expect to see a decrease in the proportion of respondents indicating that they are of Māori descent.

Imputation is planned to resolve these residual categories (including non-responses). For electoral purposes, imputation is planned for residual categories (non-response and response unidentifiable) and for 'don't know' responses. Imputation will use, in order, the response to the 2013 Census, Department of Internal Affairs birth records, if the respondent answered the iwi question, and lastly if the respondent indicated Māori ethnicity in the 2018 Census.

Iwi affiliation

Priority 2 variable
Moderate real-world change, moderate data collection change

The iwi variable collected in census has undergone some significant changes since the 2013 Census. The iwi question asked in 2018 is similar to what was asked in 2013. In 2013, the English and bilingual forms asked respondents to list their 'iwi' and 'rohe'. In 2018, the English form asks for 'iwi' and 'region' while the bilingual form asks for 'iwi' and 'rohe'. We expect that the English form may result in respondents entering regions more in line with Stats NZ geographies (eg Northland), while on the bilingual form we may get 'rohe' responses akin to Māori electorate boundaries (eg Te Tai Tokerau).

In 2006 and 2013 respondents completing the iwi question could refer to a list of iwi on the back of the individual form. In the 2018 Census the list of iwi for those filling the paper form will be on a separate sheet of paper (at the back of the guide notes), adding burden to some respondents. However, experience was optimised for online respondents, including an as-you-type list for the iwi question. When a respondent starts typing in an iwi, the response field generates possible suggestions from the Stats NZ iwi classification in a drop-down list, which respondents select from the list or write in a response if they choose. We expect that the use of the as-you-type function will improve the quality of responses.

Changes to the classification

The most significant change since the 2013 Census is the changes to the iwi classification Stats NZ uses. The updated iwi classification was released in September 2017 with 33 new iwi added to the classification. The changes to the iwi classification will likely impact on the number recorded in each iwi.

It is difficult to predict how the new classifications will change the distribution of respondents across iwi. We expect the number of respondents that know at least one of their iwi to increase as the population of Māori descent increases. In 2013, 83 percent of respondents to the Māori descent question reported at least one iwi, and we expect the 2018 census data to be comparable.

We expect that the largest iwi will continue to be the largest iwi, despite the likelihood of losing members to smaller iwi that were previously classified as part of them.

Collection methodology

The new collections approach, with targeted strategies encouraging Māori to engage in the census, may lead to an increase in the numbers of respondents reporting that they belong to one or more iwi. In particular areas such as the Far North, Whangarei, Rotorua, and Gisborne are targeted response areas and have higher proportions of Māori than New Zealand on average.

Birthplace

Priority 2 variable
Moderate real-world change, minor-moderate data collection change

Country of birth has been collected in previous censuses. The question has changed slightly – no longer asking for the current or present name of the country where the respondent was born. This change could have a small impact on the data for countries that had a former name or border changes. We expect the impact to be small, given the number of cases this is likely to affect, although this will something we will check during evaluations.

Imputation

Where possible, imputation is planned for the birthplaces variable when a respondent's intention is not clear or there is no response. Responses will be imputed from administrative data from the 2013 Census, Department of Internal Affairs' births data, or Ministry of Business, Innovation and Employment data on recent arrivals. The impact of imputation will largely depend on where the non-respondents were born.

There have been no other significant changes to the processing system.

Real-world changes

We expect moderately significant changes to the real-world data due to the relatively high levels of immigration in the last five years since the 2013 Census. We expect that the proportion of respondents born in New Zealand will decrease from the 2013 Census. The proportions of respondents in born all other regions is expected to increase with the possible exception of those born in Australia.

Usual residence one year ago

Priority 2 variable
New question, so no real-world change or data collection change

'Usual residence one year ago' is a new variable for the 2018 Census. Previously the census asked about the usual residence respondents lived at five years ago. We expect there to be some

similarities in the patterns we observe in the data between the usual residence one year ago and five years ago.

Non-response and respondent burden

In the 2013 Census, the non-response rate for the usual residence five years ago was 3.5 percent. The usual residence five years ago question was accompanied by a significant amount of respondent burden, likely contributing to the non-response rate. The respondent burden was part of the decision to ask usual residence one year ago instead of five years ago. We predict this change will lower the non-response rate for this question.

We expect the distribution of data to be similar to the 2013 Census data for usual residence five years ago, although we expect a higher proportion of respondents to be living in the same residence one year ago compared to five years ago. We expect that younger people, particularly those aged 15 to 29, will be less likely to live in the same residence one year ago when compared to older people.

Differences by age

People aged under 15 years will likely have a similar mobility pattern to their parents' cohort (30 to 64 years). In 2013, the proportion of young people aged 15 and under who were not born five years ago was 36 percent, clearly the proportion of this group not born one year ago will be much lower, with those aged between 1 and 5 years distributed between living in the same usual residence one year ago, somewhere else in New Zealand, and overseas.

We decided not impute responses to the usual residence one year ago question in the 2018 Census.

Legally registered relationship status

Priority 2 variable
Minor real-world changes, minor data collection changes

The 2018 Census collected information on legally registered relationship statuses. Legally registered relationship status differs from social marriage, and excludes de facto couples, unless respondents have been previously legally married.

The legal relationship status question has not changed since the 2013 census, and the way the data is processed for this variable has not had any significant changes.

We expect that there will be some movement in the data for legally registered relationship status as a result of real-world changes. Fewer people are getting married, meaning a larger proportion of respondents are likely to report they have never been married. We also expect the proportion of people reporting that they are divorced and separated will decline, at least in part as a result of the declining marriage rate.

Partnership status in current relationship

Priority 2 variable
Minor real-world change, minor data collection change

The variable 'Partnership status in current relationship' is derived from the living arrangements question and the legally registered relationship status questions. While there has been a small change to the living arrangements question – opposite sex and same sex civil union partners and de

facto partners are now collected in the same category as legal husbands and wives – partnership status will be output in primarily the same manner as in 2013.

Derivation

As a result of the change to the living arrangements question, we had to change the derivation used in 2013. We do not expect that this will have a significant impact on the data quality.

Imputation is not being used for this variable in 2018.

Real-world change

Real-world changes in partnering are expected to the most salient change in this variable since 2013 (with the absence of significant data collection changes). While the real-world changes are the most significant change to this variable, the changes we expect to see are likely to be relatively minor and in line with trends seen over previous censuses. It is expected that rates of marriage will continue to decline slightly, while the rates of those who are partnered but not married are likely to increase. Civil union partnerships have previously made up a very small proportion of all partnerships (0.2 percent in 2013), and this is likely to decline slightly with the legalising of same-sex marriage.

Household composition, family type, extended family type, and child dependency status

Priority 2 variable
Minor real-world change, major data collection change

Household composition, family type, extended family type, and child dependency status are a set of variables to describe the relationships between individuals within the household. These variables are derived from two questions in the dwelling form and one question on the individual form, see figure 1.

While household status, family type, extended family type, and child dependency status are the main variables derived from these questions, other outputs are derived, including number of children in a household and adult children.

The derived variables have not changed since the 2013 Census, nor have the questions used to derive these variables. The way the variables are derived has changed.

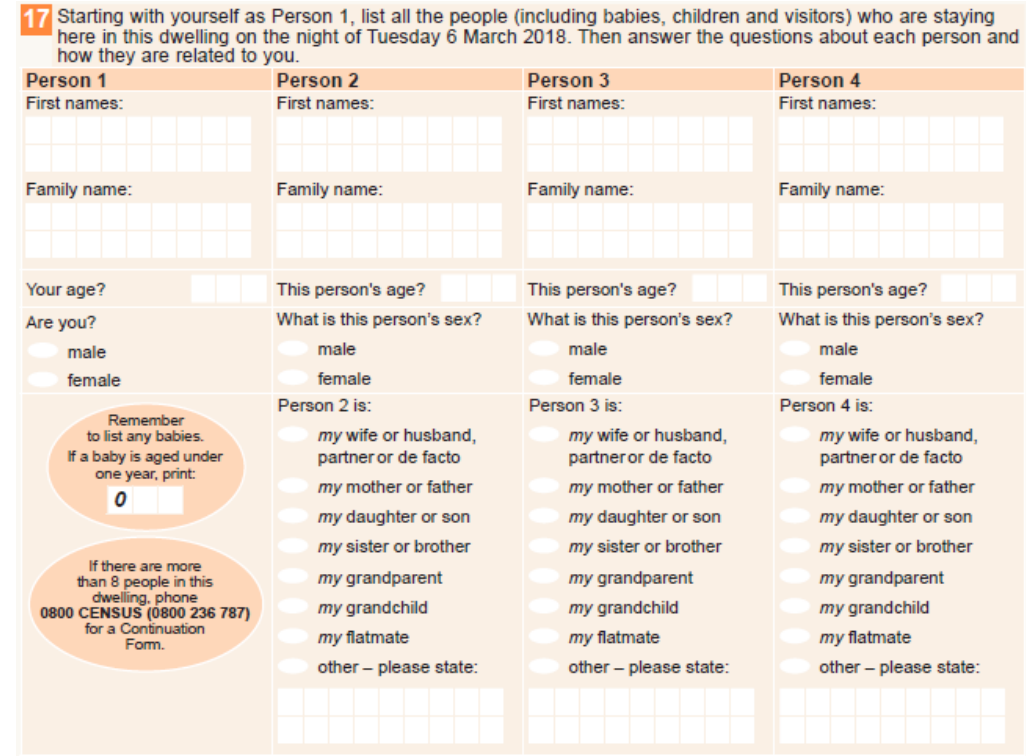
Processing system changes

We made some significant changes to the processing system in the 2018 Census. Previously, the processing system assigned each individual within a household four values according to their position and role in the household. The combination of these values, along with age, sex, and work and labour force status in some cases, were used to derive a household composition. If applicable, family type and extended family type and child dependency status were derived using the same values.

In the 2018 Census processing system, households are arranged in to matrices in order to determine their household composition, family and extended family type, and child dependency status. The new approach uses a similar approach to other Stats NZ household surveys.

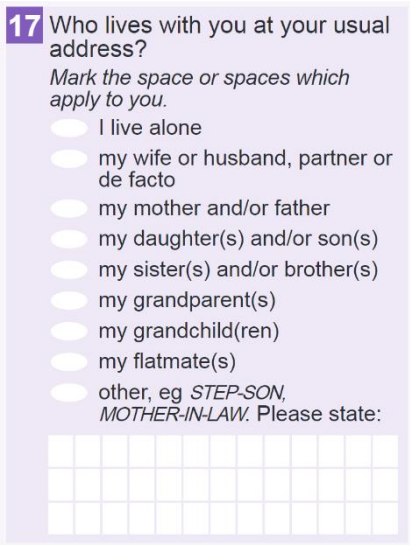
Figure 1

Household composition, family type, extended family type and child dependency status questions, dwelling form and individual form, 2018 Census



17 Starting with yourself as Person 1, list all the people (including babies, children and visitors) who are staying here in this dwelling on the night of Tuesday 6 March 2018. Then answer the questions about each person and how they are related to you.

Person 1	Person 2	Person 3	Person 4
First names:	First names:	First names:	First names:
Family name:	Family name:	Family name:	Family name:
Your age?	This person's age?	This person's age?	This person's age?
Are you? male / female	What is this person's sex? male / female	What is this person's sex? male / female	What is this person's sex? male / female
Remember to list any babies. If a baby is aged under one year, print: 0	Person 2 is: *my* wife or husband, partner or de facto / *my* mother or father / *my* daughter or son / *my* sister or brother / *my* grandparent / *my* grandchild / *my* flatmate / other – please state:	Person 3 is: *my* wife or husband, partner or de facto / *my* mother or father / *my* daughter or son / *my* sister or brother / *my* grandparent / *my* grandchild / *my* flatmate / other – please state:	Person 4 is: *my* wife or husband, partner or de facto / *my* mother or father / *my* daughter or son / *my* sister or brother / *my* grandparent / *my* grandchild / *my* flatmate / other – please state:
If there are more than 8 people in this dwelling, phone **0800 CENSUS (0800 236 787)** for a Continuation Form.			



17 Who lives with you at your usual address?

Mark the space or spaces which apply to you.

- I live alone
- my wife or husband, partner or de facto
- my mother and/or father
- my daughter(s) and/or son(s)
- my sister(s) and/or brother(s)
- my grandparent(s)
- my grandchild(ren)
- my flatmate(s)
- other, eg *STEP-SON, MOTHER-IN-LAW.* Please state:

Imputation

We plan to use imputation, as it is often not possible to determine all the relationships in the household with the information provided by the respondents alone. In these cases donor imputation will be used to fill information that is missing from the household, though this will not override the information provided by the household.

The change to the processing system is significant for the family and household variables – they have the potential to impact on data quality. Before we implemented the changes, we ran the 2013 Census data through the new processing system. The data from the new system looks comparable to the data generated from the 2013 processing system.

There are unlikely to be any significant real-world changes to the derived variables. It appears that the number of 'couples without children' is increasing in size and proportion, and numbers of 'couples with children' and 'single parents with children' are declining. However, these changes are slow and are continuing in line with previous censuses.

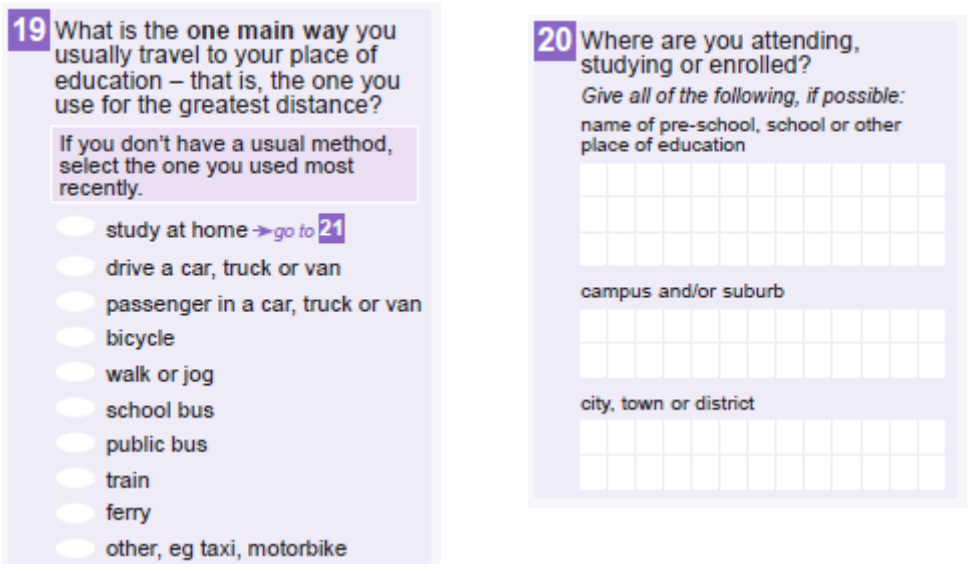
Main means of travel to education and Educational institution address

Priority 2 variable
New question, so no real-world change or data collection change

'Travel to education' and 'Educational address' are new variables for the 2018 Census. It is therefore difficult to predict how this question will be answered by respondents, and what, if any, issues there may be. The two tests run before the 2018 Census (one in 2016 and one in 2017) as well as information from the 'travel to work variable' can give some insight into how the new approaches to collection and processing may affect these variables. See figure 2 for questions from 2018 Census form.

Figure 2

Travel to education and educational address questions, 2018 Census



In the 2018 Census, we expect that most respondents will complete their census form online and that the responses from the online form will be of higher quality than those from the paper form particularly as multiple responses are not possible online. In Census test 2017 the majority of responses that could not be coded were multiple responses with the predicted high uptake in the online mode for the 2018 Census, we expect that the amount of data in 'not elsewhere included' (which is made up of response unidentifiable and not stated) will be quite minimal, and perhaps lower than seen previously for the travel to work question.

Imputation

Where there is no response to the travel to education question, we plan to use donor imputation. This is expected to further minimise the amount of response unidentifiable or non-response.

The 2018 Census also collected the address of the educational institute. Respondents were asked to provide as much detail as possible. We expect some differences in data quality between the paper and online forms for this data as the online form has an as-you-type list to help respondents complete this question accurately.

Imputation for educational institution address will also be done (using the donor method). This is expected to help minimise the amount of data in the 'not elsewhere included' category.

Highest secondary school qualification

Priority 2 variable
Minor real-world change, minor data collection change

The 2018 Census asked respondents about their highest secondary school qualification, as part of a suite of questions that aims to understand the qualifications of people living in New Zealand. The highest secondary school qualification variable has undergone only minor change since the 2013 census. As a result, we expect changes to the way the information is collected and processed will have little impact on the data.

We expect that the shift to online as the primary mode will improve the data quality. For respondents whose qualifications are not included in any of the tick box options, there will be an as-you-type list which we expect will improve the quality and usability of responses. Those aged over 75 years are less likely to respond to this question than other age groups. The propensity for older age groups to miss or skip this question, combined with their higher likelihood of completing the questionnaire on paper, may mean that there is a substantial difference in non-response rates for paper and online modes.

Imputation

Imputation is planned to rectify non-response for this variable. Administrative data held by Ministry of Education is of high quality for those who have gained New Zealand qualifications since 2001. Administrative and historic data from the 2013 census will be used, taking the highest qualification from the Ministry of Education and 2013 Census datasets. As the Ministry of Education dataset was last updated in December 2016, this will not be an issue for the majority of people who have not studied in the period between the start of 2017 and census day but we recognise we could miss a qualification gained in 2017. We assessed the risk of this to the data to be small.

Real-world change

We expect the majority of change will be as a result of real-world change, with an expected increase in the proportion of high school qualifications gained overseas, and an overall increase in the proportion of the population with some high school qualification, particularly amongst young people where more young people are getting high school qualifications.

Post-school qualification

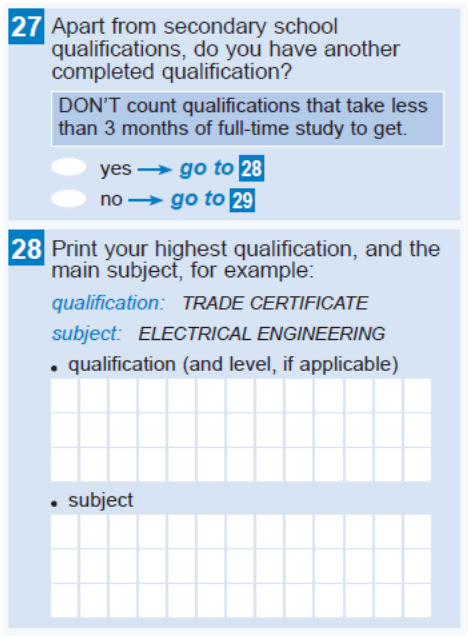
Priority 2 variable
Minor real-world change, moderate/major data collection change

Related to the highest secondary school qualification variable, and part of the suite of questions around educational attainment referred to Highest secondary school qualification section, is the post-school qualifications variable. The post-school qualification has undergone significant revision in order to minimise the burden to respondents, as well as improving the quality of the data. Figure 3 compares the 2013 and 2018 census questions.

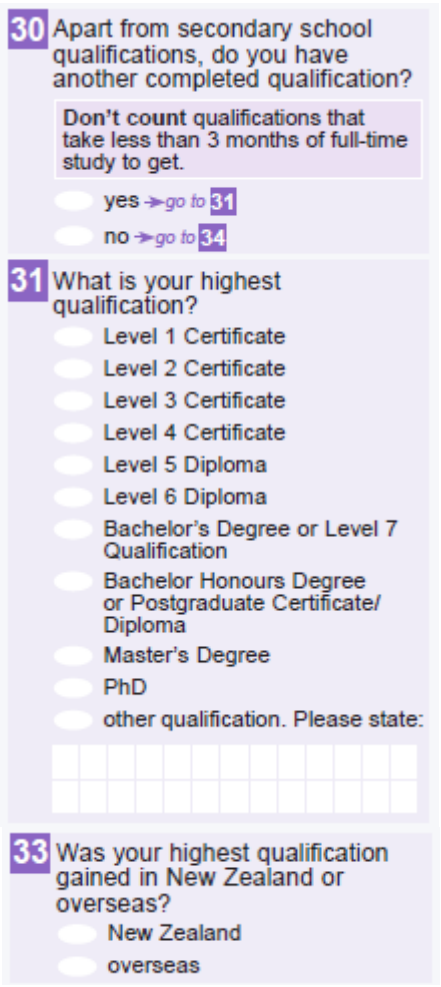
Figure 3

Post-school qualifications questions, 2013 and 2018 censuses

2013 Census



2018 Census



The most significant and notable change is the move from asking respondents to report their qualification in a free-text box, to allowing the majority of respondents to simply tick an option, with a free text box if necessary with a question to indicate whether they gained the qualification in New Zealand or overseas.

Changes as a result of 2013 Census quality issues

The quality of data collected in the 2013 census was only of a moderate quality, with a relatively large number of responses coded to unidentifiable. An example noted from 2013 was if respondents entered 'Diploma' in to the qualification box, it was difficult to determine whether this is a level 5 or 6 diploma. Thirteen percent of responses collected in 2013 were residuals, which included unidentifiable responses and non-responses. We expect that the changes to the 2018 Census question will reduce the level of residuals.

We expect that the majority of respondents, approximately 70 percent, will complete their census online. We expect that the online mode will improve the quality of responses, particularly as respondents cannot select more than one qualification.

As with the question about high school qualifications, older age groups are less likely to respond to this question. There is some evidence that older age groups, possibly because they are not familiar with the New Zealand Qualifications Authority (NZQA) qualifications, give responses that are unable to be coded at a higher rate. We tried to navigate this issue with the as-you-type list online, with the extensive list of synonyms that may be more familiar to respondents. However, as older people are more likely to complete their census on paper, responses for older populations, particularly those aged 65 and over, may be of poor quality.

Real-world change

We expect proportions of respondents with post-school qualifications and the types of qualifications to change as a result of real-world changes. Data from previous censuses indicates that we may continue to see an increase in the proportion of the population with some post-school qualification. We expect the most common qualification will continue to be a bachelor's degree and level 7 qualification, and we expect the prevalence to increase from the 14 percent of the population in 2013.

We expect to see an increase in post-school qualifications across all age groups from the 2013 census. The exception to this may be the 15–19-year-old age group, as respondents are encouraged to stay in school as long as practicable.

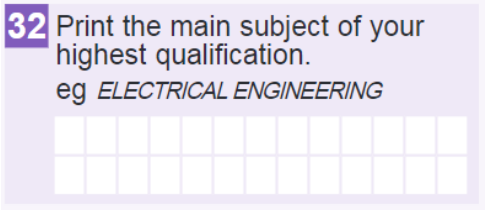
Field of study

Priority 2 variable
Minor real-world change, Moderate data collection change

The field of study variable relates to the subject of the respondents highest qualification. The concept asked has not changed since the 2013 Census, however the way the question is asked is different. In 2013 field of study was asked with the post-school qualification question, with respondents asked to report their qualification and the subject the qualification was gained in, see figure 4.

Figure 4

Field of study question, 2013 and 2018 censuses

2013 Census	2018 Census
	

Online optimisation

We anticipate that the quality of data we receive from respondents will be better in 2018 particularly as post-school qualification is now asked as a separate question with tick box options. As well as improved data as a result of questionnaire improvements, we expect the as-you-type function on the online form will improve the quality of responses – particularly as options selected from the as-you-type list are auto-coded to the classification.

In the 2013 Census, 28 percent of all respondents with a post-school qualification were unable to be classified to a field of study. We expect that the use of the online form will reduce the numbers that are unable to be coded as a result of automatic routing and the as-you-type lists. We anticipate that for respondents completing their census form on paper there will be similar quality issues to the 2013 Census.

Real-world change

We expect that there will be little change in the fields of study data as a result of real-world change. In the 2013 Census the four highest fields of study were management and commerce, engineering and related technologies, society and culture, and health. There is little compelling evidence to suggest that the proportions in each of these categories is likely to change significantly.

Highest qualification (derived variable)

Priority 2 variable
Minor real-world change, moderate/major data collection change

Highest qualification is a derived variable, derived from highest secondary school qualification and post-school qualifications. The question is not asked in the census and there have been no changes to the concept measured. Some changes to the post-school qualification variable may impact on the data quality of highest qualification, see Post-school qualification.

We expect to see some increase in the quality of this variable as a result of the predicted uptake in the online mode. One of the main advantages of the online mode is it is able to automatically route

respondents to questions that they need to answer. In relation to this variable, respondents who answer that they do not have another completed qualification (beyond a high school qualification) will not be able to answer the question about post-secondary qualifications. The use of this functionality will not allow conflicting responses for these questions. The same checks will not be in place for the paper questionnaire.

We expect to see similar trends in highest qualification as we expect to see in post-school qualifications – the trend towards higher qualifications continuing, with fewer respondents having no qualifications, and the proportion of respondents completing level 4–6 qualifications declining in favour of higher qualifications.

Sources of personal income

Priority 2 variable
Minor real-world change, moderate data collection change

The sources of personal income question is used to better understand the way New Zealander's receive income as well as prompting respondents to think about all their sources of income before they answer the total income question that follows.

The sources of personal income concept has not changed, with only a few minor changes to the questionnaire to reflect name changes of New Zealand government benefits. We do not expect the change to the questionnaire to effect the quality of the data, particularly as the names of the benefits changed in July 2013. However, there may be some confusion if respondents are not aware of the new benefit names.

Imputation

Imputation from administrative data is planned for use for the sources of income variable. We plan to use Inland Revenue (IRD) data to impute for non-response. The IRD data is of high quality for most sources of income, however there are some issues with the administrative dataset that may cause issues with data quality for those responses. The IRD dataset has high quality data for income earnt from salaries, wages, and government benefits (including superannuation), however the data is of lower quality for income earnt from investments. If imputation from administrative datasets is not possible, donor imputation will be used.

The lack of imputed information available from investments is most likely to impact on those aged over 65 as their only recorded income will likely be superannuation where they may also have substantial investment income. The proportion of non-response in the over 65 population will need to be monitored during the evaluation process to ensure that the imputation will not affect the types of income sources for this group.

Real-world change

There are also likely to be changes to the data as a result of real-world changes. Due to an ageing population, it is likely there will be an increase in the number and proportion of people collecting superannuation payments. Also, it is likely that we will see more people receiving income from wages and salary than in 2013, since New Zealand was still recovering from the Global Financial Crisis in 2013 and there has been a rise in immigration of mostly working age people.

Total personal income

Priority 2 variable
Moderate real-world change, moderate data collection change

Total personal income was collected in the 2018 Census in the same way it was collected in 2013. While there were no changes to the concept or questionnaire design there have been some other changes for this variable.

Imputation

The most significant change for this variable is the planned use of imputation in the 2018 Census. Imputation is planned using the same method as sources of personal income. As mentioned Sources of person income – imputation, the IRD admin dataset is of generally high quality, although there are some issues with the dataset. There is a higher proportion of superannuation and working age benefits in the IRD dataset than recorded in the census data. As a result there is a higher proportion of earners earning between $15,001 and $20,000 in the IRD dataset.

The IRD database does not have good coverage of income received from investments. This may mean that for some segments of the population, particularly those aged 65 and over, imputed income may be lower than actual income. Whether the data is of acceptable quality is something we need to assess during evaluations.

Real-world change

There are likely to be some changes to the data as a result of real-world changes. We anticipate that there will be an increase in incomes as a result of expected wage inflation.

Work and labour force status

Priority 2 variable
Minor/moderate real-world change, minor/moderate data collection change

Work and labour force status is an indicator derived from five questions. These questions ask the type of employment a respondent was in (if any), the number of hours worked in a main job and any additional jobs, if the respondent was not employed have they looked for paid work, what methods they used to seek paid employment, and their availability to work if a job had been available (see figure 5).

The answers to these questions determine a person's work and labour force status. Stats NZ produces outputs based on five work and labour force statuses: employed full-time, employed part-time, unemployed, not in the labour force, and work and labour force status unidentifiable.

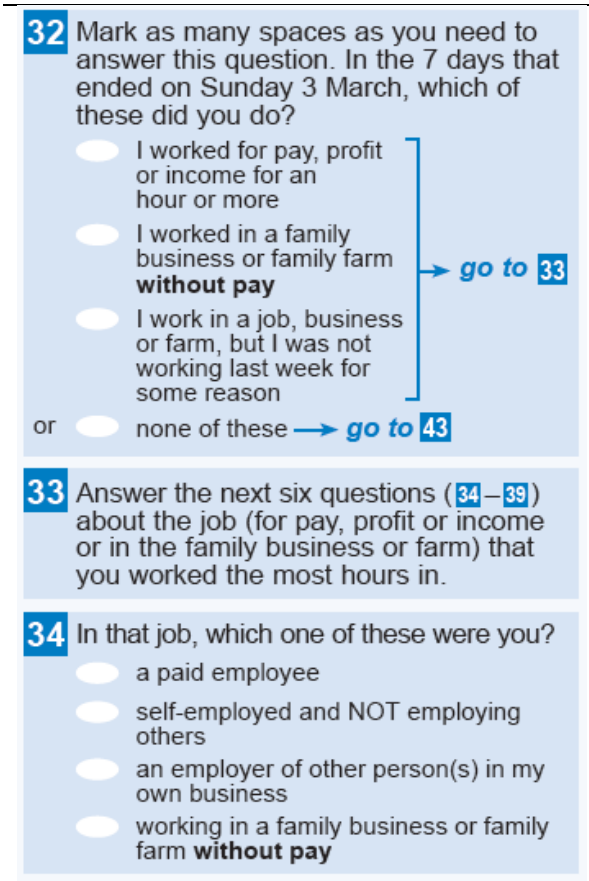
Questionnaire and concept changes

The work and labour force questions have changed slightly. The hours worked question is now asked before the occupation question. The hours worked question now asks about the hours worked in the main job rather than the hours worked in questions 34–39 as in 2013, see figure 5.

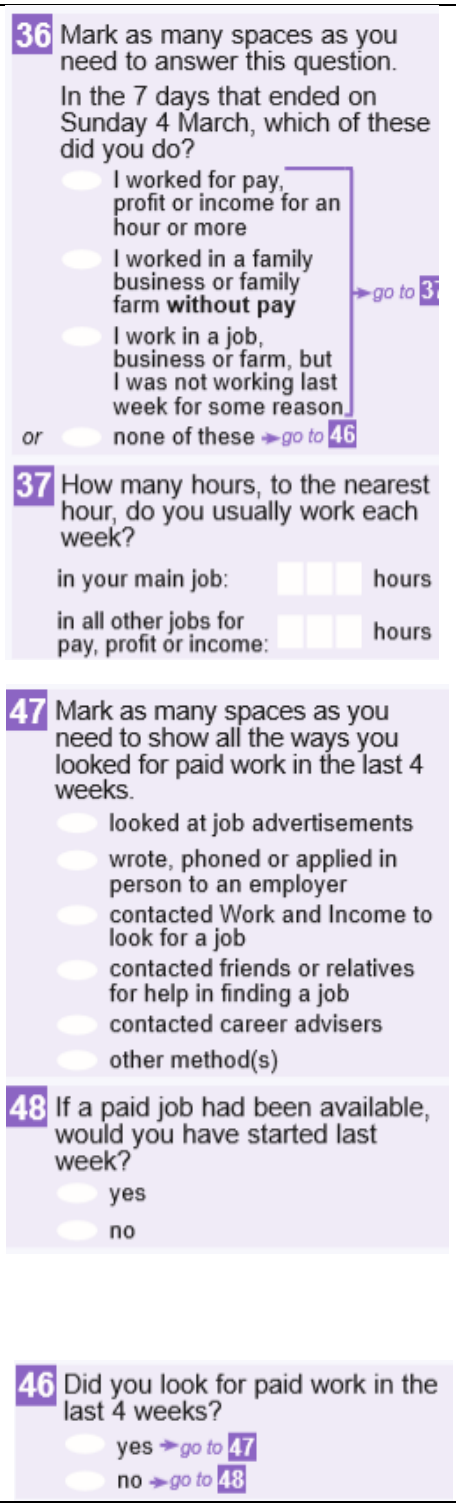
Figure 5

Hours worked questions, 2013 and 2018 censuses

2013 Census



2018 Census



42 If you have answered questions about your job → *go to* **46**

otherwise → *go to* **43**

43 Did you look for paid work in the last 4 weeks?

yes no → *go to* **45**

44 Mark as many spaces as you need to show all the ways you looked for paid work in the last 4 weeks.

- looked at job advertisements
- wrote, phoned or applied in person to an employer
- contacted Work and Income to look for a job
- contacted friends or relatives for help in finding a job
- contacted career advisers or vocational guidance officers
- other method(s), for example:
  - contacted other employment agency
  - placed an advertisement about a job
  - took steps to set up own business

45 If a paid job had been available, would you have started last week?

yes no

We do not expect these minor changes to the question will have a substantive impact on the way respondents answer questions. The uptake in online response, however, should have some positive impacts on data quality. The online questionnaire has automatic routing, meaning these respondents are less likely to skip questions which apply to them.

Imputation

While we expect that item non-response to these questions may drop due to the uptake of the online form, imputation is planned to resolve any non-response or unidentifiable responses. Donor imputation is being used to impute those responses whose work and labour force status cannot be ascertain from responses to the census questionnaire.

Real-world change

We expect the largest changes to the work and labour force status data will come about as a result of real-world changes. We expect that there will be absolute and proportional increases in the numbers of respondents being classified as employed full time and part time due to a declining unemployment rate, as evidenced by data collected in Stats NZ's Household Labour Force Survey (HLFS). Conversely we expect that there will be declining numbers and rates of people who are unemployed, while the proportion who are not in the labour force is unlikely to experience a significant decline.

Hours worked in employment per week

Priority 2 variable
Minor real-world change, minor data collection change

We made some small changes to the question about hours worked in employment per week. One change was to move the question to before the job related questions rather than after them.

To better understand what the New Zealand work force looks like, we asked respondents a series of questions about their employment. One variable contributing to this suite of questions is the hours worked in employment. This question asks about a respondent's main job, as well as any other jobs they may hold. The 2018 Census now asks about the hours worked in your main job rather than the job you have answered questions 34–39 about. We don't expect this change to have any signification impact on the data.

The non-response and response unidentifiable rate was 4.1 percent in 2013, and expected to be similar or lower in 2018. There may be some reduction in the non-response rate due to the uptake in the online form, which automatically routes respondents to applicable questions. Donor imputation will be used in the 2018 Census, but due to the expected low rate of non-response we do not expect imputation to impact on the data.

Status in employment

Priority 2 variable
Minor/moderate real-world change, minor data collection change

The status in employment is one of the variables examining employment. This variable establishes the type of employment a respondent is in.

There has been no change to this question or concept since the 2013 Census, and no significant changes to the way the data is processed that will impact the quality of the data.

In the 2013 Census, non-response for this variable was low, with only 2.2 percent of respondents not answering the question. We are expecting that with the uptake of the online mode this number will fall slightly, and donor imputation will be used to rectify non-responses.

Data from other surveys

Stats NZ holds detailed information on this variable collected in the Household Labour Force Survey (HLFS). The results from this survey may give us a good indication of what the 2018 data will look like. HLFS results indicate that there has been an 11.9 percent increase in paid employees since 2013. The number of respondents who are employers, indicated by HLFS data, has grown by 43.7 percent since 2013. The number of those who are self-employed has grown by 37.3 percent since 2013.

The HLFS data indicates that we are likely to see increases in the numbers reporting for each of the four categories. There is evidence that this increase is likely to be above and beyond the trends between the 2006 and 2013 Censuses, particularly for the paid employee, self-employed, and employer categories

Main means of travel to work

Priority 2 variable
Minor real-world change, minor/moderate data collection

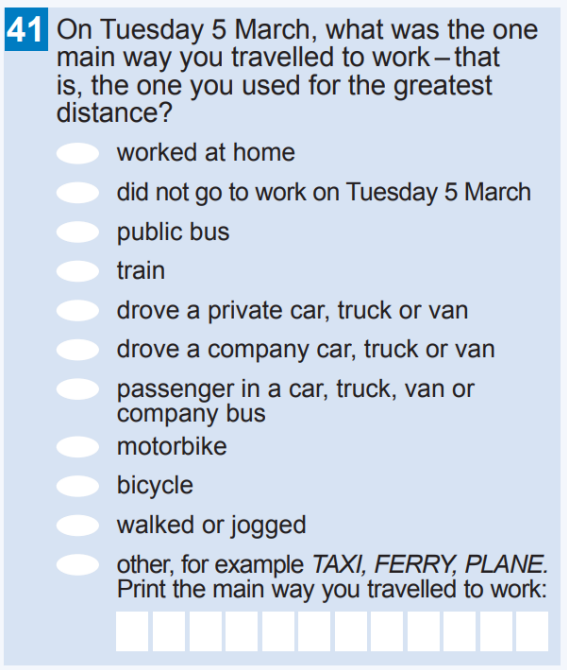
In 2018, as in previous censuses, we asked a question about how respondents travel to work. The information collected will be different from previous censuses.

Previously, we asked respondents how they got to work on census day. In 2018 we asked how they usually get to work. The response option 'did not go to work [on census day]' has been removed as it is no longer needed. As previously, the information collected still relates to the main means of travel only (the one used for the greatest distance) so this is still a single-response question and secondary means necessary to complete the full journey are not captured.

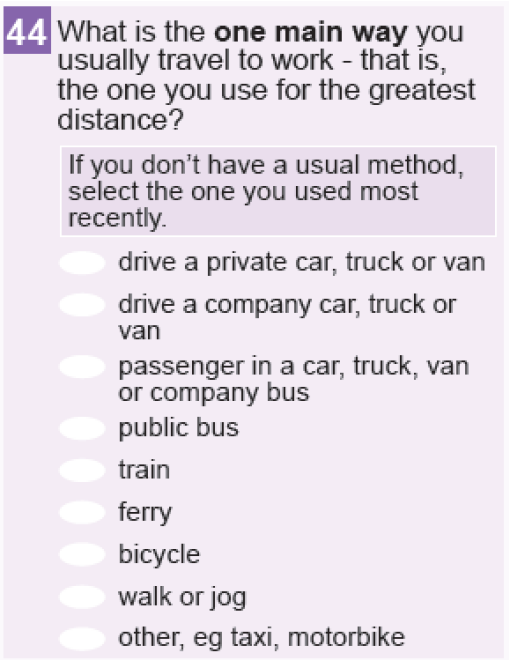
Figure 6

Means of travel to work question, 2013 and 2018 censuses

2013 Census



2018 Census



Other changes are the inclusion of the ferry option and removal of the separate motorbike response option. The free text response space has also been removed. In 2013 these responses were not part of any output.

Imputation

We plan to use donor imputation when no valid response is given, be it multiple response or non-response. This is expected to greatly reduce or eliminate the need to code data to the not stated and response unidentifiable categories as was done previously.

Real-world change and change as a result of concept changes

We do not expect the change in the type of information being collected to have a major impact on the data, but some change is possible. We anticipate that responses that would have previously been in the 'did not go to work on census day' category will be distributed amongst the various categories in similar proportions to those who did go to work on census day. The inclusion of a separate category for 'ferry' will provide a fuller picture (count) of public transport use (which is of particular in certain areas) whereas previously responses of 'ferry' were included in 'other' and so not identifiable in the data.

Occupied dwelling type

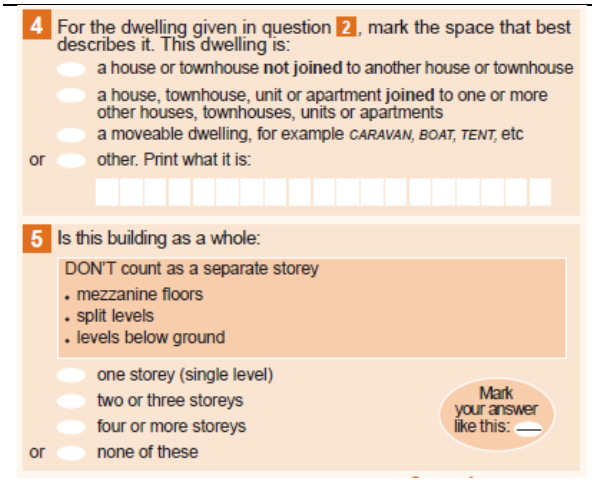
Priority 2 variable
Moderate real-world change, moderate data collection change

Occupied dwelling type is derived from three questions in the dwelling form, see figure 7. They ask about the type of dwelling (house, unit, apartment etc), whether the dwelling is joined to another dwelling, shop, or business, and how many storeys it has or, for joined dwellings, how many storeys there are in the building as a whole.

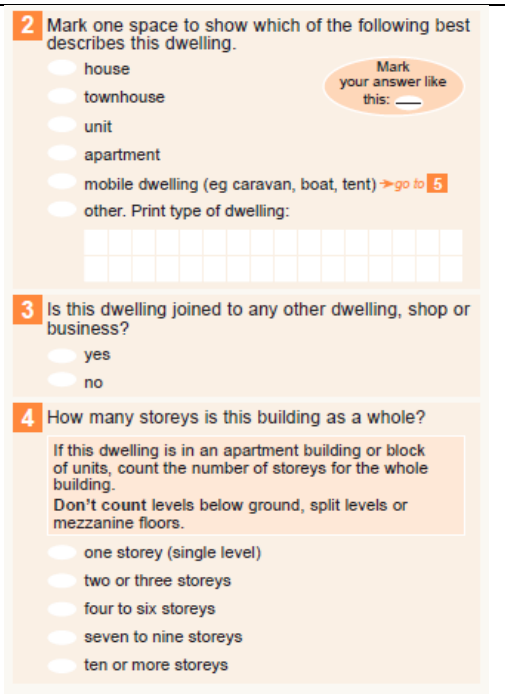
Figure 7

Occupied dwelling type question, 2013 and 2018 censuses

2013 Census



2018 Census



Questionnaire and concept changes

The information collected in the 2018 Census is changed from the 2013 Census. The 2018 Census collects more detailed information on storeys and whether the dwelling is joined to another dwelling, shop, or business, whereas in 2013 the question only asked if the dwelling was joined to another dwelling.

The questionnaire design has also changed for 2018, with three questions instead of two. This change was made to help make the questions easier for respondents to answer and improve data quality.

We hope to see some improvements in the quality of this data. These may have some effect on the comparability of the data over time. We expect to see an increase in the number and proportions of joined dwellings partly as a result of the change to the classification, but also due to real-world change (more apartments and terrace type housing developments being built over the intervening five years).

Collection methodology

The greater uptake of online forms in 2018 is expected to increase the quality of responses. However, there is some risk of non-coverage of dwellings with the new enumeration strategy, particularly those associated with other dwellings, for example “granny flats”. We hope the targeted strategies in place will mitigate this, but we don’t yet know the size of the possible undercount.

Imputation is planned for this variable where dwellings have been classified as private dwellings not further defined. In these cases, information from 2013 Census data, 2018 Census canvassing, tenancy bonds data, and building consents data will be used to impute occupied dwelling type where possible. Imputation for this variable will reduce the number of dwellings reported as ‘private dwelling not further defined’, with the aim of improving the quality and usability of the data available for our customers.

While changes to the concept, questionnaire design, and collection mode, plus the planned use of imputation, are expected to have an effect on the data, we also expect real-world changes will affect the data.

We expect the number of private occupied dwellings in New Zealand will increase, as the address canvassing has indicated.

Real-world change

We expect houses will continue to make up the majority of occupied private dwellings, but it is likely that their dominance will diminish. Due to increased housing demand over the past several years and greater diversity in the types of housing being built, we anticipate that we will see an increase in joined dwellings.

The type of dwelling, and the changes seen since 2013, are likely to be geographically concentrated. Areas of high growth (Auckland, Tauranga, Wellington, Queenstown, and Christchurch to an extent) are likely to see a greater increase in joined dwellings than rural areas or smaller urban centres.

Other private dwellings – mobile dwellings, improvised shelters – have historically made up a very small proportion of private dwellings in New Zealand. In the 2013 Census, these dwellings were 0.7 percent of the total. However, this figure was more than double that of 2001 and it is likely that the housing unaffordability crisis will lead to a further increase in this figure.

In the census we also identify non-private dwellings. Non-private dwellings include, but are not limited to, motels and hotels, prisons, hospitals, residential care facilities, night shelters, and boarding houses.

We expect that the number of non-private dwellings will be similar to the 2013 Census, although there could be a small increase. Hotels, motels, and other guest accommodation is likely to remain the largest sector of non-private dwellings. In the 2013 Census, these types of dwellings made up over half of non-private dwellings, and this is expected to continue in 2018.

Challenges identifying non-private dwellings

Boarding houses, in particular, can be difficult to distinguish from private dwellings and so can get misclassified as private in the census. There is an effort in the 2018 Census to improve identification of boarding houses in the data. If these efforts are successful, it is possible that the number of boarding houses recorded in the 2018 Census could increase from 174 in the 2013 Census.

Field operations had a specific strategy in the 2018 Census to identify and enumerate non-private dwellings.

Tenure of household

Priority 2 variable
Minor real-world change, minor data collection change

The census dwelling form asks a series of questions designed to ascertain the tenure of households in privately occupied dwellings. The variable 'tenure of household' refers to whether the household owns the dwelling they live in, has it in a family trust, rents it, or whether it is provided rent-free.

Questionnaire change

The concept of the household tenure variable has not changed since the 2013 Census, however the design of the questionnaire has had some changes, see figure 8.

The suite of questions used in 2018 are simpler, with fewer questions and fewer routing instructions. These changes were made in order to reduce respondent burden and improve data quality. By simplifying the suite of questions, and promoting online response (where individual questions are less easily skipped), it is hoped that non-response will decrease and we will observe an accompanying increase in data quality.

Imputation

Another important change for tenure of household is the use of imputation in the 2018 Census. Imputation is being planned to resolve cases where there is no response or the response is unidentifiable. The data sources used for imputation will be 2013 Census data and tenancy bonds. It is likely that this imputation will improve data quality.

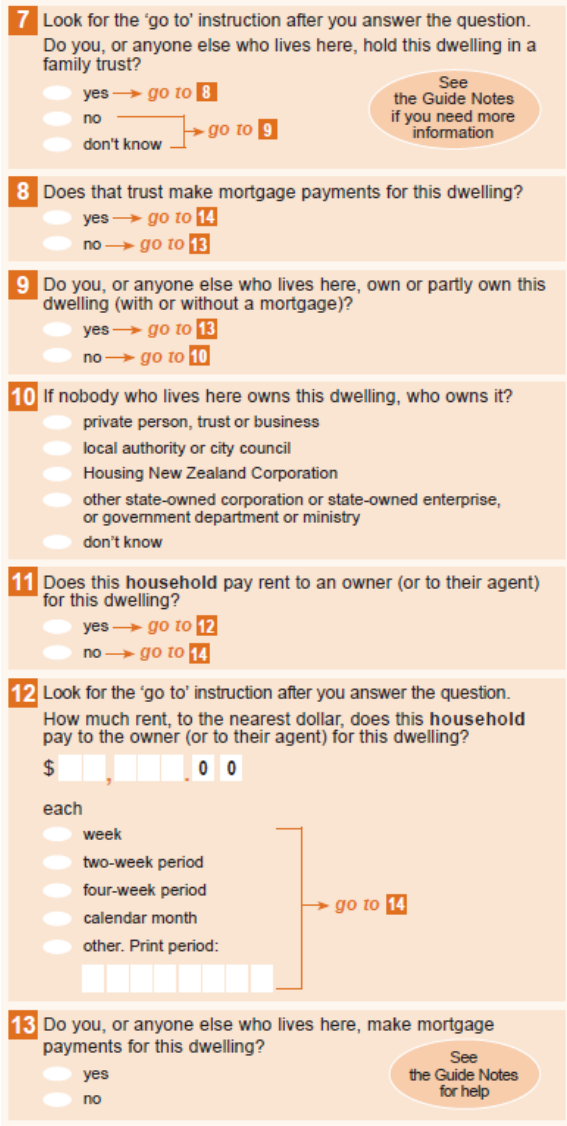
Households that rent are less likely to complete their census forms. They tend to be younger, have lower income levels, and be more transient than owner-occupier households. The use of tenancy bonds data for imputation will mean that more of these households will now be able to be identified in the census data. We anticipate that, in addition to real-world changes, this change could contribute to an increase in the proportion of households coded to a household tenure of renting.

The changes to questionnaire design, greater use of online forms, and introduction of imputation may have some effect on the comparability of the data over time, and make the data a little less comparable than previously but this is not expected to have a major impact on data quality.

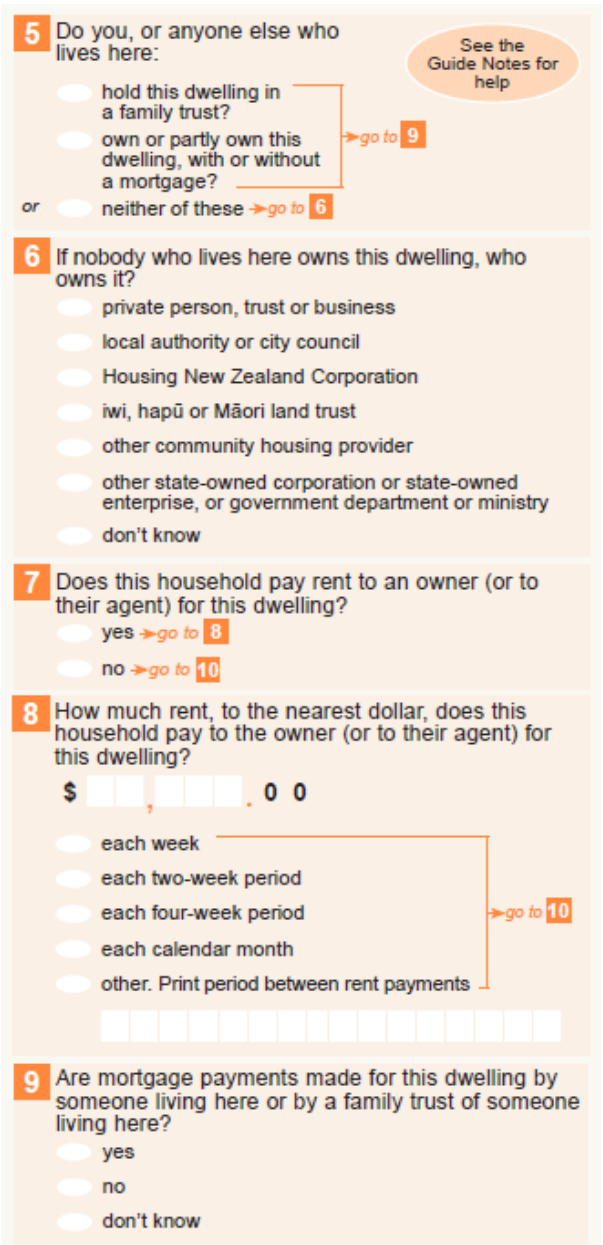
Figure 8

Tenure of household questions, 2013 and 2018 censuses

2013 Census



2018 Census



Real-world change

Real-world changes that have taken place since the last census in 2013 are also likely to have an impact on how the data looks. Increases in housing prices, in both real terms and relative to income, are likely to mean we will see a decrease in the proportion of households who own their dwelling or hold it in a family trust and an increase in those renting.

Years at usual residence

Priority 3 variable
Minor real-world change, minor data collection change

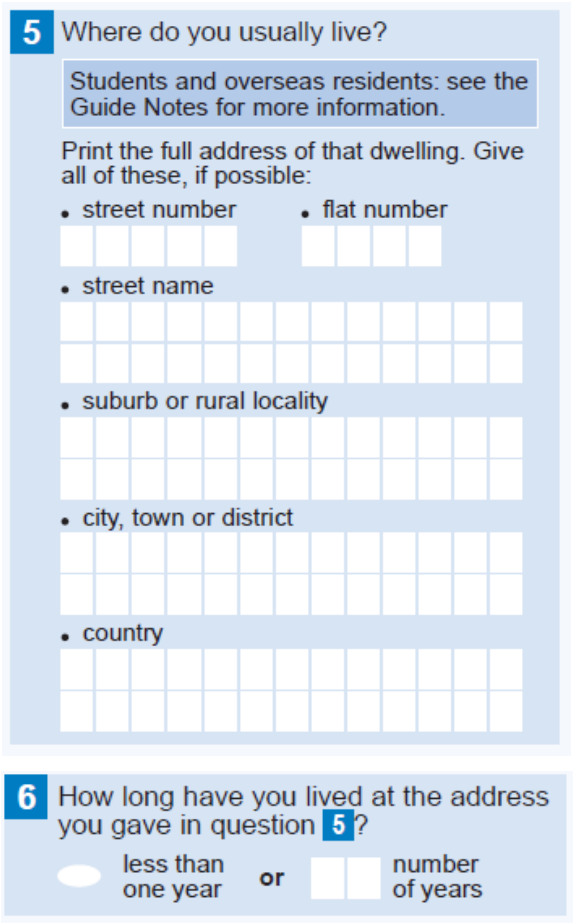
Related to the usual residence variable, the census asks respondents how long they have lived at their usual residence. There has been a small but potentially significant change to the usual residence and years at usual residence questions that could have an impact on the data as a result of how respondents interpret what is being asked of them. Aside from the changes to the wording of the question, there have been no changes to how the data is processed, and thus we don't expect any additional error to be introduced to the years at usual residence, above and beyond the 2013 Census data.

The 'tenure of household' asked respondents to enter their usual residence. If they usually live overseas they were asked to enter just the country where they usually live. In the follow-up question we asked how long the respondent lived at the address or country given. In 2013 respondents were asked to enter their usual address, whether this be in New Zealand or overseas, see figure 9.

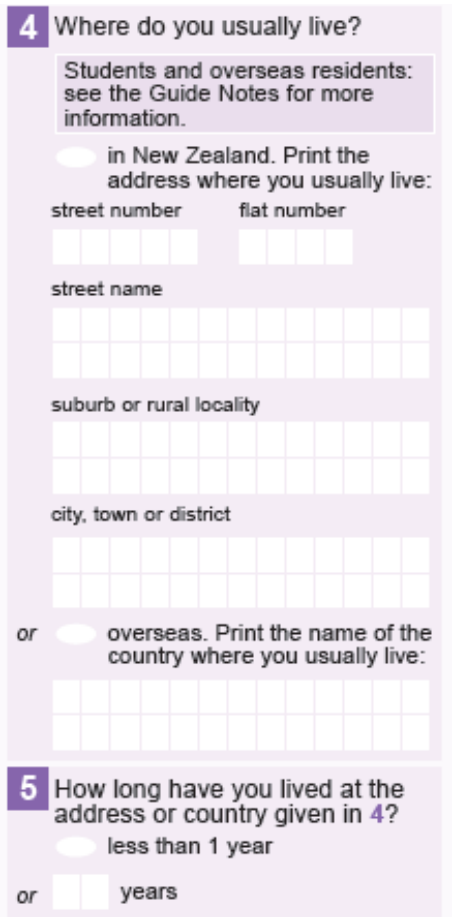
Figure 9

Years at usual residence questions, 2013 and 2018 censuses

2013 Census



2018 Census



In the Whanganui test in April 2017, the majority of responses gave similar results to the 2013 Census in Whanganui. However, a small number of respondents misinterpreted the question and answered how long they had been in the country (New Zealand) instead of how long they had been at the address where they usually live. This is a change from the 2013 Census, where respondents were asked how long they had lived at the address, with no mention of 'country'.

Due to possible misinterpretation of this question – as being 'How long have you lived in New Zealand?' – there may be an increase in the proportion of respondents who have lived at their address for a very long period of time.

We expect the number of respondents to answer this question incorrectly to be relatively small, based on the results we saw in the Whanganui test, and therefore have only a small impact on the quality of data.

Years since arrival in New Zealand

Priority 3 variable
Moderate real-world change, moderate data collection change

Related to country of birth, is the number of years since arrival in New Zealand. This variable is derived from responses to the question that asks respondents the month and year they arrived to live in New Zealand.

The question and concept are the same as in 2013 so there will be no effect of a changing question on data quality. Changes to the collection methodology and imputation are the only notable changes to this variable since the 2013 Census.

Collection methodology

Given the new collection methodology, particularly the lack of initial contact between Stats NZ field staff and many respondents, there is some risk that new migrants might not complete the census as they will not have been exposed to a New Zealand census in the past. In this case, output data could be impacted – it will appear that the average time since arrival in New Zealand is longer than is the case. Specific strategies were employed to mitigate this by engaging with migrant community leaders.

Imputation

We plan to carry out imputation for this variable in 2018. In the first instance, if the respondent's arrival details can be found in the 2013 Census records, we will use this information. If the respondent arrived after the 2013 Census, or did not enter information in the 2013 Census about their arrival date, we will use administrative records to ascertain the arrival date.

The decision to carry out imputation in 2018 is likely to have an impact on the data. While the use of imputation will provide data users with a more comprehensive picture of how long the non-New Zealand-born population has been living in New Zealand. The use of imputation is likely to lead to an increase in the number of people within each of the output classifications. In 2013, the non-response rate was 3.6 percent so we expect an increase of this magnitude across the output categories, aside from increases we expect from real-world change, that is, migration.

Languages spoken

Priority 3 variable
Moderate real-world change, minor data collection change

In the 2018 Census, respondents were asked to report which language(s) they can hold an everyday conversation in. This variable has had very little change since the 2013 Census. The question and concept has not changed, although the online form has as-you-type functionality for the 'other' language response option. This allows respondents to begin typing a language to then select from a list of suggestions that will refine as they continue to type. We expect that the as-you-type list will positively impact data quality for online respondents who speak a language that is not listed.

Imputation

Imputation is planned to rectify non-response and unidentifiable responses in the 2018 Census. Imputation was not used in the 2013 Census. Historic data from the 2013 Census will be used, where available, for non-responses. Failing this, language(s) spoken will be determined from the closest in age person living in the usual residence household. If both of these approaches are unsuccessful, donor imputation will be carried out. We know that Māori, Pacific, and CALD (culturally and linguistically diverse) populations are less likely to respond to the census than the overall population. As a result, we may see imputation having an impact on the numbers and proportions of people speaking languages that are not English.

Collection methodology

Changes to the 2018 collection model may also have an impact on the distribution of the data. We deployed targeted response strategies to encourage groups, particularly Māori and Pacific populations, to engage with the census. If successful, we could see an increase in the proportion of respondents marking that they can speak te reo and Pacific languages. However, it is unclear whether the new targeted approach will actually increase participation within these groups or just mitigate the decline in participation that could potentially happen as a result of the change to the collection model and a general decline in response rates.

Real-world change

Real-world change is also likely to have an impact on the data when compared to the 2013 Census. We expect the proportion of the population able to speak English to remain relatively stable, while the proportion of te reo Māori speakers may experience a small decline. Following trends from previous censuses, we expect that the proportion of the population speaking Asian languages will increase as the number of overseas-born-residents increases, with migration increasingly coming from Asia.

In 2013, 18.6 percent of respondents spoke more than one language. We are expecting that the proportion of the population speaking more than one language will increase by approximately one to two percent in the 2018 Census.

Religious affiliation

Priority 3 variable
Moderate real-world change, major data collection change

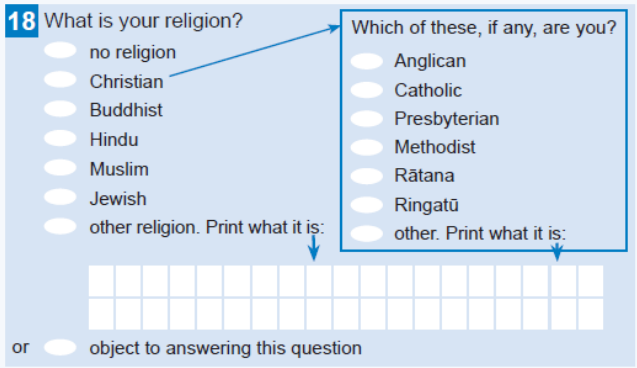
Religious affiliation has undergone some significant changes between the 2013 Census and the 2018 Census. These changes were: a new format to the religion question, changes to the classification, changes to the way imputation is used, as well as continued societal change meaning the data will look different to in the 2013 Census.

While the question and concept in the 2018 is the same as in the 2013 Census, the way respondents are asked to report their religion is different, see figure 10.

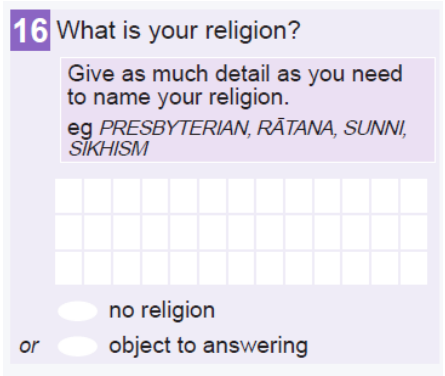
Figure 10

Religious affiliation question, 2013 and 2018 censuses

2013 Census



2018 Census



The majority of the tick box options have been removed, with only 'no religion' and 'object to answering' remaining. Removing the majority of the boxes and instead prompting those who identify with a religion to write it in, gives respondents the opportunity to give more detailed information about their religion. However some may give broader, less detailed responses. The online form has as-you-type functionality to assist respondents in writing in their religion with the appropriate level of detail.

Classification

The religious affiliation classification underwent a refresh in 2017. As a result, it includes a number of new output categories within the classification that were previously included within residual categories. This is not expected to have a large impact on the spread of reported religions or no religion overall. However, some beliefs that were previously coded as outside of scope will now be included within the standard outputs for this variable.

Imputation

Imputation is planned for this variable where there is no response. Where available, historic data from the 2013 Census will be imputed. Failing this, religion will be determined from the usual

resident who is the closest in age in the household. In cases where neither of these are possible, donor imputation will be employed.

Real-world change

We anticipate that most of the change to the data across the broad religious groupings will be as a result of real-world change. We anticipate that those stating they have no religion will continue to rise, while fewer people are likely to report that they are affiliated with a Christian denomination. We are expecting that, for the first time, those stating that they are not religious will become the largest group, overtaking Christian. While a smaller proportion identifies with Christianity each census, some other religions are experiencing increases. We expect to see a continued increase in Hinduism, Islam and Sikhism.

Study participation

Priority 3 variable
Minor real-world change, Major data collection change

Study participation has undergone some significant revisions since the 2013 Census, which will have an impact on elements of the data. The study participation question asks whether respondents are studying at a place of education, and whether this participation is full or part time. The question remains similar to the 2013 Census question, with only minor changes to the wording that are not expected to impact on how respondents answer the question. However, previously the study participation question was only asked of respondents aged 15 years and over – in 2018 the question was asked of all New Zealand usual residents. We expect the changes to the data for the 15 years and over group will be consistent with previous censuses as there are few other changes to how the variable is being treated in 2018.

Quality issues for young children

There may be some quality issues with the under-5-year age group, particularly because it is difficult to predict how respondents will answer for children in early childhood education (ECE). The Whanganui test, which provided an indication of what might be expected from the 2018 census, showed that fewer under-5s were reported as being in education (36 percent of under-5-year-olds) than would be expected based on other evidence. The proportion of school-aged children reporting that they were in education was also lower than expected, raising questions about how the suitability of the data for respondents aged under 15 years.

Imputation

Planned imputation for this variable will take advantage of administrative data in the 2018 Census. Administrative data for study participation is known to be of very high quality and was last updated in December 2016. The process of imputation will be that respondents up to 16 years of age will be imputed using donor imputation. If a respondent is over 16 and was studying in 2016, or has arrived in New Zealand within the last two years, donor imputation will be used. If someone wasn't studying in 2016 and they are over 16 we assume they are not studying as at March 2018.

Real-world change

We are expecting some minor real-world changes for the population aged 15 years and over compared to previous censuses. Overall, we expect study participation rates will be similar to those

in the 2013 Census. Trends from previous censuses show there may be a small increase in full-time study and a decline in part-time. However, as the job market continues to be strong, there may be a stabilisation in full-time study rates (tertiary enrolments tend to increase when jobs are scarce).

The new government's free education policy for first year university students may also have an impact in increasing that rates of those studying, although the magnitude of this is difficult to assess at this point.

Disability/activity limitations

Priority 3 variable
Minor/moderate real-world change, major data collection change

While the census has included questions on disability in the past, the 2018 Census disability, or activity limitations, variable will be significantly different. From the questions asked in the 2018 Census, we will be deriving a disabled indicator for the first time from the census.

In the 2013 Census, the disability question was only asked of New Zealand residents aged 15 years and over. In the 2018 Census, the subject population is New Zealand residents aged 5 years and over.

The question set being used is the Washington Group Short Set of Questions on Disability. The Washington short set questions have been collected in the New Zealand General Social Survey 2016/17, the Household Labour Force Survey June 2017 quarter and in two of the large scale census tests. While the GSS and HLFS were interviewer administered, and the two census tests were voluntary, they may give some indication of the quality and distribution of the data in the 2018 Census.

We expect the proportion of respondents to be output as disabled (those who have indicated that they have a lot of difficulty or cannot do at all in one or more of the 6 areas of activity) to be between 8–10 percent. The quality of the data from the disability variable is dependent on the quality of responses for the question set. Testing for the 2018 Census did not indicate that the questions would have higher non-response rates than for other questions at the same stage in the questionnaire forms.

There are no plans to use imputation for this variable.

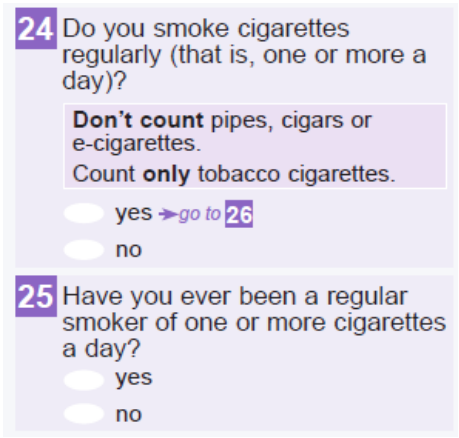
Cigarette smoking behaviour

Priority 3 variable
Moderate real word change, minor data collection changes

Cigarette smoking behaviour is being asked in the 2018 census in a similar fashion to previous censuses, see figure 11 for 2018 questions.

Figure 11

Cigarette smoking behaviour questions, 2018 Census



We made one minor change to the instruction for respondents, with e-cigarettes replacing cigarillos. We do not expect the change to the instruction to have any substantial impact on the data for 2018 – the majority of change should arise from real-world changes, with the changed collection methodology and imputation possibly having some impact too.

We made some changes to the processing system in 2018, however there have been no significant changes to the way the smoking responses will be treated. Most importantly, as in previous censuses, if a respondent indicates that they currently smoke (Q24), any response to the previously smoked question (Q25) will be ignored.

Collection methodology

We used targeted strategies in the 2018 Census to encourage groups that historically had low participation rates in the census to respond. At time of publication, we did not know whether this resulted in an improved response rate for these populations, or whether there were similar rates to the 2013 Census. If the latter occurred, the targeted strategies have mitigated the decline only as a result of the changed collection strategies and a general decline in response rates.

This is the first time we plan to use imputation for cigarette smoking behaviour. We plan to use 2013 Census data where possible, but if there isn't a response available from the 2013 census, we will use donor imputation. This will use respondents with similar demographic characteristics to impute a response. If there is a high rate of non-response or response unidentifiable, particularly if either of these are concentrated in a group with similar characteristics, the quality of the data can be compromised. It is impossible to predict the magnitude of non-response and response unidentifiable in advance, however it is an important consideration when data is being used.

Real-world change

We expect the largest effect on the smoking data to be as a result of real-world changes. The proportion of the population smoking declined between the 2006 and 2013 censuses, by about 5 percentage points. Part of this decline is reflected in a small (0.8 percentage point) increase in the population that had previously smoked regularly but did no longer do so, while the majority of the decline is in the group that never smoked regularly, increasing from 57.2 percent in 2006 to 62.0 percent in 2013. The increase in the proportion of the population that has never smoked indicates

fewer people are taking up smoking, as well as immigration largely coming from populations who are less likely to smoke.

We expect that the proportion of the population who do not currently smoke will continue to decrease, and the proportion who have never smoked will continue to increase. While we expect to see a decrease and increase, respectively, in these categories in the 2018 Census, it is likely the magnitude of change will be smaller. Part of this is likely to come about as a result of a plateauing in smoking rates that we expect to see, as well as the five year time difference between the 2018 Census and the 2013 Census, compared to the seven year intercensal period between the 2006 and 2013 censuses.

Number of children born

Priority 3 variable
Minor real-world change, minor data collection change

As in previous censuses, the 2018 Census asked for the number of children born to females. The question underwent a minor revision where the word 'alive' was removed, therefore still-born children will now be included in the responses. This change was made as a result of feedback received from interest groups who were concerned about the sensitivity of this prompt.

Concept change

There will be some changes to the number of children being reported in this question as a result of women including still-born children. However, as the number of still-born children in New Zealand is low, and testing indicates that still-born children are often included in the count by respondents regardless of the wording, we are anticipating that the change to the question will have minimal impact on the overall data.

Imputation will not be carried out, and no changes to the way the data is being processed will be made for the 2018 Census.

Real-world change

The majority of change to the 'number of children born' variable will be a result of real-world changes. We expect to see a continuation of the trends shown in the 2006 and 2013 data. This would mean females are having fewer total children, with the proportion of females having one or two children increasing, and three or more children declining. We expect that the largest group of females, around 31 to 32 percent, will have no children, with the next most populous group, between 26 and 27.5 percent, having two children.

Because the question is potentially sensitive, there is an 'object to answering' option. In 2006 and 2013 the proportion of women choosing to object to this question was between 2.5 and 3 percent of the population. We expect that with the removal of the word alive, the proportion objecting to the question will decline.

Individual home ownership

Priority 3 variable
Minor real-world change, minor data collection change

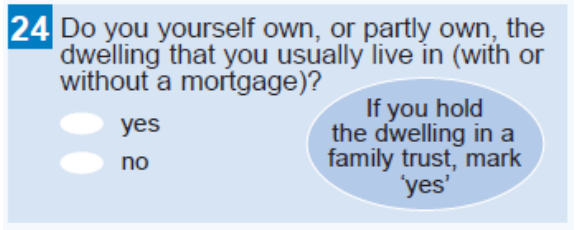
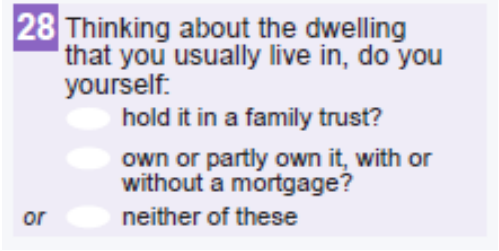
In the 2018 Census we ask a question about individual home ownership. This differs from the question in the dwelling form that asks if anyone in the household owns the dwelling.

The individual home ownership question is valuable as there are no other known alternative datasets available. There is some good quality data on the proportion of dwellings that are owner-occupied, but the distribution of property wealth amongst individuals is not collected outside of census.

The concept being asked in the 2018 Census does not differ from the 2013 Census, however the question has changed from the 2013 in order to improve the respondent comfort and help respondents answer correctly, see figure 12.

Figure 12

Individual home ownership question, 2013 and 2018 censuses

2013 Census	2018 Census
	

For Stats NZ's outputs, the family trust and owned categories can be combined to allow time series analysis. Due to the change in questionnaire design, there may be some change to the data, above and beyond real-world change. However we expect the impact on the data due to this change to be small.

Collection methodology

With the changing collection mode, there has been an increased focus on collecting from the targeted groups. Encouraging young people to complete their census form is part of this targeted approach. As young people, and the targeted populations of Māori and Pacific are less likely to own their homes than the general population, the success of the engagement may have an impact on the data we see. A successful engagement may increase the percentage in the not owned category and reduce the percentage in the owned category.

Imputation will not be used for this variable in 2018 so will not be a factor affecting the data. Non-response for the 2017 Whanganui test was very low at 1.7 percent, but it is important to note the voluntary nature of the test is likely to attract more compliant respondents. Non-response may be higher than this in the 2018 Census.

Real-world change

We expect the majority of changes seen in this data to be a result of real-world change. Particularly, we expect to see a decline in the proportion of respondents owning their home as a result of the well-documented continuing difficulties with affordability of home ownership housing crisis. The declining home ownership rates are likely to be particularly apparent in the younger age groups. As there has been little empirical research into the impact of this on ownership rates, it is difficult to predict the magnitude of the decline in home ownership but it may have dropped to around 47 percent (or possibly less) of respondents owning their own home (down from 49.8 percent in 2013, and 53.2 percent in 2006).

Occupation

Priority 3 variable
Minor/moderate real-world change, minor/moderate data collection change

As part of the suite of questions about work, respondents are asked to give their occupation. There have been some significant changes in 2018 to the way the data will be processed as well as the way the majority of respondents experienced the question.

Online optimisation and coding

For the majority of respondents, an as-you-type list on the internet form helped them select their occupation. All occupations in this list are coded to the occupation classification. Any online responses that are not selected from the as-you-type list and scanned paper responses will go through a process of auto-coding.

The auto-coding process goes through a series of rounds within the processing system, where close matches within the synonym list are compared against the response (eg common misspellings or the occupation title ordered differently – REGISTERED NURSE or NURSE REGISTERED would both be accepted). If a match is not found, then the processing system looks at occupation and industry pairs that were found in 2013 to determine if there is a match. If there is a match, the respondent's occupation will be coded to the occupation found from 2013.

We expect the as-you-type list and the automatic coding will have a positive impact on the numbers of people being coded to their correct occupation. The use of as-you-type lists and auto-coding will reduce the need for manual coding. We do not expect a significant impact on the top level of the classification (which we primarily output on), however it is likely there will be some shifting in the lower levels of the classification as more people are correctly coded to their occupations.

In 2018, occupations will only be coded and output to the ANZSCO classification, compared to in 2006 and 2013, where census data was dual-coded to both NZSCO99 and ANZSCO. The change will still allow comparability to earlier censuses so should not have an impact on the data quality.

Imputation

We are also planning to use donor imputation for the 2018 Census. Therefore we expect the proportion of respondents in the residual categories (non-response and response unidentifiable) to decrease. In 2013, 5.1 percent of responses were residuals – in 2018 these responses will be shared across the other categories.

We expect there will be a similar spread across the highest level of the occupations classification. However, it is likely (with a growing population and a declining unemployment rate) there will be an increase in the numbers of people employed in all occupation groupings, with similar proportions of the labour force in each occupation group. There may be some small changes to the proportion of respondents in each grouping – for example in Technicians and trade workers, Machinery operators and drivers, and Labourers as a result of the Christchurch (and to an extent Wellington) earthquake rebuilds. If these increases eventuate, there would be a corresponding decrease in the proportions of other occupations.

Industry

Priority 3 variable
Minor/moderate real-world change, minor/moderate data collection

The industry worked in by a respondent is derived from three questions asked on the individual form – full name of the business or employer, address of the place worked, and the main activity of the business or employer.

Online optimisation and coding

There have been no changes to the question in 2018, however there were some changes to the way the data was collected. Respondents using the online mode were able to select from an as-you-type list for the main activity of the business or employer. The options available within the as-you-type list were automatically coded using a similar approach to the occupation variable, described in Occupation – Online optimisation and coding.

For respondents who did not use the as-you-type list, or respond on paper, a number of steps will be used to ascertain the respondents correct industry. This will involve looking at the respondent's business name and comparing this to the business register. If this fails to result in coding, the respondents industry or workplace address will be used to code to an industry. If all of these approaches fail, the processing system will try a number of approaches to get a match, including fuzzy matching of the respondent's response. If this approach fails at all steps, imputation is planned – administrative followed by donor imputation.

We anticipate the uptake of the online mode will improve the quality of responses meaning that fewer responses are unidentifiable and need to be resolved with imputation.

In 2018 industries will only be coded and output to the ANZSIC06 classification, compared to in 2006 and 2013 where census data was dual-coded to both ANZSIC96 and ANZSIC06. The change will still allow comparability to earlier censuses so should not have an impact on the data quality.

We expect there to be minimal change to the distribution of industry sizes. We expect that there be an increase in the number of respondents we are confident are being coded to the correct industry, but we do not expect that this quality increase will be seen in any one industry over another. There may be some small changes to the absolute and relative sizes of some industries as a result of real-world change.

Real-world change

We expect that most industries will experience an absolute increase in the number of people employed within them as a result of population increase. Mining, Financial and insurance services, and Arts and recreation services may experience a slight proportional decrease with all other industries experiencing a modest increase or a similar proportion compared to the 2013 Census.

Sector of ownership

Priority 3 variable
Minor/moderate real-world change, minor data collection change

Sector of ownership refers to the sector where respondents are employed: local government, central government, or private sector. The sector of ownership variable has undergone minimal change since the 2013 Census with no change to the concept or questionnaire design, and no change to how the data is being processed.

Imputation

Although there is no change to the way the data is processed, imputation is planned for this variable to resolve non-responses and cases where the response is unidentifiable. Administrative imputation is planned for this variable where available. In cases where this is not possible, donor imputation is planned.

Real-world change

We expect there will be minimal changes to the data as a result of real-world change for this variable. In the 2013 Census, most respondents worked in the private sector (approximately 1.5 million people), with nearly 300,000 working for central government and about 45,000 working for local government. We expect the absolute numbers employed in each sector will increase, however, we expect that the proportions employed in each is unlikely to significantly change.

Unpaid activities

Priority 3 variable
Minor real-world change, minor data collection change

In the 2018 census we asked respondents about unpaid activities they may have done in order to capture work that takes place outside of the paid employment sector. We have not changed the way the unpaid activities variable was asked or the data will be processed. We expect the majority of responses to be collected online, however we expect this will have only a minor impact on the quality of the data.

We expect there to be some absolute increases in the numbers of people reporting that they have done each of the activities in the last four weeks as a result of population increase. Also due to population increase, we expect an increase in the number of respondents saying that they have not done any of the activities. We do not expect to see much, if any, increase in the proportion of any of the activities.

Number of rooms and bedrooms

Priority 3 variable
Minor real-world change, moderate data collection change

The census asks a question about the number of rooms and bedrooms within the dwelling in order to gain an understanding of the suitability of New Zealand's housing stock. Respondents are asked to report the number of bedrooms, lounges, dining rooms, kitchen, conservatories, and studies. The concept being measured has not changed since the previous census in 2013, however the way the question is asked has changed.

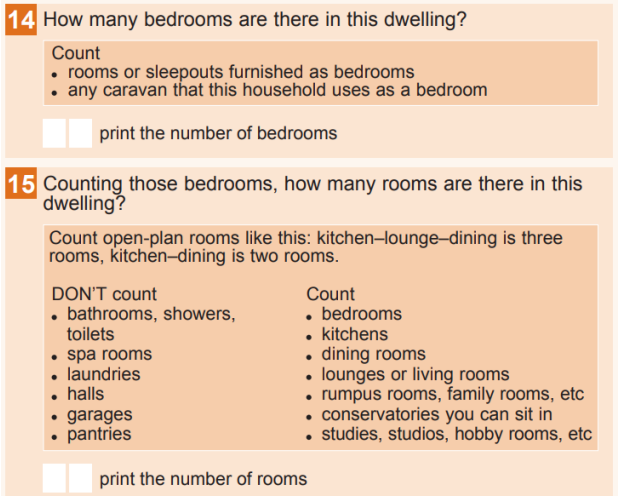
Questionnaire change

Figure 13 shows the questions from the 2013 and 2018 censuses. In 2013 respondents were asked to report the number of bedrooms and the total number of rooms. Reporting the total number of rooms can be cognitively challenging and likely led to data quality issues in the 2013 Census. To make it easier for respondents and improve the data quality for this variable, the 2018 Census asked to report the number of each type of room.

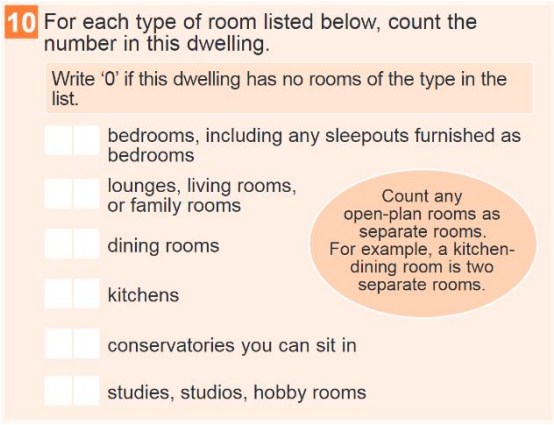
Figure 13

Number of rooms and bedrooms questions, 2013 and 2018 censuses

2013 Census



2018 Census



As respondents in the 2018 Census no longer report the total number of rooms, we created a derivation to derive the total number of rooms. This is expected to enhance data quality as the derivation may be more accurate than some respondents.

Imputation

Imputation is planned to be used for this variable for 2018 with the aim of reducing non-identifiable responses and non-response. Where possible, the number of rooms will be imputed from 2013 Census data and the number of bedrooms will be imputed tenancy bonds data, which is likely to improve data quality including data quality for households who rent.

Real-world change

There are likely to be real-world changes too, although we anticipate that these changes will be consistent with previous census trends. It is expected that dwellings with three bedrooms will continue to dominate, although there is some evidence that these may decline as a proportion while four and five bedroom dwellings become more common.

The increase in four- or five-bedroom dwellings is likely a result of high land prices and developers wanting to maximise their returns. As the prevalence of four and five bedroom dwellings is, at least in part, driven by land values, it is likely that we will see spatial patterns in the number of bedrooms emerging across the nation. With greater diversity in the types of housing being built, we may also see some increase in the proportions of one and two bedroom dwellings, particularly in inner city locations and areas dominated by the elderly.

It is likely that changes to the number of rooms within dwellings closely reflects changes in the number of bedrooms. As many dwellings will have a lounge, dining room, and kitchen, the number of total rooms is likely to be the number of bedrooms plus three. This may not be the case for larger dwellings, or dwellings with only one or two bedrooms.

In previous censuses the most common number of total rooms was six, followed closely by five and seven. In the 2018 Census data we expect to see a similar distribution of data with a possible decrease in the dominance of six-room dwellings closely related to the expected decline in the three- bedroom home.

Main types of heating

Priority 3 variable
Moderate/major real-world change, major data collection change

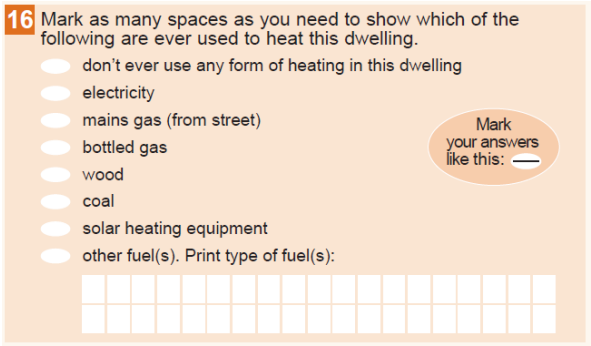
The 2018 Census asked about the ways in which people heat their homes. The information collected for 2018 is the type of appliances being used, with the types of fuel derived. A question on heating was included in the 2013 Census, but it asked respondents what type of fuels they used. As a result, the data from 2006/2013 and 2018 is not comparable, however we expect to produce data on fuels for 2018 so time series analysis of fuels data should still be possible.

The question the 2018 Census includes an ‘other’ option with a write-in box is available for respondents who use a heating appliance not listed (see figure 14). It is expected that responses to this will be minimal. Respondents completing an online census form will have an as-you-type list, which will likely improve the quality of responses for those selecting an ‘other’ option as the responses within the as-you-type list are auto-coded to the classification.

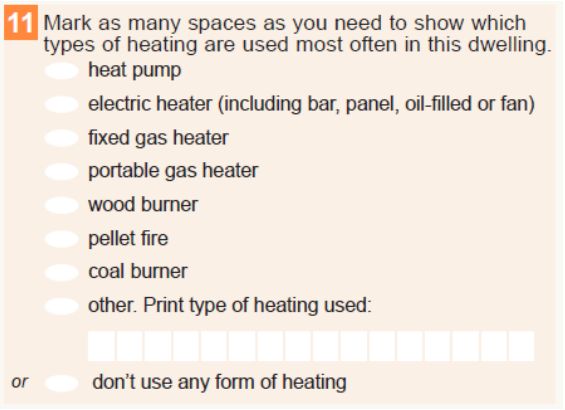
Figure 14

Main types of heating question, 2013 and 2018 censuses

2013 Census



2018 Census



Responses to the 2013 Census question on heating fuels are not comparable, but they may give us some indication of the distribution of the data.

In 2013 the most common heating source was electricity. Therefore we expect that, combined, heat pump and electric heater may be the most common response. Gas fuelled heaters are likely to be relatively uncommon with only a small proportion of dwellings having gas mains, and the dangers of portable gas heaters being well publicised.

Wood burners have been common in the past, however they are likely to be declining in popularity as their environmental impact becomes well known and councils act to impose regulations on them.

As there is little other comparable data available on heating appliances it is difficult to predict how the data will look. However, this question tested well in the last large-scale public test. Imputation is not being used for this variable.

Access to telecommunication systems

Priority 3 variable
Major real-world change, minor data collection change

In the 2018 Census respondents are asked about their access to telecommunications – specifically cellphone, [fixed line] telephone, and internet access. There has been no change to the concept being asked in the questionnaire. However, in the 2013 Census respondents were also asked about whether they had access to a fax machine in their dwelling.

We do not anticipate that removing the fax option will have an impact on the data quality of the other responses as each are independent of one another. There have been no significant changes to how the data is processed since the 2013 Census, and imputation will not be used for this variable in the 2018 Census.

Real-world change

While we do not predict that the questionnaire changes or the processing system will have an impact on the data, there are likely to be some quite significant real-world changes. We predict that the proportion of dwellings with access to a [fixed line] telephone will decline while the proportion with access to a cellphone and internet will increase.

In the 2013 Census, the proportion of households with a fixed line telephone was 81.1 percent. The proportion of dwellings with a telephone has declined from the 2006 and 2001 censuses, although slowly. We anticipate the proportion of dwellings with access to a telephone will continue to decline and that it is likely that the decline will accelerate.

As telephones and cellphones are substitutes, we expect the proportion of dwellings with access to a cellphone to increase from the 79.4 percent in the 2013 Census. It is difficult to predict the increase as while a number of surveys have examined the prevalence of cellphones within the New Zealand population, they largely focus on the number of cellphone connections, which is a different metric from what is used in the census. There is some evidence that we may have reached, or be approaching, market saturation. It is likely that the majority of new connections are to individuals with another cellphone, or into households where there is already an existing cellphone and the census measures whether there is a cellphone within the dwelling. Despite this, we do expect to see a real-world increase.

Likewise, we expect to see an increase in the proportion of dwellings with access to the internet. The proportion of households with access to the internet has increased at each census since 2001. However, the increase has been slowing from 2001 to 2006 and then again between 2006 and 2013. It is possible that like cellphones we are reaching a point of market saturation where new users are coming on board more slowly than in earlier.

Number of motor vehicles

Priority 3 variable
Minor real-world change

As part of the census' role in informing decision making in New Zealand, the census dwelling form asks about the number of motor vehicles available to the household. The concept and question collected in 2018 is identical to the one in the 2013 Census.

There have been no changes to the way the data is handled within the processing system and imputation is not being used. There are not likely to be any significant changes to the distribution of the data as a result of either decisions made by Stats NZ regarding processing or collecting of the data, or by actual changes in the real-world.

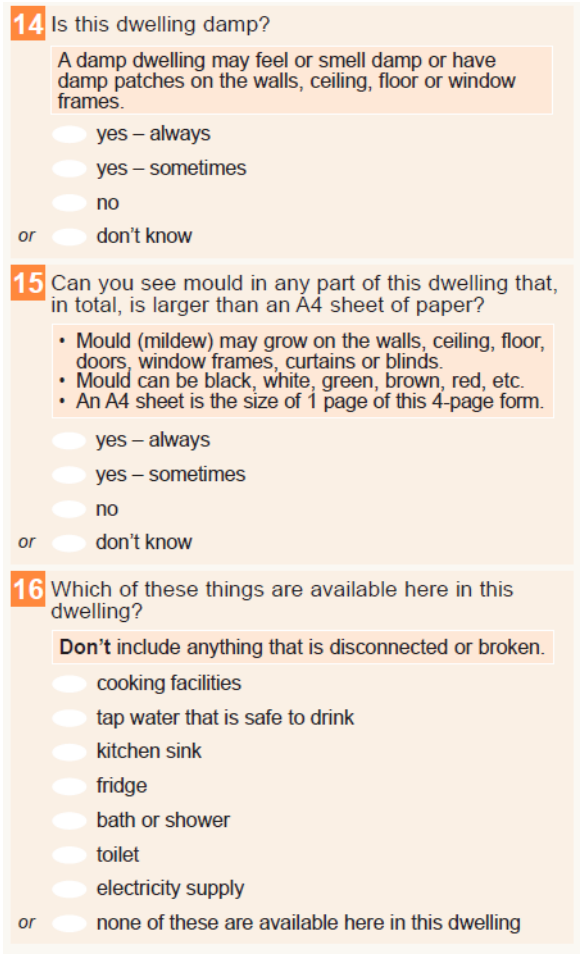
Housing quality

Priority 3 variable
New question, so no real-world change or data collection change

The new set of housing quality variables in the 2018 Census are keenly anticipated by data users. The 2018 Census is the first time questions on mould and damp have been asked in a New Zealand Census, and the first time amenities have been asked about since 1996, see figure 15.

Figure 15

Housing quality questions, 2018 Census



14 Is this dwelling damp?

A damp dwelling may feel or smell damp or have damp patches on the walls, ceiling, floor or window frames.

- yes – always
- yes – sometimes
- no

or don't know

15 Can you see mould in any part of this dwelling that, in total, is larger than an A4 sheet of paper?

- Mould (mildew) may grow on the walls, ceiling, floor, doors, window frames, curtains or blinds.
- Mould can be black, white, green, brown, red, etc.
- An A4 sheet is the size of 1 page of this 4-page form.

- yes – always
- yes – sometimes
- no

or don't know

16 Which of these things are available here in this dwelling?

Don't include anything that is disconnected or broken.

- cooking facilities
- tap water that is safe to drink
- kitchen sink
- fridge
- bath or shower
- toilet
- electricity supply

or none of these are available here in this dwelling

These three indicators can be used in conjunction with each other, or individually, by customers to better understand the quality of New Zealand's housing stock and to evaluate the living conditions of certain populations. A 'don't know' option was included for the dampness and mould question for respondents who don't have the necessary understanding of the dwelling.

While we do not have historical census data to indicate what the distribution of these indicators may look like in the 2018 Census, other surveys may give an insight. The New Zealand General Social Survey (NZGSS) and the BRANZ Housing Condition Survey have both included questions on housing quality. Both surveys indicate that rented homes are more likely to have an issue with mould and dampness. These surveys also indicate that there is likely to be a relationship between use of heating and mould and dampness – with homes that are not heated or are heated with portable gas heaters are more likely to be damp or have mould than those heated with electric heaters.

Conclusion

The new model we employed in the 2018 Census to better reflect the needs of customers has resulted in changes that may impact on the comparability of the data over time. We have outlined the anticipated impacts this report.

While we can estimate how variables may be impacted, it is impossible to ascertain the true impact on data quality until the data has been analysed, however this document offers customers some information about what they might need to consider when using census data.

In addition to this report, more metadata around data quality will be released when available and will serve as an additional resource for using and analysing census data.

References

CANCEIS (Canadian census edit and imputation system). Contact Statistics Canada.

Stats NZ (2017). 2018 Census data quality management strategy. Retrieved from www.stats.govt.nz.

Stats NZ (2017). Statistical standard for geographic areas 2018. Retrieved from http://archive.stats.govt.nz.

Stats NZ (nd). Household Labour Force Survey. Retrieved from www.stats.govt.nz.

Washington Group on Disability Statistics (nd). Washington Group short set of questions on disability. Retrieved from www.washingtongroup-disability.com

Appendix 1: Summary of changes

This table presents the likelihood of change and type of change, for each 2018 Census variable. The variables are in order of priority levels, with order in the individual form and then dwelling form within priority levels.

Appendix table 1

Change for variable, by type of change, 2018 Census								
Variable	**Concept change**	**Questionnaire design change**	**Subject population change**	**As-you-type functionality available**	**Classification change**	**Derivations used**	**Donor Imputation planned**	**Administrative imputation planned**
Age / Date of birth	N	Y	N	N/A	N	Y	Y	N
Sex	N	N	N	N/A	N	Y	Y	N
Usual residence	N	N	N	Y	Y	Y	Y	N
Census night address	N	N	N	Y	Y	Y	Y	N
Ethnicity	N	N	N	Y	N	N	Y	Y
Māori descent	N	N	N	N/A	N	Y	N	Y
Iwi affiliation	N	Y	N	Y	Y	Y	N	N
Birthplace	N	Y	N	Y	N	N	Y	Y
Usual residence one year ago	N/A	N/A	N/A	Y	N/A	Y	N	N
Legally registered marital status	N	N	N	N/A	N	Y	N	N
Partnership status in current relationship	N	N	N	N	N	N	N	N

	Concept change	Questionnaire Design change	Subject population change	As-you-type functionality available?	Classification change?	Derivations used?	Donor Imputation planned?	Administrative imputation planned?
Household composition, family type, extended family type and child dependency status	N	N	N	N	N	Y	Y	N
Main means of travel to education and Educational institution address	N/A	N/A	N/A	Y	N/A	N	Y	N
Highest Secondary School qualification	N	N	N	Y	N	Y	N	Y
Post-school qualification	Y	Y	N	Y	N	Y	N	Y
Field of study	N	Y	N	Y	N	N	N	N
Highest qualification (derived variable)	N	Y	N	Y	N	Y	N	Y
Sources of personal income	N	Y	N	N/A	N	Y	Y	Y
Total personal income	N	N	N	N/A	N	Y	Y	Y
Work and labour force status	N	Y	N	N/A	N	Y	Y	N

	Concept change	Questionnaire Design change	Subject population change	As-you-type functionality available?	Classification change?	Derivations used?	Donor Imputation planned?	Administrative imputation planned?
Hours worked in employment per week	N	Y	N	N/A	N	N	Y	N
Status in employment	N	N	N	N/A	N	N	Y	N
Main means of travel to work	Y	Y	N	N/A	N	N	Y	N
Occupied dwelling type	N	Y	N	N	Y	Y	N	Y
Tenure of household	N	Y	N	N/A	N	Y	N	Y
Years at usual residence	N	Y	N	N/A	N	N	N	N
Years since arrival in New Zealand	N	N	N	N/A	N	N	N	Y
Languages spoken	N	N	N	Y	N	N	Y	Y
Religious affiliation	N	Y	N	Y	Y	N	Y	Y
Study participation	N	N	Y	N/A	N	N	Y	Y
Disability / activity limitations	Y	Y	Y	N/A	N	Y	N	N

	Concept change	Questionnaire Design change	Subject population change	As-you-type functionality available?	Classification change?	Derivations used?	Donor Imputation planned?	Administrative imputation planned?
Cigarette smoking behaviour	N	N	N	N/A	N	N	Y	Y
Number of children born	Y	Y	N	N/A	N	N	N	N
Individual home ownership	N	Y	N	N/A	Y	N	N	N
Occupation	N	N	N	Y	Y	N	Y	N
Industry	N	N	N	Y	N	N	Y	Y
Sector of ownership	N	N	N	Y	N	N	Y	Y
Unpaid activities	N	N	N	N/A	N	N	N	N
Number of rooms and bedrooms	N	Y	N	N/A	N	Y	N	Y
Main types of heating	Y	Y	N	Y	Y?	Y?	N	N
Access to telecommunication systems	Y	Y	N	N/A	N	N	N	N
Number of motor vehicles	N	N	N	N/A	N	N	N	N
Housing quality	N/A	N/A	N/A	N/A	N/A	Y	N	N