2018 Census External Data Quality Panel
Assessment of Variables

Acknowledgements

The 2018 Census External Data Quality Panel would like to thank staff at Stats NZ for their openness and responsiveness and for their help with providing the data, analysis and reports needed to compile this report. We also thank an independent reviewer who provided useful comments on this report.

We were supported throughout the review by a small secretariat provided by Stats NZ. They have been vital to the conduct of the review and we thank them for their efforts.

December 2019

Introduction to the 2018 Census External Data Quality Panel

Stats NZ constituted the 2018 Census External Data Quality panel in August 2018.

Panel members are as follows:

- Richard Bedford, Emeritus Professor, recently retired Professor of Population Geography, Auckland University of Technology and University of Waikato (co-Chair)
- Alison Reid, Team Manager, Economic and Social Research and Evaluation, Auckland Council (co-Chair)
- Dr. Barry Milne, Director, COMPASS Research Centre, University of Auckland
- Dr. Donna Cormack, Senior Lecturer, Te Kupenga Hauora Māori, University of Auckland; Senior Research Fellow, Te Rōpū Rangahau Hauora a Eru Pomare, University of Otago, Wellington
- Ian Cope, international census expert, ex-Office of National Statistics (ONS), United Kingdom
- Len Cook, former New Zealand Government Statistician and former National Statistician of the United Kingdom
- Tahu Kukutai, Professor of Demography, National Institute of Demographic and Economic Analysis, University of Waikato
- Thomas Lumley, Professor of Biostatistics, University of Auckland.

Table of Contents

1. Introduction 10
1.1. Quality assessment 10
1.2. Data sources 11
1.3. Variables assessed 13
2. Detailed assessments - variables about people 17
2.1. Absentees 17
2.1.1. Overall assessment 17
2.1.2. Background 17
2.1.3. Coverage 18
2.1.4. Consistency 19
2.1.5. Comparability 19
2.1.6. Contemporaneity 19
2.2. Cigarette smoking behaviour 20
2.2.1. Overall assessment 20
2.2.2. Background 20
2.2.3. Coverage 21
2.2.4. Consistency 22
2.2.5. Comparability 23
2.2.6. Contemporaneity 23
2.3. Birthplace 24
2.3.1. Overall assessment 24
2.3.2. Background 24
2.3.3. Coverage 25
2.3.4. Consistency 25
2.3.5. Comparability 26
2.3.6. Contemporaneity 27
2.4. Hours worked in employment per week 28
2.4.1. Overall assessment 28
2.4.2. Background 28
2.4.3. Coverage 28
2.4.4. Consistency 29
2.4.5. Comparability 29
2.4.6. Contemporaneity 30
2.5. Industry 31

2.5.1. Overall assessment 31
2.5.2. Background 31
2.5.3. Coverage 33
2.5.4. Consistency 33
2.5.5. Comparability 34
2.5.6. Contemporaneity 35
2.6. Languages spoken 36
2.6.1. Overall assessment 36
2.6.2. Background 37
2.6.3. Coverage 38
2.6.4. Consistency 40
2.6.5. Comparability 40
2.6.6. Contemporaneity 41
2.7. Main means of travel to education 42
2.7.1. Overall assessment 42
2.7.2. Background 42
2.7.3. Coverage 43
2.7.4. Consistency 44
2.7.5. Comparability 44
2.7.6. Contemporaneity 44
2.8. Main means of travel to work 45
2.8.1. Overall assessment 45
2.8.2. Background 46
2.8.3. Coverage 46
2.8.4. Consistency 48
2.8.5. Comparability 48
2.8.6. Contemporaneity 49
2.9. Occupation 50
2.9.1. Overall assessment 50
2.9.2. Background 50
2.9.3. Coverage 51
2.9.4. Consistency 53
2.9.5. Comparability 53
2.9.6. Contemporaneity 53
2.10. Qualifications: Highest qualification 54

2.10.1. Overall assessment 54
2.10.2. Background 54
2.10.3. Coverage 55
2.10.4. Consistency 56
2.10.5. Comparability 57
2.10.6. Contemporaneity 57
2.11. Religious affiliation 58
2.11.1. Overall assessment 58
2.11.2. Background 58
2.11.3. Coverage 59
2.11.4. Consistency 60
2.11.5. Comparability 60
2.11.6. Contemporaneity 62
2.12. Status in employment 63
2.12.1. Overall assessment 63
2.12.2. Background 63
2.12.3. Coverage 64
2.12.4. Consistency 64
2.12.5. Comparability 65
2.12.6. Contemporaneity 65
2.13. Study participation 66
2.13.1. Overall assessment 66
2.13.2. Background 66
2.13.3. Coverage 67
2.13.4. Consistency 67
2.13.5. Comparability 68
2.13.6. Contemporaneity 69
2.14. Total Personal Income & Sources of Personal Income 70
2.14.1. Overall assessment 70
2.14.2. Background 70
2.14.3. Coverage 71
2.14.4. Consistency 72
2.14.5. Comparability 72
2.14.6. Contemporaneity 72
2.15. Usual residence five years ago 73

2.15.1. Overall assessment 73
2.15.2. Background 73
2.15.3. Coverage 74
2.15.4. Consistency 74
2.15.5. Comparability 75
2.15.6. Contemporaneity 75
2.16. Work and Labour Force status 76
2.16.1. Overall assessment 76
2.16.2. Background 76
2.16.3. Coverage 77
2.16.4. Consistency 77
2.16.5. Comparability 78
2.16.6. Contemporaneity 78
2.17. Years since arrival in New Zealand 79
2.17.1. Overall assessment 79
2.17.2. Background 79
2.17.3. Coverage 79
2.17.4. Consistency 81
2.17.5. Comparability 81
2.17.6. Contemporaneity 81
3. Detailed assessments - variables about dwellings and households 82
3.1. Access to telecommunications systems 82
3.1.1. Overall assessment 82
3.1.2. Background 82
3.1.3. Coverage 83
3.1.4. Consistency 83
3.1.5. Comparability 84
3.1.6. Contemporaneity 85
3.2. Counts of Dwellings 86
3.2.1. Overall assessment 86
3.2.2. Background 86
3.2.3. Coverage 87
3.2.4. Consistency 87
3.2.5. Comparability 88
3.2.6. Contemporaneity 88

3.3. Dwelling type 89
3.3.1. Overall assessment 89
3.3.2.Background 90
3.3.3. Coverage 90
3.3.4. Consistency 91
3.3.5. Comparability 91
3.3.6. Contemporaneity 93
3.4. Housing quality 94
3.4.1. Overall assessment 94
3.4.2. Background 94
3.4.3. Coverage 95
3.4.4. Consistency 95
3.4.5. Comparability 95
3.4.6. Contemporaneity 96
3.5. Main types of heating and fuel types used to heat dwellings 97
3.5.1. Overall assessment 97
3.5.2. Background 97
3.5.3. Coverage 98
3.5.4. Consistency 99
3.5.5. Comparability 99
3.5.6. Contemporaneity 100
3.6. Number of bedrooms and number of rooms 101
3.6.1. Overall assessment 101
3.6.2. Background 101
3.6.3. Coverage 102
3.6.4. Consistency 102
3.6.5. Comparability 103
3.6.6. Contemporaneity 103
3.7. Number of motor vehicles 104
3.7.1. Overall assessment 104
3.7.2. Background 104
3.7.3. Coverage 105
3.7.4. Consistency 105
3.7.5. Comparability 106
3.7.6. Contemporaneity 106

3.8. Tenure of household 107
3.8.1. Overall assessment 107
3.8.2. Background 108
3.8.3. Coverage 109
3.8.4. Consistency 109
3.8.5. Comparability 110
3.8.6. Contemporaneity 110
References 111
Appendix 1 – Stats NZ data quality assurance definitions for 2018 Census 112
Appendix 2 – Glossary 115

1. Introduction

Statistics New Zealand (Stats NZ) has adopted a rating scale to assess the quality of each variable measured in the 2018 Census of Population and Dwellings. The 2018 Census External Data Quality Assurance Panel (the panel) has summarised its findings on the variables it has assessed by using this scale as well. The scale has five options - Very High, High, Moderate, Poor, Very Poor – and there are three metrics that Stats NZ has adopted to calibrate this scale. A summary of these metrics is provided in Appendix 1; further detail can be found in Data quality assurance for 2018 Census.

The quality assessments by Stats NZ are a welcome initiative. Such measures enable users to assess the fitness for use of the statistical estimates and models that Stats NZ produce, and to have those assessments independently challenged. As a result, the richness of the information obtained in each population census can be used more critically in public policy, social and economic research and analysis, and community development.

When assessing the quality of data for several of the variables covered in this report, the panel has found a single category on the rating scale was not appropriate. For example, the data may be 'moderate' in terms of its quality at certain levels of aggregation and for certain groups in the population, but 'poor' at other levels of aggregation and for other groups. In some of the assessments in this report the panel's consensus judgement is that a fair assessment spans two categories.

The panel assessed key variables such as age, sex, usually resident population count, Māori descent electoral, and ethnicity (levels 1 and 2) in the Initial report of the 2018 Census External Data Quality Panel, which was published to coincide with the 23 September first release by Stats NZ of 2018 Census data.

The panel has assessed an additional 30 variables in this report. These variables, along with the ones assessed in the initial report, and a small number that will be commented on in the panel's final report to be published in early 2020, are listed in Tables 1.1 and 1.2. The shaded areas in these tables highlight the variables covered in this report.

1.1. Quality assessment

The panel adopted a standard format for each assessment. For each variable this assessment includes the following information:

- the panel's rating of quality;
- Stats NZ's rating of quality;
- a link to the relevant Stats NZ DataInfo+ page with more detail on each variable;
- an overall assessment;
- background information; and
- consideration of various aspects of quality (such as completeness, regional/ethnic coverage, whether the data used was contemporaneous to the census date). Note that **ethnic coverage is only investigated for variables about individuals**. Stats NZ does not usually release variables about dwelling and households stratified by ethnicity (because dwelling and households often comprise several ethnicities).

These headings were based on, and informed by, the Statistics Canada data quality framework (Statistics Canada, 2017) and are broadly consistent with the approach taken by Stats NZ. Higher levels of aggregation are likely to mask data quality issues through equal and opposite errors cancelling; conversely greater disaggregation is likely to reveal data quality problems. The panel has given greater weight to the level of disaggregation (geographic or by level of the relevant classification) at which the data are likely to be used, which has led to some variables having a panel quality rating that is lower quality than assessed by Stats NZ.

The very nature of a population census means that not all variables can be measured with the same quality expectations. This reflects the conceptual, methodological and practical limitations of census-type enquiries. In this report, the panel has sought to provide and summarise information that will be of value to users when they use 2018 Census data – pointing out:

- those variables for which the data can be used with confidence, especially at high levels of spatial aggregation (there are often caveats attached about use of the data at low levels of spatial aggregation – e.g. SA1 and SA2)
- where limitations in the data were regarded as acceptable and anticipated at the design stages
- those variables for which the data can be used, but with caution, and
- those variables for which the data should not be used.

The panel made use of three resources to assess the quality of each variable:
(i) the Stats NZ DataInfo+ page for each variable, which gives a high level summary of the quality of the variable;
(ii) In depth 'warrant of fitness' assessments undertaken by Stats NZ and made available to the panel (see Processing and evaluating the quality of 2018 Census data for a description of the 'warrant of fitness' process); and
(iii) information on the sources of data that contributed to each census variable (e.g., directly from 2018 census, 2013 census, administrative data, imputation), stratified by small area and ethnicity.

Readers should note that the panel's quality assessments are based on the information that was available when the report was compiled (October – December 2019). As at December 2019, Stats NZ are continuing to undertake assessments of the quality of a number of variables. It is not possible for the panel to take these ongoing assessments into account, nor can we speculate whether these ongoing assessments would change the panel's quality assessment or rating for any variable.

1.2. Data sources

Both the panel's and Stats NZ's rating of the quality of variables was influenced by the extent to which data for variables were obtained from sources other than individuals' responses to questions on the 2018 census forms. There were potentially four sources of data for each census variable, each with their own limitations.

First, data could be sourced directly from 2018 census responses. The limitation of this source is that due to the non-response problems for the 2018 Census information was often

not available for a sizeable part of the population. The census returns also included 'residual category' information (i.e., 'not stated', 'response outside scope', 'response unidentifiable', 'refused to answer', or 'don't know'). Data from this source was always used where it was available and did not constitute a residual category.

Second, data from the same individual's 2013 census response was used. When an individual was successfully linked to the Integrated Data Infrastructure (IDI), which included the 2013 census dataset, existing responses to the 2013 census could be copied across to fill in gaps in individual variables. For example, information on smoking and 'usual residence five years ago' were taken from 2013 census responses. Some variables (such as birthplace, or number of bedrooms) don't change over time, or do so slowly – in such instances the 2013 Census is a high quality source. The primary limitation of this method is that the data was reported at an earlier time point, so change, where it exists, will be increasingly underestimated with increasing use of 2013 census data.

Third, 'admin data' were used. When an individual was successfully linked to the IDI, existing administrative (admin) data could be copied across to fill in gaps in individual variables. For example, ethnicity was taken mostly from birth, education and health data. Limitations of this source include that (i) admin data sometimes doesn't measure exactly the same concept as the census question sought to measure; (ii) admin data sometimes isn't classified using the same classification systems as Stats NZ census variables; (iii) admin data sometimes is measured at a different time to the 2018 census measure; and (iv) individual's responses to questions may differ between administrative settings and the census (e.g. ethnicity may be reported differently across different settings).

Fourth, 'imputation' was used. The primary form of imputation used was a form of 'nearest neighbour' 'donor' imputation (i.e. find a census respondent who is similar to the census respondent with missing information for a census question, and copy cross the 'donor's response). The specific system was called CANCEIS (CANadian Census Edit and Imputation System), developed by Statistics Canada. This is described in more detail in section 3.2.5 of the panel's initial report and also in Stats NZ (2019a). For example, occupation data was imputed using CANCEIS. Imputation tends to be *unbiased*, so accurate counts are likely to be generated. However, the main limitation of imputation is that it can be inaccurate at the individual level (i.e. a response is used for an individual that does not match the response that individual would have given had they completed the 2018 census). This inaccuracy may serve to decrease estimates of association between two variables, where one or other variable used a substantial amount of imputation.

Stats NZ have applied imputation to 27 variables in 2018, but to only four variables – age, sex, usual residence and labour force status – in 2013 (Stats NZ, 2014). This means that the 2018 Census dataset is more 'complete' than for previous censuses, which in itself can impact on measures of change.

We note that use of alternative sources (i.e., second through fourth above) will have substantially improved the data compared to leaving large amounts of 'no information' (i.e. if only 2018 census responses were used). However, the data will tend to be less reliable and less timely than if higher response rates to the 2018 census had been achieved.

1.3. Variables assessed

Variables which were assessed in the panel's initial report are listed in Tables 1.1 and 1.2 but they have not been assessed further in this report. Reference should be made to the initial report for detailed assessments of the quality of the data for these variables. The panel's final report contains further analyses of ethnicity, down to Level 4 of the ethnicity classification, a section on small area data and a section on the quality of data on families and households.

The panel has not been able to assess all variables that Stats NZ will be releasing and has focussed its efforts on those which Stats NZ has rated as having data which is of Very High, High, or Moderate quality. The Stats NZ quality ratings for the variables considered by the panel in this report, its initial report and its final report are listed in Tables 1.1 and 1.2.

In addition, the usual residence five years ago variable was assessed as, uniquely, this was always designed to be generated from comparing 2013 and 2018 Census data rather than, as previously, asking a specific question in the census. One variable rated as Very Poor (Absentees) has been assessed, partly as a matter of completeness and to be able to explain the drivers for a Very Poor rating.

Tables 1.1 and 1.2 summarise the position for all variables. The first table covers variables about people (derived from the individual form), the second table covers those variables collected at the dwelling level (from the dwelling form).

The tables provide the variable name (which links to the relevant DataInfo+ page), the Stats NZ priority level, where the variable has been assessed (initial report, this report, final report, not assessed) and the panel's and Stats NZ's quality ratings.

By and large the panel endorse the quality ratings given by Stats NZ, but in several instances the panel has rated the variables of mixed quality or lower quality than Stats NZ.

As part of this assessment process the panel has identified improvements that could be made to the DataInfo+ pages – Stats NZ have implemented such changes. There are a number of questions (such as Tenure of Household; Main means of travel to work; Main types of heating and fuel types used to heat dwellings) where question changes for the 2018 Census have made comparisons with the 2013 Census difficult. For such variables we recommend **that Stats NZ should** undertake further analysis of the impact of the changes in the wording or response options of the questions., and should compare 'Main means of travel to work' data with the Ministry of Transport Household Travel Survey.

Table 1.1 Summary assessments - variables about people

Variable name	Priority level	Where covered	EDQP Quality rating	Stats NZ Quality rating	Page no.
Absentees	1	This report	Very Poor	Very poor	17

Variable name	Priority level	Where covered	EDQP Quality rating	Stats NZ Quality rating	Page no.
Activity limitations	3	Not assessed	N/A	Poor	
Age	1	Initial report	Very High	Very high	
Census night population count	1	Initial report	Moderate	Moderate	
Census usually resident population count	1	Initial report	Very High	Very high	
Cigarette smoking behaviour	3	This report	Moderate /Poor	Moderate	20
Birthplace	2	This report	High	High	24
Educational institution address	2	Not assessed	N/A	Moderate	
Ethnicity	1	Initial report and final report	Moderate	High	
Families and households: extended family type	2	Section in final report	Very Poor	Very poor	
Families and households: family type	2	Section in final report	Very Poor	Very poor	
Families and households: household composition	2	Section in final report	Very Poor	Very poor	
Hours worked in employment per week	2	This report	Moderate /Poor	Moderate	28
Individual home ownership	3	Not assessed	N/A	Poor	
Industry	3	This report	High	High	31
Iwi	2	Initial report	Very Poor	Very poor	
Languages spoken	3	This report	Very high to poor	High	36
Main means of travel to education	2	This report	Moderate	Moderate	42
Main means of travel to work	2	This report	Poor	Moderate	45
Māori descent – output	1	Initial report	High	High	
Māori descent – electoral	1	Initial report	High	High	
Number of children born	3	Not assessed	N/A	Moderate	

Variable name	Priority level	Where covered	EDQP Quality rating	Stats NZ Quality rating	Page no.
Occupation	3	This report	Poor	Moderate	50
Qualifications: highest qualification	2	This report	Moderate /Poor	Moderate	54
Qualifications: highest secondary school qualification	2	This report	Moderate /Poor	Moderate	54
Qualifications: post-school qualification level of attainment	2	This report	Moderate /Poor	Moderate	54
Qualifications: post-school qualification field of study	2	Not assessed	N/A	Poor	
Relationship status: Legally registered relationship status, and partnership status in current relationship	2	Not assessed	N/A	Poor	
Religious affiliation	3	This report	High	High	58
Sector of ownership	3	Not assessed	N/A	Moderate	
Sex	1	Initial report	Very High	Very high	
Sources of personal income	2	This report	High	High	70
Status in employment	2	This report	Moderate	Moderate	63
Study participation	2	This report	Moderate /Poor	High	66
Total personal income	2	This report	High	High	70
Unpaid activities	3	Not assessed	N/A	Poor	
Usual residence address	1	Initial report	High	High	
Usual residence one year ago	2	Not assessed	N/A	Poor	
Usual residence five years ago	2	This report	Poor	Poor	73
Work and labour force status	2	This report	Moderate	Moderate	76
Workplace address	2	Not assessed	N/A	Moderate	
Years at usual residence	3	Not assessed	N/A	Poor	

Variable name	Priority level	Where covered	EDQP Quality rating	Stats NZ Quality rating	Page no.
Years since arrival in New Zealand	3	This report	Moderate	Moderate	79

Table 1.2 Summary assessments - variables for dwellings

Variable name	Priority level	Where covered	EDQP Quality rating	Stats NZ Quality rating	Page no.
Access to telecommunication systems	3	This report	Moderate	Moderate	82
Census night address	1	Initial report	Moderate	Moderate	
Counts of dwellings	1	This report	High	High	86
Dwelling occupancy status	N/A	N/A	N/A	N/A	
Dwelling type	2	This report	Poor	Moderate	89
Housing quality: access to basic amenities	3	This report	Moderate	Moderate	94
Housing quality: dwelling dampness indicator	3	This report	Moderate	Moderate	94
Housing quality: dwelling mould indicator	3	This report	Moderate	Moderate	94
Main types of heating and fuel types used to heat dwellings	3	This report	Moderate	Moderate	97
Number of bedrooms	3	This report	High	High	101
Number of rooms	3	This report	Poor	Poor	101
Number of motor vehicles	3	This report	Moderate	Moderate	104
Sector of landlord	2	Not assessed	N/A	High	
Tenure of household	2	This report	Moderate	Moderate	107
Weekly rent paid by household	2	Not assessed	N/A	Moderate	

2. Detailed assessments - variables about people

2.1. Absentees

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/af809275-6df9-4749-84eb-f83b2537722f/?_ga=2.246967044.160130192.1572492057-1393686945.1568850831

EDQ Panel rating: **Very Poor**

Stats NZ rating: **Very Poor**

2.1.1. Overall assessment

In general the panel is not assessing variables rated as Poor or Very Poor, but has done this one as an example of drivers for a rating of very poor.

Stats NZ rate the quality of 'Absentees' as Very Poor and at this point in time are not releasing the information. The panel endorse this assessment and do not believe the data are fit for use at any level of geography.

Stats NZ state "The quality rating is based on assessment of all of the variables within the Absentees suite of questions, with particular attention to 'In NZ on Census Night' and 'Time Away from NZ' as they're considered key information specific to absentees."

The main driver of the 'Very Poor' quality rating for absentees is the high non-response to 'Absentee in NZ on census night'. As family coding does not use this variable, the very poor quality of this suite has minimal impact on families and households.

"The non-response rate for 'In NZ on Census Night' was 42.3 percent; the response rate of 57.7 percent is equivalent to a rating of 'very poor quality' coverage. …the high non-response to the 'In New Zealand on census Night' question [means there is] no information as to whether the absentee was in New Zealand on census night or not."

Stats NZ usually only output data on 'Absentee In New Zealand on Census Night', 'Absentee Time away from New Zealand' and 'Number of census night absentees' – see, for instance, table 50 in 2013 Census totals by topic. In 2018 the relevant data are severely impacted by the low response rate for the 'In New Zealand on Census Night' question.

2.1.2. Background

Absentees was a Priority 1 variable for the 2018 Census (a Priority 2 variable for 2013). Absentees have a significant role in family and household derivations in determining the usual residents away from home on census night.

Absentees is a 'suite' of questions that includes the count of absentees, as well as details (Age, Sex, Relationship to reference person, In NZ on Census Night, and Time Away from NZ) about absentees. Of these, only 'Name' was mandatory to complete on the 2018 online Household Set-up Form, which appears to have resulted in high non-response to the 'In New Zealand on Census Night' question.

Stats NZ output (limited) data about absentees: the number of census night absentees, absentee in New Zealand on census night and absentee time away from New Zealand). The rest of the variables in the suite (age, sex, relationship to reference person) are not output, but are used in creating families and household data.

A decision was made not to impute information for these records. It was recognised that this would impact on data quality for this variable but was less likely to negatively impact other data/subject populations.

There was a significant decrease from previous censuses in absentees who responded that they were 'In New Zealand on census night' to the 'In New Zealand on census night question'. This is a result of the high 'Not stated' category which is significantly greater than it was in 2013 and 2006.

For 'Time away from New Zealand' (for absentees who reported as not being in NZ on Census night), there are similar trends in 2018 compared to 2006 and 2013.

Table 2.1 below shows the data sources used for absentee in New Zealand on census night – i.e. only 2018 Census returns. Absentee time away from New Zealand comes 100 percent from 2018 Census responses.

Table 2.1. Data Sources: Absentee in New Zealand on census night – absentees from occupied private dwellings

Source	Percent
Response from 2018 Census	57.7
2013 Census data	0.0
Administrative data	0.0
Statistical imputation	0.0
No information	42.3
Total	100

2.1.3. Coverage

Was there good coverage for the overall population?

No. There is not good coverage of Absentees, with only a 57.7 percent response rate to the 'In New Zealand on Census Night' question.

Was there good coverage for ethnic groups and regions?

No, because of the large non-response rate to the key absentee questions.

Overall there was a significant increase in the number of absentees nationally and by region, exceeding expectations, although Stats NZ say that there were some concerns about the quality of the 2013 count.

All other absentee variables are consistent with expectations, with the exception of 'In New Zealand on Census Night' and subsequently 'Time away from New Zealand' which did not meet expectations. However, these are the two key variables within this suite of questions.

The main use of absentee information is to provide information about usual residents who were away from home on census night, in order to derive household and family variables. Since household and family variables do not use 'Absentee in NZ on census night', the very poor quality of this suite of questions therefore has minimal impact on household and family variables.

2.1.4. Consistency

Was a consistent classification used?

The classification of the absentee variables in the 2018 Census is consistent with the classifications used in the 2013 and 2006 Censuses.

Was data collection consistent across online and paper data collection methods?

No – The questions were not mandatory in either the 2013 or 2018 Census but there were clearly problems with the 2018 online design, with many respondents skipping these questions.

2.1.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

There is no alternative source for this information.

Is there a consistent time series with previous census, particularly census 2013?

No.

2.1.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

There was no use of administrative data for this variable – all responses were from the 2018 Census, as at the census reference period.

2.2. Cigarette smoking behaviour

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/6c30d129-d952-46ab-b02b-e3ee9c8963b9/

EDQ Panel rating: **Moderate/Poor**

Stats NZ rating: **Moderate**

2.2.1. Overall assessment

Given that smoking prevalence is declining, the use of 2013 Census data to address under-enumeration in 2018 Census would be expected to overestimate current smoking (at all ages) and underestimate 'ever smoked' in young adults. Stats NZ state that analysis indicated that the use of 2013 Census data has introduced some bias towards 'regular smokers' into the dataset, but that this bias is small. The estimated smoking prevalence is 13.2 percent for 2018 Census compared to 15.1 percent for 2013 Census.

The use of administrative enumerations has increased the population of some subgroups (e.g. Māori and Pacific and MELAA peoples, especially young adults) compared to previous censuses, and therefore introduced step changes in the trends between censuses.

Stats NZ consider that cigarette smoking behaviour is consistent with expectations at a national level of geography. However, due to data quality issues they "caution the use of this data at both TA and Regional Council Level."

Given that national estimates of smoking trends are measured in sample surveys, the information on smoking in the census is likely to be used in analyses of change for different groups in the population. The use of responses from the 2013 Census to fill gaps in responses in 2018 Census will introduce a bias in such analyses, as well as bias in estimates of smoking uptake and quitting between 2013 and 2018. Bias will be greater for groups which relied more heavily on 2013 Census data (e.g. Māori and Pacific and MELAA peoples, especially young adults, and those from Northland and Gisborne). Data should be considered **poor** for these groups.

2.2.2. Background

Cigarette smoking refers to the active smoking of manufactured or hand-rolled tobacco cigarettes. It does not include, for instance, the smoking of cigars, pipes, or e-cigarettes.

Cigarette smoking behaviour is used to monitor changes in smoking prevalence among the population of New Zealand. It is used to understand the profile of smokers, target and evaluate the success of health education programmes, examine inter-relationships between smoking and other socioeconomic variables and measure how these change over time.

Data from 2018 Census were available for 84-85 percent of individual responses, from 2013 Census for 7-8 percent and from imputation for the remaining 8.1 percent (see Table 2.1.1). In previous censuses, responses that could not be classified or did not provide the type of information asked for were grouped (along with 'not stated) and classified as 'not elsewhere included' (see next section).

Table 2.2.1. Data Sources: Cigarette smoking – Census usually resident population aged 15 years and over

	Regular smoker	Ever smoked
Source	**Percent**	**Percent**
Response from 2018 Census	84.0	85.1
2013 Census data	7.8	6.8
Administrative data	0.0	0.0
Statistical imputation	8.1	8.1
No information	0.0	0.0
Total	100	100

2.2.3. Coverage

Was there good coverage for the overall population?

Coverage is complete as there are no 'not elsewhere included' responses for this variable in 2018 Census, as a result of use of alternative sources (i.e. historic data from the 2013 census and statistical imputation). There were 9.2 percent not elsewhere included responses in 2013 and 8.6 percent in 2006. Historic data and imputation have been used more for the young adult population in 2018 Census.

Was there good coverage for ethnic groups and regions?

While there is complete data for everyone, the level of information from 2018 census responses differs by ethnic group, for both 'regular smokers' and 'ever smoked' (Table 2.2.2). This is important given the known differential pattern of cigarette smoking by ethnicity. The ethnic groups most impacted by coverage issues are the same as those that are more likely to be 'regular' or 'ever' smokers.

Table 2.2.2. Rate of response from 2018 Census form: Cigarette smoking by ethnicity – Census usually resident population aged 15 years and over

	Regular smoker	Ever smoked
Level 1 Ethnic grouping	**Percent**	**Percent**
Māori	70.6	72.6
Pacific Peoples	67.0	67.2
Asian	82.7	82.8
Middle Eastern/Latin American/African	79.1	79.8
Other Ethnicity	80.8	81.7
European	88.1	88.8

European and Asian ethnic groups have the lowest proportion of 2013 Census data and imputation, across age groups. For Māori, Pacific, and MELAA ethnic groups, young adults have the highest proportions of use of historic data and imputation.

- For Māori, for the 'smokes regularly' variable, the age group 25-29 years has the highest proportion of use of 2013 Census data and imputation combined - the figures are 2013 Census 18.7 percent and imputation 17.5 percent. For 'smokes ever' the age group 20-24 years has the highest proportion - the figures are 2013 Census 21.0 percent and imputation 15.3 percent.
- For Pacific Peoples the age group 20-24 years has the highest proportion of use of 2013 Census data and imputation combined. For 'smokes regularly' the figures are 2013 Census 21.2 percent and imputation 17.4 percent. For 'smokes ever' the figures are 2013 Census 22.4 percent and imputation 18.0 percent.
- For the MELAA ethnic group, for the 'smokes regularly' variable, the age group 25-29 years has the highest proportion of use of 2013 Census data and imputation combined - the figures are 2013 Census 7.8 percent and imputation 22.1 percent. For 'smokes ever' the age group 20-24 years has the highest proportion - the figures are 2013 Census 12.5 percent and imputation 16.2 percent.

Users analysing cigarette smoking should be aware of the high level of use of historic data and imputation for Māori, Pacific and MELAA ethnic group overall, and particularly for young adults. This level of historic data is likely to overestimate smoking rates for these groups, since smoking has declined since 2013, and Census 2013 data was the primary data source used where a valid response from the 2018 Census was not available.

Stats NZ used the New Zealand Health Survey trends in smoking to estimate the likely level of bias in smoking prevalence in the 2018 Census, given the extent of use of 2013 Census data. Based on estimates of change between 2013 and 2018 from the New Zealand Health Survey, Stats NZ estimated bias to be 0.2 percent for the overall population (i.e., smoking may be overestimated by 0.2 percent in 2018), and up to 1.0 percent for Māori (i.e., smoking in Māori may be overestimated by 1.0 percent in 2018). These estimates of bias may be slightly different for different age groups. The New Zealand Health Survey estimates that smoking in Pacific peoples has not changed since 2013; if true, use of 2013 census data for Pacific peoples would not bias estimates of 2018 census smoking prevalence.

Small-area analysis may be unreliable. Stats NZ state "Regions and TAs in those areas of higher non-response rates may show an increase in smoking rates, in particular, Kawerau, and Wairoa Districts. The bias introduced through the use of historic data sources may be more prevalent at TA level."

2.2.4. Consistency

Was a consistent classification used?

The classification of cigarette smoking behaviour in the 2018 Census is consistent with the classification used in the 2013 and 2006 Censuses.

Was data collection consistent across online and paper data collection methods?

Not quite. On the online form, those who indicated that they were a 'current smoker' did not answer the 'ever smoked' question. On the paper form, multiple responses could not be prevented. Where multiple responses were given and a valid response was not able to be determined, imputation was used. Those not in the subject population (i.e. those under the age of 15) were not asked the smoking questions on the online form but could complete the questions on the paper form. Responses for those not in the subject population were filtered out.

2.2.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

2018 Census regular smoking rates (13.2 percent) are lower than Ministry of Health estimates (14.9 percent), although there was a larger difference in 2013 (15.1 percent vs 17.7 percent).

Is there a consistent time series with previous census, particularly census 2013?

There will be consistency issues for some ethnic groups and age groups, including those who are most likely to be 'regular smokers' or 'ever smokers'. In some regions, especially those which had low 2018 Census response rates, there was a heavier reliance on data from the previous census records and imputation (e.g. Northland, Gisborne).

We note that 2013 census smoking estimates themselves may be biased. Speculatively, if 2013 smoking prevalence is underestimated (because those missed from the 2013 Census are more likely to be smokers, as they were in 2018) and 2018 smoking prevalence is underestimated (because of use of 2013 census data in 7-8 percent of cases, when smoking prevalence was higher), then this will serve to underestimate the level of smoking change between 2013 and 2018. That is, smoking may have decreased more than is apparent from census estimates in 2013 and 2018.

2.2.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

No. 7-8 percent of 2018 Census records came from the 2013 Census, which will have biased the results upwards by a small amount (given that smoking rates are in decline). These proportions are significantly higher and may have more impact in some population groups (i.e. Māori and Pacific ethnic groups, young people).

2.3. Birthplace

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/f441f28b-4215-468e-8167-236cc766633e

EDQ Panel rating: **High**

Stats NZ rating: **High**

2.3.1. Overall assessment

Birthplace is a characteristic that, unlike some others in the census, does not change over time. The use of 2013 Census or admin records (such as birth certificates and visa data) has provided high quality information on birthplace. In fact, the non-response rate after use of admin data is significantly lower than previous years at 1.2 percent of usual resident population in 2018, compared with 5.9 percent in 2013 and 4.5 percent in 2006.

The distribution of birthplace appears consistent with previous census results plus known migration in the last five years. However, it should be noted that because of improvements to the 2018 Census data, such as better coding of responses (including an As-You-Type-List online), and the use of 2013 Census and admin data to replace missing responses and residual responses (e.g. don't know), for some groups there will be a break in the time series. Care should therefore be adopted when comparing change in birthplace populations between censuses.

Stats NZ state "For this variable, data quality was checked to Level 3 (4 digit) of the classification at a National and Regional Level, using data from external migration patterns, and other relevant Census 2018 variables (Ethnicity, Languages and Years Since Arrival)." Data will be published down to SA1 Level, but have not yet been checked at that level of geography.

2.3.2. Background

Table 2.3 below shows the breakdown of the various data sources used for this variable.

Table 2.3. Data sources: Birthplace – Census usually resident population

Source	Percent
Response from 2018 Census	83.8
2013 Census data	8.6
Administrative data	6.4
Statistical imputation	0.0
No information	1.2
Total	100

The ‘no information’ percentage relates people whose birthplace data could not be sourced from the 2018 or the 2013 Censuses, or from admin data.

Data from the following administrative sources was used:
- Births register, Department of Internal Affairs
- Migration data, Ministry of Business, Innovation and Employment.

2.3.3. Coverage

Was there good coverage for the overall population?

Yes. 83.8 percent of birthplace data came from 2018 Census records, with 2013 Census and admin data filling most of the gap. Only 1.2 percent of records were recorded as ‘missing’.

Was there good coverage for ethnic groups and regions?

Yes. It is likely that some migrant communities, defined by birthplace, were disproportionately represented in 2018 Census non-response. However, the data on migrants entering and staying in NZ between 2013 and 2018 tends to be of high quality and will have filled in the gaps for recent migrants. 2013 Census data will have filled gaps for many of the migrants who arrived before 2013.

Large shares of the imputed responses for small ethnic groups end up in the ‘no information’ category for birthplace variable. This is appropriate because Stats NZ has chosen not to impute birthplace. As a consequence, it is not possible to assign a birthplace to a person who did not provide data at the 2018 census and who cannot be located in admin data. Care is therefore needed when using the birthplace variable for small ethnic groups. This matter is reviewed further in the panel’s final report.

2.3.4. Consistency

Was a consistent classification used?

Yes. There have been very minor changes to the classification of birthplace from the 2013 Census. The changes relate to changes in official country names and include:
- Burma (Myanmar) has been changed to Myanmar
- Cape Verde has been changed to Cabo Verde
- South Sudan has been added as a new country.

The birthplace variable has a 3-tiered classification (1 digit, 2 digit, 4 digit) with 276 countries of birth and residual categories at the most detailed level, including 8 residual categories. Countries are grouped according to geographic proximity. They are then grouped into progressively broader geographic areas based on their similarity in terms of social, cultural, economic, and political characteristics.

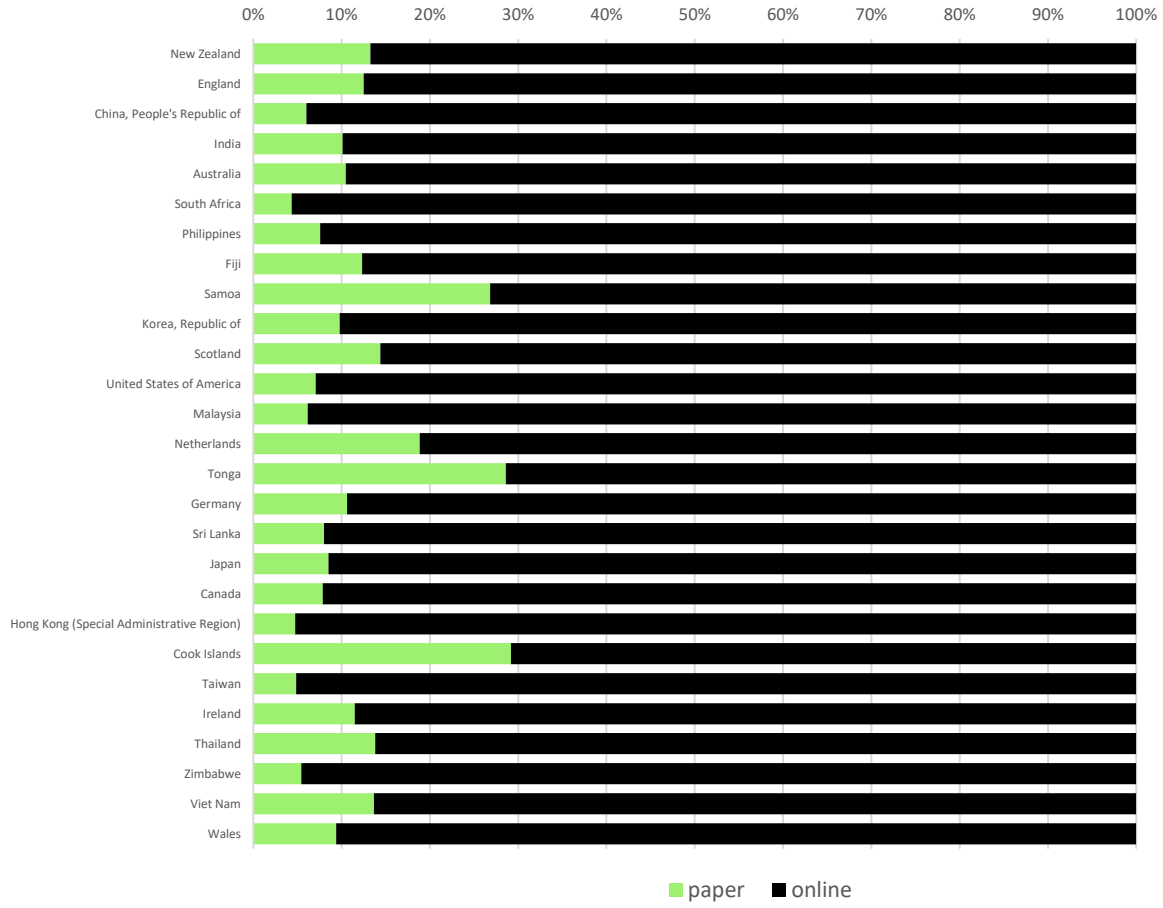
The top level (level 1, major country groups) are: Supplementary Codes, Oceania and Antarctica, North-West Europe, Southern and Eastern Europe, North Africa and the Middle East, South-East Asia, North-East Asia, Southern and Central Asia, The Americas, and Sub-Saharan Africa.

Was data collection consistent across online and paper data collection methods?

The online form had significantly reduced non-response rates to the birthplace question, likely through the use of an 'as you type list' which enabled respondents to start typing and prompted them with likely choices – thereby improving the quality of the data captured.

The chart below shows the proportion of responses on paper (green) and online (black) for the top 27 countries in level 4 of the classification. There are clear differences between online and paper forms for different countries of birth (e.g. Cook Islands 29.2 percent paper compared to Australia 10.5 percent).

Figure 2.3.1: Percentage of Mode by Birthplace (Level 4) – Top 27 countries



2.3.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

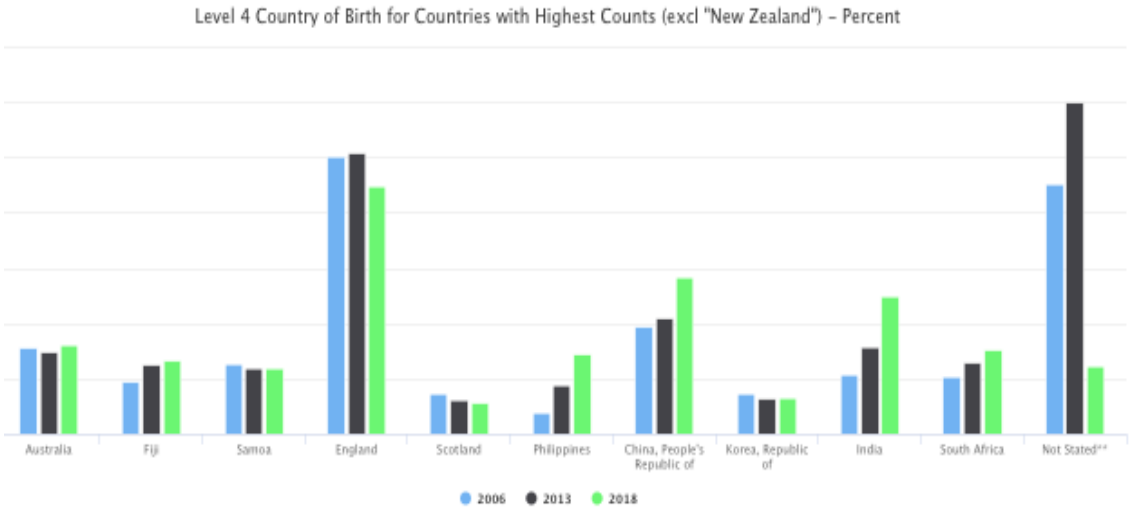
Comparisons were made against expectations based on subject matter expert knowledge, combined with External Migration reports.

Expectations were estimated at the 2-digit level of the classification, but with Permanent and Long-Term Arrivals data, by country of last permanent residence, used to supplement some categories at the 4-digit level.

Is there a consistent time series with previous census, particularly census 2013?
Comparisons have been carried out against 2006 and 2013 Censuses, including at the lowest level of the classification - see chart below, for the largest non-New Zealand countries of birth. The distribution of birthplace appears consistent with previous census results plus known migration in the last five years.

Stats NZ note that there was a fall in the 'England' categories which was more than offset by a sizable rise in the 'United Kingdom (not further defined)'.

Figure 2.3.2: Birthplace (Level 4) with highest counts (excluding New Zealand)



2.3.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?
Yes. Birthplace is not an attribute that changes during a person's life (even if the name of the birth country may change). Therefore use of 2013 Census and admin data (from birth certificates, visa applications etc) will provide information that still applies as at the census date.

2.4. Hours worked in employment per week

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/4fcf1d7b-e8e5-4f03-ac00-c942d95adcf5

EDQ Panel rating: **Moderate/Poor**

Stats NZ rating: **Moderate**

2.4.1. Overall assessment

This variable has high consistency with previous censuses and also expected distributions by gender, occupation, and labour force status. Some edits used in 2018 may have served to reduce the average number of hours worked (compared to 2013, see below), but it can be argued that these edits improved the overall quality of the variable. The overall rating of 'moderate' results from the high level of imputation (18.7 percent overall). For some groups with very high levels of imputation for "hours worked in employment per week", e.g. Pacific peoples (39.6 percent), coverage quality should be considered **poor**.

2.4.2. Background

The 'Hours worked in employment per week' variable is collected for all those 15 years or older who work at least one per week in paid employment, or in unpaid employment if the work "contributed directly to the operation of a business, farm, or professional practice operated by a relative". It is the sum of all hours worked across all jobs worked by the individual, and is the sum of two census questions: 'hours worked in main job', and 'hours worked in all other jobs for pay, profit or income'.

Overall, 81.3 percent of responses come from the Census and the remainder (18.7 percent) from CANCEIS (nearest neighbour) imputation (Table 2.4).

Table 2.4. Data sources: Hours worked in employment per week – Employed census usually resident population aged 15 years and over

Source	Percent
Response from 2018 Census	81.3
2013 Census data	0.0
Administrative data	0.0
Statistical imputation	18.7
No information	0.0
Total	100

2.4.3. Coverage

Was there good coverage for the overall population?

There were no 'response unidentifiable' or 'not stated' responses in the 2018 Census, compared to 4.5 percent in 2013. Overall, 18.7 percent of responses were sourced from a 'nearest neighbour' imputed response (i.e., not a direct response from a census respondent, but a response from another census respondent with characteristics similar to them). The

'data sources and coverage' quality rating for hours worked in employment per week is 0.94 and so is rated as **moderate** (i.e., in the range 0.90-<0.95, see Appendix 1).

Was there good coverage for ethnic groups and regions?
The level of imputation for 'Hours worked in employment per week' was far greater for Māori respondents (32.5 percent) and Pacific respondents (39.6 percent), compared to the overall population. As such, the 'data sources and coverage' quality rating for Māori is 0.90, so within the **moderate** range. The 'data sources and coverage' quality rating for Pacific is 0.88 and so should be considered **poor** (i.e., in the range 0.75-<0.90). The level of imputation was similar across regions; Gisborne had the greatest amount of imputation at 26 percent (which corresponds to a **moderate** 'data sources and coverage' quality rating of 0.92).

2.4.4. Consistency

Was a consistent classification used?
The classification used – Hours worked in employment per week - standard classification (3 numeric) V2.0.0 – is the same standard that was used in 2013 and 2006, and is the same for both census responses and imputed responses. It codes the numbers of hours worked to the nearest hour from 0 hours to 168 hours. Note, a small, but surprising, number of individuals report that they work 168 hours – every hour of the week – or close to this. This seeming anomaly is consistent with previous censuses -- a small number of respondents reported working 168 hours a week in both the 2013 and 2006 censuses.

After imputation, there are no individuals classified as 'response unidentifiable' or 'not stated'.

Was data collection consistent across online and paper data collection methods?
Yes – the same questions were used. However, the online collection ensured individuals who "indicated that they worked for pay or profit in the last 7 days, or that they usually work but were not working in the last 7 days … were routed to the hours worked question." Those completing the census on paper who completed this question but were not in work (and so not in the subject population) were filtered out of the final variable. Imputation was slightly more likely with paper responses (14.9 percent) than online responses (1.7 percent).

2.4.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
An increase in those working part-time is consistent with the 2018 Household Labour Force survey.

Is there a consistent time series with previous census, particularly census 2013?
Yes. Stats NZ state, "Checks against 2013 and 2006 data showed only minor inconsistencies with the 2018 data at the territorial authority level, mainly for 'hours worked in other jobs' above 50 hours a week." Note that the average number of hours worked is about 1 hour less in 2018 (37.1) than 2013 (38.1), which may be due to real world change, but may also be partly due to edits applied to 2018 data, where "hours worked in other job" was set to zero <u>if</u> it was equal to the "hours worked in main job" <u>and</u> was >=40.

2.4.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Yes, all data were obtained from census 2018 responses, either directly or through CANCEIS (nearest neighbour) imputation.

2.5. Industry

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/f51950b4-1ae3-4a4a-9aa8-64936c0a45a1/

EDQ Panel rating: **High**

Stats NZ rating: **High**

2.5.1. Overall assessment

2018 Census industry data is more complete than that available from past population censuses. As with income, the benefits of higher coverage are offset by the cost in increased variability in measures of intercensal change.

Where industry information is used to measure differences between population groups in 2018, data quality will most likely have improved compared to previous censuses. However, when measuring change between censuses the changes to methods for 2018 will have introduced some breaks in the series, especially for data at lower levels of the industry classification.

It is likely that the use of auto-coding in the online 2018 Census form plus the use of admin data will have produced higher quality results, but it has clearly introduced some breaks in the time series – especially at lower levels of the classification.

The distribution of industries at the higher levels of aggregation in the classification in the 2018 Census is similar to that of the 2013 Census. There are a few industries at the lower levels of the classification where the number of people has increased over 100 percent compared to 2013 Census. These changes could be due to changes in methods. Industry knowledge is needed to assess how much these will be due to changes in methods of enhancing the census database compared with structural change in the industry.

It is at these lower levels of the classification that the data are likely to be of much more variable quality and thus a quality category moderate seems more appropriate at this level. When using industry data from the 2018 Census, any assessment of quality will need to be based on the purpose for which the data are being used.

Stats NZ state that “Data has been assessed at the national and regional council level of geography and at level 1 of the classification. Some variation is possible at geographies and classifications below this level. The inclusion of admin data and statistical imputation means there is no non-response category for 2018. Care should therefore be taken if comparing absolute figures to previous years. We recommend using proportions.”

2.5.2. Background

The census data on industry relates to the industry for the main job held by an individual. This is the job in which a person worked the most hours.

Industry data is derived from ‘name of business/employer’, ‘main activity of the business/employer’ and ‘address of the place where worked’ on the individual form. The information provided in these questions is designed to allow identification of the

respondent's employer and the industry recorded for that employer in the Stats NZ Business Register. This has been the practice for several past censuses.

Table 2.5 shows the various data sources used for this variable.

Table 2.5. Data sources: Industry
– Employed census usually resident population aged 15 years and over.

Source	Percent
Response from 2018 Census	71.6
2013 Census data	0.0
Administrative data	20.8
Statistical imputation	7.7
No information	0.0
Total	100

Administrative data or statistical imputation were used not only where there was no census response, but also where there was item non-response or the response from a census responder could not be matched to an employer to obtain the industry classification. There is, therefore, no missing information for 2018 Industry data. In the 2013 and 2006 censuses, responses that could not be classified or did not provide the information asked for were grouped with 'not stated' and classified as 'not elsewhere included'. This applied to 4.0 percent of responses in 2013 and 5.6 percent in 2006.

Use of admin data

Industry always comes from the Business Register, as this holds the definitive classification for every Enterprise.

The way tax is collected in NZ means there is a direct link between a person's IRD number and their employer in the same tax filing. The Employer Monthly Schedule (EMS) is a monthly list of all employees required from all employers – mainly wage and salary earners, and the IR3 form is supplied annually for the self-employed.

The tax information in the IDI provides a link from an individual to their employer's Enterprise and then to the information about the Enterprise held on the Business Register. This linkage was developed in the mid 2000s for Linked Employer-Employee Data (LEED).

Industry in the 2018 Census was determined primarily in two main ways:

- Where an individual responded to a census questionnaire then the business identified was matched to the business register and the appropriate industry assigned.
- Where an individual had not answered the industry questions; they had answered, but a match to the business from the information provided was not possible; or there was no census return (and thus an admin record had been created for that person) then their

personal Inland Revenue (IR) number provides the link to their employer in the IR admin data. The employer information provides the link to the business register to assign an industry.

2.5.3. Coverage

Was there good coverage for the overall population?

Yes. The use of admin data from the tax system will be a high quality source. Imputation was mostly based on 2018 census returns and is acceptable compared to the level of item non-response in previous censuses

The 'data sources and coverage' quality rating for Industry is High (0.95 to 0.98). This is driven particularly by a quality rating of 1.00 (i.e. of the same quality as the census) for the 21 percent of records sourced from admin data.

The admin data quality rating of 1 for Industry and Sector is because there is little room for error. IR number is part of a person's IDI spine record. There is no linkage error between person and employer, as they are from the same IR forms. The use of this admin data results in the same business register values as used by 'census responses'.

The only issues are getting the correct Enterprise:

- If someone has two jobs Stats NZ take the one with the highest earnings (a reasonable approach)
- If a 2018 IR3 record is not available for the self-employed Stats NZ sometimes have to use a 2017 IR3. That is likely to be in the same Industry as for 2018, so this is also a reasonable approach.

Was there good coverage for ethnic groups and regions?

Yes. The completeness of the census and tax data combined will be mirrored in the census final counts for industry.

The admin data will be just as good for all ethnic groups, including Māori and Pacific. The 7.7 percent imputation aims to reduce bias for all ethnic groups, and data could be expected to be of comparable quality to 2013, which is affected by potential bias due to 4 percent missing data.

The quality of the ethnicity data by industry will be in line with the overall issues around ethnicity (see initial and final reports for more detail).

In these contexts much of the data at lower levels in the industry classification are likely to be of **moderate** rather than high quality. However, we have retained a 'high' quality assessment for this variable because of the opportunities to use reliable, timely admin data to fill gaps in the census database.

2.5.4. Consistency

Was a consistent classification used?

There are no conceptual and classification changes to the variable. The classification of industry in the 2018 Census is consistent with the classification used in the 2013 and 2006

Censuses, although the data was previously dual coded to the previous 1996 ANZSIC classification.

The Australian and New Zealand Industrial Classification 2006 (ANZSIC06) V1.0.0 Industry is a hierarchical classification with four levels. Level one (division) contains 20 categories. Level 2 contains 87 Subdivisions. Level 3 contains 219 Groups. Level 4 contains 511 Classes.

Was data collection consistent across online and paper data collection methods?
Data from the online forms is likely to be of higher overall quality than data from paper forms. There were no differences between the wording or question format in the online and paper versions of these questions.

The online form had routing which directed respondents who were usually resident and employed on census night to the industry input questions. Respondents were then provided with an 'as-you-type' list for the 'main activity of business' question. On the paper form it was possible for unemployed people or those not in the labour force to respond to the industry questions. These instances were resolved using edits.

Although most 2018 Census returns were online, there was much higher use of admin data and imputation for responses on paper than online

2.5.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
2018 Census industry data are consistent with alternative sources such as the Quarterly Employment Survey (QES) and Household Labour Force Survey (HLFS) across nearly all consistency checks, at the national and regional council levels of geography, and at most levels in the classification.

Is there a consistent time series with previous census, particularly census 2013?
The distribution of industries in the 2018 Census is similar to that of the 2013 Census. There are a few industries at the lower levels of the classification where the number of people increased over 100 percent compared to 2013 Census. The increases in these industries are likely to be due to use of auto coding (giving better classification), and/or from the use of admin data for missing people and to replace missing or residual categories.

For example, among 20,100 people in Labour Supply Services industries (which was three times the 2013 figure), more than 11,000 are from admin data. These results may be of higher quality in 2018 but will have introduced breaks in the time series at these levels. Comparisons at detailed breakdowns with previous censuses should therefore be made with care.

It is likely that the use of admin data in the 2018 Census will have produced higher quality results but may have introduced some breaks in the time series.

2.5.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Mostly. The tax admin data was used from the September 2018 refresh. EMS filings between January and March 2018 were used, and IR3 data from the March 2018 and March 2017 years.

2.6. Languages spoken

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/adaf4f53-c4ea-408c-aaf1-f32614dd9845/

EDQ Panel rating: **Very high to Poor, depending on language**

Stats NZ rating: **High**

2.6.1. Overall assessment

Stats NZ note that "census is a key source of information on languages spoken for small areas and small populations. Many other sources do not provide detail at this level."

There are three official languages in New Zealand: - English, Māori and New Zealand Sign Language. The analysis of data sources (and the ability to calculate quality ratings below the national level) in the WoF is at level 1 of the Classification, where only New Zealand sign language is identifiable – this has a quality rating of **high**. Stats NZ has provided the panel with a metric 1 quality rating for English, of **high**, and of te reo Māori, which is rated as of **poor** quality, which reflects a high degree of variability for te reo Māori responses between 2013 and 2018. No other language below Level 1 has been rated.

Whilst the assessment of language spoken at high levels of geography, and across all languages may justify a quality rating of high, this ignores the fact that a critical use of this data is to support Treaty of Waitangi obligations around te reo Māori, as well as for planning purposes. The Stats NZ assessment of quality was calculated at the most detailed level (level 4) but was not included in the WoF.

It is clear that the quality of language data varies by language, and that the appropriate **quality ratings need to be based on the quality of level 4 of the classification**. These quality ratings range from **very high** to **poor** – it is the panels view that this range best represents the quality of this dataset and that an overall quality rating at level 1 of the classification does not make sense.

The panel recommends that **Stats NZ should** calculate a quality rating (at least for metric 1) for every language at level 4 of the classification – which is the level at which this data is most likely to be used.

There are major equity issues for this variable, in relation to both the quality of the dataset itself and in terms of how quality assessment was undertaken for the WoF. 18.2 percent of data in the 'Central-Eastern-Malayo-Polynesian' language category (that includes Māori and many Pacific languages) is from historic (2013 Census) data and 16.3 percent from imputed data, with only 65.5 percent from 2018 Census individual responses (compared to 83.8 percent overall from 2018 Census data).

Assessment at the national level provides information that will more closely align with the numerically-dominant population group (i.e. NZ European). Language information is more likely to be important for language categories (and ethnic groups) that are non-dominant, or marginalised. These are at level 4 of the classification, which has only been assessed at the national level, and where only limited quality assessment has been provided.

2.6.2. Background

Information on languages spoken is used to formulate, target and monitor policies and programmes to revitalise the Māori language as an official language of New Zealand and as an indicator of iwi vitality and cultural resources. It is also used to assess the need to provide multi-lingual pamphlets and translation services in a variety education, health and welfare and to evaluate and monitor existing language education programmes and services. The census languages spoken data provides information for television and radio programmes and services and to understand an aspect of the diversity of the New Zealand population over time.

While te reo Māori is not one of the content areas that is legislatively required to be collected in the census, information on changes in te reo Māori over time are critical to the work of other government agencies, in relation to the Māori Language Act 2016, and as an official language.

Te Ture mō Te Reo Māori 2016 | Māori Language Act 2016, recognises the status of the Māori language and aims to "provide means to support and revitalise the Māori language" (Section 3(2)). As part of this, the Act notes that it "provides for Te Mātāwai and the Crown to develop Māori language strategies to support the revitalisation of the Māori language, including by promoting an increase in the number of people speaking the Māori language and improving their fluency in that language" (Section 3 (3, ii)).

The Census 2018 is an important source of data for monitoring Crown obligations under the Māori Language Act 2016. Other significant sources of data on this variable, namely Te Kupenga 2018, are likely to be impacted by the non-response issues of the 2018 Census amongst Māori (as it is a post-censal survey).

Other surveys, such as the GSS, have some language questions but the small Māori sample size limits the ability to do detailed analysis.

The table below shows the breakdown of the various data sources used for this variable.

Table 2.6.
2018 languages spoken – census usually resident population

Source	Percent
Response from 2018 Census	83.8
2013 Census data	8.2
Administrative data	0.0
Statistical imputation	8.0
No information	0.0
Total	100

The proportion of data from 2018 Census data, 2013 Census data and imputation varies by specific population groups.

Due to the use of 2013 Census data and imputation there was no missing data for 2018. For 2013 and 2006 the 'not elsewhere included' were 6.5 percent and 5.1 percent respectively.

2.6.3. Coverage

Was there good coverage for the overall population?

Yes, across all languages, there was good coverage – but this masks significant variations in the data sources for specific sub-groups of the population.

Was there good coverage for ethnic and regional subgroups?

Stats NZ note that "the language classification with the highest imputation rate (CANCEIS and probabilistic) is the Central Eastern Malayo Polynesian language group. This category includes Te Reo Māori and most of the languages spoken in the Pacific." 18.2 percent of data in the 'Central-Eastern-Malayo-Polynesian' language category (that includes Māori and many Pacific languages) is from historic data and 16.3 percent from imputed data, with only 65.5 percent from 2018 Census individual responses (compared to 83.8 percent overall) for 'Central-Eastern Malayo-Polynesian', only 65.5 percent come from 2018 Census responses, with the rest from historic or administrative data.

Based on the national data source rating scores, at level 1 of the classification all language groups are rated as High, apart from Central-Eastern Malayo-Polynesian (which includes te reo Māori) and Indo-Aryan which are rated as Moderate, and the following which are rated as Very High: Greek, Uralic, and 'artificial languages' (but there are only around 100 responses to this group).

Coverage has been checked to Level 4 of the classification only at the national level, but not reported in detail in the WOF. However, Stats NZ has calculated a quality rating for te reo Māori (0.86) which results in a metric 1 quality rating of poor.

The metric 1 quality rating is influenced by the consistency of 2018 responses with 2013 census responses. The percent of those who answered Yes to te reo Māori in 2018 and also answered Yes in 2013, is low at 56 percent. Almost equal numbers of people have replied Yes in 2018 and No in 2013 as those who have said the opposite (No in 2018 and Yes in 2013). This level of disagreement within a five-year period is unlikely to all be due to a genuine change in proficiency. There is almost no change at the aggregate level of the number of Māori speakers for those who responded in both censuses.

The WOF does not provide information on the data sources for this variable by ethnicity but based on differential coverage rates in the Census overall, it is likely that there is higher missing data for Māori and Pacific populations. Stats NZ is calculating quality rating scores for smaller language groups excluding the impact of the official languages (English, te reo and NZ sign language). The metric 1 quality rating score for these languages is 0.93, which equates to a quality rating of Moderate.

There is some information on variability of coverage by region, which shows that some regions had higher collection rates than others. As for many other variables, Gisborne (77.2 percent) and Northland (78.6 percent) regions had the lowest proportions from 2018 Census responses. Calculation of metric 1 quality ratings across all languages by region shows that each region is rated as high quality, except for Southland, which is rated as Very High.

The following charts show Māori language speakers by five year age groups – counts and percentages. They illustrate a significant increase in the recorded number of Māori speakers in 2018 at every five year age group compared to previous censuses. This is unlikely to be a genuine increase in Māori speakers but is likely due to the use of admin records to create 2018 Census records for people in Northland, Gisborne etc (the areas with the lowest response rates in 2013) and the use of imputation to assign te reo Māori language status.

Figure 2.6.1. Count of Māori language speakers in the total population by 5 year age groups

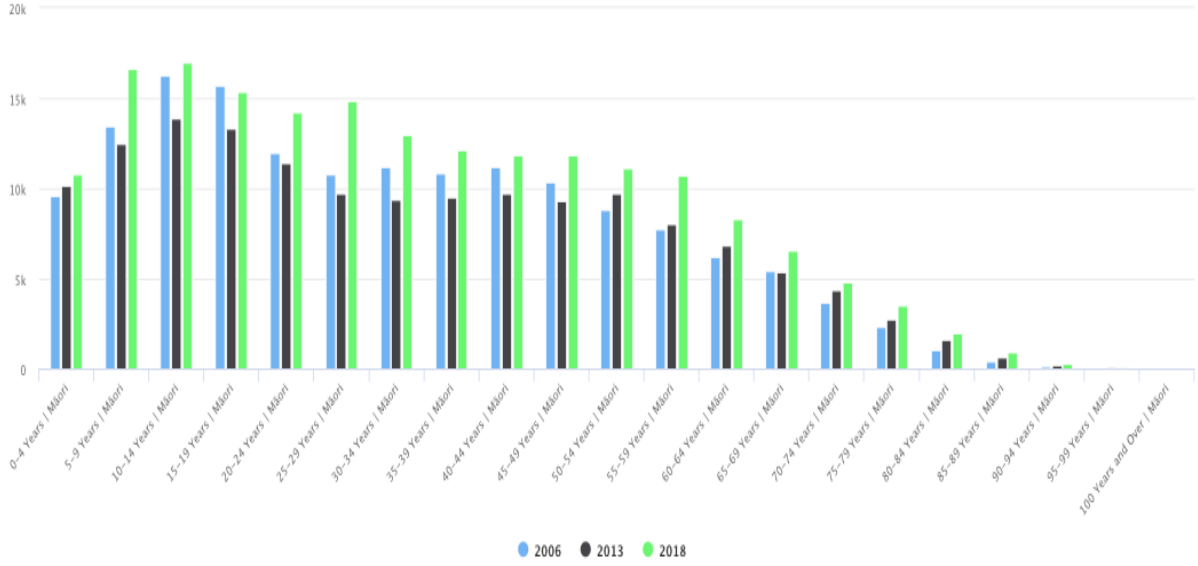
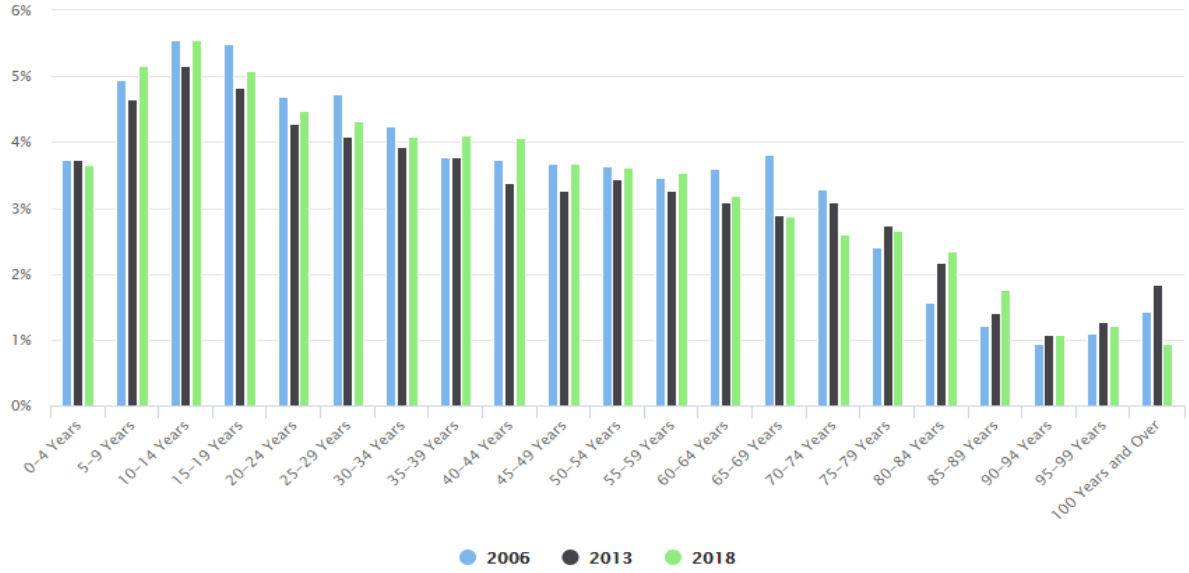


Figure 2.6.2. Percent of Māori language speakers in the total population by 5 year age groups



2.6.4. Consistency

Was a consistent standard used?

The classification of languages spoken in the 2018 Census is consistent with that of both the 2006 Census and the 2013 Census.

Languages spoken is a hierarchical classification with four levels. Level 1 has 26 categories. Level 2 contains 30 categories. Level 3 contains 49 categories. Level 4 contains 196 categories.

Specific languages such as Italian, Japanese, English, and te reo Maori are at level 4 of the classification. Up to 6 languages can be selected across all levels in a valid response.

Was data collection consistent across online and paper data collection methods?

There were no differences between the wording or question format in the online and paper versions of this question. However, there were differences in the way a person could respond.

The question was not mandatory to complete online. The online form had an 'As-you-type' functionality, which may have reduced 'unidentifiable' responses and increased detail/accuracy for online responses. This may also contribute to differential quality between online and paper forms. On the paper forms invalid responses were possible. Alternative data sources were used to replace these responses.

Coverage from census responses appears better for the online population relative to the offline population.

2.6.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

There are not good external benchmarks available for this variable. The General Social Survey (GSS) has some language questions but the small sample size limits the ability to do detailed analysis or draw strong comparisons.

The WoF notes that the "expectation report for Languages Spoken did not have quantitative expectations."

Is there a consistent time series with previous census, particularly census 2013?

Stats NZ state that "at level 1 of the classification, this data is highly comparable with the 2013 and 2006 Census data." Issues with time series have also been noted in relation to changes in the way in which residuals (non-responses) are dealt with in this dataset relative to the Census 2013.

Stats NZ checked level 4 of the classification at a national level. There were some inconsistencies with time series at this level – however this detail is not provided within the WoF – and this is the level at which individual languages, such as te reo are captured.

This will be a more significant issues when comparing with previous censuses for some ethnic groups, and at lower levels of the language classification. Time series comparisons are likely to be impacted for those languages and populations for whom a significant proportion of data are drawn from historic data, i.e. the 2013 Census.

The WoF notes that "some of the categories have inconsistencies in their timeseries. This is due to the change in collection methodology and imputation. An example is the 'Hindi' and 'Fiji Hindi' languages, however, combining these two categories can create continuity in 2018 in the timeseries." In such circumstances it will be important for users of detailed language data to carefully select groupings of languages for comparisons over time.

2.6.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Mostly. 83.8 percent was sourced from 2018 Census returns, with a further 8.0 percent from imputations based on 2018 Census returns. However 8.2 percent were sourced from 2013 Census data and will be out of date to the extent that languages spoken might have changed since then. Consistency between the 2013 and 2018 census is also affected by the variability in responses for the same level of proficiency. While the use of 2013 census language data will to some extent miss genuine change in language ability, in aggregate the impact may be relatively small. The proportion from 2013 Census is higher for some populations and some language groups - e.g. for the Central-Eastern-Malayo-Polynesian grouping (which includes Te Reo), 18.2 percent of the data is from the 2013 census.

2.7. Main means of travel to education

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/32f5d501-695b-4bc7-9812-3c4ba784ed65/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

2.7.1. Overall assessment

Overall the data are of **moderate** quality. However, greater use of imputation among Māori (27.9 percent) and Pacific peoples (32.6 percent) results in data of **poor** quality for these ethnic groups, while data for European peoples can be considered **high** quality.

Main means of travel to education is a new question for 2018 so, unlike main means of travel to work, no comparison with 2013 Census results is possible. This means the data are not affected by issues of quality arising from comparisons of the two census data sets unlike the situation relating to main means of travel to work. Stats NZ have been able to make comparisons with data from the Ministry of Transport Household Travel Survey and state that the results are "reasonably consistent."

Main means of travel to work is one of the ten variables that Stats NZ calculated imputation match rates for using 2013 Census data. The match rate was 0.44. If this quality rating weight for travel to work were applied to travel to education the rating for metric 1 (Data sources and coverage) remains at moderate. This is because there is a higher proportion of data from the 2018 Census for travel to education than there is for travel to work (see section 2.7).

Stats NZ have carried out quality assurance checks down to TA level, so there may be issues at lower levels of geography. Stats NZ state "There are some inconsistencies at lower geographical levels when this variable is checked or compared with other variables, e.g. Age (for a transport type unlikely to be used by an age group). Also, there are small counts where the cross-tab of Usual Residence and Educational Address at TA level for travel to education shows some inconsistencies, e.g. people riding a bike or walking from a usual residence TA to the Educational Address TA which are a long distance apart."

2.7.2. Background

Travel to education data is expected to be used extensively by transport planners to: plan and manage transport for schools, both in country areas for planning school transport needs, as well as in large cities where congestion is a problem; planning new schools and infrastructure around schools; monitoring investment in certain travel modes, such as investment to support walking and cycling; targeting initiatives aimed at encouraging more children and tertiary students to use public transport, and developing fare structures to promote public transport use amongst tertiary students.

2018 Census was the first time that the main means of travel to education question was asked in the census. Main means of travel to education is the usual method a person used to travel the longest distance to their place of education (for example, by bicycle, school or public bus, walking, or driving). Respondents were asked for the main way that they usually

travel to education and, if they didn't have a usual method, to select the one used most recently.

Table 2.7 shows that only two data sources were used for this variable – 2018 Census returns and imputation.

Table 2.7. Data sources: Main means of travel to education – Census usually resident population in study

Source	Percent
Response from 2018 Census	84.5
2013 Census data	0.0
Administrative data	0.0
Statistical imputation	15.5
No information	0.0
Total	100

Of those with data from a 2018 Census individual form, the initial item non-response rate was 2.2 percent, but these gaps were filled with imputed responses.

Whilst overall 84.5 percent of data came from the 2018 Census and 15.5 percent was imputed, these rates varied significantly by ethnicity.

2.7.3. Coverage

Was there good coverage for the overall population?

84.5 percent of the records were sourced from 2018 Census; imputation was used to fill in the gaps (15.5 percent).

Was there good coverage for ethnic groups and regions?

No. Overall 84.5 percent of data came from the 2018 Census and 15.5 percent was imputed, however these rates varied significantly by ethnicity. The response rate for Europeans was 88.9 percent; Asian 87.0 percent; other 87.8 percent; MELAA 85.7 percent, Māori 72.1 percent, and Pacific 67.4 percent. This gives rates of imputation from 11.1 percent for Europeans, up to 32.6 percent for Pacific ethnicities. These response patterns reflect the overall non-response patterns to the 2018 Census.

The Data sources and coverage (Metric 1) quality ratings for main means of travel to education by ethnicity are therefore **High** for European, other and Asian; **Moderate** for MELAA, and **Poor** for Māori and Pacific ethnicities.

In line with the overall pattern of non-response to the 2018 Census, imputation rates for this variable are highest in Gisborne (26.2 percent) and Northland (23.8 percent), which also reflects the distribution of ethnic groups.

Stats NZ have cross tabulated 'Usual residence TA' and 'Educational address TA' by main means of travel to education. This shows a small number of people taking an unlikely transport mode due to the long distance (e.g. people in the Far North walking, biking or taking the school bus to study in Auckland, or people in the Kaipara District taking the school bus to study in Auckland). Stats NZ state "Some of these issues are due to difficulty in coding to the correct education institution address."

2.7.4. Consistency

Was a consistent classification used?

This is a new question, so it has a new classification, which is flat and contains the following categories: Study at home; Drive a car, truck or van; Passenger in a car, truck or van; Bicycle; Walk or jog; School bus; Public bus; Train; Ferry; Other; Not elsewhere included (including the residual categories of 'response unidentifiable' and 'not stated').

Online this question only appeared to respondents if they said that they attended a place of education full-time or part-time, and only one option could be selected. If a respondent selected more than one box, the first tick response was unselected, so only one response was possible.

On paper everyone could see this question so people that were not in scope to answer it could still write in a response (although these responses would not be included in the final dataset). A respondent could tick more than one main way that they travelled to education. These multiple responses were then fixed by an edit.

Was data collection consistent across online and paper data collection methods?

Online respondents could only select one valid response, while on paper it was possible to answer multiple times, which then required editing. The imputation rates were much higher for responses on paper (27.0 percent) than online (4.6 percent).

2.7.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

Stats NZ have compared results to the Ministry of Transport Household Travel Survey and state "while this is a small sample survey the categories of Passenger, Driving and Walking by age group at the national level and for Auckland, Wellington and Christchurch were reasonably consistent."

Is there a consistent time series with previous census, particularly census 2013?

Not relevant – this is a new variable.

2.7.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Yes, all data were sourced from the census 2018 individual forms, either directly or through imputation.

2.8. Main means of travel to work

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/ab342389-7db7-46f1-b3d5-ae01504b93c1/

EDQ Panel rating: **Poor**

Stats NZ rating: **Moderate**

2.8.1. Overall assessment

Main means of travel to work is one of the ten variables that Stats NZ calculated imputation match rates for using 2013 Census data (by randomly deleting records in the actual dataset, imputing the match rates for these records, and then comparing the imputed results to the actual 2013 records). This gave a match rate of 44 percent (i.e., the imputed result matched the actual results in 44 percent of cases). Putting this value in to the data quality rating calculations (instead of the 50 percent used) gives a data sources and coverage rating of **poor**, and thus the panel rates this variable as poor overall.

In the 2013 Census, respondents were asked about the one main way travelled to work on census day; for the 2018 Census, respondents were asked about the one main way they usually travel to work, a different concept. Other things being equal, the 207,000 (10.4 percent) responses to 'Did Not Go To Work Today' in 2013 will have been reallocated to other categories in 2018. While the change should result in a more accurate 2018 measure of usual means of travel to work, it is at the expense of time series comparisons. Because of the change in concept, the panel advises that compare 2018 results to 2013 results with care and caution.

There are large changes in the levels recorded in 2018 compared to previous censuses which highlights the impact of the question change. The largest category (Drove a Private Car, Truck or Van) increased 45.4 percent between 2013 and 2018 (compared to a 2.1 percent increase 2006 to 13). It represented 57.8 percent of the means of traveling to work in 2018, compared to 48.6 percent in 2013.

Stats NZ have produced tables of 2013 Census main means of travel to work data which exclude those who did not travel to work on census day in 2013. The DataInfo+ page for 'Main means of travel to work' states that "we recommend using proportions rather than counts for timeseries comparisons and for earlier censuses only including respondents who travelled to work on census day" The use of travel to work data from the 2013 Census (and earlier) which excludes those who did not travel to work on census day will be more comparable to the data for 2018. Stats NZ should carry out further analysis of the impact of the change in the question.

Stats NZ have carried out quality assurance checks down to Territorial Authority/ Auckland Local Board level using data from previous census years. There may be issues at lower levels of geography.

Some of the changes between 2013 and 2018 will be due to real world effects, but a non-trivial proportion are clearly due to the changes in the question. The panel recommend that **Stats NZ should** carry out further analysis of the impact of the change in the question.

Added to this are other changes to the methods in 2018 leading to a higher rate of completeness of the census population count compared to previous censuses, and the use of imputation to replace the 2.5 percent item non-response in 2013. Overall, the panel do not consider that the 2018 travel to work data is comparable to that from 2013.

2.8.2. Background

Table 2.8 shows that only two data sources were used for this variable – 2018 Census returns and imputation.

Table 2.8. Data sources: Main means of travel to work – Employed census usually resident population aged 15 years and over

Source	Percent
Response from 2018 Census	81.0
2013 Census data	0.0
Administrative data	0.0
Statistical imputation	19.0
No information	0.0
Total	100

In output for previous censuses, responses that could not be classified or did not provide the type of information asked for, such as response unidentifiable, were grouped with 'not stated' and classified as 'not elsewhere included'. For 2013 this was 3.7 percent; for 2006 it was 5.3 percent. In 2018 imputed values were created to replace such categories.

In the 2013 Census, respondents were asked about the one main way they travelled to work on Tuesday 5 March. For the 2018 Census, respondents were asked about the one main way they usually travel to work. In 2013 10.4 percent of responses were 'Did Not Go To Work Today'. In 2018 there was zero percent for this category because it was not relevant given the change in question.

Other things being equal, those 2013 respondents who 'did not go to work today' will have been reallocated to other categories in 2018. While the change should result in a more accurate measure of usual means of travel to work, it is at the expense of time series comparisons. For instance, if the 10.4 percent who did not 'Did Not Go To Work Today' were reallocated evenly across all other valid 2013 responses, this could explain about one-third of the 2013-18 increase for 'Drove a Private Car, Truck or Van'.

2.8.3. Coverage

Was there good coverage for the overall population?

The travel to work variable has no missing information. 81.0 percent of records were derived from the census individual form. The remaining 19.0 percent of data were imputed using the CANCEIS 'nearest neighbour'.

Was there good coverage for ethnic groups and regions?
No. Overall 81.0 percent of data came from the 2018 Census and 19.0 percent was imputed, however these rates varied significantly by ethnicity. The response rate for Europeans was 86.0 percent; other 78.3 percent; Asian 76.9 percent; MELAA 73.8 percent, Māori 67.1 percent, and Pacific 59.7 percent. This gives rates of imputation from 14.0 percent for Europeans, up to 40.3 percent for Pacific ethnicities. These response patterns reflect the overall non-response patterns to the 2018 Census.

The Metric 1 (data sources and coverage) quality ratings for main means of travel to work by L1 ethnicity are **Moderate** for European, and **Poor** for all other ethnicities. These outcomes are the same whether a weight of 0.5 or 0.44 is used.

The percentage of data from the 2018 Census varied by region from 85.3 percent in Wellington and Canterbury regions, down to 76.8 percent in Northlands (23.2 percent imputed) and 73.8 percent in Gisborne region (26.2 percent imputed) – this reflects the distribution of ethnic groups.

While there are large changes between 2013 and 2018 overall, there are some very large changes for some ethnic groups – especially in groups which are most impacted by the use of admin data (such as Māori and Pacific peoples). This has led to a higher overall count in 2018 for these groups compared to 2013. The use of imputation to create records for all responses will also have impacted on measures of change.

Stats NZ advise that "due to the change in question, we recommend using proportions rather than counts for timeseries comparisons and for earlier censuses only including respondents who travelled to work on census day." However, it is informative to understand some of the scale of changes in the data and the errors likely to occur if this were done.

There is a 186.6 percent increase 2013 to 2018 in Māori working from home (190.5 percent for Pacific people) compared to 64.4 percent for Europeans and 71.6 percent overall for this category. In some regions these changes are even larger – e.g. a 231 percent increase in Māori working from home in Bay of Plenty (425 percent for Pacific people). There is an 84.2 percent increase 2013 to 2018 in Māori who 'Drove a private car, truck or van' (99.3 percent for Pacific people) compared to 33.5 percent for Europeans and 45.4 percent overall for this category.

There are some large increases for people of Asian ethnicity, such as 191.2 percent increase 2013 to 2018 in use of train to travel to work, compared with a 98.9 percent increase 2006 to 2013 and an overall 98.0 percent increase in train travel as main means of travel to work.

Using the Stats NZ 0.5 weight for imputation, the quality ratings for main means to travel to work at the Regional Council level gives a result of Moderate Quality for eleven Regional Councils, with five (Auckland, Bay of Plenty, Gisborne, Hawkes Bay, and Northland) being of Poor Quality. If, however, the 2013 calculated imputation match rates are used (i.e. a

weight of 0.44 rather than 0.5), then an additional one Regional Councils (Waikato) shifts from Moderate to Poor quality ratings.

The quality of the data by ethnicity and region highlight the care that will be needed when comparing 2018 results with those from 2013.

2.8.4. Consistency

Was a consistent classification used?

There were minor changes to the classification of this variable un 2018 Census:

- The category 'did not go to work on census day' was not used due to the change in concept to 'usual' means of travel. The category remains in the classification for time series purposes.
- 'Motorbike' was removed because this category was rarely used.
- 'Ferry' was added following consultation.

Main means of travel to work is a flat classification with the following categories: Worked at home; Did not go to work today; Drove a private car, truck or van; Drove a company car, truck or van; Passenger in a car, truck, van or company bus; Public bus; Train; Bicycle; Walk or jogged; Ferry; Other; Not elsewhere included.

Was data collection consistent across online and paper data collection methods?

The level of imputation far greater for paper responses (20.1 percent) than online (7.0 percent).

2.8.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

Unlike main means of travel to education, the main means of travel to work results were not compared with the results from the Ministry of Transport Household Travel Survey. This is a relevant source for comparative data and this comparison should be carried out.

Is there a consistent time series with previous census, particularly census 2013?

No. Stats NZ state "Main means of travel to work in the 2018 Census is moderately comparable with the 2013 Census data." The panel do not believe that the 2018 travel to work data is comparable to that from 2013. The change from asking about main means of travel to work on census day, to usual means of travel to work should result in a more accurate measure of usual means of transport but at the expense of time series comparisons.

Stats NZ state that "we recommend using proportions rather than counts for timeseries comparisons and for earlier censuses only including respondents who travelled to work on census day" The use of travel to work data from the 2013 Census (and earlier) which excludes those who did not travel to work on census day will be more comparable to the data for 2018.

The 15.4 percent of 2013 responses that have zero percent in 2018 (did not travel to work, Motor Cycle or Power Cycle, response unidentifiable and not stated) will have been

reallocated over other categories in 2018 and this will have an important impact on measures of change.

2.8.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Yes, all data were sourced from the census 2018 individual forms, either directly or through imputation.

2.9. Occupation

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/7889e133-a8e8-4c68-8d91-ea11ebc10c2f

EDQ Panel rating: **Poor**

Stats NZ rating: **Moderate**

2.9.1. Overall assessment

The very nature of classifying the job descriptions of people into prearranged categories involves a degree of misclassification, which has been well studied elsewhere. This affects the quality of imputation methods generally. The level of imputation for occupation in 2018 Census is very high, particularly for a variable with such a highly detailed classification.

The level of imputation overall indicates the rating is on the border between moderate and poor. For Māori, Pacific and Asian ethnic groups and for many areas of the country, the level of imputation indicates occupation should be rated as **poor**.

Imputation will generally produce unbiased estimates of numbers in occupations (especially at the upper levels of the classification) but will often assign incorrect occupations to people (e.g., an estimated 40 percent will be correct at the major group), which may impact estimates of association between occupation and other variables. Initial investigations suggest estimates of association between occupation and income are largely maintained, while estimates of association between occupation and qualifications are mostly maintained, but with slightly underestimated associations for Māori and Pacific males and Pacific females.

2.9.2. Background

Data were obtained from census individual responses for 79.7 percent of responses, and from imputation for the remaining 20.3 percent of responses (Table 2.9.1).

Table 2.9.1. Data sources: Occupation – Employed census usually resident population aged 15 years and over

Source	Percent
Response from 2018 Census	79.7 percent
2013 Census data	0.0 percent
Administrative data	0.0 percent
Statistical imputation	20.3 percent
No information	0.0 percent
Total	100 percent

Occupation was classified for those in employment (~2.4m) using Australian and New Zealand Standard Classification of Occupations version 1.2 (ANZSCO1.2), which classifies occupation using a five-level hierarchical system. For example, someone may be classified

as a Labourer as the highest level of the classification ('major group', 9 categories), a Factory Process Workers at the next level ('sub-major group', 44 categories), a Packers and Product Assembler at the 'minor group' level (99 categories), a Packer at the 'unit' level (361 categories), and a Meat Packer at the 'occupation' level (1035 categories). The full classification can be found at Australian and New Zealand Standard Classification of Occupations (ANZSCO V1.2.0).

This is the same classification as the ones used in the 2006 and 2013 Censuses, except the ANZSCO used in 2013 and 2006 had only 1,025 and 1,001 categories at the 'occupation' level.

There were some changes to collection methods that will impact results. These include:

I. online respondents were provided with 'as you type' lists of occupations that allowed them to choose among occupations that match what they have typed. This was done to reduce manual coding (which it did – from 73 percent in 2013 to 20 percent in 2018). While this may enhance accuracy, there may have been a tendency to choose the top occupation shown in a list or chose among displayed occupation as a 'joke' answer. Note, there is no way of knowing the extent to which this happened.
II. no questions were asked about tasks and duties of an occupation (as had been asked in previous censuses). Information about tasks and duties helps with manual coding, so where manual coding was used in 2018 quality may be lower (noting that manual coding was used less frequently in 2018 than in previous censuses, as stated above);
III. use of imputation means there is no non-response (non-response was 2.8 percent in 2013); and this results in increases to the proportions in all but one (of eight) major groups in 2018 compared with 2013 Census.

2.9.3. Coverage

Was there good coverage for the overall population?

The occupation variable is complete for the subject population (i.e., there is no 'no information' responses, after the addition of imputed responses).

79.7 percent of records were derived from the census individual form. The remaining 20.3 percent of data were imputed using the CANCEIS 'nearest neighbour' approach (see description in the introduction of this report).

When tested on a 4 percent sample of 2013 census forms, imputation was shown to preserve the major group distribution very well, but be largely inaccurate at replicating the precise occupation. Only 40.2 percent of imputed occupations matched reported occupation at the least granular ('major group') level, and only 15.9 percent of imputed occupations matched reported occupation at the most granular ('occupation') level.

Further analysis of associations between occupation and income, and occupation and qualifications showed that 2018 census occupation data produced very similar associations to 2013 census occupation data across all major ethnic groups. The only exceptions were that pearson correlations between occupation (using a continuous measure of occupational socio-economic status as a proxy for 'minor group' level occupation) and qualifications

(years of study) were lower in 2018 than in 2013 for Māori males (*r*=0.48 vs *r*=0.56), Pacific males (*r*=0.43 vs *r*=0.54), and Pacific females (*r*=0.49 vs *r*=0.55).

Was there good coverage for ethnic groups and regions?

The degree of imputation varied markedly by ethnicity and also by geography. At the national level, Stats NZ assigns occupation records imputed using CANCEIS a 0.5 quality weight and those derived directly from census forms a 1.0 (see Table 2.8.2). This means that all subgroups and geographies with >20 percent & <50 percent imputation should be rated as of **poor** quality (i.e., overall rating of 0.75 – 0.9 (see Appendix 1). This includes those in a Māori (33.9 percent imputation, quality rating = 0.83), Pacific (41.4 percent imputation, quality rating = 0.80) or Asian (24.7 percent imputation, quality rating = 0.87) ethnic group. It also includes those from around 30 territorial authorities. The three territorial authorities with the greatest proportion of imputation – and lowest quality rating are shown in Table 2.9.2. This means that (e.g.) >40 percent of the occupation data for someone of Pacific ethnicity will not be the occupation they recorded on their census form, but will instead be an imputed occupation from a 'donor' census respondent.

Table 2.9.2. Quality rating calculation table for sources of occupation data by ethnicity and territorial authority
– Employed census usually resident population aged 15 years and over

A. By ethnicity		***Total population***		***Māori***		***Pacific***		***Asian***	
Source	**Rating**	**Percent total**	**Score contribution**	**Percent total**	**Score contribution**	**Percent total**	**Score contribution**	**Percent total**	**Score contribution**
2018 Census form	1.00	79.7	0.80	66.1	0.66	58.6	0.59	75.3	0.75
Imputation									
Donor's 2018 Census form	0.50	20.3	0.10	33.9	0.17	41.4	0.21	24.7	0.12
Total			**0.90**		**0.83**		**0.80**		**0.87**
B. By selected territorial authorities		***Total population***		***Kawarau***		***Opotiki***		***Wairoa***	
Source	**Rating**	**Percent total**	**Score contribution**	**Percent total**	**Score contribution**	**Percent total**	**Score contribution**	**Percent total**	**Score contribution**
2018 Census form	1.00	79.7	0.80	62.8	0.63	65.3	0.65	70.2	0.70
Imputation									
Donor's 2018 Census form	0.50	20.3	0.10	37.2	0.19	34.7	0.17	29.8	0.15
Total			**0.90**		**0.82**		**0.82**		**0.85**

Due to rounding, individual figures may not always sum to the stated total(s) or score contributions.

2.9.4. Consistency

Was a consistent classification used?

The same classification, ANZSCO, has been used for censuses since 2006, and was used both for occupation completed on individual census forms in 2018, and occupation imputed using 'nearest neighbour' census forms (with only very minor changes at the lowest 'occupation' level).

Was data collection consistent across online and paper data collection methods?

Online collection allowed 'as you type' completion of occupation, which resulted in significantly less manual coding than offline collection. This may have resulted in differential quality between the two modes.

2.9.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

Compared to results from the Household Labour Force Survey (20.5 percent), census 2018 results slightly underestimated the 'Managers' major group category (18.0 percent) but was close for other major group categories.

Is there a consistent time series with previous census, particularly census 2013?

Compared to census 2013 data, census 2018 results were consistent across major groups, suggesting a consistent time series. There were inconsistencies for some 'occupation' categories – this may reflect real world change (as specific occupations become more or less common) – or represent difficulties in coding.

2.9.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Yes, all data were sourced from the census 2018 individual forms, either directly or through imputation.

2.10. Qualifications: Highest qualification

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/33e8dc17-1be8-446d-8f33-6f458e86f94c

EDQ Panel rating: **Moderate/Poor**

Stats NZ rating: **Moderate**

2.10.1. Overall assessment

The consistency of the highest qualification variable should allow for time series comparisons at the national level. Alternative data sources (2013 census as well as administrative data) were used in around 12 percent of cases for both input variables to improve coverage for this variable. In the cases of Māori and Pacific peoples this percentage was almost double at around 22 percent while in the Gisborne region it was around 18 percent. Valid responses were available for 93.5 percent of the census usual resident population aged 15 years and over, though only 86.6 percent for the Pacific census usual resident population aged 15 years and over, and around 91 percent for the Māori and Asian census usual resident populations aged 15 years and over. Quality should therefore be considered **poor** for the Māori, Pacific and Asian populations, and for some regions (particularly Gisborne).

For the two component variables that are used to derive the highest qualification variable: *Highest secondary school qualification* and *Post-school qualification level* (see below), quality should also be considered **moderate** overall, but **poor** for the Māori, Pacific and Asian populations, and for some regions.

2.10.2. Background

The highest qualification variable covers all those aged 15 years and older. It is a derived variable constructed by combining two variables: *Highest secondary school qualification* and *Post-school qualification level*.

For *Highest secondary school qualification* data were obtained from 2018 census responses in 82.4 percent of cases, from 2013 census data in 7.7 percent of cases, and from administrative data in 4.0 percent of cases (Table 2.9). Administrative data were obtained from Ministry of Education data on school enrolments and courses. For *Post-school qualification level* data were obtained from 2018 census responses in 80.7 percent of cases, from 2013 census data in 6.5 percent of cases, and from administrative data in 5.9 percent of cases (see Table 2.9). Administrative data were obtained from Ministry of Education data on course completions, Tertiary Education Commission IT learner, targeted training, and student qualifications. Ministry of Education holds authoritative information on formal qualifications gained in New Zealand. The 2013 Census data includes qualifications gained up to 2013, including those with no qualification. The highest qualification from the combination of these two sources is used for the 2018 census data.

There were two main changes to collection methods in 2018 compared to 2013:
(i) In 2018, *Post-school qualification level* had a check list of levels of qualification and a free text 'other qualification' field, whereas in 2013, respondents were asked to write the level of their qualification in a free text box.

(ii) online respondents were provided with 'as you type' lists to help provide valid responses.

Table 2.10. Data sources: Highest qualification – Census usually resident population aged 15 years and over

Source	Highest secondary school qualification Percent	Post school qualification level Percent
Response from 2018 Census	82.4	80.7
2013 Census data	7.7	6.5
Administrative data	4.0	5.9
Statistical imputation	0.0	0.0
No information	5.9	7.0
Total	100	100

2.10.3. Coverage

Was there good coverage for the overall population?

For highest qualification in 2018, 6.5 percent had 'not elsewhere included' responses, compared to 11.1 percent in 2013 (this includes the 'not stated' responses). The reduction in 'not elsewhere included' responses from 2013 to 2018 necessarily results in an increase in other categories.

For the component variables, *Highest secondary school qualification* had 6.3 percent 'not elsewhere included' responses, compared to 9.7 percent in 2013; while *Post-school qualification level* had 7.1 percent 'not elsewhere included' responses, compared to 13.1 percent in 2013. The 'data sources and coverage' quality rating for highest qualification is 0.92 and so in the **moderate** range (i.e., 0.90 - 0.95). The 'data sources and coverage' quality rating for both *Highest secondary school qualification* (0.91) and *Post-school qualification level* (0.90) should also be considered **moderate**.

Was there good coverage for ethnic groups and regions?

'Not elsewhere included' responses are higher for Pacific (13.4 percent), Māori (9.3 percent) and Asian (8.7 percent) peoples (6.5 percent for the overall population). Use of alternative sources to the 2018 census for the *Highest secondary school qualification* and *Post-school qualification level* variables is also greater for Māori (23.5 percent and 22.1 percent, respectively) and Pacific peoples (22.3 percent and 21.2 percent, respectively), compared to 11.7 percent and 12.4 percent, respectively, for the overall population. Thus, 'data sources and coverage' quality rating for highest qualification should be considered **poor** for Pacific, Māori and Asian New Zealanders.

'Not elsewhere included' responses are high in the Auckland (7.8 percent) and Gisborne (7.9 percent) regions. Use of alternative sources to the 2018 census is high in the Gisborne

region (18.0 percent for both *Highest secondary school qualification* and 17.9 percent for *Post-school qualification level*), indicating **poor** 'data sources and coverage' quality for this region.

2.10.4. Consistency

Was a consistent classification used?

The same standard classification – Census highest qualification output V2.0.0 – has been used since 2006, and was used both for data obtained from individual census forms in 2018, and from individual census forms completed in 2013. The highest qualification variable contains 13 categories:

- No Qualification
- Level 1 Certificate
- Level 2 Certificate
- Level 3 Certificate
- Level 4 Certificate
- Level 5 Diploma
- Level 6 Diploma
- Bachelor Degree and Level 7 Qualification
- Post-graduate and Honours Degrees
- Masters Degree
- Doctorate Degree
- Overseas Secondary School Qualification
- Not elsewhere included.

Highest secondary school qualification (V2.0.0) is consistent with the classification used in the 2013 and 2006 Censuses, and has the following categories: No qualification, Level 1 certificate, Level 2 certificate, Level 3 or 4 certificate, Overseas secondary school qualification, and Not elsewhere included.

Post-school qualification level (V1.0.0) is also consistent with the classification used in the 2013 and 2006 Censuses, and has the following categories: No post-school qualification; Level 1, 2 or 3 certificate; Level 4 certificate; Level 5 diploma; Level 6 diploma; Bachelor degree and level 7 qualification; Post-graduate and honours degrees; Masters degree; Doctorate degree; Level not given (but subject given); Not elsewhere included.

Data obtained from Ministry of Education administrative data was coded to these standards. Stats NZ note that "Ministry of Education data sourced from the IDI only states level 1, 2, or 3 certificates without referring to whether this is a secondary qualification or a post-school qualification. Many of these were coded as a post-school qualification. Consequently level 1, 2, and 3 certificates for highest secondary school qualifications may have been understated."

Was data collection consistent across online and paper data collection methods?

Mostly. Both online and paper forms had the same check box categories and free text spaces for both the *highest secondary school* and *post-school qualification level* questions. For free-text responses, online collection allowed 'as you type' completion, while offline

(paper) did not. Highest qualification level was correlated with data collection mode – e.g., 24.6 percent of those with no high school qualification completed responses on paper, compared to only 5.4 percent of those with a Masters degree.

2.10.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

Comparisons against non-census data collections have been undertaken for school qualifications, using the 2018 Household Labour Force Survey (HLFS). The census distribution of school qualifications was mostly consistent with the HLFS, except that the census recorded **a greater proportion of females with 'upper' secondary school qualifications** (29.0 percent vs 23.0 percent), and **a lower proportion of males with no qualifications** (12.4 percent vs 17.0 percent). No comparisons for post-school qualifications have been undertaken (and it is unclear whether the data exist to undertake such comparisons).

Is there a consistent time series with previous census, particularly census 2013?

Mostly. The proportion in each category for 'highest qualification' is as expected given census 2013 results, with the exception that there was a decrease in 'overseas school qualifications', where an increase was expected. Note that administrative data does not record non-New Zealand qualifications, which may serve to undercount 'overseas school qualifications', as well as other qualifications awarded overseas. Stats NZ also acknowledge "there was a large increase in post-graduate diplomas which may be partly explained by a change to the questionnaire format (respondents selecting their highest qualification from a check box list instead of entering it into free text box) and partly explained by real world changes."

For *Post-school qualification level*, there was an increase in level 1, 2, and 3 certificates compared to previous censuses, which may be explained by level 1, 2, or 3 certificates sourced from administrative data being coded as post-school certificates, when some may have been attained at school (see above).

2.10.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

No, for both input variables (highest secondary school qualification and post-school qualification level), slightly more than 80 percent of records were obtained from data collected as part of the census in March 2018; around 7 percent from data collected as part of the census in March 2013, and around 5 percent from administrative data that may have been updated at any time from 1994 until December 2017.

2.11. Religious affiliation

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/472884ac-0f5f-4606-827c-b7153a1a51b1/

EDQ Panel rating: **High**

Stats NZ rating: **High**

2.11.1. Overall assessment

There will have been large real-world changes in religious affiliation between 2013 and 2018 reflecting both continuing long term societal trends (e.g. towards no religion) and also the impact of immigration. Migrants from different countries and ethnic groups tend to bring their religious affiliation with them.

The trends in religion in the 2018 Census appear credible and consistent with previously observed changes. The use of 2013 Census data for 8.2 percent of 2018 records will have biased the affiliation rate to particular religions upwards slightly (due to the ongoing decline in affiliation over time).

Stats NZ state "The data is comparable with previous censuses. Guidance may be needed when comparing data at lower levels of the classification, particularly in terms of increases in responses due to the change in response format or changes to the classification." and "Lower geographies such as TA, SA2 and Auckland Local Boards have not been checked."

2.11.2. Background

The table below shows the breakdown of the various data sources used for this variable. There was no use of admin data.

Table 2.11. Data sources: Religious affiliation – Census usually resident population

Source	Percent
Response from 2018 Census	82.9
2013 Census data	8.2
Administrative data	0.0
Statistical imputation	8.8
No information	0.0
Total	100

In output for previous censuses, responses that could not be classified or did not provide the type of information asked for were grouped with 'not stated' and classified as 'not elsewhere included' (nei). In 2006 the nei rate was 7.3 percent; in 2013 it was 8.2 percent. In 2018, due to the use of 2013 Census data and imputation, there were no 'not stated' or 'not elsewhere counted' cases.

A total of 3.97m completed an individual 2018 Census form, of which 3.9m (98.1 percent) answered the religious affiliation question. Therefore the non-response to this question (for those who completed the 2018 Census form) was 1.9 percent, before the use of imputation.

There was a change to the religious affiliation question, allowing respondents to provide more detail in their response. The online form had an As-You-Type list that provided a list of options as respondents started typing in the response field. Taken together this means that there are higher numbers for some categories at lower levels of the classification.

2.11.3. Coverage

Was there good coverage for the overall population?

82.9 percent of records were derived from the census individual form; 8.2 percent from 2013 Census records, and the remainder from imputation.

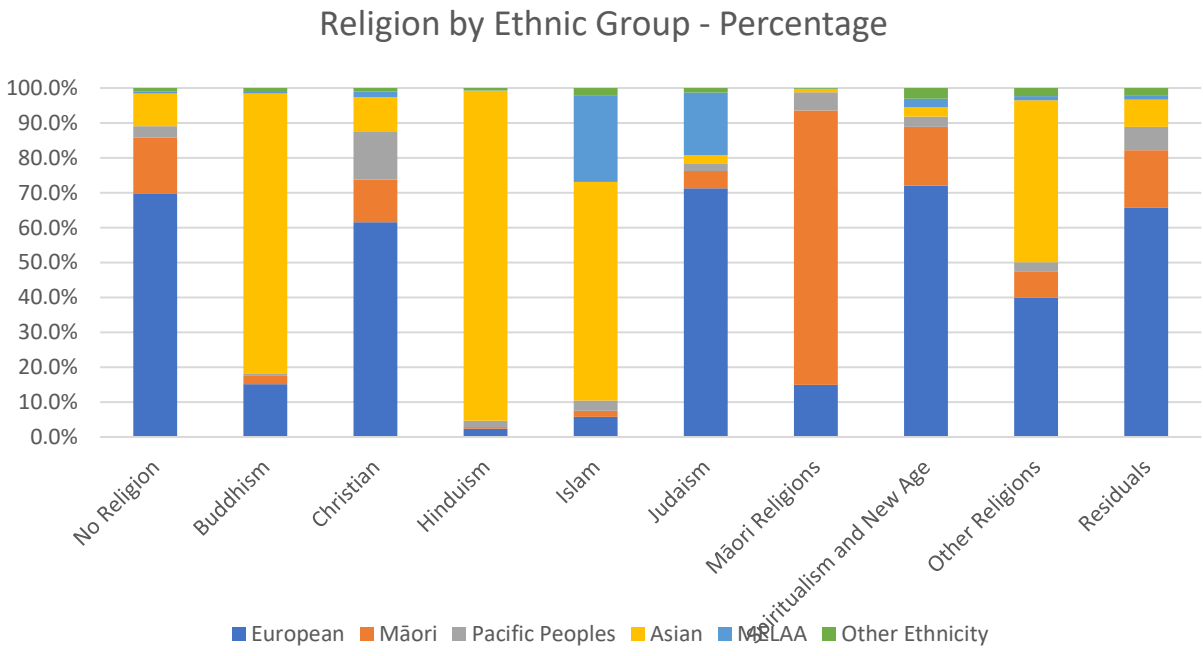
Was there good coverage for ethnic groups and regions?

There will have been large changes observed between 2013 and 2018 reflecting both continuing long term societal trends (e.g. towards no religion) and also the impact of immigration. Migrants from different countries and ethnic groups tend to bring their religious affiliation with them.

Apart from 'other ethnicity', for every major Level 1 ethnic group the analyses by religion (level 1) shows an increase in the percentage of 'No religion', with corresponding reductions elsewhere. The religious affiliation by ethnicity results appear credible:

- No Religion, Christian, Judaism, Spiritualism/New Age, and the Residual Categories are made up primarily of people of European ethnicity
- Buddhism, Hinduism, Islam are made up primarily of people of Asian ethnicity. Asian is the largest group 46.4 percent) of 'other religions'
- 78.6 percent of the records for Māori religions are for people with Māori ethnicity.

Figure 2.11.1. Religion by ethnicity



2.11.4. Consistency

Was a consistent classification used?

There have been some changes to the lowest level of the classification since 2013. New religions have been added including: Mahayana Buddhism, Arise Church, Shi'a, Sunni, Pagan, Jedi, and Church of the Flying Spaghetti Monster.

Some religions have moved or merged categories. For example: Nazarene moved from other Christian to Methodist religions, and Plymouth Brethren and Exclusive Brethren merged into one category of Plymouth or Exclusive Brethren.

Religious affiliation is a hierarchical classification with three levels. Level 1 has 10 categories, level 2 contains 51 categories and level 3 contains 167 categories. Level 1 contains: No Religion; Buddhism; Christian; Hinduism; Islam; Judaism; Māori religions, beliefs and philosophies; Spiritualism and new age religions; Other religions, beliefs and philosophies; Residual categories (includes object to answering).

Respondents could write multiple answers to the religion question. If more than one religion was reported, each response up to a maximum of four responses was counted. As this is a multiple response variable, the total number of responses in a table is greater than the total number of people stated.

Was data collection consistent across online and paper data collection methods?

There was a higher response rate to religious affiliation online than on paper – maybe reflecting the characteristics of these populations.

- Of the fully responding individuals who answered online 99.0 percent specified a religious affiliation.
- Of the fully responding individuals who answered on paper 92.2 percent specified a religious affiliation.

There were differences in mode by religion. At level 1 of the classification Māori religions, beliefs and philosophies had the highest proportion of responses on paper (30.2 percent); Hinduism had the lowest proportion of paper forms (9.5 percent).

2.11.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

There are no reliable sources of religious affiliation other than the census.

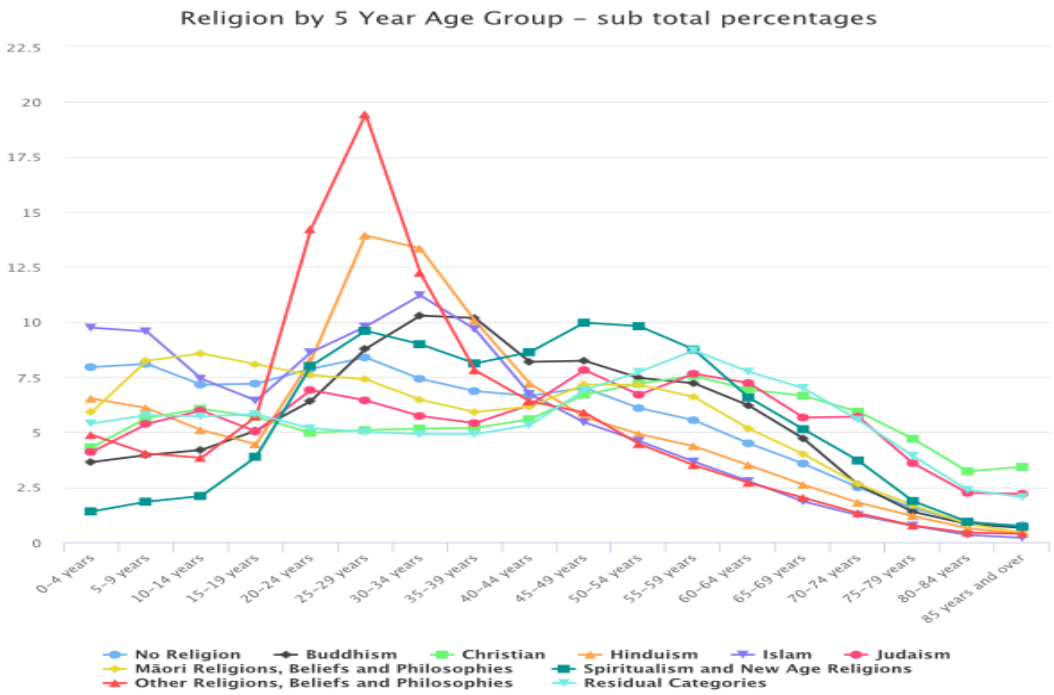
Is there a consistent time series with previous census, particularly census 2013?

The religious affiliation results are in line with historical trends and the expected impacts of migration. Stats NZ have compared 2018 Census results with 2006 and 2013. They have analysed trends in religious affiliation (level 1), and also religious affiliation by:
Sex; Ethnicity; Regional Council; 5-year age bands; and by mode of response (paper vs online). These checks appear credible, with results in line with historic trends and the major changes in migration since 2013.

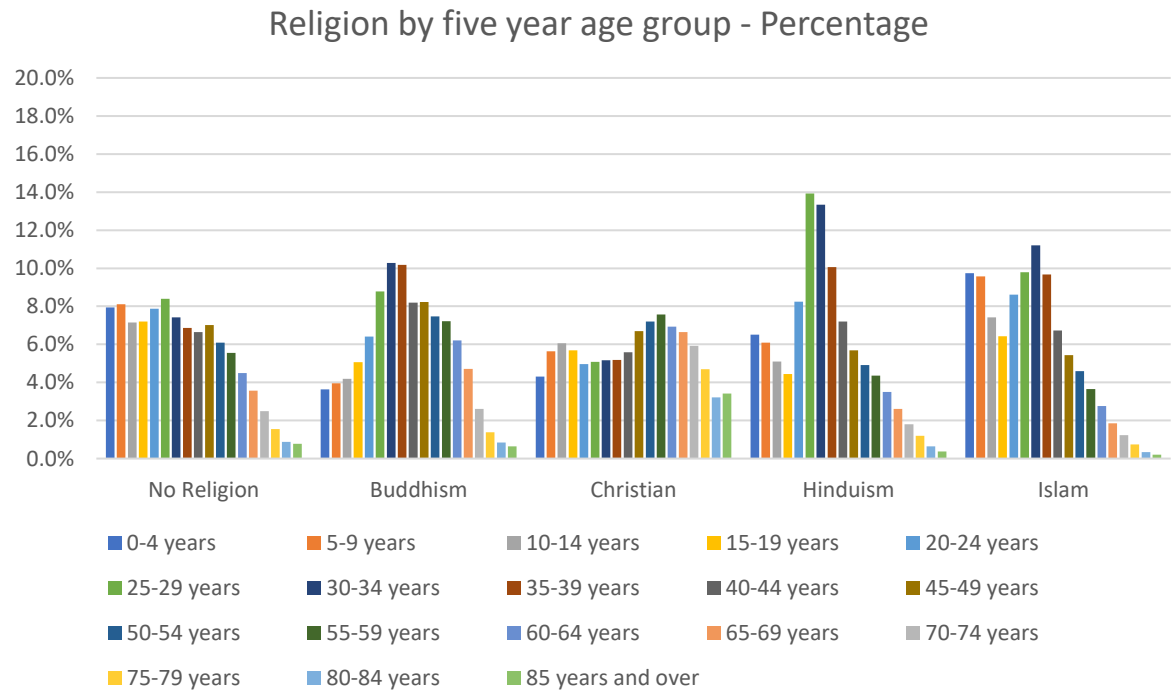
Stats NZ state "Quality checks were done between this variable and Age, Sex, Ethnicity, and Birthplace. No issues were found when checking these cross-tabs."

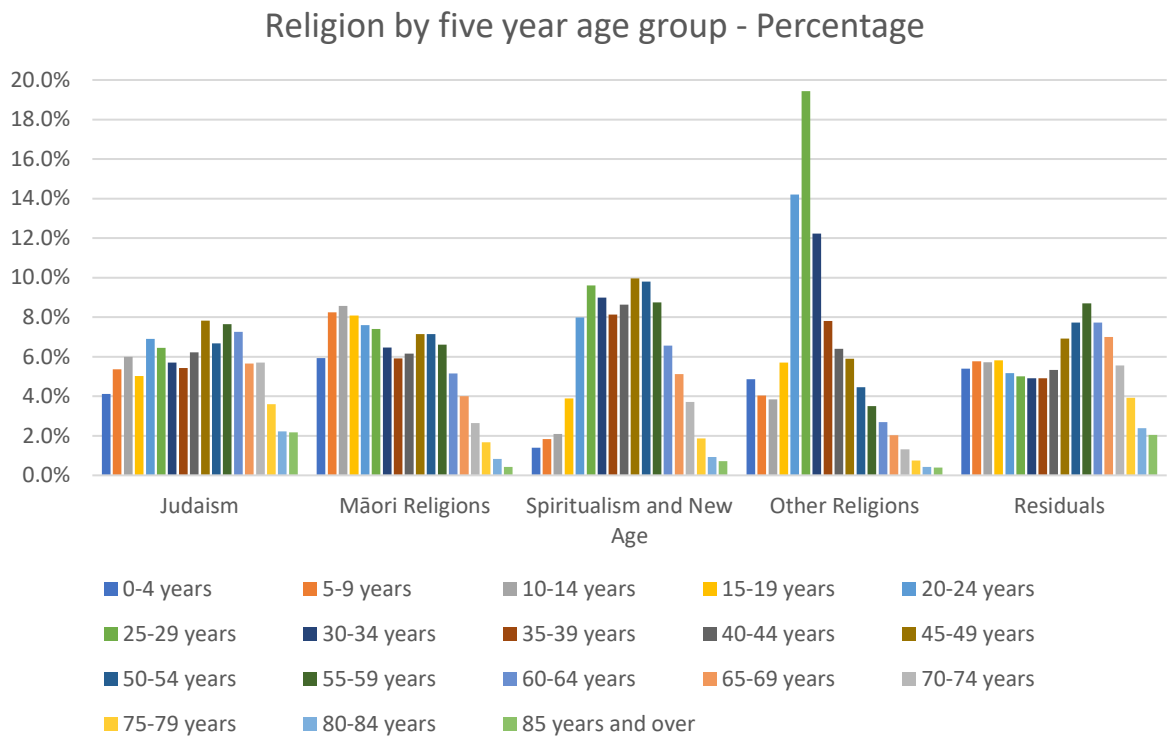
The graph below shows religious affiliation by 5-year age bands. It shows the kinds of patterns that would be expected (e.g. Christian having the highest percentage amongst older people, with 'other religions' having the highest percentage amongst young adults).

Figure 2.11.2. Religion by 5 Year Age Group



The charts below show a different representation of this data





These charts show, for instance, that Hindus in New Zealand affiliation is mostly prevalent among 20-39 years old, whereas for Christian the peak ages are 45-69.

2.11.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

82.9 percent of religious affiliation data came from individual 2018 Census forms, with a further 6.0 percent from 2018 Census based (88.9 percent from 2018 Census data overall). Nine percent of religious affiliation data came from 2013 Census data (or 2013 Census based imputation.

The validity of the use of 2013 Census data for 2018 Census results will depend on the extent to which religious affiliation is fixed or changes over time. The 2018 Census results show, for instance, 'No Religion' increasing from 41.9 percent to 48.2 percent of the results. However, overall, the 8.2 percent sourced from 2013 Census data could have generated reasonable proxies for religion in 2018.

2.12. Status in employment

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/68c95ba5-cc3b-4cad-b286-1dfdcd86291d/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

2.12.1. Overall assessment

82.1 percent of the 'Status in employment' data was sourced from the 2018 Census, with the remaining 17.9 percent from imputation. There was no use of alternative data sources. Imputation is designed to be unbiased, but will increase uncertainty, especially at lower levels of disaggregation. Imputation performs less well when imputation rates are high. High levels of imputation for Pacific peoples (38.9 percent) indicate that the 'data sources and coverage' quality rating for Pacific peoples should be treated as **poor**, while low levels of imputation for European peoples (13.0 percent) indicate that the 'data sources and coverage' quality rating for European peoples should be treated as **high**. The panel endorses Stats NZ's rating of **moderate** overall.

Whilst the proportions of status in employment by various cross tabulations (sector of employment, sex, etc) look consistent with previous censuses, the levels are not. There is a step change in the time-series, particularly in those regions which had low response in 2013. This is noticeable in the increase of just over 50 percent in people in employment in Northland, for instance. Comparisons over time will therefore need to be made with care.

Stats NZ state that "care should be taken if comparing absolute figures to previous years. We recommend using proportions."

2.12.2. Background

Status in employment contains the following categories: Paid employee; Employer; Self-employed and without employees; Unpaid family worker; Not elsewhere included.

Table 2.12 shows various data sources used for this variable. There was no use of 2013 Census or admin data – imputation was used to fill the 17.9 percent response gap.

Table 2.12. Data sources: Status in employment – Employed census usually resident population aged 15 years and over

Source	Percent
Response from 2018 Census	82.1
2013 Census data	0.0
Administrative data	0.0
Statistical imputation	17.9
No information	0.0
Total	100

In 2018, after imputation there is no missing data. In comparison, in 2013 there was 2.2 percent item non-response for employment status. There were also 4.8 percent substitutes, a proportion of whom should have been identified as employed and are therefore also missing employment status. In 2006 the corresponding figures were 2.9 percent item non-response and 3.3 percent substitutes.

2.12.3. Coverage

Was there good coverage for the overall population?

There is complete coverage due to the use of imputation.

Imputation is designed to be unbiased, and to adjust distributions to correct for this differential nature of non-response patterns. However, imputation also increases uncertainty as it does not guarantee an accurate value for each individual. Results for groups/areas with higher imputation rates will have higher uncertainty.

There was an increase from 2013 to 2018 in the number of 'Paid Employees' in all age groups, and in particular amongst 20 to 39 year olds where the increase is much higher than the trends suggest. The use of administrative records has led to a probably better-quality count for 2018 but, due to young adults being undercounted in the 2013 Census, has introduced a step change in the number of young adults captured in 2018 Census. So changes over time will not reflect observed changes.

Analyses of status in employment by sector of ownership and sex shows consistent trends and patterns as in previous censuses.

Was there good coverage for ethnic groups and regions?

Use of imputation was greater in Māori (31.5 percent), Pacific (38.9 percent) and Asian (22.1 percent) populations than in the general population, and also high in Gisborne (24.7 percent) and Northland (21.7 percent). Given this level of imputation, the 'data sources and coverage' quality rating for status in employment indicates all ethnic groups should be rated a **moderate**; except the Pacific ethnic group should be rated as **poor** (rating = 0.88); and the European ethnic group should be rated as **high** (rating = 0.96).

Analyses by Regional Council show Northland (47.5 percent) and Bay of Plenty (40.8 percent) saw the largest increase in 'paid employees' between 2013 and 2018. Gisborne had a 33.5 percent increase. All three regions had low response rates to the 2018 Census and therefore the data rely disproportionately on use of administrative records. The same areas also had lower than average response rates in the 2013 Census. This has almost certainly resulted in a higher quality count for 2018, but has created a step change in the population counts. Changes from 2013 to 2018 are therefore likely to overestimate observed changes.

2.12.4. Consistency

Was a consistent classification used?

The classification of status in employment in the 2018 Census is consistent with that of both the 2006 Census and the 2013 Census. The question used in 2018 was the same as the one used in 2013.

Status in employment is a flat classification with the following categories: Paid employee; Employer; Self-employed and without employees; Unpaid family worker; Not elsewhere included.

Was data collection consistent across online and paper data collection methods?
There were no differences between the wording or question format in the online and paper versions of this question.

- On the online form routing directed individuals to the appropriate questions so that only those in the subject population (employed usual residents, aged 15 or older) saw this question. Only one response could be selected for the status in employment question.
- On the paper form it was possible for individuals outside of the subject population to respond; multiple responses for this question were possible. These were dealt with during processing.

The majority of responses were for paid employees, with 90.5 percent of responses online. 84.5 percent of unpaid family workers responded online, but these represent small numbers.

2.12.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
The census results have been compared to the Household Labour Force Survey (HLFS) which showed far larger increases in the 'Employer' and 'Self Employed' categories between 2013 and 2018, and far smaller increases in the 'Paid employee' category than the census. Note that while there are differences in methodology between the HLFS and the census – e.g., the HLFS instructs contractors to state they are self-employed, and lists response options in a different order – this should not have impacted measures of change over time. As such, the reasons for these differences are unclear.

Is there a consistent time series with previous census, particularly census 2013?
Patterns and trends of proportions look consistent with previous censuses. The use of imputation to fill non-response gaps has led to step-changes in the time series, especially for those regions where response was low in 2013. Comparisons can be made, but with care.

2.12.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?
Yes, all data were sourced from the census 2018 individual forms, either directly or through imputation.

2.13. Study participation

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/67b643ce-0ad7-4953-8408-584bf10f5038

EDQ Panel rating: **Moderate/Poor**

Stats NZ rating: **High**

2.13.1. Overall assessment

The trends are roughly in line with what was found in 2013 (for >15 years), albeit with slightly lower study participation. However, there are four reasons why this variable should be considered of moderate (or poor) rather than high quality:

- The data are poor for <15s (data were collected on this group in for the first time at the 2018 Census);
- Use of alternative sources (i.e., administrative data and CANCEIS imputation) tended to decrease study participation. Alternative sources were used for responses for >25 percent for those aged 22-27 (when study participation is common), and use was also high for Māori and Pacific (30.5 percent and 34.8 percent, respectively), especially young Māori and Pacific where we might expect higher study participation;
- The use of administrative data only to confirm the non-studying population may have led to some undercount of those aged over 18 years who were not enrolled in December 2017, but were enrolled in study in Mar 2018.
- The extensive use of imputation and administrative data for Māori, Pacific, and several regions (e.g., Gisborne, Northland) means that data quality for these population subgroups should be considered **moderate**.

2.13.2. Background

Two changes in data collection and processing were implemented in 2018.

First, the question was asked of <15s for the first time. Stats NZ state: "The subject population was changed so that the study participation question could serve as a filter for the travel to education question." However, pre-testing suggested that parents under-estimated the study participation of <15 year olds, and Stats NZ recommends that "Study participation data for under 15-year-olds should be used with caution because there is an undercount compared with admin data, especially for early childhood (1 to 4 years). This may be due to parents considering early childhood education as childcare rather than education."

Second, imputation was used for the first time, resulting in no 'not stated' responses. CANCEIS (nearest neighbour) imputation was used for all those aged 18 years or less and all those in the country for less than three years, and in some other limited circumstances. Administrative data from educational enrolment statistics was used in other cases. As educational enrolment data were taken as at 31 December 2017 (the most recent data available), it is only used to confirm those **not enrolled** in study over 18 years of age and in the country for more than three years. There is an assumption that it is unlikely that this group will enrol in study in 2018 if they are not enrolled in December 2017. This has the potential to undercount enrolment (particularly for tertiary students as December is outside

the normal academic year). The remainder of missing responses for these groups will be imputed.

Overall, 83.0 percent of responses come from the Census, 7.4 percent from CANCEIS ('nearest neighbour') imputation, and 9.5 percent from admin data (Table 2.13). Administrative data were obtained from Ministry of Education data on Courses, Enrolments, TEC IT learners, Targeted Training.

Table 2.13. Data Sources: Study participation – Census usually resident population

Source	Percent
Response from 2018 Census	83.0
2013 Census data	0.0
Administrative data	9.5
Statistical imputation	7.4
No information	0.0
Total	100

2.13.3. Coverage

Was there good coverage for the overall population?

There were no 'not stated' responses, compared to 10.4 percent in 2013. The elimination of 'not stated' responses from 2018 to 2013 (through the use of admin data and imputation) necessarily results in an increase in other categories.

The 'data sources and coverage' quality rating for study participation is 0.95 and so is rated as **high** (i.e., in the range 0.95-<0.98, see Appendix 1 - Stats NZ data quality assurance definitions for 2018 Census)**.**

Was there good coverage for ethnic groups and regions?

There was greater use of admin data and CANCEIS imputation for Māori (14.4 percent and 16.1 percent, respectively), and Pacific peoples (16.2 percent and 18.6 percent. respectively). As such, the 'data sources and coverage' quality rating for Māori is 0.91 and for Pacific is 0.90 and so should be considered **moderate** (i.e., in the range 0.90-<0.95).

Similarly, there were a number of regions, notably Gisborne and Northland, which relied more heavily on admin data (Gisborne: 11.8 percent; Northland: 12.3 percent) and imputation (Gisborne: 12.1 percent; Northland: 10.0 percent), and so are rated as **moderate** for 'data sources and coverage' (Gisborne: 0.93; Northland: 0.94).

2.13.4. Consistency

Was a consistent classification used?

The classification has changed since 2013. The classification used in 2018 (Census Study Classification 2V3.0.0) has the following categories:

01 Full-time study
02 Part-time study
04 Not studying
99 Not elsewhere included

Note that full-time study is considered 20 hours or more a week and part-time study less than 20 hours a week.

This classification has changed in two ways since 2013. First, in 2013 a category '03 Full-time and part-time study' was used for individuals enrolled in both, but this was not used in 2018. On the paper form, those who selected both '01 Full-time study' and '02 Part-time study' were coded to '01 Full-time study', using the logic that they study 20 hours or more a week. On the online form, only one option could be selected.

Second, there was no '99 Not elsewhere included' category used in 2013 (or 2006). The '99 Not elsewhere included' category was a potential category in 2018, but this category was not populated after the addition of records from administrative sources and donor imputation.

Note also that use of Census data in 89.9 percent of cases (83.0 percent directly reported and 6.9 percent through imputation sourced from a donor's response from a 2018 Census form) uses a 'participant's perception of studying' which may include informal study, whereas admin data (9.5 percent of cases) used strict enrolment/non-enrolment criteria.

Was data collection consistent across online and paper data collection methods?
As described above, online completion did not allow the selection of both 'full-time study' and 'part-time study', whereas paper completion did. On the paper form, those who selected both '01 Full-time study' and '02 Part-time study' were coded to '01 Full-time study'. Also, online completion allowed respondents to select only one of either 'Full-time study' or 'Part-time study' or 'No – neither of these'. Paper completion allowed respondents to select 'No – neither of these' as well as another option. Where this occurred, administrative data or imputation were used to generate a response.

Note also that 99.5 percent of those completing online completed the 'study participation' field compared to 90.1 percent completing on paper.

2.13.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
Comparison of trends against the Household Labour Force Survey (HLFS) suggest the trend of slightly lower study participation may be real (i.e., the HLFS also showed an increase in study non-participation from 2013 to 2018).

Is there a consistent time series with previous census, particularly census 2013?

Trends are similar with Census 2013 (taking account of the 10 percent 'not stated' in 2013), with slightly lower study participation.

2.13.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

No, data from administrative sources (9.5 percent) were obtained from study participation records at 31 December 2017, whereas the remaining data were obtained from census responses (either directly or through imputation) in March 2018. Note also that 31 December 2017 is outside of normal study terms and semesters, so enrolment data extracted from administrative sources will have been lower than term time for tertiary enrolments.

2.14. Total Personal Income & Sources of Personal Income

DataInfo+ links:

Total Personal Income: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/4dc6188a-e884-4be0-bd53-7f03c60121a9/

Sources of Personal Income: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/ab874ce5-0889-423c-a2c0-dd4a89a355a9/

EDQ Panel rating: **High** (both 'Total Personal Income' and 'Sources of Personal Income')

Stats NZ rating: **High** (both 'Total Personal Income' and 'Sources of Personal Income')

2.14.1. Overall assessment

Income usually has a higher non-response than most other census questions (9.7 percent item non-response to total personal income in the 2013 Census). Linking to tax records provides a coverage and granularity that will be more accurate for what it measures. The income distribution from tax records has in the past differed from that obtained from the census, so the results for 2018 are likely to add some variability to measures of change between 2013 and 2018.

In addition, because of the difference in coverage, the results for 2018 are likely to add variability to Census based measures of change between 2013 and 2018, with increased uncertainty about the extent to which the changes are real. However, where census income information is used to measure differences between population groups recorded in the 2018 Census, then data quality will most likely have improved. Therefore, when using income from the 2018 Census, any assessment of quality will need to be based on the purpose to which the data are put.

2.14.2. Background

Table 2.14 shows various data sources used for these variables.

Table 2.14. Data sources: Personal income
- Census usually resident population aged 15 years and over

Source	Total personal income	Sources of income
	Percent	Percent
Response from 2018 Census	81.2	83.6
2013 Census data	0.0	0.0
Administrative data	16.5	14.1
Statistical imputation	2.3	2.1
No information	0.0	0.2
Total	100	100

For Total Personal Income data, 81.2 percent of records were obtained from census individual forms; 16.5 percent from admin data, and 2.3 percent via imputation.

For Sources of Personal Income data 83.6 percent of records were obtained from census individual forms; 14.1 percent from admin data, and 2.1 percent via imputation. There was no information for 0.2 percent of records.

The 'no information' percentage is where it was not possible to source income data for a person in the subject population.

Administrative data sources

For *total personal income* Inland Revenue income information was used consisting of the most recent filing of: Personal tax summary or Individual income return (IR3). In addition tax year summary data was derived from the following: Employer monthly Schedule (EMS); Individual income return (IR3); Company shareholders details (IR4S); IR20.

Some people may have only partial information for the 2018 tax year sourced from admin data. 2017 tax year data was sourced from admin data only when there was no information for the 2018 tax year.

For *sources of personal income*, data was derived solely from Inland Revenue, which provides multiple sources of income information from the following registers: personal Tax Summary; IR3; tax year summary.

Note that Inland Revenue collects income as actual dollar values but these were output into the income bands classification. The income data from Inland Revenue is only the taxable income for an individual. While it includes income from ACC, NZ superannuation and other main benefits, not all income is available from tax data. For example it does not include:

- superannuation, pensions, or annuities (other than NZ Superannuation, Veteran's Pension or war pensions)
- supported living payments
- other government benefits such as income support payments.

2.14.3. Coverage

Was there good coverage for the overall population?

Yes. Total Personal Income information will be comprehensive because of the scope of the tax system. Between them 2018 Census and tax data gave approximately 98 percent coverage. This compares to a 9.7 percent item non-response to this variable in the 2013 Census.

For sources of personal income 2018 Census and tax data gave approximately 98 percent coverage. There was 2.3 percent missing data before statistical imputation was applied. This compares to a 7.2 percent item non-response to this variable in the 2013 Census.

At the time of writing, household income analyses has not been possible because household and family information is not yet available due to quality issues.

Was there good coverage for ethnic groups and regions?
Yes. The completeness of the census and tax data combined will be mirrored in the census final database. The 2018 Census sources of income follows similar trends as in 2006 and 2013 across all regions, but with higher proportion of wage and salary earners compared to 2013 and 2006.

2.14.4. Consistency

Was a consistent classification used?
There were no conceptual and classification changes to total personal income variable in 2018. For sources of personal income, although there were changes in benefit categories from MSD in 2013, there are no other changes in the classification which means the data on sources is still comparable to 2006 and 2013.

Was data collection consistent across online and paper data collection methods?
The quality of online responses will be higher than on paper. Most of the income responses in 2018 are from the online form, where people can only select one income band using the online form but can select multiple income bands on the paper form.

2.14.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
For the total personal income variable, data quality was checked at regional council level, using data from Household Economic Survey (HES). The results showed the income distribution in 2018 Census is very similar with the income distribution in the Household Economic Survey. Most of the differences are within 1 percentage point. For Sources of Income, most of the data in 2018 Census follows similar trends from 2006 and 2013 censuses, except for incomes derived from wages and salaries, which will be the largest categories.

Is there a consistent time series with previous census, particularly census 2013?
Broadly. There will be some conceptual differences with 2013 because of the use of tax data (e.g. slightly different reference periods and uncertainties over whether gross or net income is being reported in the census). The ability to compare with 2006 is limited because of the change in the income scales used on the census form since 2013.

2.14.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?
Total personal income relates to 12 months ending on 31 March 2018, in line with the tax year whilst sources of income collects information for the 12 months ending 6 March 2018 (census night).

2.15. Usual residence five years ago

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/58180123-b856-4fed-9b91-b006d16e43b8/

EDQ Panel rating: **Poor**

Stats NZ rating: **Poor**

2.15.1. Overall assessment

This variable is unique as, by design, in 2018, 'usual residence five years ago' was not included as a question on the individual form as in previous censuses. This change followed consultation with users who said usual residence one year ago was critical for producing accurate population estimates and estimates of internal migration. As a result usual residence one year ago replaced usual residence five years ago on the 2018 Census form.

A new approach was used to calculate usual residence five years ago. While the categories 'Not born 5 years ago' and 'overseas 5 years ago' are derived from age and years since arrival in NZ, the main source of information to derive this variable was linked records for both 2018 and 2013 Censuses for the same individual.

Information was missing for 14.6 percent of the population which indicates that the variable is 'poor quality'.

Stats NZ state that "While using historic census data in place of a census question worked relatively well, the level of missing information (14.6 percent) was still much higher than the 2013 non-response rate of 3.5 percent." They also state that "the responses that were linked to 2013 Census data are high quality."

2.15.2. Background

This variable provides information on the migration of people within New Zealand and of those who have arrived from overseas over the five years prior to 6 March 2018. A respondent is classified as being overseas five years ago if they arrived in New Zealand less than five years ago. There is also a category for 'not born 5 years ago'. The data is used to support long-term service planning (for example housing and transport planning, public health planning) and for understanding population mobility within a region.

Consultation with users indicated that while both usual residence one year ago and usual residence five years ago held distinct value, information on usual residence one year ago was critical for producing accurate population estimates. Cognitive testing confirmed that the increase in address questions was burdensome for respondents, particularly on the paper form. As a result, Stats NZ decided to include only one question on internal migration on the 2018 Census forms. Usual residence one year ago was therefore replaced by usual residence five years ago on the 2018 Census form.

After first excluding those not born 5 years ago, and those living overseas 5 years ago, the remaining 2018 Census records or admin records were matched back to their 2013 Census responses. Their usual residence meshblock from the 2013 Census data was then used to

represent their usual residence five years ago. The 2013 and 2018 usual residence meshblocks were then compared, and the indicator and summary codes were produced.

The non-response rate to the usual residence five years ago question for the usually resident population was 2.9 percent in 2006 and 3.5 percent in 2013.

The 89 percent response to the 2018 Census and the high match rate to the 2013 census suggest that this approach could produce good quality data if the census response rate is high enough - although even if the planned 94 percent response rate had been achieved this would still have led to a gap that was around twice the non-response rate to the 2006 and 2013 Censuses. There is also potential for more of the missing information to be completed. The 'years at address' question from current census responses would indicate those who were at the same address 5 years ago. For others, administrative sources could be used to determine where they lived 5 years previously.

2.15.3. Coverage

Was there good coverage for the overall population?

Partially – the results are based on 85.4 percent of the census usually resident populations. Particular attention was paid to the 'not born five years ago' category, which almost matched the number of births in the last five years.

Was there good coverage for ethnic groups and regions?

Results for ethnicity level 1 show broadly consistent patterns in 2018 as with the 2013 and 2006 Censuses, but with slightly more people moving within New Zealand compared to previously, and slightly fewer people at the same usual residence. These patterns applied across most level 1 ethnic groups. For instance:

- 1.273m people of European ethnicity were at the same address as five years ago in 2018, compared to 1.377m in 2013. 1.323m were elsewhere in New Zealand in 2018, compared to 1.129m in 2013
- 195,000 people of Māori ethnicity were at the same address as five years ago in 2018, compared to 214,000 in 2013. 324,000 were elsewhere in New Zealand in 2018, compared to 265,000 in 2013

Of their regional population, Northland had the highest proportion (20.1 percent) of missing data; the next highest regional council was Gisborne (18.5 percent).

2.15.4. Consistency

Was a consistent classification used?

No. Differences from the 2013 Census classifications include:

- area units were replaced by SA2s.
- two residual categories 'don't know' and 'refused to answer' were removed
- a new residual category of 'unable to match 2013 Census data' was added
- SA2, TA, and RC categories were appended with 'not further defined (nfd)' to clarify their use in coding.

Was data collection consistent across online and paper data collection methods?

Not relevant as not asked in the 2018 Census.

2.15.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

There are no alternative sources. Data is consistent for births data (for not born five years ago) and migration data (for overseas five years ago).

Is there a consistent time series with previous census, particularly census 2013?

No. There was a major change in methodology for this variable and the level of missing data in the 2018 Census has led to breaks in the time-series.

It is conceivable that matching 2018 Census and admin records to 2013 meshblock might give a higher quality result than individuals recollection of their usual residence five years ago (especially for the most mobile parts of the population, which tends to be younger people and new migrants). However, this would still have created a break in the time-series compared to the 2013 Census. In addition, the level of missing data in the 2018 Census has introduced breaks in the time series.

2.15.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

All data for this variable for those who lived in New Zealand 5 years ago were sourced from matching the 2018 Census individual records to the equivalent individual from the 2013 Census and comparing the two usual residence addresses. The not born 5 years ago and overseas 5 years ago categories are obtained from the relevant 2018 Census questions or admin data.

2.16. Work and Labour Force status

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/ab229e2c-1ff2-44fc-b6be-d2479cd4e690/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

2.16.1. Overall assessment

Work and labour force status data is mostly consistent with expectations across a range of checks. The 2018 Census use of admin data to count people who had not responded and the greater use of imputation for non-response and residual categories has, however, led to a break in the time series for those employed (both full-time and part-time), particularly in those regions which had low response in 2013.

Data for those classified as 'not in the labour force' was consistent with the 2006–2013 time series and expectations.

Stats NZ state "data has been assessed to be consistent with expectations at the regional council level of geography. Some variation is possible at geographies below this level."

2.16.2. Background

Work and labour force status is used by central and local government, businesses, researchers, and community groups to: analyse the labour market position of population groups and small geographic areas; analyse occupation and industry composition, the size and characteristics of the labour market, the links between income, qualifications and labour market outcomes, and measure changes over time; provide a broad indicator of socio-economic status; develop the New Zealand Deprivation Index.

Table 2.16 shows the various data sources used for the input variables from which work and labour force status is derived.

The job indicator variable applies to everyone in the subject population and indicates whether someone is employed, or not. It has an imputation rate of 16 percent. Hours worked is imputed to determine full and part-time employment. The three remaining variables are imputed for those who are not employed to derive the unemployed and not in the labour force categories. Imputation rates are higher for the variables applied to the 'not employed' sub-group.

In output for previous censuses, if a person in the subject population did not answer a work and labour force question that was relevant to them, their work and labour force status was imputed. However, substitute records, (created when there was evidence that a person existed within a dwelling but did not complete an individual form), were not imputed and were instead coded to the residual 'Work and Labour Force Status Unidentifiable'.

In 2018, if a respondent did not answer a work and labour force status question that was relevant to them or did not provide the type of information asked for, a response was imputed to that question. The work and labour force status variable was then derived. This

means that there is nearly no 'Work and Labour Force Status Unidentifiable' in 2018. There were 4.9 percent 'Work and Labour Force Status Unidentifiable' in 2013 and 3.4 percent in 2006.

Table 2.16. Data sources: Work and labour force status input variables
Job indicator – Census usually resident population aged 15 years and over
Hours worked – Employed Census usually resident population aged 15 years and over
Seeking work/ Job search methods/ Availability for work – Unemployed Census usually resident population aged 15 years and over

	Job indicator	Hours worked	Seeking work	Job search methods	Availability for work
Source	**Percent**	**Percent**	**Percent**	**Percent**	**Percent**
Response from 2018 Census	84.0	81.3	78.8	78.9	78.9
2013 Census data	0.0	0.0	0.0	0.0	0.0
Administrative data	0.0	0.0	0.0	0.0	0.0
Statistical imputation	16.0	18.7	21.2	21.1	21.1
No information	0.0	0.0	0.0	0.0	0.0
Total	100	100	100	100	100

Due to rounding, individual figures may not always sum to the stated total(s)

2.16.3. Coverage

Was there good coverage for the overall population?

The proportions of work and labour force status look consistent with previous censuses for sex, age (except for a large increase in numbers aged 25-29 years).

Was there good coverage for ethnic groups and regions?

The proportions of work and labour force status look consistent with previous censuses, plus observed increases in employment /decreases in unemployment since 2013.

The proportions of work and labour force status look consistent between censuses by regional council – for instance a decrease in employed full time from 2006 to 2013, and then an increase into 2018 for all regions.

2.16.4. Consistency

Was a consistent classification used?

Work and Labour Force Status is a flat classification with the following categories: employed full-time; employed part-time; unemployed; not in the labour force; work and Labour force status unidentifiable. This classification has been used since the 2001 census - see : Work and Labour Force Status is a flat classification

Was data collection consistent across online and paper data collection methods?

There were no differences between the wording or question format in the online and paper versions of the questions used to derive work and labour force status.

- On the online form routing directed individuals to the appropriate questions so that only those in the subject population saw the relevant questions e.g. only if they had already said that they usually lived in New Zealand, were aged 15 years or over, and were in employment would they see the hours worked per week question.
- On the paper form, respondents were directed to the appropriate questions, however, it was still possible for individuals to provide responses to the questions that were not applicable to them. These were dealt with during processing.

2.16.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

Comparisons were made to the Household Labour Force Survey (HLFS) – census results were within the (wide) range of expectations. In June 2016 there was a change in the hours worked boundary between full time and part-time, making comparisons of HLFS over the census period problematic. See Household Labour Force Survey – Revisions to labour market estimates for more information.

Is there a consistent time series with previous census, particularly census 2013?

Patterns and trends of proportions look consistent with previous censuses. The use of imputation to plug non-response gaps has led to step-changes in the time series, especially for those regions where response was low in the 2013 Census. So even though the 2018 Census results are higher quality than those for 2013, comparisons over time will need to be made with care.

2.16.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Yes, all data were sourced from the census 2018 individual forms, either directly or through imputation.

2.17. Years since arrival in New Zealand

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/9c4be05d-634b-4502-bddf-ce4cb4abd301/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

2.17.1. Overall assessment

This variable applies to the population that is not born in New Zealand – i.e. the population born overseas. Stats NZ state that data has been assessed to be consistent with previous trends plus known migration at the regional council level of geography. Some variation is possible at geographies below this level.

2.17.2. Background

Table 2.17 shows the various data sources used for this variable.

Table 2.17. Data sources: Years since arrival in New Zealand – Overseas-born census usually resident population

Source	Percent
Response from 2018 Census	83.9
2013 Census data	7.7
Administrative data	7.1
Statistical imputation	0.0
No information	1.3
Total	100

Of those who submitted a 2018 Census individual form, the initial item non-response rate for this question was 2.6 percent (compared to 3.6 percent in 2013). In 2018 these gaps were filled with 2013 Census data or admin data. After the use of 2013 Census and admin data and admin enumeration the final non-response rate was 1.3 percent. For comparison, the non-response for 2013 was 3.6 percent and for 2006 was 3.8 percent.

The only source of admin data used relates to migration (Ministry of Business, Innovation and Employment). The earliest IDI movements data are from June 1997. This means the maximum number of years since arrival value from MBIE data will be 20. However, due to inconsistencies in the data, only movements up to 18 years since arrival are used for the 2018 Census. Where an individual first moved to NZ prior to 1997, Stats NZ will only pick up their first border movement post-1997.

2.17.3. Coverage

Was there good coverage for the overall population?

Coverage is very high, with years of arrival in NZ available for 98.7 percent of the overseas born population. A further 1.2 percent have no information on birthplace, the majority of

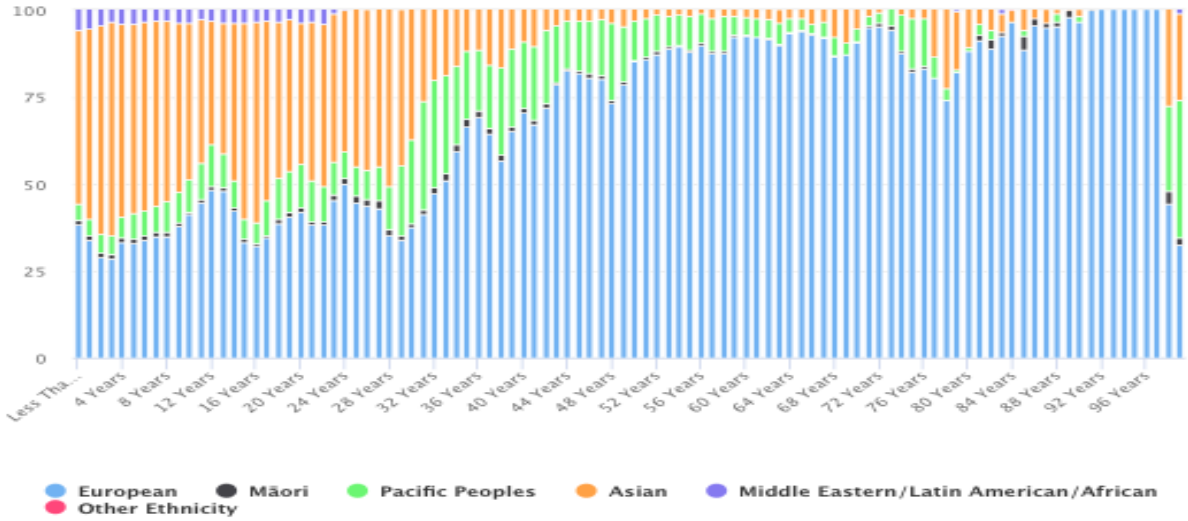
whom are likely to be born overseas. There was a higher proportion of admin data used for people aged less than20 years (7.4 percent) compared to those aged 20 years and over (7.0 percent).

Was there good coverage for ethnic groups and regions?

Given the variable response rates by ethnicity to 2018 Census higher proportions of data for the years since arrival variable will be from 2013 Census and admin sources for the Pacific ethnic group especially.

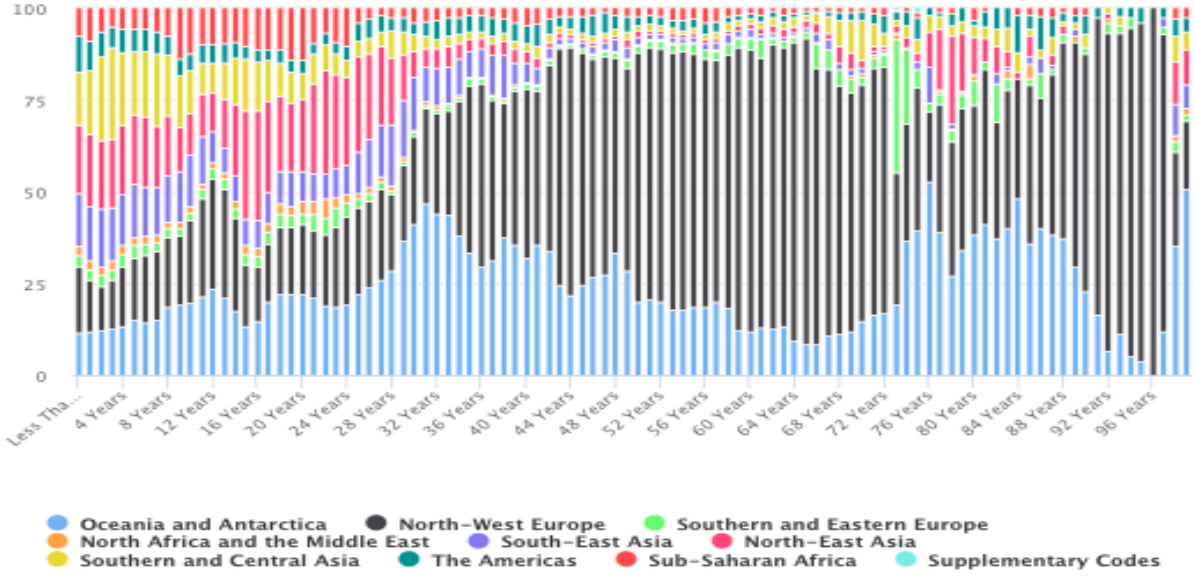
The following chart shows the high proportion of people with Asian ethnicity with less than 28 years since arrival.

Figure 2.17.1. Years since arrival by Ethnicity - Percent



The chart below shows years since arrival by birthplace, with very similar trends – with the highest proportion of migrants in the last 25-30 years being from South-East Asia, North-East Asia, and Southern and central Asia.

Figure 2.17.2. Years since arrival by Birthplace - Percent



Stats NZ have compared trends since the 2006 and 2013 Census at the Regional Council level; 2018 Census results appear consistent. The Metric 1 (Data sources and coverage) quality rating is High Quality for every region apart from Gisborne, which is Moderate.

2.17.4. Consistency

Was a consistent classification used?

Years since arrival is a flat classification with single year categories. The classification has not changed since the 2013 Census.

Was data collection consistent across online and paper data collection methods?

There were minor differences in question format and wording between online and paper forms to reflect the routing available online. The wording of the paper form question was unchanged from the 2013 Census. There were similar levels of item non-response between paper and online.

2.17.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

There is no alternative source, other than migration data for those who have arrived since the late 1990s.

Is there a consistent time series with previous census, particularly census 2013?

The WoF states that there is moderate consistency with previous censuses, which drives the overall Moderate rating. Stats NZ are carrying out further analyses of consistency with migration data since 2013, but at the time of writing the DataInfo+ page for this variable states that consistency and coherence is rated as Moderate quality.

2.17.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

No. However the use of 2013 Census data (aged by five years) for those in New Zealand on census day should be of high quality for this variable; likewise for who have arrived in New Zealand in the last 18 years and whose data was sourced from admin sources.

3. Detailed assessments - variables about dwellings and households

3.1. Access to telecommunications systems

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/42921c1a-a49d-4426-b3a9-69cfba642ba5/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

3.1.1. Overall assessment

There was a 92.3 percent response from households to this question, with no other source of data used. The level of non-response is only slightly higher than in 2013 (7.7 percent compared to 5.0 percent). The trends in the data appear consistent with expected changes in the use of technology.

Stats NZ have assessed the data at Regional Council level, but not at lower levels of geography. Stats NZ state that "Non-response rates for some regions are higher than desirable, however trends generally show expected patterns."

Stats NZ have compared 2018 Census trends (e.g. increasing levels of 'no access to landline') with industry information and the results appear in line.

3.1.2. Background

Access to telecommunications systems is used to indicate a household's ability to access services such as social and health care in an emergency to monitor the use of different types of telecommunications, and (since 2013) in the development of the New Zealand Deprivation Index. No access to the internet is one of the dimensions of deprivation used to create this index.

Table 3.1 shows that the data for this variable came only from the 2018 Census. There was no use of admin data or imputation to fill in gaps in response.

Table 3.1. Data sources: Access to telecommunication systems – Households in occupied private dwellings

Source	Percent
Response from 2018 Census	92.3
2013 Census data	0.0
Administrative data	0.0
Statistical imputation	0.0
No information	7.7
Total	100

The Access to telecommunications systems question was asked only on the dwelling questionnaire, so it does not show whether a particular household member has access to those amenities. Not every household member might have equal access.

The 'no information' percentage is where it was not possible to source access to telecommunication systems data for a household.

Responses that could not be classified or did not provide the type of information asked for (response unidentifiable) remain in the data and are included above in 'Response from 2018 Census' percentage. For output purposes, this residual category response is grouped with 'not stated' and classified as 'not elsewhere included' (nei). In 2006 nei accounted for 4.2 percent of the data; in 2013 it was 5.2 percent, and in 2018 7.8 percent.

Access to telecommunication systems is a multiple response variable and households reporting access to more than one type of telecommunication system are counted in each category that they had access to. Therefore, the total number of responses in a table is greater than the total number of households.

There were changes to the question wording and format from the 2013 Census. The 'fax access' response option was removed from the 2018 form. Both the online and paper version of the question had slightly different wording from 2013 for cellphone, as the bracketed text from the 2013 response option was dropped: 'a cellphone / mobile phone (that is here all or most of the time)'.

3.1.3. Coverage

Was there good coverage for the overall population?

There was 92.3 percent coverage of households in 2018 – i.e. 7.7 percent non-response rate. This compared to a 5 percent non-response rate in 2013.

Was there good coverage for regions?

The non-response rates for this question follow the overall non-response pattern for the 2018 Census, with the highest non-response rate in Gisborne Region (11 percent). The lowest non-response rate was 5 percent, in Southland Region.

Stats NZ have assessed the data at Regional Council level, but not at lower levels of geography. Stats NZ state that "Non-response rates for some regions are higher than desirable, however trends generally show expected patterns."

Given the variable levels of response amongst major ethnic groups, there will be data quality issues where levels of non-response are high (e.g. for Maori and Pacific peoples). Data for specific ethnic groups on access to telecommunications systems should be used with this variability clearly in mind.

3.1.4. Consistency

Was a consistent classification used?

Yes. The classification of access to telecommunication systems in the 2018 Census is consistent with the classification used in the 2013 and 2006 Censuses.

Access to telecommunication systems is a flat classification with the following categories: No Access to telecommunication systems, Access to a cellphone/mobile phone, Access to a telephone, Access to a fax machine, Access to the internet, Not elsewhere included.

'Not elsewhere included' contains the residual categories, including 'response unidentifiable' and 'not stated'.

Was data collection consistent across online and paper data collection methods?
For this question the questions were the same on paper and online. Of those who answered the access to telecommunications question, 13.7 percent responded on paper and 86.3 percent responded online.

6,500 households who completed 2018 Census online on paper said that they had 'No Access to Telecommunication Systems', compared to 10,500 households who completed 2018 Census online. While it may seem anomalous that 10,500 households who completed 2018 Census online did not have access to telecommunication systems, this might be valid if the respondent used public internet services, or online access at work, friends or family.

3.1.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
A high-level comparison with the Household Economic Survey (HES) 2015/16 expenditure data showed a similar pattern for internet access, and an explainable difference for cellphone access. There was a higher rate of telephone access in the HES compared with the 2018 Census, but the time-lag may explain this.

Stats NZ have also looked at industry and commerce commission reports on telecoms, which support the census trends:

- Spark reported a drop in landlines from 2015 to 2017 of 49,000 (from 215,000 to 166,000)
- a Commerce commission report from 2018 that states "Landline connections, including all fixed-line voice services have continued to decline in 2018. Over 40 percent of household fixed-line connections now have no voice service as more and more households are now opting to not have a home phone"
- Two reports from the Commerce commission - the '2018 Annual Telecommunications Monitoring Report – 18 December 2018' and the '2018 Telecommunications industry questionnaire results – 18 December 2018' - show the 2018 Census number of households with access to a telephone is higher than what the commerce commission reports (this might be due to the commerce commission annual monitoring reports being based on a voluntary questionnaire that is sent to the industry).

Is there a consistent time series with previous census, particularly census 2013?
The distributions of access/no access to landline, broadband, and mobile internet were compared to patterns in 2006 and 2013. The trends between these periods were in line with previous trends and market changes (e.g. increasing numbers of households giving up landlines in favour of mobile internet access):

- for access to a cellphone/mobile phone, there was an increase of 14 percent from 2013 to 2018 (compared to an increase of 18.9 percent from 2006 to 2013);
- there was a 0.53 percentage point decrease from 2013 to 2018 in households' access to any telecommunications type, compared to a 0.39 percentage point decrease from 2006 to 2013;
- access to a telephone saw a decrease of 24.2 percent fewer households with access between 2013 and 2018, compared to a 1.6 percent decrease between 2006 and 2013.

These trends can be seen through most of the areas at the regional council level.

3.1.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

There was no use of admin data for this variable. All responses (92.3 percent) were from the 2018 Census.

3.2. Counts of Dwellings

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/e38c17e6-8669-4916-af60-550c11788717

EDQ Panel rating: **High**

Stats NZ rating: **High**

3.2.1. Overall assessment

Stats NZ rate the quality of Dwelling counts as High. The panel endorse this assessment at National, Regional Council, TALB and SA1 levels of geography.

There are small numbers of SA2 and Meshblocks which have large changes in counts of dwellings. This is partly due to the incorporation of Marae and DOC huts in the Non Private Dwelling counts.

3.2.2. Background

A dwelling is any building or structure – or its parts – that is used, or intended to be used, for human habitation (see DataInfo+ link above). It can be of a permanent or temporary nature and include structures such as houses, motels, hotels, prisons, motor homes, huts, and tents. There can be more than one dwelling within a building. For example, each apartment in an apartment building is a dwelling.

The dwelling counts are based on the 'dwelling record type' at meshblock level taken from the Census Dwelling Frame (CDF). They are not derived from 2018 Census returns. The CDF is high quality and provides full coverage of New Zealand.

For the 2018 Census, Marae and Department of Conservation huts were included in the Census Dwelling Frame as Non-Private Dwellings (NPDs). This has resulted in a large increase in NPD counts compared to previous censuses.

Every dwelling on the Census Dwelling Frame has a unique identifier as well as other information that includes occupancy status, meshblock, dwelling type and sub-type, and geographic location. The CDF was derived from the Statistical Location Register (SLR) which was itself produced to support the 2018 Census - an address list to locate and enable respondents.

The Statistical Location Register is a combination of addresses and their associated attributes (e.g. x,y geo-reference). It was formed by geocoding the 2013 Census dwellings and is updated monthly by Land Information NZ (LINZ) and NZ Post data.

Other address sources were also added to the SLR and were maintained as SNZ addresses to match incoming data to a reference in the real world. Other inputs that contributed to the build of the SLR include:

- The Business Register geographic units (GEO)
- Building Consents (post 2013)

A 2018 Census Operational File was derived from these inputs using the address type attribute as the main filter e.g. private dwelling, commercial. Every dwelling in NZ had a unique ID, which is maintained through time.

The 2018 Census Operational File was verified by on-the-ground canvassing by field officers. Private dwelling addresses were divided between those which would be mailed out to (with an internet access code - 80 percent of dwellings) and those which would be visited in the field during live census operations (20 percent of dwellings). The mailout meshblocks were canvassed between June and August 2017 to check and update the address information; the field visit areas were canvassed as part of the census field operation, in February and March 2018.

The Census Operational File was updated during the census operations. The changes made indicate that initial dwelling classifications were of high quality – especially for Private Dwellings (the vast majority – 99.2 percent - of all dwellings):

- 97 percent of addresses loaded as Private Dwellings (PD) remained PD; 2 percent changed to Address Not Found (ANF)
- 89 percent of addresses loaded as Non-Private Dwellings remained NPD; 4 percent changed to PD and 3 percent became ANF
- 90 percent of addresses loaded as Commercial remained Commercial; 4 percent changed to ANF; and 2 percent to PDs.

A number of issues in the CDF were corrected between May and September 2018, the main issues being:

- Duplicate addresses
- Incorrect classification of non-private dwellings, and private dwellings at non-private dwellings (e.g. rest homes)
- Incorrect xy coordinates
- Incorrect occupancy status (or missing status)
- Significant differences from dwelling estimates.

At the conclusion of the work carried out between May and September 2018 the data in the dwelling frame, which had previously supported operational activity, was then ready for processing and evaluation. The function of the dwelling frame was now to support statistical processing and produce a clean unit record file (CURF) to enable the production of outputs from the 2018 Census data.

3.2.3. Coverage

Was there good coverage for the overall population?

Yes. There is full coverage of dwellings in New Zealand.

Was there good coverage for regions?

The dwelling counts are counts of Private and Non Private Dwellings. There is high coverage of all regions.

3.2.4. Consistency

Was a consistent classification used?

Not relevant – this is simply a count.

Was data collection consistent across online and paper data collection methods?
Not relevant – this information was not collected from census forms.

3.2.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
The 2018 Census dwelling count was 1.89 million dwellings in total (including private and non-private dwellings (both occupied and unoccupied) and dwellings under construction). The 2018 Census total private dwelling count is about 7,000 dwellings (0.4 percent) lower than the previously published 2013-based Dwelling and Household estimates.

Stats NZ collated national estimates of dwellings, which the Population Insights team produce quarterly, based on projections from the 2013 Census, updated by building consents, but it doesn't take account of demolitions.

The counts of private dwellings at national and lower geographies are very consistent with estimates. At the national level it is right in the middle of the range of the Population Insights estimate and the 2018 Census projected estimate.

Lower geography patterns look sensible – Canterbury and Auckland regions are both below the 2018 projected estimate. Stats NZ state that "We might expect this since the estimation methodology brings in building consents but does not account for demolitions."

Is there a consistent time series with previous census, particularly census 2013?
Counts of occupied private dwellings are consistent with census geography estimates at regional council level, which are calculated by extrapolating from 2013 using building consent data. The number of private dwellings is consistent with this estimate with the notable exception in Christchurch. This is because the estimate does not include demolitions. Here the estimate is much higher due to the large number of demolitions following the 2011 earthquakes.

The addition of Marae and Department of Conservation huts will have changed the NPD total.

3.2.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?
The Census Dwelling Frame, on which the dwelling counts are based was developed and quality assured in the run up to 2018 Census, was updated during the census, and then validated after the census.

3.3. Dwelling type

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/1db47d51-8bbf-4707-a97a-2092224f4a39

EDQ Panel rating: **Poor**

Stats NZ rating: **Moderate**

3.3.1. Overall assessment

Stats NZ rate the quality of 'Dwelling Type' as Moderate. However, the panel believe that this variable should be rated poor on Stats NZ's consistency and coherence quality dimension, rather than moderate, and thus poor overall.

Two inconsistencies lead the panel to this conclusion. First, between 2013 and 2018 there has been an increase in 'separate houses' and a decrease in 'joined dwellings' when the reverse was expected due to town houses, apartments, and other joined dwelling types increasing in popularity. Second, there had been a large increase from 2013 to 2018 in the number of dwellings with 'no storey information': a nearly twenty-fold increase for separate houses and a thirteen-fold increase for joined dwellings.

In the WoF for 'Dwelling type', Stats NZ state "The data from other sources used to mitigate the high levels of non-response and the classification changes improved the quality of the data. Therefore, 'dwelling type' data is fit for use for the same purposes as it was used in previous censuses." The panel do not accept this judgement.

Stats NZ have assessed data for this variable down to Regional Council; there could be issues with the data at lower levels of geography below the region level.

Stats NZ state that:

- Caution is needed when using dwelling type data for unoccupied dwellings and dwellings under construction due to the high level of imputation of dwelling type for these dwellings."

Stats NZ has identified the following caveats which means interpreting trends over time should be done with caution:

- the decrease in data in the 'private dwelling not further defined' category (from 5.8 percent in 2013 to less than 0.1 percent in 2018) may affect comparability over time
- Non-private dwelling not further defined reduced from 1.7 percent in 2013 to zero percent in 2018 and may likewise impact on estimates of change
- There is a bias in the data towards separate dwellings rather than joined private dwellings – probably not reflecting real-world changes
- There is a significant number of dwellings in the 'no storey information' – this data on number of storeys does not represent real-world trends
- Changes in the counts for residential care for older people and for residential and community care facilities are likely to be due to real-world change and improvements in identifying these dwelling types
- The increased number of boarding houses for 2018 is likely to be due to better identification of these dwellings

3.3.2.Background

The table below shows the breakdown of the various data sources used for this variable.

2018 dwelling type – occupied dwellings (private and non-private)	
Source	**Percent**
Response from 2018 Census	91.7
2013 Census data	2.4
Administrative data	3.7
Statistical imputation	2.2
No information	0.0
Total	100

The only admin source used was Tenancy Bonds from the Ministry of Business, Innovation and Employment which provides information for rented dwellings on separate houses and joined dwellings, but does not include number of storeys.

In 2018 no non-private dwellings were classified as not further defined and the percentage of private dwellings classified as not further defined was close to zero due to the use of the additional data sources described above.

Private dwelling not further defined: <0.1 percent in 2018; 5.8 percent in 2013

Non-private dwelling not further defined: 0.0 percent in 2018; 1.7 percent in 2013

The non-response rate for number of storeys was 6.8 percent in 2018. For comparison, the non-response for 2013 was 6.3 percent.

Dwelling type is derived from three questions on the dwelling form (dwelling description; dwelling joined or separate; and number of storeys) and the dwelling address type and subtype. Where there was not stated, don't know, response outside scope or response unidentifiable for the description, or one of the other two variables was not stated or unidentifiable, the record was set to nfd and then alternative sources were used where possible, and otherwise the record went through the imputation process.

Private dwellings are classified mainly by their construction. Non Private Dwellings (NPDs) by their function.

3.3.3. Coverage

Was there good coverage for the overall population?

Yes. 92 percent of records came from 2018 Census returns or census operations (for NPDs)

Was there good coverage for regions?
There were high levels of non-response to the census, which differed by region. Northland and Gisborne Regions had higher levels of non-response, and therefore a higher dependence on data from other sources, compared to regional council areas with lower levels of non-response, such as Nelson and Wellington Regions.

The Metric 1 (Data sources and coverage) quality rating is Very High Quality for every region apart from Northland and Gisborne regions, which are High.

3.3.4. Consistency

Was a consistent standard used?
There have been minor changes to the classification of this variable since the 2013 Census. These are:

- more detailed information on number of storeys for joined dwellings. The top category has been raised from 'four or more' to 'ten or more' and categories for 'four to six storeys' and 'seven to nine storeys' have been added. These changes were made to provide better information on apartments.
- dwellings joined to businesses or shops are now classified as joined dwellings instead of being included in 'occupied private dwelling not further defined' as previously.

Dwelling type is a hierarchical classification with three levels. Level one contains two categories: Private dwelling; Non-private dwelling. Level two contains seven categories and level three contains 33 categories. Private dwellings are classified mainly by their construction; Non Private Dwellings by their function.

No residual categories (i.e. categories such as not stated and response unidentifiable) were used for Level 1 of dwelling type. All dwellings were classified as private or non-private during processing. If no further information was available about what type of private or non-private dwelling it is, then the dwelling is classified as 'private dwelling not further defined' or 'non-private dwelling not further defined'.

Each independent self-care unit, villa, or house within a retirement village is classified as a private dwelling and is included in the appropriate private dwelling category according to whether it is separate or joined and the number of storeys.

Was data collection consistent across online and paper data collection methods?
The WoF contains no information on response rates by mode (online or paper).

3.3.5. Comparability

How does census 2018 data compare to recent collections of the same variable?
There are no other collections of this information. However, Stats NZ used information on Building Consents to help set expectations for private dwelling type numbers in advance of the 2018 Census. The definition of "dwellings" in Building consents issued is conceptually similar to "private dwellings" in the census (excluding "private dwellings: other"). Building consents issued also has information about the number of building consents for non-private dwellings.

Experimental Stats NZ dwelling estimates indicate, on average, that it takes roughly three quarters for a dwelling to be built after being consented, and that 97 percent of consents are built. Stats NZ took the number of new dwellings consented in the five years from June 2012 to May 2017 to reflect the increase in the private dwelling stock between the 2013 and 2018 Censuses (ignoring the fact that some private dwellings will also have been demolished). These estimates can be broken down to meshblock level. Assumptions about when dwellings were completed can be refined using data from the Quarterly Building Activity Survey.

Is there a consistent time series with previous census, particularly census 2013?
No.

Stats NZ states "Dwelling type in the 2018 Census is moderately comparable with the 2013 and 2006 Census data. Variable data is not consistent with some expectations across one or more consistency checks." However, the panel believe that this variable should be rated poor on Stats NZ's Consistency and coherence quality dimension, rather than moderate.

Stats NZ's definitions of 'moderate' and 'poor' quality for the consistency and coherence quality dimension is as follows:
Moderate: "Variable data is mostly consistent with expectations across consistency checks. There is an overall difference in the data compared with expectations and benchmarks that can be explained through a combination of real-world change, incorporation of other sources of data, or a change in how the variable has been collected."
Poor: "Variable data is not consistent overall with expectations across one or more consistency checks. There is an overall difference in the data compared with expectations and benchmarks. Where this difference occurs, this cannot be fully explained through likely real-world change, incorporation of other sources of data, or a change in how the variable has been collected."

There are some consistencies: At the top level of the 'dwelling type' classification, the numbers of occupied private and non-private dwellings have increased as expected.

- Occupied private dwellings grew from 1.56 million in 2013 to 1.66 million in 2018.
- Occupied non-private dwellings increased from 8,739 in 2013 to 9,567 in 2018.

However, there are two inconsistencies. First, between 2013 and 2018 there has been an increase in 'separate houses' and a decrease in 'joined dwellings' when the reverse was expected due to town houses, apartments, and other joined dwelling types increasing in popularity and availability in recent years. It is believed that the decrease in the percentage of 'joined dwellings' may be related to

(i) lower participation rates by those in joined dwellings and difficulties in enumerating some of these dwellings, including secure access dwellings; and
(ii) the use of administrative tenancy bond data (which has lower quality joined/separate dwelling information but is more up-to-date), preferentially ahead of 2013 Census data (which has higher quality joined/separate dwelling information but cannot capture information on dwellings that have been constructed since 2013).

Second, at level 3 of the classification there are 33 categories. As noted above, the overall non-response for number of stories in 2018 is similar to 2013. However the categories in which this is present have shifted. In 2013 most dwellings with no storey information were in the top level "Occupied Private Dwelling Not Further Defined". However in 2018, because nearly all dwellings now have a value for separate house, or joined dwelling, the missing storey information is captured within a sub-category of separate house or joined dwelling. The result is that for 'occupied private dwellings' the greatest numerical increases were for:

- 'separate house no storey information', which went from 4,479 in 2013 to 86,211 in 2018 (for Northland the increase was from 213 to 4,743); and
- 'joined dwelling no storey information' which increased from 2,055 in 2013 to 25,989 in 2018.

Although storey information should have been available from most of the (2.4 percent) of 2013 census records used for the 2018 'Dwelling Type' variable', of the (3.7 percent) of admin records used only around one in forty contained storey information. Better results would likely have been obtained if 2013 Census data had been given priority over the admin source.

Given these inconsistencies, 'Dwelling Type' is clearly closer to the definition of poor quality than moderate quality for this metric, given the following inconsistencies.

3.3.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

No – but this is probably not significant for this variable. 2.4 percent of the dwelling data was sourced from the 2013 Census, but for this variable that is probably acceptable as dwelling type is unlikely to have changed between censuses (unless a property has been redeveloped).

3.7 percent of the dwelling data is sourced from Admin data which is likely to be more recent than the 2013 Census.

3.4. Housing quality

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/ab8db4ff-c5b2-4a4f-bd2e-f2c71555d31f/

EDQ Panel rating Dampness; Mould; and Access to amenities: **Moderate**

Stats NZ rating Dampness; Mould; and Access to amenities: **Moderate**

3.4.1. Overall assessment

Housing Quality covers three topics that were new for the 2018 Census and were asked at the dwelling level: Dampness, Mould and Access to amenities. Stats NZ rate the quality of each of these variables as Moderate. The External Data Quality Panel endorse these assessments.

Stats NZ state that "data has been checked to regional council level. Some variation is possible at geographies below this level". The 'not stated' rates for damp and mould variables range between 5.9-11.6 percent approximately across all regions, which gives less confidence that these indicators truly reflect the state of housing quality regionally. As a priority three variable, Stats NZ has not carried out checking at smaller geographies.

Stats NZ state "It was expected that an overall higher response rate to the Census would likely have given a higher rate of 'damp - always' and 'damp - sometimes', based on our understanding of the characteristics of non-respondents."

It should be noted that a classification of dampness and mould relating to a dwelling is not necessarily the same as such a classification for individuals in the household. Mould and dampness can be room-specific and may affect those living in dwellings differentially.

3.4.2. Background

Table 3.4 shows that there were no alternative data sources or imputation used to replace missing responses or responses that could not be classified for the housing quality variables.

No information' in the data sources tables, is the percentage of the subject population coded to 'not stated'.

Note that responses that could not be classified or did not provide the information asked for ('response unidentifiable' and 'don't know') remain in the data. In the tables above they are included in the 'Response from 2018 Census' percentage. The rates are: dwelling dampness indicator: 3.0 percent; dwelling mould indicator: 2.2 percent; access to basic amenities: 0.1 percent. The effective response rates are therefore 0.1-3.0 percentage points lower than the 92 percent response rates would suggest.

For output purposes, these residual category responses are grouped with 'not stated' and are classified as 'not elsewhere included'. The rates are: dwelling dampness indicator: 11.0 percent; dwelling mould indicator: 10.3 percent; access to basic amenities: 8.1 percent.

Table 3.4. Data sources: Housing quality – Occupied private dwellings

	Dampness	Mould	Access to basic amenities
Source	**Percent**	**Percent**	**Percent**
Response from 2018 Census	92.0	91.9	92.0
2013 Census data	0.0	0.0	0.0
Administrative data	0.0	0.0	0.0
Statistical imputation	0.0	0.0	0.0
No information	8.0	8.1	8.0
Total	100	100	100

Due to rounding, individual figures may not always sum to the stated total(s)

3.4.3. Coverage

Was there good coverage for the overall population?

Partly. The nature of the 2018 census non-response problems is likely to have led to under-reporting from those dwellings/households more likely to have reported housing quality problems.

Was there good coverage for regions?

Partly. The problems in overall non-response to the 2018 Census will apply to this suite of variables. Dwellings occupied by more deprived households are more likely to have been missed in the census, and therefore under-represented in responses to the suite of housing quality questions.

Stats NZ state "It was expected that an overall higher response rate to the census would likely have given a higher rate of 'damp - always' and 'damp - sometimes', based on our understanding of the characteristics of non-respondents."

3.4.4. Consistency

Was a consistent classification used?

This is a new suite of questions, and so has an associated new classification.

Was data collection consistent across online and paper data collection methods?

Limited information was included in the WoF on response by mode.

3.4.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

There were a limited range of quality assurance comparisons possible with this new suite of questions.

Stats NZ have carried out analyses by region, by tenure, by sector of landlord, etc.

Damp and mould were cross-tabulated as a coherency check. Of those who reported their dwelling was not damp, a very small proportion (0.6 percent) reported they always had mould (and 4.6 percent sometimes had mould). Likewise, of those who reported their dwelling was always damp, 63.0 percent also reported they always had mould. These analyses indicate internal consistency in these questions.

The NZ Deprivation Index (2013) was included as a data quality check. The relationship between deprivation Index decile and (separately) damp and mould is as would be expected – e.g. respondents with the lowest deprivation decile score were least likely to report their home was 'damp - always' (1.0 percent), compared with those with the highest deprivation decile score (6.8 percent), etc.

Is there a consistent time series with previous census, particularly census 2013?
Not relevant. This is a new suite of questions.

3.4.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?
Yes, all data were sourced from the census 2018 individual forms.

3.5. Main types of heating and fuel types used to heat dwellings

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/c821be55-1a9f-4117-a6b7-1091b297b44d/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

3.5.1. Overall assessment

Stats NZ rate the quality of main types of heating and fuel types used as moderate, based on the 7.7 percent with no information. The panel endorse this quality rating.

There has been a change to the concept measured from 'all fuel types that had ever been used to heat the dwelling' to 'main types of heating being used in private dwellings' in response to user needs. Comparisons with previous censuses should be done with care, especially for less common fuel types (such as solar power) which have probably been most impacted by the change in concept. The panel recommend that **Stats NZ should** carry out further analysis of the impact of the change in the question.

Electricity dominates heating types accounting for 75.3 percent in 2018 (very similar to the proportion in 2013). Amongst less common fuel types used there are very large changes which are probably caused by the change in the concept measured. For instance the numbers for solar power have gone from 23,409 in 2013 to 603 in 2018.

3.5.2. Background

Information on main types of heating used is collected on the dwelling form (question 11 on the paper form). Main types of heating used is a new question for the 2018 Census. Previous censuses asked about all fuel types that had ever been used.

There has been a change to the underlying concept collected. The current concept is:

- The main types of heating being used in private dwellings, i.e. the types of heating that the respondent uses the most during periods when heating is required – that they generally rely on and use frequently/usually/most often. It excludes any types of heating that a respondent uses much less often. It excludes any types of heating that are available in the dwelling but are used infrequently or not used at all, disconnected or broken.

The previous concept was:

- All fuel types that had ever been used to heat the dwelling.

Previously information on the appliance used has not been collected. This change in concept is in response to users' indicating the importance of collecting information on heating appliances used. The previous data on all fuel types ever used was considered to add 'noise' into the data, clouding the picture of heating use.

The main types of heating used measures the types of heating that are usually used to heat an occupied private dwelling. It indicates the appliances used, and, in most cases, the fuels

used (e.g. heat pump, fixed gas heater, wood burner). The data on fuel types measures the main types of fuels used to heat an occupied private dwelling

This information is used, for example: to identify areas affected by fuel poverty; for understanding changes in energy demands; to support monitoring against the National Environmental Standards for air quality and regional air quality plans; to track heating use against clear air heating legislation, monitor the effectiveness of energy and carbon-reduction policies and inform air quality policies.

No alternative data source or imputation was used to replace missing responses or responses that could not be classified (Table 3.5).

Table 3.5. Data sources: Main types of heating and fuel types – Occupied private dwellings

	main types of heating	fuel types used to heat dwellings
Source	**Percent**	**Percent**
Response from 2018 Census	92.3	92.3
2013 Census data	0.0	0.0
Administrative data	0.0	0.0
Statistical imputation	0.0	0.0
No information	7.7	7.7
Total	100	100

The 'no information' percentage is where it was not possible to source data for a dwelling in the subject population. The 7.7 percent non-response rate is only slightly higher than the 5.1 percent non-response rate for fuel types used in the 2013 Census.

3.5.3. Coverage

Was there good coverage for the overall population?
Reasonable – 92.3 percent of dwellings were covered.

Was there good coverage for regions?
Comparisons have been made by regional council (e.g. Not using heating appliances is most common in Auckland and Northland; heat pump most common in Canterbury); by tenure; by sector of landlord, by income; etc.

Main types of heating and fuel types is a dwelling attribute and ethnicity is not an attribute of a dwelling. Stats NZ do not produce standard outputs on dwellings on the basis of ethnicity of the residents, so no analysis is available by the ethnic group of residents. Given the variable pattern of non-response to 2018 Census by different ethnic groups there will be

variable levels of coverage for main types of heating used as well, which will affect the quality of the data for this variable.

3.5.4. Consistency

Was a consistent classification used?

Main types of heating used is a new classification for the 2018 Census. This is a multiple response variable. There is a flat classification with the following categories: No heating used; Heat pump; Electric heater; Fixed gas heater; Portable gas heater; Wood burner; Pellet fire; Coal burner; Other types of heating; Not elsewhere included.

Number of heating types used is a new classification for the 2018 Census. There is a flat classification with the following categories: No heating types used; One heating type used etc, up to 'Eight or more heating types used; Not elsewhere included.

Census main types of heating used is a multiple response variable, so a classification to show combined use of heating types in a dwelling is also available. Main types of heating used single/combination is a new classification for the 2018 Census. It is a flat classification with single/combination combinations such as: Heat pump only; Electric heater only; Heat pump and electric heater; Heat pump, electric heater, and wood burner, etc.

Fuel types used in dwellings is a flat classification with the following categories: Electricity, Gas, Wood, Coal, Home heating oil, Solar power, No fuels used in this dwelling, Other fuel(s), Not elsewhere included

The fuel types used in dwelling classification has changed since 2013. Previously there were separate categories for mains gas and bottled gas. These categories have been combined because the information collected in 2018 no longer clearly distinguishes whether gas is mains supplied or accessed via bottles.

Number of heating fuels used is the same classification as used in the 2013 and 2006 Censuses. It is a flat classification with the following categories: No heating fuels used; One fuel; etc through to Seven or more fuels; Not elsewhere included

Was data collection consistent across online and paper data collection methods?

Although the vast majority of responses to these questions were online, the percentages responding online and paper by fuel type were broadly consistent – e.g. heat pumps online were 44 percent compared to 41 percent on paper; wood burner online were 32 percent compared to 31 percent on paper.

3.5.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

No comparisons were made to alternative sources of information.

Is there a consistent time series with previous census, particularly census 2013?

There has been a change to the concept measured from 'all fuel types that had ever been used to heat the dwelling' to 'main types of heating being used in private dwellings' in response to user needs. Comparisons with previous censuses should be done with care,

especially for less common fuel types which have probably been more impacted by the change in concept.

Electricity dominates fuel types used accounting for 75.3 percent in 2018 compared to 79.2 percent in 2013 and appears comparable to previous census results. Amongst less common fuel types used there are very large changes which are probably caused by the change in the concept measured. The numbers for solar power have gone from 23,409 in 2013 to 603 in 2018.

The 'Response unidentifiable' category has increased from 6,813 in 2013 to 17,565 in 2018 and 'not stated' has increased from 79,578 in 2013 to 128,913 (7.7 percent of all responses) in 2018. The 2013 Census non-response rate for fuel types used was 5.1 percent. This increase since 2013 will have impacted on comparisons over time and probably had a bigger impact on the data for less common sources of fuel.

3.5.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Yes, all data were sourced from the census 2018 dwelling forms.

3.6. Number of bedrooms and number of rooms

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/43d10e9f-1dec-4d71-b7c9-2a52a7e902ba/

EDQ Panel rating Number of Bedrooms: **High**
EDQ Panel rating Number of rooms: **Poor**
Stats NZ rating Number of Bedrooms: **High**
Stats NZ rating Number of rooms: **Poor**

3.6.1. Overall assessment

The number of bedrooms data is high quality at the National level and comparable with 2006 and 2013 data. However the quality of the data has not been assessed at lower levels of geography.

Number of rooms is poor quality and shows a systematic shift in the distribution towards dwellings with more rooms in comparison with 2013 Census and is not comparable with 2006 and 2013 data at even the National level. This is mainly the result in a change to the question in 2018, which led to a more accurate count of rooms for some respondents, but also created some issues with the online form which led to incorrect counts of very high numbers of rooms.

3.6.2. Background

This information is used, for instance, to derive household crowding measures and to estimate future demand for housing. The number of bedrooms is used in the NZ Deprivation Index, so this variable is more important than the total number of rooms variables.

Table 3.6.1 shows the various data sources used for this variable.

Table 3.6.1. Data sources: Number of bedrooms – Occupied private dwellings

	Number of bedrooms	Number of rooms
Source	**Percent**	**Percent**
Response from 2018 Census	91.1	91.1
2013 Census data	3.4	5.2
Administrative data	2.6	0.0
Statistical imputation	2.8	3.7
No information	0.1	0.1
Total	100	100

The ‘no information’ percentage is where it was not possible to source number of bedrooms or number of rooms data for a dwelling. The 2013 ‘not stated’ rates were 5.8 percent for rooms and 5.1 percent for bedrooms.

Admin data from Housing New Zealand Corporation and Tenancy Bonds (MBIE) were used for number of bedrooms.

3.6.3. Coverage

Was there good coverage for the overall population?

Yes. The majority (91 percent) of the rooms data came from 2018 Census Dwelling Forms. Stats NZ has been able to fill in gaps reliably for number of bedrooms and number of rooms. The 2013 Census provides a good source of information for both variables. Admin data is available for number of bedrooms, but no admin data sources are available for number of rooms.

Was there good coverage for regions?

Not relevant, other than for information on number of bedrooms/rooms by region. No analyses by level of geographic have been undertaken to allow an assessment of quality by region.

The quality ratings for number of bedrooms at the Regional Council level gives a result of Very High Quality for all Regional Councils, apart from Northland, Bay of Plenty, Gisborne and West Coast, which are rated High.

Given the variable response rates for 2018 Census from the major ethnic groups, data relating to these variables need to be used with considerable caution, especially for Māori and Pacific peoples.

3.6.4. Consistency

Was a consistent classification used?

The classifications of number of bedrooms and number of rooms in the 2018 Census are consistent with the classifications used in the 2013 and 2006 Censuses. The rooms question was asked in a different way in 2018 (all room types) compared to 2013 (two questions on the number of bedrooms and number of rooms).

Number of bedrooms is most often grouped for output with an upper category of five or more. Most occupied private dwellings have fewer than five bedrooms.

Number of rooms is most often grouped for output with an upper category of eight or more. Most occupied private dwellings have fewer than eight rooms.

Was data collection consistent across online and paper data collection methods?

No. There were problems with the online form that impacted on data quality. On the online forms respondents did not always press ‘tab’ to enter the next field before entering their next room count response, resulting in two responses appearing in the boxes for one type of room and a very high count for example a response of 21 bedrooms when the intended response was two bedrooms and one lounge. Edit and processing checks, and the use of

2013 Census and admin data have addressed these shortfalls for number of bedrooms, but Stats NZ were not able to fix the number of rooms/other rooms data in the time available.

3.6.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

Stats NZ have compared number of bedrooms against the 2016 General Social Survey. There is some inconsistency, however Census 2006 and Census 2013 bedroom proportions were also different (and the census data was deemed to be high quality and acceptable for those years), so this does not undermine the quality of the bedrooms data.

Is there a consistent time series with previous census, particularly census 2013?

Yes for bedrooms, no for other rooms.

Table 3.6.2 shows that the 2018 number of bedrooms - proportions are consistent with the 2013 Census.

Table 3.6.2. Distribution of number of bedrooms in 2013 and 2018 censuses – Occupied private dwellings

Number of bedrooms	Count 2013	Count 2018	2013 (percent)	2018 (percent)
One Bedroom	84,138	102,831	5.7	6.2
Two Bedrooms	283,008	318,492	19.1	19.1
Three Bedrooms	659,529	723,327	44.5	43.5
Four Bedrooms	346,941	396,981	23.4	23.9
Five or more bedrooms	108,552	121,692	7.3	7.3

The number of rooms are not consistent with the 2013 Census. There is a systematic shift in the distribution towards higher number of rooms which would be expected given the change to the question. However the large increase in the number of dwellings with 20 or more rooms is not in line with expectations.

3.6.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

The majority of data (91 percent) were from the 2018 Census dwelling form. 5-6 percent was from 2013 Census and admin data, but number of bedrooms/rooms is not a feature that changes rapidly over time (only with redevelopment etc). Use of these good quality sources has improved the overall data quality in comparison with the alternative of leaving data as missing, which is what we did in previous censuses.

3.7. Number of motor vehicles

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/e67ec537-fcbf-4858-8e89-cb23f83870c0/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

3.7.1. Overall assessment

Stats NZ rate this variable as of Moderate quality, based on the 92.2 percent response rate (7.8 percent no information). The panel endorse this rating.

The WoF states "The lower response rate to the motor vehicle question could have caused response bias towards more affluent households, and this could have contributed towards the increase in the three or more motor vehicle category. The decrease in the no motor vehicle and one motor vehicle categories could be reflective of the undercount of lower socio-economic households who were missed in the 2018 census and are more likely to be no motor vehicle access households."

Stats NZ state "Caution should be taken when looking at exact counts of motor vehicles in some smaller areas, as response rates may be lower in some areas compared with the national average."

The classification changed slightly for 2018, with 'three or more' replaced by three, four, five or more'. Users should use care when comparing this end of the distribution over time.

3.7.2. Background

Number of motor vehicles is used by local government and transport planners to plan public transport services, and in the New Zealand Deprivation Index.

Table 3.7 shows that there were no alternative data sources or imputation used to replace missing responses or responses that could not be classified for number of motor vehicles.

Table 3.7. Data sources: Number of motor vehicles – Households in occupied private dwellings

Source	Percent
Response from 2018 Census	92.2
2013 Census data	0.0
Administrative data	0.0
Statistical imputation	0.0
No information	7.8
Total	100

The ‘no information’ percentage is where it was not possible to source data for a household in the subject population. The ‘no information’ category increased from 77,800 in 2013 (5.0 percent) to 129,000 in 2018 (7.8 percent), which will have impacted some comparisons over time.

Stats NZ state that “there are … valid responses [on paper] which have been coded to response unidentifiable [due to] scanning misrecognition from instances where respondents have not written a numeric response, crossed out their answer and written a new response, or circled their numeric answer.”

This question refers to motor vehicles that are available for private use by the usual residents of private dwellings. These vehicles must be mechanically operational, but not necessarily licensed or with a current warrant of fitness.

Motor vehicles includes: business vehicles available for private use by people in the dwelling; cars, four-wheel drive vehicles, station wagons, trucks, vans, and other vehicles used on public roads; hired or long-term leased vehicles; vehicles temporarily under repair. It does not include vehicles used only for business.

3.7.3. Coverage

Was there good coverage for the overall population?
Reasonable – 92.2 percent of dwellings were covered.

Was there good coverage for regions?
The distribution of number of motor vehicles by Regional Council appears consistent with previous censuses and national trends (e.g. reduction in percentage with one motor vehicle). Stats NZ have analysed the data at Territorial Authority where, again, the data appears consistent with previous censuses and national trends.

As in 2006 and 2013, Wellington has the highest percentage of households with no access to motor vehicles (10.4 percent in 2018 compared to 11.7 percent in 2013).

3.7.4. Consistency

Was a consistent classification used?
Although there have been no conceptual changes to this variable, there have been minor changes to the classification since the 2013 Census: ‘three or more motor vehicles’ has changed to ‘three motor vehicles’; additional categories added for ‘four motor vehicles’, and ‘five or more vehicles’.

Number of motor vehicles is a flat classification with the following categories: No motor vehicles; One motor vehicle etc up to five or more motor vehicles; Not elsewhere included

Was data collection consistent across online and paper data collection methods?
There were no differences between the wording or question format in the online and paper version of this question. There were differences in the way a person could respond:

- On the online dwelling form only one response could be selected, and the numeric response box accepted values of up to 99

- On the paper dwelling form responses outside the valid range and multiple responses were possible. These were resolved using edits.

3.7.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

There was no comparison against any alternative collections for this variable. The census is the most reliable source.

Is there a consistent time series with previous census, particularly census 2013?

The patterns in 2018 appear broadly consistent with 2013, but with a decline in 'one motor vehicle' (553,000 to 515,000) and significant increases in 'two motor vehicles' (565,000 to 598,000) and 'three or motor vehicles' (237,000 to 304,000).

3.7.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

Yes, all data were sourced from the census 2018 dwelling forms.

3.8. Tenure of household

DataInfo+ link: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/4c68913e-f620-4a6e-8bd3-af2019269a27/

EDQ Panel rating: **Moderate**

Stats NZ rating: **Moderate**

3.8.1. Overall assessment

The Panel had a range of views about the 'Moderate' quality rating for this variable.

There are changes in the trends between 2006, 2013, and 2018 Censuses that might be a result of new questions and the use of alternative data, but may also be the result of independently observable changes. Stats NZ state that "The overall quality of the data is sound and the use of alternative sources means there is no missing data. 2018 can be compared with 2006 and 2013 data with discretion, noting the change in question and use of alternative sources in 2018 and the impact of missing data in previous censuses."

The use of admin data to identify households who rent their home has improved. Likewise the use of information on sector of landlord in the derivation for tenure of household is new to 2018 and will have improved the derivation of households in the 'do not own or hold in a family trust' category. These changes may have impacted on trends since 2013 at this level.

The combined proportion of owned/trust vs not owned/trust is unchanged between 2013 and 2018 at 65 percent/35 percent. What has changed is the split between owned and family trust, which has gone from 77 percent v 23 percent in 2013 to 79 percent v 21 percent in 2018. This shift is behind the increase in 'owned' dwellings. Since 91.5 percent of those are from census forms, the change to the question is likely to be the main driver of the shift in balance of owned and trust. Similarly, the question change is likely to have the biggest impact on the shift in proportions paying mortgages.

There is a relatively small increase in non-response for this question from 5.1 percent in 2013 to 8.5 percent in 2018. The use of alternative data sources, even if they are not perfect, may well lead to better data when used for the 8.5 percent in 2018 than the 2013 situation where 5.1 percent were left as missing.

The panel cannot distinguish how much of the estimated change in tenure of household in 2018 is due to the change in question form; the use of alternative data sources in 2018; bias in 2013 due to missing data; or might be driven by actual changes. Stats NZ note "The 2016 GSS [General Social Survey], saw an increase in the proportion households that 'own/partly own' between 2012 and 2016 (results for 2018 are not yet available). The increase in the proportion of households that 'own/partly own' may be partially explained by real-world changes, including access to KiwiSaver funds and lower mortgage rates bringing more new home owners into the owning categories."

Further analysis of the impact of the change in the question format, and the impact of non-response bias in 2013 should be undertaken by Stats NZ. The panel recommend that **Stats**

NZ should carry out further assessment of tenure of household; an informed analysis of recent changes in the financial, demographic and economic influences on the proportion of households that 'own/partly own' would help support or challenge the reliability of this important measure.

3.8.2. Background

Tenure of household information is used for monitoring trends and changes in home ownership rates, for formulating and monitoring of housing policy by central and local government, and in constructing the New Zealand Deprivation Index.

Table 3.8 below shows the various data sources used for this variable.

Table 3.8. data sources: Tenure of household – Households in occupied private dwellings

Source	Percent
Response from 2018 Census	91.5
2013 Census data	2.9
Administrative data	2.7
Statistical imputation	2.9
No information	<0.1
Total	100

Census responses were obtained for 91.5 percent of households, with 8.4 percent of data from the 2013 Census, administrative sources, and imputation. The '<0.1 percent no information' is where no information could be sourced. For comparison the rate of 'not elsewhere included (which covers 'not stated' and 'no valid response') was 6.3 percent in 2013 and 6.2 percent in 2006.

The following administrative sources were used to identify households as renting: Housing New Zealand Corporation and Tenancy Bonds (Ministry of Business, Innovation and Employment). Any dwellings determined to be active in Tenancy Bonds or HNZC data at 31 March 2018 were assigned as renting – however not everyone files a tenancy bond record or rents with Housing NZ. The 2013 Census was then used for missing data where possible. The 2013 census provided information on tenure categories except for mortgage payments. Nearly all the remaining 3 percent of households were assigned an imputed tenure value except for mortgage payments.

Other data sources were not able to fill the gaps for mortgage payments information for households that 'own/partly own' or 'hold in a family trust' and there is thus an increase in the proportion and counts of households in the 'mortgage arrangements not further defined' categories compared to previous years.

There have been changes to the questions used to derive tenure of household for 2018 compared with 2013. In 2018 questions on home ownership and family trusts were combined in one question; in 2013 they were separate questions. In 2018 mortgage payments questions for owned dwellings and dwellings in a family trust were combined in one question; in 2013 these were separate questions.

In 2018, information from sector of landlord was used in the derivation for tenure of household to improve identification of households in the 'do not own or hold in a family trust' categories, many of whom rent their home.

3.8.3. Coverage

Was there good coverage for the overall population?

Yes – there was high 2018 Census coverage of this variable. The use of alternative data sources and imputation means that the tenure of household variable has virtually no non-response.

The admin data used is likely to be of high quality. The use of 2013 Census data will not capture changes in tenure over the period between censuses. Statistical imputation will be better than no response but will have introduced some uncertainty.

Was there good coverage for regions?

As there is almost no missing data, coverage is very high across all ethnic groups and regional sub-groups. Data has been checked to territorial authority and Auckland local board level.

3.8.4. Consistency

Was a consistent classification used?

Yes. The classification of tenure of household in the 2018 Census is consistent with the classification used in the 2013 and 2006 Censuses.

Tenure of household is a hierarchical classification with two levels. Level 1 contains 4 categories (Dwelling owned or partly owned), Dwelling not owned and not held in a family trust, Dwelling held in a family trust, Not elsewhere included). Level 2 contains 11 categories.

Was data collection consistent across online and paper data collection methods?

Tenure of household data is derived from the following questions on the dwelling form: dwelling owned or in family trust, sector of landlord, rent indicator, rent amount (from which weekly rent paid by household is derived) and mortgage payments.

There were differences between the wording and question format online and on paper:

- On the paper dwelling form it was possible to answer all, or any combination, of the questions used to derive tenure of household and it was possible to give a rent amount higher than $99,999, although there was only space for five digits. These responses were resolved by edits.

- The online form had automatic routing that only showed questions based on previous responses (e.g. the rent amount question was only shown if the response to the rent indicator question was 'yes').

3.8.5. Comparability

How does census 2018 data compare to recent collections of the same variable?

The 2016 General Social Survey (GSS) saw (2012 to 2016) a similar increase to the census in the proportion of households that 'own/partly own'. 2018 results are not yet available.

Is there a consistent time series with previous census, particularly census 2013?

Broadly. Stats NZ state that the "overall quality of the data is sound and can be compared with 2006 and 2013 data with discretion."

3.8.6. Contemporaneity

Were all data sources used for the variable obtained at the same time?

No. 97 percent of the data was sourced from the 2018 Census, imputations, or contemporaneous admin sources (the admin data covered the period up to June/July 2018). 2.9 percent of the records were derived from the 2013 Census.

References

Census External Data Quality Panel. (2019) Initial Report of the 2018 Census External Data Quality Panel. Statistics New Zealand: Wellington. Available from: https://www.stats.govt.nz/reports/initial-report-of-the-2018-census-external-data-quality-panel

Statistics Canada (2017). Statistics Canada's Quality Assurance Framework (Third Edition, release date April 21, 2017). ISBN 978-0-660-08114-4

Stats NZ (2014). Understanding substitution and imputation in the 2013 Census. Available from www.stats.govt.nz.

Stats NZ (2019a). Data quality ratings for 2018 Census variables. https://www.stats.govt.nz/methods/data-quality-ratings-for-2018-census-variables

Stats NZ (2019b). Data sources, editing, and imputation in the 2018 Census. Retrieved from www.stats.govt.nz.

Appendix 1 – Stats NZ data quality assurance definitions for 2018 Census

Stats NZ's Data quality assurance for 2018 Census outlines the quality assurance framework and quality rating scale used by Stats NZ to assess the quality of data from the 2018 Census to determine whether it is fit for purpose and suitable for release. The following are excerpts from this report.

The 2018 quality rating scale is made up of three metrics:

- metric 1 – data sources and coverage
- metric 2 – consistency and coherence
- metric 3 – data quality.

An overall variable rating was assigned to each by taking the lowest score that variable has received from the three metrics, across the range.

Metric 1: Data sources and coverage

This metric calculates a score by rating the overall quality of the data sources used for a census output of a variable. This aims to:

- give customers clarity around what sources have gone into the combined output for a census variable
- show how the rating given to a source (which is based on the quality of the source) will then impact the total score (and quality) of a variable
- calculate an approximation of 'missingness' and uncertainty of output values for a census variable.

To calculate a score for a variable, each source that contributes to the output for that variable is rated and multiplied by the proportion it contributes to the total output.

The rating for a valid census response is defined as 1.00. Ratings for other sources are the best estimates available of their quality relative to a census response.

We calculated the ratings for admin data sources by comparing the 2018 Census received responses with the data from admin source, with a value being derived from the match rate between the two sources.

Bands for data sources and coverage ratings

The bands used for metric 1 are similar to those used in the 2013 Census metric for non-response:

Very high	0.98–1.00
High	0.95–< 0.98
Moderate	0.90–< 0.95
Poor	0.75–< 0.90
Very poor	0.00–< 0.75

Metric 2: Consistency and coherence

Stats NZ rated the level of consistency and coherence in the data on:

- comparability with the expected trends
- comparability with other sources
- contribution of other sources to the census data for this variable.

The ratings account for changes occurring for variables in the 2018 Census as a whole, including the use of admin data and, in some cases, the change in question or concept. In some cases, 2018 Census data may be moving away from expected time series trends, due to methodological changes that have brought the data closer to the 'real world' situation, by addressing historic issues, or biases within census coverage.

For new or changed variables where there is no previous census data for comparison, we used other data sources and expectation reports as the primary source of comparison. These may only be comparable at a national level.

Explainable change (see 'moderate' ratings below) could be the result of real-world change, incorporation of other sources of data, or a change in how the variable has been collected.

Priority 1 variables were assessed for consistency:

- at level 1 of the classification by territorial authority (TA) compared with the benchmarks
- at the lowest level of classification (if applicable) at a national level.

Priority 2 and 3 variables were assessed for consistency:

- at level 1 of the classification by regional council (RC)
- at the lowest level of classification (if applicable) at a national level.

Five detailed descriptions guided their assessment and categorisation of variables for this metric:

Very high	Variable data is highly consistent with expectations across all consistency checks.
High	Variable data is consistent with expectations across nearly all consistency checks, with some minor variation from expectations or benchmarks that makes sense due to real-world change, incorporation of other sources of data, or a change in how the variable has been collected.
Moderate	Variable data is mostly consistent with expectations across consistency checks. There is an overall difference in the data compared with expectations and benchmarks that can be explained through a combination of real-world change, incorporation of other sources of data, or a change in how the variable has been collected.
Poor	Variable data is not consistent overall with expectations across one or more consistency checks. There is an overall difference in the data compared with expectations and benchmarks. Where

	this difference occurs, this cannot be fully explained through likely real-world change, incorporation of other sources of data, or a change in how the variable has been collected.
Very poor	Variable data is highly different from expectations across all consistency checks. There is a large overall difference in the data compared with expectations and benchmarks that cannot be explained through real-world change, incorporation of other sources of data, or change in how the variable has been collected.

Metric 3: Data quality

This metric relates to the data produced from the census forms received and from other data sources. This includes aspects such as coding, level of detail/classification, accuracy of responses, and any other specific quality issues that may have been identified in problem reports.

Stats NZ used the same overall approach that was used in 2013 for this metric. The ratings are:

Very high	Data has no data quality issues that have an observable effect on the data. The quality of coding is very high. Other data sources used do not create any quality impacts for this variable. Any issues with the variable appear in a very low number of cases (typically less than a hundred).
High	Data has only minor data quality issues. The quality of coding and responses within classification categories is high. Any impact of other data sources used is minor. Any issues with the variable appear in a low number of cases (typically in the low hundreds).
Moderate	Data has various data quality issues involving several categories or aspects of the data, or an entire level of a hierarchical classification. The data quality issues could include problems with the classification or coding of data, such as vague responses resulting in coding issues, or responses that cannot be coded to a specific (non-residual) category, thereby reducing the amount of useful, meaningful data available for analysis. The use of other data sources may be contributing to these issues.
Poor	Significant data quality issues emerged during evaluation. Data is considered fit for use but there are limitations on how it can be used and interpreted. There are significant issues with respondent interpretation, coding, and/or classification problems.
Very poor	Major data quality problems exist. Data does not reflect reality due to respondent misinterpretation, coding and/or classification problems.

Appendix 2 – Glossary

2013 Census	Census of Population and Dwellings undertaken on 5 March 2013. For some 2018 census topics, responses from the 2013 census were used to fill in missing data.
Absentee	A person who is identified on the census online household set-up form or paper dwelling form as usually living in a particular dwelling but who did not complete a census individual form at that dwelling because they were elsewhere in New Zealand or overseas at the time of the census.
Administrative (admin) data	Data collected by government or other organisations for non-statistical reasons, such as births, tax, health, and education records. These are typically records describing events or interactions with government agencies and have been obtained in the course of some statutory obligation or service provided by a government agency.
Auckland Local Board	Statutory community-level governance districts within Auckland Council. There are 21 local boards: Albert-Eden, Devonport-Takapuna, Franklin, Great Barrier, Henderson-Massey, Hibiscus and Bays, Howick, Kaipātiki, Māngere-Ōtāhuhu, Manurewa, Maungakiekie-Tāmaki, Ōrākei, Ōtara-Papatoetoe, Papakura, Puketāpapa, Rodney, Upper Harbour, Waiheke, Waitākere Ranges, Waitematā, Whau.
CANCEIS	Canadian Census Edit and Imputation System. A method for 'imputing' (filling-in) data for missing responses/respondents. Used by a number of national statistical institutes for census imputation.
Census usually resident population count	A count of all people who usually live in New Zealand and were present somewhere in New Zealand on census night.
Classification	System of categorising the responses to questions that are not values. Many census variables use standard classifications systems (e.g., birthplace, ethnicity, occupation). The classifications used for census variables may differ from the classifications used for the equivalent administrative variable.
CURF	Clean unit record file. A finalised approved data file made available for reporting and analysis where responses have been validated and the available information will meet the confidentiality protection requirements

DataInfo+	An online repository of Stats NZ metadata (http://datainfoplus.stats.govt.nz/). The repository includes descriptions of census 2018 variables as well as assessments of their quality – a list can be found here: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/ca28210f-3fd6-415c-a162-ecc07b4a28b0#/nz.govt.stats/2ae40a5d-64c8-4704-9829-45f802d78c6c/51.
Donor imputation	Method of imputation which uses data from similar individuals or households to 'impute' (fill-in) data for missing responses/respondents
Don't know	A response given when the respondent does not know, or cannot give, an appropriate response. 'Don't know' may be a legitimate response to certain questions.
Dwelling	A building or structure using for habitation, e.g., houses, motels, hotels, prisons, rest-homes
Dwelling form	Census questionnaire with information on the dwelling. For paper forms this includes a listing of people within the dwelling and their relationship to the person completing the dwelling form. See household set-up form.
Ethnicity	A measure of cultural affiliation. It is not a measure of race, ancestry, nationality, or citizenship. Ethnicity is self-perceived and people can belong to more than one ethnic group. Stats NZ uses a hierarchical classification system for ethnicity, with • 6 categories at 'Level 1': European; Māori; Pacific; Asian; Middle Eastern, Latin American and African (MELAA); Other; • 21 categories at 'Level 2', including New Zealand European; Samoan; Chinese; Middle Eastern; • 36 categories at 'Level 3', including South Slav; Filipino; • 180 categories at 'Level 4', including Serbian; Tahitian; Malay; Kenyan; Indigenous American.
Family	A couple, with or without child(ren), or one parent with child(ren), usually living together in a household. Related people, such as siblings, who are not in a couple or parent-child relationship, are therefore excluded from this definition.
Household	One person who usually resides alone, or two or more people who usually reside together and share facilities (such as eating facilities, cooking facilities, bathroom and toilet facilities, and a living area), in a private dwelling.
Household set-up form	Online census form containing a listing of people within the household and their relationship to the person completing the household set-up form.

Integrated Data Infrastructure (IDI)	A large database maintained by Stats NZ. It contains de-identified data about people and households sourced from government agencies (i.e., administrative data), 2013 Census, Stats NZ surveys, and non-government organisations (NGOs). Data from different sources are linked together, typically at the individual (person) level.
IDI Spine	The primary person-level dataset in the IDI to which all other person-level datasets are linked. The current (prototype) spine used in the IDI is formed by linking together tax (IRD) records since 1999, New Zealand birth records from 1920, and long-term visa records from 1997.
Imputation	The process of replacing missing data with estimated values through statistical methods. For the 2018 Census, the method for estimating values was nearest-neighbour imputation methodology (NIM), which 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. See CANCEIS.
Individual form (or questionnaire)	Census questionnaire to be completed by each person in a dwelling. This includes questions about ethnicity, education, income, etc. pertaining to the individual.
IRD	Inland Revenue Department
Iwi	Māori tribe or extended kinship group, often descended from a common ancestor and/or associated with a distinct territory.
Level 1 (2,3,4) Ethnicity	See Ethnicity
Linkage	The process of combining two or more data sets so that a data set with more information can be created which can then usually be used as though the information came from the same source.
Māori descent output (variable)	Census variable that assesses the Māori descent population in New Zealand. For 2018, valid responses were "Yes", "No", and "Don't know". For 2018, data from other sources were used when a response other than "Yes", "No" or "Don't know" was given.
MELAA	Middle Eastern, Latin American and African: A grouping at Level 1 of the ethnicity classification
Meshblock	The smallest geographic units for which statistical data are reported. These vary in size from part of a city block to a large area of rural land, with an ideal size range of 30–60 dwellings (around 60–120 residents).
MOH	Ministry of Health

MSD	Ministry of Social Development
No information	Where data could not be sourced (from a response to the 2018 Census, 2013 Census data, administrative data or from statistical imputation) for units in the subject population of a variable. For example, where the number of children born could not be sourced for a female in the census usually resident population aged 15 years and over.
Non-private dwelling (NPD)	A dwelling providing communal or transitory type accommodation (e.g., hotel, campground, prison, defence barrack, rest home, university hall of residence).
Not elsewhere classified (nec)	A residual category for responses that have no appropriate category, because they are infrequent or unanticipated. These categories never appear within classifications as stand-alone descriptors, but are combined with descriptors, often taken from a higher level in the classification. For example, for Qualifications, BSc Environmental Biology would go to Biological Sciences nec.
Not elsewhere included (nei)	Used in some outputs for a combination of residuals, such as 'not stated', 'response outside scope', 'response unidentifiable', 'refused to answer', and 'don't know'. This item should have a footnote indicating its composition.
Not further defined (nfd)	A residual category used in hierarchical classifications for responses containing insufficient detail to be classified to the most detailed level of a classification, but which can be classified to a less detailed category further up the hierarchy.
Not stated	A category used when a person gave no response to a question relevant to them or when there was no alternative data source for that information, such as a 2013 Census response, administrative data or statistical imputation
Private dwelling	A dwelling accommodating one or more people who usually live independently within the community (e.g., a house or flat)
Refused to answer	A category used only when it is known that a person has purposefully chosen not to respond to the question.
Region	The first tier of local government. There are 16 regions: Northland, Auckland, Waikato, Bay of Plenty, Gisborne, Hawke's Bay, Taranaki, Manawatu-Wanganui, Wellington, Tasman, Nelson, Marlborough, West Coast, Canterbury, Otago, Southland
Response	Completion of some or all items on a census form. In line with international practice, a census 'response' in 2018 was achieved when the minimum information to count a person was received. Thus, the listing of an individual on a dwelling

	form was considered a response, even if no individual form was received for that individual.
Response outside scope	A category applied if the meaning and intent of the response are clear ('positively identified') but clearly fall outside the scope of the classification/topic as defined.
Response rate	Number of census responses expressed as a percentage of the New Zealand Estimated Resident Population (ERP). In the report, 'total response rate' considers both individual and partial responses when calculating response rate (see 'response' above); 'individual response rate' considers just individual responses when calculating response rate; 'partial response rate' considers just partial responses when calculating response rate
Response unidentifiable	A response given that is: • illegible • unclear regarding its meaning or intent. This most commonly occurs when the response being classified contains insufficient detail, is ambiguous, vague or contradictory (for example, when the tick boxes 'yes' and 'no' have both been ticked) • clear and seemingly within the scope of the classification, yet it cannot be coded as a suitable existing option in the classification or code file (such as 'not elsewhere classified' or 'not further defined').
SA1	Statistical Area 1: A geographic unit built by joining meshblocks, with an ideal size range of 100–200 residents, and a maximum population of about 500.
SA2	Statistical Area 2: A geographic unit which aims to reflect communities that interact together socially and economically. In major urban areas, an SA2 often approximates a single suburb, generally with a population of 2,000–4,000 residents. SA2s in district council areas generally have a population of 1,000–3,000 residents. In rural areas, SA2s may have fewer than 1,000 residents if they cover large areas that have sparse populations.
Statistical geography	Classification of places in New Zealand into different levels of geography. The current classification system (SSGA18) provides a range of geographic units from 'meshblock', the smallest geographic unit (roughly 30-60 dwellings) to 'region', the largest geographic unit and top tier of Local Government (e.g., Northland region, Auckland region).
Subject population	The relevant population for a particular variable indicating where a response is expected. This is determined in part by the routing on the questionnaire, for example, the census

	usually resident population is defined as those in New Zealand on census night who gave a usual residence in New Zealand.
Territorial Authority (TA)	The second tier of local government, below regions. There are 67 territorial authorities: 13 city councils (**Auckland**, Hamilton City, Tauranga City, Napier City, Palmerston North City, Porirua City, Upper Hutt City, Lower Hutt City, Wellington City, **Nelson City**, Christchurch City, Dunedin City, Invercargill City); 53 district councils (Far North, Whangarei, Kaipara, Thames-Coromandel, Hauraki, Waikato, Matamata-Piako, Waipa, Otorohanga, South Waikato, Waitomo, Taupo, Western Bay of Plenty, Rotorua, Whakatane, Kawerau, Opotiki, **Gisborne**, Wairoa, Hastings, Central Hawke's Bay, New Plymouth, Stratford, South Taranaki, Ruapehu, Whanganui, Rangitikei, Manawatu, Tararua, Horowhenua, Kapiti Coast, Masterton, Carterton, South Wairarapa, **Tasman, Marlborough**, Buller, Grey, Westland, Kaikoura, Hurunui, Waimakariri, Selwyn, Ashburton, Timaru, Mackenzie, Waimate, Waitaki, Central Otago, Queenstown-Lakes, Clutha, Southland, Gore); and the **Chatham Islands Council**. Six territorial authorities (bolded) are also regions and therefore Unitary Councils.