Post-enumeration survey 2018: Methods and results

23 September 2020, 10:45am

> **25 September 2020**: Please note that since this report was published we have corrected some numbers in the section headed 2018 Census KPI results (methods/post-enumeration-survey-2018-methods-and-results#censuskpi) .

Post-enumeration survey 2018: Methods and results presents the salient features of the 2018 PES, including its scope and methodology. The results, and the information gathered during the survey, are important for post-censal population estimation and other demographic series.

Read this paper online or download the PDF below.

Download document

Post-enumeration survey 2018: Methods and results (/assets/Uploads/Reports/Coverage-in-the-2018-Census-based-on-the-New-Zealand-2018-Post-enumeration-Survey/Downloads/Post-enumeration-survey-2018-Methods-and-results-Stats-NZ.pdf)

Portable Document Format File, 5 MB

() Purpose and summary

Purpose

The report presents background information, methods, and the main results from New Zealand's fifth post-enumeration survey (PES).

The survey was conducted shortly after the 2018 Census of Population and Dwellings. As with the first four PESs, which took place in 1996, 2001, 2006 and 2013, the main aim was to gather information on the completeness of census coverage, that is, to estimate how many New Zealand residents had been either missed or counted more than once in the census.

Stats NZ uses PES results to derive the population base for post-censal population estimates and demographic projections. This report provides users of census and population data with important information required to understand both census data and the official population estimates and projections.

The well-known difficulties experienced with the 2018 Census had a major impact on the measurement of census coverage in this census. Results of the PES provide information on one aspect only of the quality of census output and processes, namely coverage. Just as the results from the previous PESs contributed to 2018 Census enumeration plans, the 2018 PES will similarly assist Stats NZ to develop ongoing strategies to improve future censuses.

The contribution of all Stats NZ staff who were involved with the development and implementation of the 2018 PES, the processing and analysis of the PES results, and the preparation of this publication is acknowledged. The cooperation of the New Zealand public in the successful conduct of the 2018 PES is also acknowledged.

() Summary

The report is organised into five chapters, with an appendix that contains links to summary tables (in Excel) and questionnaires. The highlights summarise the survey results and changes to the methodology.

The PES and its outcomes are described in the following chapters:

- Introduction (methods/post-enumeration-survey-2018-methods-and-results#introduction) outlines the significance of the PES, its objective, its history, as well as the known sources of census miscount.
- About the 2018 post-enumeration survey (methods/post-enumeration-survey-2018-methods-and-results#about) presents details on the survey methodology, including the survey population, sample design, data collection procedures and estimation method, as well as sampling and non-sampling errors.
- Key performance indicators (methods/post-enumeration-survey-2018-methods-and-results#key) describes the key performance indicator results.
- Coverage results (methods/post-enumeration-survey-2018-methods-and-results#coverage) describes the survey's main findings. The results include an analysis of differences in census coverage by sex, age, ethnicity, and selected geographic areas.
- Looking forward (methods/post-enumeration-survey-2018-methods-and-results#looking) introduces work that is being done to ensure robust coverage measurement in preparation for the PES following the 2023 Census and to develop ongoing improvements in understanding the importance of coverage.

() Highlights

A post-enumeration survey (PES) was conducted in New Zealand to measure the completeness of census coverage achieved by the 2018 Census of Population and Dwellings, which was held on 6 March 2018. The PES was in the field from 26 April to 2 July 2018.

Coverage results from the 2018 PES showed that:

- the census counted 97.4 percent of people usually resident in New Zealand who were in New Zealand on census night with a credible interval of 97.1–97.7 percent
- national median estimated net undercount was 2.6 percent or 124,800 people
- the median estimated gross undercount was 171,800 people, and the median estimated gross overcount was 46,700 people
- while the 2018 Census usually resident population count was 4,699,755 people, the PES estimated census night population was 4,824,600 (credible interval 4,810,700–4,839,800)
- males had a higher net undercount (3.1 percent) than females (2.1 percent)
- young adults aged 15–29 years had a much higher net undercount (4.7 percent) than other age groups and also had the highest level of uncertainty in this measure (credible interval 4.0–5.4 percent)
- by far the lowest estimated net undercount was among people aged 65 years and over – people aged 65–74 years had a net undercount of 0.4 percent and those aged over 75 years a net undercount of 0.3 percent
- people born overseas were harder to reach in census, reflected in a higher net undercount (3.3 percent) than people born in New Zealand (2.4 percent)

- people of Māori descent had a net undercount of 38,900 (4.2 percent)
- people of Māori ethnicity had a net undercount of 4.4 percent (35,300 people), compared with 4.9 percent for people of Pacific ethnicities (19,600 people) and 3.3 percent for people of Asian ethnicities (24,400 people)
- Auckland Region, home to 34 percent of the population, accounted for over 38 percent of the net census under-coverage
- the total non-enumeration rate in 2018 was 17.7 percent. The total non-enumeration rate is the number of people from whom an individual form was not received plus the net census undercount as measured by the PES. This includes 4.2 percent of the population listed on household listings on the paper dwelling or online household set-up form with no corresponding individual form returned.

Note that as with all data estimated from a survey with a small sample size, the median estimates above are subject to uncertainties. The true value however is expected to lie within a range close to the median. This range is referred to as the credible interval. For more details on credible intervals associated with the figures cited above, refer to the detailed tables in this report. Credible intervals indicate the level of uncertainty in the estimation.

Methodological changes between PESs also mean that comparisons across time should be made with caution. This is especially true of 2018 PES because of the changes in the sources of the content of the 2018 Census and the changes in the definition of a ‘census response’ as discussed in this report.

Major changes in the 2018 PES were introduced to handle the changed data environment. Around 15,100 households were selected to be visited, and PES questionnaires containing information from over 37,500 people were received. After removal of incomplete and ineligible responses, around 31,600 people were included in coverage estimation.

() Introduction

A census of population and dwellings is often the largest data-gathering exercise in any country. In principle, the New Zealand census intends to count everybody in the country on census night, including visitors to New Zealand who are usually resident overseas, but excluding New Zealand residents who are not in the country on census night. Of primary focus is the population usually resident in New Zealand. The census yields a wealth of valuable information for analysing changes in the socio-demographic profile of the population, and for monitoring, planning, and decision-making at the national and local level, by government, businesses, and the general community. It is also integral to the derivation of reliable post-censal population estimates and for charting future demographic trends.

Given the strategic importance of census data and its many diverse applications, Stats NZ, like other national statistical organisations, makes concerted efforts to ensure as high a coverage as possible in its five-yearly census. However, censuses everywhere tend to miss some people or count people more than once.

There are many statistical procedures that demographers and others use to check the accuracy of census coverage (Shryock and Siegel, 1973). These include: checks against demographically derived estimates; comparison of census figures with administrative records and other sources; and a post-enumeration survey (PES). None of these in isolation gives a complete picture of accuracy.

A PES is undertaken shortly after the census to evaluate the completeness of census coverage. It involves an independent re-enumeration of a statistically designed sample of private dwellings and the people within them covered by the national census. Post-enumeration surveys are an essential feature of census-taking in many countries. The 2018 PES was the fifth to be undertaken in New Zealand since the inaugural PES in 1996. The main objective of the 2018 PES was to measure the level of national coverage (undercount and overcount) of people in the 2018 Census.

The ability to complete census forms online was introduced in the 2006 Census, and many more people took advantage of this option in 2013. Electronic forms can include a number of features that contribute to potential improvements in the quality of data collected. It was not clear whether this would also improve census coverage or not. Other census collection initiatives, such as mail-back trialled in some areas in 2013, were designed to reduce the high costs of field collections and may have also impacted on coverage in 2018.

As highlighted in the Report of the Independent Review of New Zealand's 2018 Census (reports/report-of-the-independent-review-of-new-zealands-2018-census) (Jack & Graziadei, 2019), the 2018 Census represented a significant change from its predecessor in 2013. This change was consistent with international developments and the public sector focus on digital services, aligning with Stats NZ's long-term census transformation strategy. While many aspects of the 2018 Census went well, there were key components that were not successful, and unanticipated challenges resulted in a less than optimal outcome. The mitigation methodology put in place had direct impacts on the measurement of census coverage.

As with earlier censuses, incomplete coverage continued to result from, for example, inadvertent omission of young children, difficulty in enumerating people on the move and those living in apartments, the homeless, and people not willing to cooperate with census collectors. These are known limitations of censuses in all countries and they had a direct impact on the measurement of coverage.

() 2018 Census file

For the first time in a New Zealand census, the 2018 Census used data from alternative sources to fill gaps when the characteristics of people or dwellings were not provided on census forms. These alternative sources were the previous census in 2013 and a range of administrative (admin) data sources such as birth registrations and tax information. Where high-quality alternatives were available, these provided real information about a person, but given at a different time or in a different context from the 2018 Census. The use of statistical imputation for remaining missing data was extended to a much greater range of variables, such as ethnicity and religious affiliation, than in previous censuses.

While these improvements were planned in the build-up to the 2018 Census with the aim of reducing bias caused by non-response, they took on much greater significance in light of the lower than expected response rate to the 2018 Census field collection. Administrative records were also used to count people who were missed, referred to in census literature as 'admin enumerations'.

Where people were listed as a member of a household but completed no individual form, and where people were counted through admin enumeration, alternative sources and imputation were the source of nearly all individual census characteristics.

The use of admin enumerations improved the census count over previous censuses for some subgroups of the population, and together with the alternative data sources and imputation, for many variables, helped maintain (or in some cases improve) the quality of information seen in previous censuses. However, for other variables, higher rates of, for example, missing data meant that quality was lower than in previous censuses, and those subgroups with lower response rates to the field collection, such as Māori, Pacific peoples, and young adults were more adversely affected.

The Government Statistician initiated an independent review of the 2018 Census. This review focused on the factors that contributed to the lower-than-expected response rate in order to understand the lessons that could be learned.

See Report of the Independent Review of New Zealand's 2018 Census (reports/report-of-the-independent-review-of-new-zealands-2018-census) (Jack & Graziadei, 2019) for further information.

In addition to this review an external data quality panel (EDQP) of recognised demographic and statistical experts was convened. The three EDQP reports are available on the Stats NZ website:

Initial Report of the 2018 Census External Data Quality Panel (reports/initial-report-of-the-2018-census-external-data-quality-panel)

2018 Census External Data Quality Panel: Assessment of variables (reports/2018-census-external-data-quality-panel-assessment-of-variables)

Final report of the 2018 Census External Data Quality Panel (reports/final-report-of-the-2018-census-external-data-quality-panel)

Stats NZ also provided interim guidance on its understanding of coverage rates, indicating a 1.2 percent net under-coverage based on the initial best estimate in April 2019 (before the census dataset was finalised), but revised later (July 2019) to an implied net under-coverage of 1.4 percent:

- 2018 Census: Interim coverage rates, collection response rates, and data sources (reports/2018-census-interim-coverage-rates-collection-response-rates-and-data-sources)
- A discussion of the difference between these interim coverage assessments, and the official PES net census undercount of 2.6 percent (credible interval 2.3–2.9 percent) is provided below in Interim guidance in 2019 on under-coverage and PES results (methods/post-enumeration-survey-2018-methods-and-results#interim) .

Miscount and its endemic sources

In such a large and complex exercise as a census, it is inevitable that some people will be missed and some counted more than once (Dunstan, Heyen & Paice, 1999). In this section a few of the key sources of miscount – both undercount and overcount – are listed.

For PES purposes, an error in recording the geographic location of a person does not constitute a coverage error unless the individual was counted more than once. For example, a person who is enumerated in the wrong area, but counted only once, is neither included as over-coverage for the area in which the person was enumerated nor as under-coverage for the area in which the person should have been counted.

Persisting reasons for people and dwellings being missed are many and varied, and include:

- people deliberately avoiding the census – refusing or unwilling to respond (for example, for fear that information given will be used against their interests)
- people being reluctant to open their door to strangers
- people shifting from one house to another around the time of the census
- multiple households living at the same address
- people being away temporarily (for example, work, school)

- people having no usual residence (for example, homeless people)
- newborn babies being overlooked
- recent migrants not familiar with their obligations
- dwellings entirely missed by census collectors
- occupied dwellings misclassified as vacant.

Conversely, there are situations in which people can be counted more than once. They may have completed more than one form in different locations, or completed their own form while someone else may also have inadvertently completed a form on their behalf, or they may have completed an online response correctly but completed a further form when prompted by census follow-up procedures. This situation can happen for many reasons, including:

- students living away at school or university, and also being counted at the home of their parents
- children in joint custody where both custodial parents complete a form for the same child
- people counted while away from home and another form is completed at their home
- people shifting from one house to another around the time of the census and completing forms at both addresses
- people living in institutions, as It is common for people to be enumerated both in an institution and elsewhere, especially when tenure in the institution is short
- people with more than one usual residence
- erroneous inclusion of people who have died
- babies born after census night
- residents temporarily overseas on census night who return soon after census and complete a form when they should not have done so
- people departing overseas in the lead-up to census and completing a form prior to departure
- forms completed for family pets, etc, that were not picked up during census processing.

New sources of miscount in 2018 Census

Changes in the 2018 Census methods (specifically the inclusion of admin enumerations) introduced additional sources of miscount. An understanding of these additional sources will become increasingly important for future censuses.

An additional source of missingness due to 2018 Census methods was:

- people who should be included in the population but were not selected for inclusion due to various rules set to determine eligibility for admin enumeration inclusion.

Additional sources of overcount due to 2018 Census methods were:

- erroneous inclusions of people in the admin population who were not in the target population (such as people in NZ on census night but not identifying as usual residents)
- people included as an admin enumeration in addition to a census response (duplication as a result of linkage error between census responses and the admin population).

In addition to changes in sources of miscount, several other changes in 2018 and resulting data, also had the potential to impact on coverage patterns, including:

- the move to a 'digital-first' collection model, with the majority of households encouraged to complete the census online (this had the potential to alter coverage patterns, particularly for households without easy online access)
- insufficient provision of paper forms and heavily reduced field workforce, resulting in difficulties completing the census for targeted sub-population groups (this had the potential to exacerbate traditional under-coverage patterns)
- the introduction of repatriation in census processing (people away from home and enumerated in New Zealand were repatriated to their usual dwelling where possible). This enabled census to reduce potential overcount through identifying duplicate records for people away from home on census night, and repatriation also had the potential to improve PES processing and linking of PES people with census records due to people being at their usual residence.
- where people are unable to be repatriated to their usual dwelling (and might instead be repatriated to a small geography instead), there remained a potential risk of overcount if a duplicate record in census went undetected
- an increased number of 'partially' responding households (in which people were listed on a dwelling or household summary form but did not complete individual forms). This had the potential to impact on coverage measures due to reduced information on the census side for PES to link to.

Why measure undercount

The principal objective of a census is to count everybody. This is, in practice, extremely difficult. How the actual count differs from the real count is important, so there is interest in accurately determining the level of coverage of the census. Accurate counts are required for many planning and research purposes. Measures of the difficulty of enumeration also contribute to an understanding of underlying social concerns.

Similarly, measurement of coverage helps improve both census processes and the general quality of post-censal population estimates. Population estimates are used for a variety of purposes, including:

- to allocate funds to organisations using population-based weightings

- as denominators for the calculation of rates (for example, birth and death rates) and per capita time series
- to determine population weights for various surveys
- for administration, policy-making and planning, by both central and local government
- in demographic, social and economic studies
- as a starting point for demographic projections.

People who are geographically mobile, disadvantaged, or in particular age groups, especially in areas that are difficult to enumerate, are most likely to be missed in the census (O'Hare, 2019). A PES in its current design and scope is not amenable to measuring many of these features and how they interact with each other. However, results from previous post-enumeration surveys and indirect evidence drawn from demographic analysis, along with administrative sources such as birth registrations and school enrolments, suggest that the census does miss some people and that this under-enumeration varies across different groups and by both age and sex.

PES in New Zealand

Post-enumeration surveys have been carried out for more than 60 years in a number of countries that conduct a census as a single point-in-time survey. Measuring the effectiveness of census coverage is considered to be very important internationally.

One of the earliest PESs specifically designed to evaluate a population census was carried out in the United States of America in 1950 (Marks et al, 1953). This PES grew out of earlier surveys used to measure coverage in the 1945 Census of Agriculture, 1947 Census of Manufactures and the 1948 Census of Business, and lessons were applied to a coverage survey for the 1950 US Census. The stated objectives were to measure the extent of undercount and also to attempt to define some of the reasons why censuses miss people or count people more than once. One of the concerns was the early recognition that some groups such as Black Americans had long been severely undercounted (Miller, 1922) and that this was a sub-population of major policy interest. A PES remains a cornerstone of census evaluation in the US with continuing estimation of subpopulations by race (and more recently ethnicity), age, and sex. Since this first PES, the majority of countries with traditional population censuses have developed their own PESs. This indicates that the concerns of interest to enumerators in New Zealand today are similar to concerns wherever censuses are taken.

Before the 1996 Census, Stats NZ (then the Department of Statistics) had evaluated aspects of the general quality of census data but had not attempted systematically to measure the level of undercount or overcount directly via a post-censal coverage survey. The value of a survey to measure coverage, based on work in the US referenced above, was recognised in the 1970s and plans were drawn up to do so with the 1976 and 1981 censuses but the surveys never took place. In 1990, in association with the 1991 census planning, a pilot test as a precursor to a PES was conducted, but the survey did not go ahead. In 1994, appropriate funding was approved by the

Government to allow Stat NZ to undertake a PES in conjunction with the 1996 census. The 1996 PES was the first survey of its type in New Zealand.

The inaugural PES was conducted soon after the 1996 Census. Collection of the census questionnaires was allowed to be completed before the survey went into the field to avoid having census enumerators and PES interviewers in the field at the same time (Statistics New Zealand, 1998). At that time, a two-week buffer was deemed sufficient time for census collection to be completed. The 2001 PES commenced two weeks after the 2001 Census (Statistics New Zealand, 2002). Similarly, the 2006 Census was held on 7 March, and the 2006 PES was conducted from 21 March to 3 April 2006. In 2013, the PES was held slightly later than this, not beginning until 9 April 2013. This recognised that in 2006 the number of outstanding census forms at the end of the two-week window had been higher than desirable and impacted the coverage results. This was exacerbated by an even later start in 2018 to accommodate changes in the census model.

International practice

Mechanisms to measure census coverage are carried out in most countries which have a traditional census of the type that New Zealand conducts. At least 161 countries have held censuses in the current ‘2020’ round, among which 108 countries also had a post-censal evaluation of under-coverage, and 47 countries did not, with the remaining 8 countries not providing information about how they evaluated census coverage.

New Zealand, along with Australia, South Africa, Zambia, Rwanda (PES in each case), England/Wales, Scotland (Census Coverage Survey) and the United States of America (PES/Census Enumeration Follow Up), carried out a separate survey that involved re-counting a statistically structured sample of dwellings and the people living in them. Further information can be found in the cited literature, notably in Pereira (2002), Abbott (2011) and National Records of Scotland (2013).

Other approaches are also practised. Canada, for example, carries out a census-to-census record linking back to the previous census for a sample of the population. Canada's Reverse Record Check (RRC) is used in conjunction with an Over-coverage Study to derive a net undercount. The result of these two instruments then feeds into population estimates (Kerr, 1998). Most countries, including New Zealand, also calculate an intercensal error of closure using demographic techniques to supplement information on coverage and sources of undercount.

Changing census methods, such as the use of admin data and modernisation of collection methods, requires a shift in design for coverage measurement. New Zealand is ahead of many countries in this regard and this country’s experiences are of substantial international interest. A growing number of countries are learning from our experiences.

() About the 2018 post-enumeration survey

Objectives

The main purpose of the PES is to measure the level of coverage (undercount and overcount) in the census. Coverage measures are used as key performance indicators for the 2018 Census, with the PES constituting a major part of the evaluation component for the 2018 Census. This evaluation component is a government requirement when requesting funding for large programmes, such as census. The 2018 PES maintained the sample size that was increased for the 2013 PES, along with other planned quality improvement initiatives, to enable more robust evaluation of census.

In New Zealand, census counts are not adjusted directly to incorporate errors in counting identified by a PES. This is internationally the most common practice. Coverage measures, however, are used within Stats NZ to adjust the base population (see Estimated resident population 2018: Data sources and methods (methods/estimated-resident-population-2018-data-sources-and-methods)). This base population is the estimated resident population (ERP), which is used to derive post-censal population estimates and projections. Census coverage measures improve the accuracy of the base population, which in turn leads to more accurate post-censal estimates and projections.

Census coverage is not uniformly distributed. Younger people, males, students, new migrants, geographically mobile people or disadvantaged groups are all more likely than other groups to be missed (and therefore undercounted) in the census (O'Hare, 2019). All these groups are of high interest for policy making. As the ethnic and geographic diversity of the New Zealand population increases, the importance of measuring and accounting for the differences in undercount also increases. These differences are important elements in producing robust ethnic and subnational estimates and projections, and vitally important for ensuring equality and equity of treatment of people across society.

Population estimates and projections are widely used by central and local government to assess current and future needs for facilities and services in health, education, and social welfare, and in the allocation of public funds to provide these services. Incorporating adjustments for census coverage into the base estimated resident population enables all population groups to be as accurately represented in population statistics as possible, thereby facilitating better appropriation of funds. Businesses and community organisations also use estimates and projections data, for planning, research, and marketing purposes. Population estimates are used to improve the quality of numerous other surveys, by providing benchmarks for calibration of sample survey estimates.

Scope

PES population definitions are unique compared to standard social surveys in that they need to reflect both the census target populations (all people and dwellings in New Zealand) and the census key performance indicators (KPIs). For 2018 PES we have adopted the OECD definitions for target and survey populations:

> "A target population is the population outlined in the survey objects about which information is to be sought and a survey population is the population from which information can be obtained." (OECD glossary (https://stats.oecd.org/glossary/detail.asp?ID=2079))

In order to measure the KPIs relating to under- and over-coverage, the PES target definitions need to be broader than those of census to include those people/dwellings that should have been counted but were not, as well as those that were counted but should not have been.

The target population for dwellings in the 2018 PES is defined as all private dwellings that should have been (whether counted or not) or were (including those which should not have been) counted by the 2018 Census of Population and Dwellings. This change is to reflect potential for dwelling over-coverage in the 2018 Census model. The population for inference is all private dwellings in New Zealand.

The 2018 PES survey population for dwellings is consistent with the 2013 definition at a high level, that is all private dwellings on the North Island, South Island, or Waiheke Island of New Zealand.

However, the practical inclusions and exclusions differed from 2013 in order to support the detection and estimation of dwelling over-coverage as well as the detection of misclassification error (dwelling/non-dwelling, and private/non-private).

Dwelling inclusions

For practical reasons the 2018 PES survey population for dwellings includes permanent private dwellings and permanent non-private dwellings.

Permanent private dwellings:

- Permanent unoccupied dwellings and dwellings in which people do not usually live are included both for enumeration and estimation purposes.
- Private dwellings that are under construction are in scope for enumeration and estimation purposes.
- Mobile dwellings that are fixed to one location and have permanent private residents are included for both enumeration and estimation purposes.
- Dwellings in motor camps (such as caravans, campervans, cabins etc) that are fixed to one location and have permanent private residents are in scope for both enumeration and

estimation purposes.

Permanent non-private dwellings (including those under construction):

- These are included in scope for the PES field enumeration of dwellings for the purposes of supporting the detection of misclassification of private dwellings as non-private dwellings and vice versa.

Dwelling exclusions

For practical reasons the 2018 PES survey population for dwellings excludes ‘Other’ private dwellings.

‘Other’ private dwellings:

- Mobile dwellings (for example, caravans, houseboats, house trucks, tents etc) that are intended to be transportable and movable are excluded unless they have permanent private residents and are fixed to one location. These are excluded because of their movable nature and the likelihood that they will have changed location throughout the various phases of census and PES (PES enumeration, census enumeration, census night, and PES collection).
- Improvised dwellings, vehicles lived in, vessels lived in, and places of habitation with no dwelling are also out of scope due to their temporary and movable nature.
- Dwellings in some very remote areas are excluded because of the poor cost benefit trade-off of running field enumeration and interviews in remote areas.

Coverage estimation of non-private dwellings remains out-of-scope for 2018 PES.

The 2013 PES target population for people was all usually resident people (whether in private dwellings or not) who either should have been counted or were counted by the census. This was the same for 2018 PES and was not impacted by the change in dwelling population. The population for inference was all usually resident people in New Zealand on census night.

The 2018 PES survey population for people consisted of individuals who were usually resident in a New Zealand private dwelling or staying at one during the survey period (including both New Zealand residents and overseas visitors), with the following exclusions:

- people living in non-private dwellings (for example, prisons, hospitals, hotels)
- people living in 'other' private dwellings (for example, tents, caravans, yachts)
- people who died after census night
- overseas diplomats, their families, and people living with them
- people on off-shore islands (except Waiheke Island, which is included).

This is a slight change from the 2013 PES survey population for people. In 2013 PES people overseas for all of the PES survey period were excluded, but they were included in 2018 to support census transformation testing. Similarly, babies born after census date were excluded in 2013, but it was recognised that some such babies may have been enumerated in the census, thereby contributing to over-coverage.

Overseas visitors were included in the survey population as they may have been enumerated in census as usual residents and be contributing to the over-coverage.

Survey objectives

The survey objectives of the 2018 PES were to measure the level of coverage by the 2018 Census of Population and Dwellings of permanent private dwellings and of New Zealand residents in New Zealand on census night.

The PES was designed to produce estimates of gross undercount, gross overcount and net coverage for the following key population groups:

- total usually resident population
- sex (male, female)
- age (0–14 years, 15–29 years, 30–44 years, 45–64 years, 65–74 years, 75+ years)
- ethnicity (the European, Māori, Pacific, Asian, MELAA and Other groupings of ethnicities, together with Chinese, Indian, and Samoan, which contribute coverage measures for population projections)
- birthplace (New Zealand born, overseas born) and the Māori descent population
- geographical area in three groupings:
    - North Island, South Island
    - Auckland region, Waikato Region, Wellington region, rest of North Island, Canterbury region, rest of South Island, Total New Zealand
    - Territorial Authorities and Auckland Local Board Areas.

PESs in New Zealand set out to:

- reach most people who were randomly missed by census
- reach some people who have a lower likelihood of being found.

It was not specifically designed to:

- locate people who intend to remain invisible and take sufficient steps not to be found
- look for missing people via other methods such as through administrative records (which some international statistical agencies use to improve their coverage estimates)

- address the classification difficulties with dwellings (such as what is a private dwelling) where the counting or not of a dwelling could have an impact on the PES results and could result in people being systematically missed, or double-counted, by either census or PES
- address the effects of differences in enumeration practices between the census and the PES.

Sample size and accuracy targets

The 2018 PES aimed to measure the accuracy of coverage of permanent private dwellings and of New Zealand residents in the 2018 Census. Dwelling coverage rates were of high interest to the 2018 PES because of the use of a dwelling register in census. However, the increased focus on dwelling coverage was not to be at the expense of people coverage (and in the event was not achieved).

The target sample size for the 2018 PES was approximately 15,000 private dwellings, the same as the 2013 PES sample size.

KPIs provided quantitative measures of the quality of the post-enumeration survey. In the 2013 Census cycle, these had been agreed as part of the census programme KPIs. For 2018, it was agreed that the KPIs that specifically measure aspects of the PES should be determined by the census coverage project, with input from elsewhere. The KPIs for 2018 PES were:

- survey response rate (households) = 90 percent
- achieved sample rate = 82 percent
- national sampling error coverage = +/-0.4 percent (individuals)
- sub-population sampling error coverage (individuals) by age group, by ethnicity, by region = +/-1.5 percent.

While dwelling coverage rates were considered important during 2018 planning stages, no sampling error KPIs for dwellings were able to be set because of the need to understand more about the new field methodology and processes being planned for 2018 Census. The likely impact on PES had to be determined in order to establish dwelling-level KPIs. Consequently, dwelling coverage was de-scoped from the 2018 PES.

PES sampling frame and selection units

The 2018 PES sample uses a two-stage sample selection method, similar to many Stats NZ household surveys.

The first stage of the selection consists of selecting primary sampling units (PSUs) from the Household Survey Frame (HSF). The HSF is the standard sampling frame Stats NZ uses to select samples and manage overlap controls for all of its household surveys, comprising a list of PSUs with attributes determined by census data. PSUs are geographic areas consisting of between 50

and 100 dwellings and are formed using meshblocks – Stats NZ's standard geographic classification.

The PES sample consisted of 1364 PSUs. Prior to the second stage of sample selection, PES conducted a dwelling enumeration exercise in the selected PSUs to create the sample frame of eligible dwellings within each PSU.

The dwelling enumeration exercise was carried out in January 2018, by existing survey field staff and additional staff recruited for the exercise, using a starting address list supplied by Homes.co.nz. An important requirement for the PES dwelling enumeration exercise was that the list of dwellings used be as independent as possible of the census list of dwellings. Similarly, field operations needed to maximise this independence by taking care to ensure PES staff did not overlap with census staff (in terms of both staff members and timing of field activities).

Each PSU was divided into panels containing about eleven addresses in each. The second stage of the sample selection consisted of selecting eligible dwellings within the selected PSUs by selecting one of more of these panels. One of these panels was selected to select the sample of addresses that were visited for PES interview (the remaining panels were then available to other Stats NZ surveys). Where an address had multiple dwellings associated with it, then all dwellings at the address were selected for interview (including those detected as part of the interview that were not detected during dwelling enumeration). The selection unit was the panel, the collection unit was the dwelling and the people.

Sample design

The 2018 PES sample design was purpose-built to balance the desired accuracy targets with the cost and sample size constraints. This was a change from previous PESs which had sample designs based on existing Stats NZ household surveys such as the New Zealand General Social Survey (NZGSS, the base for 2013 PES) and the Household Labour Force Survey (HLFS, the base for 2006 PES).

As mentioned in the previous section, the first stage of the sample design and selection was based on the PSUs which are formed from meshblocks. At the time of the 2018 PES sample design and selection, the NZ geographic framework consisted of 53,589 meshblocks of which 52,973 were within the household survey frame grouped into 23,174 PSUs (excluding PSU 999999 Area Outside Scope). These PSUs were further grouped into strata to improve sampling efficiency. The strata groups were created from a combination of geographic concepts that are included to manage cost of field and interview operations, and concepts that are highly related to the survey objectives (potential to be missed or miscounted by the census).

The 2018 PES sample design grouped the PSUs into 102 strata based on:

- regions
- a mix of urban and rural
- the planned mode of delivery of census forms
- socio-economic characteristics captured in the NZ Deprivation index.

Territorial authority and Auckland local board areas (TALB) were unable to be explicitly included in the strata due to sample size constraints. However, the PSU selection process was designed to implicitly account for TALBs and to maximise the probability of selecting at least one PSU in each TALB.

The final stage of the sample design was to boost the sample representation of Pacific people. This was done using the probability proportional to size (PPS) sampling strategy, where the probability of selecting a PSU depends on its size. The PPS proportion was transformed by multiplying it by a factor reflecting the proportion of adults of Pacific ethnicity in the PSU. The targeting of Pacific people was done through the PPS (and not through stratification) because it allowed an increase in the survey precision for this group without decreasing the survey precision for other groups.

Field test operation

A field test was carried out in Whanganui in March 2017 to evaluate the planning for the enumeration phase. The enumeration phase involved a sample selection of 39 PSUs, which contained an initial 4,604 addresses, in the area to be covered by the 4 April 2017 census field test in the Whanganui Territorial Authority.

The PES enumeration test was conducted between 6 and 19 March 2017, using the canvassing option of the field management software (Census TrAK tablet app). The initial address list served as a starting point and field staff added dwellings in the PSU that were not on the list, classified address and dwelling types, and recorded those that could not be found. The objective was to ensure that dwellings missing from the address list could be added (birthed) efficiently and accurately.

This was followed by an interview visit test, which involved a sample of 345 dwellings in the area covered by the 4 April 2017 census field test in Whanganui (in the Whanganui TA). This was conducted during April and May 2017, using a modified version of the Non-Response-Follow-Up option of the TrAK app as well as a PES specific questionnaire app. The test was voluntary.

The 2017 PES field test initially had six primary objectives, to test:

- new methods for enumerating dwellings
- field processes and systems

- the electronic questionnaire (all new for 2018)
- the back-up paper questionnaire (revamped from 2013)
- integration of corporate systems – Salesforce, TrAK, Pentaho, SAS programs etc
- dwelling enumeration.

The field test identified several key learnings that were valuable for the PES collection process. Key among these were limitations identified in the use of the tablets and the electronic data capture system, including the capacity to transfer and integrate data into the wider processing system. Of paramount importance was ensuring a fully certified and seamless protection of security for data collected throughout the process of collection, on the tablets in the field as well as during the transfer, and receipting and processing of information into the Stats NZ secure operating environment.

Operational constraints were also identified, notably around recruitment and training. In addition, systematic problems were identified, such as the inclusion (birthing) of unrecorded dwellings, especially in large complexes, and the handling of dwellings where there was mixed use (private and non-private) or multi-household dwellings. Software limitations meant that missing dwellings were not properly identified and processed in the field during the field test. Limitations were also identified in address geolocation, especially in rural areas, making it sometimes difficult for the field staff to locate the correct target dwelling.

Learnings from the field test were incorporated in the end-to-end planning and a sound solution was developed. However, much of this was complicated by a late change to software that impacted the PES field operation and subsequent processing.

Collection and response rate

The survey was carried out between 26 April and 2 July 2018, following the planned completion of census fieldwork. The survey period had been planned for 17 April to 20 May 2018. Those dates had been a balancing exercise – the survey needed to start late enough to avoid overlapping in the field of census enumerators and PES interviewers, as well as minimising the effect of large numbers of late census returns. The survey also needed to start close enough to census date (6 March 2018) to maximise respondent recall. However, the start was delayed until 26 April to allow additional time to improve response rates. The planned finish date was also extended until 2 July 2018 to increase PES response rates in some areas where responses were lower than expected.

Similarly, proximity to census presents a problem common to all PESs. People missed by censuses are also likely to be missed by a PES, especially when the time between the collections is short. This is referred to as the capture-recapture or correlation bias. To reduce this bias as much as possible, the PES is kept as independent of the census as possible. To ensure this independence,

the PES was conducted after the majority of census fieldwork was completed to avoid inadvertent contact between census collectors and PES interviewers.

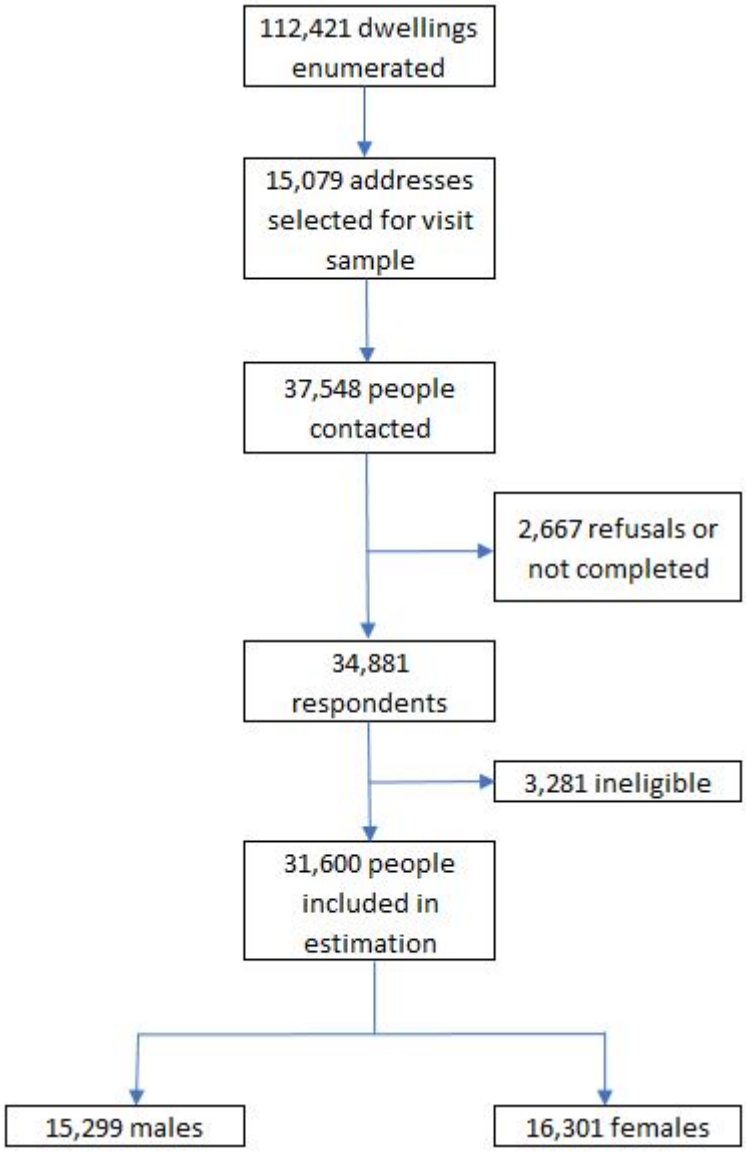
Data was collected by approximately 100 specially trained interviewers, working with 12 team leaders, using a household electronic questionnaire installed on Samsung tablets. Information on occupants of the dwelling who satisfied the scope and coverage criteria was collected through a face-to-face interview wherever practicable. When not possible, a proxy interview was conducted (that is, details were obtained from another adult in the dwelling) or a follow-up interview was done by telephone (unless the respondent insisted on a face-to-face interview).

In 2018 PES, as in 2013, we relaxed the rule of not using interviewers who had been census collectors, in order to meet field work targets. A few census collectors were selected for the PES work and were assigned to areas different from where they carried out their census work. However, most interviewers had not worked as collectors for the 2018 Census. Other countries have found that the mutual independence of census and PES is not compromised by the use and careful management of the same field staff for both surveys.

Personal details sought on the PES questionnaire included name, sex, date of birth or age, ethnicity, country of birth, Māori descent, and addresses. While the most important addresses were the usual address of the respondent and the address where they were counted on census night, the PES also collected information on any other addresses where the person might have been counted. This was to increase the chance of finding and matching the person's individual census form and, importantly, to help identify where people may have been counted more than once.

As figure 1 outlines, 15,079 addresses were selected as the sample to be visited, from the 112,421 dwellings enumerated in the target geographic areas. In addition to these dwellings, a further 134 dwellings were found and birthed in the field to give a total of 15,213 dwellings visited. At these dwellings, 37,548 people were found and contacted, which resulted in 34,881 responses (92.9 percent) being received. After the removal of 3,281 ineligible responses, there were 31,600 responses of sufficient quality to present for estimation (90.6 percent of the received responses). Of these, 48.45 percent were male and 51.6 percent were female.

Figure 1



Text alternative for figure 1: Flow of PES responses from enumeration to initiation of estimation

The figure is a flowchart that describes the flow of responses through the PES collection and processing phases to the point where they are presented to the estimation modelling process. An initial sample of 15,079 addresses was selected for visiting from among the 112,421 previously dwelling enumerated in the target geographic areas. At these addresses, a further 134 dwellings were found and birthed in the field to give a total of 15,213 dwellings visited. At these dwellings, 37,548 people were found and contacted. Any refusals or uncompleted interviews were deleted, which resulted in 34,881 responses (92.9 percent) being received. After removal of 3,281 ineligible responses, there were 31,600 responses of sufficient quality to present for estimation. Of these responses, 15,299 were male and 16,301 female.

Data processing: linking and searching

The 2018 PES responses were linked to the 2018 Census records. The objective of linking was to determine if a PES respondent was counted in the census. This included searching for them at their usual residence address, and at additional addresses where a census form may have been completed for them.

Linking of the records used a combination of automated and manual linking processes. The linking processes were designed to maximise the certainty with which a link between a PES response and a census response (or admin enumeration) could be made. In order to increase the level of certainty in linking, only high-quality PES responses were retained for linking. A high-quality response was defined as having two of the following pieces of information: name, date of birth, and address. The same high-quality definition was also applied to the census dataset.

The 2018 PES processing system used QualityStage (QS) software to complete multiple passes of automatic linking. The census data and PES data were first standardised, so that matching variables were coded compatibly.

The PES data was passed through QS up to four times, using a probabilistic linking approach. Probabilistic linking, which was formalised by Fellegi and Sunter (Fellegi and Sunter, 1969; Winkler, 1990), uses the probabilities of agreement and disagreement between a range of linking variables. Under the Fellegi and Sunter model, record pairs with probabilities above a certain cut-off value are considered links, while pairs with probabilities below another cut-off value are considered non-links, and the pairs with probabilities between the two cut-off values are set aside for subsequent passes.

Table 1 shows details of the four linking passes. In the first pass, records must share exactly the same date of birth, sex, and address ID (blocking variables). First and last names are then compared (linking variables), and records with enough similarity are accepted as links. Given the high level of information included in the blocking variables, this is a high-quality and conservative pass.

Subsequent passes used additional information such as meshblock (small geographic area) and country of birth, while allowing for some differences in these variables. However, relatively conservative cut-off weights have been chosen to minimise the number of incorrect links.

In 2018 PES, all non-exact links (indicating there are moderate differences in name or date of birth) and a sample of near-exact links were clerically reviewed to determine if the link should be retained or discarded.

The automatic linking process is followed by multiple manual linking steps for all remaining unlinked PES records. A team of reviewers searched for evidence of each PES individual in the census file using various combinations of identifying information. Additional information from the Integrated Data Infrastructure was also used where possible to provide more confidence in a given match. The detailed nature of these manual linking processes ensures that relatively few true matches were missed.

Based on a set of criteria for the similarity of PES and census information, each PES record was assigned a link status, along with a value representing the level of agreement. All automatic and manual links were retained and flowed through to PES estimation. PES records with no match identified also carried through and contributed to the estimation of under-coverage.

Table 1 Summary of link variables in 2018 PES, by pass

Pass	Blocking variables	Linking variables	Linking type
1	Date of birth Sex Address ID	First name Family name	Deterministic
2	Date of birth Sex	First name Family name Country of birth Meshblock	Probabilistic
3	Meshblock	First name Family name Day of birth Month of birth Year of birth Sex Age Country of birth	Probabilistic
4	Sex Age Soundex first name Soundex family name	First name Family name Day of birth Month of birth Meshblock	Probabilistic

A total of 30,569 links were made for eligible PES records, 29,269 through automatic linking and 1,300 through manual linking. Of the 31,600 total eligible PES respondents, 30,226 PES records were linked to a single census record, while 171 were linked to more than one census record. The remaining 1,203 PES respondents were not linked to a census record.

Dwelling coverage estimation

The 2018 PES set out to estimate the coverage in census of permanent private dwellings. Measurement of dwelling coverage has become increasingly more challenging as the census approach to counting dwellings has shifted to include greater use of administrative data.

Previously, PES would have used administrative lists of dwellings to form its dwelling list and sample frame, while census had field staff canvass the country listing all located dwellings as they went. These two approaches to create lists of dwellings were completely independent and supported PES estimating the coverage of dwellings achieved by the census.

With the 2018 Census moving to increased use of administrative data for dwellings in 2018, obtaining an independent and high-quality administrative list of dwellings was a challenge for PES. This challenge proved too much for our estimation methods and during the process we made the decision to de-scope dwelling coverage from the outputs, as we were unable to provide robust estimates of the coverage of dwellings in the 2018 Census.

Person coverage estimation

Coverage of people is the measure of the difference between the PES estimated population and the counted census resident population.

Three measures relate to this difference: gross overcount, gross undercount, and net undercount. All three are expressed as numbers, with the net undercount also expressed as a percentage of the PES estimated population.

Gross overcount is the number of people counted more than once in the census, along with people who were counted in the census, but should not have been counted.

Gross undercount is the number of people who were supposed to be counted by the census but were not counted.

The net undercount is the difference between gross undercount and gross overcount. Because the difference nearly always results in an undercount, the net undercount is a measure of how much smaller the counted census population is than the PES estimated population. Any case where the PES estimated population is smaller (that is, a net overcount) is expressed as a negative undercount.

The level of coverage is the census resident total expressed as a percentage of the PES estimate of the total population. The net coverage and the net under-coverage add to 100.

Estimation concepts

The coverage estimation of the 2018 Census dataset relates in part to how the census file was created, and how different groups of records were treated differently throughout the estimation process. In this section we:

- define the different groups making up the census file
- describe their different treatments in the coverage estimation process
- give details on how we calculate the coverage measures
- give details on how we use Bayesian models in to estimate coverage probabilities

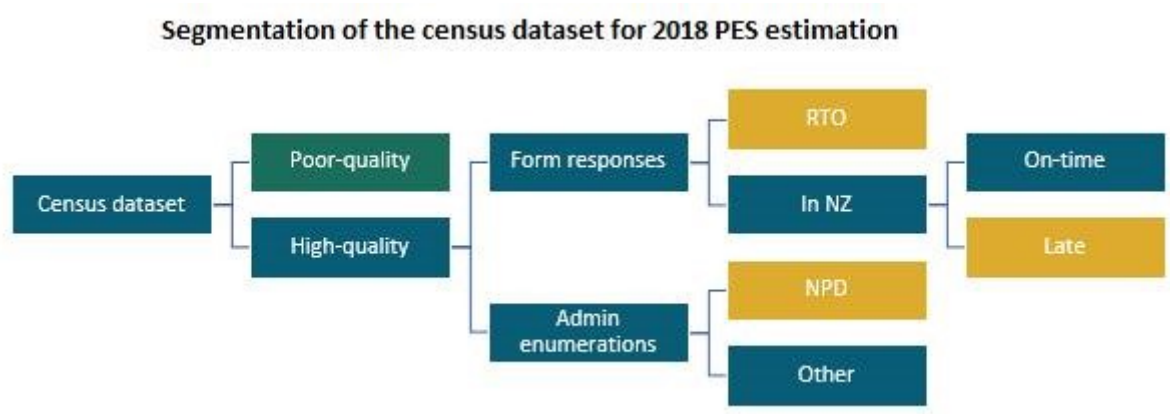
We also include a summary of changes from the previous estimation approach, along with discussion of benefits and limitations.

For coverage estimation of the 2018 Census dataset, we distinguished the following groups within the census dataset:

- quality of linking information – records in the census file are split into those with high-quality linking information and those without
- type of record – high-quality records in the census file are split into those that come from a census form response, and those that are admin enumerations
- presence in New Zealand on census night –: high-quality census form responses are split into those that were in New Zealand on census night, and those that can be identified from admin data as having been overseas on census night (residents temporarily overseas, RTO)
- time of response – high-quality census form responses for people present in New Zealand on census night are split into those that were received prior to the start of PES interviewing (26 April), and those received after (late)
- type of admin enumeration – admin enumerations into prison and defence establishments (non-private dwellings (NPD)) are separated from the remaining admin enumerations (mostly in private dwellings or at meshblock level).

Figure 2 shows how the census dataset was grouped. The blue boxes reflect the census records that are eligible for inclusion in coverage estimation. The ochre boxes show the census records that were excluded from coverage estimation for the most part, and the records in the green box were accounted for using an estimation adjustment.

Figure 2



Text alternative for figure 2: Segmentation of the census dataset for PES estimation

The figure is a tree-diagram (dendrograph) which visualises how the census dataset is grouped for coverage estimation.

From the census dataset, high quality responses received on a census form from people in New Zealand on census night

and returned prior to PES beginning field collection operations are selected. In addition, high quality admin enumerations for people not in non-private dwellings are also included. Residents temporarily overseas (RTOs), people in non-private dwellings, and late returns are for the most part excluded from coverage estimation. Responses in census which do not meet the criterion of high quality are accounted for by an estimation adjustment.

Census eligible population

The census eligible population is the high-level group of census records that are eligible for coverage estimation. This group is the combination of high-quality admin enumerations (excluding those in prison and defence establishments), and on-time census form responses for residents present in New Zealand on census night. On-time census form responses returned from non-private dwellings (NPDs) are included in this population.

Quality of linking information

Records are distinguished by the quality of their linking information in efforts to reduce the amount of linkage error between the PES and census records. Linkage error has been shown to introduce bias into coverage estimation. Records with poor quality information are treated as under-coverage throughout the estimation process since they do not have an opportunity to link. However, they are accounted for in the gross undercount calculation using an adjustment since they are not true under-coverage.

Admin enumerations in prisons and defence establishments

As part of the 2018 Census process, admin enumerations were added into prison and defence non-private dwellings, based on files of unnamed unit records provided by the Department of Corrections and Department of Defence respectively. These individuals were highly unlikely to be part of the PES survey population, which excludes people living in non-private dwellings. In addition, the use of agency-supplied records means we were confident the census counts were reliable for the prison and defence populations and did not expect the same coverage patterns as for the rest of the population.

Therefore, these admin enumerations in non-private dwellings were excluded from the PES estimation process. They are included in the total population but are assumed to have no coverage error.

Residents temporarily overseas

A resident temporarily overseas (RTO) in this context is a person who usually lives in New Zealand but was overseas on census day (6 March 2018). RTOs who responded to the census and are contained in the 2018 Census dataset are identified using admin data on border movements (dates of departure from New Zealand, and dates of arrival into New Zealand). The same approach is used for identifying RTOs surveyed by PES. As we can observe over-coverage from

RTOs directly through this admin data approach, we excluded all records identified as RTOs (in both the census and the PES) from coverage estimation. RTOs did contribute to the gross overcount.

Late census returns

A late return is a census individual form that is completed and received by census on or after the date PES field activities begin (26 April 2018). It is important to identify late returns and remove these from coverage estimation because the PES and census are required to be independent. Census forms returned after the PES went into the field may have been influenced by awareness of the PES. This would lead to a biased estimate of census coverage as the PES respondents would have a higher rate of completed census forms than the rest of the population. The problem is that we cannot distinguish between census forms returned late because of PES prompting and those returned late because of other reasons.

In general, we excluded late returns from coverage estimation and added them back into the population estimate at the end. There were two exceptions to this: late returns that contributed to over-coverage, and late returns that did not meet the requirement for high-quality linking information. Late returns that were reliably identified as duplicate-driven over-coverage were included for over-coverage estimation as they indicate a systematic problem in census regarding resolution of multiple responses. This can occur in situations in which a census respondent provides both an on-time and a late return, or multiple late returns. Late returns with poor-quality linking information were treated the same as on-time returns with poor-quality linking information, that is they were treated as under-coverage throughout.

Estimation methods 2013 and 2018

The estimation approach used for 2013 and previous PESs was a direct weighted estimation like those used by other Stats NZ household surveys. The direct weighted approach involved multiple weight adjustments and calibration steps to ensure the PES sample correctly represented the full population of NZ. More information on the 2013 PES methodology is available in Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey (http://archive.stats.govt.nz/browse_for_stats/population/census_counts/report-on-2013-post-enumeration-survey.aspx) .

Following the 2013 PES, a Bayesian modelling approach was developed to produce PES results for much finer sub-populations in preparation for producing uncertainty measures for the 2013-base estimated resident population (ERP) (Bryant et al, 2016). This modelling work has since been extended and forms the estimation approach used by 2018 PES as part of an effort to bring consistency to the methods used by the PES and the ERP.

The main benefit of retaining the direct estimation approach used previously is for consistency throughout time to allow comparison with previous PES results, and consistency with other Stats NZ household surveys. Another benefit is the computational speed with which results can be produced. In comparison, the main benefits of the modelling approach are the depth of information it can produce from the same size sample, and the flexibility to de-couple dwelling coverage from person coverage. The modelling approach allows for pooling of information across groups, whereas direct estimation treats each sub-group as a distinct population and does not allow for information to be shared between groups.

By using the modelling approach in 2018 PES, we were able to improve the granularity of geographic estimates from broad geographic regions (Auckland, Wellington, rest of North Island, Canterbury, and rest of South Island) to territorial authorities and Auckland local board areas. In addition, 2018 PES estimates include estimates of coverage by Māori descent, birthplace, and all level one groupings of ethnicity.

The additional data value that can be gained using the modelling estimation approach outweighs the need for consistency in methodology.

Estimation framework

After being linked to the census file and filtered to eligible records, the post-enumeration survey was used to infer census coverage. The PES provided a sample of the population, in which the presence, absence, and number of records in the census were recorded alongside geographic and demographic attributes. We modified the popular dual-system estimation (DSE) approach to estimate and correct census counts for under- and over-coverage. In a classical DSE framework where over-coverage is ignored, we would define a system shown in the matrix in table 2, where individuals are added to cells based on their presence or absence in each list.

To account for and quantify heterogeneity of inclusion in the census in New Zealand, the population was divided into categories by geographical areas, noted T, and demographic attributes (such as age, ethnicity, and sex), noted $\boldsymbol{X}$. Table 2 cell counts are therefore expressed as a function of $\boldsymbol{x}$ and t, which represents any subset of $\boldsymbol{X}$ and T formed by a combination of demographic and geographic attributes. Counts n_{00} ($\boldsymbol{x}$,t) and therefore N($\boldsymbol{x}$,t) are unobserved.

Table 2 Cross-tabulation of census and target population counts

		Census		
		1	0	
PES	1	$n_{11}(\boldsymbol{x},t)$	$n_{10}(\boldsymbol{x},t)$	$N_1(\boldsymbol{x},t)$
	0	$n_{01}(\boldsymbol{x},t)$	$n_{00}(\boldsymbol{x},t)$	
		$N_C(\boldsymbol{x},t)$		$N(\boldsymbol{x},t)$

Complications arise when it is not relevant to ignore over-coverage. Given the PES is designed to detect such cases of over-coverage, we can use PES to estimate the probability of a census record to be over-coverage, $\widehat{p_{ocov}}(\boldsymbol{x},t)$. Let $\widehat{p_{ucov}}(\boldsymbol{x},t) = n_{10}(\boldsymbol{x},t)/N_1(\boldsymbol{x},t)$ be the probability of being under-coverage for an individual with characteristics $\boldsymbol{X} = \boldsymbol{x}$ and $T = t$ in the population. An adjusted population count for such characteristics is defined as $N(\boldsymbol{x},t) = R(\boldsymbol{x},t) \times N_C(\boldsymbol{x},t)$, where

$$R(\boldsymbol{x},t) = \frac{1-\widehat{p_{ocov}}(\boldsymbol{x},t)}{1-\widehat{p_{ucov}}(\boldsymbol{x},t)}$$

is the coverage adjustment ratio applied to census counts in each demographic category $\boldsymbol{x}$ from $\boldsymbol{X}$ and each geographical area t from T.

As the estimation is performed at a small geographic and demographic level, using under- and over-coverage proportions observed in the PES sample as direct estimates was inappropriate. Instead, two Bayesian multilevel generalised linear models were used to estimate under-coverage and over-coverage probabilities for each cross-tabulation cell of single year of age, sex, ethnicity, New-Zealand born, Māori descent, and territorial authority and Auckland local board areas (TALB). Binary under- and over-coverage indicators were regressed against demographic covariates in two multilevel Bernoulli-logistic models (one for each under and over-coverage estimation processes). The hierarchical element of the under-coverage model incorporates the PES sampling design. For each under- and over-coverage model, 1000 draws from the posterior were used to predict the under- and over-coverage probabilities associated with each combination of covariate values that exist in the census. To obtain estimates at the TALB level, a Monte-Carlo integration over parameters related to the sampling design (household, PSU, and stratum effects) was performed. Note, the over-coverage model is simplified compared to the under-coverage model.

The geographic component of the model was reduced to the level of TALBs, where all individuals from the same TALB were assigned the same TALB-specific intercept. Included individual

covariates were four ethnicity variables, 10 age-splines, and a binary New Zealand-born indicator. The estimated under- and over-coverage probabilities were then combined into a coverage adjustment ratio R(x,t) applied to census records for each unique combination of TALB and demographic attributes. After combining eligible census counts and the coverage adjustment ratio, the output has the form of 1,000 vectors of coverage adjusted census counts for each demo-geographic combination.

Note: Some of the formulas in this section are images. If you need alternative text for them, please contact demography@stats.govt.nz (mailto:demography@stats.govt.nz) .

Estimated population for New Zealand

The PES estimated population for New Zealand is the sum of the coverage adjusted census eligible population, the late census returns, and admin enumerations in prisons and defence establishments (non-private dwellings, NPDs).

$$PES\ estimated\ population = census_{elig} \times R + census_{late} + census_{NPD}$$

where:
$census_{elig}$ = census count of high-quality admin enumerations (excluding those in prison and defence establishments), and on-time census form responses for residents present in New Zealand on census night (irrespective of responding in private dwelling or non-private dwelling)

R = coverage adjustment ratio produced from the combined under- and over-coverage models

$census_{late}$ = census count of late returns with high-quality linking information

$census_{NPD}$ = census count of prison and defence admin enumerations

From the PES estimated population, we can then produce the net undercount as the difference between the PES estimation population and the census resident population count.

$$Net\ undercount = PES\ estimated\ population - census$$

$$Net\ undercount\ (\%) = \frac{net\ undercount}{PES\ estimated\ population} \times 100$$

$$\text{The PES estimate of the census level of coverage (\%)} = \frac{census}{PES\ estimated\ population} \times 100$$

where:
census = census count of the usual resident population

Note: Some of the formulas in this section are images. If you need alternative text for them, please contact demography@stats.govt.nz (mailto:demography@stats.govt.nz) .

Estimated gross undercount of the census eligible population

Only people who should have been counted by the census can potentially contribute to the gross undercount.

In terms of PES variables, anyone who has WAS < 1 and should have been counted will contribute to the undercount. Each PES record for anyone who should have been counted is assigned a binary under-coverage indicator:

Under01 = 1 if a PES respondent SHOULD have been counted in the census but WAS NOT

= 0 if a PES respondent SHOULD have been counted in the census and WAS

The gross undercount is produced from the under-coverage model operating on the census eligible records, with an adjustment to account for records excluded from the census eligible population due to insufficient quality of linking information. This adjustment is included to account for the fact that the model estimates under-coverage for the eligible census records rather than all census records. The census records excluded from the estimation due to poor quality linking variables were, in fact, counted by census so are not considered part of the under-coverage of the full census file.

$$Gross\ undercount = census_{elig} \times (1 - p_{ocov}) \times \frac{p_{ucov}}{1 - p_{ucov}} - census_{linkrej}$$

$$Gross\ undercount\ (\%) = \frac{gross\ undercount}{PES\ estimated\ population} \times 100$$

where:

$census_{elig}$ = census count of high-quality admin enumerations (excluding those in prison and defence establishments), and on-time census form responses for residents present in New Zealand on census night (irrespective of responding in private dwelling or non-private dwelling)

$_{Pucov}$ = the estimated under-coverage probabilities for the $census_{elig}$ records

$_{Pocov}$ = the estimated over-coverage probabilities for the $census_{elig}$ records

$census_{linkrej}$ = census count of records with poor-quality linking information (includes on-time returns, late returns, and non-prison or non-defence admin enumerations).

Note: Some of the formulas in this section are images. If you need alternative text for them, please contact demography@stats.govt.nz (mailto:demography@stats.govt.nz) .

Estimated gross overcount of the census eligible population

For gross overcount, we needed to consider both the people counted more than once, and people counted when they should not have been. People may have been counted in the census more than once because they completed census forms at more than one address, someone completed a form on their behalf, or census processing included an admin enumeration for someone who had also completed a form.

Conversely, people may have been counted in the census when they should not have been. This can occur when people complete forms for babies born after census night, or forms are completed for people who were overseas on census night (either completed on their behalf, or by themselves before leaving or after returning).

In terms of PES variables:

- anyone counted more than once is any person who should have been counted and who was counted more than once (WAS > 1)
- anyone counted once (WAS > 0) when they SHOULD NOT have been.

Each PES record that was counted is assigned a binary over-coverage indicator:

Over01 = 1 if a PES respondent SHOULD have been counted in the census ONCE and WAS counted

more than ONCE

OR

= 1 if a PES respondent SHOULD NOT have been counted in the census and WAS

= 0 if a PES response SHOULD have been counted in the census ONCE and WAS counted only ONCE

Gross overcount is produced as the combination of the over-coverage probabilities for the $census_{elig}$ records and the over-coverage directly observed from RTOs:

$Gross\ overcount = census_{elig} \times {}_{Pocov} + census_{RTO}$

where:

$census_{elig}$ = census count of high-quality admin enumerations (excluding those in prison and defence establishments), and on-time census form responses for residents present in New Zealand on census night (irrespective of responding in private dwelling or non-private dwelling)

${}_{Pocov}$ = the estimated over-coverage probabilities for the $census_{elig}$ records

$census_{RTO}$ = census count of records identified as RTOs using admin data on border movements.

Credible intervals

In Bayesian statistics, credible intervals are used to summarise the uncertainty around an estimated quantity of interest. While the goal of credible intervals is similar to the frequentist confidence intervals, the statistical definition and meaning is very different and more straightforward. Credible intervals describe an interval within which the unobserved quantity of interest falls with a particular probability. For instance, a 90 percent credible interval defines a range of values within which the ’true’ value has a 90 percent probability of belonging.

In Bayesian statistics, the estimation of a quantity of interest takes the form of a probability distribution (the posterior distribution) instead of a point value. Therefore, a credible interval is a summary of the posterior distribution. Following our previous example, if we want a 90 percent credible interval for a quantity we are trying to estimate, we need to find the central portion that corresponds to 90 percent the posterior distribution. To do so, we select the 5 percent quantile (the value below which lies 5 percent of the distribution), and the 95 percent quantile (the value below which lies 95 percent of the distribution).

Care is required when interpreting coverage measure. Credible intervals indicate uncertainties. Noting these intervals is especially important when comparing results.

() Key performance indicators

Response rate

In addition to measuring against a set of census target KPIs for coverage and response, the 2018 PES set project targets for project-level response and precision of response. The survey response rate indicates what proportion of eligible households responded to the survey and contained at least one fully responding eligible person. The survey response rate (households) target was 90 percent.

The calculation of the response rate is done at the household level as opposed to the individual level. This is the standard approach to response rate calculation across all household surveys.

Survey response rate is calculated as:

$$Response\ rate = \frac{Eligible\ responding\ households}{Eligible\ households} = \frac{D}{C + D + E\left(\frac{C + D}{B + C + D}\right)}$$

where:

A = sum of the selection weights of all **ineligible pre-contact** households
B = sum of the selection weights of all **ineligible post-contact** households
C = sum of the selection weights of all **eligible non-responding** households
D = sum of the selection weights of all **eligible responding** households
E = sum of the selection weights of all **unknown eligible** households.

Note: Some of the formulas in this section are images. If you need alternative text for them, please contact demography@stats.govt.nz (mailto:demography@stats.govt.nz) .

Precision targets

The change in the estimation methodology, from direct weighted estimation to Bayesian modelling, necessitated a change in the way that uncertainty in the estimates is reflected. The primary difference is that uncertainties are reported as credible intervals rather than sample errors.

The uncertainty in our estimates is measured using the credible interval estimates that are produced by the model, reporting the 0.025 percentile (lower) and the 0.975 percentile (upper). These KPIs will be presented as the percentage difference between the PES median value and the upper/lower credible interval values. For example, a value of 1 percent means that there is an approximately 95 percent probability that the true value of the PES population is within 1 percent of the median estimate.

It is calculated as:

$$Credible\ intervals = ABS(\frac{PES\ median\ estimate - PES\ upper/lower\ estimate}{PES\ median\ estimate}) \times 100$$

Note: Some of the formulas in this section are images. If you need alternative text for them, please contact demography@stats.govt.nz (mailto:demography@stats.govt.nz) .

Background 2018 PES versus 2013 PES

KPIs are used to inform the sample design, field operational planning, and modelling of coverage data. The two main indicators for 2018 PES were the official response rates and the sampling error.

Key performance indicators provide quantitative measures of the quality of the PES. For the 2018 cycle, it was agreed that the KPIs that specifically measure aspects of PES should be determined by census coverage project (with input from other areas). This differs from the 2013 cycle, where it was part of the census programme KPIs.

The 2013 KPIs provided useful guidance in a range of areas including the sample design and modelling of coverage estimates in the 2013 cycle. The 2013 targets were not all met but keeping the same KPI target for the 2018 cycle ensured an ongoing focus on quality and improve on past performance.

In the 2013 cycle, 82 percent of the set targets were achieved (9 KPIs of 11). All the sub-population precision targets were achieved. The national target and the survey response-rate target were not achieved.

Table 3 2013 PES Key performance indicators

Variable	Target	Outcome
	Percent	
National response rate	90	87
National sampling error	0.4	0.5
Sub-population sampling errors	1.5	0.5–1.5

Key performance indicators 2018 PES

The 2018 PES KPIs are similar to the 2013 KPIs. Additional credible interval targets for broad regions were also set for 2018. The target key performance indicators for the 2018 PES are

tabulated in table 4. The achieved rates are discussed below in table 5.

Table 4 2018 PES Key performance indicators

Key performance indicators	Target percent
National response rate	90
Credible intervals	
National	+/- 0.4
Age group (years)	
0–14	+/- 1.5
15–29	
30–44	
45–64	
65–74	
75+	
Ethnicity group	
Asian	+/- 1.5
European	
Māori	
Pacific	
Regions	
Auckland	+/- 1.5
Wellington	
Canterbury	
Rest of North Island	
Rest of South Island	

2018 PES KPI results

As table 5 shows, the KPI targets were all met in the 2018 cycle.

The national response rate result was 91.1 percent, with a target of 90 percent. The national credible interval was 0.3 percent – this is relatively low compared to the credible intervals for the

sub-population groups and broad regions. The target for the sub-population and regions was a credible interval of 1.5 percent. The sub-population and region results ranged from 0.29 percent to 1.22 percent.

For the age groups, the 15–29 years age group had the highest value of 0.69 percent. The 45–64 years and 65–74 years age groups had the lower range of values, 0.44 percent and 0.45 percent, respectively. The 75 years and over age group had a credible interval of 0.56 percent.

For ethnicity, the Pacific grouping of ethnicities had the highest range of 1.22 percent. The European grouping had the lowest, 0.29 percent. The Asian grouping of ethnicities had a credible interval of 0.98 percent, with the Māori population achieving a lower value of 0.76 percent.

The regional results show that Wellington had the highest value, 0.65 percent. The rest of the North Island had the lowest value of the broad regions at 0.54 percent. Auckland and Canterbury did not vary much from Wellington, with 0.61 percent and 0.60 percent, respectively. The regions displayed a narrow range of values between 0.54 percent to 0.65 percent.

Table 5 2018 PES key performance indicator results

Variable	Target	Result
	Percent	
Response rate	90	91.1
National	+/- 0.4	0.30
Age group (years)		
0–14	+/- 1.5	0.54
15–29		0.69
30–44		0.53
45–64		0.44
65–74		0.45
75+		0.56
Ethnicity group		
Asian	+/- 1.5	0.98
European		0.29
Māori		0.76
Pacific		1.22
Regions		
Auckland	+/- 1.5	0.61
Wellington		0.65
Canterbury		0.60
Rest of North Island		0.54
Rest of South Island		0.62

() 2018 Census KPI results

2018 PES coverage results are the official response rates of the 2018 Census.

For the 2018 Census, the response rate was calculated using two methods – the traditional method, which was more comparable with previous PESs, and a new method based on a changed definition of a response:

- traditional method – in previous censuses, a census response was defined as the receipt of an individual census form and no minimum amount of information was required for the form to be counted as response
- new method – the completion of two or more of the name, date of birth, and meshblock fields on an individual, dwelling, or household census form was considered a response – this method is more consistent with international practice.

The national and sub-population response rate census KPIs were not met. The national target was 94 percent. For the traditional method, the response rate was 81.6 percent, and for the new method, the response rate was 85.8 percent.

For the sub-population groups, the target was 92 percent. For the Māori population group, the response rate was 67.4 percent for the traditional method and 73.4 percent for the new method. For the Pacific population group, the response rate was 63.7 percent for the traditional method and 72.0 percent for the new method. For the Asian population group, the response rate was 79.5 percent for the traditional method and 85.6 percent for the new method. Finally, for the 15–29 years age group, for the traditional method the response rate was 73.5 percent and for the new method 79.5 percent. None of the sub-population census KPIs achieved their targets.

The previous two paragraphs included incorrect numbers when they were published and have been corrected.

Difficulty of enumeration 2018

Difficulty of enumeration, as used here and in previous PES results, is an approximation used to illustrate the relative importance of key components of the taking census process that contribute to coverage measures. Difficulty of enumeration, as used here and in previous PES results, is an approximation used to illustrate the relative importance of key components of the census process that contribute to coverage measures. It merely adds together the under-coverage, the over-coverage, the census late returns and the admin enumerations, which in previous censuses would have been addressed by substitute records. It adds together the under-coverage, the over-coverage, the census late returns, and the admin enumerations, which in previous censuses would have been addressed by substitute records. While not statistically rigorous, this measure is consistent with that provided with previous PES results for comparability.

Net undercount is only one indicator of the difficulty of census enumeration. Other indicators include:

- gross undercount and gross overcount – undercount and overcount both reflect errors in census enumeration, and the sum rather than the net difference is a better indicator of difficulty
- late returns – census forms returned late are an indication of slow response and extended field activity

- admin records.

The sum of undercount (171,800 records) and overcount (46,700 records) for the 2018 Census totals 218,500 records. There were 95,067 forms received late.

The number of substitute records increased from 132,945 in 2006 to 203,052 in 2013. In 2018, equivalent admin records numbered 524,853. In addition, there were 203,011 people from whom no individual form was received but who were recorded on the household listing (referred to in the census reports as 'partials'). The total number of partials and admin enumerations was, therefore, 727,767 records. The new method of counting responses meant that while the 'partials' did not contribute to difficulty of enumeration, they did contribute to the calculation of the response rate.

Both the traditional method and the new method of calculating response rates are presented in tables 6–8. The traditional method is the official method for the 2018 PES, because it is more comparable to the response rates published for previous censuses.

The new method is included to provide a better indication of the quality of response data in 2018 for comparability with future censuses. The new method reflects the changes in the New Zealand census methodology. The count of people now comes from two sources: census field responses and admin enumerations. The new collection response rate reflects the people who have been counted through a census field response and the measure of coverage measure highlights the people added through admin enumeration. As noted, it is also more consistent with international practice.

The number of late returns was substantially reduced between 2006 and 2013 from around 140,000 forms outstanding at the beginning of the PES collection phase in 2006 to 21,000 in 2013. The main reason for the reduction in 2013 was that PES field interviewing did not start until 9 April 2013. In 2006 the interviewing phase started on 21 March 2006. This later start in 2013 enabled census additional time to complete collection in the field. The later start in 2013 allowed additional time to complete collection in the field. A later start was planned for the 2018 PES, but events surrounding the 2018 Census resulted in an increase to 95,100 late returns, despite more time being allowed for the completion of the field collection.

The combination of these factors, including the undercount and the overcount, indicates that the degree of difficulty with enumeration was encountered for 17.4 percent of the census usually resident population count in 2018 (or using the old method, 21.6 percent). While not strictly rigorous, that is the approach used in previous PESs and is used here for consistency. This indicates the value of the delayed start date for PES field interviewing. The difficulty with enumeration percentage in 2013 was 9.0 percent, compared with 9.6 percent in 2006 (revised), 7.4 percent in 2001, and 6.2 percent in 1996. Table 6 summarises these details for each census

where there has been a PES. Because census methodology has changed across time comparisons are indicative only.

Table 6 : Difficulty of enumeration, Censuses 1996–2018

	1996 revised	2001 revised	2006 revised	2013 revised	2018 traditional method[3]	2018 new method[4]
	Number (000)					
Estimated gross undercount	68	107	106	135	172	172
Estimated gross overcount	8	22	17	32	47	47
Late returns[1]	46	40	140	21	95	95
Substitute records	102	107	133	203	728	525
Estimated total difficulty	223	276	395	390	1,042	839
PES estimated usually resident population	3,678	3,882	4,109	4,346	4,825	4,825
	Percent					
Difficulty of enumeration[2]	6.1	7.1	9.6	9.0	21.6	17.4
1. Late returns are census responses received after the date at which the PES begins interviewing in the field.						
2. Difficulty of enumeration is an approximation used to illustrate the relative importance of key components of the taking census process that contribute to coverage measures.						
3. Traditional method includes 2018 census partials (individuals on a household form from whom an individual form was not received plus rough sleepers). This measure is historically comparable.						
4. New method excludes 2018 census partials (individuals on a household form from whom an individual form was not received plus rough sleepers). This method will be used in the next PES.						

The difficulty of enumerations at a national level is only one dimension of this measure. The difficulty of enumeration varies across the population geographically, ethnically, by age, and by sex. To illustrate the level of diversity, table 7 shows the calculations for each of the six level 1 groupings of ethnicities. Geography, age, and sex each contribute to this pattern, with the highest difficulty being among people of Māori ethnicity or people of one or more Pacific ethnicities. These are both populations made up of younger age structures and who live in geographic areas that are traditionally hard to count.

Table 7 Difficulty of enumeration 2018, by ethnicity, traditional method[3]

	European	Māori	Pacific	Asian	MELAA	Other	Total
	Number (000)						
Estimated gross undercount	85	41	24	40	3	2	172
Estimated gross overcount	24	6	5	15	1	0	47
Late returns[1]	43	26	22	19	2	1	95
Admin including partials	369	224	123	113	13	9	728
Estimated total difficulty	521	298	174	187	19	12	1,042
PES estimated usually resident population	3,359	811	401	732	72	59	4,825
	Percent						
Difficulty of enumeration[2]	15.5	36.7	43.4	25.5	26.4	20.3	21.6

1. Late returns are census responses received after the date at which the PES begins interviewing in the field.

2. Difficulty of enumeration is an approximation used to illustrate the relative importance of key components of the taking census process that contribute to coverage measures.

3. Traditional method includes 2018 census partials (individuals on a household form from whom an individual form was not received plus rough sleepers). The measure is historically comparable.

Table 8 is similar to table 7 but uses the new method. It is included to enable a comparison between the two methods used to measure the difficulty of enumeration.

Table 8 Difficulty of enumeration 2018, by ethnicity, new method[3]

	European	Māori	Pacific	Asian	MELAA	Other	Total
	Number (000)						
Estimated gross undercount	85	41	24	40	3	2	172
Estimated gross overcount	24	6	5	15	1	0	47
Late returns[1]	43	26	22	19	2	1	95
Admin including partials	268	176	90	69	9	7	525
Estimated total difficulty	421	249	141	143	14	10	838
PES estimated usually resident population	3,359	811	401	732	72	59	4,825
	Percent						
Difficulty of enumeration[2]	12.5	30.7	35.2	19.5	19.4	16.9	17.4

1. Late returns are census responses received after the date at which the PES begins interviewing in the field.

2. Difficulty of enumeration is an approximation used to illustrate the relative importance of key components of the taking census process that contribute to coverage measures.

3. New method excludes 2018 census partials (individuals on a household form from whom an individual form was not received plus rough sleepers). This measure will be used in the next PES.

The diversity in the difficulty of enumeration across sectors of the population illustrates where censuses experience the greatest challenges to obtaining high coverage. This has major implications for future census enumerations, demonstrating the one key aspect of the value of PES.

() Coverage results

This chapter presents and analyses the undercount estimates derived from the 2018 PES, summarised by age, sex, ethnicity, and geographic area.

The 2018 PES results had been intended to be released in March 2020. The delay in release until 23 September was due to required remediation work on the PES as a result of needing to further modify the methodology, to both address the complication caused by the inclusion of the census admin enumerations used to enhance the 2018 Census dataset, and also to address an issue with how dwellings were selected to be surveyed by the PES.

The revisions to the 2018 Census model, described in 2018 Census file (methods/post-enumeration-survey-2018-methods-and-results#file) above required further methodological, development and processing work for PES. Coverage is traditionally measured against valid

completed individual census returns (often referred to as the contact sector). However, many fewer than expected individual census returns were received. The magnitude of the mitigation carried out for the 2018 Census, including both admin enumerations and imputation, meant that the planned PES methodology had to be modified because the PES was not originally designed for this approach.

There were also concerns with how dwellings were selected to be surveyed by the 2018 PES. Key issues related to violations of the PES 'source of truth' and assumptions about independence of PES from census, under-coverage of the PES dwelling frame, and the propensity for dwellings missed in the census to also be missed by the PES. Violations in these assumptions would lead to bias in the coverage estimates, and a change in our approach was therefore required.

The new estimation approach focused on person coverage, removing the dependence on dwelling coverage. Estimates of person coverage are therefore still produced under the new method, however dwelling estimates are not able to be produced and there is no feasible alternative using PES data. Dwelling estimates were therefore descoped from the Census Coverage Project.

Over the last six months, we have done intensive development work resulting in an improved methodology and robust results. This additional work also entailed input from international experts in the census-coverage field to help determine the approach to resolve these issues. There has also been an external review supporting our final methodology and modelling approach.

() Interim guidance in 2019 on under-coverage and PES results

The official PES net census undercount is 2.6 percent (credible interval 2.3–2.9 percent). Initial, interim guidance on coverage rates was provided in April 2019 (before the census dataset was finalised), and this indicated 1.2 percent net under-coverage, based on an initial best estimate. A later revision followed (July 2019) that implied net under-coverage of 1.4 percent. It was expected that the official robust PES under-coverage would be higher and is the best available assessment of coverage in the 2018 Census. It was expected that the official, robust PES under-coverage would be higher and is the best available assessment of coverage in the 2018 Census.

Several factors contributed to the observed differences. There was evidence that the DSE benchmark may have underestimated the true population, and therefore overestimated coverage of the 2018 Census. It is also possible that unobserved linkage error in the PES, as described in the methodological sections above, could have resulted in PES estimates being slightly too high, though well within credible intervals.

There was a difference between the interim coverage estimates for the 2018 Census based on the DSE benchmark and newer PES estimates, particularly for the youngest and oldest age groups. These are particularly notable for the age group 0–14 years, where the PES estimate is considerably higher than the DSE benchmark. The DSE benchmark is also lower for the age group 40–59 years (particularly around age 45 for people of Pacific ethnicities and around age 55 for people of Asian ethnicities).

On balance of evidence, there was a stronger argument for the DSE benchmark estimates being too low than there was for PES estimates being too high. In either case, the patterns of coverage were broadly similar and should not majorly impact on key interpretations of results.

PES median estimated population

The 2018 PES estimated that the usually resident population in New Zealand on census night (6 March 2018) was 4,824,600 people. This is the median estimate as calculated by the methodology outlined above. This estimate is within a credible interval of 4,810,700 and 4,839,800. This represents a net undercount of 2.6 percent (credible interval 2.3–2.9).

The net census coverage is the combined effect of under-coverage and over-coverage. In 2018 gross under-coverage was 171,800 people (credible interval 157,900–186,300), which was offset by gross over-coverage of 46,700 (credible interval 42,500–50,600). Note that these are median values and are not strictly additive.

The combination of net undercount (estimated by PES) and the number of records from other sources including admin enumerations (counted within census) gives a more complete picture of the effective non-response to census, as shown in table 9. Because the European, MELAA and Other groupings of ethnicities are modelled together, these groups are combined in this table.

Table 9 2018 PES estimated population, census sources, and estimated net undercount, selected groupings of ethnicities

Source	Māori		Pacific		Asian		European/MELAA/Other	
	Number (000)	Percent of estimated total	Number (000)	Percent of estimated total	Number (000)	Percent of estimated total	Number (000)	Percent of estimated total
2018 Census individual form	551	68.0	259	64.5	594	81.2	3,019	87.0
Other sources:								
- individuals on the household listing only	49	6.0	33	8.3	44	6.0	107	3.1
- admin enumerations	176	21.7	90	22.4	69	9.4	282	8.1
Total other sources	224	27.7	123	30.6	113	15.5	389	11.2
Census usually resident population count	776	95.6	382	95.1	708	96.7	3,408	98.2
Estimated net undercount	35	4.4	20	4.9	24	3.3	63	1.8
PES estimated population	811	100	401	100	732	100	3,471	100

Coverage by age and sex

The undercount, shown in figure 3 (methods/post-enumeration-survey-2018-methods-and-results#fig3) , was higher for males than females. Males had a net census under-coverage of 3.1 percent (credible interval 2.7–3.5) compared with females at 2.1 percent (credible interval 1.8–2.5).

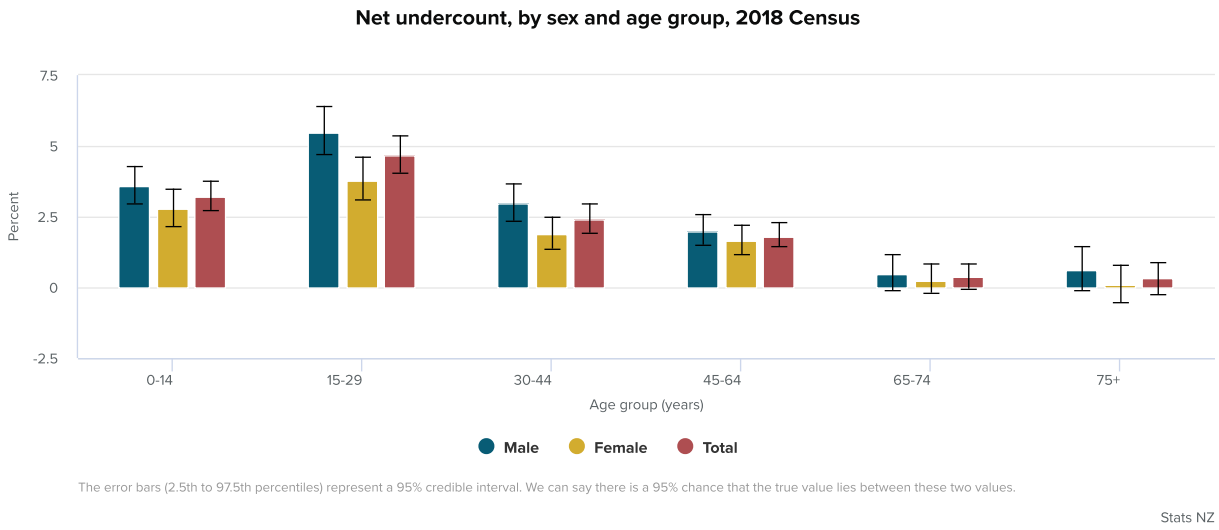
PES measured coverage for broad age groups. The age group with the highest under-coverage was found to be those aged 15–29 years at 4.7 percent (credible interval 4.0-5.4). This age group is by far the most difficult to enumerate, so this was expected, with a high degree of diversity of coverage within the age group.

The next highest under-coverage was the group aged under 15 years, at 3.2 percent (credible interval 2.7–3.7). This was perhaps less expected, as people in this age group could have been expected to have similar coverage to the age group containing the majority of their parents. are

found. For that age group, 30–44 years, the under-coverage was found to be 2.4 percent (credible interval 1.9–3.0).

Older ages were found to be much better enumerated. Older age groups were found to be much better enumerated. Those in the later working ages and older group (45–64 years) were found to have under-coverage of 1.8 percent (credible interval 1.4–2.3), the 65–74 years group had 0.4 percent (credible interval -0.1–0.8), and people aged 75 years and over had 0.3 percent (-0.2–+0.9). The credible intervals for the two oldest age groups show that the under-coverage was very small and may be a slight over-coverage. In contrast to the youngest ages, this was unexpectedly low given the enumeration difficulties experienced by the census.

() Figure 3



Coverage by ethnicity

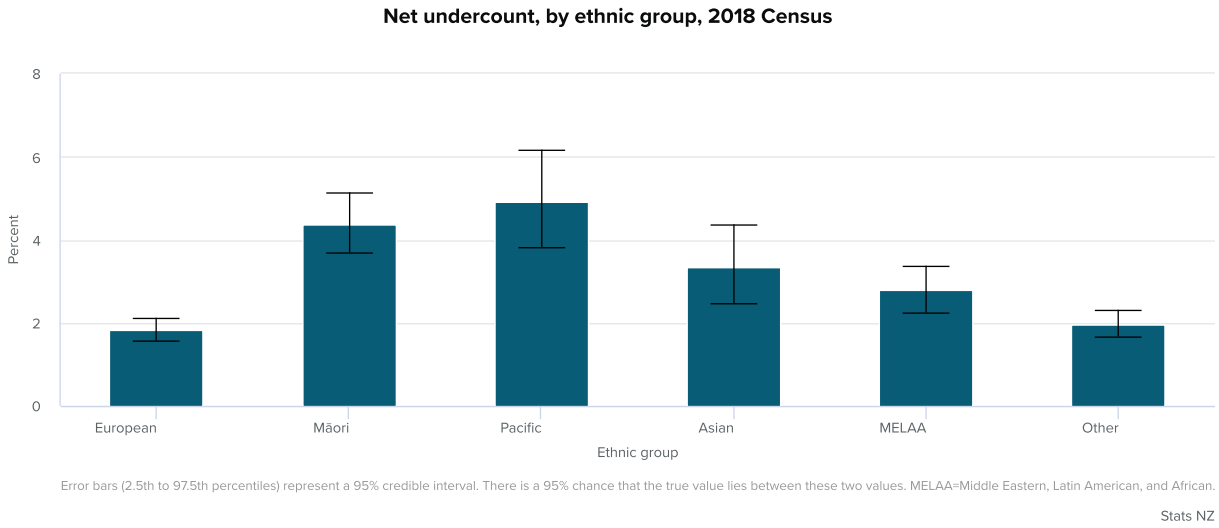
Ethnicity is one of the most important variables used in population analysis in New Zealand. Coverage of ethnic groupings in census data needs to be considered for careful population analysis and provides a core input for population estimation and the projection of future population change. It is important to observe that people may identify with ethnicities in more than one grouping. This means that groupings are not discrete entities and contain people who belong to more than one category in these data. In the case of the Māori population, more than half of the

population are multi-ethnic in census, with 66.7 percent for people of Māori aged under 15 years multi-ethnic.

In the case of the Māori population, more than half of the population were multi-ethnic in census, with 66.7 percent of people of Māori ethnicity aged under 15 years being multi-ethnic. In this youngest age group, 50.3 percent of the people of Pacific ethnicities and 30.7 percent of those of European ethnicities also identified with ethnicities in different groupings. The recording of multi-ethnic identifications is a global phenomenon (Aspinall & Rocha, 2020) and adds a level of complexity to the measurement coverage that is not possible to show fully in the available PES data, but helps explain some of the patterns that can be observed.

People of Pacific and Māori ethnicities had the highest median net under-coverage, as shown in figure 4. People of Pacific ethnicities had a net under-coverage of 4.9 percent (credible interval 3.8–6.1). People of Māori ethnicity had a net under-coverage of 4.4 percent (credible interval 3.7-5.1). This contrasts with people of European ethnicities at 1.8 percent (credible interval 1.6–2.1) and Other ethnicities at 2.0 percent (credible interval 1.7–2.3), and people of Asian ethnicities 3.3 percent (credible interval 2.5–4.3). People of MELAA ethnicities fall between Asian and European at 2.8 percent (credible interval 2.2–3.3). Care is needed in interpreting these differences because it is noteworthy that the credible intervals for Asian, Pacific, MELAA, and Māori are wide and do overlap.

Figure 4



European

People of European ethnicities account for 3,359,000 (69.6 percent) of the population, as estimated by PES in 2018. The population comprises 1,652,700 males (credible interval 1,646,400–1,659,700) and 1,706,400 females (credible interval 1,701,000– 1,712,000). In percent terms, as seen in figure 5, net under-coverage of the European population ranged from a high of 5.0 percent (credible interval 4.1–6.0) for males aged 15–29 years, to a slight net overcount of females aged 75 years and over of -0.1 percent (credible interval -0.6–0.5).

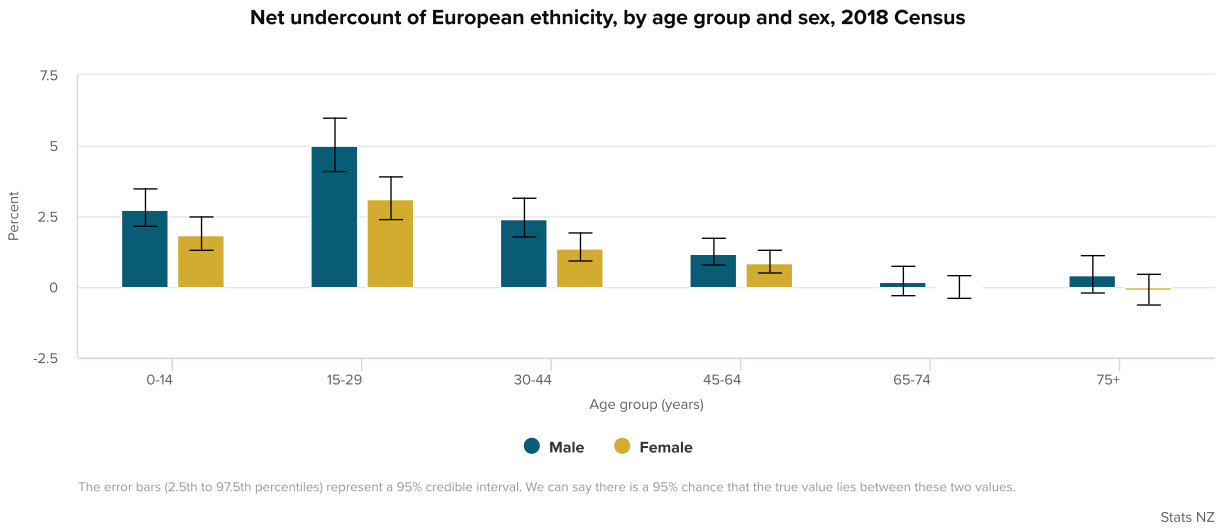
Net under-coverage for the youngest age group of males, those under 15 years, at 2.7 percent (credible interval 2.1–3.5), is only slightly higher than the 30–44 years age group, at 2.4 percent (credible interval 1.8–3.1). As with all ages, sexes, and ethnicities, the young adult age group, 15–29 years, had the highest under-coverage.

Coverage for the youngest age group of females, those under 15 years, was 1.9 percent (credible interval 1.3–2.5). This was significantly higher than the 30–44 years age group, at 1.4 percent (credible interval 0.9–1.9). As with all ages, sexes, and ethnicities, the young adult age group, 15–29 years, had the highest under-coverage at 3.1 percent (credible interval 2.4–3.9).

The coverage of females was better (that is, the under-coverage was smaller) for all ages. Among the two oldest age groups, it is notable that PES estimated that there may have been a slight overcount of females of European ethnicities, albeit with wide credible intervals, because of the

very small sample for these ages. Nevertheless, PES results suggested that census coverage was very good for these groups.

Figure 5



Māori

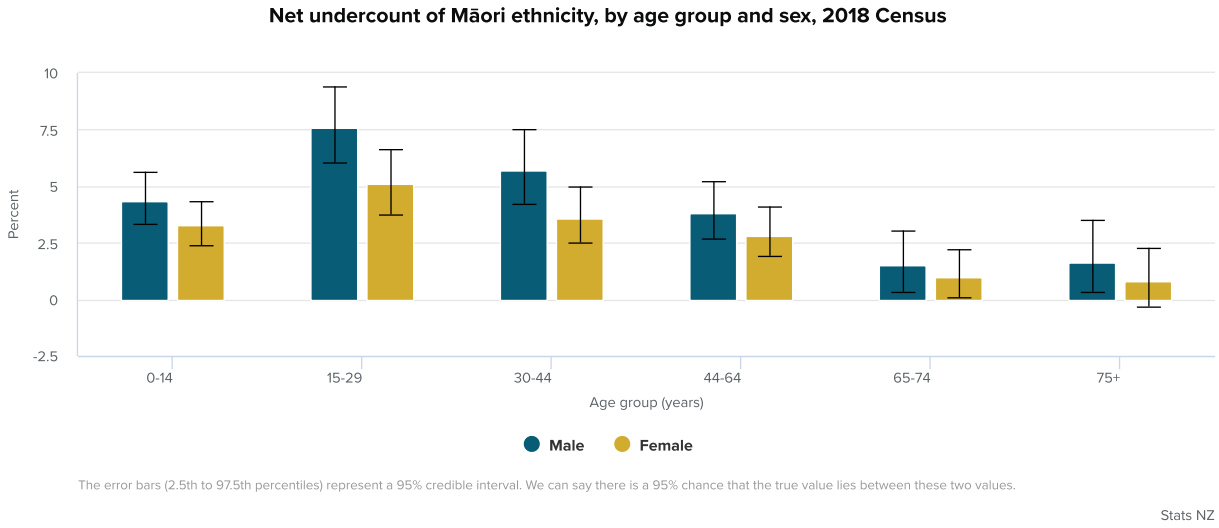
People of Māori ethnicity made up 16.8 percent of the population as estimated by PES in 2018. The Māori population had a much younger age structure than the European population, a demographic feature contributing in part to a much higher under-coverage.

People of Māori ethnicity account for 811,200 of the PES-estimated population in 2018. The population comprises 403,800 males (credible interval 400,100–408,300) and 407,300 females (credible interval 404,000–410,900).

Māori net under-coverage was higher than European under-coverage at all ages, as can be seen in figure 6. However, people of Māori ethnicity aged under 15 years had a lower net under-coverage at 4.3 percent (credible interval 3.3–5.6) for males (credible interval 3.3–5.6) and 3.3 percent for females (credible interval 2.4–4.3) than people of Māori ethnicity in the 30–44 years age group. Māori males aged 30–44 years had a net under-coverage of 5.7 percent (credible interval 4.2–7.5).

By far the highest under-coverage was, as with all groups, found in the 15–29 years age group. This is particularly important for Māori given that half the population is aged under 30 years, and a quarter are in the 15–29 years age group. This younger age structure goes some of the way, along with geographic distribution that is discussed below, to explaining a higher under-coverage for the ethnic group – among the European ethnic grouping, only 19 percent are in this age group.

Figure 6



Coverage of people of Māori descent/ancestry was also measured. The basis for coverage estimation was the census Māori descent variable (where responses of 'Don't Know' are imputed for calculating the electoral populations used to determine general and Māori electoral districts). This is because the core purpose of collecting Māori descent in PES was to support the electoral population estimates used for local government elections (as opposed to general elections). This meant that the fewer than 1 percent of people in PES, who responded that they did not know their descent status were imputed. This in no way implies that 'Don't Know' is not a valid response in census. This topic is flagged for further discussion in the next PES.

The Māori descent population is slightly older and, with an estimated population of 935,400 people (credible interval 929,400–942,100), is larger than the population of Māori ethnicity, with most people identifying with one also identifying with the other, but older ages more likely to identify ancestry but not report Māori ethnicity. Most people who identified with one also identified

with the other, but those in older age groups were more likely to identify as having Māori descent but not report Māori ethnicity. Because of these slightly different demographic features, the Māori descent population showed a slightly lower median net under-coverage at 4.2 percent (credible interval 3.5–4.8) compared with 4.4 percent (credible interval 3.7–5.1) for people of Māori ethnicity.

Pacific

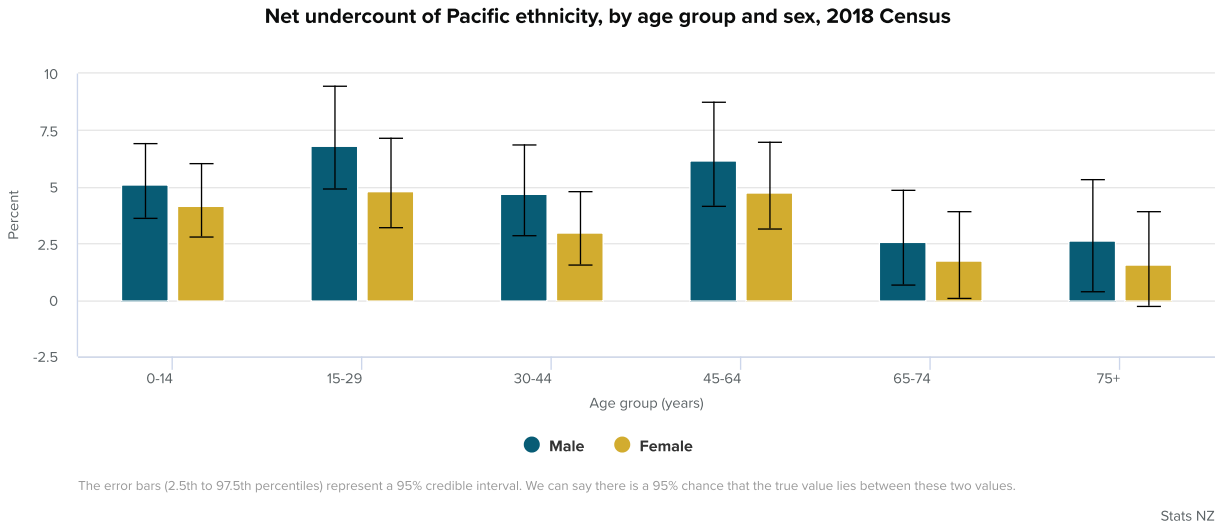
People of Pacific ethnicities form a young population, and approximately two-thirds were New Zealand born. While this is an increasingly diverse population that lives throughout New Zealand, the larger Pacific communities are found in southern Auckland. People of Pacific ethnicities make up 8.3 percent of the population as estimated by PES in 2018.

As with the Māori population, the Pacific has a much younger age structure than the European population, a demographic feature contributing in part to a much higher under-coverage. People of Pacific ethnicities account for 401,300 people. The population is made up of an estimated 202,700 males (credible interval 199,900–206,100) and 198,500 females (credible interval 196,100–201,600).

Among the Pacific population (figure 7), under-coverage of the males was higher than females for all age groups. Under-coverage for males aged under 15 years was only slightly higher than for males in the 30–44 years age group – 5.1 percent (credible interval 3.6–6.9) for the 0–14 years age group compared with 4.7 percent (credible interval 2.8–6.8) for those aged 30–44 years. Females by contrast had an under-coverage of 4.2 percent (credible interval 2.8–6.0) for those aged under 15 years, similar to their male counterparts, but a much lower under-coverage for those in the 30–44 years age group at 3.0 percent (credible interval 1.6–4.8).

As with all ages, sexes and ethnicities, the young Pacific adult ages, aged 15–29 years, had the highest under-coverage. Males again had a much higher under-coverage than females. For males it was estimated that 6.8 percent (credible interval 4.9–9.4) were missed by census, compared with 4.8 percent (credible interval 3.2–7.1).

Figure 7



Asian

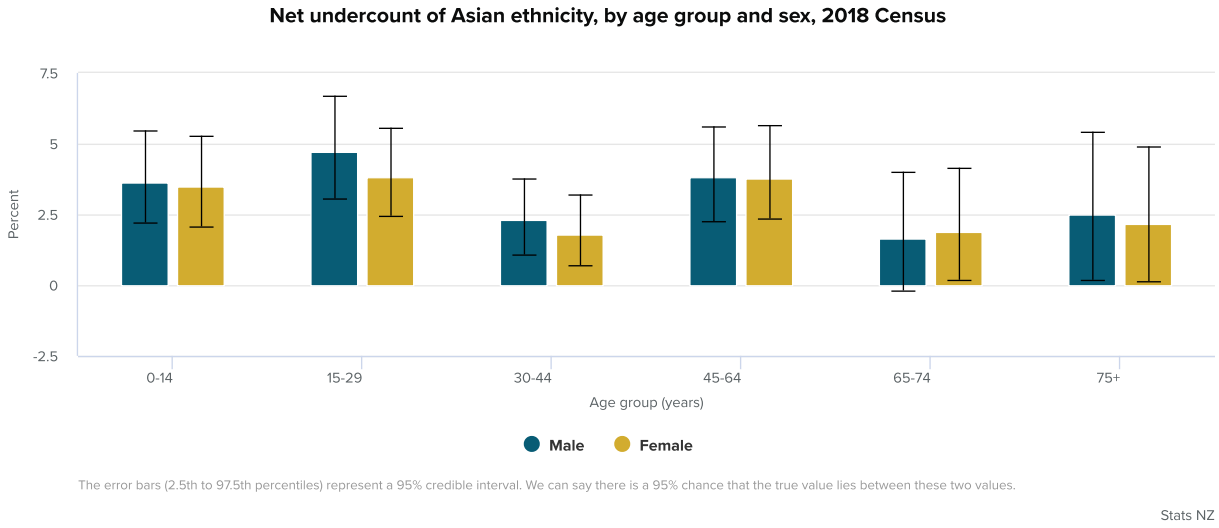
People of Asian ethnicities accounted for 732,000 people, 15.2 percent, in the population estimated by the 2018 PES. The population is young, with 74 percent aged under 44 years, and over half (54 percent) aged 15–44 years. This contributed to a slightly higher under-coverage for the Asian population than for the European population. Along with age structural differences, migration histories and internal diversity distinguishes the Asian group from the European population. The under-coverage for people of Asian ethnicities was estimated to be 3.3 percent (credible interval 2.5–4.3).

The New Zealand Asian population was estimated to have 361,700 males and 370,200 females, with females numbering slightly more than males. As stated, 54 percent of the Asian population was aged 15–44 years. Among males, 29 percent were aged 15–29 years and 27 percent were aged 30–44 years, contrasting with females where 25 percent were aged 15–29 years and 28 percent were aged 30–44 years. The difference became stronger among the 45–64 years age group where 17 percent of Asian males and 21 percent of Asian females are found.

Among the Asian population, under-coverage for males and females was much more similar than for other groupings of ethnicities. Figure 8 shows that Asian males aged under 15 years had an under-coverage of 3.6 percent (credible interval 2.2–5.5), essentially the same as for females at 3.5 percent (credible interval 2.1–5.3).

The Asian younger adult age groups also revealed higher under-coverage. In the younger 15–29 years age group, males had a higher under-coverage, at 4.7 percent (credible interval 3.0–6.7) than females at 3.8 percent (credible interval 2.4–5.5). In the older age group, those aged 30–44 years, males had an under-coverage of 2.3 percent (credible interval 1.0–3.7) and females had an under-coverage of 1.8 percent (credible interval 0.7–3.2).

Figure 8



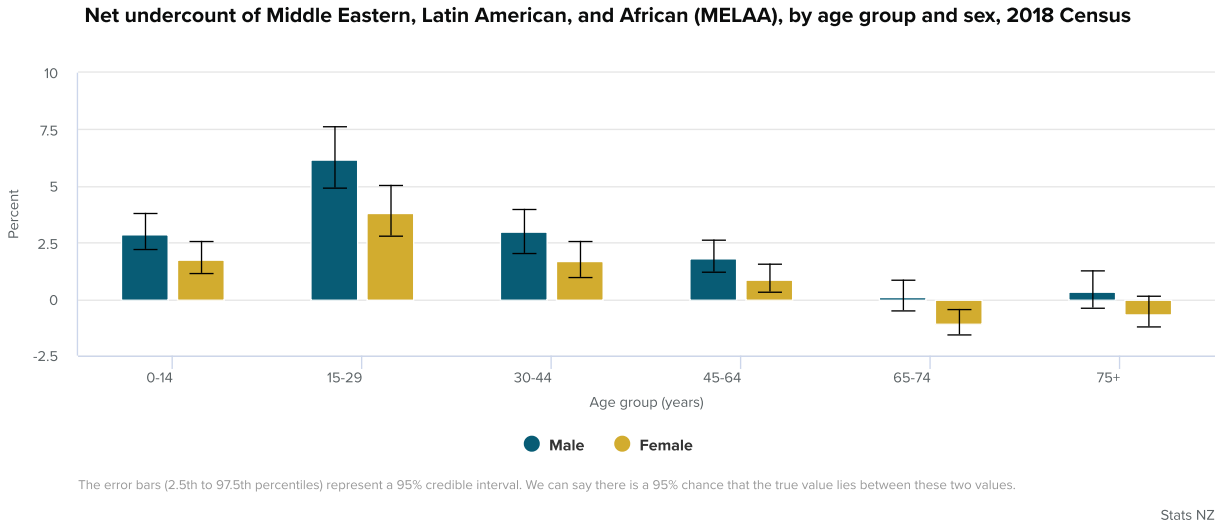
Middle Eastern, Latin American, and African (MELAA) and Other

People in the MELAA and Other groupings of ethnicities are two very much smaller populations. The coverage measures reported here are included for completeness.

The MELAA grouping is a small population (estimated 72,300 people, credible interval 71,900–72,800) and one that is hugely diverse and not forming a coherent set of communities, encompassing a very large number of ethnicities spread across the Middle East, Africa and Latin America.

The observed net under-coverage in PES for the MELAA grouping was 2.8 percent (credible interval 2.2–3.3). Figure 9 shows that males had a higher under-coverage for all ages, with by far the highest under-coverage, as with other ethnic groupings, found among those aged 15–29 years.

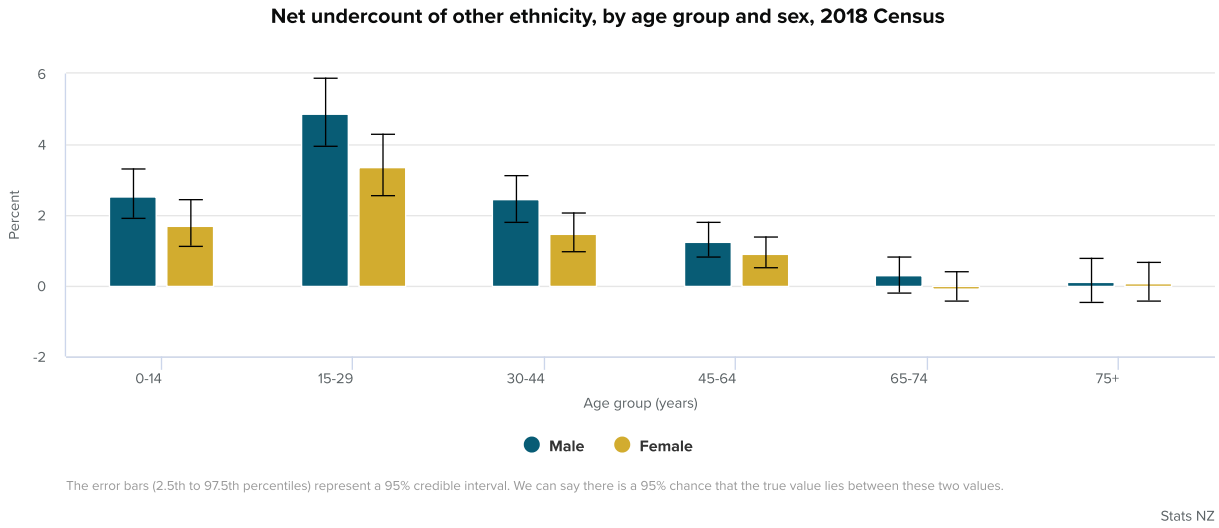
Figure 9



The vast majority of people in the Other grouping of ethnicities are people who gave ‘New Zealander’ as one of their ethnicities. In the 2018 Census, 58,053 people were counted in the Other grouping of whom 45,330 were of New Zealander ethnicity. The only other ethnic category in this grouping with significant numbers was ’other South African’ with 6,816 people coded as such. This grouping is demographically similar to the European grouping of ethnicities and coverage patterns are also similar.

The Other grouping of ethnicities had an under-coverage of 2.0 percent (credible interval 1.7–2.3). This is very similar to that for European grouping of 1.8 percent (credible interval 1.6–2.1). As figure 10 shows, the age and sex patterns were also very similar to the European patterns.

Figure 10

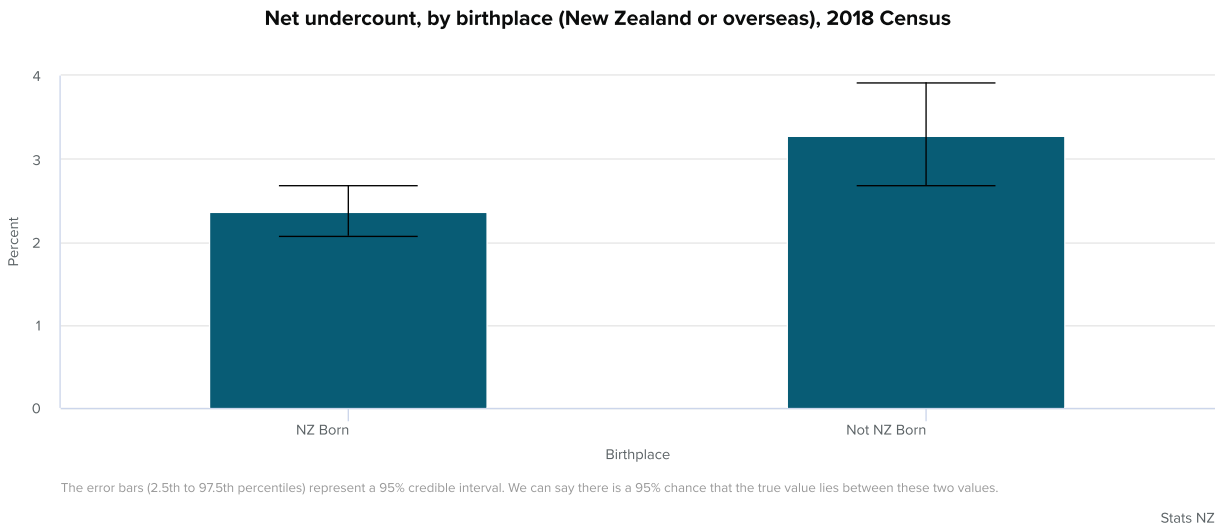


People born in New Zealand and overseas

Birthplace was collected in the 2018 PES. New Zealand has gone through a period of high immigration, associated in part with the need for temporary labour in major projects such as the Christchurch rebuild and major construction projects in Auckland. There have been ongoing shortages of skilled labour in a number of health and agricultural industries. The migrant population, including international students, are often living in New Zealand for longer than 12 months, therefore defined as usually resident, or become permanent residents of this country, but are often highly mobile and harder for census to engage with. Coverage by census of the population is therefore affected when this sector of the population grows rapidly. The effect is noticeable at the national level and also by geography and ethnicity, though these dimensions are not reported on here.

As expected, people born overseas, and this includes both recent and long settled migrants from a vast range of source countries, had a higher net under-coverage than the New Zealand born. As expected, people born overseas, and this includes both recent and long settled migrants from a vast range of source countries, had a higher net under-coverage than those born in New Zealand. People born overseas, shown in figure 11, had a net under-coverage of 3.3 percent (credible interval 2.7–3.9) compared with those born in New Zealand at 2.4 percent (credible interval 2.1–2.7).

Figure 11



Coverage by geographies

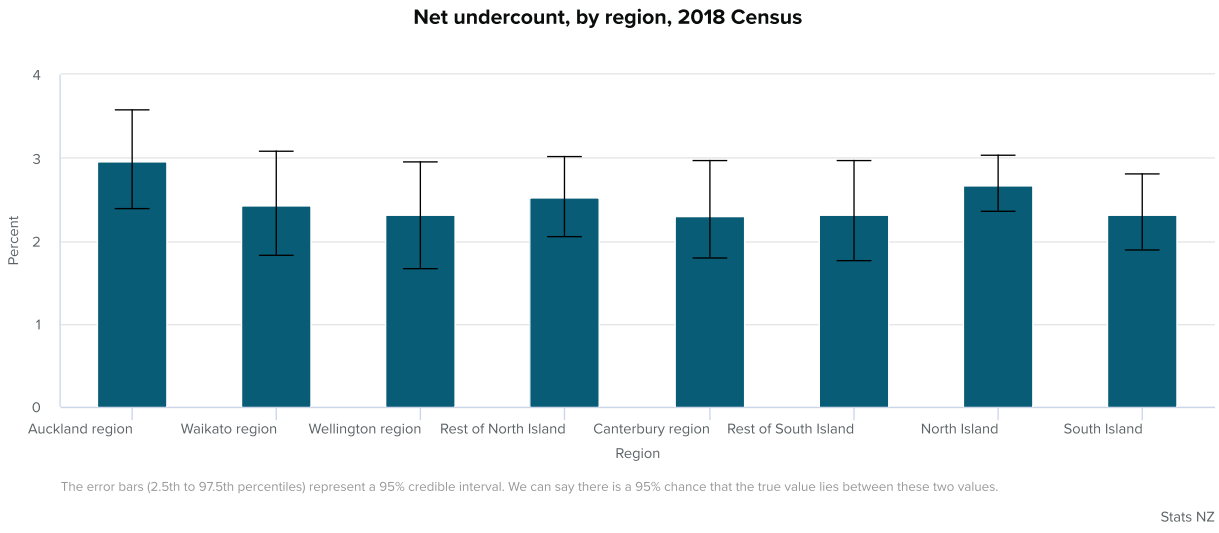
Three-quarters of the New Zealand resident population (76.5 percent) lives in the North Island. The estimated population of the North Island was 3,693,200 (credible interval 3,681,100–3,706,700), with 43.8 percent in Auckland (1,619,600 – credible interval 1,610,200–1,630,000).

The demographic composition of Auckland contributes to the higher under-coverage of the North Island. Auckland is home to 64.0 percent of the Pacific population of New Zealand, and 23.4 percent of the Māori population – both young populations with high under-coverage. Moreover, 41.7 percent of the population were born overseas, also a population with higher under-coverage. The combined effect of these demographic factors resulted in Auckland having a net under-coverage of 3.0 percent (credible interval 2.4–3.6) compared with, for example, less diverse regions such as Wellington, which had an under-coverage of 2.3 percent (credible interval 1.7–3.0) or Canterbury, also with a net under-coverage of 2.3 percent (credible interval 1.8–3.0). These results are shown in table 10, with the percentage under-coverage shown graphically in figure 12.

Table 10: Coverage, broad geographic regions, net median under-coverage 2018 PES

Selected geographies	Number of people		Net under-coverage	Credible interval	
	Census count	Estimate by PES	Median	Lower bound	Upper bound
			Percent		
Auckland region	1,571,718	1,619,600	3.0	2.4	3.6
Waikato region	458,202	469,600	2.4	1.8	3.1
Wellington region	506,814	581,800	2.3	1.7	3.0
Rest of North Island	1,057,818	1,085,300	2.5	2.1	3.0
Canterbury region	599,694	613,700	2.3	1.8	3.0
Rest of South Island	504,843	516,800	2.3	1.8	3.0
North Island	3,594,552	3,693,200	2.7	2.4	3.0
South Island	1,104,537	1,130,600	2.3	1.9	2.8

Figure 12



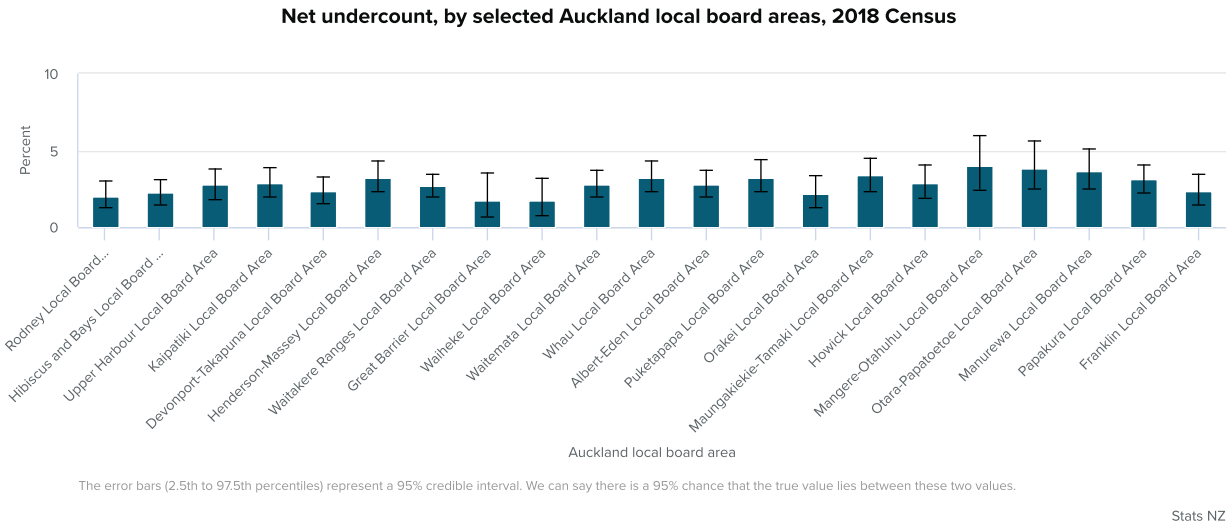
The wider Auckland demographic factors described above contribute to the wide diversity of coverage measures across Auckland. Advances in estimation methodology, describe earlier in this report, have enabled more robust estimation of coverage for Local Board Areas (LBAs) within Auckland. Measurement of coverage of these areas was an important part of evaluating outcomes experienced by the 2018 Census and for the effective planning of solutions for the 2023 Census. Southern Auckland areas in particular have been cited as of particular concern.

Concerns with coverage in Auckland are borne out in the results. LBAs such as Rodney, Ōrākei, and Devonport-Takapuna have much lower coverage (that is, higher net median undercoverage) than the more difficult to enumerate areas such as Mangere-Ōtāhuhu, Papatoetoe, and Mānurewa. Manurewa. Higher under-coverage is also seen in central areas such as Whau and Albert-Eden. Results for Auckland Local Board Areas are summarised in table 11, with the under-coverage shown graphically in figure 13.

Table 11 Coverage, Auckland Local Board Areas, 2018 PES

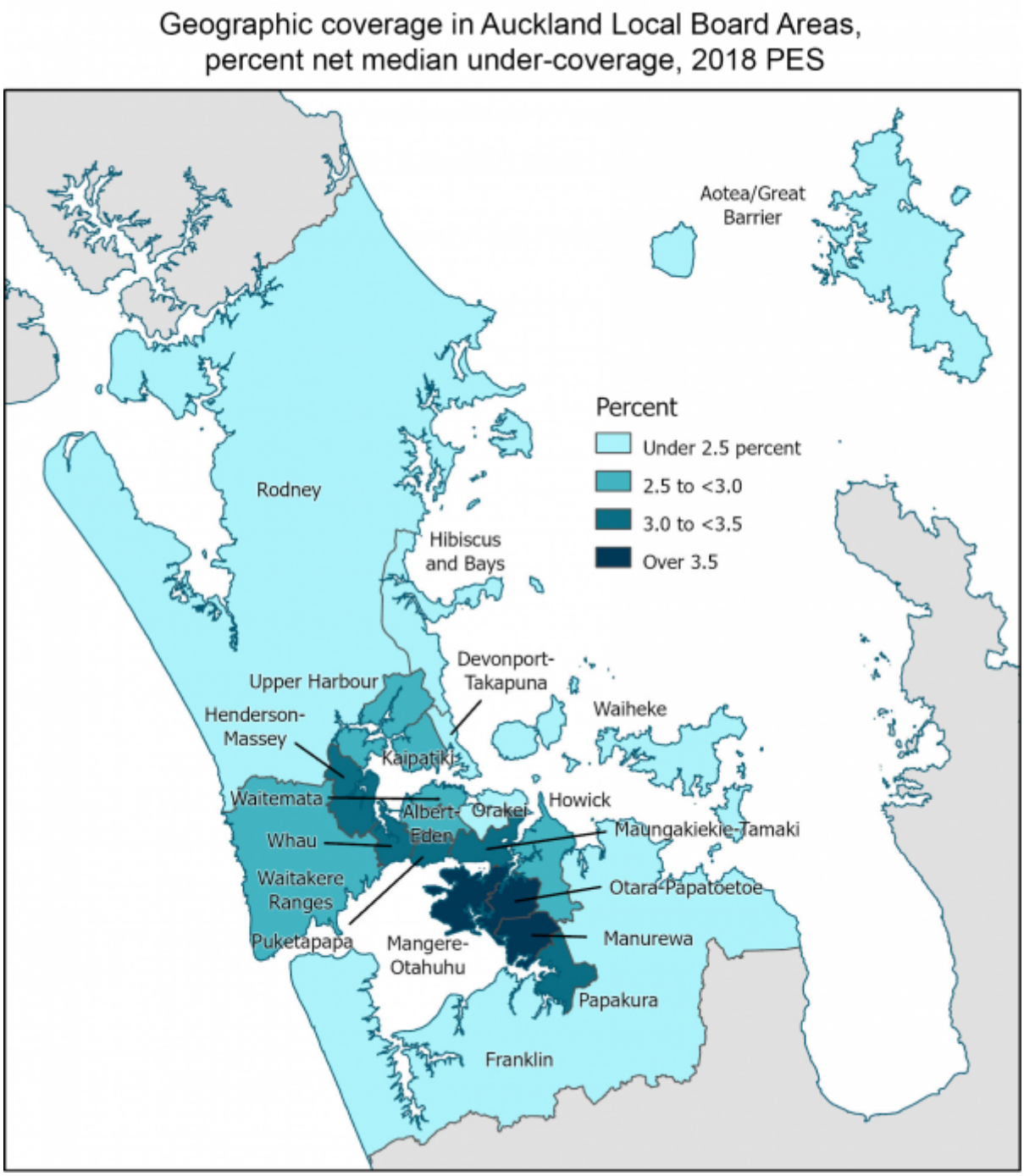
Auckland	**Number of people**		**Net under-coverage**	**Credible interval**	
Local Board Area	Census count	Estimate by PES	Median	Lower bound	Upper bound
			Percent		
Rodney	66,417	67,800	2.0	1.2	3.0
Hibiscus and Bays	104,010	106,400	2.2	1.5	3.1
Upper Harbour	62,841	64,600	2.8	1.8	3.8
Kaipatiki	88,269	90,900	2.9	2.0	3.9
Devonport-Takapuna	57,975	59,400	2.4	1.5	3.3
Henderson-Massey	118,422	122,400	3.2	2.3	4.3
Waitakere Ranges	52,095	53,500	2.7	2.0	3.5
Waiheke	9,063	9,200	1.8	0.8	3.2
Waitemata	82,866	85,300	2.8	2.0	3.7
Whau	79,356	82,000	3.2	2.3	4.3
Albert-Eden	98,622	101,500	2.8	2.0	3.7
Puketapapa	57,555	59,500	3.2	2.3	4.4
Orakei	84,318	86,200	2.2	1.3	3.3
Maungakiekie-Tamaki	76,284	78,900	3.4	2.3	4.5
Howick	140,970	145,100	2.8	1.9	4.0
Mangere-Otahuhu	78,450	81,700	4.0	2.4	6.0
Otara-Papatoetoe	85,122	88,500	3.8	2.5	5.6
Manurewa	95,670	99,300	3.6	2.5	5.1
Papakura	57,636	59,500	3.1	2.2	4.0
Franklin	74,838	76,600	2.3	1.4	3.4

Figure 13



By mapping the Local Board Areas in Auckland (figure 14), the distinct geographic pattern of median net under-coverage becomes apparent. The highest median net under-coverage is seen in a swathe of LBAs across southern and inner suburban Auckland, with median net under-coverage across the rest of Auckland around the national level. Notably, the two LBAs with the lowest under-coverage (Rodney and Waiheke) are on the periphery of Auckland. This pattern demonstrates close connection between the demography of Auckland (Spoonley, 2020) and the distribution of under-coverage.

Figure 14



Text alternative for figure 14: Geographic coverage in Auckland Local Board Areas, percent net median under-coverage, PES 2018

The figure maps the median net under-coverage of Local Board Areas in Auckland on a four-category split:

- highest net median under-coverage (3.5 percent or greater) was observed in the Otara-Papatoetoe, Manurewa and Mangere-Otahuhu Local Board Areas
- the Local Board Areas with net median under-coverage of 3.0 percent but less than 3.5 percent included Papakura, Maungakiekie-Tamaki, Puketapu, Whau and Henderson-Massey.
- the Local Board Areas with net median under-coverage of 2.5 percent but less than 3.0 percent comprised the Waitakere Ranges, Upper Harbour, Kaipatiki, Waitemata, Howick, and Albert-Eden local board areas.

- the remaining seven local board areas had a net median under-coverage of less than 2.5 percent.

() Looking forward

The high economic, social and political value of census is reflected in the fiscal commitment to achieve a census of high quality in 2023. The high economic, social, and political value of census is reflected in the fiscal commitment to achieve a census of high quality in 2023. The assurance, and demonstration, of quality is a core responsibility of the census-taking process. One of the important components of quality is coverage. This includes both broad assessment of the accuracy of the census count of people, but also the coverage of important sectors of the population. Robust measurement of coverage remains central to this assessment, to be measured in 2023 by a PES, but with sufficient flexibility to remain responsive to unplanned outcomes in census.

Dwelling coverage

Dwelling coverage will not be a priority in 2023. The 2018 PES set out to estimate the coverage in census of permanent private dwellings. Measurement of dwelling coverage has become increasingly more challenging as the census approach to counting dwellings has shifted to include greater use of administrative registers and lists. With the 2018 Census moving to increased use of an administrative address list for dwellings in 2018, PES faced challenges obtaining an independent high-quality list of dwellings. These data challenges proved too much for estimation methods and we are unable to provide robust estimates of the coverage of dwellings in the 2018 Census. Funding and methodological constraints will again limit our capacity to address this important dimension of coverage in 2023.

Person coverage

Person coverage will take centre stage. In 2023, census is planning to incorporate by design admin enumerations and admin data to mitigate missingness. PES will therefore base measurement of coverage on this combined census model with a particular focus on doing this by design.

There are a number of key elements of coverage measurement that differ when considering a combined model compared to a traditional census model. These include:

- increased risk of over-coverage of people via erroneous inclusions in the admin population resulting in admin enumerations included that should not have been (for example, not a usual resident of New Zealand on census night)
- increased risk of over-coverage of people due to linkage error between the census and admin population resulting in duplication of an existing census response

- increased risk of under-coverage of people due to over-coverage records simultaneously camouflaging under-coverage because the correct missing person has not been included.

The situations outlined above are priority areas for research and investigation ahead of 2023 PES. Work will firstly focus on assessing the robustness of assumptions made in 2018 PES, followed by improvements to design and estimation to minimise any limitations.

For the purposes of 2018 PES, it was assumed by both PES and census that all non-residents had been successfully removed from the admin data prior to selection into the census file. While this assumption has not been specifically tested in 2018, measurement any over-coverage of this sort should be tested in 2023. While this assumption has not been specifically tested in 2018, measurement of any over-coverage of this sort should be tested in 2023. Duplicates between census responses and admin enumerations should be captured by the PES linking methodology in the same way we would expect duplicate census responses are already captured. PES uses a combination of automatic and manual linking in efforts to heavily reduce linkage error, including extensive manual processes focused on resolving any false negative links. For 2018 PES, we also linked the PES records to the IDI spine and used the three-way links to detect any missed links, and to resolve any discrepancies. However, it is possible that our decision to use high quality records only for linking reduced our ability to detect duplicates due to linkage error. Future work is planned to investigate this and identify alternatives for 2023.

Additional considerations for 2023 PES estimation are the range of options available for treating late census returns, as well as the need to ensure we have an adaptive estimation design. Experience in 2018 illustrated the importance of this and we need to be prepared for the possibility that census may not perform as expected. A range of census performances and non-response levels may again require mitigations by census despite the best efforts to design for such events. PES will need to develop a design that is sufficiently adaptive to provide robust measures of coverage in case there is a repeat of the situation, or a variation of it, faced in 2018.

Beyond the measurement of coverage is the issue of the use of those measures – an analytical function rather than methodological. This includes examining patterns of under-coverage and over-coverage, analysis of the implications of these patterns, and the potential learnings that can contribute to ongoing improvements in both the understanding and the performance of census.

Central to this analysis are a set of conceptual frameworks that aim to identify positive assumptive worldviews (Lerner, 1980; Janoff-Bulmar & Carnes, 2016) which have well-established cultural and psychological underpinnings. People of Māori and Pacific ethnicities, young adults, specific geographic areas, new migrants and non-English speakers are among the many groups repeatedly identified as hard-to-count and therefore over-represented in census under-coverage, not all of which the PES has traditionally been designed to measure. People of Māori and Pacific ethnicities, young adults, specific geographic areas, new migrants, and non-English speakers are

among the many groups repeatedly identified as hard-to-count and therefore over-represented in census under-coverage, not all of which the PES has traditionally been designed to measure.

One core question is why particular groups are over-represented in missingness. The implicit assumption behind the concept of hardness-to-count is based on a blend of systematic victim-blaming and problematising the other – both well-established components of the mismatch between intentions and outcomes. From the first perspective, the assumption is that under-coverage is the fault of the people who are missed because they have specific characteristics. From the other perspective, the input presumption of the subject group being problematic contributes to the outcomes. The machinery of coverage estimation is premised on the philosophic underpinnings of the machinery of census – in both cases the same worldviews are used. There is a strong possibility that the hardness-to-count stems not from the people who census and/or PES have failed to reach, but rather stems from the collection models themselves. In other words, people may be missed not because they are poor, mobile, young, or whatever. They may be missed because the mechanisms that should include them have intrinsic features that unwittingly exclude them, in a fashion similar to that described in relation to educational marginalisation (Nakhid, 2002). Mitigation of this type of confirmation bias would be beneficial to a better understanding of both census and survey outcomes.

() References

Abbott, O & dos Santos, M (2011). The design of the 2011 Census Coverage Survey. Proceedings of the ISI Conference 2007. Southampton, University of Southampton, Southampton Statistical Sciences Research Institute.

Aspinall, P & Rocha, Z (2020). *The Palgrave International Handbook of Mixed Racial and Ethnic Classification*. Cham, Switzerland, Palgrave Macmillan.

Bryant J, Dunstan K, Graham P, Matheson-Dunning N, Shrosbree E, & Speirs R (2016). Measuring uncertainty in the 2013-base estimated resident population (http://archive.stats.govt.nz/methods/research-papers/working-papers-original/measure-uncertainty-2013-erp.aspx#gsc.tab=0) (Statistics New Zealand Working Paper No 16–04). Retrieved 4 September 2020 from www.stats.govt.nz.

Dunstan, K, Heyen, G, & Paice, J (1999). *Measuring census undercount in Australia and New Zealand*. Demography working paper 99/4. Canberra, Australian Bureau of Statistics.

Fellegi, I & Sunter, A (1969). A theory of record linkage. *Journal of the American Statistical Association, 40*, 1183–210.

Jack, M & Graziadei, C (2019). Report of the Independent Review of New Zealand's 2018 Census (reports/report-of-the-independent-review-of-new-zealands-2018-census) . Retrieved 4 September 2020 from www.stats.govt.nz.

Janoff-Bulman, R & Carnes, N (1916). Social justice and social order: binding moralities across the political spectrum. *PLoS One 11*(3), e0152479.

Kerr, D (1998). *A review of the procedures for estimating the net undercount of censuses in Canada, the United States, Britain and Australia*. Research paper 5. Canberra, Australian Bureau of Statistics.

Lerner, M (1980). *The belief in a just world.* New York, Springer.

Marks, E, Maudlin, W P, & Nisselson, H (1953). The Post-Enumeration Survey of the 1950 Census: A Case History in Survey Design. *Journal of the American Statistical Association, 48*(262), 220–43.

Miller, K (1922). Enumeration errors in Negro population. *Scientific Monthly, 14*(2), 168–177.

Nakhid, C (2002). Who do you say I am? Paper presented to the Annual Meeting of the Comparative and International Education Society, Orlando, 6–9 March, 2002.

National Records of Scotland (2013). *Census in Scotland: How the population estimates were obtained – Release 1C*. Edinburgh, Author. Available from: www.scotlandscensus.gov.uk.

OECD (2001). Glossary of statistical terms. Available from: https://stats.oecd.org/glossary/detail.asp?ID=2079 (https://stats.oecd.org/glossary/detail.asp?ID=2079)

O'Hare, W (2019). *Differential undercounts in the U.S.Census. Who is missed?*. Cham, Springer.

Pereira, R (2002). The Census Coverage Survey – the key element of a One Number Census. *Population Trends 108*: 16–31.

Shryock, H & Siegel, J (1973). *The methods and materials of demography*. Washington, US Department of Commerce.

Spoonley, P (2020). *The New New Zealand*. Auckland, Massey University Press.

Stats NZ (1998). *A report on the 1996 post-enumeration survey*. Wellington, Author.

Stats NZ (2002). *A report on the 2001 post-enumeration survey.* Wellington, Author.

Stats NZ (2005). *Statistical standard for ethnicity*. Wellington, Author.

Stats NZ (2006). 2006 Census of population and dwellings – introduction to the census (http://archive.stats.govt.nz/Census/about-2006-census/introduction-to-the-census.aspx#gsc.tab=0) . Retrieved from www.stats.govt.nz.

Stats NZ (2014). Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey (http://archive.stats.govt.nz/browse_for_stats/population/census_counts/report-on-2013-post-enumeration-survey.aspx#gsc.tab=0) . Retrieved from www.stats.govt.nz.

Stats NZ/Tatauranga Aotearoa (2019). 2018 Census: Interim coverage rates, collection response rates, and data sources (reports/2018-census-interim-coverage-rates-collection-response-rates-and-data-sources) . Retrieved 4 September 2020 from www.stats.govt.nz.

Winkler, W (1990). *String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage*. Washington, US Bureau of the Census.

2018 Census External Data Quality Panel (2019a). Initial report of 2018 Census External Data Quality Panel (reports/initial-report-of-the-2018-census-external-data-quality-panel) . Retrieved 4 September 2020 from www.stats.govt.nz.

2018 Census External Data Quality Panel (2019b). 2018 Census External Data Quality Panel: Assessment of variables (reports/2018-census-external-data-quality-panel-assessment-of-variables) . Retrieved 4 September 2020 from www.stats.govt.nz.

2018 Census External Data Quality Panel (2020). Final report of the 2018 Census External Data Quality Panel (reports/final-report-of-the-2018-census-external-data-quality-panel) . Retrieved 4 September 2020 from www.stats.govt.nz.

() Appendix: summary tables and questionnaires

See Excel summary tables in Estimated resident population (2018-base): At 30 June 2018 (information-releases/estimated-resident-population-2018-base-at-30-june-2018) .

2018 Post-enumeration survey person questionnaire (https://cdm20045.contentdm.oclc.org/digital/collection/p20045coll2/id/1095/rec/3URVIScdm20045.

2018 Post-enumeration survey household questionnaire (https://cdm20045.contentdm.oclc.org/digital/collection/p20045coll2/id/1096/rec/2INTcdm20045.co

ISBN 978-1-99-003210-3 (online)

Enquiries

Joel Watkins
03 964 8807 (http://tel+64-3-964-8807)
info@stats.govt.nz (mailto:info@stats.govt.nz)

Related pages

Post-enumeration survey: 2018 (/information-releases/post-enumeration-survey-2018)

23 September 2020

Estimated resident population (2018-base): At 30 June 2018 (/information-releases/estimated-resident-population-2018-base-at-30-june-2018)

23 September 2020

Estimated resident population 2018: Data sources and methods (/methods/estimated-resident-population-2018-data-sources-and-methods)

23 September 2020