Statistics New Zealand
TATAURANGA AOTEAROA

Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey

New Zealand Government

Crown copyright ©
This work is licensed under the Creative Commons Attribution 3.0 New Zealand licence. You are free to copy, distribute, and adapt the work, as long as you attribute the work to Statistics NZ and abide by the other licence terms. Please note you may not use any departmental or governmental emblem, logo, or coat of arms in any way that infringes any provision of the Flags, Emblems, and Names Protection Act 1981. Use the wording 'Statistics New Zealand' in your attribution, not the Statistics NZ logo.

Liability statement
While all care and diligence has been used in processing, analysing, and extracting data and information in this publication, Statistics New Zealand gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation
Statistics New Zealand (2014). *Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey.* Available from www.stats.govt.nz.

ISBN 978-0-478-42921-3 (online)

Published in August 2014 by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact
Statistics New Zealand Information Centre:
info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4610

joel.watkins@stats.govt.nz

www.stats.govt.nz

Contents

List of tables and figures

Tables

Figures

Purpose and summary

Purpose

This report describes and discusses the salient features of the 2013 PES, including its scope and methodology. The information gathered and the results are important for post-censal population estimation and other demographic series.

Summary

The report is organised into four chapters and an appendix. The highlights (below) summarise the survey results and changes to the methodology. They are described further in the following chapters.

Chapter 1 outlines the significance of the PES, its history, and the known sources of census miscount.

Chapter 2 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.

Chapter 3 describes the survey's main findings. The results include an analysis of differences in census coverage by sex, age, ethnicity, and broad geographic area. The 2013 PES results are also put into an international context.

Chapter 4 covers the work underway to improve PES estimation methodology in preparation for the PES following the 2018 Census. This chapter also advises on the relationship between the PES and the official estimated resident population.

The appendix contains a list of acronyms and definitions of key terms.

Highlights of the 2013 Post-enumeration Survey

Statistics New Zealand conducted a post-enumeration survey (PES) to measure the completeness of census coverage achieved in the 2013 Census of Population and Dwellings, which was held on 5 March 2013. The PES interviewers were in the field 9 April–10 May 2013.

Coverage results from the 2013 PES showed that:

- the census counted 97.6 percent of New Zealand residents in the country on census night with a sampling error of ±0.5 percentage points
- the national net undercount was 2.4 percent, or 103,800 people
- the net undercount resulted from an estimated undercount of 135,500 people, offset by 31,700 people being counted more than once (overcount)
- while the 2013 Census counted 4.24 million residents in New Zealand, the 2013 PES results suggest that the number of New Zealand residents in New Zealand on census night was closer to 4.35 million
- males (2.6 percent) had a higher net undercount than females (2.1 percent)
- younger adults – those aged 15–29 years – (4.8 percent) had a higher net undercount than other age groups

- people aged 65 years and over had the lowest net undercount, by age group, at 0.5 percent
- the net undercount for Māori (6.1 percent) and Pacific (4.8 percent) populations, which have young age structures, was higher than for European (1.9 percent) and Asian (3.0 percent) populations
- the northern North Island had a net undercount of 3.0 percent, compared with 1.3 percent for the rest of the North Island, and 2.2 percent for the South Island
- of the three major urbanised regions, Canterbury had the highest net undercount (2.9 percent), compared with Auckland (2.1 percent), and Wellington (0.7 percent)
- the total non-response rate to the census was 7.1 percent. Total non-response is the number of people for whom a substitute form was raised in the census, and who are therefore included in the census count, plus the net undercount as measured by the PES.

Note: All undercount estimates quoted above have a sampling error, and in most cases, differences in the undercount rates between subgroups are not statistically significant. Methodological changes in the 2013 PES also mean that comparisons across time should be made with caution.

The two major changes in the 2013 PES were the increased sample size and the introduction of automatic matching:

- Almost 15,000 households were surveyed, up from 11,000 in 2006, with completed PES forms containing information about nearly 33,000 people.
- A new initiative for 2013 was sampling in the field, which enabled interviewers to add new dwellings to the sample in certain situations, usually when a house had recently been converted into flats. This resulted in about 30 dwellings being added to the sample after the initial sample extraction.
- Automatic matching involved the electronic linking of PES to census records, improving both the efficiency and the accuracy of the PES.
- The impact of automatic matching on coverage estimates was measured by a matching impact study.

1 Introduction

A census of population and dwellings is often the largest data-gathering exercise in any country. 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.

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, business, 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, Statistics New Zealand, like other national statistical organisations, makes concerted efforts to ensure as high coverage as possible in its five-yearly census. However, censuses everywhere tend to miss some people or count people more than once.

Incomplete coverage may result from, for example, inadvertent omission of young children, difficulty in enumerating people on the move and those living in apartments, as well as people not willing to cooperate with census collectors.

New technologies also change how people interact with census collections. The ability to complete census forms online was introduced in the 2006 Census, and many more people took advantage of this option in 2013. Because the electronic forms can be effectively designed to ensure only valid data is entered in specific fields, the online option improves the quality of data collected, but it is not yet clear whether this has also improved census coverage. Other census collection initiatives, such as a mail-out district pilot carried out in parts of Oamaru and the surrounding area during the 2013 Census, are designed to reduce the high costs of field collections and may also impact on coverage.

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).

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 all 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 2013 PES was the fourth to be undertaken in New Zealand since the inaugural PES in 1996. The main objective of the 2013 PES was to measure the level of coverage (undercount and overcount) in the 2013 Census.

Miscount and its 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 general, people are becoming more difficult to contact in any census or survey collection.

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

- people deliberately avoiding the census – refusing or unwilling to respond (eg for fear that information given will be used against their interests)
- people being reluctant to open their door to strangers

- people working longer hours and being harder to find at home
- people shifting from one house to another around the time of the census
- multiple households living at the same address
- people being away temporarily (eg work, school)
- people having no usual residence (eg 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 if they complete more than one form in different locations, or complete their own form and someone else also inadvertently completes a form on their behalf. 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
- 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 either before departure or upon their return
- forms completed for family pets etc, that were not picked up during census processing
- vacant dwellings classified as occupied (non-respondent) dwellings, leading to the incorrect creation of both dwelling and individual substitute records.

An error in recording the geographic location of a person has not been considered a coverage error in the 2013 PES unless the person 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.

People who are more mobile, disadvantaged, or in particular age groups, especially in areas that are difficult to enumerate, are more likely to be missed in the census. Results from previous PESs 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. These results also suggest that this under-enumeration varies across different groups, and by both age and sex.

Why measure the 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 a need to accurately measure 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, for example:

- allocation of funds to organisations using population-based weightings
- denominators for the calculation of rates (eg birth and death rates) and per capita time series
- determination of population weights for various surveys
- administration, policy-making, and planning, by both central and local government
- demographic, social, and economic studies
- as a starting point for demographic projections.

History of post-enumeration surveys

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, Maudlin, & Nisselson, 1953). This PES grew out of earlier surveys used to measure coverage in the 1945 Census of Agriculture, the 1947 Census of Manufactures, and the 1948 Census of Business. Lessons learned from these surveys 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 groups such as African 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 sub-populations 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.

Post-enumeration surveys in New Zealand

Before the 1996 Census, Statistics NZ 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. Statistics NZ conducted a pilot test in 1990 as a preliminary to a PES, in association with the 1991 Census planning, but the survey did not go ahead. In 1994, appropriate funding was approved by the Government to allow Statistics 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 completed before the survey went into the field, to avoid having census collectors 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 began two weeks after the 2001 Census (Statistics NZ, 2002). Similarly, the 2006 PES was conducted two weeks after the 2006 Census. In 2013, the PES was held slightly later, not beginning until 9 April 2013 (five weeks after the census). This recognised that in 2006 the number of outstanding census forms was higher than expected, and may have affected the results.

International practice

Mechanisms to measure census coverage are carried out in most countries that have a traditional census of the type New Zealand conducts (United Nations, 2010). At least 168 countries have held censuses in the current round, of which 108 countries also had a post-censal evaluation of undercoverage, 51 countries did not, with the remaining 9 countries not providing information about how they evaluate census coverage (United Nations, 2010; National Institute of Rwanda, 2014; National Institute of Statistics of Angola, 2014).

New Zealand, along with Australia, South Africa, Zambia, Rwanda (PES in each case), England/Wales, Scotland (Census Coverage Survey in each case), and the United States of America (PES / Census Enumeration Follow Up) carried out a separate survey that involves 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).

Countries also use other approaches. Canada, for example, carries out a census-to-census record linking back to the previous census for a sample of the population. This Reverse Record Check (RRC) is used in conjunction with an Overcoverage 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. The error of closure measures the difference between the number of people counted by the census and the number expected to have been counted based on births, deaths, and international migration, using demographic techniques, to supplement information on coverage and sources of undercount.

2 About the 2013 Post-enumeration Survey

Objectives

The main purpose of the post-enumeration survey (PES) is to measure the level of coverage (undercount and overcount) in the census. Coverage measures are used as key performance indicators for the 2013 Census, with the 2013 PES constituting a major part of the evaluation component for the 2013 Census. The increased sample size in the 2013 PES, along with other quality improvement initiatives, will enable more robust evaluation of the 2013 Census.

In New Zealand, census counts are not adjusted directly to allow for errors in counting. Internationally the most common practice is to not adjust census counts, but to use coverage measures in deriving population estimates. Similarly, Statistics NZ uses coverage measures to adjust the base population, 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, and members of smaller ethnic groups are all more likely than other groups to be missed (and therefore undercounted) in the census. 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 population estimates and projections.

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 population statistics to more accurately represent all population groups, resulting in 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 calibrating sample survey estimates.

Scope – who made up the survey population?

The 2013 PES population **of people** consisted of those New Zealand residents who were either usually resident in a permanent private dwelling, or were staying at one during the survey period. Overseas visitors who were staying in a private dwelling at the time of the 2013 PES, and had been somewhere in New Zealand on census night, were also included. Overseas visitors were surveyed to confirm the consistency of identification of person type (ie resident or overseas visitor) between the census and the PES. For practical reasons, non-private and other private dwellings (see below for a definition) were excluded from the survey, as were dwellings in remote areas.

The 2013 PES survey population had the following exclusions:

- people living in non-private dwellings (eg prisons, hospitals, hotels)
- people living in other temporary or permanent private dwellings (eg tents, caravans, yachts, unoccupied dwellings) in which people do not usually live
- people who died after census night
- people overseas for all of the PES survey period
- babies born after census night

- overseas diplomats, their families, and other people living with them
- people on off-shore islands (except Waiheke Island, which is included).

Note: even though overseas visitors temporarily staying in a private dwelling are surveyed, they are excluded from PES estimation of census coverage.

The 2013 PES survey population **of dwellings** consisted of all permanent private dwellings in New Zealand, excluding off-shore islands other than Waiheke Island.

Survey objectives

The purpose of the 2013 PES was to measure the level of coverage by the 2013 Census of Population and Dwellings of permanent private dwellings and of New Zealand residents.

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

- total population
- sex (male, female)
- age (0–14 years, 15–29 years, 30–44 years, 45–64 years, 65+ years)
- ethnicity (European, Māori, Pacific, Asian)
- geographic area in two groupings:
  - Northern North Island (Northland, Auckland, Waikato, and Bay of Plenty regions), southern North Island (Hawke's Bay, Taranaki, Gisborne, Manawatu-Wanganui, and Wellington regions), South Island, Total New Zealand
  - Auckland region, Wellington region, rest of North Island, Canterbury region, rest of South Island, Total New Zealand.

The PES in New Zealand is intended to:

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

The PES was not specifically designed to:

- locate people who intend to remain invisible and take sufficient steps not to be counted in census
- 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 the definition of 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 by either the census or the PES
- address the effects of differences in enumeration practices between the census and the PES
- assess the quality of responses to census questions, or measure the effect of ethnic non-response in census substitute records.

Sample size and accuracy targets

The sample size for the 2013 PES was 14,915 private dwellings – 37 percent more than the 2006 sample size of 10,907 dwellings.

The accuracy targets were sampling errors of ±0.3 percentage points for the national-level estimate and ±1.0 percentage points for subgroup estimates (at the 95 percent confidence level).

Sampling frame and selection units

The 2013 PES sample uses a two-stage sample selection method, similar to many Statistics 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 Statistics NZ uses to select samples and manage overlap control for all of its household surveys, and comprises a list of PSUs with attributes determined by census data. PSUs are assigned to standard strata based on these attributes. PSUs are geographic areas consisting of between 50 and 100 dwellings and are formed using meshblocks – Statistics NZ's standard geographic classification.

Each PSU is divided into approximately seven panels, with each panel containing about 10 dwellings. The second stage of the sample selection consists of selecting eligible dwellings within the selected PSUs, by selecting one or more of these panels. Approximately one-quarter of the PSUs in the PES sample have part of a second panel selected and a small number have a full second panel selected.

All usual residents and certain visitors within a selected dwelling form part of the PES sample. The selection unit is the dwelling and the collection unit is the individual.

Sample design

The 2013 PES was based on the sample design of Statistics NZ's New Zealand General Social Survey (NZGSS). This was a change from the 2006 PES, which used the sample design of the Household Labour Force Survey (HLFS). The main reasons for choosing the NZGSS design included:

- reduced costs for enumeration and field collection – interviewers were already working in and familiar with the geographic areas used in the sample, and fieldwork made use of existing maps and street listings
- minimisation of respondent burden by controlling the overlap between the PES and other household surveys.

Moreover, the current HLFS sample design was based on 2001 Census information and was therefore less efficient than the NZGSS sample design which is based on 2006 Census information. Similar to the HLFS, the NZGSS oversamples areas that contain a high proportion of people of Māori or Pacific ethnicities in order to improve the accuracy of these groups. While the NZGSS also oversamples smaller regions to improve the accuracy of these regions, it does not do this as heavily as the HLFS sample

The sampling process was complex. The geographical framework of New Zealand consists of 46,637 meshblocks. A meshblock in urban areas is usually a block of residential dwellings surrounded by streets. In rural areas, a meshblock covers a much wider area because dwellings are sparsely spread. For sampling purposes, meshblocks are aggregated into 21,813 primary sampling units (PSUs). To improve the sampling efficiency these PSUs are stratified into 127 groups (called strata) based on region, urban/rural mix, ethnic population, and other socioeconomic variables, such as income, employment status, and population aged 65 years and over.

The PES sample consisted of 1,299 PSUs, made up of the 1,199 NZGSS PSUs plus 100 newly selected PSUs. All of the PSUs had at least one full panel selected, panel 7. There were 365 PSUs that had 1.35 panels selected (ie 100 percent of panel 7 plus 35 percent of panel 2), and 37 PSUs that had two full panels selected (panels 7 and 2). Both the 100

new PSUs and the 365 PSUs with extra part-panels were selected from the top 20 ‘difficult to count’ strata of the 2012 HSF. The 37 PSUs with two panels were targeted due to their high selection weight. The effects of the re-zoning of land and population relocation as a result of the Canterbury earthquakes were taken into account.

Preparations for collection

There were two parts of the 2013 PES PSU enumeration:

- The initial check-enumeration of the 1,199 NZGSS PSUs. This was carried out over August to December 2012 by the field staff who were in the PSUs for NZGSS interviewing.
- The enumeration of the 100 new PSUs. This process is described below.

Enumeration of the 100 newly selected PSUs

Statistics NZ drew up the address lists and maps for the 100 extra PSUs. For previous household surveys, Quotable Value (QV) had supplied the data, in conjunction with the Statistics NZ field collections team. We created an initial electronic file of individual street addresses using 2006 Census data and the street address range from the HSF. This was checked and formed the basis of the address list. Additional information, such as flat numbers and census collector's notes, was drawn from the 2006 Census field books. These provided useful information about non-private dwellings, schools, and other details of interest in the PSUs.

Collection and response rate

Collection

The survey was carried out over 9 April–10 May 2013, following the planned completion of census fieldwork. We chose the survey period to avoid overlap of census collectors and PES interviewers in the field, and to reduce the effect of large numbers of late census returns, but to be close enough to the census date (5 March 2013) to assist respondent recall. The PES interviewers were in the field 11 days longer than planned, to maximise contact and response rates.

A problem common to all PESs is that people missed by a census are also likely to be missed by a PES. We refer to this 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 is practicable. 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 120 specially trained interviewers using a household questionnaire. Information on occupants of the dwelling who satisfied the scope and coverage criteria was collected through a face-to-face interview wherever possible. Alternatively, a proxy interview was conducted (ie details were obtained from another adult in the dwelling) and a follow-up interview was done by telephone (unless the respondent insisted on a face-to-face interview).

In the 2013 PES, we relaxed the rule of not using interviewers who had been census collectors to meet fieldwork targets. A few carefully selected census collectors were used. However, the vast majority of interviewers had not worked as census collectors for the 2013 Census, and any who had were assigned to areas different from their census work. Other countries have found that the independence of the census and the PES is not compromised by the use of the same field staff for both surveys.

Personal details sought on the PES questionnaire included name, sex, date of birth or age, ethnicity, and address. While the most important addresses were the usual address

of the respondent and the address where they were 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 to help identify where people may have been counted more than once.

Response rate

The 2013 PES had a response rate of 87 percent, which is the proportion of eligible households that responded to the PES interviewer and contained at least one fully responding eligible person. This was an improvement over the 2006 PES, which achieved a response rate of 84 percent, but it was below the target of 90 percent set for the 2013 PES. In total, nearly 32,700 people responded to the survey.

Matching and searching

The objective of matching was to determine if a PES respondent was counted in the census at each address at which they stated that they had completed a census form, or at each address where a census form may have been completed for them.

In the 2013 PES, a new automatic-matching procedure used the QualityStage (QS) data integration software to link PES records to census records. This replaced the former manual search methods, which were both labour intensive and error-prone. The automatic-matching method also reduced dependence on the quality of the address information in both the census and PES. This allowed much wider and more-effective searches. Not only was this much more economic and effective, it also produced higher-quality matches. This increased quality was achieved by comparing the information given by PES respondents with the information given on census forms.

Automatic matching was a new departure in the 2013 PES, so it was necessary to identify precisely how much impact this had on the results. To quantify this, a separate matching impact study (MIS) was designed and carried out. This enabled us to estimate the effect that automatic matching had on the quality of matching, and how the results would differ had we used the former manual matching method.

Automatic matching

Statistics NZ had carried out earlier studies to investigate the feasibility of introducing automatic matching into the PES processing system. Trial studies in preparation for the abandoned 2011 Census showed very promising results, with a link rate in excess of 80 percent. The high quality of the links found in these studies was confirmed by noting that, of those records that were linked together, a negligible proportion of the links were false positive links. These false positive links all came from the probabilistic linking of the PES records within a meshblock. Similar studies undertaken by the UK Office for National Statistics and Australian Bureau of Statistics (ABS) had also shown good results.

Before the 2011 Census was cancelled, the PES processing system had been re-developed to incorporate automatic matching between PES and census records. This system formed the core of the 2013 PES automatic matching system, and included additional variables, such as name, to improve linking accuracy.

The following census variables are used for linking:

- day, month, and year of birth
- sex
- ethnicity
- name
- usual residence address

- meshblock: usual residence meshblock and census-night meshblock are used. It is possible to have two records for a single census individual if their census-night address is different from their usual-residence address.

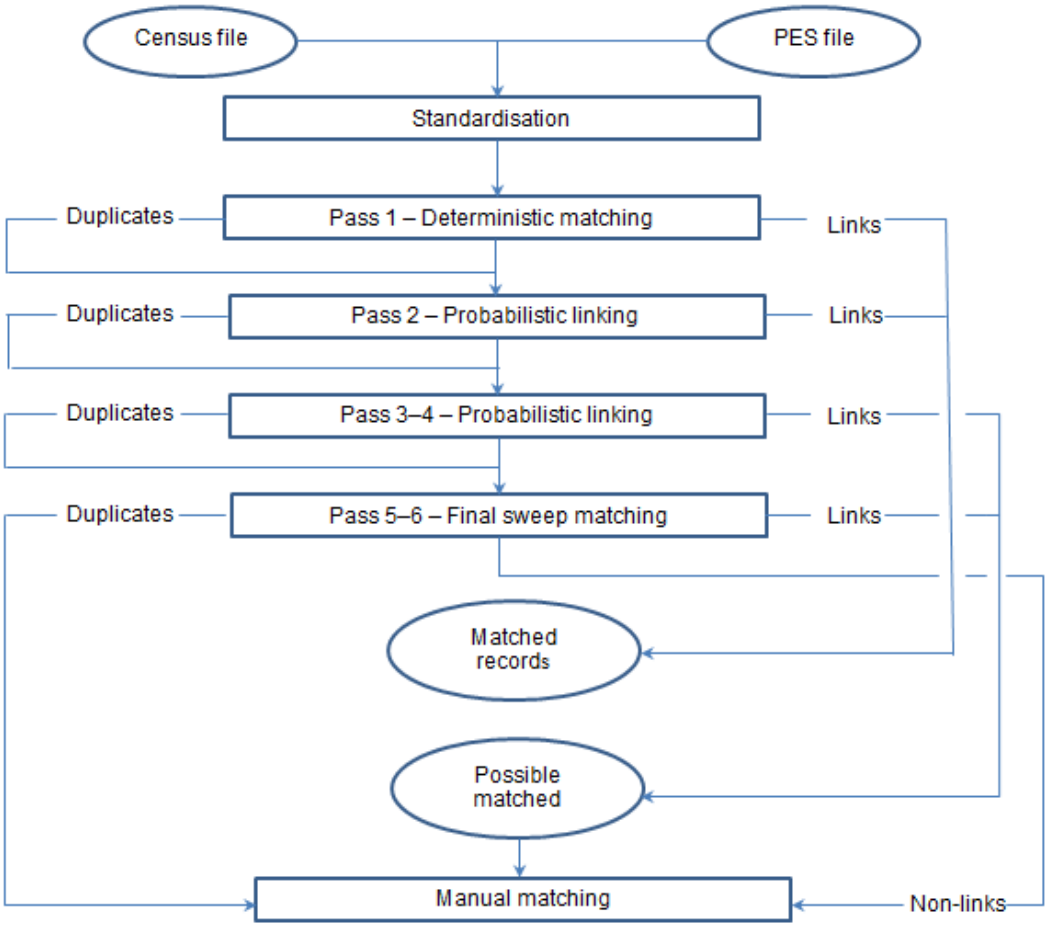
Automatic matching of 2013 PES data to 2013 Census data

The 2013 PES processing system automatically matched PES respondents to their census records using QualityStage (QS) software. Nearly 83 percent of records were matched automatically. The remaining 17 percent were matched by manual processes similar to the 2006 methods. This contrasts with previous PESs, when matching was completely manual. While we expected that automatic matching to improve matching quality, we did not know how much of an improvement was achieved. The matching impact study (MIS) was designed to measure this.

Matching was achieved by passing the data through QS up to six times. Figure 1 shows schematically the flow of data through the matching process. Two types of linking take place: deterministic and probabilistic. Deterministic linking identifies all exact links. These are cases where the PES record and the census record contains exactly the same information. The subsequent passes use probabilistic linking to make other links when the data is not exactly and uniquely identical in both records.

Figure 1

Automatic matching flow chart for the 2013 PES



The census data and the PES data is first standardised so that the matching variables are coded compatibly.

In the first pass, PES records are deterministically linked to census records using all of the following key variables:

- sex
- day, month, and year of birth
- address information.

We treated two records as a linked pair if all the following conditions were met:

- they matched exactly on all of the linking variables
- the link was unique (that is, no other record was an equally exact link)
- the responses for each of the linking variables were complete and valid.

In the second pass, we also used the address information but allowed for minor differences that may have resulted from recording or respondent errors.

We linked 68.0 percent of the records using address information in the first two passes, as shown in table 1. The first pass made exact matches deterministically, while the second pass added matches made probabilistically where the address or other information had, for example, slight differences in detail.

The remaining records were linked probabilistically and used meshblock information rather than address information. Probabilistic linking, which was formalised by Fellegi and Sunter (1969), uses the probabilities of agreement and disagreement between a range of matching 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 pairs with probabilities between the two cut-off values are set aside for subsequent passes. The probabilistic passes were designed to allow appropriate operator intervention to resolve specific difficulties. Any records unmatched after the final pass are matched where possible by manual methods. In total 82.7 percent of person addresses were matched automatically and 7.2 percent were matched manually. The remaining 10.1 percent were not matched.

Table 1

Counts of links of person addresses at each step, 2013 PES

Step	Link type	Person addresses[(1)]	Percent matched	Cumulative percentage
Pass 1	Deterministic	16,810	48.3	48.3
Pass 2	Probabilistic	6,870	19.7	68.0
	-Total matched using person address	23,690	68.0	
Pass 3	Probabilistic	4,300	12.4	80.4
Pass 4	Probabilistic	190	0.6	81.0
Pass 5	Probabilistic	530	1.5	82.5
Pass 6	Probabilistic	60	0.2	82.7
	-Total matched using meshblock	7,050	20.3	
	-Total matched by QS[(2)]	28,770	82.7	
Manual	Manual	2,510	7.2	89.9
Total matched		31,280	89.9	
Non-matched person addresses[1]		3,530		
Total person addresses		34,810	100.0	

1. Matching was undertaken at PES person-address level. PES respondents could provide more than one address where they may have completed a census form or had one completed for them, therefore the number of person addresses is greater than the number of respondents. The data is rounded to the nearest 10

2. QualityStage

Source: Statistics New Zealand

Estimation

The estimation of coverage occurs differently for the groups of records defined as the contact sector and the non-contact sector. In this section we:

- define what these sectors are, before describing the way coverage is estimated for each of these groups
- describe how coverage is estimated in each of the groups
- give details of how we calculate the PES weights, which are used to estimate coverage.

Contact and non-contact sectors

A distinction is made between contact and non-contact sectors. This terminology follows the approach the ABS used in the 2011 Australian PES. The contact / non-contact sector is defined at the date at which fieldwork for the 2013 PES began (9 April 2013).

The non-contact sector includes only people living in private dwellings on census day, but from whom no forms were received by the time the PES interviewers started work in the field. These census records comprise two categories: late returns and substitute records. The 2006 PES did not estimate coverage for this sector, whereas the 2013 PES estimates the overcount of substitute records in the non-contact sector.

The contact sector includes all people in private dwellings that are not in the census non-contact sector. The contact sector includes people who were missed at counted dwellings and people counted in missed dwellings.

Late returns are forms received by census after the PES began field interviewing on 9 April 2013, except for those that had been collected by census collectors before that date, but delivered to the census processing team later (all census collectors had finished field collection before PES started). The reason for excluding late returns is that these 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 late returns than the rest of the population, and we cannot distinguish between census forms returned late because of PES prompting and those returned late for other reasons.

There are three kinds of substitute forms raised in the census (Statistics NZ, 2014):

- A substitute individual record is raised within a real dwelling (ie we have a dwelling form), when there is a name on the dwelling form not accounted for by the individual forms received, or the number of occupants stated on the dwelling form was greater than the number of forms received.
- A substitute dwelling record is raised with real individuals attached when no dwelling form was collected but we have individual forms. In this case, the number of occupants in the dwelling is made equal to the number of individual forms received.
- A substitute private dwelling with substitute people. In this case a wholly substitute household is created and the number of occupants is imputed randomly using a donor household, from a 'nearest neighbour' imputation method.

For PES purposes, the first type of substitute can be treated as a real person since we know that the substitute record relates to that of a specific named person. The second type of substitute record is not a problem for PES individual matching as we have the real people in the dwelling to match to and we assume that no individuals are missing.

Substitutes of relevance to PES are the third type: substitute people in an all-substitute dwelling. This is the only type of substitute record that is included in the non-contact sector; the other two types are included in the contact sector.

Estimation method used for the 2013 PES

In preparing for the 2013 PES, we investigated a number of alternative estimation methods. The outcome of this investigation was that an approach based on a modified 2006 method could provide the most robust and transparent measures in the time available. The modifications adopted for the 2013 PES included an adjustment for dwelling non-response, together with an assessment of the number of substitute records that should have been raised in all-substitute households.

The 2013 method addressed the impact of an increased number of substitute records, and partly addressed the problem of internally inconsistent results between coverage estimates by ethnicity and other coverage estimates because of missing information in the non-contact sector.

Estimates of dwelling coverage

The 2013 PES set out to estimate the coverage in census of permanent private dwellings. Non-private dwellings and temporary private dwellings (caravans, tents, etc) are not included as they are not in the Household Survey Frame, which was used for the 2013 PES. Dwellings may have been demolished or been converted to a non-private dwelling between the census and the PES, but the number would be very small and covered by the post-stratification.

PES estimates of the true census dwelling count are calculated for a number of regions as well as a national estimate.

The PES estimate of the true dwelling count in region r is:

$$D' = D \times \frac{dx}{dy}$$

Where:

D' = PES estimate of in-scope private dwelling count in region r

D = census count of private dwellings in region r

dx = PES estimate of private dwellings that should have been counted in the census

$$dx = \sum_{hij} dwgwgt_{hij} * SHOULD$$

dy = PES estimate of private dwellings which were counted in the census

$$dy = \sum_{hij} dwgwgt_{hij} * WAS$$

WAS=1 if an eligible PES dwelling was counted in census (=0 otherwise)

SHOULD=1 if an eligible PES dwelling should have been counted in census (=0 otherwise)

and $dwgwgt_{hij}$ is the dwelling weight for a dwelling j in, PSU i, and stratum h within the region r

The PES estimate of net dwelling undercount is: D'-D

Dwelling overcount was not estimated in PES. There is a small risk that census collectors in adjacent areas listed the same dwelling. Examples of where this can happen include corner sections and properties with back entrances on a different street.

Dwelling non-response adjustment

A non-responding dwelling in the PES is a dwelling that is occupied, but has no responding people in the PES. A responding dwelling is one in which at least one person responds.

Weighting is necessary to calculate what proportion of the non-responding dwellings should be represented by the responding dwellings (and the people in them), and what proportion we can assume to be vacant. This proportion varies depending on the dwelling's census attribute. For example, almost 90 percent of dwellings found to be occupied in the census responded in the 2013 PES, whereas just over 30 percent of dwellings found to be unoccupied in the census responded in the 2013 PES.

The response factor is calculated within each region and census dwelling attribute. The weights of responding dwellings compensate for non-responding dwellings (refusals, partial non-contacts, etc), as well as full non-contact dwellings.

This non-response adjustment is important for deriving person weights but we do not use it for the dwelling estimates. After this stage, the dwelling weights of people in eligible

responding dwellings in the PES should represent the pool of people in all private dwellings at census time. Every person has the weight which belongs to the dwelling in which they were enumerated in the PES.

Estimating coverage for people

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

Three measures relate to this difference: gross overcount, gross undercount, and net undercount. We express these both as numbers and as percentages 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 by the census, but who 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 (ie a net overcount) is expressed as a negative undercount.

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

Gross overcount for people in the contact sector

For the overcount, we need to consider both the people counted more than once (= scope 1 in the formulae below) and people counted when they should not have been (= scope 3). People with scope = 1 may have been counted in the census more than once because they completed census forms at more than one address, or someone else also completed them on their behalf.

Conversely, people with scope = 3 may have been counted in the census when they should not have been; for example, census forms were completed for them by someone else while they were overseas, or they did so themselves before they left or after they returned.

In terms of PES variables:

- Scope 1 people – people counted more than once in the census – is anyone with the variable WAS > 1. A person should only be matched once in the census. That person's excess contribution to the total is an overcount.
- Scope 3 people – people counted by the census – is anyone who has WAS > 0 (ie were recorded in the census), but say they were not in New Zealand on census night. This contributes to an overcount.

So, working from the definition above, the formula for gross overcount is

$$Gross\ Overcount = \frac{Y}{y}O$$

Where:

Y = census count of usual resident population in the contact sector

y = PES estimate of population who were counted in census (includes both people usually resident in New Zealand and those usually resident overseas)

and

$$O = \sum_{scope=3} finwgt_i \times WAS_i + \sum_{\substack{scope=1 \\ WAS>1}} finwgt_i \times (WAS_i - 1)$$

Gross undercount for people in the contact sector

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. People with an imputed value for WAS between 0 and 1 contribute to the undercount. For example, if WAS was imputed at 0.9, then 0.1 of that person contributes to the undercount.

The formula for gross undercount is:

$$Gross\ Undercount = \frac{Y}{y}U$$

Where:

Y = census count of usual resident population in the contact sector

y = PES estimate of population who were counted in the census (includes both people usually resident in New Zealand and those usually resident overseas)

and

$$U = \sum_{\substack{scope=1 \\ WAS<1}} finwgt_i * (1 - WAS_i)$$

Estimated population for the contact sector

Demographic and regional estimates of the true census population of the contact sector are calculated as follows:

$$Y' = Y \times CAR$$

Y' = PES estimate of true census population in the contact sector

Y = census count of usual resident population in the contact sector

CAR is the coverage adjustment ratio and represents the ratio of two quantities – the weighted count of PES individuals who should have been counted in the census and the weighted count of PES individuals who were counted in the census.

$$CAR = \frac{x}{y}$$

Where:

x = PES estimate of the population who should have been counted in the census

$$x = \sum_{i \in l} SHOULD_i \times finwgt_i$$

For person i, SHOULD = 1 if usually resident in New Zealand on census night, otherwise SHOULD = 0

'finwgt' is the final person weight

and *I* = the set of individuals used in estimation. This set comprises eligible responding people usually resident in New Zealand or usually resident overseas who have NOT been matched ONLY to a substitute AND belong to the demographic or regional group of interest.

y = PES estimate of the population that was counted in the census (includes both people usually resident in New Zealand and those usually resident overseas)

$$y = \sum_{i \in I} WAS_i \times finwgt_i$$

For person i, WAS is the number of times he was counted in the census.

Estimated population for the non-contact sector

We recognised that some assessment of the validity of the substitutes raised by the census was required in the 2013 PES. Previous PESs did not evaluate substitute records at all, making the assumption that all substitute records had been correctly raised. We recognised this as a missed opportunity with major implications for the interpretation of census coverage, because any substitute record that represents the same individual as one counted elsewhere creates an overcount in the census.

So long as the number of substitute records remained relatively small in the census, this had been a relatively small problem. However, we raised a total of over 203,000 individual substitute records, or more than 4 percent of the usually resident population count, in the 2013 Census. Among these, almost 135,000 were in private households where all members and the dwelling form were substitutes (referred to, interchangeably, as an all-substitute dwelling or an all-substitute household).

Substitute individual records in all-substitute dwellings are of interest to the PES because the census imputed the number of usual occupants for these dwellings, and the number of substitute records for the dwelling may not agree with the actual population in those dwellings. In dwellings with at least one returned census form, we still assume that substitutes have been raised correctly, based on information on the dwelling form about the usual occupants.

The estimation methodology used for the 2013 PES enabled a partial investigation of the substitute records raised by the 2013 Census. We restricted this to an estimation based on the number of PES records that were matched in the census to both a substitute record in an all-substitute dwelling and to a received census form. Because we were able to match to a returned census form, we were able to identify the existence of an incorrectly raised substitute record. The 2013 PES found that the 2013 Census raised approximately 15,000 too many substitute records, suggesting that the number of individual substitute records was around 11 percent too high.

We estimate the proportion p_j of PES individuals matched to a census substitute and found elsewhere in census in region j.

$$p_j = \frac{\sum_{i \in S \cap R} WAS_i \times finwgt_i}{\sum_{i \in S} WAS_i \times finwgt_i}$$

Where:

R is the domain representing people who were matched to a substitute and not found elsewhere in the census and S is the domain of all people matched to a substitute.

j = 1, 2 geographic regions – Auckland and Northland comprising one region and the rest of New Zealand comprising the other.

The estimated number of census substitutes is then $\sum_j p_j \, Census_{S,j}$ where $Census_{S,j}$ is the census count of substitutes in region j.

The estimated population for the non-contact sector is the sum of the Late returns (census forms returned late) ($Census_L$) and the estimated census substitutes.

The PES estimated population of the non-contact sector $= Census_L + \sum_j p_j \, Census_{S,j}$

Estimated population for New Zealand

The PES estimated population is the sum of the estimated populations for the contact sector and non-contact sector.

PES estimated population = PES estimated population of contact sector + Late returns + estimated census substitutes

$$= CAR * Census_{Contact} + Census_L + \sum_j p_j \, Census_{S,j}$$

Net undercount = PES estimated population - census count of New Zealand resident population

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

Weighting

The 2013 PES estimation methodology relies on assigning a 'weight' to each dwelling selected into the PES, and then to each person from whom we obtained a PES response. The weight reflects the probability that a dwelling was selected using the PES sample design, as well as adjusting for some groups of the population being less likely to participate in the survey. These adjustments are necessary to ensure that the weighted sample is representative of the population.

The first stage of weight adjustment ensures that the adjusted weights of private dwellings selected in the PES add up to the census private-dwelling count. The purpose of this adjustment is to reduce the sampling error by making use of dwelling counts known from the census, as well as to adjust for a small proportion of dwellings being missed when a list of dwellings is created within the PSUs selected for the PES. Separate adjustments are done, based on whether or not the dwelling was occupied on census night. The average weight adjustment is also applied to dwellings that were missed in the census.

The second stage of weight adjustment accounts for selected dwellings that do not respond to the PES, and ensures that the responding PES dwellings are representative of dwellings for which we did not obtain a response. Separate adjustments are made based on whether the dwelling responded to the census, because we know that dwellings that did not respond to the census are also less likely to respond to the PES. This weighting step was a new addition to the 2013 PES.

These first two adjustments affect both dwelling weights and person weights. The third and final adjustment affects only the person weights that are used to estimate the coverage of different types of people. This adjustment ensures that the adjusted weights of people responding to the PES, who are in the contact sector and found in the census, add up to the census count of people in the contact sector. The contact sector excludes late returns and substitute records from all-substitute households.

The purpose of this adjustment is to reduce the sampling error by making use of information known from the census, and to reduce non-response bias due to some

population groups being less likely to respond to the PES. The adjustment is done by age group, region, and ethnicity.

For each of the three stages of weight adjustment, the average adjustment factor is also applied to the non-contact sector.

All the weight adjustments made required assumptions that cannot be validated by the data collected in the PES. This introduces a source of non-sampling error. The first assumption is that, within the weighting classes used, the people and dwellings responding to the PES adequately represent those that did not respond to the PES. In the first stage of weight adjustment, we are also assuming that the chance of a dwelling being missed in the census is not related to the chance of it being missed in the PES. In the third stage of weight adjustment, there is also an assumption that the groups (such as younger adults), which are under-represented in the PES contact sector, will also be under-represented in the non-contact sector.

Sampling and non-sampling errors

Estimates of undercount, overcount, and net undercount produced from the PES are all subject to both sampling and non-sampling errors. Since only a sample of dwellings was included in the PES, estimates derived from the survey may differ from figures that would be obtained if all dwellings had been included. The sampling error is a measure of how far the estimates derived from the PES might vary by chance from the true figures.

Because of the relatively small sample size, it is only possible to provide reliable estimates of undercount for broad groups of the population. The size of the sampling error is one measure of the level of reliability. Some estimates have high sampling errors and should be used with caution. In general, the sampling errors associated with subnational estimates (such as by area or age group) are larger than for the national estimate.

Non-sampling errors are different from sampling errors. Non-sampling errors in surveys have a variety of sources including non-response, imperfections in reporting by respondents, data collection, and data processing. While non-sampling errors can be reduced by careful form design, training, and supervision of interviewers and efficient operating procedures, we cannot entirely eliminate them. The matching of PES forms to census forms is another source of non-sampling error in this survey.

When we compare undercount estimates, both sampling and non-sampling errors must be kept in mind. The presence of non-sampling errors is one reason why it is reasonable to assume that the PES estimates of undercount are likely to be lower than the true value.

Matching impact study

Statistics NZ undertook a matching impact study (MIS) on a subsample of 2013 PES records. This study evaluated the new method of automated electronic matching, supplemented with manual matching, against the manual-only method used previously, in order to measure the effect of the change in method. The automatic matching was at the individual level only.

The 2013 MIS followed similar procedures to those used for the Australian PES's Statistical Impact Study (ABS, 2012) to estimate the impact of automatic matching. The MIS took a subsample of the 2013 PES records and used the 2006 manual methodology for matching to calculate an alternative estimate of coverage.

The MIS was done after the PES processing was complete, and was run in a separate database from the main PES so as to maintain independence of the PES processes. A sample of 2,000 eligible occupied dwellings was drawn from the 2013 PES sample and stratified by regional council. These dwellings had to contain at least one eligible

responding person. The dwellings were selected by proportional allocation within each stratum. The households in this sample were then processed in a manner as close as possible to the 2006 PES processing system.

Essentially, this means there was no automatic matching involved when trying to 'find' a PES respondent in the census. Manual matching depends heavily on the list of addresses where a census form may have been completed by, or for, the respondent. This method, when used alone, is only likely to identify correctly the number of times a PES respondent has been found in the census in those cases where the address list is comprehensive.

The MIS was then able to estimate the improvement in PES measures. The MIS provided a confidence interval for the additional number of individuals estimated to have been found in the census because of automatic matching.

Results of the study

The MIS verified that the 2013 PES results improved the quality of individual record matching. The new automatic-matching method reduced the net census undercount because we were able to locate more people who gave vague or inaccurate census-night addresses in the 2013 PES. The primary outcome of the study is it showed the method used in the 2013 PES produced more-robust results. The manual matching method used in the MIS estimated the national net undercount rate to be 3.9 percent (±0.6 percentage points), compared with 2.4 percent (±0.5 percentage points) using the new automatic-matching method (table 2).

Table 2

Comparison of 2013 MIS and 2013 PES estimates of net undercount

	Net undercount (percent)		Sampling error (percentage points)	
	MIS	PES	MIS	PES
Male	4.1	2.6	±0.7	±0.5
Female	3.6	2.1	±0.7	±0.5
Total	3.9	2.4	±0.6	±0.5

Source: Statistics New Zealand

Because the MIS sample is small and the sampling errors comparatively high, results for other subgroups, such as geographic area, age, and ethnicity are not available.

What the MIS implies for comparability over time

In this report we have compared 2013 PES data with 2006 PES data. We caution that data from these two PESs is not completely comparable. This is because the 2006 PES data used manual matching, while the 2013 PES data reflects the benefits of automatic matching. Also, partial validation of substitute record creation was done in 2013 but not in 2006.

Table 3 summarises one way of showing the relationship between the 2006 and 2013 PES results. The MIS results show, within the limits of the sample size, the results that would have been expected had we not used automatic matching.

Table 3

Comparison of 2006 PES, 2013 MIS, and 2013 PES net undercount rates

	Matching type	Male	Female	Total
		Net undercount (percent)		
2013 PES	Automatic	2.6	2.1	2.4
2013 MIS	Manual	4.1	3.6	3.9
2006 PES	Manual	2.1	1.8	2.0
		Sampling error (percentage points)		
2013 PES	Automatic	±0.5	±0.5	±0.5
2013 MIS	Manual	±0.7	±0.7	±0.6
2006 PES	Manual	±0.5	±0.6	±0.4

Source: Statistics New Zealand

3 2013 Post-enumeration Survey results

This chapter presents and analyses the undercount estimates derived from the 2013 PES. Included in this chapter is analysis of results by age, sex, ethnicity, and geographic area, with details of coverage measures.

Analysis of 2013 PES results

The 2013 PES estimated total population, as defined in this report, is the estimated number of New Zealand residents in New Zealand on census night expected to have been counted. The total population is therefore equivalent to the census usually resident population count plus the estimated net undercount of New Zealand residents.

Overall, the 2013 Census coverage was high, reflecting the cooperation and support of the New Zealand public. The 2013 PES estimated that about 97.6 percent of New Zealand residents in the country on census night were counted in the 2013 Census. The net undercount was about 2.4 percent, or 103,800 people. The net undercount figure resulted from an estimated gross undercount of 135,500 people (3.1 percent), offset by 31,700 people (0.7 percent) being overcounted.

The estimated net undercount for the 2006 PES was 2.0 percent, or 81,000 people. For the 2001 PES it was 2.2 percent, or 85,000 people. The first PES, in 1996, estimated the net undercount to be about 1.6 percent, or 60,000 people. A re-run of the 2006 PES data using the same method as for the 2013 PES was done for assurance purposes, but because we were not able to replicate the automatic matching and because the differences between the re-run data and the published 2006 PES results were small, we have not revised the 2006 PES results. This means that we will not revise historical published data.

While the 2013 Census counted 4.24 million residents in New Zealand, the 2013 PES results suggest that the number of New Zealand residents in New Zealand on census night was closer to 4.35 million. This figure excludes New Zealand residents temporarily away overseas on census night. Given the small size of the 2013 PES sample, some estimates of net undercount yielded by the survey are subject to sizeable sampling error margins. Therefore the undercount estimates need to be interpreted with caution.

The sampling error associated with the net undercount of 2.4 percent in the 2013 PES was ±0.5 percentage points. This means there is a 95 percent probability that the actual net undercount was between 1.9 and 2.9 percent, not allowing for non-sampling error. Similarly, there was an overall sampling error of ±0.4 percentage points in the 2006 PES, ±0.3 percentage points in the 2001 PES, and ±0.2 percentage points in the 1996 PES.

Table 4 shows the 2013 net undercount estimates and the sampling errors for these estimates. Because of the small sample size and the resulting large sampling errors, it has been necessary to aggregate age groups and geographic areas. A complete version of this table containing the time series from 1996 to 2013 is included in the accompanying set of downloadable tables.

Table 4

Net undercount estimates and sampling errors for the 2013 PES, by selected subgroup

Subgroup	People (000)		Percent	
	Census usually resident population count	Net undercount	Net undercount	Sampling error coverage[1]
	2013 PES			
Sex				
Males	2,064	56	2.6	0.5
Females	2,178	48	2.1	0.5
Total	4,242	104	2.4	0.5
Age group (years)				
0–14	866	27	3.0	0.9
15–29	845	42	4.8	1.1
30–44	830	17	2.0	0.7
45–64	1,095	15	1.3	0.5
65+	607	3	0.5	0.6
Total	4,242	104	2.4	0.5
Geographic area[2]				
Northern North Island	2,239	69	3.0	0.7
Southern North Island	998	13	1.3	0.4
South Island	1,004	22	2.2	0.9
Total New Zealand	4,242	104	2.4	0.5
Ethnicity[3]				
Māori	599	39	6.1	1.3
Pacific	296	15	4.8	1.5
Asian	472	15	3.0	1.2
European	2,969	59	1.9	0.5

1. The sampling error indicates the extent to which an estimate from the PES sample might deviate from the true value. For example, there is a 95 percent chance that the true coverage for New Zealand in 2013 is between 97.2 and 98.1 percent.

2. North and South islands exclude people living on outlying islands (other than Waiheke Island). Northern North Island includes Northland, Auckland, Waikato, and Bay of Plenty regions. Southern North Island includes Gisborne, Hawke's Bay, Taranaki, Manawatu-Wanganui, and Wellington regions.

3. Ethnicity is based on total responses and is not comparable with other data in the table.

Source: Statistics New Zealand

Non-response in the census

Net undercount obtained from the post-enumeration survey is one measure of census non-response. The census has processes that allow known unit non-response to be included. Census counts include substitute records for occupied dwellings known to be missing and for the people known to be living in those dwellings, where the census is satisfied these exist. These are the all-substitute dwellings used in the PES estimation. In addition, substitute records are also created for people known to be missing within enumerated households, as well as substitute dwelling records where people have returned individual forms but not an associated dwelling form (Statistics NZ, 2014). We are more certain about the characteristics and number of substitutes created in partly-counted households than we are about substitutes created in all-substitute households.

The combination of the net undercount (estimated by the PES) and the total number of substitutes (counted within the census) gives a more complete picture of the effective non-response to the census. In the 2013 PES, the total census non-response rate was 7.1 percent, up from 5.2 percent in 2006, 5.0 percent in 2001, and 4.4 percent in 1996. The results for censuses from 2001 to 2013 are shown in table 5, with the 1996 results included with the downloadable tables. The 2013 net undercount was higher than in 2006, and the higher number of substitute records created in 2013 resulted in a substantially higher rate of total non-response overall in the 2013 Census.

Table 5

Non-response in the census, 2001–13 PESs

Source	2001 PES		2006 PES		2013 PES	
	Number (000)	Percent of estimated total	Number (000)	Percent of estimated total	Number (000)	Percent of estimated total
Forms received by the census	3,631	95.0	3,895	94.8	4,039	92.9
Substitute individual records						
–in partly counted households	27	0.7	38	0.9	68	1.6
–in all-substitute households	80	2.1	95	2.3	135	3.1
–total substitutes	107	2.8	133	3.2	203	4.7
Census usually resident population count	3,737	97.8	4,028	98.0	4,242	97.6
Estimated net undercount	85	2.2	81	2.0	104	2.4
PES estimated population	3,822	100.0	4,109	100.0	4,346	100.0

Source: Statistics New Zealand

Difficulty of census enumeration

The net undercount is 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. 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
- substitute records.

The sum of the undercount (135,500 records) and overcount (31,700 records) for the 2013 Census totals 167,200 records. There were 21,300 forms received late (ie after the PES collection phase began). The number of substitute records increased from 132,950 in 2006, to 203,150 in 2013.

The number of late returns was substantially reduced from 2006, when there were around 145,000 forms outstanding at the beginning of the PES collection phase. The main reason for the reduction in 2013 was that PES field interviewing did not start until 9 April 2013 (five weeks after the census). In 2006 the interviewing phase started two weeks after the census on 21 March 2006. This later start in 2013 gave the census additional time to complete collection in the field.

These factors, together with the undercount and the overcount, indicate that some form of difficulty with enumeration was encountered for 9.2 percent of the census usually resident population count in 2013. This indicates the value of the delayed start date for PES field interviewing. The difficulty with enumeration percentage in 2006 was 9.3 percent, compared with 7.2 percent in 2001 and 6.1 percent in 1996.

Table 6

Difficulty of census enumeration, 1996–2013 PESs

	1996	2001	2006	2013
	Number (000)			
Estimated gross undercount	68	107	92	135
Estimated gross overcount	8	22	11	32
Late returns[1]	45	40	145	21
Substitute records	102	107	133	203
Estimated total difficulty	223	276	381	390
Census usually resident population count	3,618	3,737	4,028	4,242
Difficulty of enumeration[2]	6.1	7.2	9.3	9.2

1. Late returns for 2006 are estimated.
2. Estimated total difficulty as percentage of census usually resident population count.

Source: Statistics New Zealand

International comparison

Given the strategic significance of census data, many countries conduct surveys to measure the coverage of their population census. The Australian Bureau of Statistics, for example, has run evaluation programmes since 1966; the United States Census Bureau since 1950; Statistics Canada since 1966; and the national statistical offices for the United Kingdom and for Scotland since 1981. At least 108 countries with traditional censuses have either surveys or record matching processes to measure undercoverage.

The net undercount rate of 2.4 percent for New Zealand's 2013 Census is similar to the rate for Canada in 2011 (2.3 percent), although slightly higher than Australia's rate in 2011 (1.7 percent). Reported undercount rates vary dramatically across the world, with some countries reporting net undercount rates below 1 percent (such as the United States at 0.0 percent in 2010 and Rwanda at 0.8 percent in 2012). Others have much higher net undercount rates (such as South Africa at 14.6 percent in 2011).

Variations in undercount rates in New Zealand by sex, age, and subgroups within the population mirror differences found internationally. Many countries also measure undercount by other parameters such as language spoken or marital status, but the New Zealand PES is not designed to assess data consistency other than for counts by key variables required for population estimates and projections, such as age, sex, ethnic groupings, and geographic area.

International comparisons should be interpreted with caution because of the differences in sample design, in enumeration procedures and practices, and in the nature and size of sampling and non-sampling errors.

Diversity of undercounts

When we look at subgroups within the population, we find significant variations in net undercount by age, sex, ethnicity, and geographic area. Some population groups are more likely to be missed by the census than others. This section looks at how the 2013 PES results vary across population groups in New Zealand, and compares these with the 2006, 2001, and 1996 PES results. Due to methodological changes over time, care is needed when comparing rates from different time periods.

Direct comparisons of coverage with previous PES results should be used with caution as we are unable to assess how these results would have changed had improvements to the matching process been possible on previous surveys. However, it is still useful to compare patterns between PES years in order to see if some groups of the population are becoming relatively harder to count when compared with other groups in the population.

Figures 2–12 contain estimated rates of undercount for each subgroup. Sampling error intervals are included to give the range of the true undercount at the 95 percent confidence level.

Sex and age

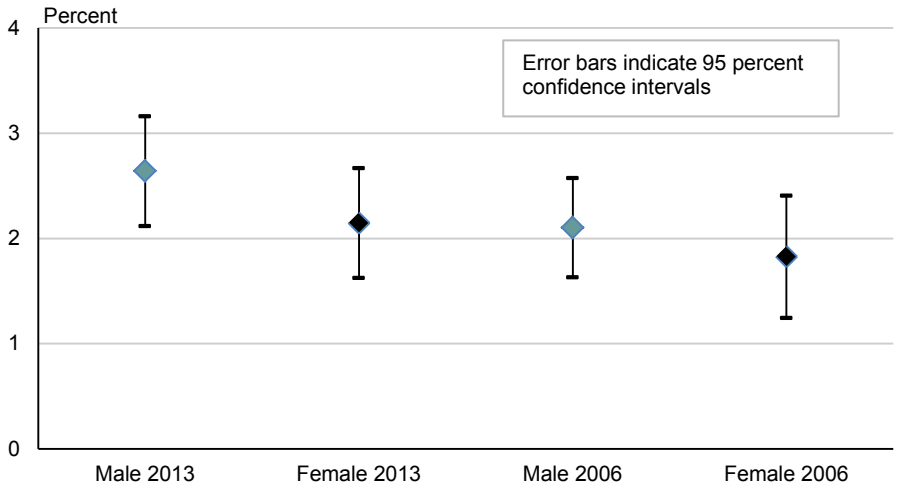
Sex

Table 4 showed that the 2013 Census missed fewer females than males – an estimated 48,000 females compared with 56,000 males. Males accounted for about 54 percent of the estimated 104,000 people missed in the census, although they account for less than 49 percent of the census usually resident population count.

The net undercount rate was 2.6 percent for males and 2.1 percent for females (figure 2). In 2006, the net undercount was lower for both males and females, estimated at 2.1 and 1.8 percent, respectively. The 2013 net undercount had the same sampling error of ±0.5 percentage points for both sexes, so the net undercount ranged between 2.1 and 3.1 percent for males, and between 1.6 and 2.6 percent for females.

Figure 2

Net undercount
By sex
2013 and 2006 Censuses



Source: Statistics New Zealand

Both overseas and New Zealand studies (eg Valentine & Valentine, 1971; Callister, Didham & Bedford, 2006) have shown that males are more likely to be missed in a census than females. For example, in Australia in 2011, the net undercount was 2.2 percent for males and 1.2 percent for females (ABS, 2012). Similarly, in 2006 the Canadian estimated undercount was 3.9 and 1.5 percent for males and females, respectively (Statistics Canada, 2013).

Age

Demographic literature also offers a substantial body of evidence on the undercounting of young children and youth in any census (eg Kerr, 1998; West & Robinson, 1999). Babies and young children are the next groups, after younger adult males, most likely to be missed by the census (Simpson & Middleton, 1997; Baffour, 2006). While the oldest age groups are also prone to being missed by the census, the overall undercount for those aged 65 years and over is very small. It includes age groups that are liable to be over-counted, especially for people who are away from home on census night and for people transitioning into care facilities, who may be counted both at home and in non-private dwellings.

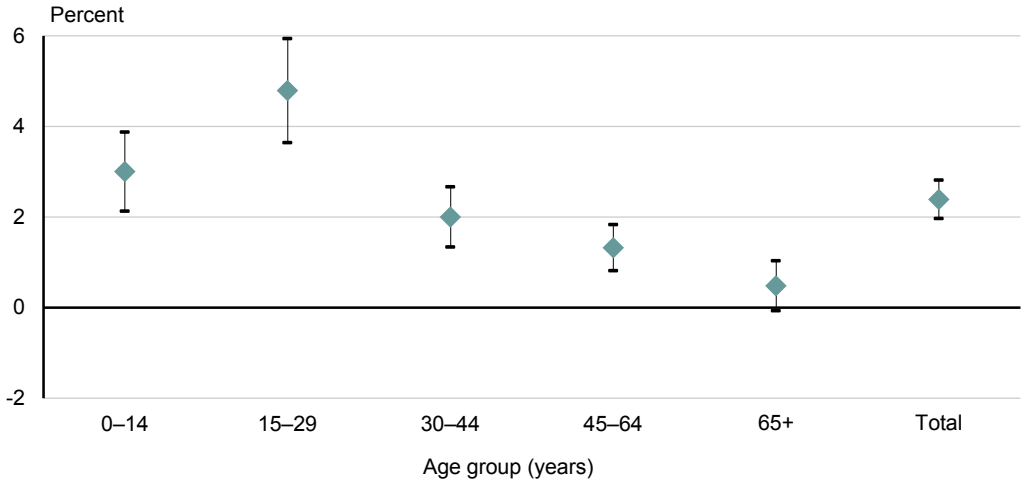
Undercount of babies is significant from a statistical perspective. This is because post-censal estimates based on unadjusted, or inadequately adjusted, base figures highlight the poorly enumerated birth cohorts, and reduce the value of early childhood estimates to planners; for example, in the early childhood education and child health arenas. As long ago as the 1920s, undercount of babies in censuses and population estimates was flagged as critical, because of the major effect this has on the measurement of rare events such as infant mortality (McPhail, 1927).

When the overall sample size is small, demographers face a difficult task in analysing the PES results by age and other characteristics, because sampling errors are relatively large. However, the larger sample size available for 2013 has enabled an additional age group to be analysed. Age data are aggregated to five broad age groups: 0–14 years, 15–29 years, 30–44 years, 45–64 years, and 65 years and over. However, in historical time series the last two groups were combined into one group aged 45 years and over.

In general, the age differentials in the undercount are more pronounced than the sex differentials outlined in the preceding section, with significant variation in the completeness of coverage of age groups (figure 3).

Figure 3

Net undercount
By age
2013 Census



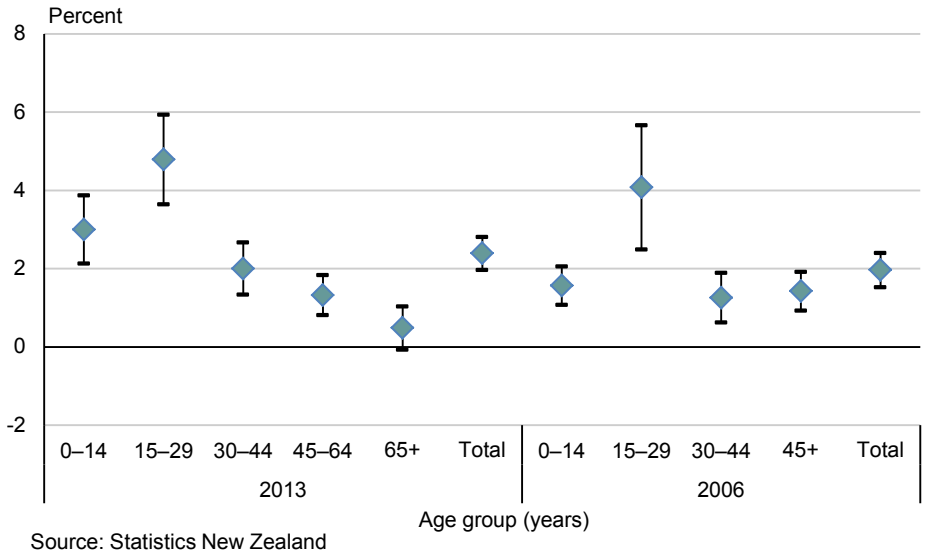
Source: Statistics New Zealand

Based on previous PES experience, younger adults (those aged 15–29 years) are highly mobile and therefore have a higher risk of being missed in a census. About 42,000 people in this age group were missed by the 2013 Census, up from 35,000 in 2006 and 24,000 in 2001. Proportionally, they accounted for 41 percent of those missed in 2013, down from 43 percent of those missed in 2006. This drop can be at least partially explained by changes in migration patterns, especially with respect to international student numbers, between 2006 and 2013. Younger adults also had the largest sampling error at ±1.1 percentage points, compared with ±0.5–0.9 percentage points for the other age groups. The true net undercount rate for younger adults is therefore between 3.7 and 5.9 percent.

The net undercount rate for children (those aged 0–14 years) in 2013 (3.0 percent) was substantially higher than for 2006 (1.6 percent) and slightly higher than that recorded in 2001 (2.7 percent). The sampling error in the 2013 PES was ±0.9 percentage points for this age group. Conversely, people in the 30–44 years, 45–64 years, and 65+ age groups had the lowest undercount rates at 2.0 (± 0.7 percentage points), 1.3 (±0.5), and 0.5 (±0.6) percent, respectively. The true net coverage for the 65+ age group lies between a net undercount of 1.1 percent and a small net overcount of 0.1 percent. Figure 4 shows that, although there were increases in the net undercount for younger age groups between 2006 and 2013, the overall pattern remains similar.

Figure 4

Net undercount
By age
2013 and 2006 Censuses



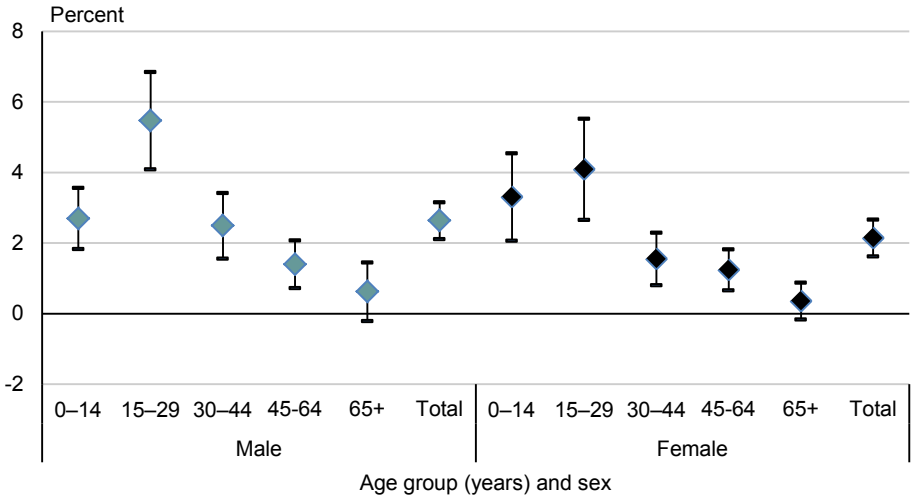
Source: Statistics New Zealand

Age and sex

The net undercount profile by age differs between males and females (figure 5). Generally, the census tends to miss males at all ages more frequently than females, although in recent decades these behavioural differences between men and women have become smaller. PESs in New Zealand have confirmed this pattern. The 2013 PES shows that females' net undercount rate is larger than males' among those aged under 15 years. However, the relative sampling errors suggest that this result is not statistically significant. It is, however, a feature that needs further investigation as it may be related to the methods used for age and sex imputation in the census.

Figure 5

Net undercount
By age and sex
2013 Census



Source: Statistics New Zealand

Ethnicity

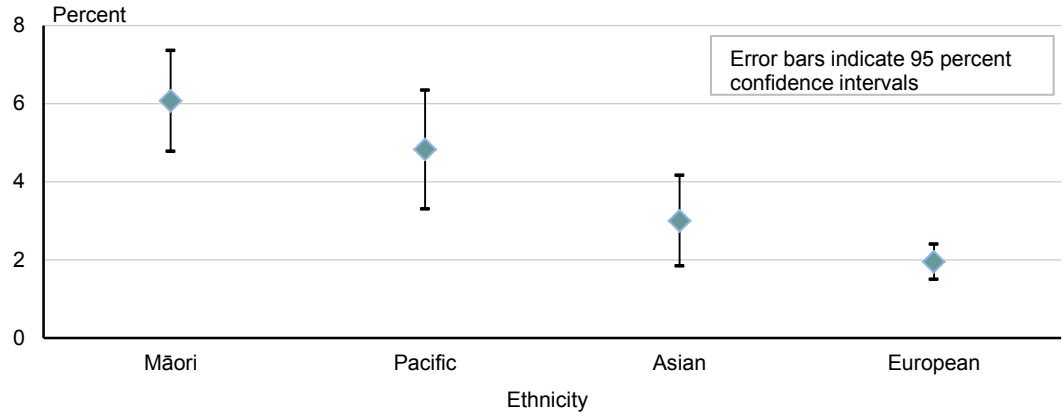
Undercount continues to vary between ethnic groupings. Care is required when comparing the ethnic undercount rates with other undercount rates in this report. Rates by ethnicity are not comparable with previous PESs, because of methodological differences. The rates given in this section however can be compared with each other, enabling us to understand the relative undercount of each ethnic grouping within the 2013 PES. Coverage data by ethnicity is confined in this report to level 1 of the Statistical Standard for Ethnicity. This means the data is restricted to people of European, Māori, Pacific, and Asian ethnicities, which are, in population terms, the four largest in New Zealand.

People may, and do, identify with more than one ethnicity. Each of these ethnicities may be in more than one level 1 grouping of ethnicities within the statistical standard for ethnicity classification. This means that people are counted in each relevant grouping and the totals for the groupings sum to more than the number of people in the subject population.

Of the six level 1 categories described in the classification of ethnicity, the grouping of ethnicities referred to as the European ethnic group is by far the largest. Almost 70 percent of the 2013 PES estimated total population (3.028 million out of 4.346 million people) were European. The net undercount for this group was found to be 1.9 percent, with a sampling error of ±0.5 percentage points (figure 6).

Figure 6

Net undercount
By ethnicity
2013 Census



Source: Statistics New Zealand

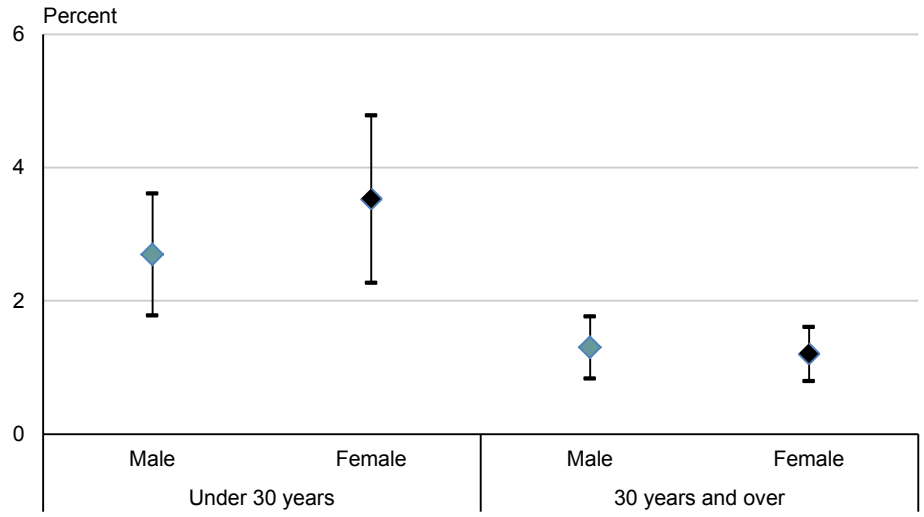
The increased sample size in the 2013 PES has made it possible to provide slightly more detailed information on undercount by ethnicity. In this section, we include details of net undercount for the level 1 ethnic groupings by age and sex. The age split is, however, limited to people under the age of 30 years, compared with those aged 30 years and over.

European net undercount

People of European ethnicities make up the majority of the population and, as expected, the pattern for this grouping is similar to that of the total population. Figure 7 shows the net percentage undercount for people of European ethnicities, split by broad age group and sex.

Figure 7

Net undercount of people of European ethnicities
By age and sex
2013 Census



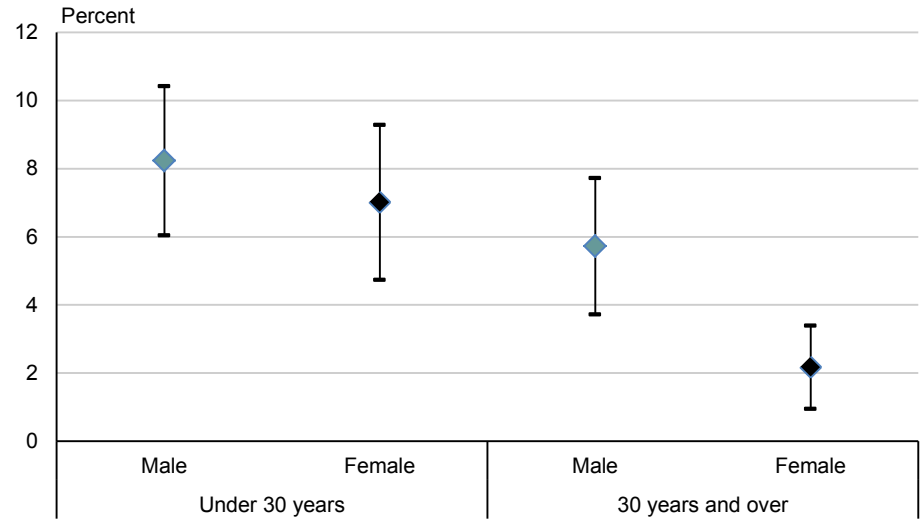
Source: Statistics New Zealand

Māori net undercount

The rate of Māori net undercount was 6.1 percent in 2013, with a sampling error of ±1.3 percentage points (figure 8). One important consideration is that the Māori population is a young population, and at least part of the undercount rate results from the predominance of people under 30 years of age and from the relatively small numbers in the older age groups. Moreover, the Māori population is more likely to have a higher undercount because many live in areas that are harder to count. The net undercount rate for Māori males is not only higher than that for Māori females, it is high for all ages, whereas the net undercount rate for Māori females aged 30 years and over is substantially lower than for those under 30 years of age.

Figure 8

Net undercount of people of Māori ethnicity
By age
2013 Census



Source: Statistics New Zealand

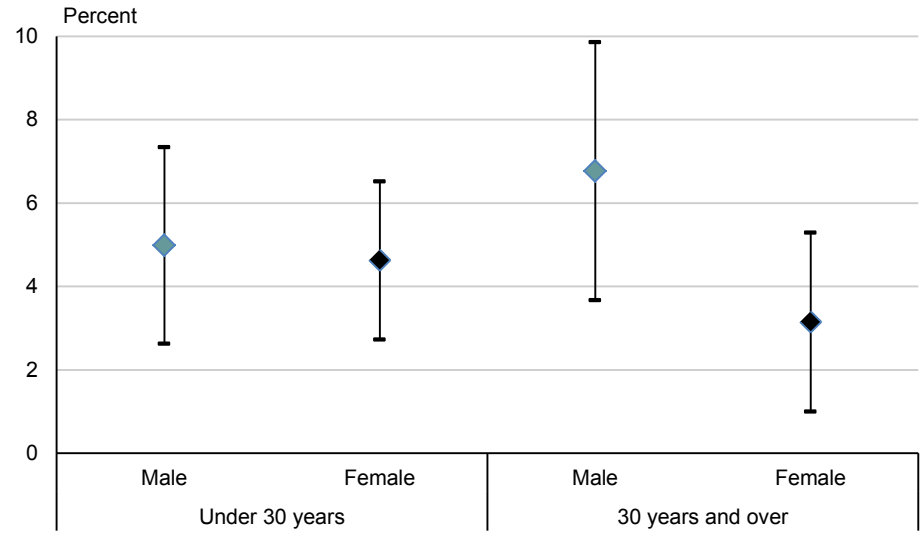
Pacific net undercount

The Pacific population, with a net undercount of 4.8 percent in 2013 (figure 9), has a young age profile and geographic distribution that are similar to those for the Māori population, although it is smaller and more urbanised. The sampling error for the Pacific population was ±1.5 percentage points.

Pacific males show a different profile from males in other ethnic groupings, with Pacific males aged 30 years and over having a higher undercount than Pacific males under 30 years of age, and indeed higher than the rate for Māori males aged 30 years and over. The small sample size in the PES partly explains this, and the large percentage sampling error is an indication that caution is required when interpreting this statistic.

Figure 9

Net undercount of people of Pacific ethnicities
By age
2013 Census



Source: Statistics New Zealand

Multiple ethnic identifications contribute to the difference between the younger and older age groups as well as to the relationship between the undercount for the Pacific ethnic grouping relative to both Māori and European ethnic groupings. In the 2013 Census, over 41 percent of Pacific males and females aged under 30 years also had ethnicities other than Pacific. For those aged 30 years and over, 17 percent of males and 19 percent of females had both Pacific and other ethnicities.

The vast majority of multi-ethnic Pacific people in the 2013 Census were either also Māori and/or European. Among the Pacific people aged under 30 years, over 55 percent were also of Māori ethnicity and 69 percent also reported one or more European ethnicities. For Pacific people aged 30 years and over, nearly 36 percent were also Māori and over 71 percent also European.

Asian net undercount

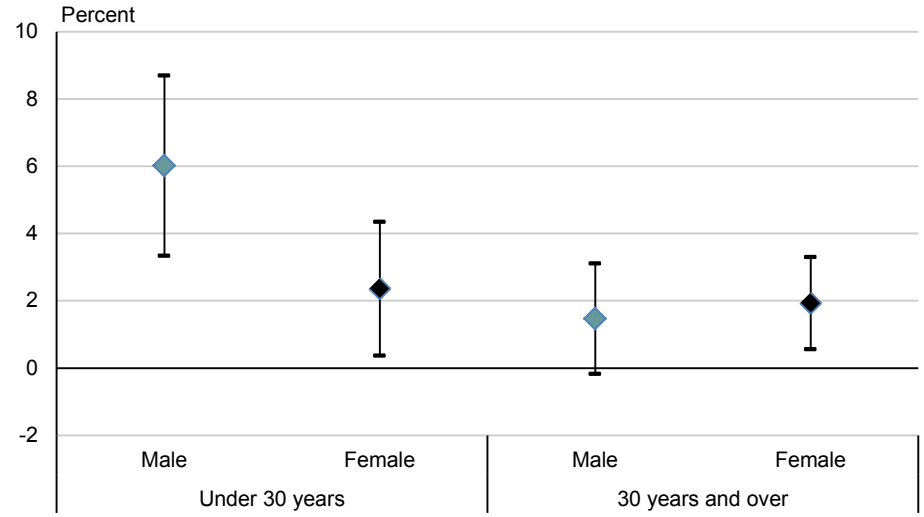
The Asian population in 2013 had a net undercount of just 3.0 percent. This was much lower than the 5.2 percent estimated in 2006. A number of factors contributed to this change. A major factor is thought to be the downturn in the number of overseas students between 2006 and 2013. Foreign tertiary students are predominantly in the hard-to-count younger ages, and tend to be more mobile than other groups. They are also more likely to be missed by the census since many would not realise that they are required to participate, and many may also have been recorded as overseas visitors if they had only recently arrived to start their courses and had not noted that they were intending to stay longer than one year.

The relative sampling errors, though, suggest we should not overanalyse the difference between 2006 and 2013. The increased sample size for the 2013 PES enabled a more robust estimate of Asian undercount, with the result that the sampling error in 2013 was reduced to ±1.2 percentage points. This means that the true net undercount percentage lies, with 95 percent probability, between 1.8 and 4.2 percent. Data users should treat the estimated undercount for the Asian population in 2006 with caution because the sampling error was ±3.4 percentage points – an undercount in the range of 1.8 to 8.6 percent for the 2006 Census.

Figure 10 shows the net undercount for people of Asian ethnicities follows a very similar pattern to other ethnic groupings, with higher undercount rates for the younger population than for the older population. While we expect younger males to have a higher rate of undercount than younger females, the difference among the Asian population is larger than expected, with males at more than double the rate of females (6.0 percent for males aged under 30 years, compared with 2.4 percent for females aged under 30 years).

Figure 10

Net undercount of people of Asian ethnicities
By age
2013 Census



Source: Statistics New Zealand

At least some of the explanation for the Asian patterns relates to social settings and migration histories. Data from the 2013 Census provides some insights. While Asian students in the younger adult ages (those aged 15–29 years) are slightly more likely to be male, females of Asian ethnicities are much more likely to be partnered and to not be studying. Undercount for single younger adults and for foreign students is higher than for other groups. The number of students, in particular, affects Asian patterns more than those of other ethnic groupings.

Net undercount of other ethnicities

The remaining two level 1 categories of the ethnicity classification (Middle Eastern / Latin American / African (=MELAA) and Other) are not covered in this report. We combined these two groupings into a single residual group for the 2013 PES analysis because the sample was too small to treat each separately. However, even the combined group in the PES sample was too small to enable any robust estimate of coverage. The sampling errors were too large for the results to have any statistical significance. The PES sample included very few in the MELAA ethnic group. The Other ethnic group is largely made up of people who gave New Zealander as one of their ethnicities. Behavioural evidence from other sources for this group suggests that there should be little difference between the European rates and the Other rates.

Need for measuring diversity

In many countries, ethnic, indigenous, or racial minorities, in particular, tend to be poorly counted. Similarly, in all countries with diverse populations, there is high importance placed on the accurate measurement of diversity. In Australia's 2011 PES, the net undercount rate for indigenous Australians was considerably higher than that of non-indigenous Australians (ABS, 2012). Similarly, in 2000 in the United States, where there

was an overall net overcount of 0.5 percent, the net undercount rate for African-American males aged 30–49 years exceeded 8 percent (United States Census Bureau, 2003). The United States also recorded higher-than-average undercount rates for other race groups (Kaneshiro, 2013).

In New Zealand, high undercount rates for ethnic minority groups have been a concern since the first PES. For example, the 2001 PES found that the net undercount was much higher for the Māori and Pacific populations (4.4 and 5.2 percent, respectively) than for the European population (1.7 percent). Consequently, for the 2006 Census, we identified a need to find more-effective methods of reaching Māori and Pacific people, as well as other ethnic and youth communities (Statistics NZ, 2006). We needed to place more focus on areas such as Auckland, which has more ethnic diversity than the rest of New Zealand, and areas with high proportions of Māori, such as Northland and Gisborne regions.

Māori and Pacific liaison officers led communications programmes to encourage people to participate, and for the first time, two youth ambassadors (māngai rangatahi) took the 2006 Census message directly to youth (further detail on the 2006 Census communication strategy is found in the Statistics NZ report Introduction to the census). These programmes were strengthened in the planning for the abandoned 2011 Census and the subsequent communications campaign for the 2013 Census.

Geographic area

In addition to age and ethnicity, geographic area of residence can be an important part of the explanation for differences in census coverage. Some areas are harder to count accurately because of their location, while others are harder to count because of the characteristics of the population, such as high levels of mobility, social deprivation, or a predominance of younger adult males. However, defining spatial variations in coverage is an important element in correctly understanding spatial inequalities observed in census data.

Spatial inequality is a central policy concern. Overseas studies indicate significant variations in census coverage both between and within areas, especially in countries with significant levels of geographic diversity. In 2006 in Canada, for example, undercount rates for provinces and territories ranged between 0.8 and 5.6 percent, while state capitals were better enumerated than the remaining state areas (Statistics Canada, 2010). When the UK Office for National Statistics was designing their 2011 Census Coverage Survey (CCS), they took into account high diversity within areas and the effect this has on the reliability of coverage estimates (Abbott & dos Santos, 2010). Areas with high levels of undercount may be disadvantaged because their needs may be inadequately identified.

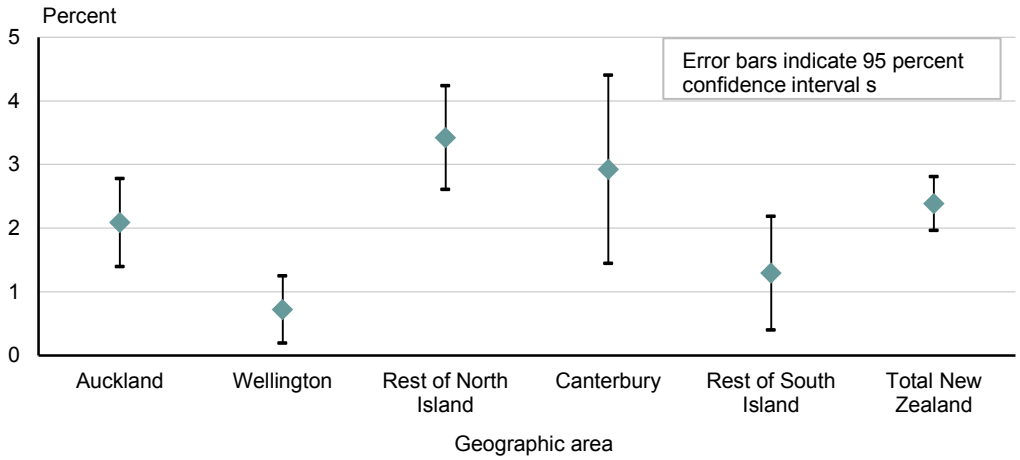
The PES, however, is limited by sample size in what can be described with statistical validity. The increased sample size in the 2013 PES made it possible to calculate census coverage rates at a slightly more detailed level than was previously possible, but still has not enabled robust results to be broken down to smaller geographic areas. In 2013, three key regions, Auckland, Wellington, and Canterbury, had samples large enough to produce reliable results. The remaining areas of the North and South islands have been aggregated. These are shown in figure 11.

The net undercount for the three regions, Auckland, Wellington, and Canterbury, each with large urban populations, was 2.1 percent, 0.7 percent, and 2.9 percent, respectively. The sampling errors for Auckland and Wellington regions were low at ±0.7 percentage points and ±0.5 percentage points. The sampling error for the Canterbury region was higher at ±1.5 percentage points, reflecting difficulties related primarily to the ongoing consequences of the Canterbury earthquakes, in particular the impact of unoccupied dwellings.

The rest of the North Island had an undercount of 3.4 percent, compared with the rest of the South Island at 1.3 percent.

Figure 11

Net undercount
By geographic area
2013 Census



Source: Statistics New Zealand

For historical comparisons, the data is also presented on the same geographic basis as was done in 2006, when the North Island was split between northern North Island and southern North Island. The northern North Island consists of the Northland, Auckland, Waikato, and Bay of Plenty regions and is home to more than half of all New Zealanders. The southern North Island comprises Gisborne, Hawke's Bay, Taranaki, Manawatu-Wanganui, and Wellington regions.

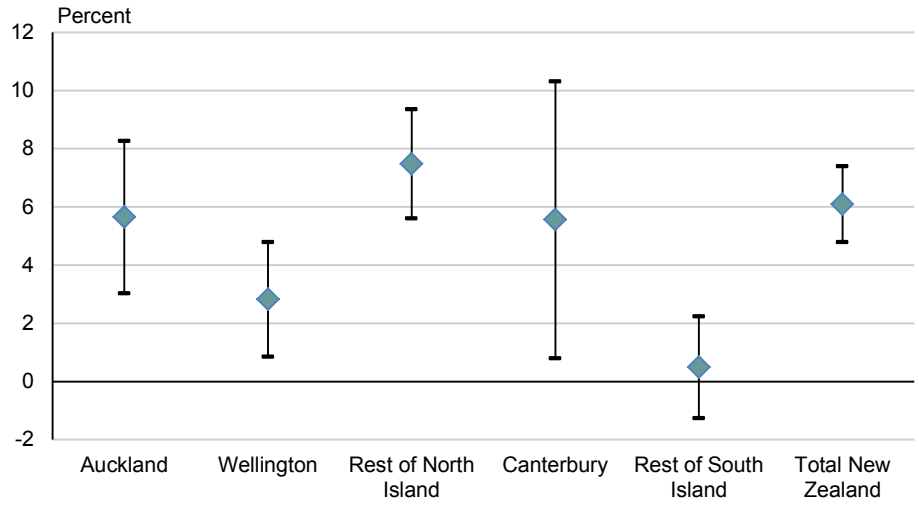
The northern North Island, comprising four highly diverse regions, recorded a net undercount of 3.0 percent in 2013. This was an increase on the 2006 PES net undercount figure of 2.4 percent. The southern North Island had the same net undercount of 1.3 percent in both 2013 and 2006.

The South Island as a whole had a net undercount of 2.2 percent in 2013. In the 2006 PES, the net undercount for the South Island was 1.8 percent.

Regional data for the Māori population is fundamentally important for Māori development. The increased sample size for the 2013 PES makes it possible to look at regional net undercount for people of Māori ethnicity. Data users should be aware that the sample for some areas is small and the results should still be treated with caution, as the size of the sampling errors, especially for Canterbury, indicate (figure 12). The results are, however, robust enough to illustrate the diversity in coverage for Māori at a subnational level, but the overall pattern geographically is very similar to the pattern shown in figure 11.

Figure 12

Net undercount of people of Māori ethnicity
By geographic area
2013 Census



Source: Statistics New Zealand

Undercoverage of dwellings

In previous PESs, coverage of dwellings was found to be very high (99.5–99.8 percent). Following the introduction of automatic matching in the 2013 PES, it was more difficult to match vacant and non-responding dwellings to their corresponding census dwelling. This is because automatic matching was centred around linking of individuals, with the dwelling match occurring as the secondary step in most cases. Where there were no individuals to match to census records it became much harder than previous manually processed surveys to determine if the dwelling was in fact a match. Vacant and non-responding dwellings showed an unexpectedly high non-match rate in 2013. This was probably a consequence of the new method rather than the census actually missing more dwellings. Dwelling undercount estimates are therefore not included in this report as they are not deemed to be sufficiently reliable under the new method.

Key performance indicators

Two sets of performance indicators are relevant to the 2013 PES: those related to the census and those related to the PES itself.

2013 PES performance indicators

The response rate for the 2013 PES is an important quality measure. The PES achieved a response rate of 87 percent, which was short of the target of 90 percent, but in line with international experience. For example, Scotland also achieved an 87 percent response rate in their 2011 CCS (National Records and Scotland, 2013), although England/Wales's CCS and Australia's PES both achieved 90 percent or more in 2011.

The New Zealand 2013 PES achieved a net undercount sampling error at the national level of ±0.5 percentage points, outside the target of ±0.3 percentage points. On the other hand, most subpopulations were within the ±1.0 percentage-point target, as expected with the increased sample size. The notable exception was the sampling error percentages for ethnic groupings, where 3 of the 4 broad ethnic groupings included in this report had sampling errors for net undercount of between ±1.2 and ±1.5 percentage points, reflecting sample size and measurement aspects.

2013 Census performance indicators

The 2013 Census established a set of indicators to measure the success of census collection and processing. Two key performance indicators for the census are measured by the PES: the coverage rate and the response rate.

The 2013 census did not meet either the coverage or the response rate targets. The 2013 PES estimated the national coverage rate of 97.6 percent was just below the target for coverage of 98.0 percent. The PES also found that the achieved census national response rate of 92.9 percent was below the target of 95.0 percent.

4 Where to from here?

Adjusting the base population

The 2013 PES provides information on census coverage for developing the base population for estimated resident populations (ERP) for the period 2013–18. The ERP is now based at 30 June 2013. We derive the ERP by adjusting census counts to account for New Zealand residents temporarily overseas on census night; along with the undercount found by the PES; and births, deaths, and international migration flows between the census and 30 June 2013.

The report *Estimated resident population data sources and methods* (Statistics NZ, 2014b; in press) describes the methods used to derive the ERP, and explains how the PES results are incorporated into the population estimates. Recalibration of estimates to the new census is an important part of maintaining robust information on population change. New experimental estimation methods require further work and will contribute to ongoing developments for the 2018 Census and 2018 PES. However, the 2013 PES was able to assess the validity of substitute-record creation in the census for the first time. These changes have provided a more complete view of census coverage, both by the key individual characteristics of age, sex, and ethnicity, and by the geographic variability in census coverage.

Investigation of a new experimental method

Data collected in the PES is used to estimate census coverage for both people and private dwellings. We undertook substantial research after the 2006 PES in order to improve the estimation methods. For the 2013 PES, we investigated a new estimation method based on methods used by the Australian Bureau of Statistics (ABS) in their 2011 PES.

The new method includes estimating the size of the whole non-contact sector in the census, including both people and dwellings missed by the census, and substitute records created by the census.

How does the new experimental method compare with the 2006 and 2013 methods?

All methods estimate census coverage. The basis of the methods is the calculation of two variables - 'SHOULD' and 'WAS':

- SHOULD represents whether a PES person SHOULD have been counted in the census (should=1) or SHOULD NOT have been counted in the census (should=0). The latter case corresponds to the situation where a PES person said they were overseas on census night.
- WAS represents whether or not, and how many times, the person was counted in the census. WAS may be imputed, in which case it takes a value between 0 and 1 based on the probability of 0 or 1, and assumes that the true value is not greater than 1.

The essential difference between the methods is that the 2006 method estimates census undercount based on the contact sector alone; the new method includes an estimate for the non-contact sector, and hence for substitute records. The 2013 method is very similar to the 2006 method but provides some adjustment to the census substitute count. Table 7 compares the derivations of the coverage adjustment ratios and the PES estimated populations using each of the methods.

Table 7

Comparison of estimation methods

	Coverage adjustment ratio (CAR)	PES estimated population
2006 method	$CAR_{2006} = \frac{\sum_i w_i SHOULD_i}{\sum_i w_i WAS_i}$ with all summations over the contact sector	$= CAR_{2006} * Census_{Contact} + Census_{Non-contact}$
2013 method	$CAR_{2013} = \frac{\sum_i w_i SHOULD_i}{\sum_i w_i WAS_i}$ with all summations over the contact sector	$= CAR_{2013} * Census_{Contact} + Census_L + \sum_j p_j \, Census_{S,j}$
New experimental method	$CAR_{new} = \frac{\sum_{i \in U} w_i SHOULD_{U,i}}{\sum_{i \in \text{Contact}} w_i WAS_i + Census_{Non-contact}}$	$= CAR_{new} * Census_{Total}$

Source: Statistics New Zealand

2006 method

The 2006 method calculated the weighted sum of 'should' and 'was' for PES eligible respondents in the contact sector only. Late returns and substitute records were not included.

The ratio of weighted sums of SHOULD and WAS gives the coverage adjustment ratio (CAR). The CAR is multiplied by the census count for the contact sector to obtain the PES estimate of the population in the contact sector. The census count for the non-contact sector is then added on to give the PES estimate of the total resident population.

The limitation with this method is that it produces a biased coverage estimate. It assumes that the census measures the non-contact sector perfectly. In particular, it assumes that the correct number of substitutes has been raised. However, it is a difficult task for a census collector to determine whether they should treat a 'non-responding' dwelling as occupied (in which case it becomes a substitute household with an estimated number of substitute individuals) or unoccupied (which includes dwellings under construction, vacant dwellings, and 'residents away'). For example, if the census raises too many substitutes, the PES estimated population will be too large and coverage will be too low.

Details of new experimental method

The new method calculated the weighted sum of SHOULD for all PES eligible respondents (ie, in both contact and non-contact sectors). Note this is different from the 2006 method.

The weighted sum of WAS is the same as in the 2006 method – that is, for the PES eligible respondents in the contact sector only.

In this method CAR is the ratio:

$$\frac{\sum_{i \in U} w_i SHOULD_{U,i}}{\sum_{i \in \text{Contact}} w_i WAS_i + Census_{Non-contact}}.$$

The PES estimate of the population is obtained by multiplying the total census count by the new CAR.

The new method is theoretically superior to the 2006 method, and has been shown to work very well in Australia. However, the larger-than-expected number of substitute records in the New Zealand 2013 Census meant that missing information for the non-

contact sector became significant. The application of weighting for individuals proved to be statistically very sensitive, and it is clear that more research is required in this area.

The main limitation in the method used in 2013, and a difficulty for the new method, is that important variables such as ethnicity are not imputed for substitute records. Many PES records matched to substitute records were from people who gave Māori as one of their ethnicities. The missing ethnic information for substitutes means the current PES has limitations in its coverage measures for ethnic groupings.

Specifically, the PES tends to overstate coverage and underestimate undercount for some ethnic groupings. We recognised that considerably more work is required to ensure we are able to produce sound results with the new experimental method in the New Zealand environment, and decided to proceed with the 2013 PES with a modification of the method used in the 2006 PES. The main modification was that we were able to test the validity of substitute records in all-substitute dwellings for the first time in PES.

Investigating the new method did delay release of results from the 2013 PES. However, the benefits derived were significant. As a result of this work we now have a better understanding of the relationship between under-coverage and substitute records in the census, and also of non-sampling errors. We also gained further knowledge about sensitivity of weighting methodologies, along with the effects of classification errors and inconsistencies between the 2013 Census and the 2013 PES. These represent important gains in our knowledge to be carried forward to the 2018 PES.

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.

Australian Bureau of Statistics (2012). Census of population and housing – details of undercount. Canberra: Australian Bureau of Statistics. Available from www.abs.gov.au.

Baffour, B (2006). Modelling census under-enumeration: A logistic regression perspective. Edinburgh: General Register Office for Scotland. Available from www.gro-scotland.gov.uk.

Callister, P, Didham, R & Bedford, R (2006). Changing sex ratios in New Zealand: Real change or a statistical problem? *New Zealand Population Review*, *32*(1), 21–33.

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. Available from www.abs.gov.au.

Felligi, I & Sunter, A (1969). A theory for record linkage. *Journal of the American Statistical Association, 64*(328), 1183–1210.

Kaneshiro, M (2013). Missing minorities? *Demography*, *50*(5), 1897–1919.

Kerr, D (1998). A review of the procedures for estimating the net undercount of censuses in Canada, the United States, Britain and Australia (Demographic Division, research paper no. 5). Ottawa: Statistics Canada. Available from www.statcan.gc.ca.

Kish, L (1965). *Survey sampling.* New York: John Wiley & Sons, Inc.

Marks, E, Maudlin, WP, & 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), 220–243.

McPhail, E (1927). Infant mortality as shown by Canadian vital statistics. *American Journal of Public Health*, *1927*, 476–484.

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

National Institute of Statistics of Angola (2014). *Censo em Angola.* Luanda: National Institute of Statistics of Angola.

National Institute of Statistics of Rwanda (2014). Fourth population and housing census, Rwanda, 2012: Post-enumeration survey report. Kigali: National Institute of Statistics of Rwanda. Available from www.statistics.gov.rw.

National Records of Scotland (2013). 2011 Census in Scotland: How the population estimates were obtained – Release 1C. Edinburgh: National Records of Scotland. Available from www.scotlandscensus.gov.uk.

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 DC: US Department of Commerce.

Simpson, L and Middleton, E (1997). Who is missed by a national census? A review of empirical results from Australia, Britain, Canada, and the USA. Manchester: University of Manchester, Centre for Census and Survey Research. Available from www.ccsr.ac.uk.

Statistics Canada (2010) 2006 Census technical report: Coverage. Ottawa: Statistics Canada. Available from www.statcan.gc.ca.

Statistics Canada (2013). Census of population – reverse record check (RRC). Ottawa: Statistics Canada. Available from www.statcan.gc.ca.

Statistics New Zealand (1998). A report on the 1996 Post-enumeration Survey. Wellington: Statistics New Zealand. Available from www.stats.govt.nz.

Statistics New Zealand (2002). A report on the 2001 Post-enumeration Survey. Wellington: Statistics New Zealand. Available from www.stats.govt.nz.

Statistics New Zealand (2005). Statistical standard for ethnicity. Wellington: Statistics New Zealand. Available from www.stats.govt.nz.

Statistics New Zealand (2006). 2006 Census of population and dwellings - introduction to the census. Wellington, Statistics New Zealand. Available from www.stats.govt.nz.

Statistics New Zealand (2013). Introduction to the census. Wellington: Statistics New Zealand. Available from www.stats.govt.nz.

Statistics New Zealand (2014a). Understanding substitution and imputation in the 2013 Census. Wellington: Statistics New Zealand. Available from www.stats.govt.nz.

Statistics New Zealand (2014b). *Estimated resident population data sources and methods.* In press. Wellington: Statistics New Zealand. Available from www.stats.govt.nz.

United Nations (2010). Post-enumeration surveys: Operational guidelines. New York: United Nations. Available from www.unstats.un.org.

US Census Bureau (2003). Technical assessment of ACE revision II. Washington DC: US Census Bureau. Available from www.census.gov.

Valentine C & Valentine, B-L (1971). Missing men: A comparative methodological study of underenumeration and related problems. Washington DC: US Census Bureau. Available from www.census.gov.

West, K and Robinson, J (1999). What do we know about the undercount of children? (Population Division working paper No. 39). Washington DC: US Census Bureau. Available from www.census.gov.

Appendix

List of acronyms

ABS – Australian Bureau of Statistics

CAR – census adjustment ratio (=should have been counted/was counted)

CCS – census coverage survey

HLFS – Household labour force survey

HSF – Household sampling frame

MIS – matching impact study

NZGSS – New Zealand General Social Survey

PES – Post-enumeration Survey

PSU – primary sampling unit

QS – QualityStage (software)

RRC – reverse record check

Definitions of terms

Contact and non-contact sectors

We have borrowed the terminology of census 'contact sector' and 'non-contact sector' from the ABS. We define the contact and non-contact sectors at 'PES time', which is the date at which fieldwork for the PES begins – 9 April 2013 for the 2013 PES.

Contact sector

The contact sector is made up of census forms that have been returned by PES time. The contact sector includes all people that were **not** specifically at non-contact sector dwellings in the census; ie it includes people missed at counted dwellings and people at missed dwellings.

Non-contact sector

The non-contact sector includes people at private dwellings located in the census, but from which no form was received by PES time. This group is separated out as of interest because the census has imputed the number of usual occupants for these dwellings, and we may want to evaluate how well that imputation reproduced the actual population in those dwellings.

The non-contact sector is made up only of people living in private dwellings at census time. This group comprises two types of people associated with the terms 'late returns' and 'substitutes'. Detailed descriptions of these are provided below.

Late returns

A late return is a census form that is returned to the office after the PES goes into the field (9 April 2013 for the 2013 PES). It is important to identify late returns and remove these from the PES estimation process, because the PES and the census are meant 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.

We do not classify census forms collected by the census collector as 'late', because the census collectors had finished field work before the PES started. This means late returns will always be either Internet or mailback forms.

Substitutes

There are three kinds of substitute forms raised in the census:

1. a substitute individual form is raised within a real dwelling (ie we have a dwelling form), when there is a name on the dwelling form not accounted for by the individual forms received.
2. a substitute dwelling form is raised with real individuals attached when no dwelling form was collected, but we have individual forms. In this case, the number of occupants in the dwelling is equal to the number of real individual forms received. This means we cannot have an instance where we have a substitute dwelling form with both real and substitute individual forms belonging to it.
3. a substitute private dwelling with substitute people. This happens when the census collector is fairly sure people are living within a dwelling (ie they have determined its status is 'occupied'), but has not been able to make contact – this information may come from a neighbour. In this case, an all-substitute dwelling is raised and the number of occupants within it are imputed randomly using a donor household, from a 'nearest neighbour' imputation method.

Substitute forms contain very little information – essentially imputed age, sex, and usual residence. For PES purposes, the first type of substitute can be treated as a real person as we can match to the name on the dwelling form – ie we know that the substitute form relates to that person specifically. The second type of substitute form is not a problem for individual matching as we have the real people in the dwelling to match to.

International PESs

A reference table of countries with censuses and/or post-censal evaluations is included with the downloadable tables.

2013 PES questionnaire form

Visit Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey for a PDF of the 2013 PES questionnaire form.