Innovations in the Census Post Enumeration Survey in Australia

Jarvis, Bjorn

Australian Bureau of Statistics

45 Benjamin Way

Belconnen, Australian Capital Territory, Australia, 2617

E-mail: bjorn.jarvis@abs.gov.au

Abstract

Whenever a Census of Population and Housing is undertaken, questions invariably arise about the completeness and accuracy of the Census count. While every effort is made to eliminate the potential causes of error, some undercount and overcount will occur in such a large operation. In Australia, estimates of net undercount in the Census are based on the results of the Census Post Enumeration Survey (PES).

Australia has a long history of conducting a PES after each Census, with a number of key innovations being introduced in recent surveys. This paper highlights some of the key changes to the Australian PES, describes plans for the 2011 survey, and outlines how the Australian Bureau of Statistics uses results from the PES in rebasing the Australian Estimated Resident Population.

1. Introduction

1.1. Why Australia conducts a PES

The five-yearly Census of Population and Housing is a valuable data source for estimating the size and geographic distribution of the Australian population, and for analysing major demographic, social and economic characteristics, particularly for small geographic regions and other small sub-populations. It provides statistics for decision-making by governments, businesses, community organisations and individuals.

Whenever a Census is undertaken, questions about the completeness and accuracy of the Census count invariably arise. In such a large and complex exercise, it is inevitable that some people will be missed and some will be included more than once (or included when they should not be). While every effort is made to eliminate these potential causes of error, some undercount and overcount will invariably occur.

Estimates of net undercount, which take into account undercount and overcount, are used to:

1. Derive an estimate of the resident population for 30 June of the Census year;
2. Provide users with an assessment of the completeness of Census counts, allowing them to take this into account when using Census information; and
3. Evaluate the effectiveness of Census collection procedures so that improvements can be made for future Censuses.

The Australian Census counts people on the basis of where they were on Census night. The ABS augments this Census count for net undercount (as measured by the PES) and for Australian residents who were temporarily absent from Australia on Census night. It then subtracts the count of overseas visitors temporarily in Australia and enumerated in the Census, to form a basis for the calculation of the Estimated Resident Population (ERP) of Australia on a place of usual residence basis.

Accurate ERPs are required for a wide range of uses, including the allocation to Australian states and territories of seats in the Federal House of Representatives, the distribution of Commonwealth payments to states and territories, and to support demographic, social and economic studies.

1.2. The history of the Australian PES

The PES has been undertaken in Australia for each Census from 1966 to 2006. From 1966 until 1976, the PES had three components. It evaluated coverage of persons (that is, the extent of the undercount of individuals), coverage of dwellings (where a separate sample of dwellings was used to compare dwellings listed by PES interviewers with those compiled by Census collectors), and the accuracy of responses to particular Census questions.

From 1981 onwards, the PES has been designed solely to obtain a measure of Census coverage, incorporating information on the undercount of persons and some information on dwellings. Some information continues to be published on classification differences between PES and Census, such as differences in Indigenous status and Country of birth.

1.3. Statistical independence of the Australian PES

The purpose of the PES is to provide an independent check on Census coverage. There are two aspects to this independence: operational independence and population independence. Operational independence requires that Census operations do not influence the PES in any way, and vice versa. ABS controls this very closely as described below. Population independence means that there should be no subgroups of the population where being missed in the Census indicates that a person or dwelling is more likely to be missed by the PES also. This is harder to achieve, but the PES estimation process can adjust for this to some extent by subdividing the population into smaller groups where the assumption of population independence is more likely to be true.

Steps are taken to maintain the operational independence of the PES from the Census at every stage, including enumeration, processing and administration. These steps include selecting the PES sample from an independent sample frame, generally using separate office staff in the PES and Census, ensuring the PES interviewers were not employed as Census field staff in the same area (and vice versa), maintaining the confidentiality of the PES sample so that Census field and office staff are not aware which areas will be included in the PES, and ensuring that Census late returns are treated differently in PES estimation.

The PES questions are asked of householders face-to-face by experienced, highly trained interviewers, whereas most Census forms are self-completed. The PES is also a much smaller scale operation (and hence easier to control) than the Census. These features enable the PES to deliver an accurate estimate of the percentage of people and dwellings missed by the Census.

The Census can also be used to form an estimate of the percentage of dwellings and people missed by the PES. The PES excludes non-private dwellings (hotels, motels, hospitals) for operational

reasons. The PES is also conducted several weeks after the Census, so a respondent's recollection of their location on Census night may not be entirely accurate. Census has special procedures for enumerating homeless people, while the PES is essentially a survey of dwellings and the people who reside in them. Thus the Census may include some dwellings and people that the PES misses. PES estimation implicitly accounts for the dwellings and people missed in the PES but counted in the Census.

On the assumption that the Census and the PES are independent, the estimate of the percentage missed by the PES but found by the Census, and the percentage missed by the Census but found by the PES, can be used to construct estimates of the percentage missed by both PES and Census.

Despite efforts to maintain independence, the likelihood of a person being missed in the PES may be related to whether they were missed in the Census. This may result in a 'correlation bias' in the PES estimates. To minimise this bias, PES estimation takes account of the fact that different groups have a different likelihood of being missed.

2. Key changes in the Australian PES in the last 10 years

2.1. PES data capture in the field

Computer Assisted Interviewing (CAI) was introduced into the PES in 2006, replacing Paper and Pen Interviewing (PAPI). CAI refers to the use of an electronic Blaise questionnaire (Blaise software was developed by Statistics Netherlands), into which the ABS Interviewer enters responses as the interview progresses. The CAI questionnaire automatically sequences the questions that the Interviewer asks, which removes the need for sequence guides. It also has the capacity for providing Interviewers with prompts and other useful information.

The introduction of CAI has generated gains in both efficiency and quality. Conducting a PES has become considerably more efficient, now that there is no longer a requirement to scan completed PES forms from the field during the data capture phase. There have also been important gains in the quality of information from direct data capture into an electronic form, due to the considerable reduction in the risk of transcription and sequencing errors and problems with character recognition.

While these gains have also benefited other household surveys, CAI offers some PES-specific gains over PAPI, in relation to the number of persons and search addresses (i.e. addresses where a person may have been counted to be explored during the matching phase) that can be entered, given it is not constrained by the physical limits of paper.

However, it is important to note that the CAI mode of collection was not considered appropriate for enumerating discrete Indigenous communities (i.e. a geographic location, bounded by physical or cadastral boundaries, inhabited predominantly by Aboriginal and Torres Strait Islander peoples) in 2006. Instead, a PAPI questionnaire was used to collect information similar to that collected from private dwellings in the ‘mainstream’ using the CAI questionnaire. The information collected on PAPI questionnaires was later transcribed to a CAI instrument. This ensured that, as with the mainstream sample, that all responses were captured electronically by ABS interviewers before the survey left the field.

In 2011 both modes of collection will be available to interviewers in the communities, to provide Interviewers with the flexibility to select the most appropriate option for a given situation.

In respect of technology, in 2006 each ABS interviewer was provided with a notebook to conduct their computer assisted interviews, which will again be the strategy for 2011. It is likely that new technology options will be explored for 2016, to take advantage of computer and telephone

technology improvements in both mobility and functionality.

2.2 Match and Search System

A key enabler for both efficiency and effectiveness in the processing of the Australian PES has been the development of systems that support the clerical matching and searching exercise. Considerable gains were made possible in 2006 by the development of a Match and Search System, a software system built specifically for PES processing.

As the name suggests, the MSS allowed processing clerks to search, view, compare, and record matches between PES and Census data. PES processing clerks used the MSS to record matches of dwellings and people between PES and Census, and to search for people included on Census forms at all alternative addresses provided.

A key innovation of the system was the use of 'snippets' of the Census collector record books, whereby a clerk could view scanned sections of Census forms electronically. The system was therefore able to directly source these snippet images from the Census data repository, removing the need to access physical forms.

This system also facilitated the recording of more detailed information about the reasons for matching decisions. This detailed information is useful in gauging the quality of matches, above simply determining that a match has been made.

Enhancements have been made to this system for 2011 to take advantage of gains in efficiency and effectiveness from automated data linking. This ensures that the strengths of clerical matching are retained as a complement to the strengths of automated linking, which are discussed later in this section of the report. The MSS is also used as a means of quality assuring automated linking, particular given this is the first instance of such linking being incorporated into PES processing in the ABS.

2.3 Automated Data Linking

Automated Data Linking (ADL) is being introduced into the Australian PES in 2011. This follows an evaluation exercise which was undertaken by linking experts within the ABS after the 2006 PES. ADL refers to the use of automated linking processes to determine possible links between Census and PES data, before any clerical matching process has begun.

ADL employs probabilistic linking techniques, using a range of personal and address characteristics, to evaluate the likelihood that a PES and Census record pertain to the same individual. ADL therefore provides the opportunity to match persons that would have been too difficult to match previously, given the constraints of previous technology and processes.

The key advantages of automated data linking include:

1. Establishing links between persons in PES and Census data that would not have been possible from previous clerical matching processes - a gain in matching *effectiveness*;
2. Identifying addresses where a person may have been counted that were not identified by PES respondents - a gain in matching *effectiveness*; and
3. Reduced clerical matching activities - a gain in matching *efficiency*.

While ADL is the next step in the evolution and continual improvement of PES processing, it is important to note that ADL cannot entirely replace the clerical decision-making process that has previously been at the core of PES processing. Clerical judgment will always be required to resolve the more complex or ambiguous cases, and be used as a means of quality assuring automated

processes. Some adjustments to the clerical match and search processes have therefore also been necessary, to ensure that the relative strengths of both ADL and MSS are fully realised.

For 2011, ADL has been incorporated into the PES processing sequence used in 2006, and therefore complements the existing processing stages. In 2016 it is likely to be the primary processing element, with the rest of the processing sequence re-engineered to draw on its strengths to the fullest extent.

2.4 Geography coding

Address information has always been important to the ability to match people between PES and Census data. Matching a person between the data sources is facilitated by the location of the dwelling where they were enumerated in PES, where they were included on a Census form and where they may have been included on a Census form. Having a consistent geography is important, both in terms of classification and level of detail.

In 2006, prior to processing PES data, each enumeration address had a Collection District (CD) associated with it, drawn directly from the sample. Each search address was also coded with a Collection District (CD), using the AddressCoder@ABS application. This CD was then used as the default starting point for dwelling and person searching using PES addresses.

Occasionally the information obtained in the PES interview did not contain all the necessary details for the AddressCoder@ABS to return a single CD. In such cases, where a search address could not be isolated to fewer than 5 CDs, a default CD was chosen by referencing the longitude and latitude coordinates for the geographic centre of the locality. The CD that encompassed these coordinates was chosen as the default CD. In situations where a search address could be identified to within five CDs by the AddressCoder@ABS, the default CD was chosen as the most likely location to commence searching. A number of adjacent CDs were also recorded if wider searching was necessary.

In 2011, the coding of addresses has become even more important to the processing of the PES, especially the precision of coding. The introduction of ADL has made it necessary to have a Census enumeration area that can be used as a filtering variable for a number of the ADL runs, which requires a single collector workload to be identified.

As in 2006, the coding of addresses is largely an automated process, using the AddressCoder@ABS application. However, the increased requirement for precision is such that some additional review and processing by clerks will be necessary, as a single Census enumeration area is required.

The coding of PES data to a number of different geographies also increases the options for producing undercount information for different geographies. For instance, evaluating net undercount by geographic remoteness is of key importance to the effective evaluation of Census field strategies targeted at remoteness.

2.5. Improvements to the quality of Indigenous net undercount

The Aboriginal and Torres Strait Islander population is of particular importance to PES; given it is a population for which coverage issues are an ongoing concern for the Census and a population for which the quality of its population estimates are closely scrutinised. Indigenous net undercount is therefore prominent within the published results and is a measure upon which the ABS has focused a range of innovations.

To improve the quality of Indigenous undercount information in 2006, the PES included remote

areas and discrete Indigenous communities for the first time. Previous cycles had excluded these areas from the scope of the survey because of operational reasons – mainly the additional cost and the need to use the same local contacts as Census, which was considered likely to compromise the independence of the PES. Inclusion of these communities in 2006 ensured the geographic scope of the 2006 PES was more complete than it has been in the past and included more Indigenous persons, recognising that around a quarter of Indigenous persons live in the communities. The risk to statistical independence was effectively managed in 2006 through interviewer training and the procedures put in place for field staff.

To support the inclusion of discrete Indigenous communities in 2006, the ABS developed a second questionnaire, which was tailored to collecting equivalent information to the 'mainstream' questionnaire, but in a more culturally appropriate form. For example, it is more common for Indigenous persons in the community to have more than one name by which they may be known, as they may have a 'community name' and a 'town name'. This tailored questionnaire ensured that the best effort was made to collect sufficient information to effectively match persons living in the communities.

A community level questionnaire was also provided to ABS Interviewers enumerating in the communities, in order to obtain community-specific intelligence, such as events that had occurred around Census enumeration that could be used as reference and information that would be useful in the matching process. Both questionnaires used in the communities have been enhanced for 2011, based on recommendations from 2006 and learnings from the 2010 dress rehearsal (conducted in the communities in August and the mainstream in July). The dress rehearsal also examined the efficacy of enumeration procedures, modes of collection, and the support documentation provided to Interviewers

In 2011, the ABS has continued to make enhancements to the PES, aimed at improving the quality of Indigenous net undercount. Building on the inclusion of communities in 2006, the number of communities selected will be increased, as will the total number of dwellings selected in this part of the sample. This increase in sample from the communities will increase the gains to Indigenous undercount information realised in 2006.

A second sample design strategy to improve Indigenous estimates for 2011 has been to include some additional sample focused on areas of Australia outside of the discrete Indigenous communities that have been identified by the ABS as having a high density of Indigenous persons. This is of particular importance, given approximately three-quarters of the Indigenous population usually reside outside of the communities, in areas covered by 'mainstream' enumeration. By focusing additional sample in high density Indigenous areas the ABS expects an increase in the number of responding Indigenous persons, while ensuring that there is minimal bias introduced.

2.6. Changes to estimation methodology

Considerable innovation has occurred in PES estimation methodology across recent cycles. The ABS has implemented a number of changes that have contributed to a more robust estimate of net undercount, given the importance of PES results in the re-basing of population estimates and in evaluation of the Census. The PES estimation methodology also receives a relatively high level of scrutiny from outside of the ABS, due to the important role of subsequent population estimates in the distribution of Goods and Services Tax revenue and electoral representation.

Following the 2001 PES, a review of the PES estimation method was commissioned, to develop an estimator for the PES that adjusts adequately for non-response and non-coverage in PES and for miscounting in the Census. The estimator used in the 2001 PES (itself replacing the estimator used

from 1966 to 1996) did not fully account for people missed in both the Census and PES.

One outcome of the review was the development of a new estimator, Prediction Regression (PREG), for use in the person weighting stage of the 2006 PES. Unlike the estimator used in 2001, the weight adjustment applied by PREG to each PES person does not depend on their Census response; it does however depend on a number of characteristics of the person, including age, sex, geography, Indigenous status and whether they were sampled in discrete Indigenous communities. The PREG estimator also allows for differences in the reporting of a person's characteristics (e.g. age, Indigenous status) between PES and Census.

Another outcome of the review was to estimate the numbers of persons in dwellings who were imputed in the Census, as these persons had previously been excluded from the scope of the PES. An improved approach for adjusting for dwellings found in the Census but not responding in the PES was also introduced. These and other changes had a larger impact on the final estimates than the change in estimator.

Further information on the changes in estimation methodology can be found in *Research Paper: An Estimating Equation Approach to Census Coverage Adjustment* (cat. no. 1351.0.55.019), which is available from the ABS website. This research paper includes comparisons with the previous estimators.

3. Plans for the 2011 PES

Apart from the new and ongoing innovations outlined in this report, the 2011 PES is largely consistent with the 2006 survey. While the development cycle has been somewhat longer than that of the 2006 cycle, the enumeration, processing and dissemination phases will closely resemble what was done in 2006.

More detailed information is available in *Information Paper: Measuring Net Undercount in the 2011 Population Census, Australia, 2011* (cat. no. 2940.0.55.001), which was released on the ABS website in July 2011.

Appendix 1: PES resources available on the ABS website

Research Paper: An Estimating Equation Approach to Census Coverage Adjustment (cat. no. 1351.0.55.019)

Census Dictionary, 2006 (cat. no. 2901.0)

Census of Population and Housing - Undercount, Australia, 2006 (Cat. no. 2940.0)

Census of Population and Housing - Details of Undercount, Australia, August 2006 (cat. no. 2940.0)

Information Paper: Measuring net undercount in the 2006 Population Census, Australia, 2006 (cat. no. 2940.0.55.001)

Information Paper: Measuring Net Undercount in the 2006 Population Census, Australia, 2007 (cat. no. 2940.0.55.001)

Information Paper: Measuring Net Undercount in the 2011 Population Census, Australia, 2011 (cat. no. 2940.0.55.001)

Information Paper: Population Concepts, 2008 (cat. no. 3107.0.55.006)

Population Estimates: Concepts, Sources and Methods, 2009 (cat. no. 3228.0.55.001)

Appendix 2: ABS contacts

Statistical area	**Contact officer**	**Phone and email contacts**
Post Enumeration Survey	Bjorn Jarvis, Director, Post Enumeration Survey	+61 2 6252 6552 or bjorn.jarvis@abs.gov.au
Post Enumeration Survey Methodology	Philip Bell, Assistant Director, Methodology Development Unit	+61 8 8237 7304 or philip.bell@abs.gov.au
Estimated Resident Population	Patrick Corr, Director, Demography	+61 2 6252 5722, patrick.corr@abs.gov.au or demography@abs.gov.au
Census of Population and Housing	Paul Lowe, Assistant Statistician, Population Census Branch	+61 2 6252 6863 or paul.lowe@abs.gov.au

Appendix 3: PES results in ERP rebasing

The estimated resident population (ERP) is the official estimate of the population of Australia calculated by the ABS at quarterly intervals each year. Following each Census, the ERP for 30 June of the Census year is rebased using Census data, adjusted by PES data and other information.

Initially, a count of Australian residents in Australia on Census night is obtained by excluding all overseas visitors enumerated in the Census. This Census count is then augmented to produce an estimate which accounts for net undercount. Net undercount is the net result of people who were missed by the Census, and people who were counted more than once in the Census (or counted at all when they should not have been). The data source for this estimate of net undercount is the PES.

ERP calculations then make an allowance for Australian residents who were temporarily absent from Australia on Census night. This estimate is made using data from completed passenger cards; and visa and passport information obtained from the Department of Immigration and Citizenship. These people are added into the Australian resident population.

The final step in calculating ERP is to backdate it to 30 June of the Census year. This is achieved by adding the deaths and subtracting the births and net overseas migration which occurred between 1 July and the Census date. Information on the calculation of rebased ERP for 30 June 2006 based on the 2006 Census was published in *Australian Demographic Statistics, December quarter 2006* (cat. no. 3101.0).

While the PES identifies people and dwellings missed in the Census, the extent to which some people are missed in both the Census and the PES may not be fully accounted for in PES estimation, resulting in a possible 'correlation bias'. As in any survey, the PES is also subject to sampling and non-sampling error. To offset the impact of correlation bias and survey error, population estimates derived from the PES are further refined using demographic adjustments based on independent sources of population information: birth and death registrations, Medicare enrolment numbers, and the estimated resident population based on the previous Census carried forward.