Australian Bureau of Statistics

Information Paper

Measuring Net Undercount in the 2011 Population Census

Australia

2011

Information Paper

Measuring Net Undercount in the 2011 Population Census

Australia

2011

Brian Pink
Australian Statistician

INQUIRIES

- For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070 or Bjorn Jarvis on Canberra (02) 6252 6552, email <bjorn.jarvis@abs.gov.au>.

CONTENTS

ABBREVIATIONS

. . . .

'000	thousand
ABS	Australian Bureau of Statistics
ADL	automated data linking
ARA	any responsible adult
ASGC	Australian Standard Geographical Classification
ASGS	Australian Statistical Geography Standard
CAI	computer assisted interviewing
CARDS	collect, analyse, reduce, de-duplicate and systematise
CD	collection district
CLW	collector workload
ERP	estimated resident population
FEBRL	Freely Extensible Biomedical Record Linking
ICF	Indigenous Community Frame
MSS	Match and Search System
NPD	non-private dwelling
PAPI	pen-and-paper interview
PD	private dwelling
PES	Census Post Enumeration Survey
PREG	prediction regression
QA	quality assurance
SA1	Statistical Area Level 1
SE	standard error

INTRODUCTION

THE CENSUS OF
POPULATION AND
HOUSING

The Population Census is a valuable data source for estimating the size and geographic distribution of the Australian population, and for analysing the major demographic, social and economic characteristics of the population, particularly for small geographic regions and other small sub-populations. It provides statistics for decision-making by governments, businesses, community organisations and individuals. A Census also provides a base for post-censal population estimates and projections, which assist in planning and policy-making at the national, state and local levels.

A Census is conducted in Australia every five years. The next Census will be held on the night of Tuesday 9 August 2011. On Census night, every person present in Australia, excluding foreign diplomats and their families, should be included on a Census form at the place where they stayed.

NET UNDERCOUNT

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 counted more than once. Some of the reasons why people may be missed (i.e. undercounted) include:

- they were travelling and were difficult to contact;
- they mistakenly thought they were counted elsewhere;
- there was insufficient space on the Census form in the household where they were staying and they did not obtain additional forms;
- the person completing the form thought that, for example, young babies, the elderly or visitors should not be included;
- they did not wish to be included due to concerns about the confidentiality of information or a more general reluctance to participate;
- the dwelling in which they were located was missed because it was difficult to find (e.g. in a remote or non-residential area); and
- the dwelling in which they were located was mistakenly classed as unoccupied.

Some of the reasons why people are counted more than once (i.e. overcounted) include:

- they were included on the Census form at the dwelling where they usually live, even though they stayed and were counted elsewhere on Census night; and
- they were overseas on Census night and so should not have been counted at all, but were included on the Census form at the dwelling where they usually live.

While every effort is made to eliminate these potential causes of error, some undercount and overcount will occur. Usually more people are missed than overcounted in Australia, so the Census count of the population is fewer than the true population. This difference is called net undercount.

Rates of net undercount can vary significantly for different population groups depending on factors such as sex, age, ethnicity (including Indigenous origin) and geographic location. The ABS obtains estimates of net undercount using information collected in a post-enumeration survey conducted immediately following the Census. Estimates of net undercount are used to:

- derive an estimate of the resident population for 30 June of the Census year;
- provide users with an assessment of the completeness of Census counts, allowing them to take this into account when using Census information; and

NET UNDERCOUNT *continued*

- evaluate the effectiveness of Census collection procedures so that improvements can be made for future Censuses.

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

PURPOSE OF THIS PAPER

The purpose of this paper is to set out the methods to be used in the conduct of the 2011 Census Post Enumeration Survey.

OVERVIEW

In Australia, estimates of net undercount in the Census are based on the results of the Census Post Enumeration Survey (PES). The purpose of the PES is to determine how many people were missed in the Census and how many were counted more than once (and how many times these people were counted).

The PES is a household survey conducted by specially trained interviewers starting about three weeks after Census night. This is a different collection methodology to the Census, where most forms are self-completed. A major advantage of interviewer-administered questionnaires is that people can be provided with assistance if they are uncertain about the meaning of any questions.

For each dwelling selected in the PES, the any responsible adult (ARA) methodology is used. This generally involves a single person in the dwelling responding on behalf of all persons present or usually resident. In addition to obtaining basic demographic information, questions are asked about each person's usual residence, location on Census night, and any other addresses where they might have been counted in the Census.

Using this address information, the corresponding Census forms are examined at the Census data processing centre to confirm how many times each respondent in the PES was counted in the Census. The results of this process are then combined and weighted to produce an estimate of net undercount in the Census.

In the 2006 PES, a sample of about 39,000 private dwellings (PDs) was selected (about half of 1% of all dwellings in Australia). The total number of people included in the PES was about 86,000. The sampling fraction varied between states and territories, with the smaller states and territories having higher sampling fractions to ensure adequate reliability of estimates.

The 2011 PES sample size is slightly larger than that for 2006, with around 41,000 dwellings selected. Additional information on the 2011 sample is provided in the Sample Design section of this paper.

INDEPENDENCE FROM THE CENSUS

One of the purposes of the PES is to provide an independent check on Census coverage. There are two aspects to this independence: population independence and operational independence.

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 than operational independence, though 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.

INDEPENDENCE FROM
THE CENSUS *continued*

Operational independence requires that Census operations do not influence the PES in any way, and vice versa. Steps are taken to maintain the operational independence of the PES from the Census at every stage of the survey, including enumeration, processing and administration. These steps include:

- selecting the PES sample from an independent sample frame;
- using separate office staff in the PES and Census where possible;
- ensuring the PES interviewers are not employed as Census field staff in the same area, and vice versa; and
- maintaining the confidentiality of the PES sample so that Census field and office staff are not aware which areas are included in the PES.

Differences between Census and the PES

Householders are asked the PES questions 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 the 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.

Late census forms

For some people who have not returned a Census form, contact from the ABS following selection in the PES acts as a reminder and possible motivator to return a completed Census form. These late returns, if not identified, would result in the PES sample having a higher proportion of Census response than in the overall population. To protect against this, all Census forms received after the start of PES field work are deemed 'late' and treated differently in PES estimation.

Correlation bias

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.

CHANGES SINCE THE
2006 PES

As a result of the continual improvements made in PES processing methodology, an increased focus on Indigenous population estimates, and geographical changes that have been implemented by the Census, the following three key changes have been implemented in the 2011 PES:

- Automated Data Linking has been introduced, to complement the existing PES Match and Search System;
- the PES sample has been expanded, to incorporate an increased focus on enumeration of discrete Indigenous communities and focused sampling in parts of the mainstream sample for which there is a relatively high proportion of Indigenous persons; and
- the Australian Statistical Geography Standard (ASGS) has replaced the Australian Standard Geographical Classification (ASGC) as the framework for PES geography.

Further information on each of these changes is outlined later in this information paper.

SCOPE

The scope of the Census is every person present in Australia on Census night (with the exception of foreign diplomats and their families). Ideally the PES would sample from all people who were or should have been in the scope of the Census, but for practical reasons there are a number of areas, dwellings and people excluded or not able to be covered by the PES. Of the people present in Australia at the time of the PES, the following are not included:

- foreign diplomats and their families;
- people in non-private dwellings (NPDs) such as hotels, motels, hospitals and other institutions;
- homeless people (as the sample selected in the PES is based on the selection of dwellings);
- overseas visitors who were not in Australia on 9 August 2011 (Census night);
- babies born after 9 August 2011; and
- people in Cocos (Keeling) Islands, Christmas Island, Australian Antarctic Territory and Jervis Bay Territory.

The PES does not obtain information about people who died between Census and the PES, however it does obtain information about Australian residents who are overseas during the PES enumeration period and who departed after the Census, provided that they usually live with people remaining in Australia.

Remote areas and discrete Indigenous communities

Remote areas and discrete Indigenous communities were first included in the scope of the PES in 2006. Prior to this, they were excluded from the PES because of operational issues associated with enumerating them. In addition, it was considered difficult in the past to implement procedures to ensure that independence of the PES from the Census was not compromised by using the same local contacts for both Census and PES.

In 2006, the risk to statistical independence was effectively managed through interviewer training and the procedures put in place for field staff. As a result, enumeration of discrete Indigenous communities will again occur in 2011, between the end of August and early October 2011.

Non-private dwellings

Non-private dwellings are establishments which provide predominantly short-term accommodation for communal or group living, and often provide common eating facilities. They include hotels, motels, hostels, hospitals, religious institutions providing accommodation, educational institutions providing accommodation, prisons, boarding houses and short-stay caravan parks. Non-private dwellings each comprise a number of dwelling units. About 2% of the Australian population live in non-private dwellings.

Non-private dwellings have previously been excluded from the PES because:

- a high number of these dwelling units are found to be vacant;
- many of the people who are contacted in these dwelling units are not in scope of the PES, or they usually live in a private dwelling and the PES coverage rules give them a chance of selection at their usual residence;
- the quality of information is often poor, especially in institutions, where information cannot be provided by individuals themselves and so is collected from administrative lists or from staff;

■ enumeration suffers from a lack of independence from the Census, because the administrative lists used by interviewers to collect data about institutionalised people are the same as those used by Census staff; and

■ the cost of enumerating them is relatively high compared to private dwellings, because of the additional time and effort required.

An investigation was conducted into the possible inclusion of non-private dwellings in the scope of the 2006 PES. The investigation found that non-private dwellings were likely to have high levels of sample loss and non-response, which meant their inclusion in the PES would not be cost-effective. Non-private dwellings were therefore excluded from the scope of the 2006 PES and will also be excluded in 2011. While this is expected to have minimal impact on overall estimates of net undercount, the inclusion of non-private dwellings will be investigated again prior to the 2016 PES.

COVERAGE RULES

The PES coverage rules are designed to give each person in the population a single chance of selection in the survey by associating each person with one and only one dwelling. The coverage rules are implemented by asking a series of questions in the PES interview. The questions cover topics such as where each person usually lives and whether they will be staying at the PES dwelling on a particular reference night.

Structuring the questionnaire so that each person has a single chance of selection in the PES ensures that data from the sample can produce estimates that are representative of the whole population.

SAMPLE DESIGN

In the PES, private dwellings (houses, flats, etc.) and discrete Indigenous communities are separately identified and sampled. In total, about 35,000 dwellings are expected to respond to the PES, including approximately 500 dwellings selected from 28 discrete Indigenous communities.

Table 1 shows the expected number of fully responding households in the private dwelling sample and discrete Indigenous community sample for each State and Territory. It is important to note that, due to sample loss and other non-response, the number of dwellings selected in the sample will be greater than shown below. Based on the results of other ABS surveys, the rate of sample loss is expected to vary between different components of the sample, which is taken into account.

TABLE 1: EXPECTED NUMBER OF FULLY RESPONDING HOUSEHOLDS

	Private dwellings	ICF dwellings(a)	Total
	no.	no.	no.
New South Wales	7 618	0	7 618
Victoria	6 450	0	6 450
Queensland	6 394	48 (3)	6 442
South Australia	3 694	18 (2)	3 712
Western Australia	4 410	33 (5)	4 443
Tasmania	1 971	0	1 971
Northern Territory	2 460	382 (18)	2 842
Australian Capital Territory	1 349	0	1 349
Australia	**34 346**	**481 (28)**	**34 827**

(a) The number in brackets is the number of discrete Indigenous communities from which the ICF dwellings will be selected.

This sample size is expected to be sufficient to provide estimates of the net undercount rate with the following standard errors (SEs):

- males and females at Australia level - less than 0.2 percentage points;
- males and females by age group at Australia level - less than 0.6 percentage points;
- persons by state/territory - less than 0.6 percentage points (except for the Northern Territory, where it is expected to be less than 1.5 percentage points); and
- Indigenous people at Australia level - less than 2.5 percentage points.

Mainstream private dwelling sample

The mainstream sample of private dwellings is derived from the 2006 Monthly Population Survey Parallel Sample frame, which is based on the Statistical Division and Subdivision structure of the Australian Standard Geographical Classification (ASGC).

This sampling framework first divides Australia into 100 geographical areas. These areas are then divided into strata according to population density, remoteness and growth, then:

- in the first stage of selection, a sample of Census collection districts (CDs) is selected (systematically, with probability proportional to size) to represent each stratum;
- in the second stage of selection, each selected CD is divided into smaller areas called blocks, one of which is selected (again systematically, with probability proportional to size); and
- in the third stage, a sample of dwellings in the selected block is taken using systematic equal probability sampling.

In less populated areas, an additional stage precedes the selection of CDs to ensure that the sample is not too geographically spread (as that would lead to unacceptable enumeration costs).

The PES sample departs from the Parallel Sample by selecting extra dwellings, and sometimes extra CDs, in areas with higher proportions of Indigenous households. This measure has been introduced in the 2011 PES to improve estimates of Indigenous undercount.

Discrete Indigenous communities sample

CDs with an identified Indigenous population of greater than 75% are classified differently from the remaining population for both operational and sampling reasons. These CDs and the Indigenous communities within them form the Indigenous Communities frame (ICF).

For the purposes of sampling, discrete Indigenous communities are grouped into 'sets' comprising main communities and their associated outstations. The selection of main communities is undertaken with probability proportional to the size of the set. The aim is to select as representative a sample as possible while also considering cost constraints, reasonable interviewer workloads and expected sample size.

If a community is selected in the sample, selection of dwellings within the community follows the same procedure as for the selection of private dwellings within selected blocks in the non-ICF component of the sample. That is, an interviewer compiles a list of all the dwellings within the community and a sample is taken using systematic equal probability sampling.

A selection of outstations associated with each selected main community is also included in the sample. Each outstation has an equal chance of selection and, once selected, all dwellings within the outstation are enumerated.

Sampling strategies to improve Indigenous population estimates

Initial attempts to improve the quality of Indigenous undercount information occurred in 2006, when remote areas and discrete Indigenous communities were enumerated for the first time. The 2011 PES strives to further improve the estimates of Indigenous undercount by increasing the number of discrete Indigenous communities selected, as well as the total number of dwellings selected within the community sample. This increase in sample is expected to decrease the sampling error on estimates of Indigenous undercount.

In addition to the increase in the discrete Indigenous communities sample, the 2011 PES will increase the number of selected dwellings from areas outside the discrete Indigenous communities that have been identified as having a relatively high proportion of Indigenous persons. This focused sampling of additional areas within the mainstream sample is a key sample design change from 2006.

The focused sampling approach will involve a combination of extra dwellings selected from within the parallel sample, as well as some additional selections from CDs outside the parallel sample.

This focus on obtaining an increase in Indigenous respondents within the mainstream sample is of particular importance, given approximately three-quarters of the Indigenous population usually reside outside of communities.

The impact of the transition to the ASGS

With the geographic classification used in the Census changing from the Australian Standard Geographical Classification (ASGC) to the Australian Statistical Geography Standard (ASGS), there will be a mixture of geographies used in the 2011 PES, with different geographic units used at different stages of the survey cycle. This development presents challenges for the ABS to manage in relation to the sample design.

The impact of the transition to the ASGS continued

As in 2006, the 2011 PES sample design will be structured by CDs. This is in contrast to the 2011 PES field operations, processing and output, which will use geographical units from the ASGS, primarily Statistical Area Level 1 (SA1). To overcome this challenge, the PES sample will be recoded to the ASGS geographies prior to the start of PES enumeration.

The second challenge posed is the spatial definition of discrete Indigenous communities, some of which may be different between the two geography frameworks. This will be managed by a mapping exercise to compare the two geographies for the 28 communities selected in the PES sample.

COLLECTION METHODS
Field procedures

Various strategies have been devised for the enumeration of the PES. Where possible, standard procedures are used when enumerating private dwellings and discrete Indigenous communities. However, these procedures are modified where necessary in discrete Indigenous communities to take account of language and cultural issues. While question wording and collection methodology may be modified in some cases, attempts will be made to ensure underlying concepts remain the same across both components of the sample.

Mainstream sample

Specially trained PES interviewers will collect data through face-to-face interviews starting around three weeks after Census night. All mainstream dwellings will be enumerated using Computer Assisted Interviewing (CAI). Interviews will be conducted with any responsible adult of the household who will be asked to respond on behalf of all household members.

Discrete Indigenous communities sample

Enumeration in discrete Indigenous communities will take place as soon as practicable after Census enumeration, while also ensuring that the PES enumeration period is as short as possible. This will limit the potential for people to be enumerated in the PES in more than one community. PES enumeration in these areas may start as early as 21 August. It is expected that PES enumeration in discrete Indigenous communities will be completed by early October, although the completion of the PES will be dependent on the completion of the Census in these areas.

In discrete Indigenous communities the PES will be conducted by specially trained ABS staff with the assistance of facilitators recruited from within the community. The facilitator's role is to assist in establishing rapport with respondents, to assist the interviewer in identifying residents of the selected households, and to interpret where necessary.

To preserve the independence of the Census and the PES, efforts will be made to recruit facilitators who were not involved as Census interviewers. Where this is not possible, Census interviewers will only act as PES facilitators at dwellings where they did not interview during the Census. If there is no option but to use a facilitator who conducted a Census interview at the same dwelling, the PES interviewer will ask the facilitator to introduce the interviewer to the household, but otherwise to take no part in the interview. Alternatively, where acceptable to the community, the PES interviewer may enumerate dwellings without the assistance of a facilitator.

Discrete Indigenous communities sample continued

For discrete Indigenous communities, the primary collection method will be a customised CAI instrument. However, due to practical considerations, a pen-and-paper interview (PAPI) questionnaire will be available to interviewers, which provides them with an effective degree of flexibility. The data collected from PAPI forms will be transcribed by the interviewer into the CAI instrument.

Census follow-up

In each Census there are always dwellings for which Census forms have not been returned within the required timeframe. For this reason, intensive Census follow-up procedures are employed at the end of Census collection.

In 2011, mainstream Census enumeration will finish on 28 August, with post collection follow-up to occur between 29 August and 4 September. To allow sufficient time for the Census follow-up and to minimise potential overlap with these activities, PES enumeration will begin on 4 September and will continue until 30 September.

For discrete Indigenous communities, Census collection will start on 29 July and end on 28 August. To avoid any overlap, PES enumeration will only begin in these communities once it has been established that all Census field activities have been completed. As a result, PES enumeration of the discrete Indigenous communities sample will be staggered, to align with the completion of Census field operations in each of the communities.

As in previous PESs, special procedures will be implemented for Census forms received after the start of the 2011 PES field work. These procedures are needed to preserve the independence of the Census and the PES as some people may be prompted to return their Census forms following receipt of the PES primary approach letter or the arrival of the PES interviewer. Any Census form received after the start of PES field work will be flagged as a 'late return'. The treatment of late returns is explained in the Estimation section of this paper.

Census and PES staffing

The independence of the Census and the PES will be strengthened by strategies designed to minimise the overlap of Census collectors and PES interviewers. These include:
- interviewers wanting to apply for District Manager or Area Supervisor positions in the Census must seek an exemption from the interviewer panel for the PES;
- PES interviewers may also work as Census collectors but they must enumerate a different area;
- any person who has worked in an area of Census processing related to dwelling or population counts may not also work on PES processing; and
- any person who has been involved in Census field collection in any capacity may not also work on PES processing.

QUESTIONNAIRE DESIGN
Private dwelling questionnaire

The PES questionnaire collects personal details (name, sex, date of birth, age, relationship in household, marital status, country of birth and Indigenous status) to facilitate matching of PES person records to Census person records and to allow accurate undercount estimates to be generated for age and sex categories, and indigenous status.

Private dwelling
questionnaire continued

The PES also asks respondents:

- whether they were included on a Census form (and if so, where);
- whether they could have been included on a Census form at other addresses (and if so, where); and
- where they stayed on Census night.

The different addresses collected in the PES are used to search Census records to determine the number of times each PES respondent was included in the Census. Visitors to households included in the PES are also asked for their address of usual residence. In addition to questions relating to Census night, the PES collects a small amount of information on dwelling tenure and structure.

Discrete Indigenous
community questionnaire

As in 2006, a specially designed PAPI questionnaire and CAI instrument will be used in discrete Indigenous communities. A number of questions in the mainstream PES questionnaire are not considered applicable to people living in discrete Indigenous communities and have been adapted to ensure that information is collected in the most culturally appropriate manner. For example, it is more common for Indigenous persons in the community to have more than one name by which they may be known. This tailored questionnaire ensures that sufficient information is collected to effectively match persons, taking into account the potentially different nature of matching.

The PAPI questionnaire will be available for use for all people, whether Indigenous or non-Indigenous, selected in the PES in discrete Indigenous communities. A community-level questionnaire will also be asked of the community contact or council officer of discrete Indigenous communities. Information collected, such as whether any significant event (for example a sports carnival) may have occurred at the time of the Census, is expected to help with respondent recall and assist in the completion of the individual questionnaires.

DATA CAPTURE

Data entered into the CAI instrument will be converted into files containing dwelling-level information (address, number of people, response status, etc.), person-level information (name, age, sex, marital status, etc.), search addresses (where the person may have been included on a Census form), and interviewer comments (e.g., to assist in dwelling and person matching). The files will be loaded into a database from which all PES processing systems draw information.

ADDRESS CODING

Address information is essential for matching 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. Prior to processing PES data, each enumeration address will have a Collector Workload (CLW) and a Statistical Area Level 1 (SA1) associated with it, drawn directly from the re-coded sample. Each 'search address', that is, an address where a person may have been included on a Census form for which the PES can search, will be coded with a CLW and SA1, using the AddressCoder@ABS application.

ADDRESS CODING *continued*

In 2011, the coding of addresses has become even more important to the processing of the PES than in previous cycles. The introduction of Automated Data Linking (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 SA1 to be identified. In addition to this coding, attention must also be given to ensuring the accuracy of other address elements (e.g., street names, suburbs and postcodes).

The CLW is important for subsequent match and search processing as it is the default starting point for dwelling and person searching.

AUTOMATED DATA LINKING

ADL is being introduced into the PES in 2011 for the first time. 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 data linking processes to determine possible links between Census and PES data before any clerical matching process has begun.

ADL employs probabilistic data linking techniques, using a range of personal and address characteristics, to evaluate the likelihood that a PES record and a 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 gains in matching effectiveness and efficiency provided by ADL include:
- the ability to conduct a more comprehensive search for PES respondents than was possible from previous clerical matching processes;
- the ability to locate PES respondents at Census night addresses that were not identified on the PES; and
- a reduced requirement for clerical matching resources.

ADL uses Freely Extensible Biomedical Record Linking (FEBRL) software developed at the Australian National University. A number of different linking runs will be used to compare PES and Census records, each focused on a slightly different combination of name, addresses and demographic variables.

Potential links are assessed by assigning weights that reflect the level of agreement on selected data items from the two records. Large positive weights indicate probable matches, while large negative weights are observed for probable non-matches.

Important to the effective use of ADL is a series of processes that are run after ADL output is obtained. The Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS) process identifies and rates the most plausible links from each ADL run for all PES respondents. The process then combines the links from all ADL runs and removes any duplicates.

The CARDS process concludes by creating dwelling links, as the PES Match and Search System (MSS) operates on a dwelling basis. Person-level links are grouped together within dwellings, and dwelling-level links are derived and rated on the basis of the numbers of persons recorded in the PES and Census dwellings and the number and quality of person-level links identified between the two dwellings.

AUTOMATED DATA
LINKING *continued*

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 also been necessary to ensure that the relative strengths of both ADL and the MSS are fully realised.

THE PES MATCH AND
SEARCH SYSTEM

The MSS is the main PES clerical review facility, built specifically for PES processing in 2006. The MSS allows processing staff to search, view, compare, and record matches between PES and Census data. PES processing staff use the MSS to record matches of dwellings and people between PES and Census, and to search for people on Census forms at alternative addresses provided in the PES. In 2011, it will also be used to assure the quality of ADL processing.

The initial phase of MSS processing involves confirming whether the ADL-identified dwelling link is correct. Once a dwelling link is confirmed, the Census person records for that dwelling are compared with the PES person records. The information compared includes name, sex, date of birth, age, marital status, Indigenous status and country of birth. The extent to which each of these items of information is the same in both the PES and the Census determines the ADL match status of the pair and the level of match.

Once all ADL links have been reviewed, the final phase in MSS processing is to conduct an intensive clerical search for persons not matched as a result of ADL information. This is done by searching CLWs (and often neighbouring CLWs) for addresses provided by respondents during the PES interview, in order to locate possible Census forms.

MSS quality assurance

To ensure the accuracy of MSS processing, quality assurance (QA) procedures are used in the match and search process whereby all PES records are processed a second time by a different operator. Where the initial and QA processing outcomes correspond, the initial match status is accepted. Where there is a discrepancy between the initial match status and the QA match status (either on dwelling match or person match code), the records are flagged for adjudication by a senior officer who reviews all information and determines which is correct. If both initial and QA records are deemed to be inaccurate, the adjudicator reprocesses the record.

*MSS processing for
discrete Indigenous
communities*

MSS processing for discrete Indigenous communities will be similar to that for the mainstream component, with an additional approach that accounts for the fact that some Indigenous Australians are associated with more than one dwelling and move between these dwellings on a regular, seasonal or random basis (Kinfu, 2005). This means that a number of people may not be staying in the same dwelling during the PES that they were during the Census. For this reason, the matching process in discrete Indigenous communities involves searching the whole community for a person match, rather than just searching within a single dwelling.

Person matching in discrete Indigenous communities will use the same rules for determining a match as in the mainstream component, but will use up to two alternative names by which the person is known when matching on name.

ESTIMATION

The PES interview process determines whether each person in the sample should have been counted in the Census, and the linking and matching process determines how many times each person was actually counted in the Census. The PES estimation process combines and weights these data to produce an estimate of the difference between the number of people who should have been counted in the Census and the number of people who were actually counted. This difference is the estimate of net undercount.

Broadly speaking, the PES estimate of the number of people who should have been counted in the Census is obtained by adding this estimate of net undercount to the number of people actually counted in the Census.

Overview of weighting for the PES

Weighting is the process of adjusting results from a sample survey to infer results for the total in-scope population. The weight can be considered an indication of how many population units are represented by the sample unit.

Essentially, the PES weighting methodology produces an estimate of the number of people who should have been counted in the Census. Weights are given to people that were counted in the Census and to people that were missed in the Census. The sum of weights (i.e. the people who were counted plus the people who should have been counted but were not) becomes the estimate of all people that should have been counted.

Weighting begins with the initial dwelling weight, which is essentially equal to the inverse of the probability that a dwelling was selected in the PES sample. Then dwelling weighting adjustments are made to compensate for the different rates of non-response in different dwelling types and areas of Australia. Person weighting adjustments are then applied to allow for undercoverage and non-response in the PES. Larger adjustments are made for categories of people (such as young adult males) who are harder to contact and interview successfully in the PES.

Dwelling weights in the PES

Each dwelling in the PES sample is given a 'dwelling weight', so that the PES sample represents all private dwellings in Australia. The PES sample is designed to ensure each private dwelling in a state or territory has a fixed and non-zero probability of selection. The inverse of this probability is the dwelling's selection weight. In practice, certain types of dwellings are more likely to be missed in the PES. Adjustment for this is made according to a dwelling's post-stratum, where dwellings are assigned to a particular post-stratum based on the following variables:

- Census response category at the start of the PES enumeration period (responding, non-responding, unoccupied on Census night);
- dwelling structure (e.g., flat, house, caravan etc.); and
- region (six states and NT divided into capital city and balance, plus ACT, giving 15 regions).

To obtain initial dwelling weights, a dwelling's selection weight is multiplied by a factor calculated as the Census count for the dwelling's post-stratum divided by the total of the selection weights for PES dwellings in that post-stratum that were counted in the Census.

These initial weights apply to all dwellings in the PES sample, even those that did not result in a PES response. Responding dwellings in the PES are further weighted up to represent non-responding dwellings which are deemed to be occupied.

Dwelling weights in the PES continued

For example, suppose that the PES selected 22 dwellings in a post-stratum, each with a probability of 0.2, and that 18 of them were counted in the Census. The selection weight for each dwelling is 5 (i.e. 1 divided by 0.2) with the total selection weight of the dwellings counted in the Census being 90 (i.e. 18 multiplied by 5). If 99 dwellings were counted in the Census, the initial dwelling weight for all these dwellings becomes 5 multiplied by 99/90, which is equal to 5.5. Furthermore, if 5 of the selected dwellings fail to respond in the PES, then the cumulative weight of these dwellings will be redistributed to the other dwellings in the sample in order for the sample to reflect the independently estimated distribution of the population. In this case, the weight for the 17 responding dwellings becomes 7.1 (i.e. 5.5 multiplied by 22/17).

Person weighting

In estimation, the weights of persons responding in PES are adjusted so that when summed across those persons counted in the Census the totals correspond to the actual Census counts within a number of benchmark categories. The benchmark categories are based on personal characteristics including age, sex, state/territory, country of birth and Indigenous status.

Estimates of the number of people who should have been counted in the Census based on the dwelling weights would only represent the population of people who were in private dwellings at the time of PES. That is, they would underestimate the private dwelling population at the time of the Census because some people in private dwellings on Census night will be in non-private dwellings, overseas, or may even be deceased at PES time. Such estimates would also not represent people living in non-private dwellings. To represent all in-scope people on Census night requires adjusting the dwelling weights to give a person weight.

The initial person weight adjustments are chosen to ensure that the PES estimates of people counted in private dwellings (other than late return or imputed dwellings) in a set of benchmark categories match the actual Census counts for these categories. The variables used to form these benchmark categories are region, sex, age (by 16 age groups), country of birth, marital status, Indigenous status, and whether sampled in an ICF dwelling. The weight adjustment applied to a person depends only on the information reported in the PES.

As a final step in weight adjustment, the initial person weights are adjusted so that the PES estimates represent people in non-private dwellings as well as private dwellings. This final step uses region, age and sex only, as information on other items is not reliable for non-private dwellings.

Intuitively, a 'good' set of weights for the PES should ensure that if the PES were used to estimate the actual Census count, the PES would get the 'right' answer. The above step makes sure that this happens. Technically this is also a desirable property for a set of PES weights to have, since there is a very strong relationship between the actual Census count, and the count that the Census should have made.

For example, consider the benchmark category consisting of ACT males aged 25 to 29 years in Census dwellings (that were not imputed or late returns). Suppose there are 120 of these records in the sample, 100 of whom were counted in the Census and 20 who were not. Also suppose the dwelling weights for each were 10, with a Census total of 2000. The dwelling weight will therefore be increased to give a person weight of 20, so

Person weighting continued

that the 100 records will add up to the Census count of 2000. However, the 20 records that were not counted will also have an equivalent weight of 20, giving us an estimate of 400 people in this category who were missed in the Census.

In 2011, the ABS will again use the Prediction Regression (PREG) estimator, which was developed and used in 2006 as an estimator that extends the Dual System Estimator approach to account for overlapping benchmark categories and the situation where people give different responses between PES and Census. A detailed description of the PREG estimator can be found in *Research Paper: An Estimating Equation Approach to Census Coverage Adjustment*, *May 2007* (cat. no. 1351.0.55.019).

Estimating the number of people in Census late return and imputed dwellings

For some people who have not returned a Census form, contact from the ABS following selection in the PES acts as a reminder and possible motivator to return a completed Census form. These late returns, if not identified, would result in the PES sample having a higher proportion of Census response than in the overall population. To protect against this, all Census forms received after the start of PES field work are deemed 'late'. For the purpose of PES estimation, the dwellings from which these forms are received are treated as though they had not been contacted in the Census and are classified to the 'non-contact sector' of the Census.

Comparability of sub-state undercount estimates over time

Estimates of net undercount below the state level are important to the production of sub-state population estimates and the evaluation of Census coverage on a more localised basis. In 2006, these sub-state estimates incorporated a capital city and balance of state/territory split, which will again be included for the 2011 results. The comparability of these estimates over time will be affected by the transition from the ASGC to ASGS. While the sub-state split is similar, it is not directly comparable.

HOW NET UNDERCOUNT CONTRIBUTES TO ERP

BACKGROUND

The estimated resident population (ERP) is the official estimate of the population calculated by the ABS at quarterly intervals each year.

Accurate estimates of the Australian resident population are essential for the allocation of seats in the Federal House of Representatives as well as distributing Commonwealth funding to states and territories. These estimates are also used for demographic, social and economic studies. The validity of these population estimates depends on a number of factors, one of which is the accurate measurement of net undercount in the Census.

Following each Census, the ERP for 30 June of the Census year is rebased using the Census data, adjusted by the PES.

CALCULATION OF ERP

Initially, a count of Australian residents who were in Australia on Census night is obtained from the Census. This Census count is adjusted on the basis of net undercount, to account for people being missed and people being counted more than once in the Census. The basis of the estimate of net undercount is the PES.

ERP calculations then make an adjustment for Australian residents who were temporarily absent from Australia on Census night. An estimate of this number is made using data from completed passenger cards, visa and passport information obtained from the Department of Immigration and Citizenship and 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. Table 2 shows the components used to calculate ERP for Australia from the 2006 Census.

TABLE 2. COMPONENTS OF ERP, AUSTRALIA, 2006

	Persons
Components of ERP (a)	'000
Census count, actual location	20 061.6
less Overseas visitors	-206.4
Census count, place of usual residence	19 855.2
plus Net undercount (b)	534.4
plus Residents temporarily overseas	345.2
ERP (a) as at 8 August 2006	20 734.8
less Births (1 July to 8 August 2006)	-29.5
plus Deaths (1 July to 8 August 2006)	15.5
less Net overseas migration (1 July to 8 August 2006)	-19.4
ERP (a) as at 30 June 2006	**20 701.4**

(a) Preliminary ERP.
(b) Includes demographic adjustments.

Information on the calculation of the ERP for 30 June 2011 based on the 2011 Census will be reported in *Australian Demographic Statistics, December quarter 2011* (cat. no. 3101.0), due for release on 21 June 2012. Additional information can also be found in *Population Estimates: Concepts, Sources and Methods, 2009* (cat. no. 3228.0.55.001).

Demographic adjustments

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 reflected 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 three sources of independent population information: the National Demographic Data Bank, Medicare enrolment numbers, and the estimated resident population based on the previous Census. These sources have different strengths and weaknesses, but where the data are considered to be most reliable, they are used for comparison with PES adjusted age and sex population distributions, and for possible minor adjustments to population estimates.

The National Demographic Data Bank is a population database maintained by the ABS using administrative data (notably births, deaths, and overseas arrivals and departures). The database is independent of Census data and contains population data back to 1925. For the 2006 PES, these data were considered to measure age-sex totals well up to about age 35, after which there were some concerns about pre-1970 international migration data. Sex ratios derived from these data are considered most reliable for ages under 28 years.

Enrolment data from Medicare (the Australian government health rebate system) are considered a good source for calculating sex ratios, but less reliable for age-sex totals. Age-sex totals are least reliable among the older ages where people may remain enrolled in Medicare after their death until Medicare Australia is notified and the record updated.

GLOSSARY

AddressCoder@ABS	A web service that assigns a geographic classification to an address or a list of addresses and is used for linking and matching activities.
Automated Data Linking (ADL)	Automated linking processes used to determine possible links between Census and PES data, before any clerical matching process has begun. It employs a probabilistic linking method that uses a range of personal and address characteristics to evaluate the likelihood that a PES and Census record pertain to the same individual.
Census Collection Districts (CDs)	The basic geographic unit of collection in the 2006 Census. It is generally an area that one collector can deliver and collect forms over a specified collection period.
Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS)	A series of processes which prioritises and organises ADL output for use by the MSS system.
Collector Workloads (CLWs)	The basic geographic unit of collection for the 2011 Census. It is generally an area that one collector can deliver and collect forms over a specified collection period.
Computer Assisted Interviewing (CAI)	A method of data collection whereby responses are recorded directly into an electronic questionnaire on a notebook computer.
Correlation bias	A bias arising when people who were not counted in the Census are more likely to be missed in the PES than people with similar values of the characteristics used in PES estimation (such as age, sex and Indigenous status) who were counted in the Census.
Coverage	The survey coverage refers to the population units which have a chance of being selected in the survey sample. For the quality of the survey estimates, it is desirable that the survey coverage matches as closely as possible the survey scope. Coverage rules are generally applied in all household surveys to ensure that each person is associated with only one dwelling, and hence has only one chance of selection.
Discrete Indigenous community	A geographic location, bounded by physical or legal boundaries, and inhabited or intended to be inhabited predominantly by Indigenous people, with housing or infrastructure that is either owned or managed on a community basis.
Estimated Resident Population (ERP)	The official measure of the population of Australia based on the concept of residence. It refers to all people, regardless of nationality or citizenship, who usually live in Australia, with the exception of foreign diplomatic personnel and their families. It includes usual residents who are overseas for less than 12 months. It excludes overseas visitors who are in Australia for less than 12 months.
ERP rebasing	Following each Census of Population and Housing, ABS rebases its official estimates of population - the Estimated Resident Population (ERP) series. This provides the basis for population estimates until the next Census. The rebased Census-year ERP together with other Census information and data on intercensal components of population change allows ERP revisions to be made back to, but not including, the previous Census.
Freely Extensible Biomedical Record Linking (FEBRL)	Probabilistic linking software developed at the Australian National University used for PES automated data linking activities.
Imputation	A statistical process for predicting values where no response was provided to a question and a response could not be derived.
Imputed dwelling (in Census)	A dwelling which is determined to be occupied in the Census and where Census data is imputed because no Census form was received.
Indigenous Community Frame (ICF)	The listing of all discrete Indigenous communities from which the sample of communities is selected.
Late return	A Census form which was returned after the start of PES enumeration.
Mainstream sample	All PES private dwelling sample selections other than those from discrete Indigenous communities.

Match and Search System (MSS)	The main PES clerical review facility, which allows processors to search, view, compare, and record matches between PES and Census data.
Non-private dwelling	An establishment which provides a communal type of accommodation, such as a hotel, motel, hospital or other institution.
Non-sampling error	Error arising from inaccuracies in collecting, recording and processing the data. Every effort is made to minimise non-sampling error by the careful design of questionnaires, intensive training and supervision of interviewers, and efficient data processing procedures. Non-sampling error also arises because information cannot be obtained from all persons selected in the survey and data are imputed based on assumption about the non-respondents.
Outstation (or homeland)	A discrete Indigenous community that has a population of fewer than 50 people AND is administered by, or linked to, an organisation such as a Resource Agency or larger parent discrete Indigenous community for the provision and maintenance of services.
Private dwelling	A residential structure which is self-contained, owned or rented by the occupants, and intended solely for residential use. A private dwelling may be a flat, part of a house, or even a room, but can also be a house attached to, or rooms above, shops or offices.
Sampling error	Error occuring because a sample, rather than the entire population, is surveyed. One measure of the likely difference resulting from not including all dwellings in the survey is given by the standard error. There are about two chances in three that a sample estimate will differ by less than one standard error from the figure that would have been obtained if all dwellings had been included in the survey, and about nineteen chances in twenty that the difference will be less than two standard errors.
Scope	Within household surveys in the ABS, survey scope is considered to be the population about which inferences are desired: that is, when the results are published, the population to which they refer.
Search address	An address where a person was reported to be staying on Census night or where a person may have been included on a Census form. PES processing attempts to locate a Census form for each search address, in order to determine the number of times (if any) a person enumerated in the PES was included on a Census form.
Unoccupied dwelling (in Census)	A structure built specifically for living purposes which is habitable but the Census Collector was certain was unoccupied on Census night.

BIBLIOGRAPHY

ABS 2006, *Information Paper: Measuring Net Undercount in the 2006 Population Census, 2006*, cat. no. 2940.0.55.001, ABS, Canberra.

ABS 2007, *Australian Demographic Statistics, December quarter 2006*, cat. no. 3101.0, ABS, Canberra.

ABS 2007, *Information Paper: Measuring Net Undercount in the 2006 Population Census, 2006*, cat. no. 2940.0.55.001, ABS, Canberra.

ABS 2007, *Research Paper: An Estimating Equation Approach to Census Coverage Adjustments, May 2007*, cat. no. 1351.0.55.019, ABS, Canberra.

ABS 2009, *Population Estimates: Concepts, Sources and Methods, 2009*, cat. no. 3228.0.55.001, ABS, Canberra.

Kinfu, Y., 2005, 'Spatial mobility among Indigenous Australians: Patterns and determinants', Demography and Sociology Program - Working Papers in Demography No. 97, Australian National University, Canberra.

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** the ABS website is the best place for data from our publications and information about the ABS.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au