



Information Paper

**Measuring net
undercount in the 2006
Population Census**

Australia

2006

Information Paper

**Measuring net
undercount in the 2006
Population Census**

Australia

2006

**Dennis Trewin
Australian Statistician**

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30AM (CANBERRA TIME) THURS 13 JUL 2006

ABS Catalogue No. 2940.0.55.001

ISBN 0 642 48197 0

© Commonwealth of Australia 2006

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior written permission from the Commonwealth. Requests and inquiries concerning reproduction and rights in this publication should be addressed to The Manager, Intermediary Management, Australian Bureau of Statistics, Locked Bag 10, Belconnen ACT 2616, by telephone (02) 6252 6998, fax (02) 6252 7102, or email: <intermediary.management@abs.gov.au>.

In all cases the ABS must be acknowledged as the source when reproducing or quoting any part of an ABS publication or other product.

Produced by the Australian Bureau of Statistics

INQUIRIES

- For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070.

CONTENTS

page

INTRODUCTION

The Census of Population and Housing	1
Net undercount	1
Purpose of this paper	2

CENSUS POST ENUMERATION SURVEY

Overview	3
Independence from the Census	3
Changes since the 2001 PES	4

METHODOLOGY OF THE 2006 CENSUS POST ENUMERATION SURVEY (PES)

Scope	6
Coverage rules	7
Sample design	7
Collection methods	9
Questionnaire design	11
Data capture	11
Matching and searching	12
Vague address imputation	13
Estimation	13

HOW ESTIMATED NET UNDERCOUNT CONTRIBUTES TO THE ESTIMATED RESIDENT POPULATION (ERP)

Background	18
Calculation of ERP	18
Further information	19
Glossary	20
Bibliography	22

ABBREVIATIONS

'000	thousand
ABS	Australian Bureau of Statistics
ARA	any responsible adult
ASGC	Australian Standard Geographical Classification
CAI	computer assisted interviewing
CD	Collection District
ERP	estimated resident population
ICF	Indigenous Community Frame
NPD	non-private dwelling
PAPI	pen-and-paper interview
PES	Census of Population and Housing Post-Enumeration Survey
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 night will be Tuesday 8 August 2006. 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 included 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 they were in was missed because it was difficult to find (e.g. in a remote or non-residential area)
- the dwelling they were in 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
- 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 invariably occur in such a large operation.

Usually more people are missed than overcounted in Australia, so the Census count of the population would be less than the true population. This difference is called net undercount.

Rates of 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

INTRODUCTION *continued*

NET UNDERCOUNT *continued*

- provide users with an assessment of the completeness of Census counts, allowing them to take this into account when using Census information
- 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 2006 Census Post Enumeration Survey. Comments on these methods are welcomed and may be provided to Merilyn Henden on (02) 6252 7166 or email: merilyn.henden@abs.gov.au.

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.

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, a responsible adult member is interviewed and asked about all persons present or usually resident in the household. 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 person 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 2001 PES, a sample of about 37,000 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 84,000. 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 2006 PES sample size (about 38,000 dwellings) is expected to result in a similar number of people. More detail on the PES sample is given later in this paper.

INDEPENDENCE FROM THE CENSUS

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 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
- maintaining the confidentiality of the PES sample so that Census field and office staff are not aware which areas are included in the PES.

INDEPENDENCE FROM THE CENSUS *continued*

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.

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 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.

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 2001 PES

Evaluation of the 2001 Census, the 2001 PES, and the subsequent rebasing of the Estimated Resident Population (ERP) highlighted a number of areas that required further investigation in preparation for the 2006 PES. These included:

- options for improving the quality of the estimated net undercount for remote areas and Indigenous people
- the possibility of improved quality of the PES by the inclusion of non-private dwellings, such as hotels, motels, hospitals and other institutions
- the possibility of improvements to 'match and search' processing
- the possibility of improvements to the PES estimation method.

As a result of these and other investigations, a number of changes are being implemented in the 2006 PES, including:

- expanding the scope of the PES to include remote areas and discrete Indigenous communities
- the use of a Computer Assisted Interviewing instrument to replace the paper questionnaire
- special procedures to limit the impact of Census extended follow-up on the PES

CENSUS POST ENUMERATION SURVEY *continued*

CHANGES SINCE THE
2001 PES *continued*

- a more computerised match and search system for processing the PES data, including the recording of more detailed information about the reasons for matching decisions
- an improved estimation method.

Additional information on these changes can be found below.

METHODOLOGY OF THE 2006 PES

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:

- people in non-private dwellings such as hotels, motels, hospitals and other institutions
- homeless people (as the sample selected in the PES is based on the selection of dwellings)
- foreign diplomats and their families
- overseas visitors who were not in Australia on 8 August 2006 (Census night)
- babies born after 8 August 2006
- 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 some time in August, provided that they usually live with people remaining in Australia.

The 2006 PES will include remote areas of Australia and discrete Indigenous communities in scope for the first time.

Remote areas and discrete Indigenous communities

Remote areas and discrete Indigenous communities have previously been 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.

Since 2001, however, the ABS has conducted a number of Indigenous surveys and has become more adept at managing the issues associated with data collection in these areas. The practice of using Indigenous facilitators drawn from the local community has been established for ABS surveys, and guidelines for enumerating discrete Indigenous communities in surveys have been developed, and continue to improve.

Testing of the proposed methodology for enumerating discrete Indigenous communities has shown it is feasible to include them in the scope of the PES. Inclusion of these communities will ensure the geographic scope of the PES is more complete than it has been in the past. Since this is new for the PES, the field procedures and resulting estimates will be scrutinised closely.

The PES in discrete Indigenous communities will be conducted as soon as is practicable after the completion of the Census in each community. The PES enumeration of discrete Indigenous communities is expected to occur between the end of August and the end of October 2006.

Non-private dwellings

Non-private dwellings (NPDs) 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 (see below) 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 respondents themselves and so is collected from administrative lists or from staff
- lack of independence from Census, because the administrative lists used by interviewers to collect data about institutionalised people are likely to be the same as the ones used by Census staff
- the relatively high cost of enumerating them compared to private dwellings, because of the additional time and effort required to set up the survey.

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. For these reasons, non-private dwellings will be excluded from the scope of the 2006 PES. This is expected to have minimal impact on overall estimates of net undercount.

COVERAGE RULES

The PES uses coverage rules 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 results 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 32,000 dwellings (0.3% of dwellings in Australia) are expected to respond to the PES, including approximately 200 dwellings selected from around 20 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

SAMPLE DESIGN *continued*

the results of other ABS surveys, the rate of sample loss is expected to vary between different components of the sample. The sample design takes this into account.

TABLE 1. EXPECTED NUMBER OF FULLY RESPONDING HOUSEHOLDS

	<i>Private dwellings</i>	<i>ICF dwellings(a)</i>	<i>Total</i>
	no.	no.	no.
New South Wales	7 050	0	7 050
Victoria	6 500	0	6 500
Queensland	5 705	13 (2)	5 718
South Australia	3 835	6 (1)	3 841
Western Australia	4 155	13 (3)	4 168
Tasmania	1 950	0	1 950
Northern Territory	1 235	180 (15)	1 415
Australian Capital Territory	1 250	0	1 250
Australia	31 680	212 (21)	31 892

(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 error (SE):

- 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
- Indigenous people at Australia level – less than 2.6 percentage points.

Private dwelling sample

The PES sample of private dwellings is obtained using a stratified multi-stage area sample. Using the Statistical Division and Subdivision structure of the Australian Standard Geographical Classification (ASGC), Australia is first divided 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)
- 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 probabilities of selection at each stage are set so that each dwelling within a state or territory has the same overall chance of selection in the PES.

Discrete Indigenous community 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 Community 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. Each dwelling on the list is given a dwelling number. A pre-determined random start point and dwelling skip is applied to select the dwellings to be included in the sample.

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.

COLLECTION METHODS

Field procedures

Various strategies have been devised for the enumeration of the PES. Where possible, standard procedures are used in 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, an attempt will be made to ensure underlying concepts remain the same across both components of the sample.

Specially trained PES interviewers will collect data through face-to-face interviews starting around three weeks after Census night. Interviews will be conducted using Computer Assisted Interviewing (CAI), which replaces the paper questionnaire used in previous PESs. Interviews will be conducted with any responsible adult (ARA) member of the household who will be asked to respond on behalf of all the members of the household.

Census follow-up

The 2001 Census saw a significant increase in non-response compared to 1996. For this reason, more intensive follow-up procedures are being implemented for 2006, which will delay the end of Census field operations in a small proportion of CDs by two weeks.

If the 2001 timetable were followed, Census enumeration would finish on 26 August and PES enumeration would begin on 29 August. In 2006, the main Census enumeration will finish on 30 August and Census extended follow-up will be conducted between 5 and 9 September. To allow for the extended Census follow-up, PES enumeration will begin on 1 September and will continue until 24 September. This, however, still results in five days of overlap between the two collections in an estimated 250 CDs (5% of CDs in the PES sample). Special field procedures are being implemented to minimise the impact of this overlap.

Census follow-up continued

As in previous PESs, special procedures will be implemented for Census forms received after the start of the 2006 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 (PAL) 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' and treated accordingly. The treatment of late returns is explained in the Estimation section of this paper. Special procedures will be implemented to ensure that any Census forms returned or collected as a result of Census follow-up activities are also flagged as late returns.

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:

- a requirement that interviewers wanting to apply for District Manager or Area Supervisor positions in the Census 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 (CD)
- any person who has worked in an area of Census processing related to dwelling or population counts may not also work on PES processing
- any person who has been involved in Census field collection in any capacity may not also work on PES processing.

Discrete Indigenous communities

Census field operations in discrete Indigenous communities are expected to begin as early as mid-July 2006 and continue until the end of September, with the majority of the field work being conducted one week either side of Census night (8 August 2006). Enumeration of the PES in these areas 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.

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 continued

Due to practical considerations, discrete Indigenous communities will be enumerated via a pen and paper interview (PAPI) questionnaire, rather than via the CAI methodology. The data collected will be transcribed later by the interviewer into a customised CAI instrument.

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 origin) 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.

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. Visitors to households included in the PES are also asked for their address of usual residence. These responses are used to search the Census forms received to determine the number of times each PES respondent was included in the Census.

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

A specially designed PAPI questionnaire and CAI instrument will be used in discrete Indigenous communities. The questionnaire will collect the following additional data items:

- up to two alternative names the person is known by (this information will assist in matching on name)
- whether the person was identified on an administrative list in the community (this information will assist in matching).

A number of questions in the PES were not considered applicable to people living in discrete Indigenous communities, and a number of questions needed to be reworded. For example, questions about addresses have been modified to cater for situations where street addresses may not exist or are not used by residents of the community. Other questions have also been modified slightly to be more conversational in tone.

The PAPI questionnaire will be used 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 at 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 assist 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 the match and search processing system draws information.

MATCHING AND
SEARCHING

The match and search process involves comparing dwellings and people enumerated in the PES with dwellings and people found in the Census. The objective of matching and searching is to determine:

- whether each PES dwelling has been counted or missed in the Census;
- whether each person at the PES dwelling was counted in the Census at that address; and
- whether any people at the PES dwelling were counted in the Census at any other addresses.

The first step in the process is to search the Census file for an address which matches the address of the PES dwelling. Once a dwelling is matched, the Census person records for that dwelling are compared with the PES person records. During person matching, information from the PES is compared with information on the Census record to determine whether a person match has been made. The information compared includes name, sex, date of birth, age, marital status, Indigenous origin 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 match status and the level of match. If a person match cannot be found at the particular Census dwelling, a wider search is conducted of the whole of the CD.

The final step in the process is to check whether the person was included on a Census form at any other address. The PES questionnaire asks respondents where each person in the household was staying on Census night, and for any other addresses where each person may have been included on a Census form. Census forms for each of these addresses are checked, and the process of person matching repeated, to determine whether the person was included on a Census form at any of these other addresses.

Census undercount occurs when a person included in the PES who was in Australia on Census night and should have been counted in the Census cannot be found on any Census form. Census overcount occurs when a person included in the PES is found on more than one Census form, or when a person is found on a Census form who should not have been included (for example, a person who was overseas on Census night).

Quality assurance

To ensure the accuracy of 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.

*Matching and searching
for discrete Indigenous
communities*

The match and search process for discrete Indigenous communities will be similar to that for the private dwelling component. However, an alternative approach will be used in person matching to account for the fact that, as reported by researchers such as Kinfu (2005), some Indigenous Australians are associated with more than one dwelling and move between these dwellings on a regular, seasonal or random basis. 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

Matching and searching for discrete Indigenous communities continued

communities involves searching the whole community for a person match, rather than just searching within a single dwelling. This is the equivalent of the 'whole of CD' search in the private dwelling sample.

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

VAGUE ADDRESS
IMPUTATION

For the 84,000 respondents to the 2001 PES, there were 6,200 search addresses (i.e. addresses other than the PES dwellings where a respondent may have been included on a Census form). A similar number of search addresses is expected in 2006. Of these additional search addresses, it is likely that some will be incomplete or too vague (e.g. Western Sydney) to allow a match status to be determined. In these cases, rather than simply assuming that the person was not included on a Census form at another address, the match status for vague search addresses will be imputed based on the characteristics of the person in question.

Using data from people who supplied search addresses which were precise enough to determine a match status, a logistic regression model will be used to impute match status for vague search addresses using the following regression variables:

- the PES question in answer to which the search address was given
- Census night address
- scope and coverage status
- Indigenous origin
- whether the respondent has already been matched to a Census form at the PES address
- whether the respondent considered that the person had been included on a Census form
- age
- marital status
- part of state
- number of search addresses
- whether selected in a discrete Indigenous community.

ESTIMATION

The PES interview process determines whether each person in the sample should have been counted in the Census, and the match and search 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 number of people who should have been counted in the Census. Net undercount is the difference between this estimate and the actual Census count.

Broadly speaking, the PES estimate of the number of people who should have been counted in the Census is obtained as a weighted sum of the number of people in the PES sample who should have been counted in the Census. More precisely, there are small adjustments which allow for late returns to the Census, and for persons imputed by the Census for non-contact dwellings.

Census late returns

Being contacted for inclusion in the PES will prompt some householders to complete and return their Census form when they otherwise would not have done so, giving the PES sample an artificially high Census response. To avoid this, PES estimates are initially based solely on the Census information available at the commencement of the PES survey period.

*Imputed persons in
Census non-contact
dwellings*

Imputation is a statistical process for predicting values where no response was provided. If no Census form is received for a dwelling, and that dwelling is determined to be occupied on Census night, the person data for that dwelling is imputed using 'hotdeck' imputation on the Census file. Hotdeck imputation is a process whereby data are copied from a dwelling with similar attributes in the same geographic area. Hotdeck imputation of person records for Census non-contact dwellings does not occur until long after the end of Census field collection, after all late returns have been received.

*Treatment of Census late
returns and imputed
dwellings*

The PES initially estimates the number of people who should have been counted in the Census at the commencement of PES enumeration. Census counts for late returns and imputed dwellings are later added. As for previous PESs, the approach adopted for 2006 assumes the Census counts from late returns and imputed dwellings are accurate. This gives a better overall estimate than using the PES sample to estimate for these late returns and imputed dwellings, as their contribution could be measured only with a relatively high error. However, undercount estimates specific to these dwellings, based on the PES sample, will be used for evaluation of this assumption.

Treatment of Census late returns and imputed dwellings continued

In mathematical terms, the treatment of late returns and imputed dwellings in the estimate of the number of people who should have been counted in the Census can be shown by:

$$T_C = T_{P,C} + LR_C + I_C$$

where:

T_C = PES estimate of the number of people who should have been counted in the Census for category C

$T_{P,C}$ = PES estimate of the number of people who should have been counted in the Census at the commencement of PES enumeration for category C (i.e. excluding imputed dwellings and late returns)

LR_C = Census count of people for category C who supplied late returns after the start of PES

I_C = Census imputed counts of people for category C in dwellings known to be occupied, but where a Census form was not obtained

$$T_{P,C} = \sum_{i \text{ in } C} W_i$$

where:

$\sum_{i \text{ in } C} W_i$ = sum of the weights of PES sampled people i who should have been counted in category C (excluding imputed dwellings and late returns)

Net undercount

The PES estimate of Census net undercount is the difference between the estimate of people who should have been counted in the Census, and the number of people who were actually counted in the Census.

The net undercount rate is the difference between the above two measures expressed as a proportion of the number of people who should have been counted in the Census.

Overview of weighting for the PES

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-contact 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.

Dwelling weights in the PES continued

Although the PES sample is designed to ensure each private dwelling in a state or territory has an equal chance of selection, in practice certain types of dwellings are more likely to be missed in the PES. Accordingly, different weights are attached to each PES dwelling counted in the Census depending on the dwelling category (or post-stratum). 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
- region (six states and NT divided into capital city and balance, plus ACT, giving 15 regions).

INITIAL DWELLING WEIGHTS

For dwellings selected in the PES that were counted in the Census, initial dwelling weights for each post-stratum are calculated as the Census count divided by the number of selected PES dwellings.

Additional post-strata are formed for dwellings selected in the PES that were not counted (as a separate dwelling) in the Census. These dwellings are given an average initial weight from the counted dwellings with the same region and dwelling structure.

ADJUSTMENT OF DWELLING WEIGHTS FOR NON-RESPONSE

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. This adjustment is made separately within similar post-strata.

Person weighting

In estimation, the population is divided into groups in which, as closely as possible, the likelihood of a person being missed in the PES is unrelated to their being missed in the Census. These groups are called 'benchmark categories', and are based on personal characteristics including age, sex, state/territory, country of birth and Indigenous status.

Initial person weights for all responding people in the PES sample are set to the dwelling weight of their PES dwelling. These person weights are then increased or decreased from the corresponding dwelling weights so that the PES estimate of the number of times people actually were counted in the Census, for each benchmark category, corresponds to the actual Census count for that category (excluding imputed dwellings and late returns).

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). Summing the dwelling weights of the PES sample of people in this category multiplied by how many times they were counted in the Census, gives an estimate of the number of times the Census counted people in this category. If this is lower than the actual Census count, then this

Person weighting continued

benchmark category is under-represented in the PES sample. Person weighting adjusts the weights so that the benchmark categories are all appropriately represented by the PES sample.

The same person weights are then applied to the number of times a person should have been counted in the Census (which can be either 0 or 1) to give an estimate of the correct Census count allowing for undercount and overcount (again excluding Census imputed dwellings and late returns).

In previous PESs, the ABS has calculated person weights using an estimator known as a Dual System Estimator (DSE). This can ensure that the PES estimate of actual Census counts equals the Census counts only for a set of non-overlapping benchmark categories known as post-strata (e.g. region by age by sex). Ideally, many fine post-strata would be used, because an assumption implicit in the estimator is that, within those post-strata, all people have the same probability of being correctly counted in the Census.

For 2006, the ABS has developed an estimator that extends the DSE approach to account for overlapping benchmark categories. The new estimator is called the Prediction Regression (PREG) estimator. For the 2006 PES, the PREG estimator will allow the region by sex by age benchmark counts to be supplemented, for example by benchmark counts classified by Indigenous status, marital status, and country of birth. Simulation studies have shown the PREG estimator to be technically superior to the DSE. A peer review of the PREG estimator has been undertaken by the Centre for Statistical and Survey Methodology at the University of Wollongong. The review confirmed the validity of the new estimator. A paper describing the theory of the PREG estimator will be available on the ABS web site later in 2006.

HOW ESTIMATED 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 demographic, social and economic studies. These estimates are also the basis for distributing Commonwealth funding to the states and territories. 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, excluding Australian residents overseas, is obtained by excluding all overseas visitors enumerated in the Census.

This Census count is then 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 this estimate of net undercount is the PES.

ERP calculations then make an adjustment for Australian residents who are 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 Multicultural Affairs (DIMA) 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 2001 Census.

Table 2. Components of Estimated Resident Population (ERP), Australia, 30 June 2001.

Components as at 7 August 2001—	Persons '000
Census count, actual location	18 972.4
<i>less</i> Overseas visitors	203.1
<i>equals</i> Census count, place of usual residence	18 769.2
<i>plus</i> Net undercount	346.2
<i>plus</i> Residents temporarily overseas	330.2
<i>equals</i> ERP as at 7 August 2001	19 445.6
Components from 1 July to 7 August 2001—	
<i>less</i> Births	25.4
<i>plus</i> Deaths	14.8
<i>less</i> Net permanent and long-term overseas movement	21.8
<i>equals</i> ERP as at 30 June 2001	19 413.2

Information on the calculation of the ERP for 30 June 2006 based on the 2006 Census will be reported in *Australian Demographic Statistics, December quarter 2006* (cat. no. 3101.0), due for release on 5 June 2007.

HOW ESTIMATED NET UNDERCOUNT CONTRIBUTES TO ERP

continued

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 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 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 2001 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.

Population estimates based on the previous Census are not entirely independent of the National Demographic Data Bank and Medicare enrolment data because of the five years of intercensal population components of growth (births, deaths and migration data) common to these datasets.

FURTHER INFORMATION

For further information on measuring net undercount in the 2006 Census, contact Meryl Henden on (02) 6252 7166 or email: merilyn.henden@abs.gov.au.

GLOSSARY

Benchmark category	A category for which the PES estimate of persons counted in the Census is forced to reproduce the actual Census count.
Census collection districts (CDs)	A Census collection district (CD) is the basic geographic unit of collection in the Census. A CD is generally a Census workload area that one collector can cover delivering and collecting forms in a specified period. On average there are about 200 dwellings per CD, however, there may be more in some urban CDs, and in rural areas a CD may contain fewer dwellings yet cover an extensive area.
Computer Assisted Interviewing	Computer-assisted interviewing (CAI) is 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 <i>were not</i> 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 <i>were</i> 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 discrete Indigenous community is defined as 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)	ERP is 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.
Hotdeck imputation	An imputation process whereby a donor record is located and relevant responses copied from the donor record to a non-responding record. The donor record will have similar characteristics to the non-responding record and must also have the required variable(s) stated. In addition, the donor record will be located geographically as close as possible to the location of the record to be imputed. The match must occur within the same Capital City or Balance of State. When a suitable match is found, then the copying of the response(s) from the donor record to the variable(s) that have missing values can occur.
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.
Late return	A Census form which was returned after the start of PES enumeration.
Non-private dwelling	An establishment which provides a communal type of accommodation, such as a hotel, motel, hospital or other institution.
Non-sampling error	Non-sampling error arises 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.

GLOSSARY *continued*

Outstation (or homeland)	A discrete Indigenous community that has a population of less 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 private dwelling is 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.
Remote areas	<p>Within the Australian Standard Geographical Classification (ASGC), the Remoteness classification comprises five categories each of which identifies a (non-contiguous) region in Australia having a particular degree of remoteness. The categories range from 'highly accessible' to 'very remote'.</p> <p>The degree of remoteness of each Collection District (CD) was determined using the Accessibility/Remoteness Index of Australia (ARIA). CDs have then been grouped into the appropriate category of Remoteness to form non-contiguous areas within each State.</p> <p>For more information, refer to <i>Statistical Geography Volume 1: Australian Standard Geographical Classification (ASGC) 2001</i> (Cat. no. 1216.0) and <i>ABS Views on Remoteness</i> (Cat. no. 1244.0).</p>
Sampling error	Sampling error occurs 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.
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.
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.
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 2003, *Information Paper: Census of Population and Housing, Data Quality - Undercount, 2001*, cat. no. 2940.0, ABS, Canberra.

ABS 2002, *Information Paper: Labour Force Survey Sample Design, 2002*, cat. no. 6269.0, ABS, Canberra.

Dunstan, K, Heyen, G & Paice, J 1999, *Demography Working Paper 1999/4: Measuring Census Undercount in Australia and New Zealand*, 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 web site is the best place for data from our publications and information about the ABS.
- LIBRARY* A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries.

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 web site, or purchase a hard copy publication. 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 ABS statistics can be downloaded free of charge from the ABS web site.

- WEB ADDRESS* www.abs.gov.au



2000001526316

ISBN 0 642 48197 0

RRP \$11.00