



2006

2940.0

CENSUS OF POPULATION AND HOUSING – UNDERCOUNT AUSTRALIA

EMBARGO: 11.30AM (CANBERRA TIME) TUES 5 JUN 2007

CONTENTS

	<i>page</i>
Notes	2

OVERVIEW

Introduction	3
What is undercount?	3
Why measure undercount?	4
Census Post Enumeration Survey (PES)	4

ESTIMATES OF NET UNDERCOUNT

Introduction	6
Australia, states and territories	6
Capital city and balance of state/territory	7
Age and sex	9
Registered marital status	10

EXPLANATORY NOTES

Scope of the 2006 Census	11
Scope and coverage of the 2006 PES	11
PES estimation	12
Census counts	12
Correction for Census imputation error in the non-contact sector	14
Independence from the Census	14
Reliability of the undercount estimates	15

APPENDICES

Appendix 1: Census Post Enumeration Survey	17
Appendix 2: PES Match and Search Processing	22
Appendix 3: PES Estimation	24

TECHNICAL NOTE: SAMPLING ERRORS

Sampling errors associated with statistics produced from the PES	27
--	----

GLOSSARY

Glossary	28
----------------	----

INQUIRIES

For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070 or Meryln Henden on Canberra (02) 6252 5489.

NOTES

ABOUT THIS PUBLICATION This publication presents estimates of net undercount for the 2006 Census, as well as information on how the estimates were calculated.

NOTES ABOUT THE ESTIMATES Information in this publication has been obtained from the 2006 Census Post Enumeration Survey (PES). The PES is a household survey conducted by the Australian Bureau of Statistics (ABS) about three weeks after Census night. People from approximately 33,000 households across Australia were asked about everyone present or usually resident in the household. In addition to obtaining basic demographic information, questions were asked about each person's usual residence, their location on Census night, and any other addresses where they might have been counted in the Census.

The results of the PES are used to determine how many people were missed in the Census and how many were counted more than once.

The estimated net undercount from the PES is used to augment the Census counts for the purpose of deriving population estimates for Australia and the states and territories.

ROUNDING As estimates have been rounded, discrepancies may occur between sums of the component items and totals.



ABBREVIATIONS

- ABS Australian Bureau of Statistics
- CD Collection District
- ERP estimated resident population
- FMC field match code
- ICF Indigenous Community Frame
- MSS Match and Search System
- PD private dwelling
- PES Census of Population and Housing Post-Enumeration Survey
- PMC person match code
- PREG prediction regression
- QA quality assurance
- SE standard error

Brian Pink
Australian Statistician

OVERVIEW

INTRODUCTION

Tuesday, 8 August 2006 was Census night in Australia. Every person present in Australia on Census night, excluding foreign diplomats and their families, should have been included on a Census form at the place where they stayed.

Whenever a Census is undertaken, questions about the completeness and accuracy of the Census count invariably arise. In such a large and complex exercise, it is inevitable that some people will be missed and some will be included more than once (or included when they shouldn't be). In Australia, the Census Post Enumeration Survey (PES) is used 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 the Australian Bureau of Statistics (ABS) shortly after the Census. The survey provides an independent check on Census coverage and also identifies the key demographic characteristics of the population that have been missed or over counted in the Census.

WHAT IS UNDERCOUNT?

In the 2006 Census, some people were missed (undercount) and some were counted more than once (overcount). As is usually the case, in 2006 more people were missed than over counted. The difference between the Census count and the number of people who should have been counted is called the 'net undercount' of the Census. (In this context, 'Census count' includes person records imputed for dwellings that were non-responding in the Census.)

Some of the reasons why people are missed in the Census 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 they were staying at and they did not obtain additional forms
- the person completing the form thought that young babies, the elderly or visitors (for example) should not be included
- they did not wish to be included due to concerns about 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)
- difficulty with completion of an eCensus form (Internet)
- the dwelling they were in was mistakenly classed as unoccupied.

Some of the reasons people are counted more than once in the Census 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.

The net undercount for a category of person is obtained by taking the PES estimate of the number of people in the category who should have been counted and subtracting the Census count of the number of people (in the category). Net undercount for a category of person is the net result of the PES estimate of gross undercount, gross overcount, differences in classification between the PES and Census (e.g. age, sex, Indigenous status) and imputation error in the Census.

Rates of net undercount vary significantly for different population groups depending on factors such as age, sex, ethnicity (including Indigenous status) and geographic location.

OVERVIEW *continued*

WHY MEASURE UNDERCOUNT?

There are a number of reasons why measuring undercount in the Census is important:

- to augment the Census count for the purpose of deriving an estimate of the resident population for 30 June of the Census year
- to provide users with an assessment of the completeness of the Census count, allowing them to take this into account when using Census information
- to evaluate the effectiveness of Census collection procedures so that improvements can be made for future Censuses.

Accurate resident population estimates are required for demographic, social and economic studies, as well as for the allocation of seats in the Federal House of Representatives, and Commonwealth payments to states and territories and local government.

CENSUS POST ENUMERATION SURVEY (PES)

The PES interview process determines whether each person in the sample should have been counted in the Census, and the category in which they should have been counted (such as age, sex, Indigenous status, region of usual residence). The match and search process determines how many times each person in the PES sample was actually counted in the Census. PES output processing and estimation combines and weights results from the match and search process 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 (including imputed persons for non-responding dwellings).

A number of improvements have been implemented in the 2006 PES. These include:

- 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 of previous PESs
- special procedures to limit the impact on the PES of Census extended follow-up
- 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, including a specific adjustment for dwelling-level non-response
- weighting for 'late return' and imputed dwellings designed to make these dwellings representative of this combined group in the Census
- new estimator (dubbed PREG), allowing:
 - enhanced model for coverage and response e.g. separate effects for persons sampled in discrete Indigenous communities
 - handling of differences in reporting (e.g. age, Indigenous status) between Census and the PES.

Estimation

PES estimation involves assigning a 'weight' to each selected PES dwelling and then to each person for whom a PES response was obtained. The PES only covers people associated with private dwellings and discrete Indigenous community dwellings available for selection during PES enumeration. However, the weights attached to persons allow the PES sample to represent the whole population of interest – that is, all people present

OVERVIEW *continued*

Estimation continued

in Australia on Census night, including people in non-private dwellings, such as hotels, hospitals and jails, which are not covered by the PES.

Dwelling weighting for the 2006 PES comprises two stages. For private dwellings selected in the PES that were found in the Census, the first stage of weighting adjusts the PES selection weight such that the adjusted weights add up to the Census private dwelling count within categories based on geography and dwelling characteristics. A first-stage weight adjustment is also applied to private dwellings selected in PES that were missed in the Census. For dwellings in discrete Indigenous communities, a similar first-stage weight adjustment is applied based on dwelling counts for communities within each state and territory.

The second stage of dwelling weighting applies a non-response adjustment so that the responding PES dwellings represent other dwellings from which no response was obtained.

The initial stage of person weighting adjusts the dwelling weights to ensure that the PES estimates of people counted in private dwellings and discrete Indigenous community dwellings (other than late return or imputed dwellings) in a set of benchmark categories match the actual Census counts for these categories. The weight adjustment applied to a person does not depend on whether they responded in the Census, but only on characteristics of the person as reported in the PES.

As a final step in weight adjustment, the initial person weights are adjusted so that the PES estimates also represent people in non-private dwellings.

Further information on PES estimation can be found in Appendix 3 of this publication.

ESTIMATES OF NET UNDERCOUNT

INTRODUCTION

This section presents estimates of net undercount of people in the 2006 Census for key demographic characteristics. Net undercount is the difference between the PES estimate of the number of people who should have been counted in the Census and the actual Census count (including imputed persons for non-responding dwellings).

In the following tables, net undercount is presented as level estimates and rates, along with their standard errors. The net undercount rate expresses the net undercount as a percentage of the PES estimate of the number of people who should have been counted in the Census.

Estimates of net undercount have been calculated on a place of usual residence basis.

AUSTRALIA, STATES AND TERRITORIES

The 2006 Census counted 19,852,973 usual residents of Australia. It is estimated that this count is 549,486 persons fewer than the number of usual residents who were actually present in Australia on Census night – a net undercount rate of 2.7%.

Table 2.1 and graph 2.2 set out the 2006 Census rates of net undercount for Australia, States and Territories, along with the rates from the 1991, 1996 and 2001 Censuses.

Table 2.1 shows some consistent patterns in the rates of undercount for the states and territories. For example, Northern Territory has had the highest net undercount rate in the past four Censuses, while the Australian Capital Territory has consistently recorded the lowest net undercount rate.

All states and territories had a higher net undercount rate in 2006 than in 2001. The greatest increases were for the Northern Territory (4.0% in 2001 to 7.6% in 2006) and Queensland (1.9% in 2001 to 3.7% in 2006).

2.1 NET UNDERCOUNT(a), State/territory of usual residence

	1991			1996			2001			2006(b)		
	Persons	Rate		Persons	Rate		Persons	Rate		Persons	Rate	
		%	SE									
New South Wales	109 200	1.9	0.1	91 400	1.5	0.2	130 106	2.0	0.2	157 578	2.4	0.4
Victoria	78 800	1.8	0.1	74 000	1.6	0.3	67 254	1.4	0.2	113 596	2.3	0.4
Queensland	52 100	1.8	0.1	57 300	1.7	0.3	68 514	1.9	0.2	148 409	3.7	0.4
South Australia	22 300	1.6	0.1	19 300	1.3	0.3	24 293	1.6	0.2	36 281	2.3	0.4
Western Australia	33 200	2.1	0.2	28 100	1.6	0.3	37 446	2.0	0.3	64 150	3.2	0.6
Tasmania	7 700	1.7	0.2	6 600	1.4	0.4	7 410	1.6	0.3	9 535	2.0	0.6
Northern Territory	4 800	2.9	0.7	5 700	3.1	1.6	7 814	4.0	0.6	15 909	7.6	1.5
Australian Capital Territory	4 100	1.4	0.2	3 400	1.1	0.3	3 282	1.0	0.4	4 027	1.2	1.0
Australia	312 300	1.8	0.1	285 800	1.6	0.1	346 119	1.8	0.1	549 486	2.7	0.2

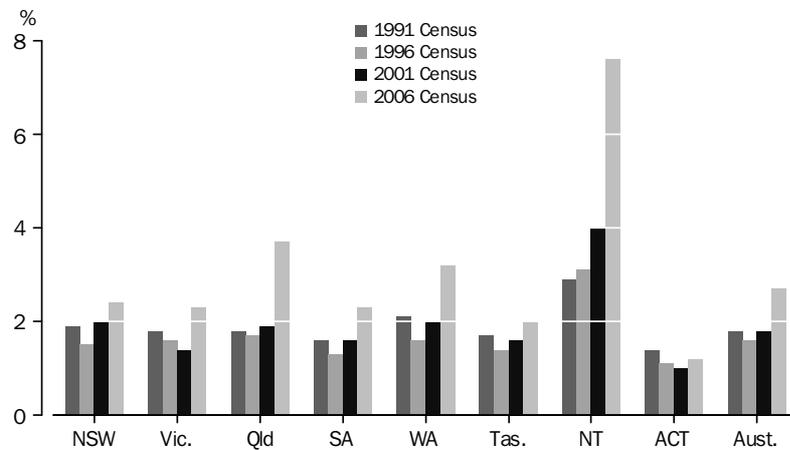
(a) The estimates were calculated from samples and are subject to sampling error, indicated by SEs provided. See the technical Note for more information on interpreting SEs.

(b) Care should be taken when comparing 2006 estimates with previous years due to changes made to PES estimation and because of the inclusion of remote areas and discrete Indigenous communities for the first time in the 2006 PES.

ESTIMATES OF NET UNDERCOUNT *continued*

AUSTRALIA, STATES AND TERRITORIES *continued*

2.2 NET UNDERCOUNT RATE, State/territory of usual residence



CAPITAL CITY AND BALANCE OF STATE/TERRITORY

The regional differences in net undercount in the 2006 Census for capital cities and the balance of each state/territory are given in table 2.3, table 2.4 and graph 2.5. Capital cities are defined as the capital city statistical division in each state and the territories. Balance of state/territory areas comprise all statistical divisions outside the capital city statistical division boundaries.

Different problems are encountered in enumerating people in urban and rural areas, and these are reflected to a certain extent in the capital city/balance of state net undercount rates for all states and territories. In urban areas, locating dwellings is generally easier, but contacting the occupants and getting their cooperation can be more difficult. In rural and remote areas where dwellings may be scattered over a wider area, locating the dwellings can cause considerable difficulties. In 2006, in all states and territories other than New South Wales and Victoria, the net undercount rate was higher in balance of state than in the capital city.

The total net undercount rates in 2006 were 3.0% for balance of state compared to 2.5% for capital cities. This difference is less than in 2001, when the net undercount rate for balance of state was 2.6% compared to 1.3% for capital cities. Northern Territory shows the largest difference in net undercount rate between the capital city and balance of state/territory.

ESTIMATES OF NET UNDERCOUNT *continued*

CAPITAL CITY AND
BALANCE OF
STATE/TERRITORY *continued*

2.3 NET UNDERCOUNT, State/territory of usual residence—By capital city/balance of state/territory

	Capital city		Balance		Total	
	no.	SE	no.	SE	no.	SE
New South Wales	124 934	22 674	32 644	15 062	157 578	25 679
Victoria	86 635	19 279	26 962	10 328	113 596	21 009
Queensland	34 258	15 468	114 151	18 070	148 409	18 181
South Australia	24 536	6 090	11 746	3 356	36 281	6 709
Western Australia	40 601	10 124	23 549	7 005	64 150	11 754
Tasmania	2 533	2 236	7 002	2 178	9 535	2 984
Northern Territory	6 247	2 065	9 663	2 590	15 909	3 300
Australian Capital Territory(a)	4 027	3 189	na	na	4 027	3 189
Australia	323 770	35 516	225 716	26 921	549 486	39 908

na not available

(a) Balance of Australian Capital Territory has been included with capital city.

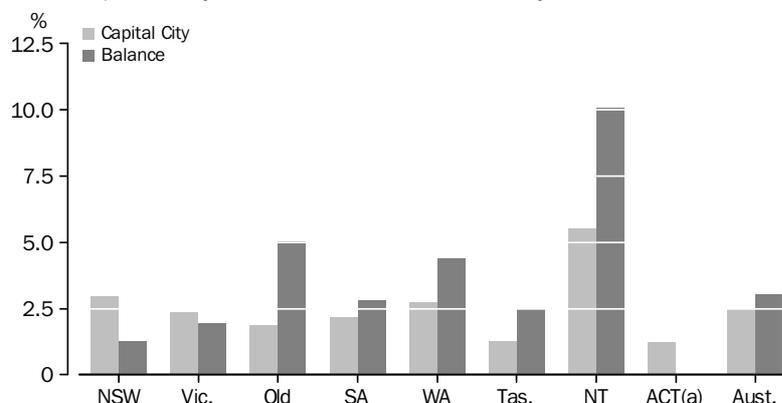
2.4 NET UNDERCOUNT, State/territory of usual residence—By capital city/balance of state/territory

	Capital city		Balance		Total	
	%	SE	%	SE	%	SE
New South Wales	3.0	0.5	1.3	0.6	2.4	0.4
Victoria	2.4	0.5	2.0	0.7	2.3	0.4
Queensland	1.9	0.8	5.1	0.8	3.7	0.4
South Australia	2.2	0.5	2.8	0.8	2.3	0.4
Western Australia	2.7	0.7	4.4	1.3	3.2	0.6
Tasmania	1.3	1.1	2.5	0.8	2.0	0.6
Northern Territory	5.5	1.7	10.1	2.4	7.6	1.5
Australian Capital Territory(a)	1.2	1.0	na	na	1.2	1.0
Australia	2.5	0.3	3.0	0.4	2.7	0.2

na not available

(a) Balance of Australian Capital Territory has been included with capital city.

2.5 NET UNDERCOUNT RATE, State/territory of usual residence—By capital city/balance of state/territory



(a) Balance of Australian Capital Territory has been included with capital city.

ESTIMATES OF NET UNDERCOUNT *continued*

AGE AND SEX

The likelihood of enumerating a person in the Census is related to the age and sex of that person. As has been observed in previous censuses in Australia, as well as censuses overseas, young adult males are the group least likely to be enumerated in the Census. Young adult females are the next least likely to be enumerated in the Census. Older adults are much more likely to be enumerated than younger adults.

Tables 2.6 and 2.7 and graph 2.8 confirm these patterns for the 2006 Australian Census. In particular, young adult males had the highest net undercount rates at 8.1% for males aged 25-29 years and 7.5% for males aged 20-24 years. The net undercount rate for young adult females aged 20-24 years has significantly increased since the 2001 Census (2.5% in 2001 to 6.0% in 2006). More generally, males had a higher net undercount rate (3.3%) than females (2.1%), and children aged 0-9 years had higher net undercount rates than children aged 10-14 years. The lowest net undercount rates were for people aged 55 years and over.

2.6 NET UNDERCOUNT, Sex by age group

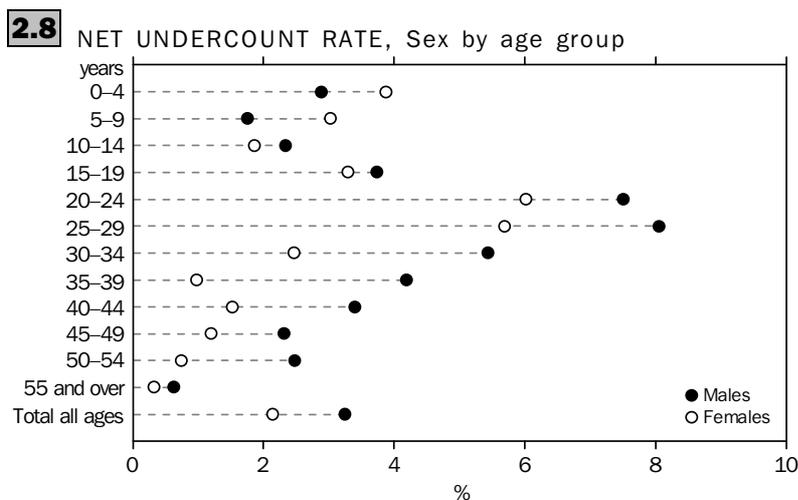
Age	Males		Females		Persons	
	no.	SE	no.	SE	no.	SE
0-4 yrs	19 234	5 674	24 706	5 985	43 941	8 576
5-9 yrs	11 949	6 000	19 836	5 663	31 785	8 429
10-14 yrs	16 798	5 912	12 638	5 842	29 435	8 438
15-19 yrs	27 014	6 464	22 492	5 652	49 506	8 389
20-24 yrs	55 353	9 258	42 570	7 889	97 923	12 687
25-29 yrs	55 683	8 516	38 642	6 706	94 325	11 372
30-34 yrs	39 324	7 338	18 122	6 519	57 446	10 200
35-39 yrs	31 286	6 899	7 396	5 990	38 682	9 475
40-44 yrs	25 429	6 597	11 565	5 976	36 994	9 135
45-49 yrs	16 790	6 001	8 887	5 606	25 678	8 490
50-54 yrs	16 467	5 439	4 968	5 718	21 434	7 934
55 yrs and over	14 197	7 205	8 139	7 511	22 336	11 012
Total all ages	329 524	24 120	219 962	21 994	549 486	39 908

2.7 NET UNDERCOUNT RATE, Sex by age group

Age	Males		Females		Persons	
	%	SE	%	SE	%	SE
0-4 yrs	2.9	0.8	3.9	0.9	3.4	0.6
5-9 yrs	1.8	0.9	3.0	0.8	2.4	0.6
10-14 yrs	2.3	0.8	1.9	0.9	2.1	0.6
15-19 yrs	3.7	0.9	3.3	0.8	3.5	0.6
20-24 yrs	7.5	1.2	6.0	1.1	6.8	0.8
25-29 yrs	8.1	1.1	5.7	0.9	6.9	0.8
30-34 yrs	5.4	1.0	2.5	0.9	3.9	0.7
35-39 yrs	4.2	0.9	1.0	0.8	2.6	0.6
40-44 yrs	3.4	0.9	1.5	0.8	2.5	0.6
45-49 yrs	2.3	0.8	1.2	0.7	1.7	0.6
50-54 yrs	2.5	0.8	0.7	0.8	1.6	0.6
55 yrs and over	0.6	0.3	0.3	0.3	0.5	0.2
Total all ages	3.3	0.2	2.1	0.2	2.7	0.2

ESTIMATES OF NET UNDERCOUNT *continued*

AGE AND SEX *continued*



REGISTERED MARITAL STATUS

Tables 2.9 and 2.10 set out the net undercount estimates and rates by registered marital status and sex.

The net undercount rates were highest for people identified as never married (4.6%). This has increased from 3.1% since the 2001 Census. The lowest net undercount rate was recorded for persons who were married (1.0%).

2.9 NET UNDERCOUNT, Registered marital status by sex

Registered marital status	Males		Females		Persons	
	no.	SE	no.	SE	no.	SE
Never married(a)	253 172	20 194	186 991	17 946	440 163	31 088
Widowed, divorced or separated	28 528	9 622	4 361	9 872	32 889	14 227
Married	47 824	10 575	28 610	10 356	76 434	19 063
Total persons	329 524	24 120	219 962	21 994	549 486	39 908

(a) Includes those who are living with a de facto partner and have never been in a registered marriage.

2.10 NET UNDERCOUNT RATE, Registered marital status by sex

Registered marital status	Males		Females		Persons	
	%	SE	%	SE	%	SE
Never married(a)	4.9	0.4	4.1	0.4	4.6	0.3
Widowed, divorced or separated	2.8	0.9	0.3	0.6	1.2	0.5
Married	1.2	0.3	0.7	0.3	1.0	0.2
Total persons	3.3	0.2	2.1	0.2	2.7	0.2

(a) Includes those who are living with a de facto partner and have never been in a registered marriage.

EXPLANATORY NOTES

SCOPE OF THE 2006 CENSUS

- 1** The 2006 Census of Population and Housing was held on 8 August 2006. The objective of the Census was to measure the number and key characteristics of people in Australia on Census night, and the dwellings in which they live.
- 2** The aim of the Census was to count every person who spent Census night in Australia. This includes Australian residents in Antarctica and people in the territories of Jervis Bay, Cocos (Keeling) Islands and Christmas Island. The other Australian External Territories, Norfolk Island and minor islands such as Heard and McDonald Islands, are outside the scope of the Australian Census. The only people who spend Census night in Australia but are excluded from the Census are foreign diplomats and their families.
- 3** Visitors to Australia are counted regardless of how long they have been in the country or how long they plan to stay. Australian residents out of the country on Census night are out of scope of the Census. People outside Australia who are not required to undertake migration formalities, such as those on oil and gas rigs off the Australian coast, are included. People were counted where they stayed on Census night. This means that the Census was conducted on an 'actual location' or 'place of enumeration' basis.
- 4** All private dwellings, except diplomatic dwellings, are included in the Census, whether occupied or unoccupied. Caravans in caravan parks, manufactured homes in manufactured home estates and self-care units in accommodation for the retired or the aged are counted only if occupied. Occupied non-private dwellings, such as hospitals, prisons, hotels, etc., are also included.
- 5** Details about the 2006 Census content, collection operations, confidentiality and privacy protection, processing and evaluation activities are contained in *2006 Census Nature and Content* (cat no. 2008.0).

SCOPE AND COVERAGE OF THE 2006 PES

- 6** 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 counted in the Census – thus this set of people constitute the theoretical scope of the PES. 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.
- 7** The PES also 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 (in private dwellings).
- 8** The 2006 PES included remote areas and discrete Indigenous communities for the first time. Previous PESs have excluded these areas from the coverage of the survey due to operational reasons – mainly the additional cost and the need to use the same local contacts as Census, which was considered likely to compromise the independence of the PES. Inclusion of these communities in 2006 ensures the geographic coverage of the PES is more complete than it has been in the past. In practice, the PES is used to produce estimates for the full Census scope, even though its actual coverage is somewhat less.

EXPLANATORY NOTES *continued*

SCOPE AND COVERAGE OF THE 2006 PES *continued*

9 The PES interview process determines whether each person in the sample should have been counted in the Census, and a few of the various categories in which they should have been counted (such as age, sex, Indigenous status, country of birth, region of usual residence, etc). The match and search process involves comparing dwellings and people enumerated in the PES with dwellings and people counted in the Census. The objective is to determine how many times each person in the PES sample was actually counted in the Census. PES output processing and estimation combines and weights results from the match and search process to produce an estimate of the number of people who should have been counted in the Census.

PES ESTIMATION

10 Following the 2001 PES, a review of the PES estimation method was commissioned to develop an estimator for the PES that adjusts adequately for non-response and non-coverage in PES, and for miscounting in the Census. The estimator used in the 2001 PES did not fully account for people missed in both Census and PES.

11 One outcome of the review was the development of a new estimator, Prediction Regression (PREG), for use in the person weighting stage of the 2006 PES. Unlike the estimator used in 2001, the weight adjustment applied by PREG to each PES person does not depend on their Census response. The PREG estimator also allows for differences in reporting of a person's characteristics (e.g. age, Indigenous status) between PES and Census and enables a correction for errors in Census imputation.

12 Another outcome of the review was an improved approach to adjusting for dwellings found in the Census but not responding in the PES. This and other changes described in this publication had a larger impact on the final estimates than the change in estimator.

13 The 2006 PES introduced a number of methodological improvements that allow more aspects of the Census to be appropriately measured. A side-effect of these changes is that the sampling error on the overall population estimates has increased, with previously unmeasured potential errors now being measured by the survey and included in adjustments.

14 Further details of PES estimation are in Appendix 3. Technical details of the PREG estimator were reported in *Research Paper: An estimating equation approach to Census coverage adjustment* (cat. no. 1351.0.55.019) released on 7 May 2007. Further descriptions of PES estimation can also be found in *Information Paper: Measuring Net Undercount in the 2006 Population Census, Australia* (cat. no. 2940.0.55.001) released on 7 May 2007.

CENSUS COUNTS *Usual residence*

15 The Australian Census counts people where they actually were on Census night, rather than where they usually live. There is, however, a need for data based on place of usual residence, and Census counts are available on this basis.

16 For usual residents of Australia, 'place of usual residence' for the 2006 Census is defined as the address at which a person has lived or intends to live for six months or more in 2006. While for most people their usual residence was the same as their actual location on Census night, some people spent Census night at a place other than where they usually lived. Thus, their 'place of enumeration' and their 'place of usual residence' were different.

17 People visiting Australia on Census night are included in the Census counts on a place of enumeration basis but not those on a place of usual residence basis.

18 Usual residents of Australia who are temporarily overseas on Census night are not included in Census counts on either a place of usual residence or place of enumeration basis. However, counts of these people are accounted for in the estimated resident population of Australia (ERP). For information on the calculation of ERP, see the ABS

EXPLANATORY NOTES *continued*

Usual residence continued

publication *Australian Demographic Statistics, December quarter 2006* (cat. no. 3101.0), released on 5 June 2007.

19 Estimates presented in this publication are on a place of usual residence basis.

Census late return and imputed dwellings

20 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.

21 The non-contact sector also contains dwellings which were non-responding in the Census – that is, dwellings where the Census never obtained a return, and which could not be established as having been unoccupied on Census night. These non-response dwellings are given imputed values (using 'hot-deck' imputation) during Census processing, based in many cases on information provided by the Census collector about the dwelling and its residents. Inevitably, the imputed values, at the dwelling and aggregate level, differ from the true, but unknown, values. The imputed records constitute the majority of the Census non-contact sector records; late returns (as defined here) are only a small component of the overall Census non-contact sector. Given that late returns prompted by PES would otherwise have been classed as non-response in the Census, the PES sample is representative of the whole non-contact sector, even though it cannot split late returns from non-responses in a manner comparable to the Census.

Resolution of Census not-stated values

22 For Census purposes, age, sex, marital status and state of usual residence are imputed (statistical process for predicting values where no response was provided) during Census processing where these items have been left blank, including where a whole person record has been imputed. Missing values for any other items remain 'not-stated' in the final version of Census counts.

23 The PES uses Census data items to form benchmark categories for weighting and estimation purposes. Two such data items are Indigenous status and country of birth. In cases where these items have been left blank in the Census, a value was imputed during PES processing so that these items could be used for 2006 PES benchmarks.

24 The imputation method used involved imputing both variables together. This enforced the assumption that anyone who is imputed as Indigenous is born in Australia.

25 For benchmarking purposes, 12 categories of persons were defined:

- Indigenous, Australian-born
- non-Indigenous, Australian-born
- non-Indigenous, born in other country – 1 to 9 ranked (in terms of population) countries of birth from the 2006 Census
- non-Indigenous, born elsewhere.

26 The full range of 10 non-Australian country of birth classes were used in benchmarking at the Australia level; for most regions in Australia they were collapsed to 2 classes.

27 These are non-overlapping categories, but persons with not-stated values in the Census may be imputed as partially in two or more of the categories (with the sum of all categories adding to 1 for each person).

28 Imputation was performed separately for each person within non-overlapping imputation classes, whereby data are imputed by the proportion of stated values of respondents in the same imputation class. In order of importance, these classes were:

EXPLANATORY NOTES *continued*

*Resolution of Census
not-stated values continued*

- Statistical Local Area – geographical areas, in almost all cases, identical with whole legal local government areas, comprising cities, district councils, community government councils, municipalities, shires, rural cities, and towns.
- Census form type
- Age Group – 5 year age groups to 75+
- Sex.

Differences in classification

29 Occasionally the answers obtained for a person in the PES interview were not consistent with the answers obtained for the equivalent questions in the Census. There are a number of reasons a response may differ, including:

- a person may have difficulty answering a question for themselves or another household member, either in the Census or the PES
- a person may interpret the question differently in the Census, where forms are self-completed for the majority of Australia, than in the PES, which is administered by an interviewer
- different people may provide the Census and PES answers
- the correct answer could change between the Census and PES. Changes in age can be taken into account using the actual date of birth, but other changes, for example if the person married or divorced, may not be identified.
- the Census may contain a 'not-stated', or imputed, response while the PES will have a valid response.

30 The PREG estimator adjusts the weights of responding persons according to their PES-reported categories. This ensures that PES persons not responding in the Census get the same weight adjustments as similar persons who did respond. The Census categories are used in producing sample-based estimates of numbers of persons counted in the Census. The weighting ensures that these estimates match the actual Census counts for all benchmark categories.

CORRECTION FOR CENSUS
IMPUTATION ERROR IN THE
NON-CONTACT SECTOR

31 The standard error (SE) on the PES estimate of the population in the non-contact sector has been calculated, and used in comparing the accuracy of the PES estimate with that of the unadjusted Census count for this sector. Analysis showed that the overall population estimates are considerably more accurate if the PES estimates are used for this sector rather than using the Census counts (which in this sector were mostly imputed).

32 The decision to use the PES to estimate the population in the non-contact sector allows the sampling error in this sector to be measured and included in the published standard errors. This is the major reason that SEs for the 2006 PES exceed those of previous PESs. The increased SE represents uncertainty in estimating the contribution of the non-contact sector. This is a much better outcome than the alternative of including an unmeasured and potentially major bias arising from unadjusted inaccuracies in the Census imputation process. The ability to measure this sector is one of the key improvements made in the 2006 PES.

33 Net undercount estimates presented in this publication incorporate the PES estimates for the population in the non-contact sector.

INDEPENDENCE FROM THE
CENSUS

34 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

EXPLANATORY NOTES *continued*

INDEPENDENCE FROM THE CENSUS *continued*

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.

35 Steps were taken to maintain the operational independence of the 2006 PES from the Census at every stage of the survey, including enumeration, processing and administration. These steps included:

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

36 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.

37 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.

38 The Census can also be used to form an estimate of the percentage of dwellings and people missed by the PES. The PES excludes non-private dwellings (hotels, motels, hospitals) for operational reasons. The PES is also conducted several weeks after the Census, so a respondent's recollection of their location on Census night may not be entirely accurate. Census has special procedures for enumerating homeless people, while the PES is essentially a survey of dwellings and the people who reside in them. Thus the Census may include some dwellings and people that the PES misses. PES estimation implicitly accounts for the dwellings and people missed in the PES but counted in the Census.

39 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.

40 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.

RELIABILITY OF THE UNDERCOUNT ESTIMATES

41 As the estimates of undercount are based on data from a sample survey, they are subject to sampling error. Some of the estimates presented in this publication have high SEs, and these estimates should be used with caution. For more information about SEs see the Technical Note.

42 The estimates of undercount are also subject to non-sampling errors which occur in all collections, censuses and surveys. Examples of this kind of error include imperfections in reporting by respondents, errors made in collection of data, and errors made in processing the data. Every effort is made in the Census and PES to reduce non-sampling error to a minimum by careful design of forms, training and supervision of collectors and interviewers, and by using efficient operating procedures. Types of

EXPLANATORY NOTES *continued*

RELIABILITY OF THE UNDERCOUNT ESTIMATES *continued*

Census Data Quality

non-sampling error arising from the way the PES is conducted and the way estimates are derived from the survey are discussed below.

43 A potential weakness in the PES method is its dependence on matching as a means of deciding whether or not a given person or dwelling has been counted in the Census. The difficulties associated with the matching process mean that there is a risk of failing to match people who are actually included in the Census. The effect of not matching when there should have been a match would be to overstate net undercount in the Census.

44 While the Census and PES are conducted independently of each other, they are very similar in many respects. Thus, some weaknesses in the Census may also be shared by the PES leading to an understatement of net undercount. For example, dwellings missed by a Census collector are often difficult to find and so could possibly be missed by a PES interviewer as well. Also, people who avoid being included in the Census may also avoid being included in the PES. The use of benchmarks in estimation is designed to control the effect of this 'correlation bias'.

45 Census data are subject to a number of inaccuracies resulting from errors by respondents or mistakes in collection or processing. Whilst many of these are corrected by careful processing procedures, some still remain. The effect of the remaining errors is generally slight, although it may be more important for smaller groups in the population. Using the characteristics reported in PES is intended to correct for such errors.

46 The main kinds of errors occurring in the Census are:

- Partial non-response: In some cases where an answer is not provided to a question an answer is imputed (often from other information on the form). In other cases a 'not stated' code is allocated.
- Processing error: While such errors can occur in any processing system, quality management is used continuously to improve the quality of processed data, and to identify and correct data of unacceptable quality.
- Random adjustment: Table cells containing small values are randomly adjusted or suppressed to avoid releasing information about particular individuals, families, or households. The effects of these changes is statistically insignificant.
- Respondent error: Because processing procedures cannot detect or repair all errors made by people in completing their forms, some remain in final data.
- Undercount: Although the Census aims to count each person, there are some people who are missed and others are counted more than once.

47 For further information on sources of error in the Census, refer to the appropriate entries in the *2006 Census Dictionary* (cat.no. 2901.0) released on 26 May 2006.

APPENDIX 1 CENSUS POST ENUMERATION SURVEY

BACKGROUND

The Census Post Enumeration Survey (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 people usually complete the forms themselves. 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. Form completion by interviewers also helps ensure no questions are left unanswered.

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 undercount rate for persons in the PES sample is thus available. PES estimation then attaches a 'weight' to each person in the sample so that together they represent the whole population - accounting for underenumeration in both Census and PES. The outcome is a reliable estimate of net undercount in the Census.

SAMPLE DESIGN

In the PES, private dwellings (houses, flats, etc.) and discrete Indigenous communities are separately identified and sampled.

Table A1.1 shows the sampling fractions used in the 2006 PES sample design for each state and territory. These sampling fractions represent the probability that a dwelling in a state or territory is selected in the PES sample.

A1.1 PES STATE SAMPLING FRACTIONS

	NSW	Vic.	Qld	SA	WA	Tas.	NT	ACT
PES sampling fraction	1 in 321	1 in 270	1 in 239	1 in 149	1 in 165	1 in 90	1 in 37	1 in 86

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, Australia is first divided into around 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% (based on 2001 Census data) 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 and CDs. 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.

This sampling strategy ensured that within each state or territory, all private dwellings and discrete Indigenous community dwellings had the same chance of selection.

FIELD RESPONSE

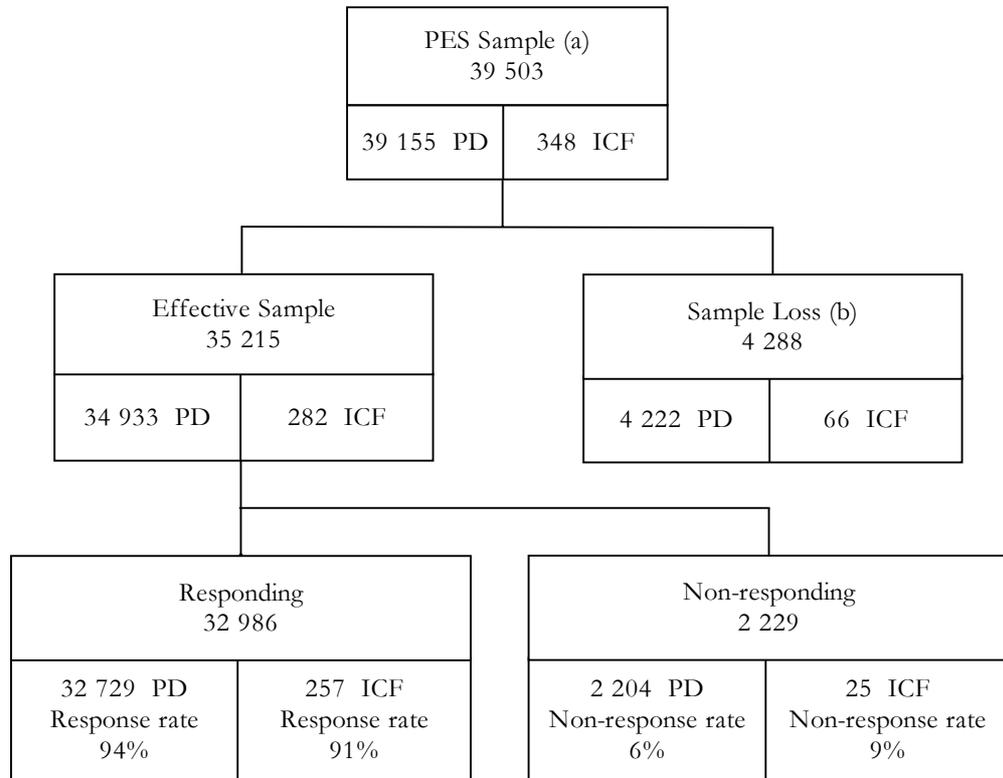
For the private dwelling (PD) sample, interviews were conducted between 1 September and 24 September 2006, with an initial sample of just under 40,000 PDs. For the ICF component, interviews were conducted between 11 September and 12 October 2006, with a sample of around 350 dwellings.

After allowing for sample loss (e.g. dwellings selected with no residents in scope, vacant dwellings, dwellings under construction), the combined effective sample was 35,215 dwellings. The enumeration of the 2006 PES PD sample achieved a response rate of 94% (compared to 96% in 2001). The response rate for the ICF sample was 91%.

The diagram below shows the number of dwellings in the two components of the PES sample for the main response types.

APPENDIX 1 CENSUS POST ENUMERATION SURVEY *continued*

Figure 1. DWELLINGS, PES Sample by main response type



(a) Number of dwellings selected.

(b) Sample loss includes: Dwellings in which all persons were out on scope/coverage; Vacant dwellings; Dwellings converted to non-dwellings; Derelict dwellings; Demolished dwellings; Dwellings under construction.

FIELD RESPONSE *continued*

Table A1.2 shows the expected and actual dwelling selections for the PES sample, by PD and ICF sample type.

A1.2 DWELLINGS, PES sample—Expected and actual dwelling selections

	PD SAMPLE				ICF SAMPLE				Total actual selections
	Expected selections	Actual selections	Expected fully responding	Actual fully responding	Expected selections	Actual selections	Expected fully responding	Actual fully responding	
	no.	no.	no.	no.	no.	no.	no.	no.	no.
New South Wales	8 632	8 750	7 050	7 333	8 750
Victoria	7 775	7 698	6 500	6 450	7 698
Queensland	6 912	7 001	5 705	5 914	20	24	13	16	7 025
South Australia	4 521	4 652	3 835	3 800	8	9	6	6	4 661
Western Australia	4 994	5 200	4 155	4 290	27	30	13	18	5 230
Tasmania	2 330	2 437	1 950	1 971	2 437
Northern Territory	1 625	1 914	1 235	1 622	275	285	180	217	2 199
Australian Capital Territory	1 415	1 503	1 250	1 349	1 503
Australia	38 204	39 155	31 680	32 729	330	348	212	257	39 503

.. not applicable

APPENDIX 1 CENSUS POST ENUMERATION SURVEY *continued*

Census response type

Although all persons resident in Australia should be counted in the Census, not all dwellings can expect to receive a Census form. This is because not all dwellings are habitable or, in the case of a diplomatic dwelling, do not contain people within the scope of the Census.

Census defines 'private dwellings' as structures established for self-contained accommodation. Private dwellings may be 'occupied' or 'unoccupied', where:

- 'Occupied' refers to private dwellings that are occupied by one or more persons on Census night. Non-permanent, removable structures such as caravans, tents, manufactured homes or houseboats are treated as private dwellings and included in the Census only if they are occupied on Census night. Occupied improvised dwellings, such as sheds and garages, are also included in the Census.
- 'Unoccupied' refers to private dwellings that are capable of being lived in but are not occupied on Census night. This includes unoccupied holiday houses, vacant houses to let, and unoccupied apartments. Unoccupied non-permanent or improvised structures, such as caravans, converted garages, tents, manufactured homes and houseboats, are not counted in the Census. Houses under construction and derelict houses are not included in the Census.

Private dwellings may also be classified as a 'non-contact' or 'refusal' to the Census, where:

- 'Non-contact' refers to private dwellings where the Census collector was unable to make contact with a householder after five visits, and was unable to verify that the dwelling was unoccupied on Census night.
- 'Refusal' refers to private dwellings where the householder has refused to participate in the Census.

For PES purposes, dwellings that provide a Census return after the commencement of PES enumeration are flagged as 'late returns' and have special treatment during PES processing. See paragraph 20 of the Explanatory Notes for further information.

During Census processing, non-contact and refusal dwellings will have the number of usual residents for that dwelling imputed, as well as some personal characteristics. See paragraph 21 of the Explanatory Notes for further information.

Table A1.3 shows the Census count of dwellings by PES response and categories of Census response type.

A1.3 DWELLINGS, PES response by Census response type

	CENSUS RESPONSE TYPE							Total
	Occupied	Un-occupied	Late returns	Imputed records	Missing	ICF not missing	ICF missing	
<i>PES Response</i>	no.	no.	no.	no.	no.	no.	no.	no.
Responding	29 782	1 339	412	819	377	248	9	32 986
Non-Responding	1 472	365	26	197	144	23	2	2 229
Vacant dwelling	623	1 681	16	181	544	46	11	3 102
Other Sample								
Loss	169	211	1	32	764	6	3	1 186
Total	32 046	3 596	455	1 229	1 829	323	25	39 503

APPENDIX 1 CENSUS POST ENUMERATION SURVEY *continued*

Census response type
continued

In total, just over 88,000 people responded to the PES. Table A1.4 shows the count of PES respondents by state and sample type.

A1.4 PERSONS, Number of PES respondents by state and sample type

	SAMPLE TYPE		
	<i>PD</i>	<i>ICF</i>	<i>Total</i>
	no.	no.	no.
New South Wales	19 668	. .	19 668
Victoria	17 192	. .	17 192
Queensland	16 092	57	16 149
South Australia	9 592	42	9 634
Western Australia	11 273	95	11 368
Tasmania	4 879	. .	4 879
Northern Territory	4 431	1 239	5 670
Australian Capital Territory	3 685	. .	3 685
Australia	86 812	1 433	88 245

. . not applicable

For further information on the field phase of the 2006 PES, see *Information Paper: Measuring Net Undercount in the 2006 Population Census, Australia* (cat. no. 2940.0.55.001) released on 7 May 2007.

APPENDIX 2 PES MATCH AND SEARCH PROCESSING

THE PES MATCH AND SEARCH SYSTEM (MSS)

The main PES processing facility was called the 'Match and Search System' (MSS), a software system built specifically for PES processing. As the name suggests, the MSS allowed processors to search, view, compare, and record matches between PES and Census data. PES processors used the MSS to record matches of dwellings and people between PES and Census, and to search for people included on Census forms at all alternative addresses provided.

DWELLING MATCHING

An attempt was made to determine whether every dwelling selected in the PES had been counted in the Census. The process involved searching through Census information (e.g. images of Census Collector Record Books) to locate the address of the PES dwelling. This searching was performed at the CD level. The CD number was the key search unit for the MSS.

Strict procedures were set to ensure consistency and accuracy when performing the dwelling matching. In order for a dwelling non-match to be recorded, all reasonable attempts had to be made to locate the dwelling in the expected CD, or the neighbouring CDs. All PES dwelling non-matches were confirmed by the PES processing supervisor.

PERSON MATCHING

Once a PES dwelling had been matched to the Census, person matching was carried out for all people in the PES dwelling. In the majority of cases the same people were in the same dwelling during both the PES and Census enumeration, and hence a dwelling match was found to be a strong indicator that the person matching would be successful within the dwelling. When this was not the case, a further search for the PES respondent was carried out at any alternative addresses provided during the PES interview.

In situations where a corresponding Census dwelling could not be found for a PES dwelling, other dwellings in the CD were searched in an attempt to locate the residents of the PES dwelling.

To determine if a PES person matched to a Census person, a comparison was made of responses to key variables common to both the PES and Census forms. A field match code (FMC) was recorded to indicate the strength of the match against each of the following fields:

- Name
- Sex
- Date of Birth
- Age
- Marital Status
- Indigenous Status
- Country of Birth
- Relationship in household.

The MSS incorporated an algorithm which combined field match codes to produce a final person match code (PMC), of which there are approximately 20. The MSS required manual confirmation of the PMC calculation prior to progressing to the next record. Fields, such as name, sex, and age had a greater importance in determining the PMC. Other variables, such as marital status, Indigenous status, country of birth and relationship in household were considered less important in determining a PMC but still played a vital part in informing the decision of whether a match existed.

QUALITY ASSURANCE AND ADJUDICATION PROCESSES

The Quality Assurance (QA) process for the 2006 PES included a 100% recoding exercise by a different processor. There was no identifier on the workloads that allowed the PES processors to know whether they were processing an 'original' or QA workload. Any discrepancies at either the dwelling or person level found between the original and QA processing were automatically flagged by the MSS, and both records were presented to the PES supervisor for adjudication. Adjudication processing was performed in the MSS to ensure consistent procedures were applied when evaluating or reprocessing records.

APPENDIX 2 PES MATCH AND SEARCH PROCESSING *continued*

QUALITY ASSURANCE AND ADJUDICATION PROCESSES *continued*

The PES supervisor reviewed the discrepancies between the original and QA coding and decided which was correct. If the PES supervisor decided that both the original and QA coding were incorrect, the PES record was reprocessed by the PES supervisor.

The QA process was also useful in identifying potential processing issues or areas where processors were having difficulty. This allowed ongoing constructive feedback to be provided to the PES processors and contributed to the overall quality assurance of the PES processing.

APPENDIX 3 PES ESTIMATION

OVERVIEW

The PES interview process determines whether each person in the sample should have been counted in the Census, and the category in which they should have been counted (such as age, sex, Indigenous status, region of usual residence). The match and search process determines how many times each person in the PES sample was actually counted in the Census (and the categories in which they were counted). PES output processing and estimation combines and weights results from the match and search process 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 (including imputed persons for non-responding dwellings).

Following the 2001 PES, a review of the PES estimation method was commissioned to develop an estimator for the PES that adjusts adequately for non-response and non-coverage in PES and for miscounting in the Census. The estimator used in the 2001 PES did not fully account for people missed in both Census and PES.

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

Technical details of the PREG estimator were reported in *Research Paper: An estimating equation approach to Census coverage adjustment* (cat. no 1351.0.55.019) released on 7 May 2007.

PES ESTIMATION

Broadly speaking, PES estimation involves assigning a 'weight' to each selected PES dwelling and then to each person for whom a PES response was obtained. 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.

The net undercount for a category of person is obtained by taking the PES estimate of the number of people in the category who should have been counted and subtracting the Census count of the number of people (in the category). Net undercount for a category of person is the net result of the PES estimate of gross undercount, gross overcount, differences in classification between the PES and Census (e.g. age, sex, Indigenous status) and imputation error in the Census.

Dwelling weighting

Dwelling weighting for the 2006 PES comprises two stages. For private dwellings selected in PES that were found in the Census, the first stage of dwelling weighting adjusts the PES selection weight such that the adjusted weights add up to the Census private dwelling count. These adjustments are made within a set of Census dwelling categories based on region, dwelling structure and type of Census response (single occupant, multiple occupant, unoccupied, non-contact sector). A first-stage weight adjustment is also applied to private dwellings selected in PES that were missed in the Census. These dwellings receive the average adjustment across all categories of type of Census response (for the same region). For dwellings in discrete Indigenous communities, a similar first-stage weight adjustment is applied, except that Census dwelling categories are based only on state or territory of the community.

The second stage of dwelling weighting applies a non-response adjustment so that the responding PES dwellings represent other dwellings from which no response was obtained. These adjustments are also made within a set of Census dwelling categories based on region, dwelling structure and type of Census response. For discrete Indigenous communities, a single overall adjustment is applied in each state and territory.

APPENDIX 3 PES ESTIMATION *continued*

Dwelling weighting continued

Improvements made to the PES dwelling weighting for 2006 include:

- New Census response categories have been introduced: Occupied dwellings are now split into 'single-occupant' and 'multiple-occupant', and a category has been introduced for discrete Indigenous community dwellings.
- Dwelling structure is introduced as a benchmark category in weighting (for non-ICF dwellings matched to the Census).
- Partially-responding dwellings are treated as responding dwellings, so that the responding people in these dwellings can be included in person weighting and estimates.
- Weighting for 'Late Returns and Imputed Dwellings' is now designed to make these dwellings representative of this combined group in the Census.
- Northern Territory is treated as two regions (Darwin/Balance of Territory).

Person weighting

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 and discrete Indigenous Community 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 (from the PREG estimator) are chosen to ensure that the PES estimates of people counted in private dwellings and discrete Indigenous Community 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 does not depend on whether they responded in the Census, but only on characteristics of the person as reported in the PES. In 2001, the PES estimator used for weighting made weight adjustments reliant on whether a person responded in the Census, and assumed people were recorded in the same categories in PES and Census (i.e. same age, sex, region, etc.).

As a final step in weight adjustment, the initial person weights are adjusted (using the PREG estimator) so that the PES estimates represent people in non-private dwellings as well as private and ICF dwellings. This final step uses region, age and sex only, as information from non-private dwellings for the other benchmark variables is not sufficiently reliable.

For the initial stage of person weighting, data items used to form benchmark categories include Indigenous Status and Country of Birth. However, one or both of these may be 'not-stated' (i.e. blank) for a number of Census records where people have not responded to these items in the Census. In these cases a value needed to be imputed so that these items could be used for initial person weighting. See paragraphs 22 to 28 of the Explanatory Notes.

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.

APPENDIX 3 PES ESTIMATION *continued*

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

The non-contact sector also contains dwellings which were non-responding in the Census – that is, dwellings where the Census never obtained a return, and which could not be established as having been unoccupied on Census night. These non-response dwellings are given imputed values (using 'hot-deck' imputation) during Census processing, based in many cases on information provided by the Census collector about the dwelling and its residents. Inevitably, the imputed values, at the dwelling and aggregate level, differ from the true, but unknown, values. The imputed records constitute the majority of the Census non-contact sector records; late returns (as defined here) are only a small component of the overall Census non-contact sector. Given that late returns prompted by PES would otherwise have been classed as non-response in the Census, the PES sample is representative of the whole non-contact sector, even though it cannot split late returns from non-responses in a manner comparable to the Census.

In previous Censuses, only the Census contact sector was corrected for under- and over-count by using the PES estimates. Effectively, late returns and imputed dwellings (the Census non-contact sector for PES purposes) were treated as being reported accurately. While this assumption is imperfect, this was considered the most feasible way to calculate accurate net undercount estimates at the level of detail needed for producing estimated resident population counts in Australia.

For the 2006 PES, the person weighting step in PES processing calculates weights for all PES records, including those that correspond to the Census non-contact sector. Hence the PES can provide an estimate of the total population in Census late return and non-response dwellings on Census night. This is a change from previous PESs, in which persons selected in non-contact sector dwellings were excluded from matching and from estimation. The inclusion of these persons in the 2006 PES is an innovation made possible by the development of appropriate methods for representing them in estimation.

PES estimates of the population in the non-contact sector have relatively high sampling errors because of the small sample size (there are relatively few non-contact dwellings selected by chance in the PES sample); and also because person counts for this sector are not available to use as a weighting 'benchmark'. This lack of Census person counts also means that, while the dwelling weights used for the non-contact sector are estimated from the sector itself, the adjustments applied to provide final person weights depend strongly on information observed in the contact sector. This is a potential source of non-sampling error, as is any bias arising from peculiarities of the non-respondents in this sector. Both these sources of non-sampling error are expected to be small compared to the sampling error of the population estimate for the non-contact sector.

Using PES estimates for the population of the non-contact sector leads to a rise in the standard error of the overall population estimates. On the other hand, the alternative, where this sector is not measured by PES but is treated as accurately represented by the Census figures, can have a bias associated with Census imputation for non-response.

The standard error of the PES estimate of the population in the non-contact sector was used to compare the accuracy of the PES estimate with that of the unadjusted Census count for this sector. Analysis showed that the overall population estimates are considerably more accurate if the PES estimates are used for this sector rather than using the Census counts (which in this sector were mostly imputed).

As a result, estimates of net undercount for the 2006 Census incorporate the PES estimate of the number of people who should have been counted in the non-contact sector.

TECHNICAL NOTE SAMPLING ERRORS

SAMPLING ERRORS ASSOCIATED WITH STATISTICS PRODUCED FROM THE PES

1 Statistics produced from the PES are subject to sampling error. Since only a sample of dwellings is included in the PES, estimates derived from the survey may differ from figures which would have been obtained if all dwellings had been included in the survey. One measure of the likely difference is given by the standard error (SE) which indicates the extent to which an estimate might have varied by chance because only a sample was included.

2 The particular sample selected for the PES was only one of a number of possible samples. Each possible sample would yield different estimates. The SE measures the variation of all the possible sample estimates around the figures which would have been obtained if all dwellings had been included.

3 Given an estimate and the SE on that estimate, there are about two chances in three that the sample estimate will differ by less than one SE 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 SEs.

4 The following example illustrates the use of the concept of SE. If an estimate of 1.3% has a SE of 0.1 percentage points there are two chances in three that the figure that would have been obtained if all dwellings had been included in the sample is in the range $1.3\% \pm (1 \times 0.1\%)$ or 1.2% to 1.4%, and nineteen chances in twenty that the figure is in the range $1.3\% \pm (2 \times 0.1\%)$, that is, between 1.1% and 1.5%.

5 For ease of use, the SEs corresponding to the net undercount rates are given next to the rates in the tables throughout this publication.

SAMPLING ERRORS ON ESTIMATES OF DIFFERENCES

6 The sampling error on the difference between two estimates can be derived from their SEs. For the difference between two estimates x and y produced from the PES the SE of the difference may be approximated by the following formula:

$$SE(x - y) = \sqrt{[SE(x)]^2 + [SE(y)]^2}$$

7 This approximation will be exact for differences between estimates in different states, for Capital City versus Balance of state, or for differences between estimates from different Censuses. However, for estimates within the same region there will tend to be a negative correlation between the rates so that the approximation will tend to underestimate the true SE.

8 For example, if the estimates of the rate of net undercount for usual residents in NSW capital city and balance of state are 3.0% and 1.3%, with SEs of 0.5 and 0.6 percentage points respectively, using the formula above the SE on the difference (1.7 percentage points) is:

$$\sqrt{(0.5)^2 + (0.6)^2} = 0.78 \text{ percentage points}$$

9 Therefore there are nineteen chances in twenty that the difference between the rates of undercount for usual residents between these two regions is within the range $1.7 \pm (2 \times 0.78)$ or 0.14 to 3.26 percentage points.

GLOSSARY

Benchmark category	A category of dwelling or person for which the PES estimate of the Census count is forced to reproduce the actual Census count.
Capital city or balance of state/territory	Capital cities are defined as areas covered by the capital city statistical division in each state and territory. The balance of state comprises all statistical divisions in the state or territory other than the capital city statistical division.
Census collection district (CD)	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 (CAI)	Computer-assisted interviewing (CAI) is a method of data collection whereby responses are recorded directly into an electronic questionnaire on a notebook computer.
Contact sector	The Census contact sector comprises all Census dwellings, excluding late return and imputed dwellings.
Coverage	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.
Dwelling	<p>A dwelling is a building or structure in which people live. This can be a building, such as a house; part of a building, such as a flat; or it can be a caravan or tent, humpy or a park bench. Houses under construction, derelict houses, vacant tents, or converted garages, are not counted as dwellings in the Census. There are private and non-private dwellings.</p> <p>A private dwelling is normally a house, flat, part of a house, or even a room. Private dwellings can be either occupied or unoccupied. The following accommodation types are also classed as private dwellings: a house attached to, or room above, shops or offices; an occupied caravan or unit in a caravan park or on a residential allotment; an occupied boat in a marina; an occupied dwelling in a manufactured home estate; an occupied self-care unit in a retirement village; a houseboat; or a tent if it is standing on its own block of land.</p> <p>Non-private dwellings are those dwellings not included above which provide a communal or transitory type of accommodation. These dwellings include hotels, motels, guest houses, prisons, religious and charitable institutions, defence establishments, hospitals and other communal dwellings. Only occupied non-private dwellings are included in the Census.</p>
Estimated Resident Population (ERP)	Estimated Resident Population (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.
Gross overcount	The number of people in the Census who should not have been counted, either because they had already been counted or because they were overseas and should not have been counted at all. If a person was counted in the Census three times, for example, they would contribute two counts to the gross overcount (assuming they should have been counted in the Census).
Gross undercount	The number of people who should have been counted in the Census but were not.

GLOSSARY *continued*

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. 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.
Indigenous Community Frame (ICF)	<p>The Indigenous Community Frame (ICF) is a listing of all discrete Indigenous communities from which the sample of communities was selected for the 2006 PES. The selection unit on the ICF is a community set. A set usually comprises a main community, a number of out-stations and the non-community dwellings within the CDs associated with the main community. If a community set is selected then a selection of the dwellings in the main community, all the dwellings in some outstations, and a selection of the non-community dwellings will be enumerated in the survey.</p> <p>Once a community set has been selected the main community dwellings and outstations that will be enumerated are determined by the community skip and a random start. For instance, suppose a selected community set has 25 dwellings, 8 outstations and a skip of 5. The random start assigned is 3. The dwellings selected for enumeration are 3, 8, 13, 18, and 23, and all of outstations 3 and 8. The non-community dwellings will be enumerated with a skip of 5.</p>
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 considered to be occupied in the Census, and where Census data is imputed because no Census form was received (i.e. the dwelling was classified as non-contact or refusal).
Late return	A Census form which was returned after the start of PES enumeration.
Mainstream	The term 'mainstream' refers to the PES private dwelling sample, that is, all PES selections other than those from discrete Indigenous communities.
Net undercount	The difference between the actual Census count (including imputations) and an estimate of the number of people who should have been counted in the Census. For a category of person, net undercount is the net result of the PES estimate of gross undercount, gross overcount, differences in classification between the PES and Census (e.g. age, sex, Indigenous status) and imputation error in the Census.
Non-contact sector	The Census non-contact sector comprises late return and imputed dwellings.
Non-private dwelling	An establishment which provides a communal type of accommodation, such as a hotel, motel, hospital or other institution. Non-private dwellings were not included in the 2006 PES sample.
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.
Other territories	The Other territories comprise Jervis Bay Territory and the external territories of Christmas Island and Cocos (Keeling) Islands.
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.
Part of state	Parts of state are the capital city and balance of state in each state and territory. In a number of processes, such as estimation, the different parts of state are dealt with separately.

GLOSSARY *continued*

Place of enumeration Census count	People are counted according to where they were on Census Night. Overseas visitors are included and Australians overseas are excluded from the counts. No adjustment is made for Census undercount.
Place of enumeration net undercount	This is the net undercount of the place of enumeration Census counts. It is the net percentage of people present (in Australia, or a particular state or territory) on Census Night who were not counted.
Place of usual residence Census count	People are counted according to their stated place of usual residence in Australia. Overseas visitors are excluded, as are Australians residents who are temporarily overseas.
Place of usual residence net undercount	This is the net undercount of the usual residence Census counts. It is the net percentage of usual residents (of Australia, or a particular state or territory) present in Australia on Census Night who were not counted.
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.
Standard error	A measure of the likely difference between the true value and the estimate. For more details see the technical note.
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.
Usual residence	The usual residence for resident of Australia is defined for the purpose of the Census as being the place where they have lived or where they intend to live for 6 months or more in 2006.
Usual resident or visitor status	A person was a usual resident for the Census if they were at their usual residence on Census night. Visitors were people who were not at their usual residence on Census night.

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 statistics on the ABS web site can be downloaded free of charge.

- WEB ADDRESS* www.abs.gov.au



2294000001060

ISBN 9780642483256

RRP \$11.00