



**CENSUS OF POPULATION AND HOUSING  
– DETAILS OF UNDERCOUNT AUSTRALIA**

EMBARGO: 11.30AM (CANBERRA TIME) THURS 21 JUN 2012

**C O N T E N T S**

	<i>page</i>
Notes .....	2
 <b>1.OVERVIEW</b>	
Undercoverage in the Census .....	4
What is net undercount? .....	4
Key uses of net undercount .....	5
The importance of effective statistical independence .....	5
 <b>2.ESTIMATES OF NET UNDERCOUNT</b>	
Australia .....	7
Age and sex .....	8
States and territories .....	10
Greater capital city/Rest of state region .....	11
Registered marital status .....	12
Indigenous status .....	12
Country of birth .....	14
 <b>3.UNDERCOUNT ADJUSTMENT FACTORS</b>	
Adjustment factors .....	16
 <b>4.COMPONENTS OF NET UNDERCOUNT</b>	
Components of net undercount .....	18
 <b>5.CONDUCTING THE PES</b>	
Survey enumeration .....	22
Linking and matching .....	29
Estimation .....	35
 <b>EXPLANATORY NOTES</b>	
Explanatory notes .....	39
 <b>TECHNICAL NOTES</b>	
1. Improvement in collection of Indigenous status .....	44
2. Differences in classification between the PES and Census .....	46
3. People and dwellings missed in the Census .....	48
4. Identifying Census late returns .....	51
5. Statistical impact of ADL .....	54
6. Sampling errors .....	56
 <b>GLOSSARY</b>	
Glossary .....	58

**I N Q U I R I E S**

For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070 or Bjorn Jarvis on Canberra (02) 6252 6552.

## NOTES

**ABOUT THIS PUBLICATION** This publication presents estimates of net undercount for the 2011 Census of Population and Housing. It also details how the 2011 PES was conducted and how net undercount estimates have been calculated.

**NOTES ABOUT THE ESTIMATES** Information in this publication has been obtained from the 2011 Census Post Enumeration Survey (PES). The PES is a household survey conducted by the Australian Bureau of Statistics shortly after each Census, in order to provide an independent measure of Census coverage. The 2011 PES included people from approximately 36,000 responding households across Australia. Information was collected for 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 (undercount) and how many were counted more than once (overcount), which together with the Census counts, are used to determine the Census *net undercount*.

PES estimates of net undercount are used to augment the Census counts for the purpose of deriving resident population estimates for Australia and its states and territories.

It is important to note that where Census counts are presented in this publication, they refer to Census counts which correspond to the scope of the PES, and may differ slightly from aggregate counts in other Census products. For further information on the scope of the PES, refer to *Survey enumeration* (section 5).

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

Brian Pink  
Australian Statistician

## ABBREVIATIONS

---

'000	thousand
ABS	Australian Bureau of Statistics
ADL	automated data linking
ARA	any responsible adult
ASGC	Australian Standard Geographical Classification
ASGS	Australian Statistical Geography Standard
CAI	computer assisted interviewing
CARDS	collect, analyse, reduce, de-duplicate and systematise
CD	collection district
CLW	collector workload
DLR	dwelling link rating
ERP	estimated resident population
FEBRL	Freely Extensible Biomedical Record Linking
ICF	Indigenous Community Frame
MSS	Match and Search System
NPD	non-private dwelling
PAPI	pen-and-paper interview
PD	private dwelling
PES	Census Post Enumeration Survey
PLR	person link rating
PREG	prediction regression
QA	quality assurance
SA1	Statistical Area Level 1
SA2	Statistical Area Level 2
SE	standard error

## OVERVIEW

---

### UNDERCOVERAGE IN THE CENSUS

Tuesday, 9 August 2011 was Census night in Australia. All people present in Australia on this night, with the exception of foreign diplomats and their families, should have been included on a Census form at the place where they stayed.

The Census of Population and Housing is the largest statistical collection undertaken by the Australian Bureau of Statistics (ABS) and one of the most important. Its objective is to accurately measure the number of people in Australia on Census night, their characteristics and the dwellings in which they live. Due to its size and complexity, whenever a Census is conducted it is inevitable that some people will be missed and some will be counted more than once.

It is for this reason that the Census Post Enumeration Survey (PES) is run shortly after each Census, to provide an independent measure of Census coverage. The PES determines how many people should have been counted in the Census, how many were missed, and how many were counted more than once. It also provides information on the characteristics of those in the population who have been missed or overcounted.

Some of the reasons why people may have been missed in the Census (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 certain people, for example, young babies, the elderly or visitors, should not be included;
- they did not wish to be included due to concerns about confidentiality or a more general reluctance to participate;
- the dwelling in which they were located was missed because it was difficult to find (e.g. in a remote or non-residential area); and
- the dwelling in which they were located was mistakenly classed as unoccupied.

Some of the reasons why people may have been counted more than once (i.e. overcounted) or in error include:

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

### WHAT IS NET UNDERCOUNT?

While every effort is made to eliminate these potential causes of error, some undercount and overcount will inevitably occur. As is usually the case in Australia, in the 2011 Census more people were missed than overcounted and so the Census count of the population is fewer than the true population. This difference is referred to as *net undercount*.

Net undercount for any category of person 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 in non-responding Census dwellings).

## OVERVIEW *continued*

### KEY USES OF NET UNDERCOUNT

Net undercount is the primary measure of Census coverage, and as such, is used in the following ways:

- to augment Census counts, in order to derive the most robust estimate of the resident population (ERP) for 30 June of the Census year;
- to provide users with an assessment of the completeness of the Census counts, allowing them to take this into account when using Census information; and
- to evaluate the effectiveness of Census collection procedures so improvements may be made for future Censuses.

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

For more information on the calculation of ERP for 30 June 2011 based on results from the 2011 Census and PES, see the ABS publication *Australian Demographic Statistics, December quarter 2011* (cat. no. 3101.0), released on 20 June 2012.

### THE IMPORTANCE OF EFFECTIVE STATISTICAL INDEPENDENCE

The PES is designed to be an independent check of Census coverage. Therefore, it is critical that the statistical independence between the PES and the Census is effectively managed, to ensure the PES is a robust check on Census coverage.

There are two aspects to statistical independence, both of which were effectively managed throughout the 2011 PES cycle: population independence and operational independence.

*Population independence* refers to the principle that there should be no sub-groups of the population where being missed in the Census indicates that a person or dwelling is also more likely to be missed by the PES. Although the PES estimation process adjusts for this to some extent, by subdividing the population into smaller groups where the assumption of population independence is more likely to be true, population independence is always more difficult to achieve than operational independence.

Selection to participate in the 2011 PES was based on a sample of private dwellings, meaning that those persons who were not living in, or visiting, a private dwelling at PES time were unavailable for selection. Therefore, although the PES has shown it is very effective in assessing *overall* Census coverage, its usefulness for estimating the undercount of certain sub-populations is limited, such as fly-in fly-out workers, who often live in non-private dwellings.

*Operational independence* requires that Census operations do not influence the PES in any way, and vice versa. The operational independence of the PES from the Census was effectively monitored at every stage of the 2011 cycle, including enumeration, processing and administration. Steps taken to ensure this independence included:

- selecting the PES sample from an independent sample frame;
- using separate office staff in the PES and Census;
- ensuring PES interviewers were not employed as Census field staff in the same area;
- maintaining the confidentiality of the PES sample so Census field and office staff were unaware of which areas were selected in the PES; and

## OVERVIEW *continued*

---

THE IMPORTANCE OF  
EFFECTIVE STATISTICAL  
INDEPENDENCE *continued*

- ensuring Census forms received after PES enumeration commenced were excluded from PES estimation, thereby protecting the PES sample from having a higher proportion of Census response than in the overall population (due to contact from PES prompting respondents to return their Census forms).

## ESTIMATES OF NET UNDERCOUNT

### INTRODUCTION

#### *Understanding the Census count*

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. The Census count includes persons who have been imputed in non-responding dwellings in the Census, that is, the Census count equals persons counted on Census forms *plus* imputed persons for non-responding Census dwellings.

Net undercount is therefore a measure of the combined outcome of Census enumeration and data processing. For more information about imputed persons for Census non-responding dwellings and the adjustments made for them in the PES estimates, see *Components of net undercount* (section 4) and the *Identifying Census late returns* Technical Note in this publication.

#### *2011 PES net undercount estimates*

In the following tables, net undercount is presented as both level estimates of persons and rates, together with their associated standard errors (SEs). The net undercount rate expresses the net undercount (i.e. undercount minus overcount) as a percentage of the PES estimate of a given population (i.e. as a percentage of the number of people who *should* have been counted in the Census). All estimates of net undercount based on geography have been calculated on a *place of usual residence basis*, meaning they are based on the location where a person lived, or intended to live, for six months or more in 2011.

### AUSTRALIA

The 2011 Census counted 21,504,721 usual residents of Australia (including imputed persons in non-responding dwellings). This was around 374,540 persons fewer than the estimated population usual residents who were present in Australia on Census night. This equates to a net undercount rate of 1.7%. In other words, 98.3% of the usually resident population were included in 2011 Census counts.

While the net undercount rate decreased from 2.7% in 2006 to 1.7% in 2011, it is important to note that PES estimates of net undercount are not strictly comparable over time due to changes in both Census and PES methodologies. The PES is designed to provide the best measure of Census coverage at a single point in time rather than as a time series, with improvements made to the PES and Census in each cycle. This is particularly true for 2011 with the introduction of Automated Data Linking (ADL) in the PES which has made it difficult to directly compare level estimates and proportions of net undercount from one Census to another. Compositional analysis is therefore much more illustrative.

The ABS has estimated that the introduction of ADL for 2011 has resulted in a net undercount that was 246,985 persons *less* than if the 2006 methodology had been used. This estimate is subject to sample error. For further details see the *Statistical Impact of ADL* Technical Note.

Table 1 shows the net undercount rates and associated standard errors (SEs) for Australia for each Census from 1971 to 2011. The 2011 estimate of 1.7% continues the historical trend of almost complete coverage in Australian Census counts.

## ESTIMATES OF NET UNDERCOUNT *continued*

### 1 NET UNDERCOUNT RATE, Australia—1971–2011

	1971	1976	1981	1986	1991	1996	2001	2006(a)	2011(b)
	%	%	%	%	%	%	%	%	%
Net Undercount	1.4	2.7	1.9	1.9	1.8	1.6	1.8	2.7	1.7
Standard Error (SE)	0.1	0.04	0.1	0.1	0.1	0.1	0.1	0.2	0.2

- (a) Care should be taken when comparing estimates from 2006 onwards with previous years due to changes made to PES estimation and the inclusion of remote areas and discrete Indigenous communities in the PES sample from 2006.
- (b) Care should be taken when comparing 2011 estimates with previous years due to changes made in PES linking and matching methodology. For more information see Linking and matching (section 5).

#### AGE AND SEX

The likelihood of counting a person in the Census has traditionally varied according to age and sex. As has been observed in previous Censuses (both in Australia and overseas), young adults are the age group who are most likely to be missed in the Census, with young adult males being more likely to be missed than their female counterparts. In contrast, older adults are much more likely to be counted.

Tables 2 and 3, and graph 4, show that this was also true for the 2011 Australian Census. In particular, males aged 20–24 years again had the highest net undercount rate (7.8%) followed by males aged 25–29 years (7.5%). The net undercount rate for females was also highest for those aged 20–24 years (6.0%). While the undercount rate for 25–29 year old females (4.0%) was higher than most age groups, in contrast to 2006 it was noticeably lower than the rate for females in their early 20s.

The lowest net undercount rate was for people aged 55 years and over (-0.1%).

In general, males had a higher net undercount rate (2.2%) than females (1.2%).

### 2 NET UNDERCOUNT (a), Sex by age group—2011

Age	Males		Females		Persons	
	no.	SE	no.	SE	no.	SE
0–4 yrs	8 488	5 925	8 775	5 114	17 263	8 017
5–9 yrs	13 559	5 914	6 367	5 045	19 926	7 781
10–14 yrs	4 465	5 124	1 449	5 538	5 914	7 664
15–19 yrs	18 005	6 396	17 890	6 577	35 895	9 044
20–24 yrs	62 594	9 392	45 555	7 730	108 148	12 812
25–29 yrs	61 035	10 613	31 602	7 647	92 636	13 444
30–34 yrs	30 419	7 225	14 527	6 528	44 946	10 407
35–39 yrs	15 746	7 330	1 462	5 909	17 207	9 424
40–44 yrs	8 768	6 307	12 255	6 077	21 023	8 856
45–49 yrs	11 367	6 141	-5 876	5 448	5 491	8 063
50–54 yrs	11 848	6 279	1 900	5 428	13 747	8 411
55 yrs and over	-5 498	7 396	-2 161	7 626	-7 659	11 500
<b>Total all ages</b>	<b>240 796</b>	<b>24 534</b>	<b>133 744</b>	<b>20 772</b>	<b>374 540</b>	<b>38 315</b>

- (a) A negative value indicates a net overcount.

# ESTIMATES OF NET UNDERCOUNT *continued*

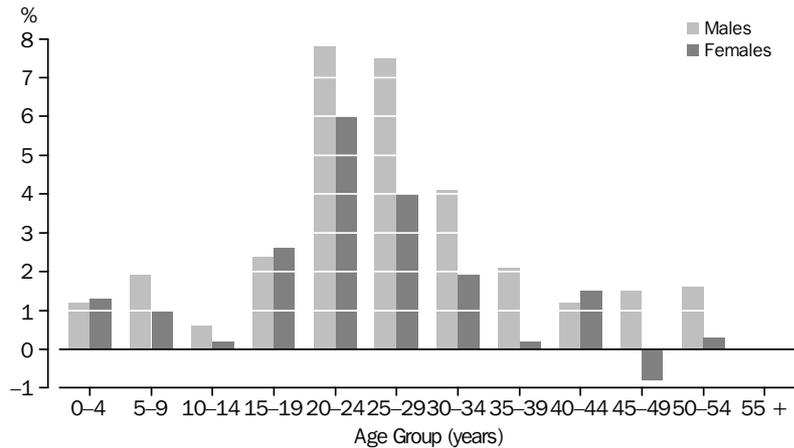
AGE AND SEX *continued*

## 3 NET UNDERCOUNT RATE(a), Sex by age group—2011

Age	Males		Females		Persons	
	%	SE	%	SE	%	SE
0-4 yrs	1.2	0.8	1.3	0.7	1.2	0.6
5-9 yrs	1.9	0.8	1.0	0.8	1.5	0.6
10-14 yrs	0.6	0.7	0.2	0.8	0.4	0.6
15-19 yrs	2.4	0.8	2.6	0.9	2.5	0.6
20-24 yrs	7.8	1.1	6.0	1.0	6.9	0.8
25-29 yrs	7.5	1.2	4.0	0.9	5.8	0.8
30-34 yrs	4.1	0.9	1.9	0.9	3.0	0.7
35-39 yrs	2.1	0.9	0.2	0.8	1.1	0.6
40-44 yrs	1.2	0.8	1.5	0.8	1.3	0.6
45-49 yrs	1.5	0.8	-0.8	0.7	0.4	0.5
50-54 yrs	1.6	0.9	0.3	0.7	0.9	0.6
55 yrs and over	-0.2	0.3	-0.1	0.3	-0.1	0.2
<b>Total all ages</b>	<b>2.2</b>	<b>0.2</b>	<b>1.2</b>	<b>0.2</b>	<b>1.7</b>	<b>0.2</b>

(a) A negative value indicates a net overcount.

## 4 NET UNDERCOUNT RATE, Sex by age group—2011



## ESTIMATES OF NET UNDERCOUNT *continued*

### STATES AND TERRITORIES

The challenges facing Census enumeration vary between states and territories. Table 5 shows the rates of net undercount for Australian states and territories for Censuses from 1991 to 2011.

As in previous Censuses, in 2011 the Northern Territory recorded the highest net undercount rate of all states and territories (6.9%), while the Australian Capital Territory continued to record the lowest net undercount rate (0.7%).

While the two territories reflected the minimum and maximum net undercount rates, Victoria and South Australia continued to show relatively low rates (both 1.1%). Western Australia had the highest rate for a state (2.5%), emphasising the continued coverage challenges in that state. All states and territories had a lower net undercount rate in 2011 than in 2006, except for Tasmania which was relatively consistent at 2.0%. The greatest decreases were for Queensland (3.7% in 2006 to 1.8% in 2011) and Victoria (2.3% in 2006 to 1.1% in 2011).

It is important to note the effect of the introduction of Automated Data Linking (ADL) when considering the changes from 2006. For more information, see the *Statistical Impact of ADL* Technical Note in this publication.

### 5 NET UNDERCOUNT, State/territory of usual residence—1991-2011

	1991			1996			2001			2006(a)			2011(b)		
	Persons		Rate												
	no.	%	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	136 647	1.9	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	56 906	1.1	0.3
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	77 215	1.8	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	17 283	1.1	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	57 918	2.5	0.5
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	10 261	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	15 716	6.9	1.3
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	2 595	0.7	0.8
<b>Australia</b>	<b>312 300</b>	<b>1.8</b>	<b>0.1</b>	<b>285 800</b>	<b>1.6</b>	<b>0.1</b>	<b>346 119</b>	<b>1.8</b>	<b>0.1</b>	<b>549 486</b>	<b>2.7</b>	<b>0.2</b>	<b>374 540</b>	<b>1.7</b>	<b>0.2</b>

(a) Care should be taken when comparing estimates from 2006 onwards with previous years due to changes made to PES estimation and the inclusion of remote areas and discrete Indigenous communities in the PES sample from 2006.

(b) Care should be taken when comparing 2011 estimates with previous years due to changes made to PES linking and matching methodology. For more information see Linking and Matching (section 5).

## ESTIMATES OF NET UNDERCOUNT *continued*

### GREATER CAPITAL CITY/REST OF STATE REGION

The regional differences in net undercount in the 2011 Census for greater capital cities and the rest of state regions are presented in tables 6 and 7.

Greater capital cities are represented by Greater Capital City Statistical Areas (GCCSAs) and represent a socio-economic definition of each of the eight state and territory capital cities. This means each greater capital city includes people who regularly socialise, shop or work within the city but live in small towns and rural areas surrounding the city. 'Rest of state' regions are the areas within each state or territory not defined as being part of the greater capital city.

Different problems are encountered in enumerating different areas of Australia and these are reflected, to a certain extent, in the net undercount rates. In urban areas, locating dwellings is generally easier but contacting occupants and gaining their cooperation can be more difficult. In contrast, in rural and remote areas where dwellings may be scattered over a wider area, locating the dwellings can cause considerable difficulties. In 2011, New South Wales, Tasmania and the Northern Territory had higher net undercount rates in their rest of state regions compared with their greater capital cities. All other states and territories had a lower net undercount in their rest of state regions (compared to their greater capital cities).

At the broad Australia level, the total net undercount rates in 2011 were slightly lower for the rest of state regions (1.7%) compared to greater capital cities (1.8%). This contrasts with 2006, when the net undercount rate for balance of state/territory (3.0%) was higher compared to the capital cities (2.5%). As was the case in 2006, in 2011 the Northern Territory showed the largest difference in net undercount rate between its greater capital city and rest of state region (3.7% and 10.9% respectively).

It is important to note that for the 2011 PES, the Australian Statistical Geographical Standard (ASGS) replaced the Australian Standard Geography Classification (ASGC) as the framework for PES geography. The move to the new geography will allow for improvements in the quality of small area time series data from the Census. However, the change has resulted in an unavoidable break in series and care should be taken when comparing the 2011 greater capital city/rest of state net undercount estimates to the 2006 capital city/balance of state net undercount estimates.

### **6** NET UNDERCOUNT, State/territory of usual residence—By greater capital city/rest of state region—2011

	<i>Greater capital city</i> .....		<i>Rest of state region</i> .....		<i>Total</i> .....	
	no.	SE	no.	SE	no.	SE
New South Wales	79 825	21 972	56 821	17 139	136 647	26 865
Victoria	55 079	16 029	1 827	9 823	56 906	17 864
Queensland	41 898	11 244	35 316	13 997	77 215	17 431
South Australia	14 096	5 584	3 186	3 033	17 283	6 045
Western Australia	51 610	10 690	6 308	6 557	57 918	12 004
Tasmania	1 308	2 141	8 953	3 014	10 261	3 076
Northern Territory	4 637	2 254	11 078	2 157	15 716	3 074
Australian Capital Territory	2 595	3 055	na	na	2 595	3 055
<b>Australia</b>	<b>251 050</b>	<b>31 913</b>	<b>123 490</b>	<b>24 771</b>	<b>374 540</b>	<b>38 315</b>

## ESTIMATES OF NET UNDERCOUNT *continued*

GREATER CAPITAL  
CITY/REST OF STATE  
REGION *continued*

### 7 NET UNDERCOUNT RATE, State/territory of usual residence—By greater capital city/rest of state region—2011

	Greater capital city		Rest of state region		Total	
	%	SE	%	SE	%	SE
New South Wales	1.8	0.5	2.2	0.7	1.9	0.4
Victoria	1.4	0.4	0.1	0.7	1.1	0.3
Queensland	2.0	0.5	1.5	0.6	1.8	0.4
South Australia	1.1	0.4	0.9	0.8	1.1	0.4
Western Australia	2.9	0.6	1.2	1.3	2.5	0.5
Tasmania	0.6	1.0	3.1	1.0	2.0	0.6
Northern Territory	3.7	1.7	10.9	1.9	6.9	1.3
Australian Capital Territory	0.7	0.8	na	na	0.7	0.8
<b>Australia</b>	<b>1.8</b>	<b>0.2</b>	<b>1.7</b>	<b>0.3</b>	<b>1.7</b>	<b>0.2</b>

REGISTERED MARITAL  
STATUS

Table 8 shows net undercount estimates and rates by registered marital status by sex. The net undercount rates were highest for people identified as never married (3.7%) and lowest for people widowed, divorced or separated (a net overcount of 0.8%). It is important to consider the strong relationship with age when interpreting net undercount estimates by registered marital status.

### 8 NET UNDERCOUNT(a), Registered marital status by sex—2011

	MALES				FEMALES				PERSONS			
	Persons		Rate		Persons		Rate		Persons		Rate	
	no.	SE	%	SE	no.	SE	%	SE	no.	SE	%	SE
Never married (b)	234 038	20 842	4.2	0.4	151 557	16 596	3.1	0.3	385 595	29 990	3.7	0.3
Widowed, divorced or separated	-8 837	10 274	-0.8	1.0	-15 434	9 878	-0.8	0.5	-24 271	14 630	-0.8	0.5
Married	15 594	10 970	0.4	0.3	-2 378	9 966	-0.1	0.2	13 216	18 737	0.2	0.2
<b>Total persons</b>	<b>240 796</b>	<b>24 534</b>	<b>2.2</b>	<b>0.2</b>	<b>133 744</b>	<b>20 772</b>	<b>1.2</b>	<b>0.2</b>	<b>374 540</b>	<b>38 315</b>	<b>1.7</b>	<b>0.2</b>

(a) A negative value indicates a net overcount.

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

INDIGENOUS STATUS

Special procedures are used in the Census to support the enumeration of the Aboriginal and Torres Strait Islander population, as counting this population continues to present a number of challenges.

The 2011 Census counted 548,147 persons who had been identified as being of Aboriginal and Torres Strait Islander origin, which was 21% more than the 454,799 persons in 2006. A summary of Aboriginal and Torres Strait Islander Census counts is presented in *Census of Population and Housing - Counts of Aboriginal and Torres Strait Islander Australians* (cat. no. 2075.0), which was also released on 21 June.

The 2011 PES estimated that 662,335 Aboriginal and Torres Strait Islander persons should have been counted in the Census, compared with 513,977 persons in 2006.

## ESTIMATES OF NET UNDERCOUNT *continued*

### INDIGENOUS STATUS

*continued*

Table 9 shows net undercount estimates by Indigenous status. In 2011, the net undercount rate was 17.2%, compared with 11.5% in 2006. The net undercount for 2011 was estimated to be 114,188 persons, which was almost double the 2006 estimate of 59,178 persons.

**9** NET UNDERCOUNT, Indigenous status—2011

	Persons		Rate	
	no.	SE	%	SE
Indigenous	114 188	14 274	17.2	1.8
Non-Indigenous	1 318 799	37 272	6.2	0.2

It is important to note that these measures refer to the undercount of persons according to their Indigenous status, regardless of whether or not they were actually counted in the Census. In other words, persons who were counted in the Census and had a 'not-stated' Indigenous status will not be included in the Census counts of either Aboriginal and Torres Strait Islander or non-Indigenous persons, but are, instead, a separate category for this classification. They will, however, be included in Census counts for other key categories, such as Age and Sex.

In order to understand the differences between the 2006 and 2011 PES results, the ABS undertook an extensive quality assurance process. The results of this process are summarised in the *Improvement in collection of Indigenous status* Technical Note in this publication.

This quality assurance process has led the ABS to advise caution when comparing net undercount for Aboriginal and Torres Strait Islander persons between 2006 and 2011. Analysis of data indicates that the main contributing factor for the difference between the 2006 and 2011 estimates was improved PES methodology and procedures, which resulted in better identification of Aboriginal and Torres Strait Islander people in the 2011 PES.

Historically, the ABS has a program of continuous improvement in its survey methodologies. Improvements to the PES in 2006 and 2011 are summarised in *Survey Enumeration, Linking and Matching and Estimation* (section 5). While the individual impacts of all improvements made in 2011 cannot be measured, they have resulted in a change in the Indigenous status classification of Aboriginal and Torres Strait Islander persons in the PES and the Census between 2006 and 2011. This has in turn resulted in a noticeably different net category change for Indigenous status in net undercount estimates, and accounted for most of the change in the estimate of net undercount for Aboriginal and Torres Strait Islander persons between 2006 and 2011.

It is also important to note that Indigenous status, as collected in both the Census and PES, is based on responses to a question related to information that some people will consider personal and sensitive. Respondents can choose to indicate in the Census that they are Indigenous or non-Indigenous, or they can choose to not answer the question at all. If no answer is provided, the Census does not impute for this missing response (which is also the case for imputed persons). The Census count is therefore a count of

## ESTIMATES OF NET UNDERCOUNT *continued*

### INDIGENOUS STATUS

*continued*

those who were identified by a respondent as Indigenous (i.e. those without a response are excluded).

### COUNTRY OF BIRTH

As Census forms are generally completed by one or more persons in a household, those who have come to Australia from other countries and whose first language is not English may find completing a Census form more difficult than other Australians. For several Censuses, special strategies have been employed to promote an understanding of the Census among migrants, in particular that the Government is not using their information for anything other than statistical purposes and to provide assistance in a range of languages.

Tables 10 and 11 show the undercount estimates and rates by country of birth. The countries displayed were the 10 highest ranked (in terms of population residing in Australia) according to the 2011 Census. There were 1,195,432 people (5.6% of the Census count) whose country of birth was not stated in the Census. Since Census does not impute a Country of birth for these people, the PES estimates of net undercount are not adjusted to take account of any imputed persons. As with Indigenous status, these people, while counted in the Census, do not contribute to the Census counts for these categories but do count to PES estimates of their population. For further information about Census not-stated responses and their impact on estimates of net undercount see *Components of net undercount* (section 4).

Of those countries listed, persons born in China had the highest net undercount (55,965 persons) followed by New Zealand (46,536 persons). China also had the highest net undercount rate (14.9%) followed by India (9.7%). Persons born in Scotland had the lowest net undercount rate (1.2%) followed by those born in England and Italy (both 4.6%).

**10** NET UNDERCOUNT(a)(b), Country of birth by sex—2011

Country of Birth	Males		Females		Persons	
	no.	SE	no.	SE	no.	SE
Australia	584 461	19 579	477 937	16 850	1 062 398	30 156
England	26 257	4 870	17 938	4 549	44 195	7 049
New Zealand	30 504	4 323	16 032	3 770	46 536	6 344
China	28 239	5 375	27 726	5 033	55 965	8 934
India	22 309	4 371	9 386	2 856	31 694	6 075
Italy	4 880	1 631	4 004	1 654	8 884	2 513
Vietnam	8 020	2 736	8 075	2 326	16 095	4 313
Philippines	6 441	1 952	10 665	2 417	17 106	3 603
South Africa	7 099	2 138	4 613	1 791	11 712	3 383
Scotland	-1 348	1 511	2 929	1 674	1 582	2 208
Other overseas	151 369	11 502	122 435	8 990	273 804	15 894

(a) Net undercount is based on Census counts for a category. In the Census, Country of birth was set to not-stated where the response was blank and where imputed person records were created for non-responding dwellings. Hence net undercount estimates for Country of birth do not sum to the Australia total.

(b) A negative value indicates a net overcount.

## ESTIMATES OF NET UNDERCOUNT *continued*

COUNTRY OF BIRTH  
*continued*

### 11 NET UNDERCOUNT RATE(a)(b), Country of birth by sex—2011

<i>Country of Birth</i>	<i>Males</i>		<i>Females</i>		<i>Persons</i>	
	%	SE	%	SE	%	SE
Australia	7.3	0.2	5.9	0.2	6.6	0.2
England	5.4	1.0	3.8	0.9	4.6	0.7
New Zealand	11.1	1.4	6.3	1.4	8.8	1.1
China	16.6	2.6	13.5	2.1	14.9	2.0
India	12.0	2.1	6.7	1.9	9.7	1.7
Italy	4.9	1.6	4.2	1.7	4.6	1.2
Vietnam	8.6	2.7	7.5	2.0	8.0	2.0
Philippines	9.1	2.5	9.1	1.9	9.1	1.7
South Africa	9.0	2.5	5.9	2.2	7.4	2.0
Scotland	-2.1	2.4	4.2	2.3	1.2	1.6
Other overseas	11.2	0.8	8.8	0.6	10.0	0.5

- (a) Net undercount is based on Census counts for a category. In the Census, Country of birth was set to not-stated where the response was blank and where imputed person records were created for non-responding dwellings.
- (b) A negative value indicates a net overcount.

## UNDERCOUNT ADJUSTMENT FACTORS

---

### UNDERCOUNT ADJUSTMENT FACTORS

While estimates of net undercount are important for an effective understanding of the completeness of Census counts, undercount adjustment factors are the means for adjusting Census counts.

The undercount adjustment factor is the ratio of the PES population estimate to the actual Census count. This factor can be applied to the Census count for any category to indicate how many people should have been counted in that Census category. Table 12 provides corresponding PES population estimates, Census counts and net undercount adjustment factors for the categories previously discussed.

The undercount adjustment factor should not be used alone to derive an alternative measure of the Estimated Resident Population (ERP). Official population estimates include additional data and adjustments for usual residents of Australia who are temporarily overseas. For information on the calculation of ERP, see the ABS publication *Australian Demographic Statistics, December quarter 2011* (cat. no. 3101.0), released on 20 June 2012.

## UNDERCOUNT ADJUSTMENT FACTORS *continued*

### 12 NET UNDERCOUNT, Population estimates, Census counts and Undercount adjustment factors—2011

	PES population estimate(a)		Census count(b)	Net undercount		Undercount adjustment factor	
	no.	SE		no.	SE	no.	SE
<b>Australia</b>	<b>21 879 261</b>	<b>38 315</b>	<b>21 504 721</b>	<b>374 540</b>	<b>38 315</b>	<b>1.017</b>	<b>0.002</b>
State/territory of usual residence							
New South Wales	7 054 301	26 865	6 917 654	136 647	26 865	1.020	0.004
Victoria	5 410 949	17 864	5 354 043	56 906	17 864	1.011	0.003
Queensland	4 409 948	17 431	4 332 733	77 215	17 431	1.018	0.004
South Australia	1 613 850	6 045	1 596 567	17 283	6 045	1.011	0.004
Western Australia	2 297 111	12 004	2 239 193	57 918	12 004	1.026	0.005
Tasmania	505 616	3 076	495 355	10 261	3 076	1.021	0.006
Northern Territory	227 675	3 074	211 959	15 716	3 074	1.074	0.015
Australian Capital Territory	359 812	3 055	357 217	2 595	3 055	1.007	0.009
Sex							
Male	10 872 858	24 534	10 632 062	240 796	24 534	1.023	0.002
Female	11 006 403	20 772	10 872 659	133 744	20 772	1.012	0.002
Age group (years)							
0–4	1 438 176	8 017	1 420 913	17 263	8 017	1.012	0.006
5–9	1 371 672	7 781	1 351 746	19 926	7 781	1.015	0.006
10–14	1 376 784	7 664	1 370 870	5 914	7 664	1.004	0.006
15–19	1 441 530	9 044	1 405 635	35 895	9 044	1.026	0.006
20–24	1 568 558	12 812	1 460 410	108 148	12 812	1.074	0.009
25–29	1 605 452	13 444	1 512 816	92 636	13 444	1.061	0.009
30–34	1 498 425	10 407	1 453 479	44 946	10 407	1.031	0.007
35–39	1 537 123	9 424	1 519 916	17 207	9 424	1.011	0.006
40–44	1 563 684	8 856	1 542 661	21 023	8 856	1.014	0.006
45–49	1 509 411	8 063	1 503 920	5 491	8 063	1.004	0.005
50–54	1 460 951	8 411	1 447 204	13 747	8 411	1.009	0.006
55+	5 507 492	11 500	5 515 151	-7 659	11 500	0.999	0.002
Registered marital status							
Never married(c)	10 490 920	29 990	10 105 325	385 595	29 990	1.038	0.003
Widowed, divorced or separated	2 915 337	14 630	2 939 608	-24 271	14 630	0.992	0.005
Married	8 473 004	18 737	8 459 788	13 216	18 737	1.002	0.002
Indigenous status							
Indigenous	662 335	14 274	548 147	114 188	14 274	1.208	0.026
Non-Indigenous	21 216 926	37 272	19 898 127	1 318 799	37 272	1.066	0.002
Not stated(d)	—	—	1 058 447	—	—	—	—
Country of birth							
Australia	16 078 800	30 156	15 016 402	1 062 398	30 156	1.071	0.002
England	955 734	7 049	911 539	44 195	7 049	1.048	0.008
New Zealand	529 937	6 344	483 401	46 536	6 344	1.096	0.013
China	374 935	8 934	318 970	55 965	8 934	1.175	0.028
India	327 051	6 075	295 357	31 694	6 075	1.107	0.021
Italy	194 287	2 513	185 403	8 884	2 513	1.048	0.014
Vietnam	201 125	4 313	185 030	16 095	4 313	1.087	0.023
Philippines	188 325	3 603	171 219	17 106	3 603	1.100	0.021
South Africa	157 387	3 383	145 675	11 712	3 383	1.080	0.023
Scotland	135 002	2 208	133 420	1 582	2 208	1.012	0.017
Other overseas	2 736 677	15 894	2 462 873	273 804	15 894	1.111	0.006
Not stated(d)	—	—	1 195 432	—	—	—	—

— nil or rounded to zero (including null cells)

(a) PES estimate of the number of people who should have been counted in the Census.

(b) Includes imputed persons in non-responding dwellings.

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

(d) People whose Census form was either not received or partially completed.

## COMPONENTS OF NET UNDERCOUNT

---

### COMPONENTS OF NET UNDERCOUNT

While net undercount estimates measure Census coverage, and adjustment factors are a means of adjusting Census counts for coverage, it is also important to understand the contributing components of net undercount.

The components of net undercount for a category are:

- undercount (the number of people in that category who were missed in the Census);
- overcount (the number of people in that category counted in the Census when they should not have been);
- differences in classification between the PES and Census (for example, where age, sex or Indigenous status information does not match); and
- imputation error (either under or over-imputation) in the Census.

This section provides additional information on these components to assist in the interpretation of estimates of net undercount presented in this publication.

### *Contact and non-contact sector*

Table 13 provides estimates of the key components of net undercount, the definitions of which follow the table. For the purposes of PES estimation, Census dwellings deemed occupied on Census night are divided into a 'contact' sector (dwellings for which a Census form was received before the commencement of PES enumeration) and a 'non-contact' sector (where no Census form was received or a Census form was received after the commencement of PES enumeration – that is, a 'late return').

## COMPONENTS OF NET UNDERCOUNT *continued*

### 13 COMPONENTS OF NET UNDERCOUNT(a)(b), Contact and non-contact sector—2011

#### UNDERCOUNT IN THE CONTACT SECTOR

	Persons missed in the Census	Net difference in classification	Persons with Census category not-stated	Total undercount in the contact sector		Undercount in non-contact sector		Net undercount	
	no.	no.	no.	no.	SE	no.	SE	no.	SE
<b>Australia</b>	<b>634 579</b>	—	—	<b>634 579</b>	<b>25 380</b>	<b>-260 039</b>	<b>27 611</b>	<b>374 540</b>	<b>38 315</b>
State/territory of usual residence									
New South Wales	217 639	4 651	—	222 290	18 507	-85 644	18 012	136 647	26 865
Victoria	120 917	-5 098	—	115 819	11 283	-58 912	13 630	56 906	17 864
Queensland	139 820	-3 269	—	136 550	11 612	-59 336	12 466	77 215	17 431
South Australia	32 446	-803	—	31 644	4 727	-14 361	4 344	17 283	6 045
Western Australia	87 034	2 211	—	89 246	7 760	-31 327	8 800	57 918	12 004
Tasmania	13 839	1 041	—	14 880	2 369	-4 619	1 709	10 261	3 076
Northern Territory	16 609	1 952	—	18 561	2 642	-2 845	1 402	15 716	3 074
Australian Capital Territory	6 274	-685	—	5 589	2 209	-2 994	2 027	2 595	3 055
Sex									
Male	393 555	1 664	—	395 219	17 603	-154 423	16 281	240 796	24 534
Female	241 023	-1 664	—	239 360	14 827	-105 615	14 508	133 744	20 772
Age groups (years)									
0-4	49 140	-22 080	—	27 060	6 344	-9 797	4 772	17 263	8 017
5-9	29 582	-586	—	28 996	6 680	-9 069	4 118	19 926	7 781
10-14	19 312	729	—	20 042	6 511	-14 127	3 761	5 914	7 664
15-19	46 229	-2 473	—	43 756	7 534	-7 861	4 865	35 895	9 044
20-24	108 284	7 686	—	115 970	10 635	-7 821	6 638	108 148	12 812
25-29	95 330	-188	—	95 142	10 487	-2 505	7 812	92 636	13 444
30-34	59 418	3 548	—	62 966	8 415	-18 019	5 450	44 946	10 407
35-39	49 768	-15 574	—	34 194	7 650	-16 987	4 897	17 207	9 424
40-44	37 122	5 794	—	42 915	7 644	-21 892	4 100	21 023	8 856
45-49	37 607	-6 756	—	30 851	7 013	-25 360	3 687	5 491	8 063
50-54	30 646	7 866	—	38 513	7 131	-24 765	4 242	13 747	8 411
55+	72 142	22 032	—	94 174	9 610	-101 834	7 172	-7 659	11 500
Registered marital status									
Never married(c)	425 154	34 811	—	459 966	22 591	-74 371	18 606	385 595	29 990
Widowed, divorced or separated	74 578	-66 701	—	7 877	12 757	-32 148	6 714	-24 271	14 630
Married	134 847	31 889	—	166 736	15 225	-153 520	10 761	13 216	18 737
Indigenous status(d)									
Indigenous	56 650	5 128	7 009	68 787	11 982	45 402	5 969	114 188	14 274
Non-Indigenous	577 928	-5 128	280 012	852 812	25 388	465 987	26 476	1 318 799	37 272
Country of birth(d)									
Australia	405 053	14 448	307 088	726 589	20 883	335 809	20 578	1 062 398	30 156
England	21 653	-15 325	22 106	28 434	6 270	15 761	3 082	44 195	7 049
New Zealand	23 881	-3 398	8 368	28 851	4 894	17 685	3 766	46 536	6 344
China	33 905	-4 261	6 002	35 646	6 782	20 319	5 062	55 965	8 934
India	12 117	-2 246	5 336	15 207	3 829	16 487	4 811	31 694	6 075
Italy	2 904	-2 248	6 775	7 431	2 169	1 453	1 067	8 884	2 513
Vietnam	9 389	-1 765	4 458	12 082	4 075	4 013	1 984	16 095	4 313
Philippines	6 090	3 171	3 318	12 579	3 053	4 527	1 884	17 106	3 603
South Africa	3 496	-895	2 289	4 890	2 013	6 822	2 755	11 712	3 383
Scotland	1 268	-7 096	3 506	-2 322	1 733	3 904	1 374	1 582	2 208
Other overseas	114 821	19 616	53 599	188 036	12 411	85 768	9 553	273 804	15 894

— nil or rounded to zero (including null cells)

(a) Estimates are based on a sample survey, and hence are subject to sampling error indicated by the SEs provided. For further information see the Sampling Error Technical Note.

(b) A negative value indicates a net overcount.

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

(d) Net undercount is based on Census counts for a category. In the Census, Indigenous status and Country of birth were set to not-stated where the response was blank and where imputed person records were created for non-responding dwellings. Hence components of undercount for Indigenous status and Country of birth do not sum to the Australia total.

## COMPONENTS OF NET UNDERCOUNT *continued*

### Contact and non-contact sector *continued*

#### PERSONS MISSED IN THE CENSUS IN THE CONTACT SECTOR

For a given category of person as reported in the PES (e.g. 'Males' recorded in the PES), 'Persons not counted in the contact sector' is the difference between the number of people in that category that should have been counted in the Census and the number of people in that (PES) category that were counted on Census forms (irrespective of their Census category).

#### NET DIFFERENCE IN CLASSIFICATION IN THE CONTACT SECTOR

Occasionally, the responses obtained for a person in the PES interview are not consistent with those obtained for the same questions in the Census. In addition, where a value has been imputed in the Census for a missing response (for example, Age, Sex, or State/territory of usual residence), it may differ from the PES response.

The net difference in classification for persons actually counted in the Census contact sector is equal to the estimated number of people in the given category as reported in PES, minus the number counted in that same category in the Census. For more information see the *Differences in classification* Technical Note.

#### PERSONS (IN THE CONTACT SECTOR) WITH CENSUS CATEGORY NOT-STATED

The Census contact sector contains some dwellings which were responding in the Census but returned only a partially completed Census form. For Census purposes, values for Age, Sex, State/territory of usual residence and Marital status are imputed during Census processing in cases where these items have been left blank. Missing values for other items remain 'not-stated' in the final version of Census counts (including Indigenous status and Country of birth). Persons in the contact sector with a Census category value of 'not-stated' contribute to net undercount estimates for the category in which they should have been counted as reported in the PES (e.g. Indigenous status).

#### UNDERCOUNT IN THE NON-CONTACT SECTOR

Non-responding dwellings in the Census are dwellings where the Census never obtained a return, and the dwelling could not be established as having been unoccupied on Census night. During Census processing, a 'hot-deck' imputation method (see *Glossary*) was utilised to impute people and their Age, Sex, Marital status and Place of usual residence into these dwellings. These values were, in many cases, based on information provided by the Census collector about the dwelling and its residents. Values for all other variables (e.g. Indigenous status) were set to 'not-stated' or 'not applicable', depending on the imputed value for Age.

Inevitably, the imputed values differ from the true but unknown values.

Imputed records made up the majority of the 2011 Census non-contact sector. Late returns made up only a small proportion of this sector.

An estimate of the undercount in the non-contact sector for a category of person is obtained by calculating the difference between the PES estimate of the number of people who should have been counted in the non-contact sector in that category (as reported in the PES) and the Census count of people in the non-contact sector for the category (including imputed person records within the category). For a category of persons

## COMPONENTS OF NET UNDERCOUNT *continued*

---

Contact and non-contact  
sector *continued*

UNDERCOUNT IN THE NON-CONTACT SECTOR *continued*

classified by Age, Sex, Marital status and State/territory of usual residence, this component of net undercount largely represents an estimate of imputation error in the Census for this category of imputed persons. For the Indigenous status and Country of birth categories, this component largely represents an estimate of persons who should have been counted in this category in non-responding dwellings.

## SURVEY ENUMERATION

### SCOPE OF THE 2011 PES

For Census and PES purposes, scope refers to the group of people about which information is required. Usually a set of rules is applied to determine whether a selected person is within this population of interest.

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. However, for practical reasons there were a number of areas, dwellings and people outside the scope of the 2011 PES. Some people who were selected in the PES sample did not meet these scope rules, however were identified during the PES interview and excluded from the PES population. Of the people present in Australia at the time of the PES, the following were not included:

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

The PES does not obtain information about people who died between Census and PES. However, it does obtain information about Australian residents who were overseas during the PES enumeration period and who departed after the Census, provided these people usually live with people remaining in Australia in a private dwelling selected in the PES.

In practice, the PES is used to produce estimates for the full Census scope, even though its actual scope and coverage is somewhat less.

#### *Remote areas and discrete Indigenous communities*

The 2011 PES included remote areas and discrete Indigenous communities. These areas and communities were first included in the 2006 PES. Prior to this, these were excluded from PES coverage due to operational difficulties around enumeration and around ensuring the independence of the PES from the Census, given the important role of local facilitators in these areas and communities.

Inclusion of these remote areas and discrete Indigenous communities in 2011 ensured a more complete geographic coverage of the PES. In 2011, the risk to statistical independence in these areas and communities was effectively managed through interviewer training and field staff procedures. No PES interviewer collected for Census in the same area where they were enumerating for PES.

#### *Non-private dwellings*

The 2011 PES sample excluded people living in non-private dwellings, as has been the case in previous PES cycles. Non-private dwellings are establishments which provide predominantly short-term accommodation for communal or group living, and often provide common eating facilities. They include:

- hotels;
- motels;

## SURVEY ENUMERATION *continued*

### *Non-private dwellings continued*

- 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. For further information on the operational difficulties around effectively enumerating non-private dwellings for the PES, see *Information Paper: Measuring Net Undercount in the 2011 Population Census, Australia* (cat. no. 2940.0.55.001).

### COVERAGE OF THE 2011 PES

Coverage refers to a set of rules designed to give each person in the PES sample, who is also in scope, a single chance of selection in the survey. These rules are implemented by associating each person with a single dwelling through a series of questions in the PES interview, such as where each person usually lives and whether they (or anyone else) are staying at their usual residence on 'PES night'.

An example of the need for coverage rules is for a visitor staying at a dwelling selected in the PES. If they report that someone else is staying at their usual residence during PES enumeration, then there is a possibility that this visitor could be included twice in the PES, once at their usual residence by another person, and again at their current location. Regardless of the low probability of both dwellings being selected in the PES, this respondent will be deemed out of scope as there is a chance that they would otherwise be included more than once in the PES.

### PES COLLECTION METHODOLOGY

#### *Field procedures*

Various field strategies were devised for the enumeration of the 2011 PES. Where possible, standard procedures were used when enumerating private dwellings and in discrete Indigenous communities. However, in some cases it was necessary to modify these procedures, particularly in discrete Indigenous communities, to account for language and cultural issues, while still ensuring the underlying concepts remained the same across both sample components.

The collection methodology of the PES was tested in a formal dress rehearsal, held after the Census dress rehearsal in August 2010. This allowed PES collection procedures to be tested. Efficiencies identified as a result of the dress rehearsal were then implemented before 2011 PES enumeration.

#### *Mainstream sample*

Specially trained PES interviewers collected data through face-to-face interviews which started around three weeks after Census night. Some telephone interviews were conducted by office staff, where the respondent made contact with the office and asked to complete the interview on the spot. All mainstream dwellings were enumerated using Computer Assisted Interviewing (CAI). Interviews were conducted with any responsible adult of the household who was asked to respond on behalf of all household members.

This collection methodology differed to the way Census collected its information, where most forms were self-completed.

## SURVEY ENUMERATION *continued*

### *Mainstream sample continued*

A major advantage of interviewer-administered questionnaires is that people can be provided with assistance if they are uncertain about the meaning of questions, and help is also given to ensure no questions are left unanswered.

To ensure a high response rate was achieved, the number of repeat visits made to non-contact dwellings was twice that of most other ABS household surveys.

### *Discrete Indigenous communities sample*

Interviews in discrete Indigenous communities were conducted by specially trained ABS interviewers with the assistance of facilitators recruited from within the selected community. The facilitator assisted in establishing rapport with respondents, helped identify residents of the selected households, and provided interpretation when needed.

To preserve the independence of the Census and the PES, every effort was made to recruit facilitators who were not involved in Census collection. Where this was not possible, PES interviewers ensured facilitators had only a limited role in the PES interview and provided assistance only where necessary, such as with language interpretation. In some communities it was acceptable to enumerate without the help of a facilitator, but generally they played an important role in being able to effectively enumerate in the communities.

The primary collection method in discrete Indigenous communities was also face-to-face interviewing using a CAI instrument. However, for practical reasons, a pen-and-paper interview (PAPI) questionnaire was also available, which provided interviewers with flexibility in situations where it was difficult to use a computer. Where PAPI forms were used, interviewers later transcribed the information into the CAI instrument, generally while still in the field.

### *Census follow-up*

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

Census follow-up periods for both the mainstream areas and discrete Indigenous communities were closely monitored by PES staff to ensure that the risk of overlap between the Census and PES was effectively managed. The timing of PES enumeration was based around the dates of Census follow-up, and was designed to start as soon as possible after Census activities had concluded in each area.

PES enumeration in mainstream areas began on Sunday 4 September 2011 and continued until Monday 3 October 2011. Almost all Census follow-up activities had been completed by the time PES enumeration began, although PES enumeration was delayed by a few days in a small number of areas to avoid overlap with final Census follow-up.

PES enumeration in discrete Indigenous communities began in early September 2011 and was completed by early October 2011. As PES enumeration of discrete Indigenous communities had a staggered start date (based on when Census field operations in each community were complete), there was no overlap between PES enumeration and Census collection in the selected communities.

### *Census follow-up continued*

As in previous PES cycles, special procedures were implemented for Census forms received after the start of the 2011 PES field work. These procedures ensured the independence of the Census and the PES was maintained, as receiving the PES primary approach letter or the arrival of a PES interviewer may have prompted the return of uncollected Census forms.

Any Census form received after the start of PES field work was flagged as a 'late return'. The treatment of late returns is explained in the *Identifying Census late returns* Technical Note.

### QUESTIONNAIRES

#### *Private dwelling questionnaire*

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

The PES also asked respondents:

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

The different addresses collected in the PES were used to search Census records to determine the number of times each PES respondent was included in the Census. Visitors to households included in the PES were also asked for their address of usual residence.

In addition to questions relating to Census night, the PES collected a small amount of information on dwelling tenure and structure.

#### *Change to the collection of Indigenous status*

The most significant change to the Private dwelling questionnaire was a change in how Indigenous status was collected.

In order to ensure that Indigenous status was effectively collected in the 2011 PES and that all Aboriginal and Torres Strait Islander persons were identified, the ABS removed a household 'screening' question that had been previously used in the PES. This meant that the Indigenous status question was collected for everyone in the dwelling, on a person by person basis. This aimed to reduce an assumed response bias in the previous questionnaire.

This represented a departure from the format of previous PES questionnaires, but was considered by the ABS to be a necessary departure based on trials of the change in the 2010 PES Dress Rehearsal. It was also well supported by a robust Interviewer training program and support materials, which reinforced the importance of collecting and recording Indigenous status correctly for all respondents.

This brought the Private dwelling questionnaire into alignment with the Discrete Indigenous Community questionnaire.

## SURVEY ENUMERATION *continued*

### *Discrete Indigenous community questionnaire*

As in 2006, a number of questions in the mainstream PES questionnaire were not considered applicable to people living in discrete Indigenous communities and were adapted to ensure information was collected in the most culturally appropriate manner. For example, it is more common for Indigenous persons in a community to be known by more than one name. The tailored discrete Indigenous community questionnaire ensured that sufficient information was collected to effectively match persons, taking into account the potentially different nature of matching for people in discrete Indigenous communities.

A community-level questionnaire was also completed for each selected community with the help of the community contact or council officer. Information collected, such as whether a significant event occurred at the time of the Census (for example, a sports carnival), helped with respondent recall and assisted in the completion of parts of the individual questionnaires.

### THE PES SAMPLE AND 2011 RESPONSE RATES

#### *PES sample*

The 2011 PES sample comprised two components; private dwellings (houses, flats, etc.) within the mainstream sample, and private dwellings within discrete Indigenous communities. As described earlier, non-private dwellings (hotels, motels, hospitals) were again excluded from the PES sample for operational reasons.

In total, approximately 35,000 dwellings were expected to respond to the PES, including around 500 dwellings from 28 discrete Indigenous communities. This was more than the 2006 PES sample, where approximately 32,000 dwellings were expected to respond, with 200 of these from 20 discrete Indigenous communities.

#### *Focused sampling*

In order to improve the estimates of Indigenous undercount, the 2011 PES increased the number of discrete Indigenous communities selected as well as the total number of dwellings selected within these communities. Also, since approximately three-quarters of the Indigenous population usually reside outside of discrete Indigenous communities (based on 2006 Census data), the number of selected dwellings in the mainstream sample was increased within areas identified as having a relatively high proportion of Indigenous persons.

These two methods were employed with the aim of being able to obtain a more accurate estimate of net undercount for Australia's Indigenous population by decreasing sampling error on the estimates.

For more information on the PES sample design, see the *Explanatory Notes*.

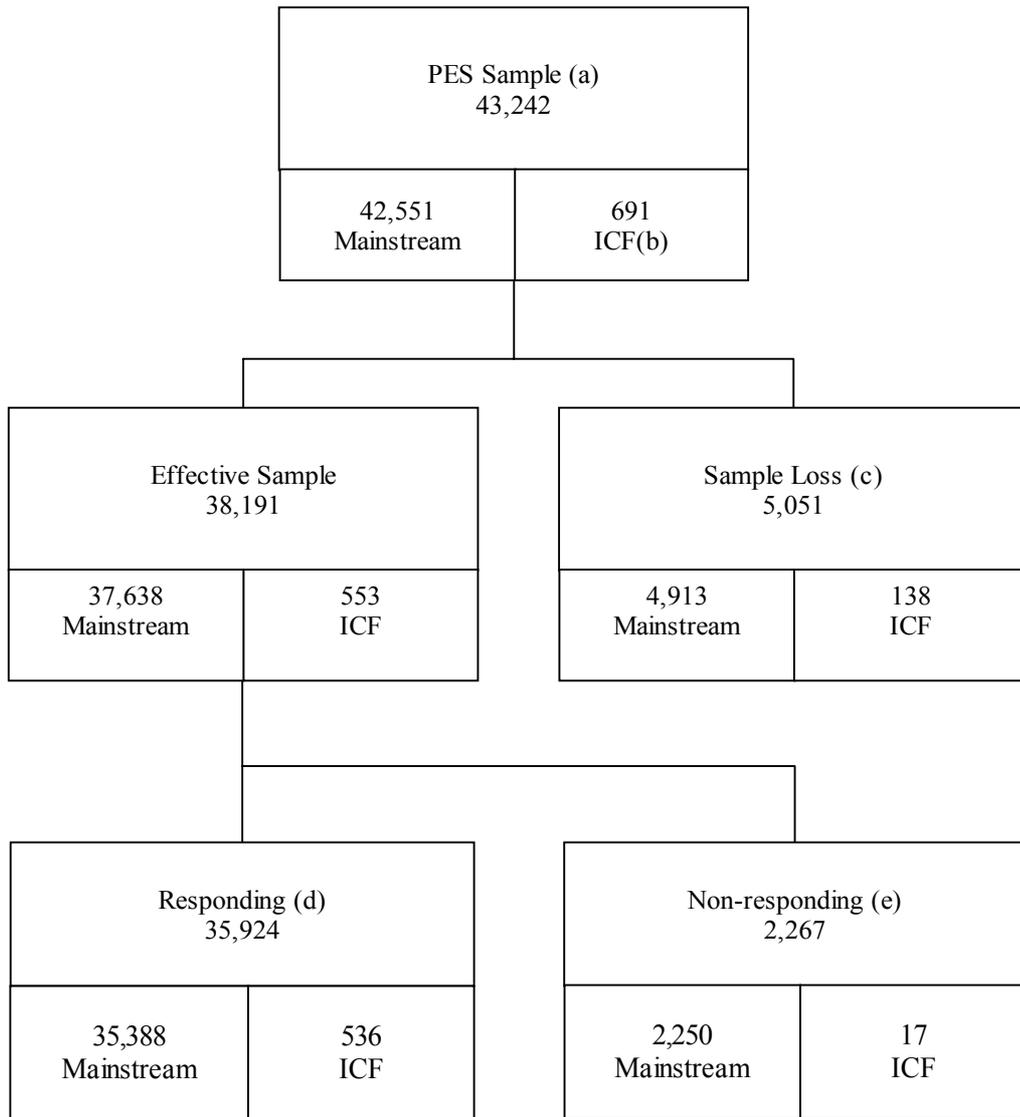
#### *2011 Response Rates*

After allowing for sample loss, the total number of fully responding dwellings (for both sample components) was almost 36,000. This represented a response rate of approximately 94% for the mainstream sample (the same rate as was achieved in 2006) and 96% for the discrete Indigenous communities (up from 91% in 2006).

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

**SURVEY ENUMERATION** *continued*

Figure 1. DWELLINGS, PES sample by main response type



(a) Number of dwellings selected.

(b) ICF refers to the Indigenous Community Frame, the sampling frame from which the discrete Indigenous community sample was selected.

(c) 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.

(d) Responding includes: fully responding and partially responding dwellings.

(e) Non-responding includes: full non-contact, language problems, death/illness, occupational health and safety, refusals, and incomplete dwellings.

## SURVEY ENUMERATION *continued*

2011 Response Rates  
*continued*

The expected and actual dwelling selections for the 2011 PES Mainstream and ICF samples are shown in Tables 14 and 15 respectively.

### 14 DWELLINGS, PES sample - Expected and actual dwelling selections—Mainstream sample

	<i>Expected selections</i>	<i>Actual selections</i>	<i>Expected fully responding</i>	<i>Actual fully responding</i>
	dwellings	dwellings	dwellings	dwellings
New South Wales	8 967	9 504	7 618	7 805
Victoria	7 614	8 039	6 450	6 727
Queensland	7 608	7 858	6 394	6 468
South Australia	4 380	4 592	3 694	3 869
Western Australia	5 382	5 576	4 410	4 607
Tasmania	2 373	2 469	1 971	2 063
Northern Territory	2 923	2 975	2 460	2 466
Australian Capital Territory	1 564	1 538	1 349	1 368
<b>Australia</b>	<b>40 790</b>	<b>42 551</b>	<b>34 346</b>	<b>35 373</b>

### 15 DWELLINGS, PES sample - Expected and actual dwelling selections—ICF sample

	<i>Expected selections (a)</i>	<i>Actual selections</i>	<i>Expected fully responding</i>	<i>Actual fully responding</i>
	dwellings	dwellings	dwellings	dwellings
New South Wales	—	—	—	—
Victoria	—	—	—	—
Queensland	56(3)	53	48	43
South Australia	21(2)	19	18	12
Western Australia	39(5)	40	33	25
Tasmania	—	—	—	—
Northern Territory	451(18)	579	382	453
Australian Capital Territory	—	—	—	—
<b>Australia</b>	<b>568(28)</b>	<b>691</b>	<b>481</b>	<b>533</b>

— nil or rounded to zero (including null cells)

(a) The number in the brackets refers to the number of discrete Indigenous communities from which the ICF dwellings were selected.

## LINKING AND MATCHING

### OVERVIEW

While the PES questionnaire collected information on whether a person was counted in the Census, the information was only used as a means of sequencing respondents through the questionnaire. Whether someone was missed, counted or counted more than once was determined through a linking and matching exercise where connections between PES information and the related Census information were established. This process involved a range of automated and manual processes, focused on linking and matching close to 100,000 PES records to their counterparts within around 22 million Census records.

This section describes the various processes that were used in the 2011 PES, beginning with input processing (where the data were prepared for linking) through to the final matching outcomes.

### INPUT EDITING - ADDRESS CODING

Address information was essential for matching people between the PES and Census. This was facilitated by identifying and coding all addresses collected in the PES to the Australian Statistical Geographic Standard (ASGS). Addresses were coded to a Collector Workload (CLW) and a Statistical Area Level 1 (SA1) by an automated program, the AddressCoder@ABS application.

Address coding was even more important to the processing of the 2011 PES than in previous cycles. The introduction of Automated Data Linking (ADL) made it necessary to have a Census enumeration area that could be used as a filtering variable for a number of the ADL runs, requiring the positioning of each address within a single SA1 (where possible).

In addition to this coding, the accuracy and consistency of other address elements (such as street names, suburbs and postcodes) had to be checked. The CLW was also important for subsequent clerical match and search processing, as it was the default starting point for clerical dwelling and person searching.

PES addresses were divided into two categories:

- Enumeration Addresses - the address at which the PES interview took place; and
- Search Addresses - including the usual address of visitors, the address at which PES respondents were located on Census night, the address at which respondents were included on a Census form, and any other addresses where the respondent may have been included on a Census form.

The PES allowed up to seven search addresses to be recorded, however the greatest number of search addresses recorded in the field for a single respondent in 2011 was two. Search addresses comprised around 10% of the total number of addresses recorded in the PES.

Table 16 shows that for every enumeration state, 70-85% of search addresses were located within the same state as the enumeration address (with the exception of the ACT), which allowed PES respondents to be linked to their Census location in a *state-based* run of ADL. The remainder were linked in non-state-based ADL runs, and were distributed predominantly throughout the three most populous states of New South Wales, Victoria and Queensland.

## LINKING AND MATCHING *continued*

INPUT EDITING -  
ADDRESS CODING *continued*

**16** SEARCH ADDRESS STATE, State of enumeration by state of search address—2011

Enumeration address	STATE OF SEARCH ADDRESS							
	NSW	Vic.	Qld.	SA	WA	Tas.	NT	ACT
	%	%	%	%	%	%	%	%
NSW	78.0	4.6	10.6	0.9	2.5	0.2	1.3	1.9
Vic.	7.8	75.0	10.6	1.6	2.7	1.1	0.6	0.6
Qld.	8.0	3.7	83.9	0.8	1.4	1.1	0.7	0.4
SA	5.0	4.7	7.6	76.1	3.8	0.8	1.1	1.1
WA	5.2	3.9	3.2	0.5	85.5	0.4	1.2	0.1
Tas.	4.1	8.7	11.9	2.2	3.9	68.8	0.2	0.2
NT	3.1	5.7	6.3	2.2	3.6	0.7	78.0	0.4
ACT	28.7	8.5	9.4	1.2	3.5	0.3	1.8	46.8

Search address data were collected directly from PES respondents and related to locations at which they were present up to two months previously. As such, the detail and accuracy of this information varied, ranging from perfectly spelt out addresses with street number, suburb, city and postcode, to 'vague addresses' such as "a motel in Sydney". Therefore, in order to code search addresses successfully, an additional two-stage process was carried out, as detailed below.

Address repair was conducted on all search addresses, that is, any address given in the PES that differed from the enumeration address. This was done manually by a team of coders who reviewed the address text fields and amended them through a variety of techniques. Quality assurance was then conducted.

Address coding was undertaken after address repair with the aim of identifying the correct geographic areas (Meshblock [MB], CLW and SA1) for all addresses (enumeration and search addresses), according to the ASGS. This was done by first running all addresses through the AddressCoder@ABS application. Quality assurance for this automated process involved a complete review of the addresses that were amended by the automated coder in order to fit into a geographic classification, and retention of all original addresses. Those records which were not automatically coded were then sent to a coding team for manual processing. This manual process utilised various methods, including mapping software, to thoroughly scrutinise addresses and achieve the most accurate geographic coding possible. Further quality assurance was then undertaken.

INPUT EDITING - ITEM  
DERIVATIONS

Most data on the PES file were of a sufficient quality to feed into both linking and matching processes and later output processing, without further detailed editing. However, certain validation processes highlighted issues that required amendments to be made.

Derivations were used to correct Age/Date of Birth (DOB) and Marital Status responses. Where one data field was missing (e.g. Age), but a similar one was available (e.g. DOB), the missing field was derived and populated. Derivations were also created by examining individual 'person level' records to derive 'dwelling level' information for the relevant dwelling (e.g. the number of Usual Residents and Visitors in the dwelling, or whether the dwelling contains any Indigenous respondents).

### INPUT EDITING - STANDARDISATION

In preparation for ADL, PES data were repaired and standardised through a three-stage process, converting it into a format that could then be compared with similarly standardised Census data through both automated and manual systems.

- Data Repair was conducted to clean the data by removing non-alphabetic characters and capitalising the remainder, and by removing additional spaces.
- Name standardisation involved converting common nicknames, abbreviations, misspellings or variations on a name to their 'origin name' (e.g. Beth, Eliza or Libby were converted to Elizabeth).
- Data transformation/recording was undertaken to ensure that each variable was comparable to its Census counterpart (e.g. ensuring PES numeric identifiers for Indigenous status matched to those of Census). Additional variables were then created from the existing PES data.

### AUTOMATED DATA LINKING - LINKING

Automated Data Linking refers to the use of probabilistic linking methods to determine possible links between Census and PES data in an automated fashion, and was used as the primary linking method in 2011. Its introduction followed an evaluation exercise undertaken by linking experts within the ABS after the 2006 PES.

ADL uses a range of personal and address characteristics, to evaluate the likelihood that a PES record and a Census record pertain to the same individual. The software used in both the 2006 quality study and the 2011 PES was Freely Extensible Biomedical Record Linking (FEBRL), which was developed at the Australian National University.

ADL provided the opportunity to match persons in the 2011 PES with those in the 2011 Census who would have previously been too difficult to match, given the constraints of prior technology and processes. The key gains in matching effectiveness and efficiency provided by ADL in 2011 included:

- the ability to conduct a more comprehensive search for PES respondents than was possible from previous clerical matching processes;
- the ability to locate PES respondents at Census night addresses that were not identified in the PES; and
- a reduced requirement for clerical matching resources.

A number of different linking runs were used in 2011 to compare PES and Census records, each of which focused on a slightly different combination of name, address and demographic variables. At the beginning of each run, a list of PES and Census records was obtained by selecting a subset of the PES and Census datasets based upon agreement on a small number of variables. This process, called 'blocking', was used to stratify identified links (i.e. links at earlier runs took precedence), and to reduce the quantity of poor quality links returned in each run. Table 17 shows the ADL runs and the relevant 'blocking' fields used in each run.

AUTOMATED DATA  
LINKING - LINKING  
*continued*

### 17

 ADL RUNS AND RELEVANT BLOCKING FIELDS

<i>ADL run</i>	<i>Blocking field</i>
1A	SA1 (Statistical Area Level 1)
1B	CLW (Collector Workload)
2	Postcode, Year of birth
3	State, Initial letter of standardised first name, Initial letter of surname, Marital status
4	Date of Birth (Day, Month, Year), Marital status

Potential links were then assessed by assigning weights that reflected the level of agreement on selected data items from the two records. Large positive weights indicated probable matches, while large negative weights were observed for probable non-matches. These weights were then grouped and organised in the processes of CARDS and DLR, which we now describe.

AUTOMATED DATA  
LINKING - CARDS AND  
DLR

Important to the effective use of ADL were a series of processes run after ADL output was obtained. The Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS) process collated, processed, identified and rated the most plausible links from each ADL run for all PES respondents. The process then combined the person links from all ADL runs and removed any duplicates. The resulting output was a single numeric 'Person Link Rating' (PLR) for each individual linked pair (a PES respondent and a Census respondent) ranging from 0.1 to 10.0 based upon agreement on various characteristics.

Person links were then grouped into Platinum, Silver and Tin categories, based on their PLRs.

- Platinum - those links which were so strong that clerical examination was not required;
- Silver - those which were convincing links, but required some clerical review; and
- Tin - those which were linked on broad fields (e.g. surname and age) and which were not considered informative.

The CARDS process concluded by identifying and rating dwelling links through the Dwelling Link Rating system. In order to identify dwelling links, all person links within one PES dwelling were grouped together into a 'dwelling'. Dwelling links were then created between that PES dwelling and the Census dwelling(s) of the linked Census respondents. A 'Dwelling Link Rating' (DLR) was then assigned to each dwelling link based on the number of people linked between the PES and Census dwellings proportional to the number (if any) that were not linked, and the PLRs of the links.

Similar to the person links, dwelling links were then stratified into Platinum, Silver and Tin categories based upon their DLRs, allowing strong links (e.g. those with many person links and high PLRs) to be investigated before weaker links (e.g. with few person links and low PLRs). For a dwelling link to be rated as platinum, all its persons had to have a platinum PLR and be linked to Census persons within a single dwelling. If there were missing people, in either the PES or Census dwelling, or not every person had a Platinum link, the maximum rating the dwelling could be assigned was Silver. As with person links, the remainder of dwellings were placed into either Silver or Tin, based on the quality of the person links within.

## LINKING AND MATCHING *continued*

### PROCESSING IN THE PES MATCH AND SEARCH SYSTEM (MSS)

While ADL was the next step in the evolution and continual improvement of PES processing, ADL could not entirely replace the clerical decision-making process that has previously been at the core of PES processing. Clerical judgment will always be required to resolve the more complex or ambiguous cases and be used as a means of quality assuring automated processes. Some adjustments to the clerical match and search processes were necessary in 2011 to ensure that the relative strengths of both ADL and the MSS were fully realised.

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

The initial phase of MSS processing involved confirming whether the ADL output was correct. Once a dwelling link was confirmed, the Census person records for that dwelling were clerically compared with the PES person records. The information compared included name, sex, date of birth, age, marital status, Indigenous status and country of birth. The extent to which each of these variables was the same, in both the PES and the Census, determined the ADL match status of the pair and the level of match.

### AUTOMATED DATA LINKING - LINK UPGRADING

Link Upgrading was a process of secondary examination after the main runs of ADL and MSS clerical review were completed for each state. Once MSS had been run on the Silver links for each state, the highest rated tin links for those PES people who were not matched were extracted (i.e. effectively upgraded) and entered into a second run of MSS processing.

### INTENSIVE SEARCH ACTIVITIES

Once all ADL links were reviewed, the final phase of MSS processing was to conduct an intensive clerical search for persons and dwellings not matched as a result of ADL-enabled processing. This was done by searching CLWs (and neighbouring CLWs) for addresses provided by respondents during the PES interview (search addresses), in order to locate possible Census forms where that person was included. This followed 2006 methodology, which is described in *Census of Population and Housing - Undercount, 2006* (cat. no. 2940.0) and *Census of Population and Housing - Details of Undercount, 2006* (cat. no. 2940.0).

### MSS QUALITY ASSURANCE AND ADJUDICATION PROCESSES

To ensure the accuracy of MSS processing, quality assurance (QA) procedures were used in the match and search process whereby all PES records processed in MSS were processed a second time by a different clerk. There was no identifier on the workloads that allowed the PES processors to know whether they were processing an 'original' or a QA workload. Where the initial and the QA processing outcomes corresponded, the initial match status was accepted. Where there was a discrepancy between the initial match status and the QA match status, the records were flagged for adjudication by a senior officer who reviewed all information and determined which match status was correct. Where both the initial and QA records were deemed to be inaccurate, the adjudicator reprocessed the record.

## LINKING AND MATCHING *continued*

**MSS QUALITY ASSURANCE AND ADJUDICATION PROCESSES** *continued*      The QA process was also useful in identifying potential processing issues or areas where processors were having difficulty. This allowed ongoing feedback to be provided to the PES processors and contributed to the overall quality assurance of PES processing.

**DISCRETE INDIGENOUS COMMUNITY PROCESSING**      MSS processing for discrete Indigenous communities followed the 2006 approach and involved searching the entire community for a person match, rather than just searching within a single dwelling. Person matching in discrete Indigenous communities used the same rules for determining a match as in the mainstream component, but allowed for the use of up to two alternate names for each person when matching on name.

**CONFIDENCE OF MATCH DECISIONS**      Table 18 shows the matching outcomes from the 2011 PES linking and matching processing. Of the 94,539 total mainstream matches, 52,398 (or 59.8%) were matched without clerical review, 34,653 (or 39.5%) were matched after clerical review of ADL links, with the remaining 594 (or 0.7%) matched as a result of intensive search processing.

**18** MATCHING OUTCOMES (a)—2011

	Matches.....	
	no.	%
<b>Mainstream</b>		
Matched		
Used in estimation		
ADL Platinum (not clerically reviewed)	51 791	59.9
ADL Silver (clerically reviewed)	34 081	39.4
Intensive search	549	0.6
Not used in estimation		
ADL Platinum (not clerically reviewed)	607	49.6
ADL Silver (clerically reviewed)	572	46.7
Intensive search	45	3.7
Total matched		
ADL Platinum (not clerically reviewed)	52 398	59.8
ADL Silver (clerically reviewed)	34 653	39.5
Intensive search	594	0.7
Not matched	6 894	7.1
Total mainstream	94 539	100.0
<b>ICF</b>		
Matched	2 528	87.1
Not matched	373	12.9
Total ICF	2 901	100.0
<b>Total</b>	<b>97 440</b>	<b>100.0</b>

(a) This table includes multiple matches for persons matched more than once. Therefore, totals do not sum to the total number of matched persons.

**STATISTICAL IMPACT STUDY PROCESSING**      In order to assess the impact of ADL on 2011 PES estimates, a Statistical Impact Study was conducted during linking and matching processing. For further information see the *Statistical Impact of ADL* Technical Note.

## ESTIMATION

### OVERVIEW

Weighting is the process of adjusting results from a sample survey to infer results for the total in-scope population. The weight can be considered an indication of how many population units are represented by a sample unit. Essentially, the PES weighting methodology attaches weights to each responding person in the PES, to enable estimates of undercount to be produced.

The representation of PES weighting and estimation has been simplified in this section, to provide a broad overview. It should be considered illustrative, with the methods applied discussed in more detail in *Research Paper: An Estimating Equation Approach to Census Coverage Adjustment, May 2007* (cat. no. 1351.0.55.019).

PES weighting began with a 'selection weight' for each dwelling, which was essentially equal to the inverse of the probability that a dwelling was selected in the PES sample. Dwelling weighting adjustments were then made to compensate for the different rates of PES non-response in different dwelling types and areas of Australia. Next, person weighting adjustments were applied, allowing for undercoverage and non-response within dwellings. Larger adjustments were made for categories of people (such as young adult males) who were harder to contact and interview successfully in the PES, as is generally the case with most surveys.

### DWELLING WEIGHTING

Each dwelling in the PES sample was given a 'dwelling weight' so that the PES sample represented all private dwellings in Australia. The PES sample was designed to ensure each private dwelling in a state or territory had a known non-zero probability of selection. The inverse of this probability is the dwelling's selection weight. In practice, certain types of dwellings are more likely to be missed in the PES. Adjustment for this was made according to a dwelling's post-stratum, where dwellings were assigned to a particular post-stratum (i.e. one of many different groups) based on the following variables:

- Census response category at the start of the PES enumeration period (responding, non-responding or unoccupied on Census night), plus a category "missed" for PES dwellings that were not successfully matched to a Census dwelling (with a small degree of category collapsing required, based on the matching outcome outlined in the *People and dwellings missed* Technical note);
- Dwelling structure; and
- Region (six states and NT divided into capital city and rest of state, plus ACT, that is, 15 regions).
- Separate post-strata were formed for dwellings sampled in discrete Indigenous communities (ICF).

Selection-weighted estimates of dwelling numbers were obtained for each post-stratum as the total of the selection weights for PES dwellings in that post-stratum. These will typically understate the actual Census counts. Initial dwelling weights were obtained by multiplying the selection weights by a factor to adjust for this. For dwellings matched to a Census-responding dwelling, the factor used was the Census count divided by the selection-weighted estimate for the dwelling's post-stratum. For other dwellings, a factor from similar Census-responding dwellings (those with the same region and dwelling structure) was applied. These initial weights are applied to all dwellings in the PES sample, even those that were not fully responding in the PES.

## ESTIMATION *continued*

### DWELLING WEIGHTING

*continued*

*Initial dwelling weight (IDW) =*

*Selection weight (S) x Factor for similar Census-responding dwellings (F)*

*F = Census count for post-stratum (C) / Sum of selection weights for post-stratum (E)*

$$IDW = S \times F = S \times C / E$$

A final dwelling weight was derived by weighting up responding dwellings in the PES in each post-stratum to represent non-responding dwellings deemed to be occupied:

*Final dwelling weight (FDW) =*

*Initial dwelling weight (IDW) x Sum of initial dwelling weights for occupied*

*dwellings (T) / Sum of initial dwelling weights for responding PES dwellings (R)*

$$FDW = IDW \times T / R$$

For example, consider a Census-responding post-stratum in which a group of 27 dwellings were selected by PES, each with a selection probability of 0.25. The selection weight (S) for each dwelling is 4 (i.e. 1 / 0.25). The total selection weight of the dwellings counted in the Census is E = P \* S (i.e. 27 \* 4) which equals 108. If 135 dwellings were counted in the Census (C), the factor F to be applied is 135/108. The initial dwelling weight (IDW) for all these dwellings is S \* F (i.e. 4 \* 135/108), which is equal to 5.

$$\begin{aligned} IDW &= (1/0.25) \times 135 / (27 \times 4) \\ &= 4 \times 135 / 108 \\ &= 5 \end{aligned}$$

Furthermore, suppose that of the 27 dwellings selected by the PES in this post-stratum (total weight T=27 x 5=135), only 25 dwellings responded in the PES (total weight R=25 x 5=125). The cumulative weight of the 2 non-responding dwellings will be redistributed to the other dwellings in the sample in order for the sample to reflect the independently estimated distribution of the population. The adjustment factor to account for non-response is T/R (i.e. 135/125), which is 1.08. This factor is applied to the initial weight (IDW) to derive a final weight (FDW) for this dwelling.

$$\begin{aligned} FDW &= IDW \times (27 \times 5) / (25 \times 5) \\ &= 5 \times 1.08 \\ &= 5.4 \end{aligned}$$

### PERSON WEIGHTING

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

Estimates of the number of people who should have been counted in the Census based on dwelling weights would only represent the population of people who were in private dwellings at the time of the PES, given the scope of the PES was limited to private dwellings only. That is, they would underestimate the private dwelling population at the time of the Census because some people in private dwellings on Census night would have been in non-private dwellings, overseas, or may have even been 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, adjustments were made to the dwelling weights in order to give a person weight. The initial person weight adjustments were chosen to ensure that the PES estimates of people counted in private dwellings (other than late return or imputed dwellings - see *Components of net undercount*, section 4) in a set of benchmark categories matched the actual Census counts for these categories. The weight adjustment applied was such that all persons with identical PES category values received the same weight adjustment, whether or not they were counted in the Census.

The variables used to form these benchmark categories were:

- Region;
- Sex;
- Age (by 16 age groups);
- Country of birth;
- Marital status;
- Indigenous status;
- Whether sampled in a discrete Indigenous community (ICF) dwelling;
- Whether sampled in the Northern Australia region (see Glossary for definition of Northern Australia region); and
- Whether sampled in a 'hard to enumerate' area (see Glossary for definition of 'hard to enumerate').

For information on the resolution of Census not-stated values for Indigenous status and Country of birth for use in benchmarking, see the *Explanatory Notes*.

As a final step in weight adjustment, the initial person weights were adjusted so that the PES estimates represented not only people in private dwellings but also people in non-private dwellings, to match the scope of the Census. This final step used only region, age and sex as information on other items was not reliable for non-private dwellings.

Intuitively, a 'good' set of weights for the PES should ensure that if the PES were used to estimate the actual Census count, the PES would get the 'right' answer. The above step ensured this was the case. Technically, this is 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.

## ESTIMATION *continued*

---

### PERSON WEIGHTING

*continued*

For example, consider the benchmark category consisting of Queensland females aged 35 to 39 years. Suppose that the dwelling-weighted estimate of persons counted in the Census in this category (in dwellings that were not imputed or late returns) was 10,000, but the actual Census count of such persons was 10,500. This would lead to adjusting the person weights of all Queensland females aged 35 to 39 in the PES sample to exceed their dwelling weights by about 5 percent ( $= 10,500/10,000$ ). This adjustment is applied regardless of whether they were counted in the Census. In practice, persons contribute to a number of benchmark categories, so the actual weight change for an individual person in this benchmark category could be lower or greater than 5 percent.

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

### ESTIMATES OF NET UNDERCOUNT

The weighted estimate of population for a category of persons is obtained as the sum of the person-weights of persons who should have been counted in that category in the Census. The final PES estimate adjusts this figure for consistency with Census counts, by adding on the Census count for the category from the responding Census dwellings and subtracting the weighted estimate of this (i.e. the sum of the person-weights of persons who were actually counted in the category in responding Census dwellings).

Net undercount for any category of person is the difference between the final PES estimate of population (i.e. the number of people who should have been counted in the Census) and the actual Census count (including imputed persons in non-responding Census dwellings). This calculation takes into account the components that are described in *Components of net undercount* (section 4).

## EXPLANATORY NOTES

### SCOPE OF THE 2011 CENSUS

- 1** The aim of the Census is to count every person who spent Census night in Australia, with the exception of foreign diplomats and their families. Visitors to Australia are in scope 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 in the Census.
- 2** The scope of the Census 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. People outside Australia who are not required to undertake migration formalities, such as those on oil and gas rigs off the Australian coast, are in scope.
- 3** 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.
- 4** Details about the 2011 Census content, collection operations, confidentiality and privacy protection, processing and evaluation activities are contained in *2011 Census Nature and Content* (cat. no. 2008.0).

### CENSUS COUNTS

#### *Usual residence*

- 5** The Australian Census counts people where they stayed on Census night. This means that the Census was conducted on an 'actual location' or 'place of enumeration' basis. There is, however, a need for data based on place of usual residence, and Census counts are available on this basis.
- 6** For usual residents of Australia, 'place of usual residence' for the 2011 Census is defined as the address at which a person has lived or intends to live for six months or more in 2011. 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.
- 7** 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.
- 8** 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 publication *Australian Demographic Statistics, December quarter 2011* (cat. no. 3101.0), released on 20 June 2012.
- 9** Estimates presented in this publication are on a *place of usual residence* basis (unless otherwise noted).

#### *Imputation*

- 10** Census counts include persons whose Census form was partially completed. During Census processing, values for age, sex, marital status, and state/territory of usual residence are imputed if they were left blank on the Census form. Values for all other variables left blank (e.g. Indigenous status, country of birth) are set to not-stated or not applicable, depending on the (possibly imputed) age of the person.
- 11** Census counts also include imputed person records for non-responding dwellings (for further information around imputed persons see *Components of net undercount* (section 4) and the *Identifying Census late returns* Technical Note). These records have values imputed for age, sex, marital status, and place (CLW, SA1 and State/territory) of

## EXPLANATORY NOTES *continued*

### *Imputation continued*

usual residence. Values for all other variables (e.g. Indigenous status, Country of birth) are set to not-stated or not applicable, depending on the imputed value for age.

**12** Therefore, care should be taken when interpreting undercount estimates for Indigenous status and country of birth. For example, there were 1,058,447 persons (4.9% of the Census count) whose Indigenous status was not stated in the 2011 Census. These persons do not contribute to either the Indigenous or non-Indigenous Census count, although they do contribute to total Census counts by age, sex, marital status, and state/territory of usual residence.

### *Resolution of Census not-stated value*

**13** 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.

**14** 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 2011 PES benchmarks.

**15** The imputation method involved imputing both variables together. This required the assumption that anyone who is imputed as Indigenous is born in Australia.

**16** 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 2011 Census
- non-Indigenous, born elsewhere.

**17** 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.

**18** 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).

**19** 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 increasing importance, these classes were:

- SA2 (Geographic unit averaging 10,000 persons; see *Glossary*);
- Census form type;
- Age Group – 5 year age groups to 75+; and
- Sex.

### *Census response type*

**20** Although all persons resident in Australia should have been counted in the Census, not all dwellings would have received a Census form. This is because not all dwellings were habitable or, in the case of a diplomatic dwelling, did not contain people within scope of the Census.

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

## EXPLANATORY NOTES *continued*

### Census response type *continued*

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

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

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

**23** For PES purposes, dwellings that provided a Census return after the commencement of PES enumeration were flagged as 'late returns' and treated accordingly during PES processing. For information around the treatment of late returns, see the *Identifying Census late returns* Technical Note.

**24** During Census processing, the number of usual residents for non-contact and refusal dwellings was imputed, as well as some personal characteristics.

### Census Data Quality

**25** 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.

**26** 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.
- *Respondent error*: Because processing procedures cannot detect or repair all errors made by people in completing their forms, some remain in final data; and
- *Undercount*.

**27** For further information on sources of error in the Census, refer to the appropriate entries in the *2011 Census Dictionary* (cat. no. 2901.0), released on 23 May 2011.

**28** The three major changes to the 2011 PES were the following:

- Introduction of Automated Data Linking (ADL) to complement the existing clerical Match and Search System. For more information about ADL see *Linking and Matching* (section 5).

### MAJOR CHANGES TO THE 2011 PES

## EXPLANATORY NOTES *continued*

### MAJOR CHANGES TO THE 2011 PES *continued*

- Expansion of the PES sample to incorporate an increased focus on enumeration of discrete Indigenous communities and focused sampling in parts of the mainstream sample for which there was a relatively high proportion of Indigenous persons. Enhancements to questionnaires (in particular the removal of a screening question, to improve Indigenous identification), field materials and training were also made in support of the sample design. For more information about the 2011 PES sample and sampling procedures see *Survey Enumeration* (section 5).
- Transition to the Australian Statistical Geography Standard (ASGS), which replaced the Australian Standard Geographical Classification (ASGC) as the framework for PES and Census geography.

### *Impact of the transition to the ASGS*

**29** With the geographic classification used in the Census changing from the Australian Standard Geographical Classification (ASGC) to the Australian Statistical Geography Standard (ASGS), a mixture of geographies was used in the 2011 PES, with different geographic units used at different stages of the survey cycle. This development presented challenges for the ABS in relation to the sample design, however these challenges were well managed throughout the complete PES cycle.

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

**31** The second challenge posed was the spatial definition of discrete Indigenous communities, some of which differed between the two geography frameworks. This was managed by a mapping exercise which compared the two geographies for the 28 communities selected in the PES sample.

### 2011 PES SAMPLE

#### *Mainstream private dwelling sample*

**32** The mainstream sample of private dwellings was derived from the 2006 Monthly Population Survey Parallel Sample frame, which is based on the Statistical Division and Subdivision structure of the Australian Standard Geographical Classification (ASGC). This sampling framework first divides Australia into 100 geographical areas. These areas are then divided into strata according to population density, remoteness and growth, then:

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

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

**34** The 2011 PES sample departed from the Parallel Sample by selecting extra dwellings, and in some cases extra CDs, in areas with higher proportions of Indigenous households. This measure was introduced in the 2011 PES to improve estimates of Indigenous undercount.

#### *Discrete Indigenous communities sample*

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

## EXPLANATORY NOTES *continued*

### *Discrete Indigenous communities sample continued*

**36** For the purposes of PES sampling, discrete Indigenous communities were grouped into 'sets' comprising main communities and their associated outstations. The selection of main communities for the 2011 PES was undertaken with probability proportional to the size of the set. The aim was to select as representative a sample as possible while also considering cost constraints, reasonable interviewer workloads and expected sample size.

**37** Where a community was selected in the PES sample, selection of dwellings within that community followed the same procedure as for the selection of private dwellings within selected blocks in the non-ICF sample component. That is, the interviewer compiled a list of all the dwellings within the community and a sample was taken using systematic equal probability sampling.

**38** A selection of outstations associated with each selected main community was also included in the sample. Each outstation had an equal chance of selection and, once selected, all dwellings within the outstation were enumerated.

### *Sampling strategies to improve Indigenous population estimates*

**39** Initial attempts to improve the quality of Indigenous undercount information occurred in the 2006 PES when remote areas and discrete Indigenous communities were enumerated for the first time. The 2011 PES strived to further improve the estimates of Indigenous undercount by increasing the number of discrete Indigenous communities selected (28 in 2011 compared with 20 in 2006), as well as the total number of dwellings selected within the community sample (around 700 dwellings in 2011, compared with 350 in 2006). As predicted, this increase in sample resulted in a decrease in the sampling error of estimates of Indigenous undercount.

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

**41** The focused sampling approach involved a combination of extra dwellings selected from within the parallel sample, as well as some additional selections from CDs outside the parallel sample. This focus on obtaining an increase in Indigenous respondents within the mainstream sample was of particular importance, given approximately three-quarters of the Aboriginal and Torres Strait Islander population usually reside outside of discrete communities. This strategy was successful in increasing the number of Indigenous respondents in the PES in 2011 (6,401 persons, compared with 3,354 persons in 2006).

### *Reliability of undercount estimates*

**42** 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. The estimates are also subject to non-sampling error. For more information about sampling and non-sampling error see the *Sampling Error* Technical Note.

# TECHNICAL NOTE 1 IMPROVEMENT IN COLLECTION OF INDIGENOUS STATISTICS

## OVERVIEW

**1** The ABS has undertaken an additional and extensive quality assurance process to understand the components of Indigenous net undercount in 2006 and 2011. This process identified that the main explanatory factor behind the change was the improved PES methodology and procedures, which led to better identification of Aboriginal and Torres Strait Islander persons in the 2011 PES and a noticeably different net category change outcome. Net category change is explored in detail in *Components of net undercount* (section 4) and in the *Differences in classification between the PES and Census* Technical Note.

**2** As a consequence of this important improvement in methodology and procedures, the 2006 and 2011 PES estimates are not directly comparable, with the 2011 methodology providing an improved estimate of net undercount for Aboriginal and Torres Strait Islander people.

## CHANGE TO THE QUESTIONNAIRE

**3** In order to ensure that Indigenous status was effectively collected in the 2011 PES and that all Aboriginal and Torres Strait Islander persons were identified, the ABS removed a household 'screening' question that had been previously used in the mainstream PES questionnaire. This meant that the Indigenous status question was collected for every person in the dwelling. This aimed to address a suspected response bias in the 2006 questionnaire.

**4** This represented a departure from the format of previous PES questionnaires, but was considered by the ABS to be a necessary departure based on trials of the change in the 2010 PES Dress Rehearsal. It was also well supported by a robust Interviewer training program and support materials, which reinforced the importance of collecting and recording Indigenous status correctly for all respondents.

## IMPACT

**5** The impact of the improved collection of Indigenous status in the PES is evident through analysis of the components of the 2006 and 2011 net undercount for Aboriginal and Torres Strait Islander people. As the following table shows, the net category change in 2006 effectively reduced the net undercount rate in the contact sector by around 6% (or 30,797 persons), in contrast to 2011, where it resulted in a slight increase of 0.8% (or 5,128 persons). The remainder of the net undercount rate was therefore similar between 2006 (17.5%, or 11.5% plus 6%) and 2011 (16.5%, or 17.2% minus 0.8%, in rounded terms), a non-statistically significant difference.

## COMPONENTS OF NET UNDERCOUNT(a), Indigenous status—2006 and 2011

.....							
..... UNDERCOUNT IN CONTACT SECTOR .....							
	PES population estimate	Persons missed in the Census	Net difference in classification	Persons with Census category not-stated	Total undercount in the contact sector	Undercount in non-contact sector	Net undercount
<b>2006</b>							
Indigenous (no.)	513 977	54 056	-30 797	7 989	31 248	27 930	59 178
Proportion of PES population estimate (%)	100.0	10.5	-6.0	1.6	6.1	5.4	11.5
<b>2011</b>							
Indigenous (no.)	662 335	56 650	5 128	7 009	68 787	45 402	114 188
Proportion of PES population estimate (%)	100.0	8.6	0.8	1.1	10.4	6.9	17.2
.....							

(a) 2006 and 2011 estimates are included in this table to support decompositional analysis only, and are not strictly comparable.

# TECHNICAL NOTE 1 IMPROVEMENT IN COLLECTION OF INDIGENOUS STATUS *continued*

IMPACT *continued*

**6** The increase in the number of respondents in the 2011 PES who were identified as Aboriginal and Torres Strait Islander also resulted in a higher incidence of differential identification of Indigenous status between the PES and Census, upon which the category change is based. As the following table shows, in 2006 there were 116 PES respondents who were Aboriginal and Torres Strait Islander in the PES and non-Indigenous in the Census, compared with 283 respondents with the reverse combination. In contrast, in 2011 there were 268 and 316 respondents. While the category change is based on weighted estimates, rather than unweighted sample counts, the change in the relationship between these differential classification categories is clear.

**7** In 2006, the identification of Indigenous status in the PES suggested that Census counts of Aboriginal and Torres Strait Islander persons were greater than they should have been and a net category change of -30,797 was applied to the estimate of net undercount in the contact sector. In 2011, the relationship between the two categories of differential classification was close to balanced, with a net category change of just 5,128 applied to the estimate of net undercount in the contact sector.

## DIFFERENCE IN CLASSIFICATION, Indigenous status, Responding PES persons — 1991–2011

	PES Indigenous, Census non-Indigenous	PES non-Indigenous, Census Indigenous	Ratio of two differentially identified groups	Net difference in classification
	no.	no.	%	no.
2011(a)	268	316	0.85	5 128
2006(b)	116	283	0.41	-30 797
2001	133	246	0.54	na
1996	86	227	0.38	na
1991	174	192	0.91	na

- (a) Care should be taken when comparing 2011 estimates with previous years due to changes made to PES linking and matching methodology. For more information see Linking and Matching (section 5).
- (b) Care should be taken when comparing estimates from 2006 onwards with previous years due to a new estimator, Prediction Regression (PREG), being implemented in 2006, as well as the inclusion of remote areas and discrete Indigenous communities in the PES sample from 2006. For further information see the Explanatory Notes of Census of Population and Housing - Undercount, 2006 (cat. no. 2940.0).

**8** The 2011 unweighted sample count ratio of the two differentially identified groups (0.85) was closest to the ratio seen in the 1991 PES (0.91) and was noticeably different to the ratio seen in the previous 3 surveys. This suggests that in 2011, prior to estimation, both types of differential identification occurred roughly a similar number of times, reflecting a different classification relationship between the two collections over time.

**9** In weighted terms, the differential classification in 2011 contributed to the net category change of 5,128. That is, in 2011 PES identification of Indigenous status suggested the Census counts of Aboriginal and Torres Strait Islander persons were less than they should have been. This accounts for most of the difference between 2006 and 2011 estimates of net undercount.

## TECHNICAL NOTE 2 DIFFERENCES IN CLASSIFICATION

### DIFFERENCES IN CLASSIFICATION BETWEEN THE PES AND CENSUS

- 1 Occasionally, the answers obtained for a person in the PES interview are not consistent with the answers obtained for the equivalent questions in the Census. There are a number of reasons why 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, compared to the PES which is administered by a trained interviewer;
  - different people may provide Census and PES responses;
  - the correct response 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 is married or divorced, may not be identified; and
  - the Census may contain a 'not-stated', or imputed, response while the PES will have a valid response.
- 2 This difference in classification is most prevalent for the Indigenous status and Country of birth responses, as these are missing in a number of Census forms, and imputation is not carried out for either of these variables in the Census.
- 3 The two tables below contain unweighted PES counts of the differences in classification between PES and Census for Indigenous status and Country of birth responses. These tables are counts of all responding people in the PES and include a record for each match made between the PES and Census where a difference in Indigenous status or Country of birth classification has occurred. The match status in the tables only includes those matches used in estimation.
- 4 The majority of person matches between PES respondents and their corresponding Census records show the same responses for Indigenous status. For example, 93.6% of matched people who reported as Indigenous in the PES also reported as Indigenous in the Census, and 98.4% of those who reported as non-Indigenous in the PES also reported as non-Indigenous in the Census.

#### DIFFERENCE IN CLASSIFICATION (a), Indigenous status—Responding PES persons—2011

Census response	PES RESPONSE					
	Indigenous		Non-Indigenous		Total	
	no.	%	no.	%	no.	%
<b>Matched</b>	<b>5 428</b>	<b>100.0</b>	<b>83 517</b>	<b>100.0</b>	<b>88 945</b>	<b>100.0</b>
Indigenous	5 083	93.6	316	0.4	5 399	6.1
Non-Indigenous	268	4.9	82 160	98.4	82 428	92.7
Not stated	77	1.4	1 041	1.2	1 118	1.3

(a) This table excludes 1,104 not matched Indigenous persons and 7,358 not matched non-Indigenous persons.

## TECHNICAL NOTE 2 DIFFERENCES IN CLASSIFICATION *continued*

### DIFFERENCES IN CLASSIFICATION BETWEEN THE PES AND CENSUS *continued*

5 Similarly, for Country of birth, the majority of person matches between PES and the corresponding Census record show the same response, with 97.3% of matched people who reported as overseas born in the PES also counted as overseas born in the Census, and 97.9% of matched people who reported as Australian born in the PES, also counted as Australian born in the Census.

#### DIFFERENCE IN CLASSIFICATION(a), Country of birth—Responding PES persons—2011

Census response	PES RESPONSE					
	Australia		Overseas		Total	
	no.	%	no.	%	no.	%
<b>Matched</b>	<b>68 484</b>	<b>100.0</b>	<b>20 461</b>	<b>100.0</b>	<b>88 945</b>	<b>100.0</b>
Australia	67 039	97.9	179	0.9	67 218	75.6
Overseas	210	0.3	19 906	97.3	20 116	22.6
Not stated	1 235	1.8	376	1.8	1 611	1.8

(a) This table excludes 5,737 not matched persons who responded in the PES as Australian-born and 2,725 not matched persons who responded in the PES as Overseas-born.

6 The PES used an estimation technique to adjust the weights of responding persons according to their PES-reported categories. This ensured that PES persons not responding in the Census received the same weight adjustments as similar responding persons. The Census categories were then used in producing sample-based estimates of the number of persons counted in the Census. The weighting ensured these estimates matched the actual Census counts for all benchmark categories (for information on benchmarking, see *Estimation*, section 5).

## TECHNICAL NOTE 3 PEOPLE AND DWELLINGS MISSED

### PEOPLE AND DWELLINGS MISSED IN THE CENSUS

#### *People missed in the Census*

**1** Given the purpose of the PES is to check the coverage of the Census and provide estimates of how many people were missed (or counted more than once), the details of people and dwellings missed in the Census are necessary in order to obtain estimates of gross undercount, which then feed into net undercount estimates. This section presents a more detailed analysis of the data available on the people and dwellings missed in the Census.

**2** In the PES, people are asked a number of questions about where they usually live, and addresses where they might have been counted in the Census, including:

- 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.

**3** For processing purposes, it was recorded whether the Census night address was the 'PES enumeration address' or a 'search address', that is, a different address supplied during the PES interview.

**4** The Census form corresponding to the PES enumeration address was checked to determine if the person was counted in the Census at that address. Any search addresses were then checked to determine if the person was counted elsewhere in the Census. A key difference in the quality of this information is that exact address information was available for PES enumeration addresses, while search addresses relied on details provided by respondents. In 2011, Automated Data Linking (ADL) also identified, for the first time in PES processing, search addresses that had not been volunteered by respondents using probabilistic linking. For further information about ADL see *Linking and Matching* (section 5).

**5** Based on results from PES linking and matching processing, there were a total of:

- 5,241 people (5.4% of the total number of people responding to the PES) who should have been counted in the Census but were missed (i.e. undercounted people); and
- 1,932 people who were PES respondents not counted in the Census and should *not* have been. This included diplomats, people who were overseas on Census night, and babies born after Census night.

**6** The table below shows whether PES respondents were counted in the Census and whether they thought they were included on a Census form. Data have been cross-classified by characteristics of their Census night address. The data in this table are unweighted and hence do not correspond to weighted estimates included in other parts of this publication.

## TECHNICAL NOTE 3 PEOPLE AND DWELLINGS MISSED *continued*

People missed in the Census  
*continued*

PES RESPONDING PERSONS, Whether counted in the Census and Whether thought they were included on a Census form(a)

	WHETHER THOUGHT THEY WERE INCLUDED ON A CENSUS FORM					Total(c)
	Yes - counted at PES address	Yes - counted at UR address	Yes - counted at other address	Don't know if counted in Census	No - not counted in Census(b)	
<i>Census night address</i>	no.	no.	no.	no.	no.	no.
<b>Counted in Census(d)</b>						
PES address	81 495	32	84	653	518	82 782
Usual residence address	14	1 266	2	176	15	1 473
Other address	525	8	4 031	309	222	5 095
Overseas	0	0	0	0	1	1
<b>Total(e)</b>	<b>82 035</b>	<b>1 306</b>	<b>4 117</b>	<b>1 140</b>	<b>756</b>	<b>89 356</b>
<b>Missed in the Census</b>						
PES address	2 087	6	27	330	1 095	3 545
Usual residence address	2	112	0	89	26	229
Other address	190	1	474	255	531	1 451
Overseas	2	1	0	2	4	9
<b>Total(e)</b>	<b>2 281</b>	<b>120</b>	<b>505</b>	<b>678</b>	<b>1 656</b>	<b>5 241</b>
<b>Out of PES scope coverage</b>						
All addresses	535	0	62	130	1 205	1 932
<b>All persons(f)</b>						
PES address	83 651	38	111	990	1 624	86 414
Usual residence address	16	1 378	2	265	41	1 702
Other address	716	9	4 535	578	767	6 605
Overseas	467	1	32	111	1 185	1 796
<b>Total(e)</b>	<b>84 851</b>	<b>1 426</b>	<b>4 684</b>	<b>1 948</b>	<b>3 617</b>	<b>97 047</b>

(a) As reported in the PES.

(b) Includes people who did not know whether they were included on a Census form.

(c) Includes people who did not report (or were not asked) whether they thought they had been included on a Census form.

(d) Includes people who should not have been counted in the Census.

(e) Includes people who did not provide a Census night address.

(f) Includes people who were not counted and should not have been counted in the Census.

**7** The majority of people were expected to have been counted in the PES at the dwelling where they stayed on Census night. The Census night address matched the PES enumeration address for 3,545 persons (around 68%) of *missed* people, compared to 82,782 persons (around 93%) of *all* people counted in the Census.

**8** Almost a third (1,656 persons) of *missed* people were believed to have not been included on a Census form, with a further 678 persons (around 13%) not sure if they had been included on a form. The remaining missed persons were believed to have been included on a Census form, with most thought to have been counted at their PES enumeration address.

**9** Of the people who were *counted* in the Census (89,356 persons), less than one percent (756 persons) were thought to have not been counted in the Census, with a further 1.3% (1,140 persons) not sure if they had been included.

## TECHNICAL NOTE 3 PEOPLE AND DWELLINGS MISSED *continued*

*Dwellings not counted in the Census*

**10** For a variety of reasons, not all dwellings received a Census form. This may have been because the dwellings were not habitable, the Census collector did not identify them, or in the case of a diplomatic dwelling, did not contain people within scope of the Census. Census field procedures required that all dwellings, apart from the types listed below, be recorded, irrespective of a received (or returned) Census form. Hence all dwellings in Australia should have been counted in the Census, with only the following exclusions:

- Dwellings under construction;
- Derelict dwellings;
- Vacant tents;
- Unoccupied converted garages;
- Unoccupied caravans; and
- Unoccupied dwellings within marinas/manufactured home estates.

**11** In 2011, the ability of the PES to match PES non-responding dwellings and sample loss dwellings to their corresponding Census dwellings was reduced. This was due to the introduction of ADL which changed the linking and matching processing methodology to focus on persons initially, then dwellings. Instead of dwelling matching being a necessary initial step in linking and matching, most dwellings were matched through links established for persons in ADL. This therefore altered the systematic nature of dwelling matching that took place in the previous PES methodology. For more information on ADL see *Linking and matching* (section 5).

**12** Estimates of dwelling undercount have not been provided in this publication due to some concerns over the representativeness of PES non-responding and sample loss dwellings (which show a higher than expected proportion of missed dwellings, and a lower than expected proportion of unoccupied dwellings). Dwelling matching outcomes for the dwellings responding in the PES are included below.

### DWELLINGS, PES response by Census response type—2011

	<i>Occupied</i>	<i>Un-occupied</i>	<i>Late returns</i>	<i>System created records</i>	<i>Missing</i>	<i>ICF not missing</i>	<i>ICF missing</i>	<i>Total</i>
	no.	no.	no.	no.	no.	no.	no.	no.
Responding	32 228	1 504	377	656	549	551	41	35 906

## TECHNICAL NOTE 4 IDENTIFYING CENSUS LATE RETURNS

### LATE CENSUS FORMS

**1** For some people who have not returned a Census form, contact from the ABS following selection in the PES acts as a reminder to return a completed Census form. These late returns, if not identified, would have resulted in the PES sample having a higher proportion of Census response than in the overall population. To protect against this, and maintain the statistical independence of the PES from the Census, all Census forms received after the start of 2011 PES field work were deemed 'late' and treated differently in PES estimation.

**2** For the purpose of PES estimation, the dwellings from which 'late' Census forms were received were treated as though they had not been contacted in the Census and classified to the Census 'non-contact sector'.

### CENSUS NON-CONTACT SECTOR

**3** As detailed in *Components of net undercount* (section 4), the non-contact sector also contains dwellings which were non-responding in the Census. During Census processing, a 'hot-deck' imputation method (see *Glossary*) imputed people into these dwellings, together with imputed values for Age, Sex, Marital status and Place of usual residence. These values were, in many cases, based on information provided by the Census collector about the dwelling and its residents. Values for all other variables (e.g. Indigenous status, Country of birth) were set to 'not-stated' or 'not applicable', depending on the imputed value for age.

**4** Inevitably, the imputed values differ from the true but unknown values, at both the dwelling and aggregate level. Imputed records made up the majority of the 2011 Census non-contact sector records, with late returns (as defined above) making up only a small component of this 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.

### ESTIMATING THE NUMBER OF PEOPLE IN CENSUS LATE RETURN AND IMPUTED DWELLINGS

**5** As was began in 2006, the 2011 PES processed the data returned from Linking and Matching, assigning a weight to each record, including those that corresponded to the Census non-contact sector. Hence the PES provides an estimate of the total number of people who should have been counted in the Census non-contact sector (i.e. late return and non-response dwellings).

**6** PES estimates of the population in the non-contact sector have relatively high sampling errors because of the small sample size (there were relatively few Census non-contact dwellings selected by chance in the PES sample); and also because Census person counts for this sector were 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 were 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.

**7** Using PES estimates for the non-contact sector population leads to a rise in the standard error of the overall population estimates, representing the uncertainty associated with estimating the contribution of the non-contact sector. On the other hand, the alternative, as was done prior to 2006, 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.

### ESTIMATING THE NUMBER OF PEOPLE IN CENSUS LATE RETURN AND IMPUTED DWELLINGS *continued*

**8** The standard error of the PES estimate of the population in the non-contact sector has been 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 2011 Census incorporate the PES estimate of the number of people who *should* have been counted in the non-contact sector.

### THE IMPORTANCE OF CORRECTING FOR CENSUS IMPUTATION ERROR

**9** The majority of Census non-contact sector records are imputed records, with the remainder made up of late returns. The Census count of late returns is considered an accurate measure, given the special procedures in place for identifying 2011 late returns. By subtracting the Census count of late returns from the PES estimate of people who *should* have been counted in the non-contact sector, a PES estimate (and corresponding standard error) of the number of people who *should* have been imputed in the Census can be calculated. The *Census imputation error* is the difference between this PES estimate and the actual Census count of imputed people.

**10** It is important to understand the level of accuracy associated with Census imputation in order to properly understand the impact it has had on PES estimates of net undercount. Consider the equation below.

**11** The PES estimate of the number of people who *should* have been counted in the Census (*A*) less those who were counted in the Census, including imputed persons (*B*), will give the net undercount (*C*) i.e.:  $A - B = C$ .

**12** Census imputation can impact on net undercount estimates in two ways:

- If the Census count (*B*) increases due to a *higher* number of imputed persons but all else remains the same, net undercount (*C*) will decrease.
- If the Census count (*B*) decreases due to a *lower* number of imputed persons but all else remains the same, net undercount (*C*) will increase.

**13** Given the direct impact that imputation can have on estimates of net undercount, it is important that PES corrects for Census imputation error in the non-contact sector and reports this number. This can then be used to better understand estimates of net undercount.

## TECHNICAL NOTE 4 IDENTIFYING CENSUS LATE RETURNS *continued*

### 2011 CENSUS LATE RETURNS AND IMPUTED RECORDS

**14** In the 2011 Census, there were 878,691 persons in the non-contact sector, who comprised 4.1% of the final 2011 Census count. The majority of the non-contact sector (around 87%) is made up of imputed person records. The following table provides Census late return and imputed person records by state/territory of usual residence, for 2006 and 2011.

### CENSUS LATE RETURNS AND IMPUTED DWELLINGS, State/territory - Person records(a)

	2006			2011		
	Census count	Non-contact sector(b)	Non-contact sector	Census count	Non-contact sector(b)	Non-contact sector
	no.	no.	%	no.	no.	%
New South Wales	6 549 177	309 261	4.7	6 917 654	286 575	4.1
Victoria	4 932 422	203 639	4.1	5 354 043	195 002	3.6
Queensland	3 904 534	172 384	4.4	4 332 733	189 517	4.4
South Australia	1 514 337	47 958	3.2	1 596 567	48 545	3.0
Western Australia	1 959 087	104 431	5.3	2 239 193	115 093	5.1
Tasmania	476 479	14 476	3.0	495 355	13 625	2.8
Northern Territory	192 900	15 778	8.2	211 959	17 131	8.1
Australian Capital Territory	324 037	12 843	4.0	357 217	13 203	3.7
<b>Australia</b>	<b>19 852 973</b>	<b>880 770</b>	<b>4.4</b>	<b>21 504 721</b>	<b>878 691</b>	<b>4.1</b>

(a) Data refer to state/territory of usual residence.

(b) Data include people from late-return dwellings and imputed persons.

## TECHNICAL NOTE 5 STATISTICAL IMPACT OF ADL

### OVERVIEW

**1** The introduction of Automated Data Linking (ADL) was arguably the most significant change introduced in the 2011 PES, a key innovation that improved the PES linking and matching methodology which is so fundamental to obtaining a net undercount measure. For further information on ADL see *Linking and Matching* (section 5).

**2** Based on the outcome of a feasibility study undertaken after the 2006 PES, the ABS expected the new methodology to deliver a significantly better linking and matching outcome than the methodology employed in previous years, and consequently a better, and significantly lower, estimate of net undercount. The main implication of a significant improvement in the estimate of net undercount was that there would be an increase in the discrepancy between 2006 and 2011 based population estimates, directly related to this change.

**3** Therefore, the decision to conduct a Statistical Impact Study arose primarily out of the need to explain the extent to which intercensal discrepancy (i.e. the change between 2006 and 2011 based population estimates) could be attributed to a change made in the PES methodology. That is, the study aimed to assess the extent to which the discrepancy was explained by a change in how Census coverage was being measured as opposed to Census coverage itself.

**4** Since the introduction of ADL represented a change to the processing of statistical inputs, rather than the inputs themselves, it was possible to estimate the impact of the new methodology through processing a sample of 2011 records using a close approximation of the methodology that was used in the 2006 PES. The matching outcomes of this process could then be compared with the outcomes achieved through ADL.

**5** A random sample of 2,158 dwellings was selected, stratified by state, containing around 5,700 persons or approximately 6% of the total 2011 responding PES sample, for the study.

### MATCHING METHODOLOGY

**6** The Statistical Impact Study records were separately processed using a methodology that was a close approximation of that used in the 2006 PES. The matching methodology used in 2006 is outlined in Appendix 2 of the 2006 PES publication: *Census of Population and Housing - Undercount, 2006* (cat. no. 2940.0).

### ESTIMATING THE IMPACT OF ADL

**7** Matching outcomes for the Statistical Impact Study sample were compared with the outcomes from the ADL-enabled process, to determine where the two methodologies differed. In order to effectively estimate the impact, some imputation was required. ADL-enabled matches that were not made using the 2006 processing were classified into six imputation groups, based on:

- whether the match was at the place of enumeration, a respondent-provided search address or another address; and
- the quality of the link.

**8** The Statistical Impact Study sample provided the average number of matches for each ADL-enabled match within these imputation groups, as well as an average number of Statistical Impact Study matches for the small number of persons unmatched in ADL-enabled processing. These averages were then applied to the rest of the PES sample, together with the outcomes of some vague address modelling (which was a feature of 2006 processing made redundant by ADL).

**9** Once the imputation work was complete, the file was then run through the same person weighting and estimation processes used to produce the published 2011 PES estimates. The format of the estimates output was therefore directly comparable to the main estimates produced via the ADL-enabled processing method.

IMPACT

**10** The ADL Statistical Impact Study estimated that the use of ADL in 2011 PES linking and matching resulted in a net undercount that was 246,985 persons *less* than the 2006 PES matching methodology would have delivered.

**11** The Statistical Impact Study result has a standard error of 43,000. A common approach to assessing the variability inherent in estimates is to examine the 95% confidence interval (which is two standard errors either side of the estimate). Using this approach, there is a 95% chance that the true value of the statistical impact of ADL on net undercount in 2011 is between 160,985 and 332,985 persons.

**12** It is important to remember that the Statistical Impact Study estimate was not designed to provide an alternative measure of net undercount for 2011, in 2006 terms, but only to identify the impact of the ADL methodology. There are a range of PES and Census changes that are not related to ADL that will affect comparability between 2006 and 2011. For instance, the reduced level of Census imputation in 2011 will directly affect the comparability of net undercount measures with their 2006 equivalents.

**13** The ADL Statistical Impact Study reinforces the value in the ABS continuing to make innovative changes to the PES, especially in developing linking and matched methods that provide a more accurate estimate of the completeness of Census coverage.

## TECHNICAL NOTE 6 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 was 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 2.5% 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  $2.5\% \pm (1 \times 0.1\%)$  or 2.4% to 2.6%, and nineteen chances in twenty that the figure is in the range  $2.5\% \pm (2 \times 0.1\%)$ , that is, between 2.3% and 2.7%.

**5** For ease of use, the SEs corresponding to the net undercount rates are given next to the estimates 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 greater capital city versus rest of state regions, or for differences between estimates from different Censuses. However, for estimates within the same region there will be a negative correlation between the rates so that the approximation will generally underestimate the true SE.

**8** For example, if the estimates of the rate of net undercount for usual residents in QLD greater capital city and rest of state are 2.7% and 1.4%, with SEs of 0.35 and 0.5 percentage points respectively, using the formula above the SE on the difference (1.3 percentage points) is:

$$SE(x-y) = \sqrt{(0.35)^2 + (0.5)^2} = 0.61\%$$

**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  $1.3 \pm (2 \times 0.61)$  or 0.08 to 2.52 percentage points.

### NON SAMPLING ERROR

**10** The estimates of undercount are also subject to non-sampling errors which occur in all collections, whether censuses or surveys. Examples of this kind of error include imperfections in reporting by respondents and errors made in collection and processing of data. Every effort is made in both the Census and PES to minimise non-sampling error by careful design of forms, training and supervision of collectors and interviewers, and by using effective operating procedures. Types of non-sampling error arising from the way the PES is conducted and the way estimates are derived from the survey are discussed below.

NON SAMPLING ERROR  
*continued*

**11** A potential weakness in the PES method is its dependence on linking and matching as a means of deciding whether or not a given person or dwelling in the PES has been counted in the Census. The difficulties associated with the linking and matching process mean that there is a risk of failing to match people who were 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. However, the introduction of ADL in the 2011 PES processing phase has helped to reduce the likelihood of this type of error occurring.

**12** 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 be missed by a PES interviewer as well. In addition, people who avoid being included in the Census may also avoid being included in the PES. The use of benchmarks in estimation helps to control for the effect of this 'correlation bias'.

## GLOSSARY

<b>AddressCoder@ABS</b>	A web service that assigns a geographic classification to an address or a list of addresses and is used for linking and matching activities.
<b>Automated Data Linking (ADL)</b>	Automated linking processes used to determine possible links between Census and PES data, before any clerical matching process has begun. It employs a probabilistic linking method that uses a range of personal and address characteristics to evaluate the likelihood that a PES and Census record pertain to the same individual.
<b>ASGC</b>	The Australian Standard Geographical Classification (ASGS) was the geography standard in use during the 2006 Census. It was built using the Census Collection District (CD), which was then used to build up larger Statistical Local Areas (SLA), which in turn built larger areas. The ASGC is in the process of being replaced by the newer Australian Statistical Geography Standard (ASGS). In the 2011 PES, the ASGC was used in conjunction with the ASGS, principally in relation to the 2011 PES sample design, which was based upon the CD.
<b>ASGS</b>	The Australian Statistical Geography Standard (ASGS) is the new geographical standard developed by the ABS for the collection and dissemination of geographic statistics. It is a hierarchically structured classification with a number of spatial units to satisfy different statistical purposes.
<b>Benchmark category</b>	A benchmark category is a PES estimate (either dwelling or person based) of the Census count in a category (e.g. Country of Birth) based upon the Census response to that category regardless of the PES response (i.e. if a record has a Census COB of 'Australia' but is recorded as having a COB of 'Other Overseas' in the PES, then the record will be in the 'Australia' benchmark category).
<b>Census count</b>	The number of people who were counted in the Census for a category. The Census count includes imputed persons for non-responding dwellings. However, some categories (such as Indigenous status and Country of birth) do not include any imputed persons, as Census assigns a not-stated category to all the imputed person records.
<b>Census Collection District (CD)</b>	A Census collection district (CD) was the basic geographic unit of collection in the 2006 Census, but was also used in several aspects of the 2011 PES. A CD was generally a Census workload area that one collector could cover, delivering and collecting forms in a specified period.
<b>Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS)</b>	A series of processes which prioritised and organised ADL output for use by the MSS system.
<b>Collector Workloads (CLWs)</b>	A Collector Workload (CLW) is a geographic grouping of 450 dwellings (on average) which define an area in which, generally, a single Collector operated in the 2011 Census.
<b>Computer Assisted Interviewing (CAI)</b>	Computer-assisted interviewing (CAI) is a method of data collection whereby responses are recorded directly into an electronic questionnaire on a notebook computer.
<b>Contact sector</b>	The Census contact sector comprises all Census dwellings, excluding imputed dwellings and those from which late return Census forms were received.
<b>Correlation bias</b>	A bias arising when people who were not counted in the Census are more likely to be missed in the PES than people with similar values of the characteristics used in PES estimation (such as age, sex and Indigenous status) who were counted in the Census.
<b>Coverage</b>	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.

## GLOSSARY *continued*

<b>Discrete Indigenous community</b>	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.
<b>Dwelling</b>	A dwelling is a building or structure in which people may 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 unoccupied converted garages, are not counted as dwellings in the Census. There are private and non-private dwellings.
<b>Dwelling Link</b>	A link between a PES dwelling and a Census dwelling based upon at least one linked person between the two.
<b>Dwelling Link Rating</b>	The Dwelling Link Rating is a numeric indicator of the strength of a dwelling link and used to rank the links for preference in later processing. These are derived primarily by combining the Person Link Rating of individuals in the dwelling with the addition of some other criteria.
<b>Estimated Resident Population (ERP)</b>	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.
<b>Freely Extensible Biomedical Record Linking (FEBRL)</b>	This is the software application that was used to run the Automated Data Linking process in the 2011 PES.
<b>Greater Capital city/rest of state/territory</b>	Greater Capital Cities are defined as areas covered by the Greater Capital City Statistical Area (GCCSA) in each state and territory. GCCSAs are aggregates of Statistical Area 4s (SA4). The GCCSAs combined with the Rest of State regions cover the whole of Australia and aggregate directly to States and Territories.  For more information, see <i>Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas</i> (cat: 1270.0.55.001), Australia, July 2011.
<b>Gross overcount</b>	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).
<b>Gross undercount</b>	The number of people who should have been counted in the Census but were not.
<b>Hot-deck imputation</b>	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.
<b>Indigenous</b>	Refers to people who identified themselves, or were identified by another household member, as being of Aboriginal and/or Torres Strait Islander origin.
<b>Indigenous Community Frame (ICF)</b>	The Indigenous Community Frame (ICF) is a listing of all discrete Indigenous communities from which the sample of communities was selected for the 2011 PES. The selection unit on the ICF is a community set. A set usually comprises a main community, a number of outstations 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.

## GLOSSARY *continued*

<b>Imputation</b>	A statistical process for predicting values where no response was provided to a question and a response could not be derived.
<b>Imputed dwelling (in Census)</b>	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).
<b>Late return</b>	A Census form which was returned after the start of PES enumeration.
<b>Mainstream Sample</b>	The term 'mainstream' refers to the PES private dwelling sample, that is, all PES selections other than those from the discrete Indigenous communities sample.
<b>Match and Search System (MSS)</b>	The main PES clerical review facility, which allows processors to search, view, compare, and record matches between PES and Census data.
<b>Net undercount</b>	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 imputations). The estimated net undercount for a category of person is the net result of undercount, overcount, differences in classification between the PES and Census (e.g. age, sex, Indigenous status) and imputation error in the Census.
<b>Non-contact sector</b>	The Census non-contact sector comprises late-return and imputed dwellings.
<b>Non-private dwelling</b>	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.
<b>Non-sampling error</b>	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 people selected in the survey.
<b>Other territories</b>	The 'Other territories' comprise Jervis Bay Territory and the external territories of Christmas Island and Cocos (Keeling) Islands.
<b>Outstation (or homeland)</b>	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.
<b>Part of state</b>	Parts of state refer to the Greater Capital City or Rest of State in each state and territory. In a number of processes, such as estimation, the different parts of state are dealt with separately.
<b>Person link</b>	A link between a single PES respondent and a single Census respondent based upon agreement on certain criteria.
<b>Person Link Rating</b>	The Person Link Rating (PLR) is a numeric indicator of the strength of a person link. This is generated through the CARDS process and used to stratify ADL links by quality. It is used through matching processing and contributes to the Dwelling Link Rating (DLR) of the PES dwelling in which the person was enumerated.
<b>PES address</b>	The address at which the respondent was enumerated in the PES.
<b>PES population estimate</b>	An estimate (based on PES and Census data) of the number of people who should have been counted in the Census.
<b>Private dwelling</b>	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.

## GLOSSARY *continued*

<b>Remote areas</b>	<p>Within the Australian Standard Geographical Classification (ASGC), the Remoteness Structure classification comprises six categories called Remoteness Areas (RAs). Each RA is created from the grouping of Collection Districts (CDs) identifying a (non-contiguous) region in Australia having a particular degree of remoteness. The categories range from 'Major Cities of Australia' to 'Very Remote Australia'.</p> <p>The degree of remoteness of each Collection District (CD) was determined using the Accessibility/Remoteness Index of Australia (ARIA+), which was recalculated using 2006 Census statistics. It is envisaged that ARIA+ will be recalculated after the 2011 Census and the results will be used to construct the 2011 Australian Statistical Geography Standard (ASGS) Remoteness Structure based on aggregations of Statistical Areas Level 1 (SA1s).</p>
<b>Statistical Areas 1, 2, 3 &amp; 4</b>	<p>Statistical areas are the geographic areas under the ASGS geographic classification used in the PES. Each area is built from aggregates of the lower area and cover Australia without gaps or overlap.</p> <ul style="list-style-type: none"><li>■ Statistical Area Level 1 (SA1s) are the smallest unit at which census data will be released and are built from aggregated meshblocks. They generally comprise 200 to 800 people in urban and regional areas (with an average of 400 persons in urban areas and fewer in regional areas), or a single Discrete Indigenous Community (and sometimes some/all of its outstations).</li><li>■ SA2s are general-purpose medium-sized areas designed to represent a community that interacts together socially and economically. They comprise between 3000 and 25,000 persons.</li><li>■ SA3s are areas which have similar regional characteristics. They generally comprise between 30,000 and 130,000 persons, however, as they are regionally rather than statistically based, these numbers are flexible.</li><li>■ SA4s are the largest sub-State regions in the Main Structure of the ASGS. They comprise a minimum of 100,000 persons (generally 1-3 times this size in rural areas and 3-5 times this size in urban areas).</li></ul> <p>For more information, see <i>Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas</i> (cat: 1270.0.55.001), Australia, July 2011.</p>
<b>Sampling error</b>	<p>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.</p>
<b>Search address</b>	<p>An address (other than the PES address) where a person was reported to be staying on Census night or where a person may have been included on a Census form (including non-reported addresses found by ADL). PES processing uses these addresses to locate a Census, in order to determine the number of times (if any) a person enumerated in the PES was included on a Census form.</p>
<b>Scope</b>	<p>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.</p>
<b>Standard error</b>	<p>A measure of the likely difference between the true value and the estimate.</p>
<b>Undercount adjustment factor</b>	<p>The undercount adjustment factor is the ratio of the PES population estimate to the Census count. This factor can be applied to the Census counts to indicate how many people should have been counted in the Census for that category.</p>
<b>Usual residence</b>	<p>The usual residence for a 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 2011.</p>





## FOR MORE INFORMATION . . .

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

### INFORMATION AND REFERRAL SERVICE

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

*PHONE*                      1300 135 070

*EMAIL*                      [client.services@abs.gov.au](mailto:client.services@abs.gov.au)

*FAX*                              1300 135 211

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

## FREE ACCESS TO STATISTICS

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

*WEB ADDRESS*      [www.abs.gov.au](http://www.abs.gov.au)