Survey Methodology, December 1988
Vol. 14, No. 2, pp. 173–189
Statistics Canada

Adjusting the 1986 Australian Census Count for Under-Enumeration

C.Y. CHOI, D.G. STEEL and T.J. SKINNER[1]

ABSTRACT

In Australia, population estimates have been obtained from census counts, incorporating an adjustment for under-enumeration in 1976, 1981 and 1986. The adjustments are based on the results of a Post Enumeration Survey and demographic analysis. This paper describes the methods used and the results obtained in adjusting the 1986 census. The formal use of sex ratios as suggested by Wolter (1986) is examined as a possible improvement of the less formal use made of these ratios in adjusting census counts.

KEY WORDS: Census under-enumeration; Post-enumeration survey; Demographic estimates; Sex-ratios.

1. INTRODUCTION

The population census provides the basic information from which estimates are made of the population of the nation, each of the eight States and sub-State local government areas. In Australia, these population estimates are required for the determination of the number of seats each State will have in the Federal House of Representatives, the allocation of funds to each State, and the funding of local government authorities. Population estimates are also used in their own right as indicators of population growth and distribution and as denominators for various demographic, social and economic indicators. Because population estimates are used in such important ways, a high level of accuracy is required.

In Australia, it is known that the level of under-enumeration at the census is significant and that this level is related to important variables such as birthplace, geographic area and age/sex. Because of this, an adjustment for under-enumeration is made to census counts used for population estimates.

The adjustment of census counts for under-enumeration is a recent practice in Australia. Prior to the 1976 Census, census counts without adjustment for under-enumeration were used directly for population estimation purposes. The need to make this adjustment was recognised when the 1976 Census count fell considerably below the population estimates for the 1976 Census date which were updated from the 1971 Census, and when the 1976 Post Enumeration Survey (PES) showed a high under-enumeration rate of 2.6 per cent compared with 0.5 per cent in 1966 and 1.3 per cent in 1971. The 1976 PES also showed significant variations in under-enumeration between States and Territories, ranging from 4.2 per cent for the Northern Territory to 1.1 per cent for Tasmania. In 1986, the level of under-enumeration is estimated to be 1.9 per cent. As in 1976, there were significant variations between States and Territories. The adjustment of 1976 and subsequent census counts has been well received and no challenges have been raised to the appropriateness of doing so or the accuracy of the methods used. This is in contrast with the high level of controversy experienced in the United States of America on the appropriateness of making adjustments to the 1980 census counts for under-enumeration.

[1] C.Y. Choi, D.G. Steel and T.J. Skinner, Australian Bureau of Statistics, P.O. Box 10, Belconnen, ACT, 2616, Australia.

Data for the assessment of the level of under-enumeration are primarily derived from a census PES. Results of the PES are assessed by comparing these with estimates based on demographic statistics and other independent data such as statistics on school enrolments, on children whose parents receive government family allowances, and on persons registered with the government Medicare insurance system. In Australia, school enrolments for children aged 6-15 years are compulsory and until means-testing was introduced in November 1987, family allowances had been universally paid to mothers of all children of ages less than 17. Medicare insurance is also compulsory and universal for all residents. These independent statistics are therefore helpful as a check of the PES results and demographic estimates.

Although population estimates include an adjustment for under-enumeration, no adjustment is made for other census data. Census counts are published without adjustment.

2. THE 1986 POST-ENUMERATION SURVEY

In its five yearly population census, the Australian Bureau of Statistics (ABS) employs census collectors for the delivery of forms to each household and for the collection of completed forms from each household. The census is conducted on the basis of enumerating people where they are located on census night.

This collector-based field system allows the census collection phase to be completed two weeks after the census date. This allows a census PES to be conducted reasonably close to the census date – in 1986 within 4-5 weeks of census night. Because the PES asks a number of questions requiring detailed answers referring to a person's location on census night, its conduct close to census date minimises recall error and also reduces the number of exclusions due to deaths and overseas travel.

As the PES provides the basis for adjusting the census counts for under-enumeration, it is important that the PES be statistically independent of the census. The Appendix describes the steps taken to ensure independence.

The basic approach adopted in the 1986 PES was to select a sample of people independently of the census through a multi-stage area sample of private dwellings. The information required of each person in the selected households was obtained by personal interview of any responsible adult by trained field staff from the ABS regular interview panel. Matching of PES and census records to determine whether each person in the sample should have been included in the census and how many times the person was in fact included was undertaken by clerical staff employed in the Census Data Transcription Centre. The procedures used are described in the Appendix.

From the survey, the ratio of the number of persons who should have been included in the census (x) to the number of persons who were estimated to have been in fact included (y) can be estimated. This ratio is the net adjustment factor which accounts for both over and under-enumeration of individuals.

This adjustment factor, after weighting, is then applied to the actual census count (Y) to produce an estimate of the population (X), *i.e.* $X = Y\ (x/y)$.

To allow for differences in expected and actual sample take in the PES, this procedure was applied at the age (5 year groups), sex and geographic area (capital city statistical division/rest of State) level. PES estimates are produced on both an actual location at the census date and usual residence basis. The estimation also includes an adjustment for the small level of non-contact and non-response in the PES. For example the estimate of usual residence population for geographic area (s) and age sex cell (a) is:

$$X_{sa} = Y_{sa}\, x_{sa}/y_{sa}$$

where

$$x_{sa} = \sum_{gc} \frac{D_{gc} + d_{gc}}{D_{gc}} \cdot \frac{x_{sagc}}{f_g}$$

and

$$y_{sa} = \sum_{gc} \frac{D_{gc} + d_{gc}}{D_{gc}} \cdot \frac{y_{sagc}}{f_g} .$$

In these estimation formulae the subscript c denotes the response status of the PES dwelling in the census and the subscript g denotes the geographic area in which the person was selected in the PES. D_{gc} is the number of responding dwellings and d_{gc} is the number of non-contact/non-responding dwellings in area g and census response category c. The sampling fraction varies between states and is denoted f_g.

In this form the estimator is a post-stratified ratio estimate. Ignoring for the moment that people may be enumerated in the census incorrectly or more than once, the estimator is the estimator obtained from a dual-record system or a capture-recapture approach discussed, for example, in Bishop, Fienberg and Holland (1975, pp231-234). This is shown in the diagram below where under the assumption of independence the estimate of the total population is Y (x/y) which is the ratio estimate X.

PES

		Counted	Missed
Census	Counted	y	Y
	Missed		
		x	

The 1986 PES, however, was *designed* to collect information on both the number of persons missed by the census and the number of persons over-enumerated, *i.e.* included in the census erroneously or included more than once. The estimate X takes into account both over and under-enumeration at the same time. In this respect, the approach adopted is different from the traditional capture-recapture methodology.

Variance estimation was based on treating X as a ratio estimate derived from a multi-stage sample. The relative standard errors on the PES estimates of the population are given in Table 1. From this table and tables 2 and 4 we see that standard errors are considerably less than the adjustments implied by the PES national age by sex estimates and State by sex estimates.

Table 1

1986 Census: Relative Standard Errors of PES Estimates of the Population

Age	Males	Females	Persons
	%	%	%
0- 4	0.29	0.36	0.24
5- 9	0.29	0.30	0.22
10-14	0.28	0.29	0.21
15-19	0.32	0.32	0.24
20-24	0.49	0.43	0.34
25-29	0.49	0.36	0.32
30-34	0.39	0.34	0.27
35-39	0.36	0.30	0.24
40-44	0.38	0.32	0.26
45-49	0.37	0.30	0.25
50-54	0.43	0.38	0.30
55-59	0.38	0.30	0.25
60-64	0.41	0.38	0.29
65-69	0.43	0.37	0.29
70-74	0.53	0.41	0.34
75+	0.47	0.39	0.31
All ages	0.12	0.10	0.08
State	**Males**	**Females**	**Persons**
	%	%	%
NSW	0.21	0.18	0.14
VIC	0.23	0.21	0.16
QLD	0.27	0.24	0.19
SA	0.27	0.20	0.17
WA	0.29	0.25	0.19
TAS	0.36	0.31	0.25
NT	1.65	1.53	1.22
ACT	0.61	0.74	0.55

For a more detailed description of the 1986 Post-Enumeration Survey and the estimation procedures, see Appendix.

3. DEMOGRAPHIC ESTIMATES OF CENSUS UNDER-ENUMERATION

An alternative method for the estimation of census under-enumeration is through the use of past demographic data including those from previous censuses, births and deaths registers, and overseas migration statistics. For example, estimates of the population at a certain date can be made by updating a previous census using data on births, deaths and overseas migration. The more distant is the previous census which serves as the base, the longer is the time series of reliable vital and migration statistics required, and the less reliance there needs to be on the accuracy of the census base. This is because estimates of persons born after the relevant census date will be affected only by the reliability of data on births, deaths and migration. Internal migration data in Australia are not sufficiently reliable to enable the use of demographic methods for estimating census under-enumeration at sub-national levels. Use of demographic estimates for census evaluation is therefore limited to Australian totals.

Australian data on births and deaths are available as a time series going back to the 19th century and it is unlikely that there have been significant omissions. Successive reports by the Australian Commonwealth Statistician after each population census from 1911 to 1961 claimed that the registration of births and deaths in Australia was substantially complete although it was recognised that some omissions were possible and that there were time lags in registrations. The Statistician's Report was discontinued after 1961. However, there is no evidence that the level of coverage of birth and death registrations has deteriorated since then.

Australia has also maintained comprehensive and reliable statistics on overseas arrivals and departures over a long period of time. These statistics cover all movements including permanent, long term and short-term movements. However, there are several deficiencies in the statistics on overseas arrivals and departures which limit their usefulness for the evaluation of the census data. First, there have been periods in the past when arrivals and departures were suspected of being inaccurately recorded (*e.g.* during World War II and the period immediately following the war). Second, because of the increase in overseas short-term movements since the 1960's only a sample (of about 1 in 20) of the arrivals and departure records has been processed for statistical purposes since 1971. Third, errors can occur in the classification of travellers into permanent, long-term and short-term categories. To avoid these errors of classifications the comparison of demographic estimates, census counts and PES estimates of the population at census date is made on the basis of actual location, which include all three categories of overseas movements.

For the assessment of under-enumeration at the 1986 Census, demographic estimates of the population as at census date 1986 by age and sex were made using births, deaths and overseas migration data going back to 1921 together with results of the 1921 Census. Demographic estimates of the population to age 65 years are therefore based solely on birth, deaths and migration data and would not be affected by the accuracy of the 1921 Census.

4. VALIDATION OF THE 1986 PES ESTIMATES

The following table shows the estimated population as at 30 June 1986 by age and sex based on demographic analysis and based on the 1986 PES. Medicare enrolments by age and sex are also shown.

There is a very high level of correspondence between PES and demographic estimates of the male population, particularly for those aged under 30. However, there is a large discrepancy for males aged 30-34, the demographic estimates being 20,000 higher than PES estimates. This can be attributed to a large net gain in the number of males of these ages from short-term movements into and out of Australia in the period 1981-86. Net gains from short-term movements of this magnitude are not detectable in the adjacent age-groups and therefore may reflect some error in overseas arrivals and departures statistics. With the volume of overseas movements being very high (over 6 million in 1986), a small error in reporting of age or in processing can lead to a relatively large discrepancy in the demographic estimate in net absolute term. The possibility of error in demographic estimates is further illustrated by the very high implied under-enumeration rate of 5.3 per cent for this age group compared with much lower rates for the surrounding age group.

It is, of course, quite likely that under-enumeration of overseas visitors was not adequately measured by the PES. However, in either case, errors in estimating the visitor component of the population should not affect the accuracy of official population estimates because these are based on the concept of usual residence and do not include visitors.

Table 2

Estimates of 1986 Population by Age and Sex Based on the 1986 PES and Demographic Analysis, and Medicare Enrolment

Age	Males ('000)									
	Population				Difference from Census			Percent Under-enumeration		
	Census (a)	PES(a)	DE(b)	Medicare	PES	DE	Medicare	PES	DE	Medicare
0- 4	608.3	616.4	612.8	611.4	8.0	4.5	3.1	1.3	0.7	0.5
0- 5	594.9	602.4	603.0	612.3	7.5	8.1	17.4	1.2	1.3	2.8
10-14	660.8	670.4	668.4	674.3	9.6	7.6	13.5	1.4	1.1	2.0
15-19	673.1	688.4	687.7	693.1	15.3	14.6	20.0	2.2	2.1	2.9
20-24	648.5	679.5	681.3	681.3	31.0	32.8	32.8	4.6	4.8	4.8
25-29	649.2	677.7	675.0	688.5	28.5	25.8	39.3	4.2	3.8	5.7
30-34	615.5	630.0	650.1	647.5	14.5	34.6	32.0	2.3	5.3	4.9
35-39	622.2	634.2	632.7	646.2	12.0	10.5	24.0	1.9	1.7	3.7
40-44	504.2	512.6	517.0	522.3	8.4	12.8	18.1	1.6	2.5	3.5
45-49	419.8	427.0	416.5	436.8	7.2	-3.3	17.0	1.7	-0.8	3.9
50-54	363.7	371.2	371.4	377.9	7.5	7.7	14.2	2.0	2.1	3.8
55-59	373.4	379.5	384.9	386.6	6.1	11.5	13.2	1.6	3.0	3.4
60-64	341.1	347.0	348.1	350.6	5.9	7.0	9.5	1.7	2.0	2.7
65-69	259.6	263.6	251.8	265.5	4.0	-7.8	5.9	1.5	-3.1	2.2
70-74	204.2	208.2	200.8	213.0	4.0	-3.4	8.8	1.9	-1.7	4.1
75+	229.5	233.0	181.5	250.1	3.5	-48.0	20.6	1.5	-26.4	8.2
Total	7768.3	7941.0	7883.1	8057.3	172.7	114.8	289.0	2.2	1.5	3.6

Age	Females ('000)									
	Population				Difference from Census			Percent Under-enumeration		
	Census (a)	PES(a)	DE(b)	Medicare	PES	DE	Medicare	PES	DE	Medicare
0- 4	579.7	591.0	583.8	580.9	11.3	4.1	1.2	1.9	0.7	0.2
5- 9	565.1	572.4	565.5	582.1	7.3	0.4	17.0	1.3	0.1	2.9
10-14	628.0	636.8	630.2	641.8	8.8	2.2	13.8	1.4	0.3	2.2
15-19	644.1	657.4	651.4	666.3	13.3	7.3	22.2	2.0	1.1	3.3
20-24	633.1	652.5	644.4	670.4	19.4	11.3	37.3	3.0	1.8	5.6
25-29	648.7	660.7	665.4	684.1	12.0	16.7	35.4	1.8	2.5	5.2
30-34	618.1	627.8	631.2	643.9	9.7	13.1	25.8	1.5	2.1	4.0
35-39	612.1	619.1	600.2	626.3	7.0	-11.9	14.2	1.1	-2.0	2.3
40-44	482.6	488.6	489.6	495.4	6.0	7.0	12.8	1.2	1.4	2.6
45-49	399.1	403.0	397.9	411.6	3.9	-1.2	12.5	1.0	-0.3	3.0
50-54	349.1	354.6	343.9	358.6	5.5	-5.2	9.5	1.6	-1.5	2.6
55-59	362.6	366.5	362.4	372.4	3.9	-0.2	9.8	1.1	-0.1	2.6
60-64	358.2	364.4	351.3	365.3	6.2	-6.9	7.1	1.7	-2.0	1.9
65-69	298.2	302.2	301.9	306.7	4.0	3.7	8.5	1.3	1.2	2.8
70-74	259.0	262.9	262.2	269.7	3.9	3.2	10.7	1.5	1.2	4.0
75+	396.2	404.7	385.0	434.1	8.5	-11.2	37.9	2.1	-2.9	8.7
Total	7833.8	7964.6	7866.2	8109.6	130.8	32.4	275.8	1.6	0.4	3.4

(a) Actual location basis.
(b) Demographic estimates based on 1921 Population Census and post 1921 demographic events.

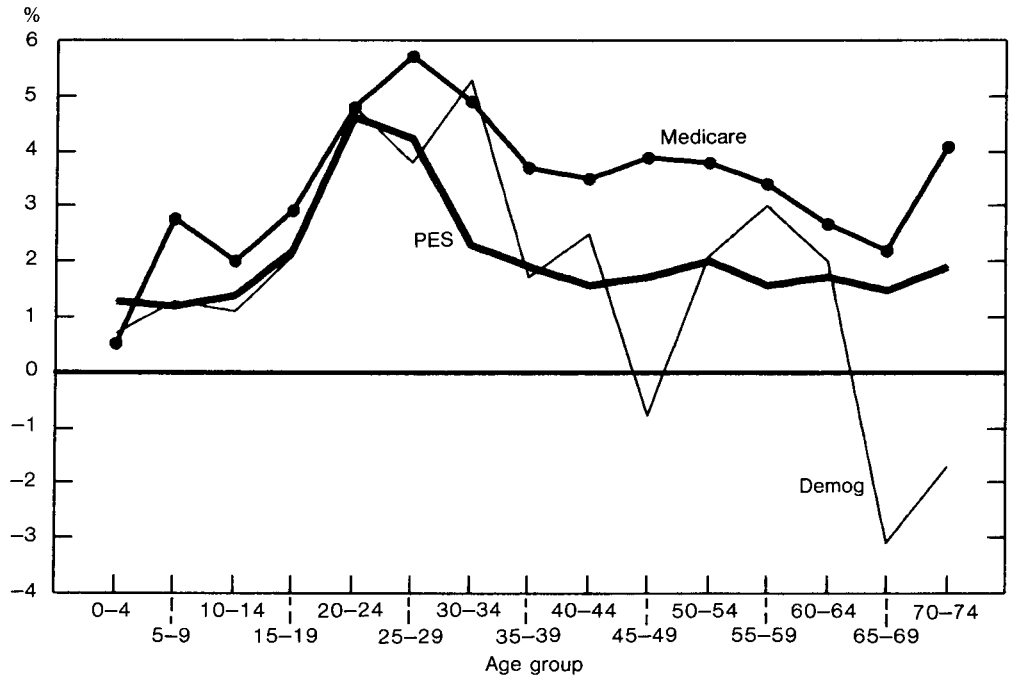

Figure 1. Percentage under-enumeration at the 1986 Census: Post-Enumeration Survey, Demographic Estimates and Medicare Enrolment-MALES.

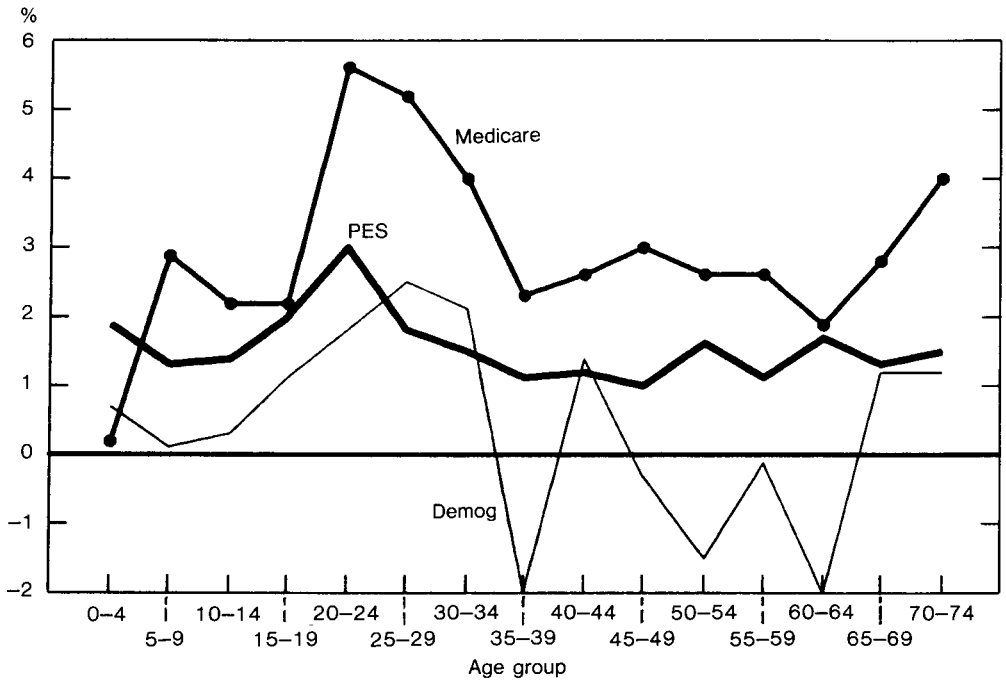

Figure 2. Percentage under-enumeration at the 1986 Census: Post-Enumeration Survey, Demographic Estimates and Medicare Enrolment-FEMALES.

For females, the level of correspondence between PES results and demographic estimates for ages below 35 is satisfactory. However, the demographic estimates for some age groups are considerably lower than PES estimates, and for those aged 35 to 39, and 45 to 64, they are lower than the unadjusted census count. Demographic estimates for these groups appear to be too low. This supports the view that demographic estimates are not sufficiently accurate for the production of population estimates and should be used only to assess PES results.

PES under-enumeration rates by age show a pattern which is smooth and much less erratic than that shown by demographic estimates. The higher PES rates for young adults aged 20-29 compared with those for other ages are as expected, given the higher rates of mobility among young adults, particularly males.

Medicare registrations are considerably higher than PES estimates and demographic estimates, except for the 0-4 age-group. Studies of registration practice show that the lower number in the 0-4 age group for medicare registration reflects the delays in births being registered with Medicare, and the higher numbers in other ages reflect delays in deleting from the Medicare register deaths and persons who have emigrated from the country.

Comparisons of PES estimates with estimates from family allowance registration and school enrolments for selected age-groups also show satisfactory correspondence. These results give some confirmation of the accuracy of the PES estimates in so far as the younger ages are concerned.

Although there is a satisfactory level of correspondence between PES estimates and other estimates of the population, there are two remaining problems which require consideration before the PES estimates can be accepted. The first emerges from an analysis of the PES estimates of census under-enumeration rates by age and sex. These rates are shown in Table 2.

Except for those aged 0-4 and 75 +, male under-enumeration rates are generally higher than female rates. While the rates for those aged 75 + could be affected by small sample size, the rate for females aged 0-4 appears too high, 1.9 per cent compared with 1.3 percent for males of the same age and for females aged 5-9. The number of females aged 0-4 estimated by the PES to have been under-enumerated was 11,300 compared with about 7,000 for the age group 5-9. This large difference in under-enumeration between those aged 0-4 and those aged 5-9 for females does not exist for males.

The PES sex ratio for persons aged 0-4 is 104.3 males to 100 female, lower than the census count ratio of 104.9 and the ratio of 105.0 males to 100 females estimated from demographic data.

On the above evidence, it appears that the PES has over-estimated females aged 0-4, although it is difficult to see how the PES could have over-estimated this group more so than other groups.

The second problem relates to the very high PES under-enumeration rate estimated for the Northern Territory. As shown in Table 4 it is 9.97% on an actual location basis and 6.45% on a usual residence basis. Northern Territory is a sparsely populated area (the census count in 1986 was 154,800 in an area of 1.3 million square kilometers) with a highly mobile population. The PES estimate of the population of Northern Territory is considerably higher than that based on the 1981 Census. Comparisons of PES estimates for the Northern Territory with independent estimates such as the number of children on the family allowance register and the number of school enrolments, also show that PES estimates are high. While these independent estimates may very well contain errors, it appears very likely that the PES has over-estimated the rate of under-enumeration for the NT.

The PES questionnaires were checked for the Northern Territory and were found to be satisfactory except for one collection district where problems with unreliable addresses and difficult terrain exposed inadequacies in field procedures and led to difficulties with matching.

Table 3

Comparison of 1986 PES Results with Independent Estimates

	PES Estimates	Demographic Estimates	Family Allowance	School Enrolment
		Persons ('000)		
0- 4	1207.3	1196.6	1204.8(a)	-
5- 9	1174.8	1168.5	1177.0	-(b)
10-14	1307.2	1298.6	1304.2	1289.6

(a) Family allowance registration for age 0 is understated because of the time lag in births being registered for family allowance. An adjustment was made by substituting the family allowance figure for age 0 by an estimate from the demographic analysis.
(b) School enrolment not compulsory for children aged 5 years.

Table 4

PES Under-Enumeration Rates (%) by State

	Actual location basis	Usual Residence basis
New South Wales	1.54	1.51
Victoria	1.59	1.77
Queensland	2.68	2.43
South Australia	1.54	1.59
Western Australia	2.32	2.26
Tasmania	1.32	1.16
Northern Territory	9.97	6.45
Aust. Capital Territory	1.95	1.61
Australia	1.91	1.84

A judgement was made that the PES over-estimation of females aged 0-4 and of the NT population should be corrected by adjusting the PES results. The adjustment to females aged 0-4 was made by using the sex ratio from demographic estimates and applying this to the PES estimates of males aged 0-4. Essentially, this amounted to replacing the PES estimate of females aged 0-4 by a better estimate using the PES estimate of males and the sex ratio. The result of this adjustment was to reduce the estimates of this group by 4,000 to 587,000.

The problem with the NT estimates was handled by not using data from the problematic collection district. This reduced the Northern Territory under-enumeration rate to 9.1 per cent (on an actual location basis) and 5.5 per cent (on a usual residence basis).

The two adjustments to PES results reduced the overall national under-enumeration rate from 1.91 per cent to 1.87 per cent (on an actual location basis), or from 1.84 per cent to 1.81 per cent (on a usual residence basis). Table 5 shows PES estimates by age and sex after the above adjustments were made to the estimates for NT and for females aged 0-4.

Table 5

Census Count 1986 Adjusted for Under-enumeration by Age and Sex

Age	On the basis of 'actual location'					
	Males		Females		Persons	
	No. ('000)	% under enumer-ation	No. ('000)	% under enumer-ation	No. ('000)	% under enume-ation
0- 4	616.3	1.30	586.6	1.17	1202.9	1.24
5- 9	602.4	1.24	572.4	1.27	1174.8	1.26
10-14	670.1	1.39	636.8	1.38	1306.9	1.39
15-19	688.3	2.19	657.3	2.02	1345.6	2.11
20-24	679.4	4.54	652.4	2.95	1331.8	3.76
25-29	677.5	4.17	660.7	1.81	1338.2	3.00
30-34	629.9	2.29	627.8	1.55	1257.7	1.92
35-39	634.0	1.87	618.9	1.11	1252.9	1.49
40-44	512.6	1.64	488.5	1.21	1001.1	1.43
45-49	426.9	1.66	403.0	0.98	829.9	1.33
50-54	371.2	2.04	354.6	1.56	725.8	1.80
55-59	379.5	1.62	366.5	1.06	746.0	1.34
60-64	347.0	1.70	364.4	1.70	711.4	1.70
65-69	263.6	1.52	302.3	1.35	565.9	1.43
70-74	208.2	1.92	262.9	1.47	471.1	1.67
75+	233.0	1.49	404.7	2.08	637.7	1.86
All ages	7940.1	2.16	7959.7	1.58	15899.8	1.87

Age	On the basis of 'usual' residence					
	Males		Females		Persons	
	No. ('000)	% under enumer-ation	No. ('000)	% under enumer-ation	No. ('000)	% under enume-ation
0- 4	615.3	1.29	585.9	1.22	1201.2	1.26
5- 9	601.3	1.23	571.2	1.22	1172.5	1.23
10-14	668.5	1.29	635.7	1.36	1304.2	1.33
15-19	685.6	2.11	654.3	1.97	1339.9	2.04
20-24	673.1	4.33	646.9	2.83	1320.0	3.59
25-29	672.6	4.02	657.2	1.80	1329.8	2.92
30-34	626.6	2.21	625.6	1.53	1252.2	1.87
35-39	630.9	1.78	616.7	1.05	1247.6	1.41
40-44	510.3	1.59	487.0	1.19	997.3	1.39
45-49	424.7	1.52	401.7	0.98	826.4	1.26
50-54	369.6	1.97	353.0	1.52	722.6	1.75
55-59	377.7	1.52	364.0	0.92	741.8	1.22
60-64	345.6	1.74	361.6	1.61	707.3	1.67
65-69	262.1	1.47	300.2	1.31	562.3	1.38
70-74	207.2	1.89	261.3	1.46	468.5	1.65
75+	232.4	1.52	403.3	2.01	635.7	1.83
All ages	7903.6	2.08	7925.5	1.54	15829.1	1.81

5. ESTIMATING SUB-NATIONAL POPULATIONS

Internal migration data are not sufficiently reliable for demographic estimates of the population at sub-national levels to be used to assess census under-enumeration. However, a comparison of the 1986 PES estimates of the number of children aged 1-15 was made with the corresponding number receiving family allowance by State/Territory. This comparison shows a general agreement except for Northern Territory where the percentage difference was more than 2%.

Given this general agreement between PES estimates and family allowance data, and in the absence of reliable independent data on higher ages for comparison with PES estimates, the PES estimates (after adjustments) of the State and Territory populations were accepted.

Population estimates at the State/Territory level by age and sex, and at the local government area level were not derived directly from the PES. The 1986 PES was a sample survey and the results are subject to sampling error. Sampling errors at the State/Territory level by age and sex and at the local government area level are high, many unacceptably high, relative to the amounts of adjustment for under-enumeration which need to be made. An alternative indirect method, using an iterative proportional fitting (IPF) procedure, was used to produce State/Territory estimates by age and sex from those higher level PES estimates with a low sampling error. For a description of the IPF procedure, see Purcell and Kish (1979). This procedure involved taking the national population estimates by age and sex and the State/Territory estimates within each sex and adjusting the census age by State/Territory counts to these two margins.

The IPF procedures involves the following cycles $n = 0, 1, \ldots$.

$$X_{gas}^{(2n+1)} = X_{gas}^{(2n)} \frac{X_{as}}{X_{as}} \; (2n)$$

$$X_{gas}^{(2n+2)} = X_{gas}^{(2n+1)} \frac{X_{gs}}{X_{gs}^{(2n+1)}}$$

and $X_{gas}^{(o)} = Y_{gas}$ the census count for state g, age category a and sex s. The procedure converges to a unique solution.The use of IPF procedures, of course, assumes that the relationship between the variables within the assocation structure is valid and that this relationship is preserved.

For estimates for local government areas, the problem with high sampling error is more acute and results of the PES are not sufficiently reliable to make direct estimates of under-enumeration for each local government area. Based on the premise that under-enumeration is age/sex and birthplace (Australian born/Overseas born) selective, and that it differs between States/Territories and between capital city and the rest of the State, adjustments for under-enumeration at the local government area level were made to reflect under-enumeration differentials by age, sex, capital city/rest of State and Australian-born/overseas-born.

6. PROBLEMS WITH THE PES ESTIMATION

As pointed out by Bailar (1985), for example, the bias and consistency of the PES estimates is affected by errors in the matching process, any correlation between a person being missed in the census and in the PES, and erroneous inclusions in either the census or the PES. It is

because of the possible effects of these factors that the results of the PES are assessed using demographic and administrative data in the ways described above.

Errors in matching will bias the PES estimates. Failure to match records that in fact should match will lead to the creation of apparently under-enumerated persons and the PES estimate will be an over estimate. The effect of false matches will be the reverse.

Erroneous inclusions in either the census or PES will inflate the values of Y or x and hence the PES estimate. The US Bureau of the Census conducts a special "E-sample" selected from the census to estimate the extent of erroneous inclusions in the census which can then be incorporated in the estimate by adjusting the census count Y. For a description of the E sample, see Fay, Passel and Robinson (1988). The matching and estimation procedures used by the ABS attempt to adjust for some of the effect of erroneous inclusions by determining not only whether or not someone has been included but whether they should have been included and if they have been included more than once. For example in the 1986 PES, 250 people were determined to have been included twice and four persons had been included three times. Cases were also found where persons had been included but should not have been. In this way viewing the PES estimation as a ratio estimator rather than a dual system estimator enables the accounting for some erroneous inclusions.

The dual system estimation method makes the assumption that whether or not someone is missed in the PES is independent of whether or not that person is missed in the census. Whilst all practical steps have been taken in ensuring that the two field and processing systems involved in the collections are completely separate and independent it is still possible for correlation to exist. Positive correlation will mean that the PES estimate based on the assumption of independence will be an under-estimate, negative correlation leads the PES estimate to over-estimate. Negative correlation would occur if being included in the census led people to be hard to enumerate in the PES but we have no clear evidence for this; the final response rate for the PES (95%) is in line with other household surveys conducted by the ABS. Positive correlation seems more likely, and there appears to have been some evidence of this in the 1981 Census. If such positive correlation exists then the PES based adjustments will have not gone far enough but will have been in the right direction.

7. ALTERNATIVE METHODS OF ESTIMATION (WOLTER 1986)

The idea of combining PES data and demographically derived sex ratios or sex ratios obtained from other sources is the basis of methods suggested by Wolter (1986). Wolter suggests several models and associated methods which formally combine sex ratios and PES estimates. These methods are attempts to loosen the assumption of independence inherent in the PES estimation methods.

Wolter considers two models. In the first it is assumed that the degree of association in under-enumeration between the PES and the census (as measured by the cross-product ratios in tables such as the diagram shown earlier in this paper) is the same for males and females within each age category. In the second model independence is assumed for females and an externally derived sex ratio is used to obtain the male figure. It is then possible to calculate the cross-product ratios implied for males.

From an initial evaluation of these methods applied to Australian data, it was found that the first model produced very erratic estimates of the cross product ratios, with approximately 50% being negative. This was greatly reduced under the second model although some remained negative and were set to zero in a modified model. The problem with negative cross-product

Table 6

Sex Ratios: Males per 100 Females

Age	Alternative	PES
0- 4	105.0	104.3
5- 9	105.2	105.2
10-14	105.2	105.3
15-19	104.7	104.7
20-24	104.1	104.1
25-29	102.6	102.6
30-34	100.3	100.3
35-39	102.4	102.4
40-44	104.5	104.9
45-49	105.2	106.0
50-54	104.2	104.7
55-59	103.0	103.5
60-64	95.2	95.2
65-69	87.1	87.2
70-74	78.8	79.2
75+	57.9	57.6

ratios was also identified by Wolter (1986, p. 7). The second model, modified, was then applied to 1986 data. For age groups 5-9 up to 35-39, the sex ratio obtained from the PES were in line with expectations and those sex ratios were used giving exactly the PES estimate. For the 0-4 age group the sex ratio obtained from demographic estimates was used and for the 40-44 to 75+ age groups, an alternative estimate of the sex ratios based on census counts was used. The sex ratios are given in Table 6.

The sex ratio used and the PES sex ratios are not greatly different so applying Wolter's second model leads to only small changes in the PES estimates. For the 0-4 and 75+ age groups the estimates of males are increased by 0.7% and 0.5% respectively. For the 45-49 and 70-74 age groups the estimates are reduced by between 0.7% and 0.5%. This analysis suggests that the differences in biases between sexes in the PES estimation method due to the combined effect of the potential problems discussed above, are relatively small. It could be the case that any biases are affecting males and females to an approximately equal degree so that PES sex ratios are broadly acceptable.

Our experience in 1981 and 1986 demonstrated the need to use sex ratios in assessing measures of under-enumeration and we believe the Wolter method is a useful way of generating alternative estimates against which the Census count and direct PES estimates can be judged. The general acceptability of the PES sex ratios in 1986 has meant that using this method made little difference. The acceptability of the PES sex ratios in 1986, except for the 0-4 age group contrasts with the experience in 1981, where an adjustment to the PES estimates was considered necessary for a number of age groups based on alternative sex ratios. These differences in the 1981 and 1986 experience may reflect a reduction in correlation between under-enumeration in the census and the PES in 1986.

8. CONCLUSION

While the ABS has adjusted the past three censuses for under-enumeration, our confidence in the basic reliability of the PES stems from its general consistency with other data sources. No fundamental change in approach is anticipated for the next census to be conducted in 1991. However, we believe there is a need to investigate further potential causes of bias, in particular the adequacy of the clerical matching procedures, and methods to overcome correlation bias. It is also planned to investigate the possibility of creating a demographic data bank on a usual residence basis, so that the effects of the large volume of short-term movements can be eliminated or reduced.

ACKNOWLEDGEMENT

We are grateful to the reviewers of this article for their comments and suggestions for further research.

APPENDIX

THE 1986 POST-ENUMERATION SURVEY

General

The 1986 PES was conducted in the 4th and 5th weeks after census night. The survey involved interviews with a sample of the population from about 35,000 private dwellings (2/3 of one percent of dwellings) across Australia involving about 100,000 persons. The sampling fraction varied between States and Territories, with the smaller States and Territories having higher sampling fractions. Personal data on name, age, sex, marital status and birthplace were obtained by interviewers for matching with information on the census form. For each person in the survey, information was sought on their place of usual residence, where they spent census night, their address before and after census night and any other address where they might have been included on a census form. At each given address, the personal information was matched to census forms to establish whether a person was missed, counted once or the number of times counted if counted more than once.

Scope and Sample Structure of the PES

Except for the special cases mentioned below, the PES included in its scope all persons who should have been enumerated in the census, except those who had gone overseas or died between the census and PES dates. Diplomatic representatives and persons in diplomatic dwellings were not included in the census. These persons were excluded from the survey as were babies born after census night. Persons in the survey who were overseas on census night were matched to census forms to determine whether they were incorrectly included in the census.

For practical reasons, very sparsely settled areas were not included in the PES. In these areas, special census procedures were used to contact and enumerate Aboriginal groups, people in mining camps, cattle stations, etc. The PES in these areas would need to rely on the same contacts and procedures adopted for the census and therefore could not accurately and independently measure under-enumeration. Consequently, the scope of the PES excluded these areas.

Non-private or special dwellings such as hospitals, hotels, and motels also were not included in the PES. The vast majority of residents in non-private or special dwellings would have been short-term residents and, according to normal ABS survey rules short-term residents would have a chance of being included in the survey at their place of usual residence where information on such persons would be obtained. A relatively small number of long term residents of these dwellings were consequently not included in the PES. For estimation purposes, populations out-of-scope were assumed to have the average capital or non-capital city rate of under-enumeration for each State as appropriate and the average Territory rate for each of the two Territories.

As non-private or special dwellings and sparsely settled areas contained less than 3% of the total population, any differences in under-enumeration of these areas compared with areas covered by the PES would be unlikely to have a significant effect on the overall estimated level of underenumeration at the State or National level.

Interaction Between the Census and the PES

It is important that the PES be conducted as independently of the census as possible. Otherwise, the factors that led to a person being missed or overcounted in the census may also be present in the PES, resulting in biased estimation of the under-enumeration. Furthermore, knowledge of the areas to be included in the PES might influence the performance of census collectors in these areas so that the PES sample would not be a representative sample of the under-enumeration. For these reasons the field and office staff used in the census and PES were totally separate. PES interviewers were not employed as census collectors or census group leaders, and census field staff were not told which areas were included in the PES.

Independence was further guaranteed in two ways – by ensuring the operational independence of the field systems, and by adopting special procedures for census forms received by mail after the PES field work commenced.

To ensure operational independence, PES field work commenced after all available census forms had been collected from the field. Thus census collectors were not in the field at the same time as PES interviewers and there was no possibility of interaction, even unintentional, between census and PES field staff.

Special procedures for census forms received after the PES commenced were required to overcome the effects PES fieldwork may have had on householders who were late returning their census forms. In some cases, PES interviewers discovered census forms still uncollected. This situation was possible because some people had preferred to post in their census forms and had not yet done so, or the census collector had been unable to make contact to collect them. Some of these people who were included in the PES may have been prompted to post their forms in, where they would not otherwise have done so. To overcome this potential bias, any census form returned by mail after Monday 20 July 1986 (the day PES interviewing commenced) was considered a late form. Special procedures for the treatment of late forms are described later in this Appendix.

Matching procedures of the PES

Matching for the purpose of determining whether a person was missed, counted once or the number of times counted if counted more than once, was conducted in two stages. Both these stages were clerical processes undertaken by staff at the census Data Transcription Centre.

The first stage was the locating of census forms for the addresses of households selected in the PES. Processing of 1986 Census forms were centralized in Sydney. Staff at the Population Census Data Transcription Centre were requested to compare the address on the front of the PES interview form with all addresses given in the record book of the census collector

who was responsible for the collection district (CD) in which the PES household was located. The record book was used as a control in the delivery and collection of census forms, and contained information such as name, address and number of persons for all households in the CD.

To assist identification of households where addresses were sometimes vague, for example in rural areas, processing staff were asked to also use names of the householders, property names etc. In addition, staff were instructed to check through all addresses in the record book so that any duplicate census forms were identified. Addresses in record books of adjacent CDs were also checked if the address of the household selected by the PES was near the boundary of the CD.

The second stage was person-matching and this was based on the name and demographic details of the persons listed on the census and PES forms. In this matching process, a search form was generated for each address reported in the PES for any person in the household, other than the address of the PES selected dwelling. A search form was treated the same way as a PES interview form and an attempt was made to locate the census form which corresponded to the search form address.

In most cases, the person-matching procedure was straight forward. There were, however, cases of spelling errors and insufficient details on addresses to identify a clear match on name. In these cases, a judgement on whether or not a person was counted was made based on other information such as age, sex, marital status, birthplace and relationship to other members in the census household. For doubtful cases, processing staff were required to consult their supervisor.

The PES also asked the respondent whether each person was included on a census form. When matching failed because of lack of adequate information, the respondent's statement about whether or not the person was counted was accepted. There were a few cases where even this information was unavailable. These cases were considered not counted in the census.

After matching, the data was entered onto computer tapes, edited and reformatted to produce a clean unit record file giving the number of times person in the PES sample were counted in the census.

Treatment of Late Census Forms and 'Dummy' Census Forms

In forming the estimation equation:

$X = Y\,(x/y)$, where
X = estimated census count adjusted for underenumeration
Y = raw census count, unadjusted
x = PES estimate of the number of persons who should have been included in the census and
y = PES estimate of the number of persons who were included in the census,

two categories of census forms were treated as missed in the census. These are 'dummy' census forms and late census forms.

Dummy census forms were created during census fieldwork for dwellings at which households were known to be residing, did not return their census forms and could not be contacted. Census collectors were instructed to exercise extreme care in creating these dummy forms and they needed to be satisfied that there was concrete evidence that the dwellings were occupied on census night. The collectors were instructed also to obtain as much information as possible regarding the number and the demographic characteristics of these residents.

When a PES address was matched to a dummy census form, the lack of name and reliable personal characteristics on the census form made it impossible to perform the matching operation satisfactorily.

It is also necessary to handle late census forms differently from normal census forms. Because late census forms might have been prompted by a PES interviewer calling, their inclusion could lead to a bias in the estimation of under-enumeration.

In the 1986 Census, there were 115,000 persons recorded on dummy census forms or late census returns, or 0.7 per cent of the population. Both dummy and late census forms were excluded from the raw census count (Y) and the PES estimate of the number of persons who were counted in the census (y), but were included in the PES estimate of the number of persons who should have been counted in the census (x). In other words, persons on dummy and late forms were treated as missed and adjusted for by (x). The adjustment factor (x/y) is exaggerated because of the exclusion of dummy and late forms from (y), but this exaggeration is compensated for by the exclusion of these forms from the raw census count (Y).

Estimation Procedure

The estimation procedure was applied at the age by sex by geographic area (capital city statistical division/rest of state) level. Adjustment factors were included in the estimation formulae to partly account for non-responding and non-contact households. These factors adjust both of the main estimates, x and y, by effectively imputing, for each non-contact or refusing household, the average number of persons per household, and, for each person so imputed, the average rate of under-enumeration at the relevant age by sex by area level. To reduce the bias from the use of such adjustment factors, the factors were calculated for various subgroups of households by the status of enumeration at the census (such as occupied dwelling, late returned form). This enumeration status was considered to be related to what non-response was encountered in the PES.

REFERENCES

BAILAR, B.A. (1985). Comments on "Estimating the population in a Census Year: 1980 and beyond" by E.P. Ericksen and J.B. Kadane, *Journal of the American Statistical Association,* 80, 109-114.

BISHOP, Y.M.M., FIENBERG, S.A., and HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge: MIT Press.

FAY, R., PASSEL, J.S., and ROBINSON, J.G. (1988). The coverage of population in the 1980 Census. Evaluation and Research Report, PHC 80-E4, United States Bureau of the Census, Washington D.C.

PURCELL, N.J., and KISH, L. (1979). Estimation for small domains. *Biometrics*, 35, 365-384.

WOLTER, K.M. (1986). Capture-recapture estimation in the presence of known sex ratio. SRD Research Report, United States Bureau of the Census, Washington, D.C.