Adjustment of Post-censal Population Estimates for Census Undercount

Research Report # 3

Adjustment of Post-censal Population Estimates for Census Undercount

Research Report #3

Statistics NEW ZEALAND Te Tari Tatau

Published in February 1998 by
Statistics New Zealand
Te Tari Tatau
Wellington, New Zealand

Preface

A post-enumeration survey was conducted in New Zealand following the 1996 Census of Population and Dwellings. Its main objective was to measure the completeness of census coverage - to gauge the extent to which individuals were missed or counted more than once in the census. Following a careful appraisal of the post-enumeration survey results, it was decided to adjust the population base for estimated net undercount, for the purpose of deriving post-censal population estimates and projections. Alternative methodologies for adjusting the population base were investigated and a draft paper carrying the main results of these investigations was subsequently circulated among an external advisory group (including the four regional health authorities) and others for review and comments. This revised report incorporates the comments received by Statistics New Zealand.

This research report was authored by Kim Dunstan, Michael Ryan and Ian Westbrooke under the editorial direction of Mansoor Khawaja, Chief Demographer. I would like to express my appreciation to them and other staff of Statistics New Zealand, especially those in the Demography and Analytical Support divisions, who critically reviewed the proposed methodology.

My thanks are also due to the members of the external advisory group, especially Professor Alastair Scott of the University of Auckland, Alistair Gray of Victoria University of Wellington, Dwayne Crombie of North Health, Grant Johnston of the Ministry of Health, Jit Cheung of Midland Health, David Scott of Central Health, and John Paice and Bill Allen of the Australian Bureau of Statistics, for their many helpful comments and suggestions.

Len Cook

Government Statistician

Contents

1. Introduction

In New Zealand there is growing interest in accurately determining the level of coverage of census and improving the general quality of post-censal population estimates. It is well known that all censuses of population fail to achieve the ideal of complete enumeration of all people intended to be counted.

In 1996 Statistics New Zealand carried out the first Post Enumeration Survey (PES) to be held in conjunction with a New Zealand Census of Population and Dwellings. The expressed objective of the survey was to provide estimates of the extent of undercount (of dwellings and people) in the 1996 Census of Population and Dwellings. Following a careful evaluation of the survey results, it was decided to adjust the population base for net undercount, for the purpose of deriving post-censal population estimates and projections. This report describes and discusses the methodology developed for adjusting the population base. It also gives an overview of the 1996 PES and describes the principal results as a preliminary to the discussion on the adjustment methodology.

The census is the largest statistical collection of information on the demographic, social and ethnic characteristics of New Zealanders and has traditionally provided the base population for deriving post-censal population estimates and demographic projections. These post-censal population estimates and projections are used for a variety of purposes including:

- allocation of national funds to organisations such as regional health authorities
- denominators for deriving fertility and mortality rates and other demographic indices as well as per capita time series
- determination of population weights for various surveys
- policy-making, planning, administration and other applications
- demographic, social and economic studies.

Users of population estimates and projections include planners, policy-makers and researchers in government departments (health, education, social welfare, law and order, energy, housing, employment), territorial authorities, market research firms, private companies and community organisations.

The justification for adjusting the population base essentially lies in the incomplete coverage of censuses everywhere. According to the standards prescribed by the United Nations (1952), where reasonable and possible, census data should be adjusted for known or estimated deficiencies:

> "Inconsistencies may be created by changes in the degree of completeness of enumeration from one census date to another, or by changes in the completeness of registration of births and deaths, or the completeness of recording of migration, over a period of time. The elimination of such inconsistencies is, of course, part of the problem of evaluating the accuracy of the statistics and correcting them for use in population estimates".

2. Main inputs and outputs

The main inputs for the post-censal population estimates and projections are:

- The 1996 Census of Population and Dwellings. This provided detailed demographic and other information on all people enumerated.
- The 1996 Post Enumeration Survey. This survey provided estimates of the level of coverage in the 1996 Census. These estimates have low sample errors at the overall level, but become less precise for smaller subgroups.
- Prior demographic knowledge about census coverage from New Zealand data, other demographic processes, and overseas experience (notably the results of the Australian Post Enumeration Survey).

The main output of this research project was the methodology to determine a base population for post-censal demographic estimates and projections. It was essential that the proposed methodology:

(a) retained the information for small population subgroups from the 1996 Census;

(b) was consistent with the direct coverage estimates from the 1996 PES;

(c) provided populations that always add to population totals at all levels; and

(d) provided all the base populations needed for demographic work, for a range of age, geographic and ethnic group combinations, some of which cannot be readily foreseen.

From a demographic perspective, Statistics New Zealand introduced the “adjustment methodology” as a restricted statistical initiative. There was, for example, no intention to revise the 1996 Census figures as such. The overriding aim was to adjust the population base, which will be used for deriving annual post-censal population estimates (national and subnational) as well as demographic projections.

3. 1996 Post Enumeration Survey

It is appropriate to briefly review here the background and principal findings of the 1996 Post Enumeration Survey. Details about the survey's scope, the information gathered and the results are published in Statistics New Zealand's media release, *Post Enumeration Survey (PES): 1996* (June 1997) and technical report, *A Report on the 1996 Post Enumeration Survey*.

The 1996 PES was based on a stratified sample of about 10,400 permanent private dwellings throughout New Zealand. The survey population was roughly 25,000 New Zealand residents either usually resident or staying in a New Zealand private dwelling during the survey period (18-31 March). Following overseas practices, temporary and non-private dwellings were excluded. These dwellings can be categorised into those which:

(a) are quite stable, where residents can be counted accurately (eg long-term care rest homes for the elderly);

(b) are very transient, where the occupiers are likely to be different two weeks after the census (eg tourist accommodation); and

(c) require special enumeration procedures (eg people with no fixed abode).

One of the principles of the PES was to have independent procedures from census. Cost-benefit considerations were also important, and very sparsely populated areas were excluded. In summary, the survey excluded the following groups:

- overseas visitors
- people living in non-private dwellings (eg hospitals, prisons, hotels)
- people living in temporary private dwellings (eg tents, caravans, yachts)
- people who died after census night
- babies born after census night
- overseas diplomats, their families and people living with them
- people on offshore islands (except Waiheke Island which was included).

As permanent private dwellings account for 99 percent of all occupied dwellings and 95 percent of the resident New Zealand population, the contribution of other dwellings to errors in coverage was likely to be relatively small. Therefore, a somewhat arbitrary assumption was made that net undercount rates were the same among all dwellings, which is broadly consistent with the practice followed by the Australian Bureau of Statistics. However, the age, sex and ethnic structure of persons in temporary and non-private dwellings was taken into account in the final adjustment of the population.

3.1 Survey results

Main results from the 1996 Post Enumeration Survey (PES) are set out in Table 1, and these indicate that about 1.2 percent of New Zealand residents were undercounted at the 1996 Census. This represents a net undercount of about 43,000 people, resulting from an undercount of 49,000 offset by an overcount of 6,000 (people being counted more than once). The 1996 Census counted 3.62 million residents in New Zealand. The resident population would be 3.66 million, if an adjustment is made for the estimated net undercount. This adjusted figure, however, excludes New Zealand residents temporarily overseas on census night.

Table 1

1996 Post Enumeration Survey Results

Variable	Census Population	Net Undercount		Sample Error (1)
	(000)	(000)	Percent of Population	Percent of Population
Total	3,618	43	1.2	0.2
Sex				
Male	1,777	25	1.4	0.3
Female	1,841	18	1.0	0.2
Age Group (years)				
0-14	832	12	1.4	0.4
15-29	808	18	2.1	0.5
30-44	834	7	0.9	0.3
45+	1,144	7	0.6	0.3
Ethnic Group				
Maori	523	16	2.9	0.7
Pacific Island	173	6	3.1	1.1
European	2,595	21	0.8	0.2
Geographic Area				
Regional Health Authority				
Northern	1,219	16	1.3	0.4
Midland	724	9	1.3	0.5
Central	891	9	1.0	0.4
Southern	783	6	0.8	0.3
Northern North Island (2)	1,780	23	1.3	0.3
Southern North Island (3)	938	9	1.0	0.3
North Island	2,718	33	1.2	0.3
South Island	899	7	0.8	0.3

(1)The extent to which an estimate from the PES might have varied by chance because only a sample of dwellings was included. For example, there is a 95 percent chance that the true estimate of undercount for New Zealand would have been between 1.0 and 1.4 percent (35,000 and 50,000 people) if all dwellings had been included.

(2)Consists of Northland, Auckland, Waikato and Bay of Plenty Regions.

(3)Consists of Gisborne, Hawke's Bay, Taranaki, Manawatu-Wanganui and Wellington Regions.

Note: Owing to rounding and unspecified categories, individual figures may not sum to give the stated totals.

The PES also found that the 1996 Census missed more males than females - about 25,000 males compared with 18,000 females. The undercount was 1.4 percent for males compared with 1.0 percent for females. Variations in undercount by age were also marked. Young adults (15-29 years), who are quite mobile and have a greater risk of being missed in the census, had the highest undercount of 2.1 percent. The undercount for those aged 45 years and over was only 0.6 percent.

There were also significant ethnic differentials. The highest undercount rate was among the Pacific Island ethnic group at 3.1 percent, closely followed by the New Zealand Maori ethnic group at 2.9 percent. For the remainder of the population (predominantly European) it was 0.8 percent. The sample size was too small to give reliable information for other ethnic groups. Variation in regional census coverage was somewhat less pronounced. The rate of undercount varied from 1.3 percent in the northern half of the North Island (Northland, Auckland, Waikato and Bay of Plenty Regions) to 0.8 percent in the South Island.

The estimated number of permanent private dwellings undercounted by the 1996 Census was about 0.5 percent or 6,000 nationally. The 1996 Census counted almost 1.3 million dwellings. The sensitivity of undercount rates to the PES matching classification is discussed in Appendix 5.

3.2 International comparison

Many countries conduct surveys to check the coverage of their population census. The Australian Bureau of Statistics, for example, has run evaluation programmes since 1966, the US Census Bureau since 1951, Statistics Canada since 1966, and the National Statistical Office of the United Kingdom since 1981. Net undercount rates for these countries and New Zealand are compared in Table 2.

Table 2

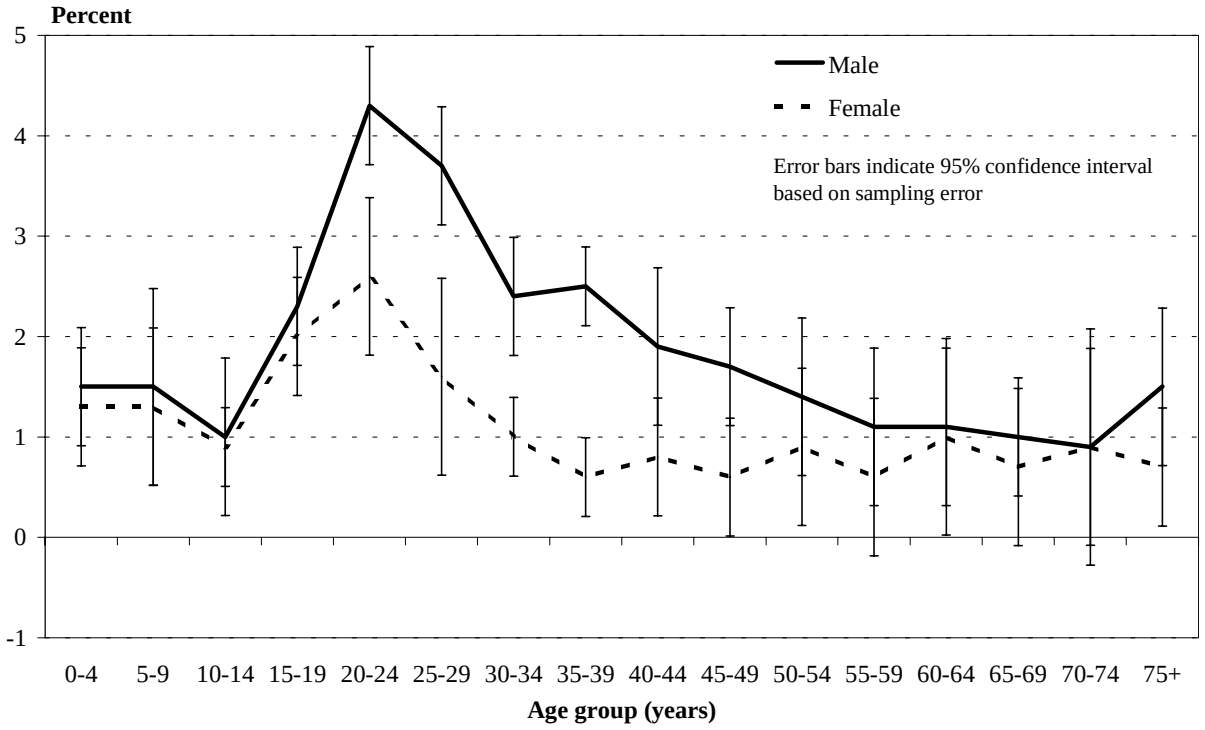
Estimated Census Net Undercount
New Zealand and Selected Countries

Country	Census Year	Number (000)	Percent
Australia	1986	300	1.9
	1991	320	1.8
	1996	290	1.6
Canada	1981	400	1.7
	1986	700	2.7
	1991	810	2.9
NZ	**1996**	**43**	**1.2**
UK	1981	1,250	2.3
	1991	1,050	1.9
USA	1980	2,700	1.2
	1990	4,700	1.8

The undercount rate of 1.2 percent in New Zealand's 1996 Census compares favourably with the rates in these other countries. New Zealand's differentials by sex and age are broadly similar to those found in other countries. For example, the 1996 Australian PES estimated a net undercount of 2.0 percent for males and 1.1 percent for females (Australian Bureau of Statistics). The estimated net undercount of males and females in the 1991 Canadian Census was 3.4 and 2.4 percent, respectively (Statistics Canada 1994). In the 1996 Australian Census, males had a higher rate of undercount than females in all age groups, and the rates for people aged 20-29 years was over twice that of other age groups (Figure 1).

Figure 1

Census Net Undercount
1996 Australian PES



3.3 Demographic estimates of undercoverage

Estimates of undercoverage are discussed in Appendices 1 and 2. Generally, the estimates of the resident population derived by updating the initial population base for births, deaths and permanent and long-term migration in the ensuing period are affected by category jumping in external migration. There are, therefore, inconsistencies in the quality of this demographic data which requires more detailed analysis.

4. How to create an adjusted population

It is instructive to outline here the statistical procedure adopted by the Australian Bureau of Statistics, because it faces a problem broadly similar to that confronting Statistics New Zealand. The ABS adjustment methodology involves a standard statistical procedure - an iterative proportional fit (IPF). The procedure starts from a table of census populations cross-classified by age, sex and state of residence. The populations are adjusted to meet the PES estimates for total Australian populations for each age group, and for the age-sex group total for each state. (In practice, some modifications of the PES estimates based on demographic analysis are worked out at the national level and applied at the state level.) The IPF procedure produces adjusted population estimates for each cell in the table by forcing the initial estimates to agree with the higher level constraints, in a cyclic iterative fashion. For a more detailed description of the IPF procedure, see Purcell and Kish (1979).

A similar process is used within most states to produce adjusted populations for the capital city and the remainder of the state. The age-sex populations for smaller areas within each state are simply scaled. That is, the adjustment factor for each age-sex group is assumed constant within parts of each state. This is a *simple* synthetic approach. The information from the PES is only used directly at the highest level, and applied to the lower levels through the IPF and scaling models. Underlying the IPF is an implicit log-linear model of the structure of the census-based population.

Much work in this area of statistics - small domain estimation - has focused on the best ways of combining direct survey information (in this case from the PES) with a synthetic model approach. The AB S has chosen not to adopt the *combined* approach for two main reasons:

- it requires a huge investment of resources for, probably, small gains in the quality of population estimates
- it is difficult to ensure additivity and consistency for all population estimates at all levels. Indeed, attempting to make sophisticated estimates both additive and consistent may undo any advantages gained from the combined approach.

We have adopted an alternative simple synthetic approach for creating adjusted population estimates. It is broadly similar to the approach used by the ABS, but we were working with different inputs and our output requirements were different.

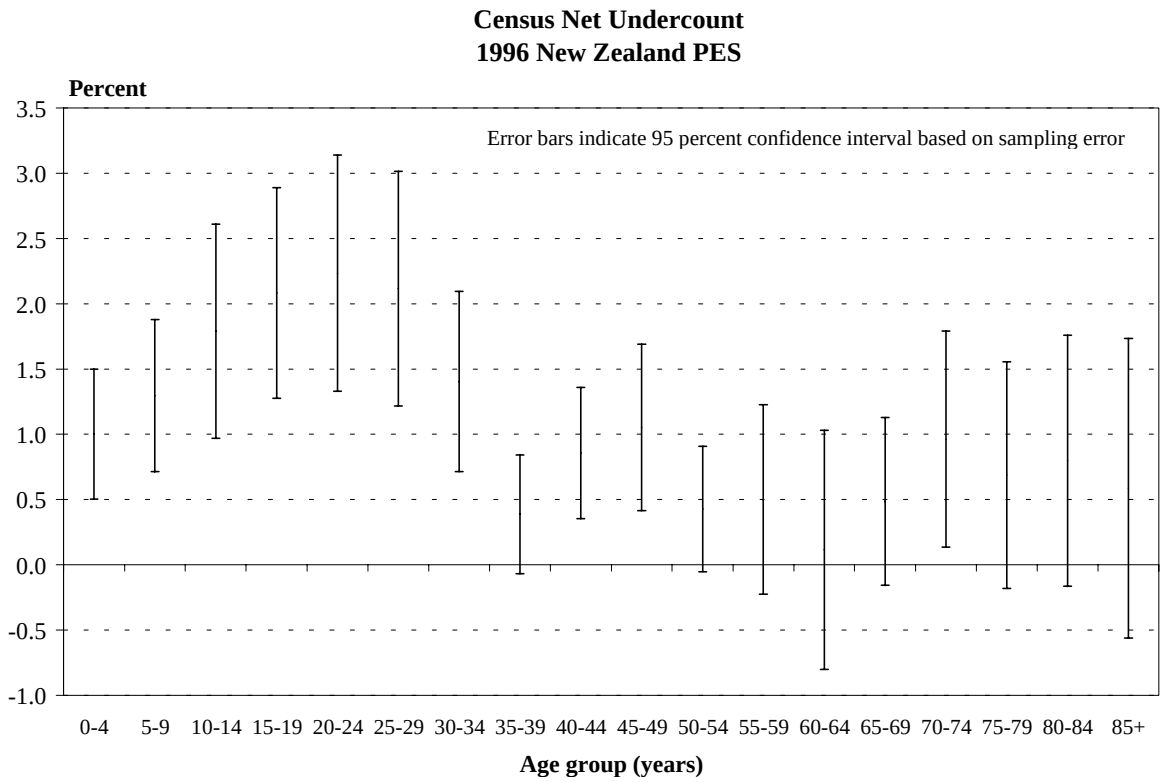
4.1 Comparison with the Australian approach

Age and sex were key factors for modelling New Zealand census coverage, as in Australia and most other countries. However, we found that ethnicity was a critical factor for New Zealand, while region had much less importance once we accounted for age, sex and ethnicity. Therefore, we used age, sex and ethnicity in developing our statistical model.

We needed a statistical procedure that would produce consistent population bases for many possible combinations of age groups, ethnic groups and areas that cannot necessarily be specified in advance. A demographic ‘wish list’ asked for a cross-tabulation of 1,700 area units by 100 age groups, by 11 ethnic groups by 2 sexes - approaching 4 million cells. These are needed to be reasonably confident of meeting all the possible demands for projections and estimates in the post-censal period. It may be possible to produce such a table using the iterative proportional fit approach, but in our case it would be pushing this approach far beyond the level at which the ABS applied it.

The New Zealand PES had a smaller sample size than the Australian one (10,000 compared with 36,000 dwellings). While the Australian PES gives quite a smooth graph of undercount for five-year age groups by sex, the New Zealand PES gave a less smooth graph with greater sampling errors (Figure 2).

Figure 2



Statistics New Zealand found that the size of the PES meant that direct estimates were of sufficient quality to justify publication for only four broad age groups (<15, 15-29, 30-44 and 45+ years). Using direct PES estimates only in the adjustment would have meant adjusting at this level of groupings. This would have given substantial discontinuities in the adjustment factors at the breaks in age groups, for example, between ages 14 and 15, or ages 29 and 30. Such discontinuities are implausible, and their location would be entirely dependent on arbitrary age groupings.

5. Choosing a synthetic model

Instead of an iterative proportional fit we decided to produce very fine age group adjustment factors for a limited number of ethnic groups by sex. New Zealand's national population estimates require quarter year of age breakdowns, and our approach allowed us to extract factors at this level. We found that given the limited amount of data for modelling age group structure, two broad ethnic groupings worked best for modelling - Maori and Pacific Island combined and 'Other'. For each ethnic-sex group, we smoothed the PES data using a parametric model for the age structure of the adjustment factors. Congdon (1993) gives an exposition of the philosophy underlying the approach adopted here. The details of the Rogers migration function are given in Appendix 3.

The Rogers migration function was used to capture our expectations of the age structure of undercount in the New Zealand PES. This prior knowledge drew on the structure of undercount in other countries with larger PES samples (notably Australia) and the age dependency of other related demographic phenomena (especially migration). The overall features of the curve came from the model, but the specifics of shape and height were determined from the PES data through estimating the function parameters. Due to the high sample errors in the New Zealand PES data, some of the parameters needed to be constrained, and in one case we gave zero weight to PES data in a five-year age group which caused the model fit to be demographically implausible.

Use of the statistical model provided much finer age structure, and reasonable adjustment factors for smaller ethnic groups, than could have been achieved using an iterative proportional fit. Our aim was to summarise the PES undercount results in models that made sense demographically and have only the parameters needed to capture the essential features of PES undercount patterns. Details of the modelling are given in Appendix 4.

5.1 Evaluating the statistical model

We checked the Rogers modelling of the PES results against the PES sample data. The PES can be used to derive two important estimates on the coverage of individuals:

- those who were counted in the census
- those who should have been counted in the census.

We evaluated proposed adjustment factors by applying the adjustment factors to those in the PES who were counted in the census, and then comparing these adjusted estimates with the estimates of those who should have been counted in the census. Comparisons for the modelled variables (age, sex and ethnicity) provided checks on the fitted Rogers functions. Comparisons for the regional variables not modelled enabled validation of the fitted Rogers functions. If there were important regional effects not explainable by age, sex and ethnicity, then poor comparisons would occur. Conversely, good comparisons support the assumption of no regional effects. We judged the closeness of the comparisons by reference to the sample error (1.96 times the standard error of net undercount, expressed as a percent).

Two aspects of fit provided the basis for evaluating the Rogers model:

- the difference between fitted and direct PES undercount rates given in Table 3
- the demographic plausibility of the shape of the age models (see Figure 11).

Table 3 shows that the model provided good fits for sex and ethnicity, which were two of the three variables in the model. Although the age and regional variables gave larger percent differences, these differences were all smaller than the corresponding sample errors. These results tend to confirm that there were no important regional effects in the PES data beyond that due to the variables age, sex and ethnicity. No regional factors were included in the model, and the reasonable regional fits support the view that they were not needed.

Table 3

Differences Between PES Estimates and Rogers Model Estimates

Variable	In Rogers Model	Categories	PES Net Undercount Direct Estimate (Percent)	PES Sample Error (Percent)	PES Estimate Minus Model Estimate (Percent)
Sex	Yes	Male	1.4	0.3	0.0
		Female	1.0	0.2	0.0
Age Group	Yes (constrained)	0-14	1.4	0.4	0.4
		15-29	2.1	0.5	0.1
		30-44	0.9	0.3	-0.3
		45+	0.6	0.3	-0.2
Ethnicity	As Maori & PI	Maori	2.9	0.7	-0.1
	As Maori & PI	Pacific Island	3.1	1.1	0.1
	Yes	'Other'	0.8	0.2	0.0
RHA	No	Northern	1.3	0.4	-0.1
		Midland	1.3	0.5	0.1
		Central	1.0	0.4	-0.1
		Southern	0.8	0.3	-0.2
Urban	No	Auckland	1.3	0.5	0.0
		Ham/Wgtn/Chch/Dun	0.8	0.4	-0.3
		Rest of North Island	1.2	0.3	0.1
		Rest of South Island	0.9	0.4	-0.1
Island	No	North	1.2	0.3	0.0
		South	0.8	0.3	-0.2

The critical variable to evaluate was age. Our preferred model reduced the undercoverage adjustment in the 0-14 age group. Based on the Australian patterns and on demographic knowledge, we observed a model that allowed undercoverage rates to rise much before age 15 as implausible. We constrained the model parameters for location and shape to be similar to the better defined Australian model parameters. This resulted in the model giving broad age group adjustment factors that were close to the boundaries of the published 95 percent confidence intervals for the PES for two of the four age groups. Not surprisingly, the biggest difference was for ages 0-14.

Further details of the model fit, including graphs of the fits for the 1991 Australian undercount by sex, and for the New Zealand undercount by sex and two broad ethnic groups, are set out in Appendix 4.

In summary, we were looking for parameters that:

(a) reproduced closely the results of the PES sample based estimates;

(b) fitted the PES data well; and

(c) made demographic sense.

Our modelling showed this was possible. We believe that the simple practical approach we took to estimation provided adequate adjustment factors. Further research on the estimation of the parameters, more directly from the PES data, may be desirable in the future.

Once the basic parameters were calculated, it was relatively straightforward to use the Rogers modelling to derive a base population corrected for net undercount. The appropriate model-based adjustment factors were applied to each age (single year), sex (male, female) and ethnic group (Maori, Pacific Island, Other) in the base population. Any adjusted population required will be consistent and additive.

6. Summary and discussion

There were three key elements in considering a possible adjustment of the population base for deriving post-censal population estimates.

1. The PES demonstrated the grounds for adjusting population estimates for census undercoverage at the total New Zealand level. It also showed statistically significant differences in undercoverage rates by age, sex and ethnicity.

2. Users of population estimates need the estimates to add up at every level. Therefore, if an adjustment was made, consistent adjustments should be made at every level.

3. This report proposes a statistical model which facilitates such an adjustment with some confidence.

The Post Enumeration Survey showed under-enumeration of the population at the national level in the 1996 Census. It also clearly showed higher undercount for males than females, for Maori and Pacific Island ethnic groups compared with European ethnic groups, and for young adults. These results were consistent with expectations and the experience of other countries, notably Australia.

The overall undercount of between 1.0 and 1.4 percent, according to the PES 95 percent confidence interval, was slightly lower than the 1.6 percent estimate for the 1996 Australian Census.

The adjustment of the population base for census undercount has improved national estimates of population as well as estimates for large subgroups, such as for males and females and for Maori, Pacific Island and European ethnic groups. One cannot be certain of the adjustment for smaller subgroups, as there was insufficient sample in the PES for small groups, but both the absolute and relative adjustments are minor for small subgroups. Adjustments had to be made at every level if we were to achieve the additivity that our users need.

In creating an explicit model we have drawn on demographic knowledge and Australian undercount patterns. This enabled us to make better use of the limited information available from the 1996 Post Enumeration Survey than if we had used the PES results in isolation. There were statistical justifications for adjustment and this report outlines an appropriate methodology for adjustment.

References

Australian Bureau of Statistics (1995), *Census of Population and Housing, 6 August 1991, Census 91: Data Quality - Undercount*, Catalogue No. 2940.0.

Australian Bureau of Statistics, personal communication.

Chan M (1994), *Census undercount in the New Zealand Population*, Statistics New Zealand Internal Paper.

Congdon P (1993), "Statistical Graduation in Local Demographic Analysis and Projection", *Journal of the Royal Statistical Society*, Series A, 156, p237-270.

Purcell N J and Kish L (1979), "Estimations for Small Domains", *Biometrics*, 35, p365-384.

Rogers A and Castro L J (1986), "Migration" in *Migration and Settlement: A Multiregional Comparative Study*, (ed. Rogers A and Willekens F J), Reidel, Dordrecht.

Stillwell J and Congdon P (1991), *Migration Models: Macro and Micro Approaches*, Belhaven Press, London and New York.

Statistics New Zealand (1996*), 1996 Post Enumeration Survey & Post-Censal Population Estimates*.

Statistics Canada (1994), *Coverage: 1991 Census Technical Reports*, Catalogue 92-341E.

Statistics New Zealand (1997*),* "Post Enumeration Survey: 1996", *Hot Off the Press* release.

Statistics New Zealand (1998*), A Report on the 1996 Post Enumeration Survey*, Wellington.

United Nations (1952), "Methods of Estimating Total Population for Current Dates", *Population Studies*, No. 10, p7.

Appendices

1. Estimated usually resident net undercount in the 1981, 1986 and 1991 Censuses
2. The consistency of undercount identified by the PES with other data
3. Rogers migration function
4. Modelling net undercount rates
5. Sensitivity of undercount rates to match classification

Appendix 1: Estimated usually resident net undercount in the 1981, 1986 and 1991 Censuses

This appendix carries some indicative estimates of net undercount for earlier censuses. These are rough figures only and were derived by working backwards from 1996. Essentially, this involved subtracting the relevant population components between each census. Estimates of undercount for earlier censuses are, therefore, subject to the limitations inherent in the estimates of those various population components, including category jumping in migration statistics, and should be interpreted with due reservation.

Table 4

Exploration of Usually Resident Undercount in 1981-91 Censuses

What the 1996 undercount implies for earlier undercount

	Intercensal increases				Status of 1996 numbers
	1981-86	1986-91	1991-96		
	1981	**1986**	**1991**	**1996**	
1. Census UR	3,143,300	3,263,300	3,373,900	3,618,300	actual
2. RTO	23,300	23,800	35,000	39,800	estimate
3. Increase due to births, deaths, PLT migration	94,000	100,800	234,200		actual
4. Net # category jumpers	26,600	24,600	-10,000		estimate
5. Incr + CJ	120,600	125,400	224,200		
6. Undercount (back calculation)	*66,200* (back calc.)	*66,300* (back calc.)	*69,900* (back calc.)	**44,900**	estimate
7. Undercount %	*2.1%* ↑	*2.0%* ↑	*2.0%* ↑	*1.2%* ↓	derived
8. Total UR (back calculation)	*3,232,800* (back calc.) ←	*3,353,400* (back calc.) ←	*3,478,800* (back calc.) ←	*3,703,000*	sum

Note: PLT = permanent and long-term.

In the processing of migration statistics the classification of each passenger as a 'short-term' or 'long-term' migrant is primarily determined by the passenger's response, on the arrival or departure card, to the question on intended or actual length of stay / absence. If the person's intention changes later during the trip, then they may also change their migrant category. For example, a person may come to New Zealand with the declared intention of settling permanently, but in fact return overseas after a few months, and therefore become a 'category jumper'. Category jumping counts are net; positive when they add to the New Zealand UR population.

Census UR (row 1) in this exploratory Table 4, gives census counts of persons usually resident in New Zealand, who were in New Zealand on census night. **RTO** (row 2) gives estimates of New Zealand residents temporarily overseas (RTO) at census. These two rows sum to give the New Zealand usually resident population at census date if New Zealand residents temporarily overseas on census night are brought back, but net undercount of usual residents is ignored.

Rows 3 to 5, labelled **Increase due to births, deaths, PLT migration**, **Net # category jumpers** and **Incr + CJ** are respectively:

- the intercensal increase of population, not corrected for category jumping;
- estimates of the net numbers of category jumpers in permanent and long-term (PLT) migration between consecutive censuses; and
- the intercensal increase, corrected for category jumping.

The nature and volume of category jumping varies over time. The estimates of category jumping cited in Table 4 are somewhat crude and imprecise, and estimated to range between 25,000 and 28,000 during 1981-86; 22,000 and 27,000 during 1986-91; and -6,000 and -14,000 during 1991-96.

The UR net undercount figure (row 6) for the 1996 Census (in bold) basically drives the estimates of net undercount for the three earlier censuses (row 7). These estimates are essentially 'residual' figures, obtained after various components of population change have been accounted for and an allowance for category jumping has been made. Estimates for 1981-91 are shown in italics to distinguish them from the 1996 input figure. Note that the undercount estimate of 44,900 is the sum of the undercount adjustment factors applied to the Maori, Pacific Island and 'Other' populations (see Figure 11). The undercount percentage (row 7) is defined in relation to the estimated true census count of persons usually resident in New Zealand (**Census UR** row 1 + **Undercount** row 6).

Row 8, **Total UR**, gives the derived estimate of the New Zealand usually resident population at census date if due allowance is made for the New Zealand residents temporarily overseas at census and for any net undercount of usual residents. For example, for 1996:

3,703,000 = 3,618,300 + 39,800 + 44,900.

Basically, Table 4 involves two steps in estimating the net undercount for previous censuses.
For example, for 1991:

1.	Total UR 1991 (row 8)	=	Total UR 1996 (row 8)	-	intercensal increase between 1991 and 1996 (row 5)		
	3,478,800	=	3,703,000	-	224,200		
2.	Estimated UR net undercount 1991 (row 6)	=	Total UR 1991 (row 8)	-	Census UR 1991 (row 1)	-	RTO 1991 (row 2)
	69,900	=	3,478,800	-	3,373,900	-	35,000

In summary it is worth reiterating that this demographic accounting approach means that imperfections in any of the input figures (including in the components of population change) will flow through to all estimated figures. The results in the above table suggest that the net UR undercount of 1.2 percent in the 1996 Census translates to a net UR undercount of about 2.0 percent for the 1981, 1986 and 1991 Censuses. Given the uncertainty in the population components, the above table also suggests that the estimated net UR undercount in the 1981-91 Censuses is in the range 1.7 to 2.4 percent.

Appendix 2: The consistency of undercount identified by the PES with other data

This appendix provides a broad assessment of the consistency of the undercount in the 1996 PES with alternative information on undercount derived by demographic analyses.

Total undercount

The net undercount of the New Zealand usually resident (UR) population was estimated by the 1996 PES to be about 43,000 persons or 1.2 percent (Table 5).

Table 5

Net Undercount
New Zealand Usually Resident Population
1996 Census

	UR Population in Permanent Private Dwellings	UR Population in Temporary Private and Non-Private Dwellings	Total
Census count	3,452,000	166,000	3,618,000
PES estimate	3,493,000	168,000	3,661,000
Net undercount	41,000	2,000	43,000
Sample error	7,000		
Relative error (%)	17		

Chan (1994), considering errors of closure for the 1945-91 Censuses, calculated that the de facto (DF) net undercount at the 1991 Census was likely to have been between 2.1 and 3.5 percent. The smaller figure of 2.1 percent assumed no DF undercount in the 1945 Census; the larger figure of 3.5 percent assumed that the net undercount rate was 3.0 percent in 1945. The DF population includes overseas visitors and it is conceivable that the increasing proportion of tourists in the DF population might be associated with a higher level of undercount because of their highly mobile and temporary nature.

The estimated UR net undercount of 43,000 people at the 1996 Census implies an estimated UR undercount of about 70,000 people in 1991 (Appendix 1: Estimated Usually Resident Net Undercount in the 1981, 1986 and 1991 Censuses).Using external migration data, overseas visitors were undercounted by an estimated 12,500 people in 1991. Our best estimate is that the 1991 Census DF population of 3.435 million was an undercount of about 82,500 or 2.4 percent. This figure of 2.4 percent falls within the plausible range of 2.1-3.5 percent calculated by Chan.

Undercount of children (0-14 years)

The 1996 Census gave a total of 832,100 children aged under 15 years resident in New Zealand. The PES estimate of net undercount for children was about 1.4 percent, or 11,500 (Table 6). Thus, using the census and the PES we estimate the number of resident children at 5 March 1996 as 843,600.

Table 6

Net Undercount
Usually Resident Children Aged 0-14 Years
1996 Census

	UR Population in Permanent Private Dwellings	UR Population in Temporary Private and Non-Private Dwellings	Total
Census count	807,200	24,900	832,100
PES estimate	818,400	25,200	843,600
Net undercount	11,200	300	11,500
Sample error	3,000		
Relative error (%)	30		

The PES estimate of 843,600 children can be compared to the demographic estimate starting in 1980 and built up using births, deaths and permanent and long-term migration in the 0-14 age group through to 1996. The population derived using demographic components will be a valid estimate of the New Zealand resident population for ages 0-14 at 5 March 1996 provided there is no category jumping in migration (between the short-term and the long-term categories) by New Zealand residents at ages 0-14. The expected population estimated this way was 850,400 New Zealand residents aged 0-14 at 5 March 1996.

Thus the demographic estimate exceeds the PES estimate by about 6,800. This is outside the range encompassed by the sample error (roughly two times the standard error). The z-statistic for this difference is 4.0. If the assumption underlying the demographic population aged 0-14 years is true then the PES result is too low. However, counts of New Zealand residents from the 1996 Census strongly suggest that the assumption of no net category jumping is unrealistic.

Figure 3 indicates that the demographic estimates are larger than the 1996 Census counts for ages 0-9 years, but smaller than the counts for ages 10-14 years. If the demographic estimates for ages 10-14 do not suffer from any net category jumping then this implies that the 1996 Census overcounted at ages 10-14. We believe that this is highly unlikely to have occurred. The PES did not show any clear evidence for net overcount for any of the 5-year age groups (see Appendix 4). The demographic estimates of net undercount for 1986 and 1991 show different patterns. Our best interpretation is that the assumption of no net category jumping at ages 0-14 is not valid. If this is accepted then there is no logical basis for using the demographic estimate for giving an independent assessment of the New Zealand UR 0-14 population at the 1996 Census.

An overall undercount of New Zealand residents of about 1 percent would imply that school rolls (data provided by the Ministry of Education) are consistent with the census and PES for ages 5-9 but understated for ages 10-14 years. At age 15 it is accepted that school rolls data does not include all usually resident 15-year-olds. This view is confirmed by Figure 3. However, there are other inconsistencies in these demographic data which limit the conclusions that can be drawn without more detailed analysis.

Figure 3

Estimated Net Undercount of Census Population Compared with Demographic Estimates

Appendix 3: Rogers migration function

As stated in the section on synthetic modelling (page 11) we used the Rogers migration function to capture our expectations (or prior knowledge) of the age structure of undercount. This model is outlined in Rogers and Castro (1986), where they propose decomposing migration by age into the sum of four components:

$$\mathbf{m_x = a_1\ exp(-\alpha_1 x) + a_2\ exp\{-\alpha_2(x-\mu_2) - exp[-\lambda_2(x-\mu_2)]\}}$$
$$\mathbf{+a_3\ exp\{-\alpha_3(x-\mu_3) - exp[-\lambda_3(x-\mu_3)]\} + c}$$

The first term models the relatively high migration at younger ages that drops through school ages, and the third term models an upturn in migration around retirement ages. The last constant term models the background level, regardless of age. The key second term models the most obvious feature of migration by age - a sharp rise in activity for young adults, dropping more slowly into maturity. For this central hump, the Rogers model incorporates a well-established model for the distribution of age at first marriage. The full Rogers model has been used by the UK Office of Population Census and Surveys to generate migration assumptions for projecting populations of subnational areas in England (Stillwell and Congdon, 1991).

We do not need the first and third terms for describing how undercount patterns vary by age, so dropping these terms gives:

$$\mathbf{m_x = a_2\ exp\{-\alpha_2(x-\mu_2) - exp[-\lambda_2(x-\mu_2)]\} + c}$$

We have modified this equation to a parametric form that is more suitable for modelling:

$$\mathbf{f(x) = \alpha + \beta\ exp\{\gamma(1-\lambda(x-\mu) - exp(-\lambda(x-\mu)))\}}$$

where
f(**x**) is net undercount rate at age **x**
exp is the exponential function
α is the background level of undercount, a constant regardless of age
β is the maximum value of the double exponential component of the above equation
μ is the age at which the double exponential component achieves its maximum (ie the mode)
λ and **γ** are shape parameters for the double exponential component
and all parameters are positive.

The main modification from the reduced Rogers function is that we have reparameterised slightly, replacing **α** by **γ** and **λ** by a new **λ**. This ensures that μ is the mode and eases the task of modelling. The other modifications include renaming parameters, and adding 1 to the first exponent to ensure that **α**+**β** is the maximum value of f(x).

Appendix 4: Modelling net undercount rates

Fitting the Rogers model presented some challenges - conceptual and practical. At the conceptual level, there were two types of error involved. First was the sampling error for the Post Enumeration Survey. Second was the error in the modelling process itself. Attempting to deal with both sets of errors simultaneously was beyond the resources of this project, although it may be a desirable area for future research. Therefore, we dealt with the parameter estimation in two stages. To reduce the impact of the sampling error, we took the undercount rates by five-year age groups. This is a standard age level for demographic work, and was the level at which the Australian PES data was published. We then used non-linear least squares (Proc NLIN in SAS) to fit the Rogers model separately for selected ethnic groups by sex. Figures 4 and 5 show the Rogers curves fitted to 1991 Australian PES undercount data.

Figure 4

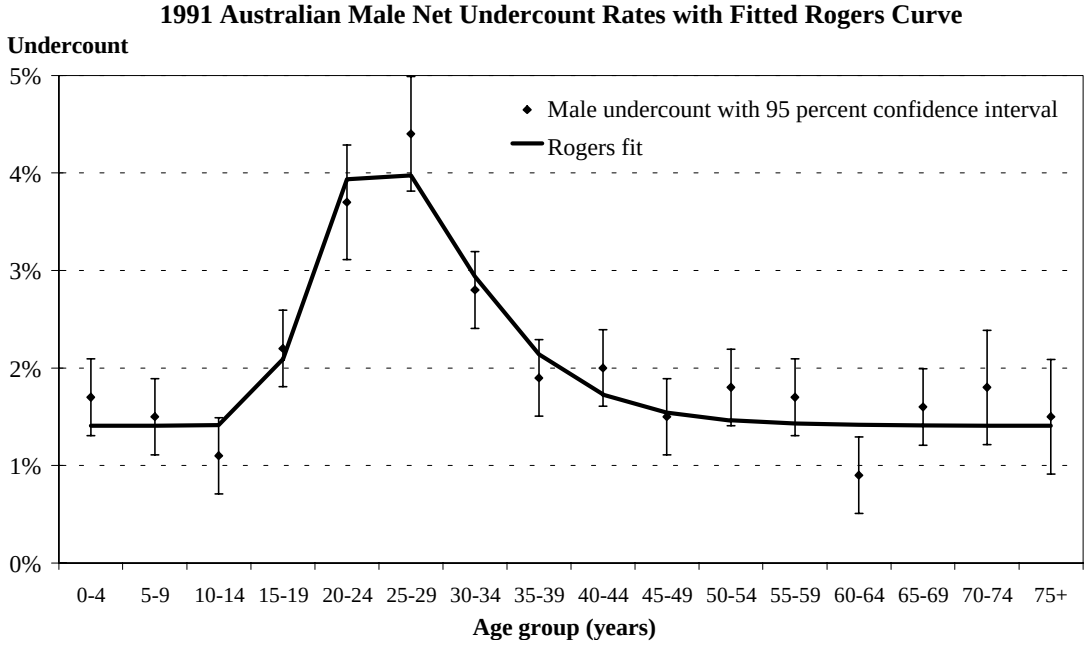
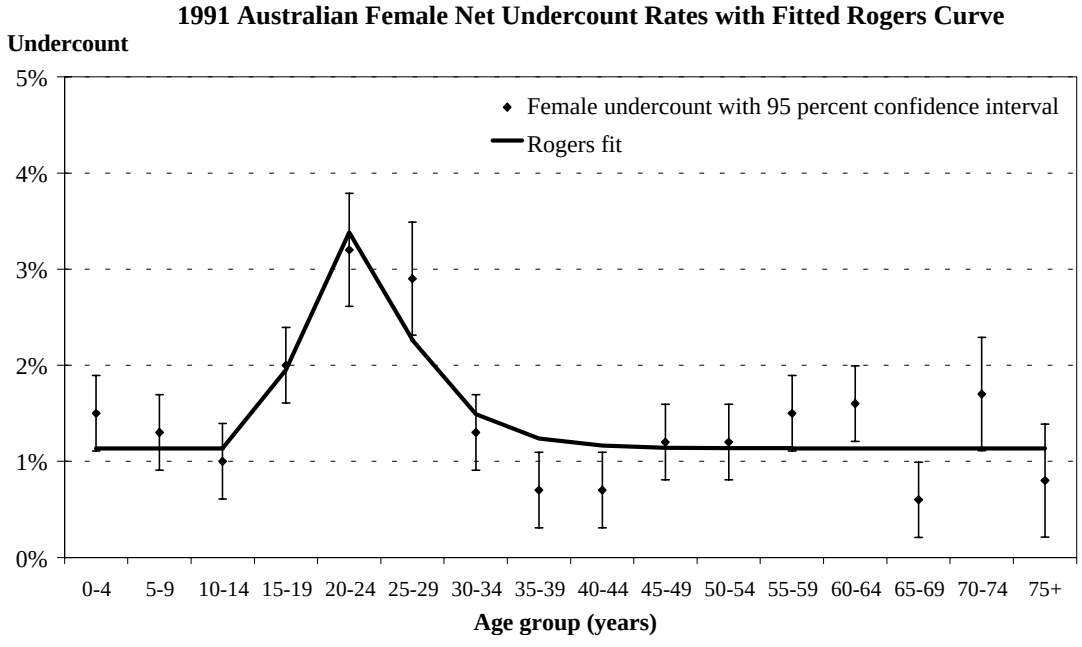


Figure 5



The Rogers function derived for Australian males fitted very well (the curve passes through all but one of the 95% confidence limits). The Rogers function derived for Australian females did not fit quite as well as that for males. It does capture the obvious peak in the 20-29 age group, but the curve misses four of the 95% confidence intervals in the 35-69 age range. However, the curve provides a reasonable summary of the female net undercount rates by age.

For the Australian data, β was estimated to be almost exactly twice α. We found a similar but less clear result for New Zealand data by broad ethnic groups. We decided it was appropriate to reduce the number of parameters by fixing $\beta=2\alpha$. This set the peak undercount rate at three times the base rate. This left us with 3 parameters for the model fit: λ, μ and γ.

Demographic imperfections were apparent in the preliminary fitting of Rogers functions to New Zealand PES data. The curves for total males and females were high at ages 10-14 years (rising significantly above the base rate towards the maximum rate). This was different from the Australian curves at ages 10-14 years, which show fitted Rogers values close to the base level of undercount. The errors bars show that the Rogers parameters were not well determined from the PES data. Therefore, there was some scope to derive parameters that made demographic sense, in particular that were closer to the Australian patterns.

We found that using two groupings of ethnicity, Maori and Pacific Island combined, and 'Other' (non-Maori / non-Pacific Island), gave satisfactory results. The overall undercount rates for the Maori and Pacific Island groups were not significantly different from the direct PES results (2.9 and 3.1 percent, respectively). The pooled data for Maori and Pacific Island groups gave acceptable fits for the constrained Rogers model. 'Other' females also fitted satisfactorily, although the curve begins to rise at an earlier-than-expected age.

We had much greater difficulty in fitting the curve to 'Other' males. Here the PES gave a high rate for age 10-14 years, resulting in a very broad curve rising from a young age. This did not fit well with the model or Australian data. In the end we decided to fit the Rogers curve without using the 10-14 age group data to gain more demographically plausible adjustment factors. Our model fits for ethnic groups by sex are given in Figures 6 to 9.

Figure 6

Fitted Rogers Curve for Maori/Pacific Island Males

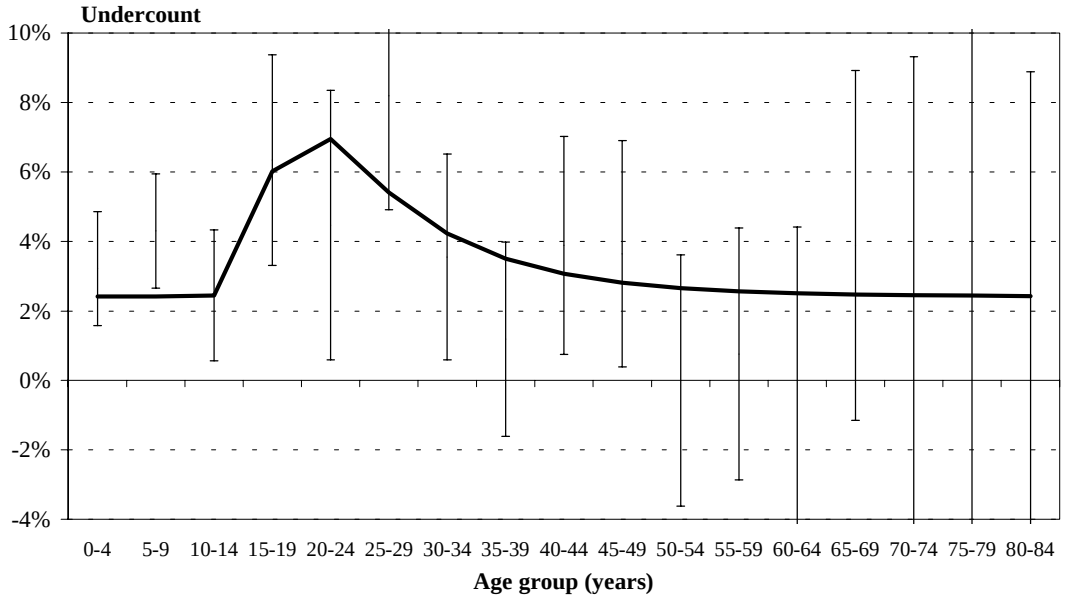


Figure 7

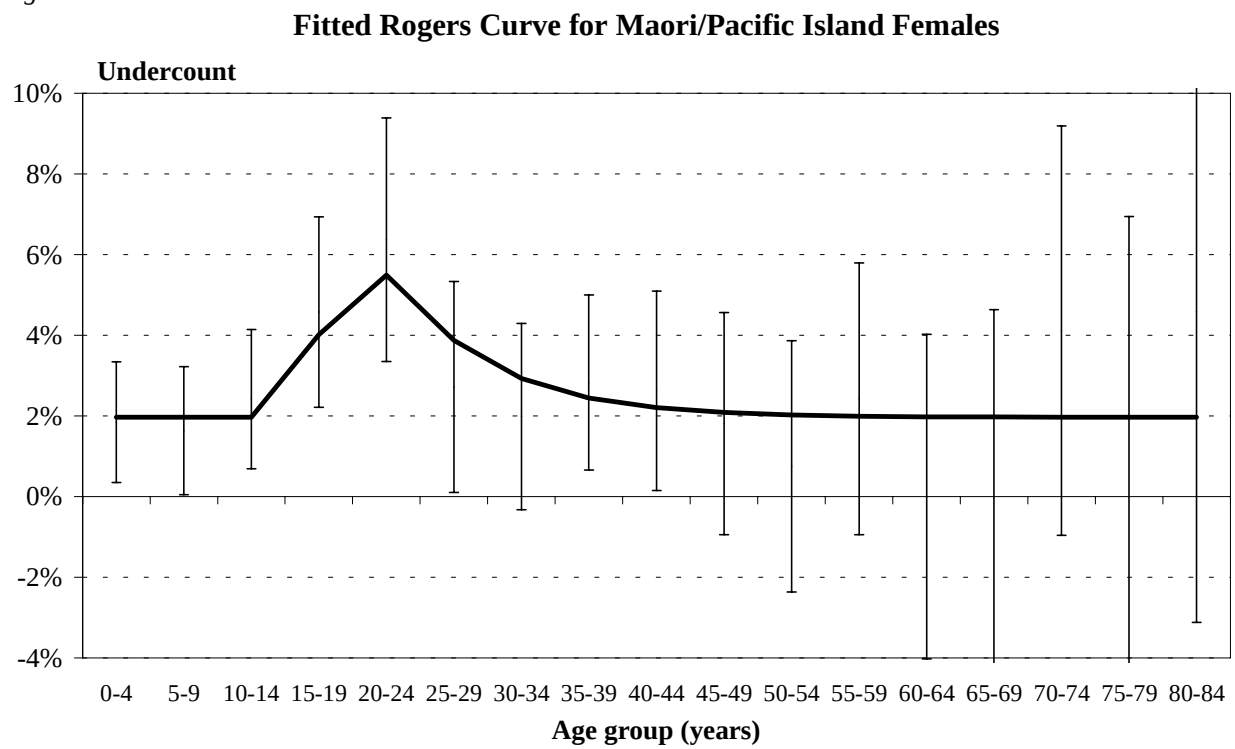


Figure 8

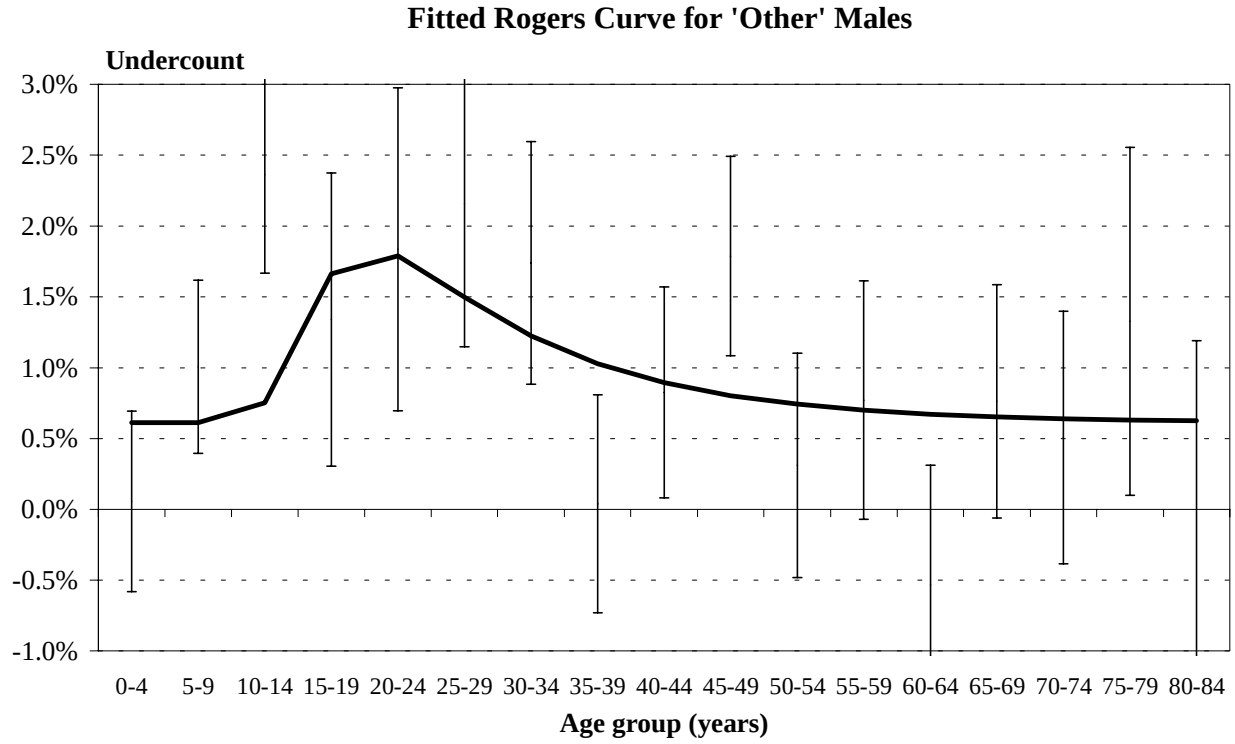


Figure 9

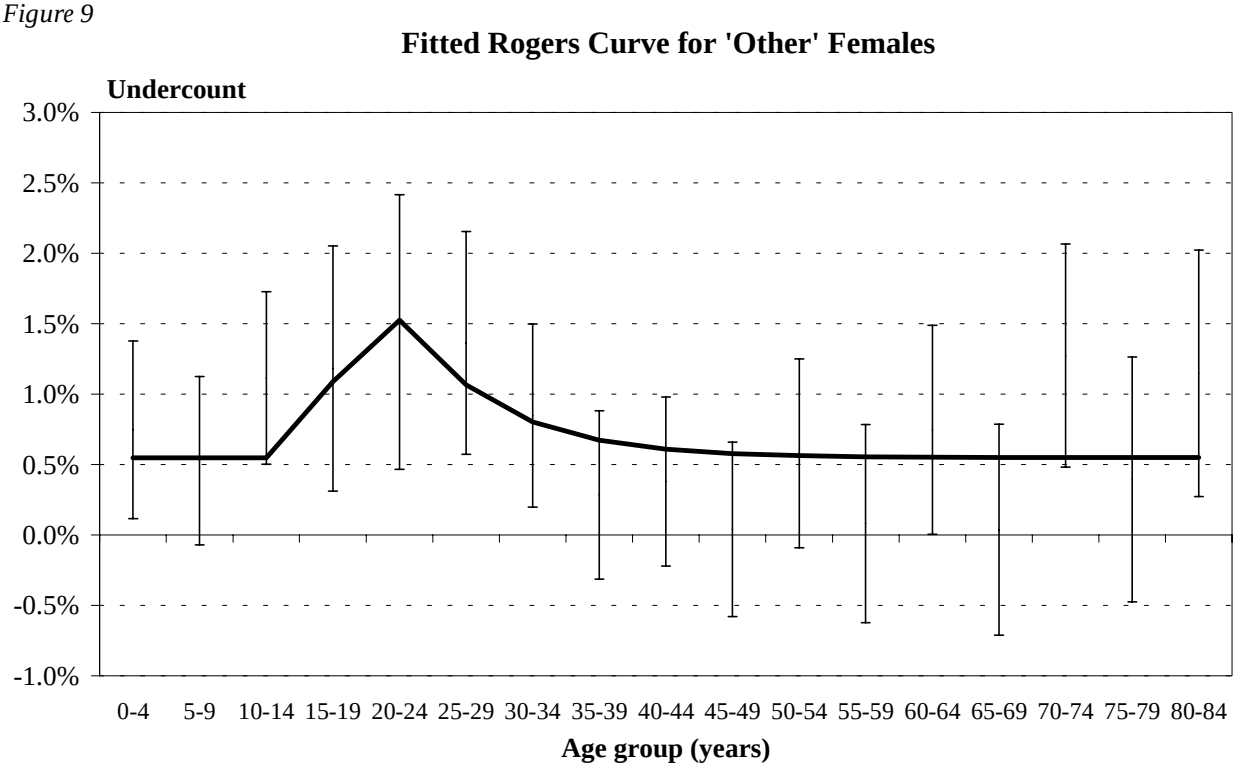
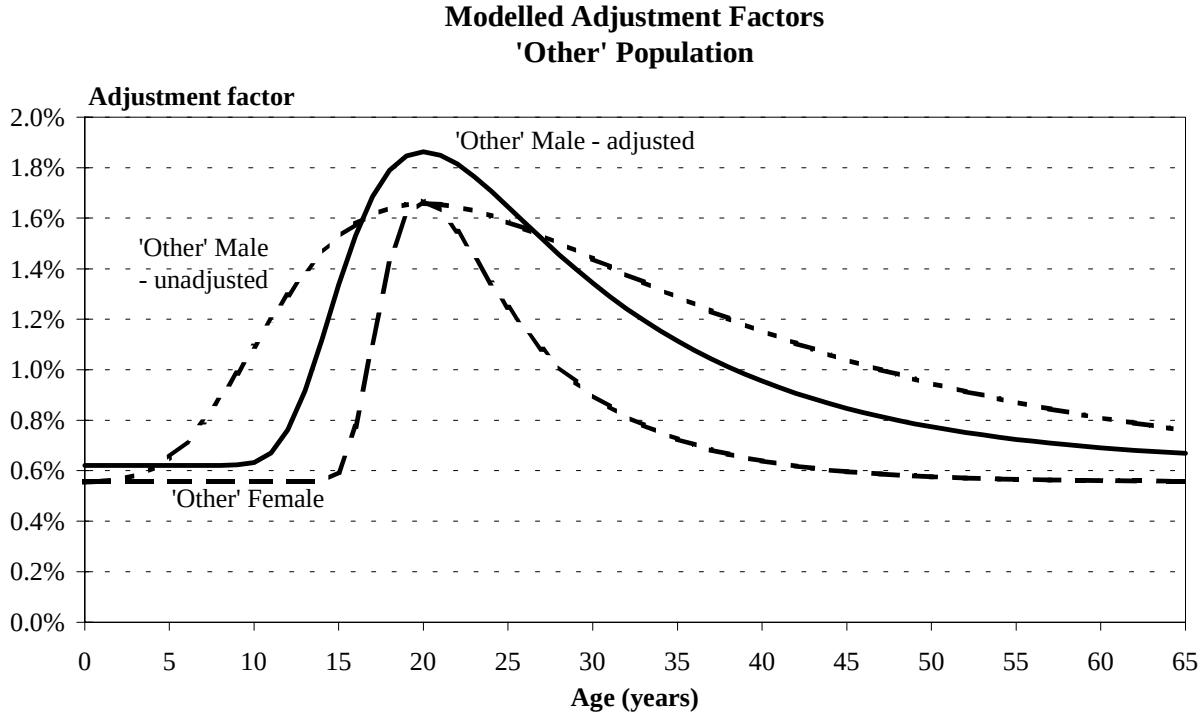


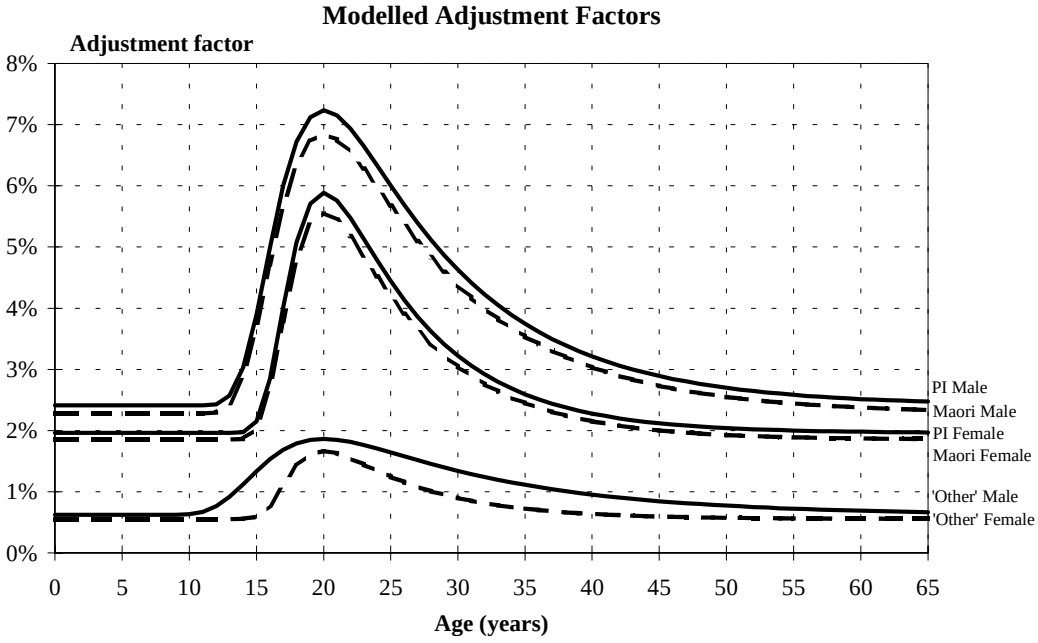
Figure 10 illustrates the modelled adjustment factors by single year of age for the 'Other' ethnic group. The broad flat shape of the 'Other Male - unadjusted' curve contrasts with the (unadjusted) curve for 'Other Female', as well as the curves for other ethnic groups (Figure 11). The 'Other Male - adjusted' curve was therefore proposed in preference to the 'Other Male - unadjusted' curve. The difference in overall national adjustment between the two 'Other Male' curves was about 1,500 people (or less than 100 persons in any single year of age). This is well within the bounds of the sampling errors.

Figure 10



Finally, minor scaling factors were applied to each of the curves so the overall adjustment factor for each major ethnic group matched that of the direct PES estimates: 2.9 for New Zealand Maori, 3.1 for Pacific Island and 0.8 for 'Other'. The final adjustment factors are given in Figure 11.

Figure 11



Appendix 5: Sensitivity of undercount rates to match classification

The objective of matching was to decide if the PES respondent was counted in the census at each address they stated that they had completed a census form, or at each address a census form may have being completed for them (search address). This was achieved by comparing the information given by PES respondents to the information given on census forms. Guidelines for matching were developed following consultation with statisticians from the Australian Bureau of Statistics.

Matching determines whether a person was counted or not counted in the census at each search address. The most important variables available for comparison were:

- name
- date of birth (or age)
- sex.

Other supplementary information used included:

- ethnic group
- usual resident or visitor
- household structure and relationships
- where the respondent completed a census form.

If the combination of variables indicated a person match, then the decision was that the person was counted. If the combination of variables did not indicate a person match, then the decision was that the person was not counted. In either case a match code was assigned as follows:

Match code	Personal match details	Match (M)/ Non-Match (NM)
1	Matched on name, age and sex.	M
2	Name matched, age or sex does not match, but the structure of the PES and census forms plus other variables indicate a match.	M
3	Name partially matched, age and sex match, and / or structure of the PES and census forms other variables indicate a match.	M
4	Name blank on one form or name not matched, age and sex match, and the structure of the PES and census forms plus other variables indicate a match.	M
5	Name only matched with all other variables missing, but structure of the PES and census forms indicate a match.	M
6	Name matched, but the structure of the PES and census forms plus other variables indicate a non-match.	NM
7	Name partially matched, and the structure of the PES and census forms plus other variables indicate a non-match.	NM
8	Name blank on one form or name not matched, and the structure of the PES and census forms plus other variables indicate a non-match.	NM
9	Name or household non-match.	NM
10	Insufficient address information to find dwelling to make a person match.	—

To test the sensitivity of the undercount rate to upward alterations of match codes (higher match codes meaning fewer matches), both codes 3 and 4 would have to be set to non-match before there was any significant impact on the undercount rates. Match code 4 are mainly matches against census dummies (about 500 matches) which are excluded from the undercount calculation because of the assumption that census dummies have full coverage. Match code 3, despite being only a partial name match is still a very strong match overall and cannot realistically be set to a non-match. Hence, the PES undercount rates are not sensitive to any upward alterations in match codes.

If altering the codes downward (lower match codes meaning more matches), codes 6 to 8 (containing less than 50 individuals) would need to change to match before there was any significant change in the undercount rate. Codes 6 and 7 are partial matches on name, usually the family name only, while other information strongly suggested a non-match.

The undercount rate from the 1996 Post Enumeration Survey is therefore deemed insensitive to any changes between match and non-match codes. Furthermore, the basic matching procedure assessed each form on the basis of “is there a match?”, before deciding what kind of match / non-match there was. Thus, every respondent was assessed individually on all available information and altering the cut-off between match and non-match is not applicable.