Creating a sense of 'closure': Providing confidence intervals on some recent estimates of Indigenous populations

B.H. Hunter and M.H. Dungey

No. 244/2003

ISSN 1036-1774

ISBN 0 7315 5619 4

Boyd Hunter and Mardi Dungey are Fellows at the Australian National University. Boyd works at the Centre for Aboriginal Economic Policy Research, while Mardi works at the Economics Division of the Research School of Pacific and Asian Studies.

Table of Contents

Tables

Acronyms and abbreviations

ABS	Australian Bureau of Statistics
ACE	Accuracy and Coverage Evaluation
AEC	Australian Electoral Commission
ANU	The Australian National University
ATSIC	Aboriginal and Torres Strait Islander Commission
CAEPR	Centre for Aboriginal Economic Policy Research
CD	Collection District
DSE	Dual System Estimator
PES	Post-Enumeration Strategy
QCPR	Queensland Centre for Population Research
SIF	Special Indigenous Form

Abstract

The 'error of closure' is the population growth that cannot be accounted for either by natural increase or by quantifiable non-demographic factors. The term is somewhat misleading: since it incorporates all unquantifiable components of the increase in a population count, it is unlikely ever to be 'closed'.

This study highlights the significance of variability of Indigenous population estimates by calculating standard errors, one of the conventional measures of reliability of statistics. That is, with tongue firmly in cheek, a sense of 'closure' is created in the debate by documenting the variability of estimates.

We introduce the Dual System Estimator method for estimating the Indigenous population, and review the international literature on its strengths and weaknesses. Once Australia's Indigenous population has been estimated using this method, confidence intervals are compared to those produced using the traditional undercount method. The main conclusion is that Dual System Estimates of the Indigenous population are reasonably accurate at the national level. Unfortunately, this conclusion may need to be revised when regionally disaggregated data are examined.

The central theme of this paper is that policy makers need to take into account the fact that Indigenous population statistics from the census are merely estimates. Given the importance of this data in the horizontal fiscal equalisation funding formula, it is particularly important to have accurate estimates for each of the States and Territories. Obviously, Dual System Estimates are not a panacea, but they are an alternative method of benchmarking the Indigenous population, and hence provide an appreciation of the reliability and bias of existing estimates.

Acknowledgments

Participants at a CAEPR seminar made useful comments on a verbal presentation of an earlier draft of this paper. Matthew Gray, Yohannes Kinfu and John Taylor provided useful comments on an intermediate written draft. Special thanks are due to Len Smith who provided generous feedback on the project from its inception. Frances Morphy and Michelle Cochrane assisted with editing and proofreading, and Wendy Forster with layout. To all, we are greatly indebted.

Introduction

One of the most contentious issues in regard to the demography of Indigenous Australians is the increasing extent to which people are likely to identify in the census and other official statistical collections as Indigenous (Gray 1997a, 1997b; Ross 1999; Taylor 1997a, 1997b). This phenomenon is not confined to Australia and there is a growing international literature addressing the problem (Sandefur, Rindfuss & Cohen 1996; Snipp 1997). The term coined to describe the extent to which population growth cannot be explained either by biological factors or by spatial mobility is the 'error of closure' (Passel 1996; Ross 1999; Taylor 1997a).

In the context of Indigenous populations, the error of closure methodology measures the difference between the expected population count and the actual population count, assuming that there is zero net immigration and that the only source of growth is natural increase. The methodology has been extended to include quantifiable non-demographic components such as changes in editing procedures (Ross 1999) and under-coverage of the population (Guimond 1999). That is, the error of closure is the population growth that cannot be accounted for either by natural increase or by quantifiable non-demographic factors. There are several possible sources of error in the error of closure calculation, for instance poor coverage of births or deaths, and differential undercount rates at each census. In addition, there are two factors that are normally cited as constituting the non-demographic increases. These are improvements in census methodology (Gray 1997a), and changes in self-identification as 'Indigenous' on the census form (Passel 1976; Ross 1999).

The term 'error of closure' is somewhat misleading. Given that it incorporates all unquantifiable components of the change in a population count, it is unlikely to ever be 'closed'. Unless variables were measured with absolute precision, and the underlying statistical model could perfectly predict the population count, it would be surprising if the error was ever equal to zero (i.e. to be 'closed'). Indeed, it would probably be an error to seek closure in the 'error of closure' because of residual randomness not accounted for in the model.

While most studies have concentrated on the point estimates of the Indigenous population in order to shed light on the potential for future increases in identification, this study seeks to highlight the role of variability of the estimates in such a process by calculating standard errors. That is, with tongue firmly in cheek, we create a sense of 'closure' in the debate by documenting the variability of estimates.

The debate over increasing Indigenous identification is at times emotive, especially given the tendency to conflate the numbers of people identifying as Indigenous with the definition of that identity (Gardiner & Bourke 2000; Gray 1997a). This paper does not seek to engage in such debates; rather it seeks to place bounds on the potential Indigenous population using a methodology that is somewhat under-utilised by Australian researchers—the Dual System Estimator

(DSE) method. In the context of human populations, the technique was pioneered more than 50 years ago by Sekar and Deming (1949).

After a brief introduction to the trends in the error of closure in estimates of the Indigenous population, the remainder of this paper reviews the Sekar–Deming method for estimating populations from survey data, and introduces the international literature on the strengths and weaknesses of calculating human populations using the method. Once Australia's Indigenous population has been estimated using this method, confidence intervals are compared to those produced using the traditional undercount method. The conclusion develops the policy consequences and puts forward suggestions for future research.

Revisiting the 'error of closure'

The error of closure is the remainder after measurable demographic and administrative factors are taken into account. The starting point is what demographers sometimes refer to as the 'balancing equation' (Shyrock, Siegel & Associates 1976: 4):

$$P_t = P_0 + B_t - D_t + NM_t + CP_t + E_t \quad (1)$$

Where P_t is population at time t; P_0 is the initial population estimate, B_t D_t and NM_t are the births, deaths and net migration since the initial period, and CP_t is the population changes resulting from census procedures. The residual in this equation is denoted E_t, and forms the basis of the error of closure. In the case where P_t refers to the Indigenous or non-Indigenous population only, E_t includes the effect of the changing propensity to identify as Indigenous.[1]

The term 'balancing equation' is rather misleading because it carries the connotation of an accounting framework. However, the estimates are only balanced by adding an error term. Furthermore, since all of the terms on the right-hand side are estimates, and hence are measured with error, all these errors are reflected in the residual of equation 1 (above). Hence the 'error of closure' is an amalgam of errors. One should be surprised, and even suspicious if it were ever closed!

Table 1 presents the Indigenous census counts collected since national statistics were first provided in 1901. The first thing to note is the variability of the annual population growth rate (column 4). Taylor (2002a) claims that these counts partially reflect highly variable methodology and the intent of the governments of the day. This notwithstanding, the annual growth rate of the Indigenous population was relatively low until 1971. In a period of significant growth for the rest of the population, it is particularly noteworthy that the Indigenous population often appeared to have negative growth rates. It is probably not a coincidence that 1971 was the first census year after the 1967 referendum. Arguably, non-Indigenous Australians displayed considerable public goodwill towards Indigenous people when they formally recognised, with an overwhelming endorsement of the referendum proposals, the need for a national policy

addressing Indigenous issues. It seems plausible, then, that a substantial number of Indigenous people were choosing not to identify in the official statistical collections before the 1967. Since these missing generations will themselves have had progeny, there may be a considerable potential population who may not have identified as Indigenous in the 1960s or even the 1970s (Taylor 2002b). Note that the low growth rates between the 1976 and 1981 censuses probably reflect the idiosyncrasies of those censuses rather than a genuine trend in the Indigenous population (see Choi & Gray 1985).[2]

Table 1. Indigenous census counts, 1971–2001

Census year	Count	Intercensal period	Annual growth rate (%)
1901	93,333	1901–1911	-1.0
1911	84,527	1911–1921	-1.6
1921	71,836	1921–1933	1.0
1933	80,721	1933–1947	-0.4
1947	75,965	1947–1954	-0.2
1954	75,040	1954–1961	1.7
1961	84,470	1961–1966	-1.0
1966	80,207	1966–1971	7.6
1971	115,953	1971–1976	6.8
1976	160,915	1976–1981	-0.1
1981	159,897	1981–1986	7.3
1986	227,593	1986–1991	3.1
1991	265,371	1991–1996	5.9
1996	352,970	1996–2001	3.0
2001[a]	410,003		

Note: a) The 2001 estimates are based on Kinfu and Taylor (2002) and authors' calculations.
Source: Ross (1999: 10).

Similar observations of low growth rates in the Indigenous population can be made for the USA in the first half of the twentieth century. This also resulted in a 'pent-up' population pressure from a potential Indigenous population with the numbers of American Indians increasing dramatically after 1950 (see Ross 2002).

While the error of closure has only been estimated for the last few censuses, it appears to be generally positive and somewhat variable. Combining the effects of births, deaths, census editing and migration, the error of closure for Australia during the 1991–96 period is 44,356 persons, or 12.6 per cent of the total Indigenous count in 1996 (Ross 1999). In comparison, the error of closure for the American Indian population, was 9.2 per cent for the most recent intercensal period, 1980–90 (Harris 1994), and 25.2 per cent for the 1970–80 period (Passell & Berman 1986).

The largest proportional increases in recent Australian census counts, and associated errors of closure, were in the highly urbanised south-eastern states. These areas experienced early European settlement and their rates of intermarriage resulted in a larger pool of persons of mixed ancestry than in areas where European settlement occurred later. It is this mixed ancestry group for whom it is likely that identification as Indigenous has changed over time (Ross 1999).

The primary force behind the unexplained component of the increase in Indigenous population appears to be related to what has been termed by Passel and Berman as 'recruitment' or changes in self-identification (1986: 164). Ross (1999) claims that the changes in the transmission of Indigenous identity from parents to children in mixed couple families is evidence that 'recruitment' into the Indigenous population is not only possible but did occur between 1991 and 1996. Data on age structures in Ross (1999) and elsewhere (Gray 1997a, 1997b; Hunter 1998b) suggest that such 'recruitment', if it occurred, must have taken place in adult age groups as well.[3]

The proportion of Indigenous population growth between 1991 and 1996 that could not be accounted for by demographic factors was just over half (51%) of the overall intercensal increase (Ross 1999: 24). For the most recent period, the Australian Bureau of Statistics (ABS) has estimated a reduction in this error of closure to 26 per cent (ABS 2002). This still leaves a sizeable error of closure, although the ABS estimate is preliminary (Kinfu & Taylor 2002). For example, no account is taken of change to census edits for Indigenous status in the ABS calculation, and incomplete vital registration data are used for births and deaths. By correcting for census edits, and by applying adjusted estimates of births and deaths, Kinfu and Taylor provide an alternative set of estimates.[4]

According to Kinfu and Taylor, 69 per cent of the increase in the census count of Indigenous Australians can be explained in terms of natural increase. The unexplained component, which amounts to 19,243 persons, accounts for 31 per cent of the intercensal population change and represents about 4.7 per cent of the total Indigenous count in 2001. This improved estimate of the error of closure is still substantially lower than that observed for the 1991–96 period (Gray 1997a; Ross 1999).

Gray and Tesfaghiorghis (1993) reported that the unexplained component of growth in the Indigenous population was relatively small for the 1986–91 period. That is, there is some evidence that the error of closure for the Indigenous population varies significantly between successive censuses. As noted earlier, a substantial number of people may have hidden their identity before 1967, and the progeny of such people may reveal their heritage in an uneven manner. Such 'revelation' may in turn be influenced by the political climate and the extent to which other relatives openly identify as Indigenous. As noted earlier, a similar observation can be made for US estimates of error of closure, which experience substantial variations over time.

Hunter (1998b) observes that the potential Indigenous population is probably more stable than these apparent changes indicate, and that its pattern of growth is not totally out of kilter with biological and demographic factors. However, this assertion requires that a valid method be identified to estimate the potential population. Fortunately, a relevant methodology has been established, initially in the biological sciences. In the context of human populations, it is often referred to as the Dual System Estimator (DSE), or sometimes 'dual survey estimators' or 'dual record systems'. The basic approach is to estimate the number of people missing from any particular enumeration using a follow-up survey. Such a survey is undertaken after major censuses in most developed countries; in Australia it is known as the Post-Enumeration Survey (PES).[5]

A brief history of the methods for estimating potential human populations

The statistical methods underpinning DSEs have a long history, having first been applied in the study of wildlife populations before being adapted for other purposes (Le Cren 1965). In ecology, the method is generally called the Petersen method or 'capture–recapture' technique because of Petersen's work associated with fish populations, in 1894.[6]

The simplest 'capture–recapture' model is the so-called two-sample model, used solely to estimate the size of a population of unknown size. The first sample provides the individuals for marking or tagging and is returned to the population, while the second sample provides the recaptures. Using the numbers of individuals caught in both samples (the recaptures) and the numbers caught in just one sample, it is possible to estimate the number not caught in either sample, thus providing an estimate of the total population size. The assumptions required for such estimate to be valid are that:

1. there is no change to the population during the investigation (i.e. the population is closed);
2. individuals can be matched from capture to recapture;
3. for each sample, each individual has the same chance of being in the sample (i.e. the population is 'homogenous'); and
4. the two samples are independent.

The application of these methods to the study of epidemiological problems came relatively late in this history, and thus has been able to draw on advances in the other areas as well as in statistical methods more broadly. Sekar and Deming were the first to adapt the method for human populations when they used it to estimate birth and death rates, and the extent of their registration in 1949, with hospital data. Using a similar approach, Shapiro (1949) applied the technique to birth registration in the USA using census data. There is also a substantial literature going back to the 1940s, under the general heading of DSE, dealing with the application of the two-sample method to census data (Fienberg 1992). By taking another sample in addition to the census, the method can be used for

estimating undercount by the census. The history of the application of DSEs to the US census are described by Hogan (1993).

In terms of the validity of assumptions for estimating the potential numbers of Indigenous Australians, it is necessary to confine our attention to closed populations. Even populations with high mobility, such as people in remote Indigenous communities, may be considered 'closed' so long as the PES or follow-up survey takes place shortly after the initial survey or census (Paradies et al. 2000).

With respect to assumption (2), matching will depend on the quality of the census records and the uniqueness of respondents' names. The ABS (1997: 7) gives a detailed assurance that all due care is taken to match and search the respective census and PES responses. In the PES, respondents are asked a number of questions about where they live, and addresses where they might have been counted in the census, including:

- if they thought they were counted on a census form (and if so where); and
- where they were on census night (ABS 1997: 25).

Another of the assumptions required for DSEs to be valid is the homogeneity of the population (assumption (3) above). That is, all the members in the population should have the same chance of being sampled in the follow-up survey. However, this assumption could be violated if, for example, variations in socioeconomic status within the population mean that members of the population may actually differ considerably in the chance of being ascertained by a particular source. Such heterogeneity, within the population, may cause problems in DSE analyses. While data that relies on voluntary reporting should be used cautiously, it should be noted that it is compulsory to answer ABS surveys under the *Census Act 1905* (Cwlth). Also note that survey weights are constructed to allow ABS to equate probability of inclusion in a sample—this reinforces our confidence in the validity of assumption (3) in the current context.

The question of independence is discussed by Sekar and Deming (1949) in some detail. While the assumption of independence may not be valid in many circumstances, the ABS applies procedures to preserve the independence of the census and PES.[7] For example, given that some people may have been prompted to return their census forms by the arrival of the PES preliminary letter or the interviewer (ABS 1997: 8), the follow-up survey is held three weeks after the official date of the census to ensure that all forms have been completed and returned.[8]

Clearly, the DSE method is easily adapted to a situation where there is incomplete information on Indigenous status (see Appendix A). The use of dual surveys in examining relatively rare populations is not unusual with applications extending well beyond the early biological applications (see Levy & Lemeshow 1991). As indicated, in many countries, including Australia, the secondary survey exists in the form of the PES.[9]

Recent controversy in the USA

The DSE methodology seems relatively non-controversial, and has been used routinely to estimate populations in the USA for several decades. However, the issue of whether to adjust the census has recently become so contentious that, beginning in 1990, it produced a number of lawsuits against the Census Bureau, some of which ended up before the US Supreme Court in 2001. More recently, its use in adjusting local population measures in the USA in the 2000 Census provoked heated debate. US Census Bureau officials recommended against adjusting the 2000 Census to compensate for people who were missed in local areas—dashing the hopes of Democrats and civil rights leaders that such an adjustment could be used in redrawing political boundaries. Since the Republicans controlled the House of Representatives by only a few seats, both sides have pushed hard for their preferred methodology. The media reports at the time reflected the nature of the debate ('Republicans gleeful' and 'Game. Set. Match'—*Washington Post*, 1 March 2001).

Ultimately, US Census officials could not guarantee that adjusted census numbers would be more accurate than results from mail-in questionnaires and a door-to-door count last year. Note that the DSE methodology is still considered valid for the nationally based estimate of the population, it was just local adjustments that were seen as too unreliable.

The decision not to adjust was based on an analysis of a 314,000-household survey, the Accuracy and Coverage Evaluation (ACE). Like its Australian counterpart (the PES), the ACE provides a quality check on the head count in the census.[10] The evaluation of the ACE concluded that the 2000 Census had done a better job than in the one in 1990, though many people had probably been left out (Robinson 2001).[11] The Census bureau decided not to adjust re-districting data because of timing issues imposed by legislation for drawing political boundaries, and possible issues with ACE, especially arising from problems with duplicates. However, the issues for the ACE identified in Robinson (2001) were neither significant nor consequential for the overall applicability of DSEs (e.g. issues arising from matching procedure, independence, classifying movers and imputation of missing data).

Some analysts remain critical of any revised estimates of the US census, especially in the context of changing ethnic identification (Skerry 2000). However, their analysis tends to yield insights into the difficulties in describing the (largely psychological) processes underlying ethnic identification rather than to undermine the validity of the DSE methodology.

The obviously political dimension to the US debate means that it has limited implications for analysis of the Australian case. It should be re-emphasised that Australian undercount estimates are based on a different methodology to that used in the USA (and Canada). The North American methodology relies heavily on DSEs that evolved from the Sekar–Deming methodology (Alho et al. 1993; Darroch et al. 1993; Hogan 1993). In contrast, the Australian undercount is a

more mechanical adjustment of census counts (see ABS 1997: 8). However, the ABS devote considerable resources to identifying duplicates, an area identified as a weakness in the recent review of the US ACE (Robinson 2001).

After the recent controversy, the US Census Bureau revised the methodology to account for the duplicate issue. While the undercounts for Black and Hispanic Americans fell significantly, there was not much change for the American Indian population (see Table 2). However, the reliability of estimates declined slightly (i.e. the standard errors increased). Notwithstanding, the standard errors are reasonably similar to those estimated for the Australian undercount.

Table 2. US undercounts for 2000 census revisited

	ACE (March 2001)		Revised Preliminary	
Race/Ethnicity	Estimate	Standard Error	Estimate	Standard Error
American Indian and Alaskan Native	3.66	1.03	3.44	1.60
Black or African American	2.17	0.35	0.78	0.45
Hispanic	2.86	0.38	1.25	0.54
Total	1.18	0.13	0.06	0.18

Source: Robinson (2001).

In the rest of this paper, DSEs are used to estimate the Australian Indigenous population since 1991, and these estimates are benchmarked against the standard Australian undercount measures of the population.

Responses to the questions on Indigenous status for Australian PES, 1991–2001

DSEs require that an individual's response to the Indigenous status question varies between a census and follow-up survey. This section documents the responses to the Indigenous question in the censuses that occurred between 1991 and 2001. Table 3 re-presents the 1991 data required to estimate the potential Indigenous population, with all numbers being expressed as a percentage of the total number who respond to the PES and provide a valid response to the question on Indigenous status. The first thing to note is that only 1.4 per cent indicated they were Indigenous in both the survey and the 1991 Census. Equal numbers of respondents (0.3%) changed from Indigenous to Non-Indigenous between census and PES, and vice versa.

Table 4 shows that there was less variation between the census and PES in 1996. While there was a slight increase in the percentage who indicate that they were Indigenous in both, there was a slight reduction in the number who changed from being non-Indigenous in the census to Indigenous in the PES.

Table 3. Percentage of PES respondents changing their answer to the census question on Indigenous status, 1991

	PES response		
1991 Census response	Non-Indigenous	Indigenous	Total
Non-Indigenous	98.1	0.3	98.3
Indigenous	0.3	1.4	1.7
Total	98.3	1.7	100.0
Numbers	64,080	1,078	65,158

Notes: The 'not stated' category is distributed proportionately across the responses to the Indigenous question. Column and row totals may not add up due to rounding error.
Source: Hunter (1998b).

One particularly interesting feature is that the percentage of people who identify as non-Indigenous in both the PES and the census is identical in both 1991 and 1996 (98.1%). Inverting this, the percentages of people who identify as Indigenous in some way were identical in both of these years. Thus, the proportion and hence number of people who identify as Indigenous appears to be more stable than the prima facie evidence indicates. Arguably, the propensity to identify as Indigenous also became more reliable, with fewer people changing their response to the Indigenous status question. However, while there was some minor variation in changes in response, the results are broadly similar between 1991 and 1996.

Table 4. Percentage of PES respondents changing their answer to the census question on Indigenous status, 1996

	PES response		
1996 Census response	Non-Indigenous	Indigenous	Total
Non-Indigenous	98.1	0.1	98.2
Indigenous	0.3	1.5	1.8
Total	98.3	1.7	100.0
Numbers	79,535	1,336	80,871

Notes: The 'not stated' category is distributed proportionately across the responses to the Indigenous question. Column and row totals may not add up due to rounding error.
Source: Hunter (1998b).

The 2001 Census saw a small, but potentially important, fall in the proportion of people who identified as non-Indigenous in both the census and the PES (from 98.1% to 97.3% in Table 5). The reliability of the 2001 estimates is likely to have been diminished by the reduction in size of the PES survey in that year. However, the fall in non-Indigenous identification is consistent with there being some 'recruitment' into the potential Indigenous population. Given the prolonged period of relatively low population growth, it would not be surprising if this were the

case. Other possible explanations lie in the higher fertility rates in the Indigenous population or changes in the coverage of the respective censuses.

Table 5. Percentage of PES respondents changing their answer to the census question on Indigenous status, 2001

	PES response		
2001 Census response	Non-Indigenous	Indigenous	Total
Non-Indigenous	97.3	0.2	97.5
Indigenous	0.4	2.1	2.5
Total	97.7	2.3	100
Numbers	55,662	1,316	56,978

Notes: The 'not stated' category is distributed proportionately across the responses to the Indigenous question. Column and row totals may not add up due to rounding error.
Source: ABS (2002: 20).

The fall in the percentage of the PES that is 'always' non-Indigenous coincides with an increase in the percentage who are 'always' Indigenous in both 1996 and 2001. While there was an increase in the percentage who changed their non-Indigenous status in the census to Indigenous in the last PES, the overall percentage changing from Indigenous to non-Indigenous (and vice versa) was the same in the 1991 and 2001 collections.

A digression on who identifies as Indigenous, 1986–2001

The focus on the large changes in the Indigenous population is sometimes confused with the question of compositional change. As a result some commentators imply that the newly identified Indigenous people might be less likely to experience disadvantage (see Gray & Auld 2000).[12] Another possibility sometimes canvassed when there is large increase in Indigenous counts in the census is that some non-Indigenous people are deliberately providing misleading information for the Indigenous question—this 'bogus' identification, in turn, would induce distortions into measured Indigenous statistics. While it is not possible to track individuals across time using census data, it is possible to track cohorts. Amongst other things, this section of the paper tracks the Indigenous cohorts across time to provide a 'natural experiment' for testing the hypothesis of compositionally equivalent cohorts in successive censuses.

Before moving to the cohort analysis, it is worth revisiting who changes identification as Indigenous between the census and the PES. Hunter (1998b) shows that there is little change in the age profile between 1991 and 1996 for either the 'marginal' (i.e. those who change Indigenous status between the census and the accompanying PES) or more 'consistent' Indigenous population. That is, the age profile of the marginal Indigenous population is still much closer to that of the 'consistent' Indigenous population than it is to the non-Indigenous profile.

This result is consistent with the findings of Gray (1997a: 14–16) and Kinfu and Taylor (2002), using aggregate data, that the 'new' Aborigines had approximately the same age distribution as the 'old' Aborigines (i.e. concordant with the higher adult mortality and fertility rates characteristic of Indigenous populations).

Unfortunately, it will not be possible to replicate the Hunter (1998b) analysis for the 2001 Census data until the PES data is publicly available later in 2003. However, given the relatively small increase in the potential Indigenous population it is unlikely that these data would lead to a different analysis.

Hunter (1998b) also assesses the effect of changing identification between the last few censuses by tracing the changes in the characteristics of single-year cohorts of individuals between 1986 and 1996.[13] If the composition of the cohort changes over time, then this will manifest itself as apparent change in the characteristics of the cohort. Fixed effect regression models of the 'age left school' variables were estimated on a 'synthetic' panel of Indigenous and non-Indigenous cohorts between 1986 and 1996. The null hypothesis of compositional change was rejected for all cohort groups. That is, the characteristics of the single-year cohorts did not change significantly over time.

Unfortunately, the 'age left school' variable was left out of the 2001 Census in favour of a new variable, 'highest level of schooling completed', which exploited the recent efforts to coordinate the secondary school curricula across the various State and Territories. This change means that it is not possible to directly replicate the cohort analysis in Hunter (1998b). Nevertheless, it may be possible to exploit 'intuitive' correspondences between the completion of school to various years and the age left school. Table 6 assumes that a person who left school before age 14 either had no schooling or only attended school until 'year 8' (or lower). The sense of the following analysis is not altered significantly by small variations in this assumption.

Table 6 confirms that there was no evidence of compositional change in the Indigenous (or in the non-Indigenous) population between 1986 and 1996. However, the proportion of the Indigenous cohort aged 20 and over in 1986 was around 10 percentage points more likely to leave school early in 2001 than in 1996. The fixed effect regression analysis of the single year cohorts, analogous to that conducted in Hunter (1998b), reveals that there was a significant change in these schooling variables in the last intercensal period for the Indigenous cohorts, but not for the non-Indigenous population. That is, if compositional issues are analysed using both age left school and highest level of schooling, there are no significant differences between non-Indigenous cohorts, but Indigenous cohorts were different in 2001, compared to 1996.

The Indigenous population thus appeared to become more *unlike* the rest of the population, since the average level of Indigenous educational attainment fell from its already low level. How do we interpret this result? If it reflected true compositional change in the Indigenous population, then one would expect the opposite result. The most likely explanation is that this result is driven by changes in the information collected that differentially affected cohorts living in

different States. Since Indigenous people are disproportionately concentrated in particular States, it would not be surprising if Indigenous cohorts were affected differently by the revisions to the ABS classification system.

Table 6. Per cent who left school before age 14, persons aged 20 and over, 1986–2001

	1986	1991	1996	2001
Indigenous females	42.2	41.1	40.6	51.7
Non-Indigenous females	24.2	22.7	23.3	20.7
Indigenous males	47.1	46.2	45.8	54.3
Non-Indigenous males	24.7	23.5	24.1	20.4

Note: The change in the code of the 'age left school' variable followed the revisions to ABS Classification of Qualifications, which lead to the Australian Standard Classification of Education being used for the 2001 Census. The assumptions required to compare the results of previous censuses with 2001 are described in the text. A sensitivity analysis of this assumption was conducted but the sense of the result did not change.
Source: ABS Census 1986, 1991, 1996, 2001.

Regrettably, Australian census data do not provide any other time-invariant characteristics that can be readily used to test the hypothesis of compositional change among Indigenous cohorts. But at the very least it can be said that the census data on school completion does not support the hypothesis of a convergence in the characteristics of the Indigenous and non-Indigenous populations in the last four censuses. That is, on the balance of probabilities, it is possible to discount 'bogus' identification as a factor explaining the growth of the Indigenous population over the last four censuses.

An unfortunate implication of this analysis is that it will not be possible to conduct any synthetic cohort analysis for the Indigenous population using 2001 data on highest level of schooling completed. For example it will not be possible to replicate Gray and Hunter's (2002) 'longitudinal' analysis of Indigenous labour force status because the crucial education variables are not comparable between the 2001 and previous censuses.

This compositional analysis provides indirect evidence that the people who now identify as Indigenous are similar to those who previously identified. That is, they are drawn from a similar population. The next section attempts to identify the size of this potential population using a DSE.

DSE estimates of the Indigenous population

The mathematical formula provided in Appendix A can be used to derive new population estimates for Indigenous Australians from the data presented in Tables 3–5. However, the PES does not cover the whole of Australia, and it is

therefore necessary to make some distributional assumptions about the Indigenous population in areas outside the scope of the PES.

The PES is not conducted in remote areas (defined as being where there were fewer than 0.57 dwellings per km^2) or in discrete Indigenous communities. Because of this it is not possible to analyse directly the changes in identification of remote residents. However, only 0.5 per cent of the population enumerated on the remote area Special Indigenous Forms (SIFs) in 1991 were non-Indigenous (Evans, Kahles & Bate 1993: 25), so changes in identification are unlikely to be an important factor in such areas. Also, the extent of a 'DSE-style' undercount of the Indigenous population is constrained in the sense there are relatively few non-Indigenous people to change their response to the Indigenous status question.[14] This assumption can also be justified on the grounds that remote areas had a very low error of closure between 1991 and 1996 Ross (1999).

The population in these remote areas was calculated using the 1991 and 1996 counts in Collection Districts (CDs) which did *not* use the SIF. The populations for each census category for Indigenous status were allocated to a cell on the basis of the proportion of each census sub-group who claimed they were Indigenous or non-Indigenous in the PES. This approach was adopted to account for the differential rates of Indigenous identification in PES and non-PES areas. That is, for simplicity's sake, and in the absence of any better information, the rate of Indigenous identification in non-PES areas for the census was taken as given.[15] This assumption will obviously not be valid in all areas, and can be relaxed later.[16]

Table 7 provides two estimates for 2001 because the ABS has not yet released any data on population in PES areas to date. Consequently the 2001 estimates are bounded by two extreme assumptions: firstly, that the population in non-PES areas did not change and, secondly, that all population growth in remote areas occurred in non-PES areas. Note that there is little difference in the DSE estimates for 2001, irrespective of the assumption used.

When using this method, the estimate of the complete Indigenous population increases from 304,751 to 372,983 between 1991 and 1996. The 2001 estimates are both around 446,000. The 95 per cent confidence intervals of the estimates are a band of between 3,500 and 5,000 around these estimates.

In addition to the point estimates and the 95 per cent confidence intervals, Table 7 also presents the growth rate of the potential population, which is more constant than the gross rates of growth reported in Table 1. That is, the potential Indigenous population grew by around 20 per cent in the two most recent intercensal periods. The growth rate declined slightly in the last intercensal period, probably because of the apparent decline in Indigenous fertility rates (Kinfu & Taylor 2002).

Table 7. DSE estimates of potential Indigenous population, 1991–2001

	1991	1996	2001a	2001b
Indigenous population	304,751	372,983	445,714	446,902
95% confidence intervals (based on $\hat{N}$ ± 1.96*standard errors)				
Upper	307,442	374,703	448,377	449,654
Lower	302,060	371,262	443,051	444,150

	1991–96	1996–2001a	1996–2001b
Intercensal growth	22.4%	19.5%	19.8%

Note: The act of distributing the 'not stated' category for the question on Indigenous status makes virtually no difference to the estimates reported in this table.

As indicated above, this methodology treats the Indigenous population of non-PES areas as being fixed (i.e. having a standard error of zero). This assumption can be relaxed, but the sensitivity analysis is not reported in any detail in order to simplify exposition. Suffice to say that the confidence interval would only increase by between 1,000 and 2,000 if the Indigenous population were scaled to the total population counts rather than only to the counts for PES areas (and then adding the Indigenous population for non-PES areas). That is, confidence intervals only increase marginally if the Indigenous population estimates are assumed to have the same variance in PES and non-PES areas.

A side effect of scaling the PES results to the total Australian population is that it imposes an Indigenous undercount result in non-PES areas, and hence increases the estimate by about 17,000. In 1996, this increases the size of the estimate to a little over the experimental estimate in ABS (1998). This effect is undesirable if one accepts the argument that there is a limited scope for an undercount in non-PES areas. However, since it is not possible to discount an undercount in such areas, one approach might be to extend the confidence intervals to take in this possibility. While this is somewhat arbitrary, it should be noted that the DSE estimates of the confidence intervals are exceptionally tight, probably as a result of both the treatment of non-PES areas and the simplifying assumption of normality (of the statistical distribution) that has to be made when using grouped data from the PES. We will return to these points in the concluding discussion.

Having identified several weaknesses in the DSE estimates of the Indigenous population, it is important to benchmark them against alternative estimates. The most obvious alternative is the standard ABS undercount estimates of the Indigenous population that permit the analyst to construct confidence intervals around the population.

Estimates of the population of Indigenous Australians using the ABS undercount

The ABS undercount refers to the number of people missed in the census for one of the following reasons: they were difficult to contact (e.g. travelling); they mistakenly thought they were counted elsewhere; there was insufficient space on the census form; the person completing the census form assumed that babies, the elderly and visitors should not be included; respondents were reluctant to participate (e.g. confidentiality concerns); or the dwelling was mistakenly classified as unoccupied. Sander's chapter in Martin et al. (2002) identified that certain census collectors simply overlooked the existence of a number of individuals in the town camps of Alice Springs. An 'overcount' refers to where people are counted too many times (e.g. duplicates). The net undercount is the difference between gross undercount and overcount.

One relatively intuitive means of estimating variability of the Indigenous population is to use the standard errors on the census undercount (ABS 1997). The ABS define the undercount rate as:

$$\text{final undercount rate} = \frac{\text{final population estimate - census count}}{\text{final population estimate}} \quad (1)$$

We have data available on the final population estimate and the initial census count. However, the ABS also makes a number of adjustments to this data to account for demographic irregularities and inconsistencies with alternative, often administrative, data sources. Hence the undercount rate calculated from the published ABS data on final population and census data does not produce the official ABS undercount rates. The final population estimate for Indigenous people in 1996 was 386,049 (ABS 1998: Table 7, 19).[17]

Appendix B shows how the undercount for the respective censuses can be used in conjunction with the census counts from Table 1 to produce estimates of the Indigenous population. For the 1996 Census we have an Indigenous persons undercount rate of 7.1 per cent with a standard error of 1.04 per cent. For 1996 data, the population is estimated using this technique to be around 379,950. Given the emphasis of this paper on the reliability of estimates, it is important to note that we can be 95 per cent confident that the 'true' Indigenous population lies somewhere between 371,800 and 388,500. That is, there is a band of approximately: 8,000 (or about 2.1% of the total Indigenous population) on either side of the point estimate.

The 1991 undercount for the Indigenous population was 3.6 per cent (Evans, Kahles & Bate 1993). The standard error for the undercount is remarkably high, at 7.2 per cent. Given the unadjusted population count reported in Table 1, the above technique gives a 95 per cent confidence interval that the true population lies between 240,155 and 322,443. Part of this range is less than the count in Table 1, implying that there may have been an overcount. However, this is more likely to reflect sampling issues in the PES. In view of this large confidence

interval, it is not surprising that the probable array of DSE estimates for 1991 lies in this range.

Note that the results reported in this section use the initial census counts to estimate the confidence intervals for the Indigenous population as the DSE estimates in the previous section are also scaled by these counts. This assumption also avoids any circularity in the logic between the ABS's demographic adjustments and the use of an adjusted census count. For the purposes of this paper, the difference between using the initial or adjusted census is not important because the following analyses are produced solely for the purposes of benchmarking the DSE estimates and associated standard errors.

The bottom line is that the DSE estimate is almost entirely within the confidence interval estimated using undercount data, even if it is a little on the low side in 1996. This relatively small confidence interval might be expected given that DSEs do not take into account other (demographic) information included in the ABS estimates. However, it should also be possible for the ABS to make similar augmentations to the DSE estimates that facilitate more direct comparisons with other estimates of the Indigenous population. The important thing is to get a sense of the variability of population estimates so policy makers can identify the potential sensitivity of the assumptions underlying their policy prescriptions.

Policy and methodological implications

Accurate Indigenous population estimates are essential for ensuring policy is effective and for estimating service delivery requirements (Taylor & Hunter 1998). This paper illustrates that the population estimates provided by the ABS are somewhat variable, but reasonably accurate at the national level over the last two censuses.

This paper has aimed, nevertheless, to illustrate the importance of assessing the reliability of Indigenous population estimates through providing standard errors. While confidence intervals have been provided on our estimates, it is important to recognise their limitations. In addition to ignoring other demographic information, no attempt has been made to take into account sampling error and survey design effects arising from the PES. Only the ABS is in a position to provide such information. They could do this by either using a DSE (and related standard errors) to estimate the Indigenous population themselves, or by being more transparent in their undercount publications in ways that permit an estimate of sampling error and design effects in the PES. Not only could the ABS develop an alternative set of population estimates for Indigenous Australians, but this may also be a worthwhile exercise for many sub-populations of interest to policy makers (e.g. Australians from a non-English speaking background).

By providing more information about the distribution of estimates from the PES, the ABS will facilitate sensitivity analysis of the assumption of normality in DSE, which we were forced to impose. For example, given the small number identifying as Indigenous in the PES, the distribution of the estimates of people who change

their response to the Indigenous status question is probably closer to a Poisson distribution than the Normal distribution.[18] The distinct properties of the Poisson distribution (such as asymmetry) mean that appropriate standard errors will not be consistent with those derived assuming normality. One remedy for this situation is to simulate the standard errors and confidence intervals using the bootstrap method (see MacKinnon 2002). Unfortunately, confidence intervals cannot be bootstrapped, unless one has access to unit record data for the PES. For this reason, the ABS should be encouraged to estimate DSEs of the Indigenous population.

In view of the non-sampling error potentially arising from census methodology in remote Indigenous communities (Martin et al. 2002), a bootstrap methodology might theoretically allow the design of the census to be taken into account in deriving population estimates (e.g. variable training of SIF interviewers). Unfortunately, this might not be a viable or cost effective strategy because it would probably entail a considerable expansion of the PES.

It may not be possible to get away from reliance on the standard techniques, due to the ongoing problems in matching Indigenous individuals in remote areas (Martin et al. 2002). It would also be extremely difficult to ensure that all the census forms were collected in remote areas by the time of the PES. Increasing the time between the census and the PES does not solve this problem as it has the drawback that it decreases the prospect that the population is closed.

In any event, the recent controversy in the USA illustrates that even a large sample in the follow-up survey does not guarantee reliable estimates, especially if the aim is to estimate local populations. If estimating local population level is intractable, then one could question the worth of releasing Indigenous data for small areas such as Indigenous Locations (ABS 2002). The problem is that there is no sense of the variability of such estimates because they are based simply on actual counts. Note that the reported population distribution across local areas may be misleading and biased if the undercount rate differs in the various areas. While augmenting the national PES sample is unlikely to be the answer, the ABS should consider conducting a small scale DSE in selected remote areas, in order to explore the differences in the undercount rates between remote and other areas. This would help in assessing the magnitude of the biases involved. There are surprisingly few studies that examine the validity of the census counts in remote Indigenous communities, and consequently this situation needs to be remedied urgently (Taylor & Bell 2002).

The use of DSE (and other demographic) adjustment to the population estimates tends to be less biased than unadjusted census counts, but the estimates are more unreliable because they are based on survey data. Zaslavsky (1993) proposes incorporation of a third piece of evidence as a means to reduce both variance and bias. It is argued that better estimators are obtained by combining PES data with a further independent source of data. For example, conditioning on age distribution and male/female ratios can improve DSE-based estimates (see Bell 1993; Robinson et al. 1993a, 1993b; Zaslavsky 1993).

There are relatively large differences between the confidence intervals of the DSEs presented here and the population estimates based on the ABS undercount methodology. This is understandable since, strictly speaking, they may not be measuring the same thing. For example, the undercount methodology is really about addressing problems with census enumeration, whereas DSEs partially abstract from enumeration issues to measure the population who would identify as Indigenous in some way if asked the question enough times. Incorporating demographic data from a third party is likely to be an effective method of reducing bias and variance of the estimators for both techniques.

DSE can be construed as an estimate of the potential Indigenous population, which appears to be relatively stable over time and consistent with recent demographic data (e.g. Kinfu & Taylor 2002). The main error in the error of closure literature is in seeking too much *closure*. Population estimates are just that—estimates which have a random component. It is particularly important to understand the sources of variability, including administration procedures relating to births and deaths. Since DSEs complement traditional demographic techniques, there is also a need for an Australian debate on the relative merits of alternative techniques, as in the debate in the US context featured in Vol. 88 (1993) of the *Journal of the American Statistical Society*.

The central theme of this paper is that policy makers need to take into account the fact that Indigenous population statistics from the census are merely estimates. Given the importance of Indigenous population data in the horizontal fiscal equalisation funding formula, it is particularly important to have accurate estimates for each State and Territory. Obviously, DSEs are not a panacea, but they are an alternative method of benchmarking the Indigenous population, and hence provide an appreciation of the reliability and bias of extant estimates.

This paper also has important implications for the accuracy of estimates of the overall population, especially for certain jurisdictions. The recent decision of the Australian Electoral Commission (AEC) to reduce the number of federal members of parliament for the Northern Territory from two to one is based on the ABS revision of the Territory's population, which now falls under the specified threshold by only 295 people.[19] However, just as the under-enumeration of the Indigenous population is subject to error, so too is the overall population estimate. Obviously, the reliability of population estimates, and hence subsequent projections of the estimated resident population, are particularly questionable in areas where the Indigenous population is substantial. If confidence intervals were constructed around the ABS projections, it is doubtful whether there would be a significant difference between the AEC quota and the Northern Territory's population.

Another related issue is that initial population adjustments for the Northern Territory are calculated using PES data, which are only collected in non-remote areas. Hence the ABS's operational assumption that the structure of underenumeration is the same in remote and non-remote areas will detract from the reliability of revised estimates. This means that the revised Northern Territory

population is even less likely to be significantly different from the threshold set by the AEC.[20] Clearly, policy makers need to be aware that population estimates are not entirely accurate and could be subject to revision if more information were collected. A state of 'closure' will not be attained until the issues of potential bias and reliability of estimates are fully addressed in both remote and other areas.

Notes

1. The change in propensity to identify only affects the size of sub-populations. That is, overall population size is not affected by changes in the composition between Indigenous and non-Indigenous peoples.
2. In Australia, processing decisions may explain the lower annual average increases in the 1976–81 and 1986–91 intercensal periods (Ross 2002). As a result of budgetary decisions, 1976 Census data for most variables, including race, were based on a 50% sample of dwellings. Indigenous status is a clustered variable, i.e. Indigenous people are likely to live in the same households as other Indigenous people rather than being equally likely to be found in all households in an area. As a result, sampling of the racial origin question by dwelling was likely to have overstated the Indigenous count. If this were true, the increase from 1971–76 would be lower in reality, and the slight decrease from 1976–81 would be transformed into an increase.
3. Gardiner and Bourke (2000) provide a critique of the notion of 'recruitment' into the Indigenous population.
4. Kinfu and Taylor describe their procedure as follows: 'During the processing of census data, the Indigenous status variable is edited to change responses of "Aboriginal and/or Torres Strait Islander" to "non-Indigenous" for persons who the ABS considers unlikely to be Indigenous. For the 2001 Census, the variables used for these edits were "Birthplace of male parent" and "Birthplace of female parent". For the 1996 Census these edits were based on the variable "Birthplace of individual", as well as the parents' birthplaces. Had the 2001 edits been applied to the 1996 Census, then 4,302 individuals who were counted as Indigenous would have been reclassified as non-Indigenous. Proper accounting of intercensal change needs to incorporate these changes either by adding 4,302 to the 2001 count or subtracting that number from the 1996 count. The ABS did neither; we chose to delete from 1996' (2002: 17, n. 3).
5. The Australian PES is an interviewer-based survey conducted three weeks after census night which allows us to compare the responses in the census and the PES to identify whether they have changed. Information is collected to determine whether persons have been missed or double counted in the census and whether dwellings were missed. The PES collects personal information on Indigenous origin, age, sex, marital status and birthplace. Note that there are several differences between the census and PES collections. For example, the census question on Indigenous status is based on self-identification whereas the PES involves an interviewer. In addition there were slight differences in the wording of the question. More importantly, the

PES question is asked of the entire household whereas the census is asked of each person individually.

6. See also the capture–recapture website <http://www.pitt.edu/~yuc2/cr/main.htm> maintained by the Global Health Network, University of Pittsburgh.
7. For example, hospital patients tend to be heterogeneous with regard to being 'caught' on a list. If certain doctors refer their patients to certain hospitals, then hospital admissions and doctors' records will not give two independent lists.
8. It should be noted that the PES is not conducted in all areas. In remote communities, a 'rolling count' is applied because of the complexity and time-consuming nature of the census collection. Hence, it is not possible to conduct the PES on the same basis as is done in other areas.
9. A good discussion on the debate over the PES, and its application to estimating the 'true' populations of self-identified groups, is given in the papers contained in Vol. 88 of the *Journal of the American Statistical Society* (1993), particularly Hogan (1993). For a discussion of some of the problems in enumerating the undercount, see also Clogg and Himes (1993), Passel (1993) and Robinson et al. (1993a, 1993b).
10. The main difference between the Australian PES and the ACE appears to be that the PES stressed the matching of records and elimination of duplicates. Another difference is that the US survey places greater reliance on mail-back responses than does the PES.
11. The 2000 Census said the country had 281 million people. The ACE survey said there were 284 million. But analysis using some of the other official documents (i.e. birth, death and immigration records) said there were 279 million. Thus, the census may have overcounted rather than undercounted the country.
12. The implication is that the newly identified Indigenous population is either more 'acculturated' or, in some sense, less 'Indigenous'.
13. The fixed effect regression includes both a dummy variable for each census year and an interaction between each cohort category type (Indigenous males, Indigenous females, non-Indigenous males and non-Indigneous females) and the variables for each year. These interaction terms can be used to test whether there has been a change in the composition of a cohort category between two censuses. The results were not sensitive to variations in the age groups included (for example, including those aged less than 25 or greater than 65 in 1986) nor to the specification of the 'age left school' variable. Thus there was no evidence of compositional change for Indigenous or non-Indigenous cohorts who left school at either 15, 16, 17 or 18 years of age. This technique is analogous to the so-called 'natural experiments' which have become increasingly popular in the economics literature (Myers 1995). For example, the fixed effect model estimated tests for compositional change as embodied in the changing effect on 'age left school' resulting from a possible increase propensity of non-Indigenous people to identify as Indigenous in the respective censuses (i.e. arising from an potentially exogenous change in cohort structure).

14. However, there is always the potential for Indigenous people to be missed from the census collection altogether. See Sanders (2002) for a critical analysis of the standard ABS undercount method, which attempts to address this issue.
15. Therefore the response to the Indigenous question in the census is taken to be correct. The estimated Indigenous population in the non-PES areas is close to Butterfield's (1998) estimates of the number of people receiving the SIF in the last two censuses. Butterfield estimates that 54,928 people used the SIF in 1991 and 62,884 did so in 1996.
16. Martin and Taylor (1996) found that an ethnographically-based enumeration was 17 per cent higher than the census count in Aurukun on Cape York Peninsula.
17. Note that this is not the same figure as given in the Estimated Residential Population in ABS (1998: Table 11, 23), which gives breakdowns of adjustments and includes information from other sources.
18. The Poisson distribution, first described by the French mathematician Siméon-Denis Poisson, indicates the probability that a random event will occur in a time or space interval under the conditions that the probability of the event occurring is very small, but the number of trials is very large so that the event actually occurs a few times. As the number of events becomes larger (and the probability of event occurring is less small), the Normal distribution (the most common distribution used in statistical studies) provides a reasonably description of the empirical distribution.
19. AEC press release, 'Commissioner Issues Federal Electoral Determination', 20 February 2003.
20. Note that it is entirely possible that the estimates would be lower if the ABS benchmarked its undercount assumptions for remote areas.

Appendix A. Applying the DSE method to the potential population of Indigenous Australians

The Sekar–Deming method of estimating populations was originally developed with reference to medical applications where survey data were to some extent inaccurate or incomplete with respect to the population. For example, it was initially used to compare hospital birth and death registrations with a household survey of births and deaths. The key to this method is an ability to match individual records, on some different criteria to the one of interest, and then check the observation of interest for consistency. In a two-outcome situation, such as a yes/no question, four potential outcomes occur, as illustrated in Table A1. First, the record can be 'yes' on both the initial and second surveys, designated by the cell x_{11}. Second, the record can be 'yes' on the first and 'no' on the second, designated by the cell x_{12}. Third, the record can be 'no' on the first and 'yes' on the second, denoted by cell x_{21}, and finally the record can be 'no' on both surveys, given by x_{22}. This method cannot, of course, pick up information that has been incorrectly recorded on both surveys (e.g. respondents answering 'yes' on both surveys when the true observation was 'no').

Table A1. A two outcome example of DSE methodology

	Response A		
	Yes	No	Total
Response B			
Yes	x_{11}	x_{12}	$x_{11} + x_{12}$
No	x_{21}	x_{22}	$x_{21} + x_{22}$
Total	$x_{11} + x_{21}$	$x_{12} + x_{22}$	$x_{11}+x_{12}+ x_{21}+ x_{22}$

DSEs rely on there being a difference in response between one collection and the next. Using the Sekar–Deming (1949) formula, the revised population estimate is:

$$\hat{N} = x_{11} + x_{12} + x_{21} + x_{12}x_{21}/x_{11} \qquad (1A)$$

If Table A1 refers to the response to a question about Indigenous status, then only x_{22} people always deny they are Indigenous. Consequently, Hunter (1998a) referred to the potential Indigenous population as being equal to $x_{11} + x_{12} + x_{21}$. However, some of the x_{22} people may also admit to being Indigenous in other circumstances. The 4th term on the right-hand side of equation 1A is the number expected to identify as Indigenous at least once if surveys are 'independent' (in statistical terms).

The variance of $\hat{N}$ can be estimated using the standard binomial approach as

$$Var(\hat{N}) = \sqrt{(\hat{N}\,\hat{q}_1\,\hat{q}_2/\hat{p}_1\,\hat{p}_2)} \qquad (2A)$$

where

$$\hat{p}_1 = x_{11}/(x_{11} + x_{21})$$
$$\hat{p}_2 = x_{11}/(x_{11} + x_{12})$$
$$\hat{p}_1 + \hat{q}_1 = \hat{p}_1 + \hat{q}_2 = 1$$

That is, $\hat{p}_1$ refers to the proportion of people who answered 'yes' in both surveys compared to the total number of people who answered 'yes' in survey A. Similarly $\hat{p}_2$ refers to the proportion of people who consistently answered 'yes' in both surveys compared to those who answered 'yes' in survey B. The proportions, $\hat{q}_1$ and $\hat{q}_2$ reflect the alternative scenario in each case.

In earlier work, a natural approach to the construction of confidence intervals was to assume that $\hat{N}$ was asymptotically normal and use $\hat{N} \pm 1.96$ standard errors. Unfortunately the distribution of $\hat{N}$ is often skewed in practice and the above interval can give misleading results. To get around this one might try and find a suitable transformation that would make the distribution look more like a normal random variable (Chao 1989). Another approach is to use the boot-strap method for numerically estimating an appropriate confidence interval (Greene 2000).

Appendix B. Constructing confidence intervals on population estimates using ABS undercounts

The results reported in the text use the initial census counts to estimate confidence intervals for the Indigenous population, since the DSE estimates are also scaled by these counts. Another reason for this procedure is to avoid any duplication or circularity in the logic between the ABS's demographic adjustments and the use of an adjusted census count. For the purposes of this paper, the difference between using the initial or adjusted census is not important because the following calculations are produced solely for the purposes of benchmarking the DSE estimates and associated standard errors.

For the 1996 Census we have an Indigenous persons undercount rate of 7.1 per cent with a standard error of 1.04 per cent.

If denote final undercount rate = r

final population estimate = y

census count = x,

then $r = 1 - x/y$,

which can be re-expressed:

$x = (1-r)*y$ and $y = x/(1-r)$ (1B)

Therefore, the census counts from Table 1 and the undercounts for the respective censuses can be plugged into equation 1B to produce our estimates of the Indigenous population. For 1996 data, the estimate is 379,950. A band of two standard errors (or rather 1.96 times the standard errors to be more precise) either side of the estimate will produce 95 per cent confidence intervals. The band for the estimate of the undercount in this case is $1.96*1.04 \pm 7.1$, giving an upper

bound of 9.14 per cent and a lower bound of 5.06 per cent. Using our estimate of x above we calculate y_u and y_l as:

y_u = 352,970/ (1-0.0914) = 388,500

y_l = 352,970/ (1-0.0506) =371,800

That is, a band of approximately 8,000 on either side of the ABS point estimate (or about 2.1% of the total estimated Indigenous population).

References

Australian Bureau of Statistics (ABS) 1997. *Information Paper: Census of Population and Housing: Data Quality—Undercount*, Cat. no. 2940.0, ABS, Canberra.

—— 1998. *Experimental Estimates of the Aboriginal and Torres Strait Islander Population*, Cat. no. 3230.0, ABS, Canberra.

—— 2002. *Population distribution, Aboriginal and Torres Strait Islander Australians 2001*, Cat. no. 4705.0, ABS, Canberra.

Alho, J.M., Mulry, M.H., Wurdeman, K. and Kim, J. 1993. 'Estimating heterogeneity in the probabilities of enumeration for dual system estimation', *Journal of the American Statistical Association*, 88 (423): 1130–6.

Bell, W.R. 1993. 'Using information from demographic analysis in post-enumeration survey estimation', *Journal of the American Statistical Association*, 88 (423): 1106–18.

Butterfield, M. 1998. 'Changes in census methodology and the impacts on counts', paper presented to the Workshop on Establishing Guidelines for Interpreting Indigenous Census Data, 26 February 1998, University House, ANU, Canberra.

Chao, A. 1989. 'Estimating population size for sparse data in capture–recapture experiments', *Biometrics*, 45: 427–38.

Choi, C. and Gray, A. 1985. 'An evaluation of census counts of the Aboriginal population, 1971, 1976 and 1981', *Occasional Paper No. 2/1985*, ABS, Canberra.

Clogg, C.C. and Himes, C.L. 1993. 'Comment: Uncertainty in demographic analysis', *Journal of the American Statistical Association*, 88 (423): 1072–6.

Darroch, J.N., Fienberg, S.E., Glonek, G.F.V. and Junker, B.W. 1993. 'A three-sample multiple-recapture approach to census population estimation with heterogenous catchability', *Journal of the American Statistical Association*, 88 (423): 1137–48.

Evans, J., Kahles, D. and Bate, C. 1993. '1991 Census data quality: Aboriginal and Torres Strait Islander counts', *Census Working Paper No. 93/6*, ABS, Canberra.

Fienberg, S.E. 1992. 'Bibliography on capture–recapture modelling with application to census undercount adjustment', *Survey Methodology*, 18 (1): 143–54.

Gardiner, G. and Bourke, E.A. 2000. 'Indigenous populations, "mixed" discourses and identities', *People and Place*, 8 (2): 43–52.

Gray, A. 1997a. 'The explosion of Aboriginality: Components of Indigenous population growth 1991–96', *CAEPR Discussion Paper No. 142*, CAEPR, ANU, Canberra.

—— 1997b. 'Growth of the Aboriginal and Torres Strait Islander population, 1991–2001 and beyond', *CAEPR Discussion Paper No. 150*, CAEPR, ANU, Canberra.

—— and Tesfaghiorghis, H. 1993. 'Aboriginal population prospects', *Journal of the Australian Population Association*, 10 (2): 81–101.

Gray, M.C. and Auld, A.J. 2000. 'Towards an index of relative Indigenous socioeconomic disadvantage', *CAEPR Discussion Paper No. 196*, CAEPR, ANU, Canberra.

—— and Hunter, B.H. 2002. 'A synthetic cohort analysis of labour market outcomes of Indigenous and non-Indigenous Australians, 1986–96', *Australian Economic Review*, 35 (4): 391–404.

Greene, W.H. 2000. *Econometric Analysis*, Prentice Hall, New Jersey.

Guimond, É. 1999. *Ethnic Mobility and the Demographic Growth of Canada's Aboriginal Populations from 1986 to 1996*, Cat. no. 91–209-XPE, Statistics Canada, Ottawa.

Harris, D. 1994. 'The 1990 Census count of American Indians: What do the numbers really mean?' *Social Science Quarterly*, 75 (3): 580–93.

Hogan, H. 1993. 'The 1990 post-enumeration survey: Operations and results', *Journal of the American Statistical Association*, 88: 1047–60.

Hunter, B.H. 1998a. 'Assessing the utility of 1996 Census data on Indigenous Australians', *CAEPR Discussion Paper No. 154*, CAEPR, ANU, Canberra.

—— 1998b. 'Assessing the validity of intercensal comparisons of Indigenous Australians, 1986–96', *Journal of the Australian Population Association*, 15 (1): 51–67.

Kinfu, Y. and Taylor, J. 2002. 'Estimating the components of Indigenous population change: 1996–2001', *CAEPR Discussion Paper No. 240*, CAEPR, ANU, Canberra.

Le Cren, E.D. 1965. 'A note on the history of mark–recapture population estimates', *Journal of Animal Ecology*, 34 (3): 453–4.

Levy, P.S. and Lemeshow, S. 1991. *Sampling of Populations: Methods and Applications*, John Wiley and Sons, New York.

MacKinnon, J. 2002. 'Bootstrap inference in econometrics', *Canadian Journal of Economics*, 35 (4): 51–67.

Martin, D.F., Morphy, F., Sanders, W.G. and Taylor, J. 2002. *Making Sense of the Census: Observations of the 2001 Enumeration in Remote Aboriginal Australia*, CAEPR Research Monograph No. 22, CAEPR, ANU, Canberra.

—— and Taylor, J. 1996. 'Enumerating the Aboriginal population of remote Australia: Methodological and conceptual issues', *Journal of the Australian Population Association*, 13 (1): 17–33.

Myers, B.D. 1995. 'Natural and quasi-experiments in economics', *Journal of Business and Economic Statistics*, 13 (2): 151–61.

Paradies, Y., Huppatz, S., Warnsey, J. and Barnes, T. 2000. 'Population and globalisation: Australia in the 21st century', Paper delivered at the 10th Biennial Conference of the Australian Population Association, Melbourne.

Passel, J.S. 1976. 'Provisional evaluation of the 1970 Census count of American Indians', *Demography*, 13: 397–409.

—— 1993. 'Comment on Robinson et al.' *Journal of the American Statistical Association*, 88 (423): 1074–7.

—— 1996. 'The growing American Indian population, 1969–1990: Beyond demography', in G.D. Sandefur, R.R. Rindfuss and B. Cohen (eds), *Changing Numbers, Changing Needs: American Indian Demography and Public Health*, National Academy Press, Washington DC.

—— and Berman, P.A. 1986. 'Quality of 1980 Census data for American Indians', *Social Biology*, 33 (3/4): 163–82.

Robinson, J.G. 2001. *Accuracy and Coverage Evaluation: Demographic Analysis Results*, DSSD Census 2000 Procedures and Operations Memorandum Series B-4, US Census Bureau, Washington DC.

——, Ahmed, B., Das Gupta, P. and Woodrow , K.A. 1993a. 'Estimation of population coverage in the 1990 United States Census based on demographic analysis', *Journal of the American Statistical Association*, 88 (423): 1061–71.

——, ——, —— and —— 1993b. 'Rejoinder', *Journal of the American Statistical Association*, 88 (423): 1077–9.

Ross, K. 1999. *Occasional Paper: Population Issues, Indigenous Australians*, Cat. no. 4708.0, ABS, Canberra.

—— 2002. 'Recent trends in the demography of the indigenous populations of Australia, New Zealand, Canada and the United States of America', in G. Briscoe and L. Smith (eds), *The Aboriginal Population Revisited: 70,000 Years to the Present*, Aboriginal History Monograph No. 10, Aboriginal History, Canberra.

Sanders, W.G. 2002. 'Adapting to circumstance: The 2001 Census in the Alice Springs town camps' in Martin, D.F., Morphy, F., Sanders, W.G. and Taylor, J. *Making Sense of the Census: Observations of the 2001 Enumeration in Remote Aboriginal Australia*, CAEPR Research Monograph No. 22, CAEPR, ANU, Canberra.

Sandefur, G.D., Rindfuss, R.R. and Cohen, B. 1996. *Changing Numbers, Changing Needs: American Indian Demography and Public Health*, National Academy Press, Washington DC.

Sekar, C. and Deming, E.W. 1949. 'On a method of estimating birth and death rates and extent of registration', *Journal of the American Statistical Association*, 44 (1): 101–15.

Shapiro, S. 1949. 'Estimating birth registration completeness', *Journal of the American Statistical Association*, 45 (2): 261–4.

Shyrock, H.S., Siegel, J.S. and Associates 1976. *The Methods and Materials of Demography*, Academic Press, London.

Skerry, P. 2000. *Counting on the Census: Race, Group Identity and the Evasion of Politics*, Brookings Institution Press, Washington DC.

Snipp, M.C. 1997. 'Some observations about racial boundaries and the experiences of American Indians', *Ethnic and Racial Studies*, 20 (4): 667–89.

Taylor, J. 1997a. 'Changing numbers, changing needs? A preliminary assessment of Indigenous population growth 1991–1996', *CAEPR Discussion Paper No. 143*, CAEPR, ANU, Canberra.

—— 1997b. 'The contemporary demography of Indigenous Australians', *Journal of the Australian Population Association*, 14 (1): 77–114.

—— 2002a. 'The context of observation', in D.F. Martin, F. Morphy, W.G. Sanders and J. Taylor (eds), *Making Sense of the Census: Observations of the 2001 Enumeration in Remote Aboriginal Australia*, CAEPR Research Monograph No. 22, CAEPR, ANU, Canberra.

—— 2002b. 'Indigenous enumeration in the late twentieth century: Emerging issues for population analysis' in G. Briscoe and L. Smith (eds), *The Aboriginal Population Revisited: 70,000 years to the Present*, Aboriginal History Monograph No. 10, Aboriginal History, Canberra.

—— and Bell, M. 2002. 'Benchmarking ABS population estimates for Queensland Aboriginal and Torres Strait Islander communities', *QCPR Discussion Paper No. 1*, QCPR, University of Queensland, St Lucia.

—— and Hunter, B.H. 1998. *The Job Still Ahead: Economic Costs of Continuing Indigenous Employment Disparity*, Office of Public Affairs, ATSIC, Canberra.

Zaslavsky, A.M. 1993. 'Combining census, dual-system, and evaluation study data to estimate population shares', *Journal of the American Statistical Association*, 88 (423): 1092–105.

Centre for Aboriginal Economic Policy Research Publications

For information on earlier CAEPR Discussion Papers and Research Monographs please contact:

Publication Sale, Centre for Aboriginal Economic Policy Research, The Australian National University, Canberra, ACT, 0200, Telephone: 02–6125 8211, Facsimile: 02–6125 2789. Information on CAEPR, abstracts or summaries of all CAEPR print publications and those published electronically, can be found at the following WWW address: **http://online.anu.edu.au/caepr/**

MONOGRAPH SERIES

7. *Mabo and Native Title: Origins and Institutional Implications*, W. Sanders (ed.), 1994.
8. *The Housing Need of Indigenous Australians, 1991*, R. Jones, 1994.
9. *Indigenous Australians in the Economy: Abstracts of Research, 1993–94*, L.M. Roach and H.J. Bek, 1995.
10. *Native Title: Emerging Issues for Research, Policy and Practice,* J. Finlayson and D.E. Smith (eds), 1995.
11. *The 1994 National Aboriginal and Torres Strait Islander Survey: Findings and Future Prospects*, J.C. Altman and J. Taylor (eds), 1996.
12. *Fighting Over Country: Anthropological Perspectives,* D.E. Smith and J.D. Finlayson (eds), 1997.
13. *Connections in Native Title: Genealogies, Kinship and Groups,* J.D. Finlayson, B. Rigsby and H.J. Bek (eds), 1999.
14. *Land Rights at Risk? Evaluations of the Reeves Report*, J.C. Altman, F. Morphy and T. Rowse (eds), 1999.
15. *Unemployment Payments, the Activity Test and Indigenous Australians: Understanding Breach Rates,* W. Sanders, 1999.
16. *Why Only One in Three? The Complex Reasons for Low Indigenous School Retention*, R.G. Schwab, 1999.
17. *Indigenous Families and the Welfare System: Two Community Case Studies*, D.E. Smith (ed.), 1999.
18. *Ngukurr at the Millennium: A Baseline Profile for Social Impact Planning in South East Arnhem Land*, J. Taylor, J. Bern and K.A. Senior, 2000.
19. *Aboriginal Nutrition and the Nyirranggulung Health Strategy in Jawoyn Country*, J. Taylor and N. Westbury, 2000.
20. *The Indigenous Welfare Economy and the CDEP Scheme*, F. Morphy and W. Sanders (eds), 2001.
21. *Health Expenditure, Income and Health Status among Indigenous and Other Australians*, M.C. Gray, B.H. Hunter and J. Taylor, 2002.
22. *Making Sense of the Census: Observations of the 2001 Enumeration in Remote Aboriginal Australia*, D.F. Martin, F. Morphy, W.G. Sanders and J. Taylor, 2002.

RECENT DISCUSSION PAPER SERIES

224/2001 *A regional CDEP for four remote communities? Papunya, Ikuntji, Watiyawanu and Walungurru*, W. Sanders.

225/2001 *Building Indigenous learning communities*, R.G. Schwab and D. Sutherland.

226/2001 *Sustainable development options on Aboriginal land: The hybrid economy in the twenty-first century*, J.C. Altman.

227/2002 *The Indigenous population of Cape York Peninsula, 2001–2016*, J. Taylor and M. Bell.

228/2002 *Urban CDEPs as Indigenous Employment Centres: Policy and community implications*, M. Champion.

229/2002 *Three years on: Indigenous families and the welfare system, the Kuranda community case study*, R. Henry and D.E. Smith.

230/2002 *Towards an Indigenous order of Australian government: Rethinking self-determination as Indigenous affairs policy*, W. Sanders.

231/2002 *Population futures in the Australian desert, 2001–2016*, J. Taylor.

232/2002 *Autonomy and the Community Development Employment Projects scheme*, W.S. Arthur.

233/2002 *Jurisdictional devolution: Towards an effective model for Indigenous community self-determination*, D.E. Smith.

234/2002 *Indigenous community stores in the 'frontier economy': Some competition and consumer issues*, S. McDonnell and D.F. Martin.

235/2002 *Some competition and consumer issues in the Indigenous visual arts industry*, J. C. Altman, B.H. Hunter, S. Ward and F. Wright.

236/2002 *Indigenous residential treatment programs for drug and alcohol problems: Current status and options for improvement*, M. Brady.

237/2002 *Journey without end: Reconciliation between Australia's Indigenous and settler peoples*, W. Sanders.

238/2002 *Decentralisation, population mobility and the CDEP scheme in central Cape York Peninsula*, B.R. Smith.

239/2002 *Welfare and the domestic economy of Indigenous families: Policy implications from a longitudinal survey*, A. Daly, R. Henry and D. Smith.

240/2002 *Estimating the components of Indigenous population change, 1996–2001*, Y. Kinfu and J. Taylor.

241/2002 *Reforming Indigenous welfare policy: Salutary lessons and future challenges for Australia from the US experience*, A. Daly and D.E. Smith.

242/2002 *Philanthropy, non-government organisations and Indigenous development*, R.G. Schwab and D. Sutherland.

243/2003 *Options for benchmarking ABS population estimates for Indigenous communities in Queensland*, J. Taylor and M. Bell.

WORKING PAPER SERIES

Available at no cost on WWW at http://online.anu.edu.au/caepr/

3/1999 *Dealing with alcohol in Alice Springs: An assessment of policy options and recommendations for action*, M. Brady and D.F. Martin.

4/1999 *Aboriginal people in the Kakadu region: Social indicators for impact assessment*, J. Taylor.

5/1999 *Reforming the Northern Territory Land Rights Act's financial framework into a more logical and more workable model*, J.C. Altman and D.P. Pollack.

6/2000 *Governance and service delivery for remote Aboriginal communities in the Northern Territory: Challenges and opportunities*, N. Westbury and W. Sanders.

7/2000 *What's in it for Koories? Barwon Darling Alliance Credit Union and the delivery of financial and banking services in north-west New South Wales*, N. Westbury.

8/2000 *The relative social and economic status of Indigenous people in Bourke, Brewarrina and Walgett*, K. Ross and J. Taylor.

9/2001 *Indigenous communities and business: Three perspectives, 1998–2000*, J.C. Altman.

10/2001 *Indigenous Australian arrest rates: Economic and social factors underlying the incidence and number of arrests*, B.H. Hunter.

11/2001 *Sensitivity of Australian income distributions to choice of equivalence scale: Exploring some parameters of Indigenous incomes*, B.H. Hunter, S. Kennedy, and D.E. Smith.

12/202 *Indigenous Australians and competition and consumer issues: An interim review of the literature and an annotated bibliography*, J.C. Altman, S. McDonnell, and S. Ward.

13/2002 *The rise of the CDEP scheme and changing factors underlying Indigenous employment*, B.H Hunter.

14/2002 *Institutional factors underpinning Indigenous labour force participation: The role of the CDEP scheme and education*, B.H. Hunter.

15/2002 *Generating finance for Indigenous development: Economic realities and innovative options*, J.C. Altman.

16/2002 *The spatial context of Indigenous service delivery*, J. Taylor.

17/2002 *The potential use of tax incentives for Indigenous businesses on Indigenous land*, O. Stanley.

18/2002 *Banking on Indigenous Communities: Issues, options, and Australian and international best practice*, S. McDonnell and N. Westbury.

19/2002 *One size fits all?: The effect of equivalence scales on Indigenous and other Australian poverty*, B.H. Hunter, S. Kennedy and N. Biddle.