4. Selected methodological issues for analysis of the 2002 NATSISS

Nicholas Biddle and Boyd Hunter

The most commonly used data for information on Indigenous Australians is the Census of Population and Housing. The five-yearly census allows us to generate reasonably reliable social statistics about Indigenous people as a by-product of the introduction, in 1971, of a question which asked whether people identify as Indigenous. However, the census is a blunt instrument that is designed primarily to count the national population rather than to measure and track changes in complex socioeconomic conditions of population sub-groups. Furthermore, census questions are limited in their number and scope by the exigencies and costs involved in collecting information from the entire population. Surveys are more flexible and cost-effective instruments for collecting a wide range of information, even though the resulting data is subject to sampling error.

In 1991 the Royal Commission into Aboriginal Deaths in Custody recommended a national survey of the Indigenous population. It was the dearth of information with which to inform the Royal Commission that resulted in the first NATSIS in 1994. This survey provided the first nationwide inter-censal estimates of Indigenous socioeconomic status.

The 2002 NATSISS is the second major nationwide survey specifically targeted to collect a large range of information on Indigenous Australians. Carried out between August 2002 and April 2003, it collected information from 9359 individuals aged 15 years and over from 5887 households. Some of the information had never been collected before for the Indigenous population, whereas a number of the questions were broadly comparable to the 1994 NATSIS.

The 2002 survey was also conducted more or less concurrently with the 2002 General Social Survey (GSS) which collected information about the total adult Australian population (the Indigenous and non-Indigenous populations are not separately identifiable in the GSS). Many of the data items in the 2002 NATSISS are comparable with the GSS, but the GSS did not collect information in very remote areas and was limited to individuals 18 years and over.

The ABS has a program (or cycle) of Indigenous household surveys, with the next NATSISS survey scheduled for 2008. Other ABS data collections with significant Indigenous components planned before then are the 2004–05 Indigenous Health Survey, the 2006 Community Housing and Infrastructure

Needs Survey (CHINS) and, of course, the 2006 Census of Population and Housing.

The NATSISS survey was designed to 'enable analysis of the interrelationship of social circumstances and outcomes, including the exploration of multiple disadvantage' (ABS 2005c: 1). Information is provided across a range of topics:

- demographic characteristics of the individuals and household and geographic characteristics of the area in which they live
- cultural and language information and the family and community context
- health and disability
- education participation and achievement
- employment
- income, housing and financial stress, and
- information technology, transport and law and justice.

This chapter seeks to outline a number of the methodological issues concerning the 2002 NATSISS, with a particular focus on helping readers of this monograph understand the remainder of the empirical results and analysis presented. The next section will outline the survey methodology, including the scope, sample selection and survey design and implementation. Section 3 will provide an overview of some of the potential issues one might need to take into account in an empirical analysis of the NATSISS, while section 4 will provide a brief critical analysis of the existing ABS outputs from the 2002 NATSISS. Finally, section 5 will summarise and highlight the main implications of this chapter.

Survey methodology

Scope of the survey

The 2002 NATSISS collected information from Indigenous Australians aged 15 years and older who were usual residents in private dwellings at the time of the survey. In line with standard ABS household survey scope, it excludes visitors to the randomly selected private dwellings. [1] The survey was carried out across Australia with the aim of collecting enough information to make conclusions at either the State/Territory level or by the Australian Standard Geographic Classification (ASGC) remoteness classification. [2] The coverage of the 2002 NATSISS data is different to the GSS (which only collected information on people aged 18 and over in private dwellings) as well as the 1994 NATSIS (which collected information from people aged 13 years and over in both private and non-private dwellings).

---

[1] Some information may have been collected from these visitors if their usual residence was also selected as part of the collection.
[2] All geographic definitions of remoteness in this paper are based on the ASGC remoteness structure.

The difference in age structures is reasonably easy to take into account by re-weighting, though it should always be kept in mind during comparative analysis. However, given that the 2002 NATSISS did not collect information on people in non-private dwellings, differential coverage may be more problematic when making comparisons with the 1994 NATSIS or the 2001 Census, or when drawing conclusions about the total Indigenous population. Non-private dwellings include hotels, motels, hostels, hospitals, short-stay caravan parks and—perhaps most importantly—prisons and other correctional facilities. Such dwellings can be identified in both the 1994 NATSIS and the 2001 Census.

According to the ABS (2005c), at 31 December 2002 there were an estimated 19 320 Indigenous people living in non-private dwellings, or about 4 per cent of the entire Indigenous population. The following discussion gives some information from the 1994 NATSIS on people who were usual residents of non-private dwellings. The sample size is 375 adults.

By definition, all prisoners in the 1994 NATSIS data were in non-private dwellings. Biddle and Hunter (2004) provided a profile of Indigenous people in non-private dwellings using the original weights for the 1994 NATSIS. Residents of non-private dwellings were more likely than Indigenous residents in private dwellings to have been arrested in the last five years. They were also concentrated outside capital cities, and were more likely to be male and young. Also, a higher percentage of respondents from non-private dwellings were taken from their natural families. Without access to accurate sampling errors for the 1994 NATSIS data, it is difficult to definitively claim that the differences between private and non-private dwellings are significant. However, it seems reasonable to assert that the people living in private and non-private dwellings are drawn from different populations.

Therefore, when statements are made regarding data from the 2002 NATSISS, it may not always be appropriate to say 'the Indigenous population has a given characteristic' but rather make a more qualified statement about 'the Indigenous population living in private dwellings' or, if referring to law and justice data, 'that portion of the Indigenous population that is not currently in prison or other non-private dwelling'.

Given the substantial growth in the Indigenous population since 1994, it is important that comparisons between the 1994 and 2002 surveys use the re-weighted 1994 data set (when the ABS makes it available as a CURF). Customised cross-tabulations provided by the ABS will use the re-weighted data as a matter of course.

Sample selection and survey design

The overall sample was spread across States/Territories in order to produce estimates that have a relative standard error of no more than 20 per cent for

characteristics that are relatively common in the Indigenous population (for example, that at least 10% of the population would possess). However, there were two components to the 2002 NATSISS sample designs. The first (in parts of Queensland, South Australia, Western Australia and the Northern Territory) was based on a sample of discrete Indigenous communities and the outstations associated with them. This is the Community Area (CA) sample. In the remainder of these four States and Territories, as well as in all of New South Wales, Victoria, Tasmania and the Australian Capital Territory, the survey methodology and sample design was somewhat different. The data from these other areas are described as the Non-Community Area (NCA) sample. Around 30 per cent of the sample came from the CAs and 70 per cent from the NCAs. Those in NCAs were interviewed using Computer Assisted Interviewing, whereas those in CAs were interviewed using a pen and paper interview.

The differences between survey questions and survey technique raise one of the most important issues for people analysing 2002 NATSISS data. Before documenting such issues, it is also necessary to briefly discuss how individuals were selected in each of the two types of areas.

The CA sample was obtained from a random selection of discrete Indigenous communities and outstations. The sample frame used to design the survey was based on both 2001 Census counts, and information collected in the 2001 CHINS (ABS 2005c: 4). Once the communities had been selected, a random selection of dwellings was made.

Dwellings—and therefore individuals—in NCAs were selected using a stratified multi-stage area sample based on the 2001 Census. A random selection of dwellings within selected census Collection Districts (CDs) was then screened to assess their usual residents' Indigenous status. An insufficient number of households with Indigenous Australians was initially collected, so additional CDs were sampled during February to April 2003.

Before moving on to the CA and NCA survey design, it is important to make clear the difference between the CA sample and the concept of remote areas. The sampling in non-remote areas (i.e. major cities, inner regional and outer regional areas) was carried out entirely under the NCA methodology. This included 5242 of the surveyed individuals.

In remote areas (which includes the remote and very-remote Accessibility/Remoteness Index of Australia classifications), both CA and NCA sampling methodology was used. Remote areas that were not identified as 'discrete communities' used the same sampling methodology and interviewing techniques as were used in non-remote areas (i.e. the NCA methodology). In remote areas where NCA methods were used, there were 1997 respondents. In discrete communities and outstations, CA sampling was used, with information collected on 2120 individuals. Although the distinction between CAs and remote

areas is not entirely clear in published record to date, according to correspondence with the ABS, the majority of those collected under the CA sample were from very remote areas rather than remote areas.

Questionnaire design and output

The questionnaire for the 2002 NATSISS was designed with the assistance of a special advisory group. Although the questions in the CAs and NCAs were broadly similar, there were still differences in what was asked, and the way the data was presented for publication by the ABS. The variables that were affected by such decisions are listed in Table 4.1 and can be classified into three main categories: those that were collected in both CAs and NCAs but that have different output categories; those collected in NCAs only; and those collected—but not released—in remote areas.

Table 4.1. Differences in data collection in CA and NCA areas

Restriction	Variable
Collected in both NCAs and CAs, with different categories outputted in remote areas	Main reason for last move
	Type of stressor in last 12 months
	Type of social activities in last three months
	Presence of neighbourhood/community problems
	Neighbourhood/community problems
	Whether used formal child care in last four weeks
	Type of child care used in last four weeks
	Main reason for not using (more) formal child care in last four weeks
	Type of organisation undertook unpaid voluntary work for in last 12 months
	Disability status
	Tenure type
	Type of major structural problems
	All sources of personal income
	Principal source of personal income
	Where used computer in last 12 months
	Where used internet in last 12 months
	Modes of transport
	Type of legal services used
	Attendance at cultural events in last 12 months*
	Self-assessed health*
	Type of government pension/allowance (auxiliary)*
	Whether working telephone at home*
Collected in NCAs only	Whether has an education restriction
	Whether has an employment restriction
	Disability type
	Multiple job holder
	Cash flow problems
	All types of cash flow problems
	Number of types of cash flow problems
Collected, but not released in remote areas	Whether ever used substances
	Type of substances ever used
	Whether used substances in last 12 months
	Type of substances used in last 12 months

Source: ABS (2004d)
Notes: An * refers to variables where the only difference between the CAI and PAPI samples is the presence or absence of a 'not stated' option.

For those variables in the first part of the table, it is not always clear, from the published record, why the ABS chose different output categories in remote and non-remote areas. After reading the questionnaires and through correspondence with the ABS, it would appear that for these questions, a reduced set of options was available to interviewees. It also seems that this was done mainly because the ABS, on advice from stakeholders and their testing processes, felt that these options would not be relevant to those living in CAs, so the benefits of having a more streamlined survey could be gained without adversely affecting the quality of the survey. However, the ABS needs to be clearer and more transparent about how it came to the conclusion that these particular variables needed

different data outputs, and hence potentially limited the ability to use such data items with total confidence.

For example, full information on where computers and internet were used in the last 12 months is not available in remote areas. The missing category was internet/cyber cafes. While it could be argued that there are no internet cafes in CAs, it presumes that community residents are not mobile and have not visited urban settings where such cafes may exist. If the ABS has information that this is the case, then it needs to be made clear. Even though it should be acknowledged that the final survey content for CAs were a product of advice from ABS stakeholders and testing which identified items that were either inappropriate or did not work in the household interview environment in CAs, the ABS does not necessarily provide the level of detail that more sophisticated users may require (e.g. ABS 2004c: 20–26).

The second category of variables in Table 4.1 is those that were collected in NCAs, but not in CAs. Obviously, the structure of the survey prevents comparisons for these variables, but it is worth a brief reflection on why these decisions may have been made. For example, it is unlikely that many respondents held more than one job in CAs, so it does not make much difference that there is no relevant data in such areas. With respect to the lack of cash flow data in CAs, it is arguable that the notions of cash flow and poverty differ in remote and non-remote areas, so will differ between CAs and NCAs (Altman & Hunter 1998). Once again, though, it would be useful for the ABS to publish why such decisions were made without requiring analysts to speculate themselves, with incomplete information.

The disability variables warrant special mention, as quite different variables were constructed for the NCA and CA samples. As a result of field testing and consultative processes, data items and questions for a range of topics, including disability, were modified to take account of language and particular circumstances of people living in very remote communities. In addition, Indigenous stakeholders advised that attempts to measure psychological disabilities in remote communities required development of an appropriate instrument sensitive to the circumstances of people in these areas. The interim instrument developed by the ABS and used with Indigenous stakeholder endorsement in the recent 2004–05 National Aboriginal and Torres Strait Islander Health Survey is the initial response to that requirement. In the longer term, a culturally appropriate social and emotional wellbeing question module is being developed by Indigenous stakeholders.

Full disability was collected in NCAs, and this is comparable to data items in the GSS. A modified set of disability questions that did not include psychological disability was collected in CAs. This question was combined with the relevant options from the NCA sample to create a new variable which can be used across both samples (see ABS 2005c), though not in comparative analysis with the GSS.

While the different nature of the labour markets in remote and non-remote areas may mean that comparative data on education and employment restrictions caused by a disability may not be very meaningful (that is, given the binding constraints evident in the lack of employment prospects outside the CDEP scheme), this issue must still be of concern for the remote areas using NCA methodology.

The last category in Table 4.1 is the substance use variables for which data was collected but not released in remote areas. The relevant questions in the 2002 NATSISS were based on the National Drug Strategy Household Survey (NDSHS) and had a response rate of over 90 per cent. In NCAs, a voluntary self-enumerated form was used to collect this information, whereas in CAs, respondents were required to respond verbally to questions asked by an interviewer. The low prevalence of substance use reported in CAs has been assumed, by the ABS, to be the result of the use of direct questioning in CAs (ABS 2005c). It is further assumed that this led to a significant adverse effect on both the level of response and the quality of responses to questions on substance use. For this reason, information on substance use in remote areas was considered to be unreliable and has not been released.

Interviewing techniques

Not only were some of the questions different in the CAs and NCAs, so too were the interviewing techniques. As indicated above, interviews in NCAs were conducted using a CAI where interviewers use a notebook computer to read the questions and to record the data gathered. If respondents were asked to choose from a range of options, then prompt cards were used. For the substance use questions, a voluntary self-enumerated form was used with a response rate of 90 per cent.

In CAs, the surveying techniques were modified to take into account the cultural and language differences predicted for these areas. Firstly, the interviewing was conducted by more traditional pen and paper interviews. In addition, Community Information Forms (CIFs) were used to collect information about the community from the local council office. In every community, Indigenous facilitators were used to improve the validity of the data. However, not all interviews in CAs were conducted in the presence of facilitators. These facilitators 'explained the purpose of the survey to respondents, introduced the interviewers, assisted in identifying the usual residents of a household and in locating residents who were not at home, and assisted respondents in understanding questions where necessary' (ABS 2005c).

While the *differential* use of facilitators may have introduced potential interviewer bias in the response to some of the questions, accurate records were not kept by the ABS as to when facilitators were used. This means it is not possible to control

the analysis of 2002 NATSISS for the effect of the presence of facilitators. However, it is still important to appreciate the relative importance of so-called 'non-sampling error' (which includes interviewer bias when facilitators are and are not present) and sampling error (which is present, to a greater or lesser extent, in all survey data).

Remote/non-remote and CA/NCA

The issue of remote/non-remote and CA/NCA is a confusing one, especially with regard to analysis of the CURF. The biggest issue is that on the CURF, those 1997 individuals in remote areas collected under the NCA methodology cannot be distinguished from those 2120 who were collected under the CA methodology. This is problematic when the questions asked across the two methodologies are quite different. Consider the question used to obtain information on whether a person was a victim of assault. In the NCA sample, the person was asked, 'In the last 12 months, did anyone, including persons you know, use physical force or violence against you?'. In the CA sample, on the other hand, the person was asked, 'In the last year, did anybody start a fight with you or beat you up?'.

There can be arguments made for or against the results from the two questions being comparable or not. However, from a methodological point of view, the point is that it is impossible for analysts using the CURF data to make that decision themselves. To repeat, users of the CURF are unable to identify whether people were surveyed under the NCA or CA sample, only whether they lived in remote or non-remote areas. The ABS does have such information available on the 2002 NATSISS MURF, and it may be possible for analysts to access it through customised tables. Ultimately, it is up to users as to whether they need to pursue such options to investigate data quality issues in greater detail.

Sampling and non-sampling issues

This section looks at a number of issues one must take into account when analysing the 2002 NATSISS or when interpreting results published elsewhere. It begins with a discussion of sampling error and non-sampling error so that analysts appreciate the relevant issues when interpreting the data. The section then benchmarks data from the 2002 NATSISS to other comparable sources. The remainder of the section then looks at a number of non-sampling issues which arise when comparing data items through time, comparing data to other contemporary sources, or interpreting results from ABS publications. These issues are not necessarily all the potential pitfalls involved with analysing the data. Rather, this section uses several examples in an attempt to highlight that analysts need to be aware of the context of the survey before making conclusions based on the data.

Sampling error

Sampling error arises from the fact that samples differ from their populations in that they are usually small sub-sets of the total population. Therefore, survey sample results should be seen only as estimations. Henry (1990) note that sampling errors cannot be calculated for non-probability samples, but they can be determined for probability samples, such as the 2002 NATSISS. First, to determine sample error, look at the sample size. Then, look at the sampling fraction—the percentage of the population that is being surveyed. The more people surveyed, the smaller the error. This error can also be reduced, according to Fox and Tracy (1986), by increasing the representativeness of the sample.

The 2002 NATSISS gathered information from 9359 Indigenous Australians. This was estimated to represent about 1 in 30 Indigenous Australians aged over 15 at the time of the survey (ABS 2005c). Such a large sample allows reasonably accurate inferences to be made about the population as a whole, as is shown by the quite similar results from 2002 NATSISS and other data sources (for which sampling error is not an issue). For example, the number of CDEP participants as recorded by the 2002 NATSISS is almost identical to administrative records (2005c: 27). However, irrespective of the concordance of results with administrative or population data sources, all data in the 2002 NATSISS is subject to sampling error and hence analysts must be aware of the reliability of estimates.

A standard error is the extent to which we expect a calculation from a sample of individuals to differ from what the same calculation would be if we had information for the whole population. Alternatively, it can be thought of as the amount of variation around the estimated value due to the fact that only a sample of people was taken.

Fortunately, sampling error can be quantified with reasonable accuracy. To the credit of the ABS, and in comparison to previous surveys, information is now given in the survey publications and CURF to enable outside users to make their own estimates of the possible sampling error. Therefore, in contrast to much of the existing analysis of the 1994 NATSIS, it is now possible to put some figures around how confident we can be about certain conclusions we make from the data.[3] That is, we are able to calculate the standard error of an estimate.

Standard errors for estimates of means and proportions can be used in two ways: constructing confidence intervals and performing hypothesis tests. Strictly speaking, a confidence interval refers to what one would expect to occur under repeated sampling. So, a 95 per cent confidence interval means that, if a given value was the true population value, then in 95 per cent or 19 out of 20 repeated

---

[3] The ABS is attempting to calculate replicate weights for the 1994 NATSIS CURF. This is a non-trivial exercise that has been facilitated by the development of computing processing since 1994 and driven, in part, by the increasing sophistication of users.

samples, we would expect the estimated value from a sample to lie within the range of the confidence interval. Although it is not strictly true, a 95 per cent confidence interval can also be thought of as the area bounding an estimate within which there is only a 5 per cent chance that the true population percentage is outside this range (but a sample of this size came up with the estimate that it did).

Instead of putting a range around the estimate, we may also want to test whether a certain estimate is different from another value. This value may be one we come up with, or another estimated value from the survey. To see whether this difference is significant in a statistical sense, a researcher can perform a hypothesis test. They would come up with a hypothesis, and then decide whether they can or cannot reject this hypothesis based on the estimated value from the sample, and the associated standard error. Note that one can only say conclusively that the population value is not a given percentage, never equivocally that it is.

There are three main ways in which one can estimate standard errors and hence estimate confidence intervals or perform hypothesis tests. Firstly, the ABS provides an approximation of the standard errors in tabular form which can be used with an Excel spreadsheet created by CAEPR for this conference. Secondly, the ABS provides a spreadsheet for estimates from the initial results publication that gives (relative) standard errors. Users can perform their own calculations with these estimates. Finally, on the CURF there are replicate weights that can be used to construct exact estimates of standard errors which users can then use to perform hypothesis tests. [4] There are formulas for deriving standard errors from replicate weights provided in ABS (2005c: 19).

Benchmarking 2002 NATSISS data against population estimates in the 2001 Census

This section benchmarks some selected results from the 2002 NATSISS using similar data from the 2001 Census to provide an indication of the underlying population estimates. The census estimates are calculated using almost the entire Indigenous population (i.e. less a relatively small undercount), and hence are reasonably reliable (from a sampling point of view). As indicated above, 2002 NATSISS data have sampling error associated with them, and hence the benchmarking exercise uses standard errors to calculate the 95 per cent confidence intervals around the results. These confidence intervals are represented as 'whiskers' around the 2002 NATSISS estimates in the following figures.

Strictly speaking, one should only compare 2002 NATSISS results to private dwellings in the census. However, the non-private dwelling estimates are also

---

[4] These replicate weights have the added benefit of being useful in calculating standard errors for estimates from regression style models that take into account the design of the survey.

presented to allow us to see who is missing from the sample. Figure 4.1 shows the proportion of the population aged 15 and over with various levels of attainment at secondary school in non-remote areas. Figure 4.2 illustrates the same results for remote areas.

Residents in non-private dwellings in both remote and non-remote areas appear to be more heterogeneous than those in private dwellings. That is, non-private dwellings are more likely to contain both people who did not go to school and people who have Year 12 or equivalent. This reflects the fact that non-private dwellings include a diverse range of residences: prisons and hostels that contain a disproportionate number of people who are extremely disadvantaged; and nursing homes and hotels that may include people with better access to resources (e.g. people with later-year secondary qualifications).

Figure 4.1. School attainment in non-remote areas, 2001 and 2002
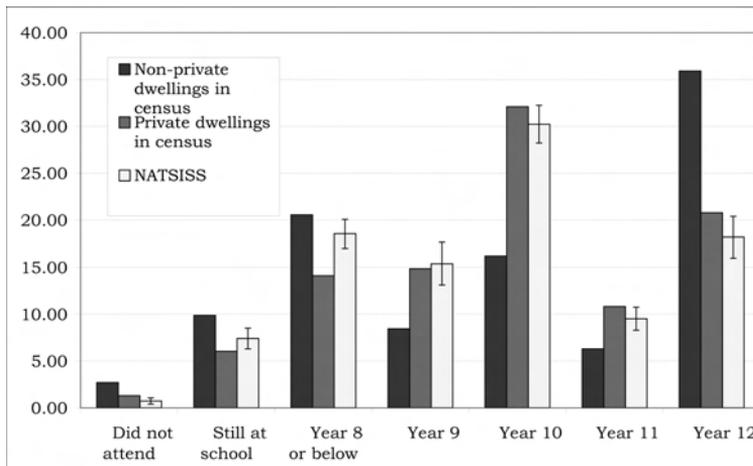
Source: Customised cross-tabulations.

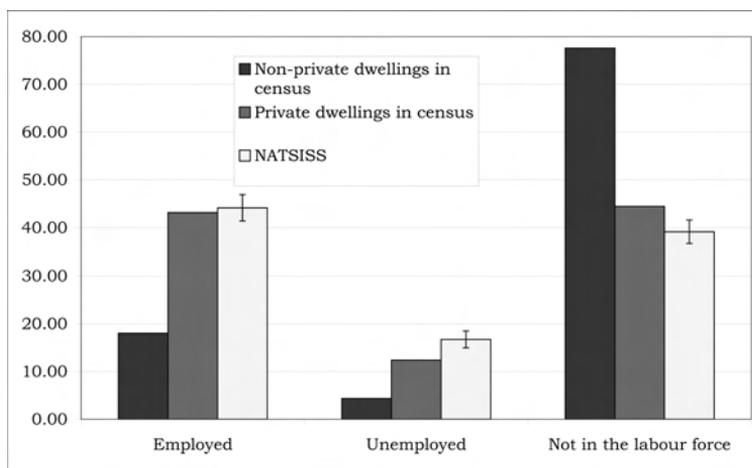**Figure 4.2. School attainment in remote areas, 2001 and 2002**
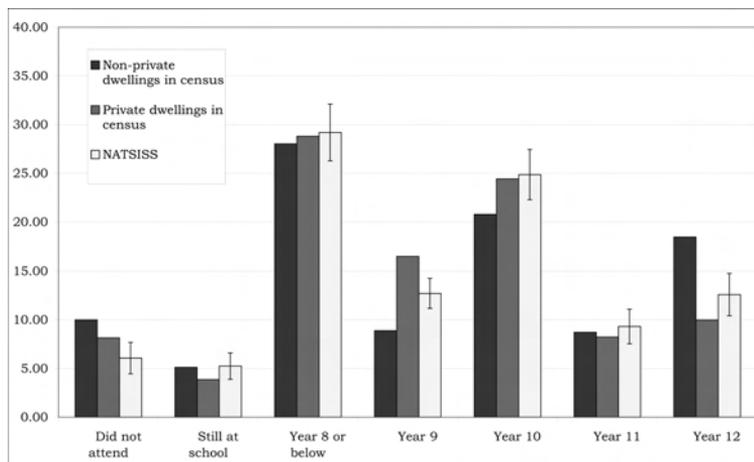
Source: Customised cross-tabulations.

In non-remote areas, it would appear that compared to the census, 2002 NATSISS had slightly lower completion rates for the later years of schooling. That is, there was a significantly lower proportion of Indigenous residents in private dwellings who have completed Year 11 or Year 12, no significant difference between those who completed Year 10 or Year 9, and a significantly higher proportion who were reported to have completed Year 8 or less (all at the 5% level of significance). 2002 NATSISS respondents were significantly more likely to be at school but less likely to have not gone to school at all.

In remote areas, there appeared to be less systematic variation. There is no significant difference between 2002 NATSISS estimates and private dwellings in the census for Still at school, Year 8 or below, Year 10, and Year 11. However, there was a significant difference between 2002 NATSISS and comparable census results for Year 12, Year 9 and for 'Did not go to school'.

Figures 4.3 and 4.4 report labour force outcomes in the 2001 Census and 2002 NATSISS, once again presented for non-remote and then remote areas. Residents of non-private dwellings tend to have lower labour force participation rates than those in private dwellings, despite the apparent heterogeneity noted above. This probably reflects that being in jail and nursing homes tends to constrain an individual's opportunities to work or look for employment, irrespective of whether they completed the later years of secondary school.

Figure 4.3. Labour force status in non-remote areas, 2001 and 2002

Source: Customised cross-tabulations.

Figure 4.4. Labour force status in remote areas, 2001 and 2002
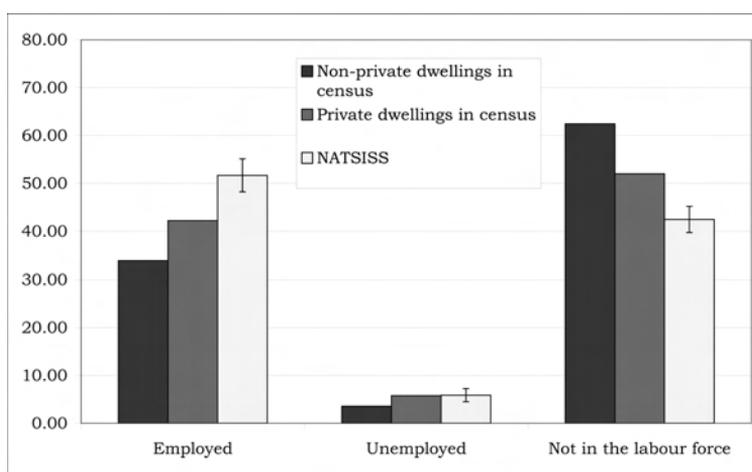
Source: Customised cross-tabulations.

In non-remote areas, the proportion employed in the 2002 NATSISS is not significantly different from the 2001 Census estimates for private dwellings. However, compared to the census, more NATSISS respondents are classified as unemployed, but fewer are classified as being not in the labour force. This could either represent a change in the outcomes over time, seasonality differences caused by the respective collections occurring at different times in the year, or the greater opportunity within a face-to-face interview to clarify what is meant by unemployment.

In remote areas, there appears to have been a shift from respondents being not in the labour force in the census to being employed in the 2002 NATSISS. This

apparent shift occurs despite the fact that 2002 NATSISS provides accurate estimates of CDEP scheme employment that is disproportionately concentrated in remote areas. Once again, it is unclear as to whether the difference between 2002 NATSISS and the census is a result of differential timing of the collections (although the magnitudes make that unlikely) or a greater chance of clarification in a face-to-face interview compared to the largely self-enumerated census. What we do know, though, is that the differences we have reported in the labour market and educational status are unlikely to have been caused by sampling error alone.

Notwithstanding this, overall there appears to be less difference between the labour force status estimates from the 2002 NATSISS and the adjacent census than was evident in the 1994 NATSIS. Hunter and Taylor (2001) showed that there was a substantially higher employment and unemployment rates in the earlier survey when compared to the 1996 Census. While participation rates were again higher when survey methodology was used in 2002, unemployment-to-population ratios were less different in 2002 than they were in 1994. Indeed, unemployment ratios in remote areas were not significantly different in the 2002 NATSISS and the census. This may reflect the fact that the estimates of CDEP scheme employment appeared to be accurately measured in the 2002 NATSISS, and consequently there was less scope for mis-classification of CDEP scheme employment as unemployment.

Benchmarking these variables showed a number of differences that are unlikely to have been explained by sampling error alone. Although it is only possible to speculate, these differences could have been caused by:

- the question sequencing in the two data collections
- the use of face-to-face interviewing in the 2002 NATSISS as opposed to self-reporting in the census
- real changes in the year or so between the collections, or
- differences caused by the timing of the census and the survey (i.e. seasonality).

These can all be classified, to a certain extent, as non-sampling error (either in the census or the 2002 NATSISS). Although the differences were mostly small, the remainder of this section looks at three areas where non-sampling error is perhaps a bigger issue.

Non-sampling errors

Non-sampling errors can be defined as errors arising during the course of survey activities rather than resulting from the sampling procedure.[5] Unlike sampling errors, there is no simple and direct method of estimating the size of non-sampling

[5] Note that non-sampling errors can be present in both sample surveys and censuses.

errors. In most surveys, it is not practical to measure the possible effect on the statistics of the various potential sources of error arising from things other than the statistical sample. However, there has been a considerable amount of research on the kinds of errors that are likely to arise in different kinds of surveys. By examining the procedures and operations of a specific survey, experienced survey analysts may be able to assess its quality. Rarely will this produce actual error ranges, as for sampling errors. In most cases, the analyst can only state that, for example, the errors are probably relatively small and will not affect most conclusions drawn from the survey, or that the errors may be fairly large and inferences are to be made with caution. In rare instances, researchers may be able to say with some confidence in what direction the error might be.

Non-sampling errors can be classified into two groups: random errors whose effects approximately cancel out if fairly large samples are used; and biases which tend to create errors in the same direction and thus cumulate over the entire sample. With large samples, systematic errors, and resultant biases, are the principal causes for concern about the quality of a survey. For example, if there is an error in the questionnaire design, this could cause problems with the respondent's answers, which in turn, can create processing errors, etc. These types of errors often lead to a bias in the final results and analyses. In contrast to sampling variance and random non-sampling error, bias caused by systematic non-sampling errors cannot be reduced by increasing the sample size.

Non-sampling errors can occur because of problems in coverage, response, non-response, data processing, estimation and analysis. The following discussion is adapted from an excellent exposition on the subject from the Statistics Canada web site (see the non-sampling error section of Statistics Canada 2005).

An error in coverage occurs when there is an omission, duplication or wrongful inclusion of the units in the population or sample. Omissions are referred to as under-coverage, while duplication and wrongful inclusions are called over-coverage. These errors are caused by defects in the survey frame: inaccuracy, incompleteness, duplication, inadequacy and obsolescence. There may be errors in sample selection, or part of the population may be omitted from the sampling frame, or weights to compensate for disproportionate sampling rates may be omitted. Coverage errors may also occur in field procedures (e.g. if a survey is conducted but the interviewer misses several households or people).

Response errors result from data that have been requested, provided, received or recorded incorrectly. The response errors may occur because of inefficiencies with the questionnaire, the interviewer, the respondent or the survey process. Subject matter experts are often in a good position to identify flaws in such aspects of the survey.

Poor questionnaire design is a common aspect of non-sampling error. It is essential that sample survey or census questions are worded carefully in order to avoid

introducing bias. If questions are misleading or confusing, the responses may end up being distorted.

As alluded to above, an interviewer and facilitators can influence how a respondent answers the survey questions. This may occur when the interviewer is too friendly or aloof or prompts the respondent. To prevent this, interviewers must be trained to remain neutral throughout the interview. They must also pay close attention to the way they ask each question. If an interviewer changes the way a question is worded, it may impact on the respondent's answer.

Respondents can also provide incorrect answers by their own volition. Faulty recollections (recall bias), tendencies to exaggerate or underplay events, and inclinations to give answers that are more 'socially desirable', are several reasons why a respondent may provide a false answer. Individuals may conceal the truth out of fear or suspicion of the survey process and the institutions sponsoring it (i.e. governments and their agencies). Other respondent errors may arise through a failure to understand the underlying concepts or a basic lack of knowledge about the information requested.

Non-sampling errors can also arise from the survey process. Using proxy responses (taking answers from someone other than the respondent) or a lack of control over the survey procedures are just two ways of increasing the possibility of response errors. Processing errors sometimes emerge during the preparation of the final data files. For example, errors can occur while data are being coded, captured, edited or imputed. Coder bias is usually a result of poor training or incomplete instructions, variance in coder performance (e.g. tiredness or illness), data entry errors, or machine malfunction (some processing errors are caused by errors in computer programs). Sometimes, errors are incorrectly identified during the editing phase. Even when errors are discovered, they can be corrected improperly because of poor imputation procedures.

Non-response errors—another category of non-sampling error—can also result from having not obtained sufficient answers to survey questions. Complete non-response errors occur when the survey fails to measure some of the units in the selected sample. Reasons for this type of error may be that the respondent is unavailable or temporarily absent; the respondent is unable or refuses to participate in the survey; or the dwelling is vacant. If a significant number of people do not respond to a survey, the results may be biased, since the characteristics of the non-respondents may differ from those who have participated. Given the high rates of mobility among Indigenous people, it is difficult to discount these issues in the 2002 NATSISS (and indeed other surveys involving mobile populations, Hunter & Smith 2002). Taylor and Kinfu's chapter on mobility will explore these issues in some detail.

Researchers and policy makers need to make themselves familiar with the discussion of non-sampling error in ABS (2004c: 56–7). The discussion on

imputation issues for 2002 NATSISS in ABS (2004c: 61–2) is also relevant: analysts should pay careful attention to the list of variables where imputed data was used. To summarise that discussion, a small amount of missing data was imputed for a range of educational and training variables due to errors in the CAI instrument. While the ABS has been admirably candid about non-sampling errors in their publications, it is necessary to explore a few specific examples, to assist researchers and policy-makers in interpreting 2002 NATSISS data.

Possible non-sampling issues for selected variables in 2002 NATSISS

This section looks at three sets of variables in the 2002 NATSISS for which the possibility of non-sample error is likely to be pronounced. By non-sample error, we mean aspects of the survey methodology that are likely to result in differences between results from the 2002 NATSISS and other published data. The variables examined are: hunting, gathering and fishing; substance use and high-risk drinking; and education variables. This is not an exhaustive list of variables that are potentially problematic—rather, the discussion provides several examples of some of the issues users need to be on top of when interpreting 2002 NATSISS data.

Hunting, gathering and fishing can be an important part of the social life of a person's community, as well as their wider social relations. Although this may be true for many non-Indigenous Australians (especially in rural areas), it is particularly true for Indigenous Australians. However, for many Indigenous Australians—especially in remote areas—hunting, gathering and fishing also provide an important supplement to cash income and are hence an important part of their economic life. This is also true, although perhaps to a lesser extent, in non-remote areas (Altman, Gray & Halasz 2005).

Altman, Buchanan and Biddle's chapter on the customary sector illustrate that the information on the customary economy in the 1994 NATSIS was potentially biased by the structure of the questionnaire. Uncertainty about the definition of voluntary work meant that many people who may have engaged in hunting, fishing and gathering probably did not indicate this—for example, in Kununurra the incidence of both variables was zero, an implausible finding. That is, there was a strong geographic correlation between the incidence of voluntary work and hunting, fishing and gathering. This is probably driven by the response to the voluntary work question, as the anthropological evidence points to the existence of substantial hunting and gathering by the Indigenous people in the area. Consequently, the 1994 measures of hunting, fishing and gathering tend to understate the extent of the customary economy.

The way in which information on hunting, gathering and fishing was collected in the 2002 survey also makes it difficult to obtain a meaningful measure of the

extent to which individuals participate in such activities. In 2002, the question was asked as part of a group cultural activity, and refers to 'fishing or hunting in a group'. In the 1994 NATSIS, the questions were asked as part of the category of voluntary work. However, an analysis of the Wallis Lake Catchment (which covers non-remote areas of Forster/Tuncurry) showed that local Indigenous people did not necessarily see hunting, gathering and fishing as being either voluntary work or a group cultural activity (Altman, Gray & Halasz 2005).

The ABS was probably right not to include hunting, gathering and fishing as a question incorporated under voluntary work, but collecting the information as a group cultural activity is not an adequate alternative. Needless to say, given that different questions were asked in 1994 and 2002, and the fact that both were problematic, the change in hunting, fishing and gathering between the two surveys is uninformative, if not meaningless.

The scope for non-sampling error is also apparent in the substance use questions. In NCAs, these questions were asked via a voluntary self-enumerated form. Pilot testing by the ABS concluded that due to English literacy problems in CAs, better information could have been obtained by asking respondents to respond verbally to questions asked by the interviewer. However, the very low prevalence of substance use in CAs led the ABS to conclude that the information obtained was not reliable and, as such, information is only available for NCAs.

The prevalence of high-risk drinking in 2002 NATSISS is substantially lower than was found in both the 2001 NHS (conducted by the ABS) and the 1994 NDSHS, Urban Aboriginal & TSI Peoples Supplement. The NDSHS had 3000 confidential interviews from metropolitan and other urban areas.

The relatively low prevalence of alcohol abuse identified in the chapter by Brady and Chikritzhs may have been caused by the different survey methodology (e.g. populations sampled, sample size, survey method, alcohol questions). Unlike the NDSHS, the alcohol questions in the 2002 NATSISS were neither confidential nor self-completed. Although a one-to-one interview was invited, respondents often answered questions in the presence of other family members and may have been reluctant to give accurate estimates. That is, asking other members to leave the room is an implication in itself. Another issue may be recall problems when questions refer to a longer time period (e.g. 12 months rather the last two weeks).

The important issue arising from the discussion of substance abuse is that non-sampling error is likely to be particularly important when asking sensitive questions about embarrassing issues or illicit activities

There were sequencing errors in the section of the NCA questionnaire on education, and the ABS needed to impute values for these variables based on the responses to other questions and information from the 2001 Census. The first sequencing error affected the 733 respondents aged 20–24 who were not studying

full-time. These individuals were not asked whether they were currently studying, nor were they asked the type of education institute they were attending. Those individuals in that age group who were studying part-time would therefore not have been recorded as such. Using data on Abstudy receipts and distributional information from the census, the ABS coded 4 per cent of this sample with missing information as studying part-time.

An additional sequencing error occurred with the 1399 respondents who had used employment support services in the last 12 months. These individuals were sequenced past four questions on vocational training. According to the ABS, imputation for the five vocational training questions was conducted using 'donor records' (ABS 2005c: 19), where information from another person was matched to records based on sex, age and labour force status. These donor records were derived from other individuals in the sample. The ABS should indicate which categories of individuals were used to impute records so that analysts can assess whether they can discount the possibility of any significant bias arising from the imputation procedure.

In general, it would be useful for the ABS to publish the exact imputation techniques for both of these problems, so other researchers using the CURF can test the sensitivity of their results. It may be the case that the methodology used by the ABS is as robust as possible for most purposes, but certain applications or modelling frameworks may require a different technique.

A further issue with regard to the education questions is that we can see a difference that cannot be explained by sampling error alone. This is clear when comparing the data on the proportion of people with non-school qualifications in the 2001 Census with the results from the 2002 NATSISS. This can be shown in Figure 4.5, where once again the 'whiskers' refer to the 95 per cent confidence interval.

Figure 4.5. Non-school qualifications, 2001 and 2002
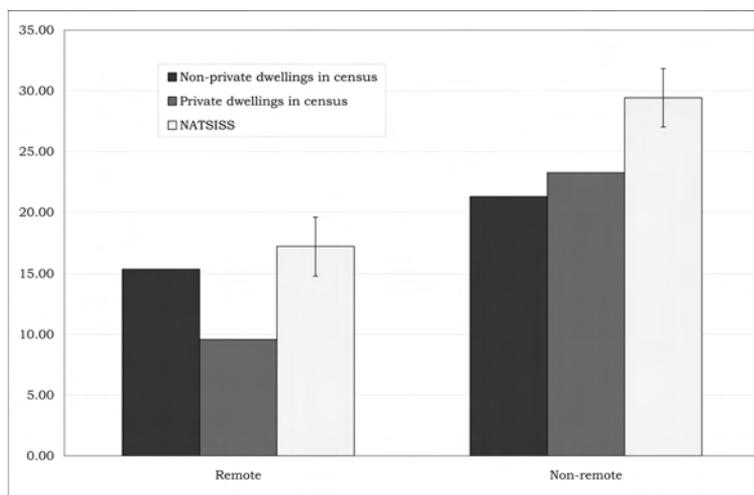
Source: Customised cross-tabulations.

This discrepancy between the 2002 NATSISS and the 2001 Census data is too great to be explained by the year or so gap in between the two data collections. The difference has, however, been noted in ABS (2005c) and is explained as a consequence of census reporting being undertaken by people on behalf of those being enumerated. That is, the non-sampling error in this comparison quite possibly arises from the census methodology.

Outputs from the 2002 NATSISS

There are three broad categories of output from the 2002 NATSISS that individuals are likely to use in their analysis of the data. These are: ABS publications (mainly the initial results publication), customised tables ordered from the ABS, and the CURF. We critically explore the first and the last of these categories as the chapter by Webster, Rogers and Black (in this volume) documents the process of securing customised tables in some detail.

ABS publications using the 2002 NATSISS

In June 2004, the ABS released the initial results from the 2002 NATSISS (ABS 2004c). This document included a 15-page summary of findings as well as 23 tables presenting the data in different ways and making comparisons with other data sources. The summary of findings focuses on selected changes through time (i.e. comparing the 1994 NATSIS and the 2002 NATSISS) as well as making explicit comparisons between the Indigenous population and the non-Indigenous population (as measured by the GSS).

While the depth of data available is impressive, and the provision of standard errors commendable, one criticism of ABS (2004c) could be made. In Table 9 of

that publication, the proportion of people with various characteristics is presented, broken down by their equivalised household income. Based on analysis of the non-Indigenous population (ABS 2003c), the ABS outputs data for the second and third decile as a measure of 'low income', arguing that the lowest income decile has characteristics closer to those with higher incomes. ABS (2004c) use this 'low income' group as an indicator for 'poverty' in their discussion of results. This appears to be an implicit endorsement of criticism of income-based measures of poverty, especially the claim that measurement error (i.e. under-reporting) is pronounced for low income earners, especially those who indicate they have an income less than or equal to zero (Tsumori, Saunders & Hughes 2002). One group in which it is particularly difficult to accurately measure income is the self-employed. The recent public debate on the difficulties in identifying a socially accepted minimum standard of living illustrates that this criticism is not totally uncontested (Harding, Lloyd & Greenwell 2001; Saunders 2002; Tsumori, Saunders & Hughes 2002). One reason to be cautious about adopting the ABS definition of 'low income' without question is that self-employment is not prominent in the Indigenous population (Hunter 2004a: chapter 5).

Even if one can make a case for using this 'low income' indicator for the non-Indigenous population, a cursory examination of Table 4.2 shows that this assumption is suspect for the Indigenous population. Compared to those in the second and third decile, those in the first decile are less likely to be employed and own or purchase a home, and are more likely to have fair or poor health. There is no significant difference between the bottom decile and the ABS 'low income' group in terms of qualifications.

Table 4.2. Implications of ABS definition of low income[a]

	1st decile	2nd–3rd decile	4th–10th decile	Not stated
Employed	16.9	30.2	76.0	48.0
Has qualification and/or completed Year 12	17.7	21.7	44.5	27.6
Owns or purchasing home	11.9	15.1	43.8	29.0
Health fair or poor	30.5	26.5	16.6	23.4

a. The difference between the values for the 1st decile and the 2nd-3rd decile is always significant at the 5% level of significance.
Source: Customised cross-tabulations

Other variables not reported in Table 4.2 indicate that the bottom decile respondents are more disadvantaged than the 'low income group', as they are more likely to have not completed Year 12, be unemployed or outside the labour force, have been arrested in the last year, and have transport difficulties. The ABS definition of low income therefore tends to understate the incidence of Indigenous disadvantage, and should not be used.

Another criticism of ABS (2004c) is the lack of clarity regarding the difference between the concept of remote areas and the different sampling methodology employed in CAs. One example is that information on cash flow problems were collected in NCAs, but not in CAs (ABS 2004d: 32). However, in Table 4 of ABS (2004c), the footnote says the data was collected in non-remote areas only. However, this begs the question of what information was collected from the 1997 individuals who were collected as part of the NCA enumeration strategy but whose usual residence is in a remote area. The ABS probably has retained this information in the MURF and people may be able to access it when requesting customised tables. If this is the case, it would be helpful if the ABS made this clear. If the data is not available because of reliability concerns, it would minimise misunderstanding by explaining why this is the case.

There also needs to be more discussion about the underlying models and theory that appear to motivate the layout and choice of variables for some of the tables. For example, if you compare the age breakdowns in tables 10 and 11 from ABS (2004c), quite different age groups are used. While a reader could speculate as to why the different age groups are chosen, the publication needs to provide links to the previous research or theoretical models the ABS used to inform their judgments. In addition, the ABS needs to provide justification for why certain variables were deemed to be amenable to age standardisation in the official publication. The age standardisation of health is a reasonably standard technique, but it is unclear why employment status is better suited to age standardisation than, say, education participation (ABS 2004c: Table 5).

The ABS has also issued a number of publications and related tables for each State and Territory, but these follow the same basic conventions and formats used in ABS (2004c). Webster, Rogers & Black (this volume) give more detail about future planned output from the NATSISS.

The 2002 NATSISS CURF

On 7 June 2005 the ABS released the CURF for the 2002 NATSISS. Containing a unique record for all those households and individuals who were part of the survey, the data set enables the researcher to run their own cross-tabulations (subject to constraints) as well as more detailed statistical analysis of the data. Unlike the 1994 NATSIS and the GSS, the 2002 NATSISS is only available via the RADL. That is, individuals who have access to the data submit Statistical Analysis Software (SAS) or Statistical Package for the Social Sciences (SPSS) programming code which is then checked to make sure confidential information is not released. This is then run by the ABS and the output posted on the user's section of the RADL web site.

There are two separate files available via the RADL. The first has household information and a weight for the 5887 households in the sample. The second has

information on each of the 9359 individuals, a household identifier that enables household information to be merged with the data, and a person level weight. These person and household weights can (and should) be used to turn information on the sample into estimates for the population, taking into account a person's or household's chance of selection. Furthermore—and this is a substantial improvement over the 1994 NATSIS—each household and person has 250 replicate weights which can be used to generate standard errors for estimates (as shown in ABS 2005c).

Another major improvement is that the 2002 NATSISS CURF has some information that has not been available to researchers on other ABS data sets. For example, previous data sets that included significant Indigenous populations, including the 1994 NATSIS, 2001 NHS and past censuses, have not included continuous income data. The 2002 NATSISS, on the other hand, contains both continuous individual data and household income data, up to a cut-off beyond which income data is censored.[6] This will enable distributional analysis outside the ABS that has not been possible before. Household rent and mortgage payments are also given as continuous values, although they are also right censored.[7]

There are, however, data items which have been made confidential, restricting the type of analysis that is possible. A number of employment variables are outputted in ranges. These include duration of unemployment and CDEP employment, as well as the number of hours usually worked per week. This latter variable places particular restrictions on analysis of hourly—as opposed to weekly—income.

More importantly, apart from in Queensland, one cannot separately identify Aboriginal Australians from Torres Strait Islanders. This is despite the fact that the Torres Strait Islander population was over-sampled in order to produce reliable (separate) estimates of the characteristics of Torres Strait Islanders living in the Strait and the rest of Queensland (ABS 2005c: 53).

The largest number of—and arguably the most confusing—restrictions on CURF data are imposed on the geographic variables. Apart from the variables mentioned in Table 4.1 that have different categories for CA/NCA or remote/non-remote areas, there are also restrictions on the type of geographic breakdowns that are possible using the 2002 NATSISS CURF. In particular, it is possible to undertake analysis by State/Territory or by remoteness, but not both simultaneously. As an example of the implication of such restrictions, it is possible to examine whether income is higher in remote versus non-remote areas throughout

---

[6] This right censoring of income appears to affect 0.65% of individuals and no households (at least when equivalised household income is used). These estimates of the proportion who are right censored are for those individuals or households who have stated their income (& rent/mortgage in the next endnote). For income, it includes those with negative or nil income, as well as those 'not applicable'.

[7] Censoring of weekly rent and mortgage repayments affects 3.45% and 12.96% of households respectively.

Australia, or higher in NSW versus other States. However, it is not possible to examine whether income is higher in non-remote versus non-remote areas within NSW specifically. While this restriction is understandable, given the imperative to protect the confidentiality of respondents, it does restrict the ability to make meaningful interstate comparisons (that is, by constraining the ability to control for the level of remoteness in the respective States).

If users want to obtain some sort of cross-classification of State/Territory and remoteness structure, the ABS insists they use a specific variable (STXREM) when attempting to do this on the RADL. Users are not permitted to submit programs that include both State/Territory and remoteness structure under any circumstances.

An additional restriction on geographic analysis is that Tasmania and the Australian Capital Territory (ACT) are included as one variable. This is despite the statement that 'the NATSISS was designed to provide reliable estimates at the national level and for each State and Territory' (ABS 2005c) and that ABS (2004c) outputted data for each State separately. Within these two combined localities, there were 736 individuals from Tasmania and 330 from the ACT, which when weighted represented about 10 900 and 2600 individuals respectively. The State governments in these two States have quite different policies, and the presence of the federal government and two large universities in Canberra (among other things) make these two populations quite different. In addition, the geography of these two regions in terms of remoteness and access to resources are quite dissimilar. Thus the joining of these two localities somewhat constrains policy analysis and restricts a modeler's ability to control for geography when analysing the relationship between other variables.

At the moment, one of the biggest constraints for analysing the 2002 NATSISS using the RADL is the requirement to undertake the analysis in only two statistical programs: SAS (version 8.2) or SPSS. While SAS performs most of the standard types of analysis, there are a number of things it either cannot do or that are not done as easily as in other statistical packages (such as the calculation of 'robust' standard errors and the easy output of marginal effects from estimation on the probit model). Perhaps more importantly, SAS is expensive. This may not be as big an issue in large government departments, or even perhaps large research centres, where the fixed costs can be spread across a number of users, but for smaller organisations or individual researchers, the costs can become prohibitive. [8] SPSS is more affordable, but it does not have the range of applications that other programs have. Users may be able to visit a data laboratory at the ABS site where they have been able to install and run the widely-used

---

[8] CAEPR cannot afford to use SAS and only a handful of people in the ANU now use SAS because of cost constraints. Stata and other cost effective statistical packages that allow more sophisticated statistical analyses are now preferred by many users.

statistical program, Stata. However, it is likely that this option will cost the user extra money because the ABS may need to employ a staff member to ensure the user does not compromise confidentiality. The ABS hopes to have a fully operational version of Stata on the RADL in the near future.

Concluding remarks

To summarise the discussion in this paper, there are three main messages that we feel users of the data should keep in mind when making conclusions.

Firstly, researchers and policy-makers need to be aware of the different survey methodology used in CAs and NCAs. Given that no identifiers for CAs are publicly released on the CURF or the RADL, this effectively means that analysts have to be aware that a large proportion of remote data in the 2002 NATSISS was collected from CAs. It may be possible to purchase customised cross-tabulations from the ABS if people are particularly concerned about potential non-sampling error arising from the differing methodology in the two types of areas. [9] Whatever one's position, some caution is warranted when making comparisons between the CAs and NCAs, and possible remote and non-remote areas.

The second issue that should be kept in mind, especially when making comparisons with either the 1994 NATSIS or the 2001 Census, is that 2002 NATSISS only collects information from Indigenous people aged 15 and over living in private dwellings. That is, the statistics collected do not provide information on the entire population of Indigenous Australians. Where other data sources are used, analysts need to ensure they are comparing the same age groups; non-private dwellings need to be excluded; and, in the case of the 1994 NATSIS, the re-weighted data set must be used once it becomes available on the CURF.

The final point is that given that the ABS has released information that enables users to reasonably accurately estimate standard errors, it is no longer acceptable to make statements about differences in means or proportions without taking into account sampling error. Analysts need to make themselves aware of the formula for calculating standard errors and confidence intervals. More sophisticated users will have to work out how to use replicate weights. Given that there is considerable room for programming errors, especially when using arrays in SAS, the ABS should consider releasing draft programming code (particularly, if such code is already available) to minimise the possibility of non-sampling error arising from analysts. Ultimately, though, analyst error is the responsibility of the analyst.

---

[9] CA records are separately identified on the 2002 NATSISS MURF.