Australian Bureau of Statistics

1352.0.55.102

Research Paper

Estimating Population Totals by Combining Household Surveys

New Issue

Research Paper

Estimating Population Totals by Combining Household Surveys

James Chipperfield, Julia Chessman
and Russell Lim

Analytical Services Branch

Methodology Advisory Committee

19 June 2009, Canberra

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) THURS 29 OCT 2009

ABS Catalogue no. 1352.0.55.102

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Mr Jonathon Khoo, Analytical Services Branch on Canberra (02) 6252 5506 or email <analytical.services@abs.gov.au>.

ESTIMATING POPULATION TOTALS BY COMBINING HOUSEHOLD SURVEYS

James Chipperfield, Julia Chessman and Russell Lim
Analytical Services

QUESTIONS FOR THE COMMITTEE

Is future work in the area of combining surveys a worthwhile objective for the Australian Bureau of Statistics?

1. Does this paper address the critical barriers to combining surveys?
2. Does this paper appropriately measure the quality of the estimates obtained from combining surveys?
3. Is there sufficient evidence to support the case study's conclusion that combining the Labour Force Survey and the National Aboriginal and Torres Strait Islander Health Survey is worthwhile?

CONTENTS

ESTIMATING POPULATION TOTALS BY COMBINING HOUSEHOLD SURVEYS

James Chipperfield, Julia Chessman and Russell Lim
Analytical Services

ABSTRACT

The Australian Bureau of Statistics is always under pressure from its clients to improve the accuracy of its estimates about the Australian population. In response to this pressure, the ABS has long exploited the potential to combine its surveys in various ways. This has typically been achieved within a design based framework but requires the assumption that *the value of a common data item, collected from the surveys which are to be combined, does not depend upon the survey in which it is collected*. This assumption is somewhat relaxed in this paper by assuming a measurement error model that relates data items from the different surveys. Inference is then over the sample design and measurement model. This paper uses diagnostics to test the validity of the measurement model which is used to combine the surveys. We describe an application of combining the Labour Force Survey and the National Aboriginal and Torres Strait Islander Health Survey to estimate employment characteristics about the Indigenous population. The findings suggest that combining these surveys is beneficial.

1. INTRODUCTION

The Australian Bureau of Statistics is always experiencing demand from its clients to improve the accuracy of its estimates about the Australian population. In response to this demand, the ABS has long exploited the potential to combine its surveys in various ways. Perhaps the most significant example of this since the late 90s is the use of the Labour Force Survey (LFS) to produce estimates of the number of households, which are in turn used as benchmarks for many ABS surveys.

Currently, combining surveys within the ABS is typically developed within a design based framework and is supported by substantial literature dealing with estimation issues. These design based applications make the assumption that *the value of a common data item, collected from the surveys which are to be combined, does not depend upon the survey in which it is collected*.

This assumption limits the range of situations in which surveys can be combined in the ABS. The main focus of this paper is addressing issues that arise from relaxing this assumption.

For instance, in the presence of *differences* in the way the survey data are collected, there is no framework for deciding whether an estimate obtained by combining surveys is more accurate than an estimate based on a single survey. Such a framework is required to answer, for example, questions like: is an estimate of employment status from the LFS alone more accurate than an estimate obtained by combining the LFS with another ABS survey, which defines employment status in a slightly different way?

This paper addresses the problem of combining non-overlapping surveys, where the data items collected by the surveys are similar but are known to have some differences. The approach taken here is to:

(i) Develop a measurement model that relates the data items from the different surveys;

(ii) Produce estimates by combining the surveys. This is done so that the estimates are unbiased over the sample design and measurement model; and

(iii) Test the validity of the measurement model and the quality of the combined estimates using a set of diagnostics.

This is a useful starting point because there are many potential applications, discussed later, for combining surveys in the ABS that require such a framework. While efforts are made to ensure consistency between survey estimates, differences in the way survey data are collected are common. These differences arise from collecting the same characteristic (e.g. employment status) using a different conceptual definition, using different data collection approaches (e.g. questionnaire design, interviewer procedures) and collecting the data within different enumeration periods.

The impact of differences in the way survey data are collected was illustrated in a 2004 ABS information paper that aimed to explain the difference between two estimates of property break-in prevalence. The General Social Survey estimate was 12% and the National Crime and Safety Survey estimate was 7.4%. The conclusions of the report were that: the sample design and selection, scope and coverage, questionnaire format and content, survey procedure and non-response were factors that contributed to the difference between the estimates; and it was not possible to measure the individual contribution of these factors to the difference between the estimates.

Section 2 summarises the ABS household survey program, gives examples of how its surveys have been combined, and briefly mentions approaches from other statistical organisations. Section 3 reviews the relevant literature. Section 4 describes a model-based framework for combining surveys. Section 5 gives a list of diagnostics for measuring the quality of the measurement model that is used to combine the surveys. Section 6 describes an application of combining the LFS and the National Aboriginal and Torres Strait Islander Health Survey (NATSIHS) to estimate employment characteristics about the Indigenous population. Section 7 suggests changes to ABS survey designs that would improve the reliability of estimates obtained from combining surveys.

2. REVIEW OF CURRENT PRACTICE

2.1 ABS household surveys

To illustrate the context of the problem we now describe ABS household surveys. The two survey vehicles for household surveys in the ABS are Special Social Surveys (SSS) and the Monthly Population Survey (MPS), both of which have multistage sample designs. The SSS and the MPS are designed to minimise overlap at the dwelling level to minimise respondent burden concerns.

A SSS will generally cover one broad subject matter in detail (e.g. health or income and expenditure), occur about every three to six years, have enumeration periods that range between three and twelve months and can have a sample size as high as 12,000 dwellings (Appendix A summarises the SSS survey program for 2008–09).

The MPS has seven out of eight of its dwellings in common for any two consecutive months – this is achieved by rotating one of the eight rotation groups each month. The MPS consists of the Labour Force Survey (LFS) and two supplementary surveys.

The LFS collects information about employment and unemployment each month and has a sample size of about 54,900 people (as of June 2008). The two MPS supplementary surveys are: the monthly supplementary survey and the Multi-purpose Household Survey (MPHS).

The monthly supplementary survey asks seven out of eight units of the LFS sample a small set of questions that aims to take less than three minutes to complete. The topics covered vary from month to month. The topics are usually employment-related but do cover other topics such as the environment (Appendix B summarises the monthly supplementary survey program for 2008–09).

The MPHS comprises one-third of the outgoing LFS rotation group (or 1/24-th of the LFS sample) each month and is designed to provide statistics annually on a small number of labour, social and economic topics. Topics for the 2007 survey were 'Environmental Views and Behaviour', 'Household Use of Information Technology', 'Personal Fraud', 'Educational Qualifications' and 'Personal and Household Income'. The annual sample for 2007 was 14,000 dwellings.

2.2 ABS examples

Examples of combining ABS household surveys can broadly be categorised into one of four types, mentioned below. The ABS has substantial experience with the range of issues (e.g. conceptual, design and estimation) arising from each type of application.

(a) Combining different cycles of the same survey

The main benefit of combining different cycles of the same survey is to reduce the sampling variability associated with survey estimates. We now mention three ABS examples:

1. LFS estimates of employment status. The data used to calculate the LFS estimates for the current month are made up of the sample from the current month and the sample from the previous seven months. The estimation procedure is an application of composite estimation (Bell, 2001).
2. Annual Indigenous Labour Force estimates. The data used to calculate the annual estimates are made up of all Indigenous records in the LFS during the period of a year.
3. LFS estimates of the number of households. These estimates are calculated by applying a smoothing filter to the monthly series of the estimated number of households. The filter is designed to reduce the volatility in the household estimates due to sampling error.

(b) Benchmarking a small survey to an estimate obtained from a large survey

One benefit of benchmarking a small survey to an estimate from a large survey is that there will be some level of consistency between the small and large survey estimates. Another benefit is a reduction in standard error. A common ABS example is benchmarking a SSS to the household estimates.

(c) Combining two different surveys of the same population

The sample for the 2004-05 National Aboriginal and Torres Strait Islander Health Survey (NATSIHS) was made up of Indigenous people selected in the 2004-05 National Health Survey (NHS) and a non-overlapping supplementary sample which was designed to targeted the Indigenous population. The NHS sample was regarded as too small to provide reliable estimates about the Indigenous population.

(d) Combining surveys to increase the scope

ABS and DoHA conducted a data pooling trial (see Kumar, 2008), combining state health survey data to produce national health estimates. The conclusion of this investigation was that “pooling of jurisdictional data is a viable proposition provided all states/territories collect and provide data according to prescribed specifications and standards”.

2.3 Potential for combining household survey data

Some potential applications of combining data include:

- (Example 1) Combine the Survey of Education and Training 2005 and the Adult Literacy and Life Skills Survey 2006 and the Survey of Education and Work 2006 to obtain more accurate education statistics, particularly for how these change over time;
- (Example 2) Combine the LFS 2006 and the Survey of Income and Housing 2005–06 to obtain more accurate labour and income statistics for population subgroups, such as low income earners;
- (Example 3) Use the Census to obtain a benchmark for use by surveys; and
- (Example 4) Combine the LFS and SSSs to obtain improved estimates of Indigenous employment status.

An interesting feature of all these potential applications is that, while the surveys to be combined collect information on the same characteristic (e.g. employment), there may be differences in the conceptual definition of the characteristic. Example 4 is explored in Section 6.

2.4 Overseas agencies: some examples of combining surveys

This subsection gives a brief summary of some ways in which overseas national statistical agencies combine survey data.

Office for National Statistics (ONS) (United Kingdom)

In 2008, the Office for National Statistics embarked on a program of integrating some of its surveys in order to, amongst other things, standardise the collection and processing of its surveys so that their estimates could be compared more reliably (see ONS, 2004). The ONS acknowledged that without integration:

> "estimates of the same variables across the different surveys cannot be combined and, despite the use of common questions, small but statistically significant differences occur between those estimates."

Statistics Netherlands

Statistics Netherlands (see Houbiers *et al.*, 2003) constructed a social statistical database from administrative and survey data containing information about individuals. Statistics Netherlands developed an estimation procedure, called repeated weighting, that ensures, as much as possible, numerical consistency between the survey estimates and improves the accuracy of estimates "due to a better use of auxiliary information".

The methodology is not well-suited to the ABS household survey situation, as it is generally not possible to link individual-level survey responses to administrative sources. While this method may have applications to ABS economic surveys, investigating this is out of scope of this paper.

Statistics Norway

Thomsen and Holmfly (1998) give an in-depth description of how survey and administrative data are combined by Statistics Norway. The data are linked at the individual level and, for the same reason as mentioned above, the methods are not well-suited to the ABS household survey situation.

Australian Bureau of Statistics

It is worthwhile pointing out here that the broad task of identifying efficiencies in the ABS' household survey program was considered in 2006 by the *Ivan King review*. The review suggested that ABS surveys collect an expanded set of core data items, covering topics such as employment status, income and education (for more information on this see Appendix C). Combining the surveys would enable population estimates on these core set of data items to be produced and used as benchmarks for an individual survey, thereby reducing the sample error.

3. REVIEW OF STATISTICAL LITERATURE

There has been a lot of work in the statistics literature on combining one or more non-overlapping surveys, which collect a common set of data items, for the purpose of estimating finite population totals. A key assumption often made is *the value of a common data item, collected from the surveys which are to be combined, does not depend upon the survey in which it is collected*.

With this key assumption, some alternatives approaches are now mentioned.

- First, if the population total for the common data item is of interest, the problem becomes one of estimation with multiple surveys from multiple frames (for example, see Hartley, 1962; Bankier, 1986; Lohr and Rao, 2000).
- Renssen and Nieuwenbroek (1997) and Merkouris (2004) consider the problem where an estimate of the population totals for the common data items is used as a benchmark for the surveys. The benefits are more reliable estimates of the population total for the common data items, more reliable survey-specific estimates, and improved consistency between the surveys' estimates resulting from the use of a common benchmark.
- Schenker and Raghunathan (2007) and Godbout and Grondin (2005) model the relationship been the common data items and the data items of interest from one survey. They then apply the model to obtain an imputed value for the data items of interest for the other survey.

4. SITUATIONS FITTING INTO THE FRAMEWORK

This paper will assume that there are only two surveys to be combined, though extensions to combining more than two surveys are straightforward. The two surveys are denoted as A and B. Survey A collects data item y and Survey B collects data item x.

Surveys A and B are primarily designed to estimate $Y = \sum_{i \in U_A} y_i$ and $X = \sum_{i \in U_B} x_i$ respectively, where U_A and U_B are the population of units in scope of surveys A and B respectively.

Survey A's estimate of population total $Y = \sum_{i \in U_A} y_i$ is denoted by $\hat{Y} = \sum_{i \in s_A} w_{Ai} y_i$ where s_A is the sample from U_A and w_{Ai} is the weight for the i-th unit in Survey A. Similarly for Survey B define X, U_B, $\hat{X}$ and s_B.

We define two mutually exclusive domains for the population U_A – the population common to surveys A and B defined as $U_C = U_A \cap U_B$ and the non-common population defined as $U_{\tilde{C}} = U_A \cap \tilde{U}_B$, where $\tilde{U}_B$ is the complement of U_B.

We denote the population totals for the population U_C by X_C and Y_C. The samples for surveys A and B falling into U_C are denoted by $s_{A,C}$ and $s_{B,C}$ respectively. The sample for Survey A falling into $U_{\tilde{C}}$ is denoted by $s_{A,\tilde{C}}$.

The aim is to improve upon the accuracy of $\hat{Y}$ using Survey B's sample, $s_{B,C}$. It is only useful to exploit the information collected by Survey B if there is a strong and identifiable relationship between y and x. The framework in this paper allows for three such relationships:

Case 1: Data item x on Survey B can be deterministically mapped to y on Survey A. This situation would arise, for example, if x is detailed employment status (long-term unemployed, short-term unemployed, not-in-the-labour force, full-time employed and part-time employed) and y is standard employment status (employed, unemployed and not-in-the-labour force). For example, a response to x in Survey B of either long term or short term unemployed is treated as a response to y in Survey A of unemployed.

Case 2: Data item y can be stochastically mapped to x. An example of a stochastic mapping is when 90%, 8% and 2% of people reported as employed in Survey A would have reported as employed, unemployed and not-in-the-labour force, respectively, if enumerated by Survey B. Such a stochastic model would generally be identified from a sample where and x and y were jointly observed.

Case 3: Data item x can be stochastically mapped to y (the reverse of Case 2).

The estimation method for problems for Cases 1, 2 and 3 above, are discussed in Sections 4.1, 4.2 and 4.3 respectively.

4.1 Case 1: x can be deterministically mapped to y

We denote the mapped variable by $y^* = f(x)$. After the deterministic mapping has been applied, the variable y is effectively available to surveys A and B. The measurement model is simply

$$y_i^* = y_i . \tag{M1}$$

This means that y_i^*, the mapped variable for unit i in Survey B, is equal to y_i, the value for y that would have been obtained from unit i if it was enumerated by Survey A.

The problem then becomes one of estimation with multiple surveys that have been selected from multiple frames. One such estimator for Y, given by Hartley (1962), is:

$$\hat{Y}^{(1)} = \hat{Y}_{\tilde{C}} + \hat{Y}_C^{(1)}$$

where $\hat{Y}_{\tilde{C}} = \sum_{i \in s_{A\tilde{C}}} w_{Ai} y_i$ is an estimate of the non-common population total $Y_{\tilde{C}} = \sum_{i \in U_{\tilde{C}}} y_i$,

and $$\hat{Y}_C^{(1)} = \alpha \hat{Y}_C + (1-\alpha) \hat{Y}_C^*$$

where $\hat{Y}_C = \sum_{i \in s_{A,C}} w_{Ai} y_i$ and $\hat{Y}_C^* = \sum_{i \in s_{B,C}} w_{Bi} y_i^*$ are estimates of the common population total Y_C

and $$\alpha = \mathrm{Var}\left(\hat{Y}_C^*\right) \left[\mathrm{Var}\left(\hat{Y}_C\right) + \mathrm{Var}\left(\hat{Y}_C^*\right) \right]^{-1}$$

This choice of α minimises $\mathrm{Var}(\hat{Y}_C^{(1)})$. Hartley (1962) requires a different α for each data item, which means a sample unit will have a different weight for each data item. Using different sample weights for each population estimate may compromise any comparisons made between them. Lohr and Rao (2000) suggest an alternative approach that uses a single weight, thereby avoiding this problem. The Jackknife estimator (see Shao and Wu, 1995) can be used to estimate $\mathrm{Var}(\hat{Y}^{(1)})$.

4.2 Case 2: y can be stochastically mapped to x

Almost all data items collected by ABS household surveys are categorical. Assume that x and y are categorical variables with J and K categories, where the $x = 1, .., J$ and $y = 1, .., K$.

We define $x_{ij} = 1$ if $x_i = j$ and $x_{ij} = 0$ otherwise, and $X_{Cj} = \sum_{i \in U_{AC}} x_{ij}$.

The measurement model, denoted by ξ, is:

$$\begin{aligned} \Pr_{\xi}\left(X_{ij} = 1 \middle| y_i\right) &= \pi_{ij} = \hat{x}_{ij} \\ \mathrm{Var}_{\xi}\left(X_{ij} \middle| y_i\right) &= \sigma_{ij}^2 = \pi_{ij}(1-\pi_{ij}) \end{aligned} \tag{M2}$$

where π_{ij} is the probability that unit i, with a response of y_i in Survey A, would have reported $x = j$ if it were selected in Survey B.

After applying this model to respondents of Survey A, the variable x is effectively available from both surveys A and B while the variable y is available from only Survey A. Hidiroglou (2001) refers to this design as a non-nested two-phase sample design. Here we treat this as a classical two-phase design so that the standard two-phase estimator applies (see for example, Särndal, Swensson, and Wretman, 1992).

Accordingly, the estimator for Case 2 is

$$\hat{Y}_k^{(2)} = \hat{Y}_{\tilde{C}k} + \hat{Y}_{Ck}^{(2)}$$

where $$\hat{Y}_{\tilde{C}k} = \sum\nolimits_{i \in s_{A\tilde{C}}} w_{Ai} y_{ik}$$

and $$\hat{Y}_{Ck}^{(2)} = \sum\nolimits_{i \in s_{AC}} w_{Ai}^{*} y_{ik}.$$

w_{Ai}^{*} is obtained by minimising

$$\sum\nolimits_{i \in s_{A,C}} \left(w_{Ai}^{*} - w_{Ai}\right)^2 w_{Ai}^{-2}$$

subject to the constraint that

$$\sum\nolimits_{i \in s_{A,C}} w_{Ai}^{*} x_{ij} = \hat{X}_{A \cup B, Cj} \text{ for all } j.$$

$$\begin{aligned} \hat{X}_{A \cup B, Cj} &= \gamma \hat{X}_{A,Cj}^{*} + (1-\gamma)\hat{X}_{B,Cj}, \\ \hat{X}_{A,Cj}^{*} &= \sum\nolimits_{i \in s_{A,C}} w_{Ai} \hat{x}_{ij}, \\ \hat{X}_{B,Cj} &= \sum\nolimits_{i \in s_{B,C}} w_{Bi} x_{ij}, \end{aligned}$$

and $$\gamma = \mathrm{Var}_{s\xi}\left(\hat{X}_{B,Cj_0}\right)\left[\mathrm{Var}_{s\xi}\left(\hat{X}_{B,Cj_0}\right) + \mathrm{Var}_{s\xi}\left(\hat{X}_{A,Cj_0}\right)\right]^{-1}$$

is a constant, and j_o takes a particular value of j.

We now define the terms in the expression for γ.

It is easy to show that

$$E_{s\xi}\left(\hat{Y}_{Ck}^{(2)}\right) = Y_{Ck} \ ,$$

which means the estimate is unbiased jointly over the measurement model ξ and sample design, s. From the independence of the sampling process and the measurement model, it follows that

$$\text{Var}_{s\xi}\left(\hat{Y}_{Ck}^{(2)}\right) = \text{Var}_{s}\left(\hat{Y}_{Ck}^{(2)}\right) + \text{Var}_{\xi}\left(\hat{Y}_{Ck}^{(2)}\right) \tag{1}$$

where the first term is the variance due to the sampling error and the second term is due to the uncertainty due to the measurement model. (Särndal, 1992, gives a full description of (1) as well as the underlying assumptions required for it to be valid). The term $Var_s(\hat{Y}_{Ck}^{(2)})$ can be estimated with a standard jackknife estimator where the response values π_{ij} are fixed (i.e. treated as is they were reported values) and the term $Var_{\xi}(\hat{Y}_{Ck}^{(2)})$ can be estimated by the bootstrap (see Rao and Wu, 1988)

$$\text{Var}_{\xi}\left(\hat{Y}_{Ck}^{(2)}\right) = B^{-1}\sum\nolimits_{b=1}^{B}\left(\hat{Y}_{Ck}^{(2)}(b) - \hat{Y}_{Ck}^{(2)}\right)^2$$

where $\hat{Y}_{Ck}^{(2)}(b)$ is an estimator with the same form as $\hat{Y}_{Ck}^{(2)}$ except that π_{ij}in $\hat{X}_{A,Cj}^{*}$ is replaced by $\pi_{ij}(b)$, and $\pi_{ij}(b)$ is the b-th independent outcome of a binomial distribution with parameter π_{ik} (e.g. $\pi_{ij}(b=1)=1,\ \pi_{ij}(b=2)=0, \ldots$ etc.).

4.3 Case 3: x can be stochastically mapped to y

Again here we assume that x and y are categorical variables, where $y_{ik}=1$ if $y_i = k$ and $y_{ik}=0$ otherwise and $Y_{Ck} = \sum_{i\in U_{AC}} y_{ki}$. The measurement model here is:

$$\begin{aligned} \Pr_{\xi}\left(Y_{ik} = 1 \middle| x_i\right) &= \pi_{ik} = \hat{y}_{ik} \\ \text{Var}_{\xi}\left(Y_{ik} \middle| x_i\right) &= \sigma_{ik}^2 = \pi_{ik}(1-\pi_{ik}) \end{aligned} \tag{M3}$$

where π_{ik} is the probability that unit i, with a response of x_i in Survey B, would have reported $y_i = k$ if it was selected in Survey A.

Estimation is in two steps:

- Step 1. Estimate Y_{Ck} from Survey A using $\hat{Y}_{Ck} = \sum_{i\in s_{AC}} w_{Ai} y_{ki}$ and estimate Y_{Ck} from Survey B using $\hat{Y}_{Ck}^{**}$ (defined below); and
- Step 2. Combine the estimates $\hat{Y}_{Ck}$ and $\hat{Y}_{Ck}^{\#\#}$ in an optimal way.

Step 1

Consider an estimate of Y_{Ck} from Survey B, given by $\hat{Y}_{Ck}^{\#\#} = \Sigma_i w_i \hat{y}_{ik}$.
It is easy to show that

$$E_{s\xi}\left(\hat{Y}_{Ck}^{\#\#}\right) = Y_C \ ,$$

which means the estimate is unbiased jointly over the measurement model ξ and sample design, s. From the independence of the sampling process and the measurement model (see Särndal, 1992), it again follows that

$$\mathrm{Var}_{s\xi}\left(\hat{Y}_{Ck}^{\#\#}\right) = \mathrm{Var}_s\left(\hat{Y}_{Ck}^{\#\#}\right) + \mathrm{Var}_\xi\left(\hat{Y}_{Ck}^{\#\#}\right)$$

where the first term is the variance due to the sampling error and the second component reflects the uncertainty due to the measurement model. The term $Var_s(\hat{Y}_{Ck}^{\#\#})$ can be estimated with a standard jackknife estimator where $\hat{y}_{ik}$ is fixed (i.e. $\hat{y}_{ik}$ is treated as if it was the reported value y_{ik}) and the term $Var_\xi(\hat{Y}_{Ck}^{\#\#})$ can be estimated by $V\hat{a}r_\xi(\hat{Y}_{Ck}^{\#\#}) = \Sigma_{i\in s_B} w_i^2 \sigma_{ki}^2$, which assumes the sample fraction is negligible.

Step 2

Again, following the same development for the estimator $\hat{Y}^{(1)}$, we define

$$\hat{Y}_k^{(3)} = \hat{Y}_{\tilde{C}k} + \hat{Y}_{Ck}^{(3)}$$

where $$\hat{Y}_{Ck}^{(3)} = \psi\, \hat{Y}_{Ck} + (1-\psi)\, \hat{Y}_{Ck}^{\#\#}$$

and $$\psi = \mathrm{Var}\left(\hat{Y}_{Ck_0}\right)\left[\mathrm{Var}\left(\hat{Y}_{Ck_0}\right) + \mathrm{Var}\left(\hat{Y}_{Ck_0}^{\#\#}\right)\right]^{-1}$$

is a constant that minimises $Var_{s\xi}(\hat{Y}_{Ck_o}^{(3)})$ and k_o is a particular value of k. Given surveys A and B are independent, the variance of $\hat{Y}_k^{(3)}$ can be obtained by noting that

$$\mathrm{Var}_{s\xi}\left(\hat{Y}_k^{(3)}\right) = \mathrm{Var}_s\left(\hat{Y}_{\tilde{C}k}\right) + \psi^2 \mathrm{Var}_s\left(\hat{Y}_{Ck}\right) + (1-\psi)^2 \mathrm{Var}_{s\xi}\left(\hat{Y}_{Ck}^{\#\#}\right)$$

where $Var_s(\hat{Y}_{Ck})$ can be estimated by the jackknife.

5. DIAGNOSTICS AND QUALITY INDICATORS

This section discusses some diagnostics that can be used to assess the quality of the estimates obtained from Section 4.

5.1 Small area estimate diagnostics

Comparing model based estimates for small areas with design based estimates is often used to test whether they are consistent. This paper suggests three such diagnostics from Brown *et al.* (2001) to determine whether the estimates from surveys A and B are consistent, conditional on a measurement model, over a set of domains $d = 1, 2, ..., D$.

Case 1 assumes there are no differences between y, the data item of interest collected from Survey A, and $y^*(x)$, the data item available to Survey B (see model M1). To test this assumption we compare the estimates of Y_d from Survey A, given by $\hat{Y}_d = \Sigma_{i \in s_{Bd}} w_{Ai} y_i$, and from Survey B, given by $\hat{Y}_d^* = \Sigma_{i \in s_{Ad}} w_{Ai} y_i^{\#}$ where s_{Ad} and s_{Bd} denote the sample in surveys A and B falling in domain d. If the assumption is correct then:

#1. the regression of $\sqrt{\hat{Y}_d}$ against $\sqrt{\hat{Y}_d^*}$, given by $\sqrt{\hat{Y}_d} = a + b\sqrt{\hat{Y}_d^*}$, will give $\hat{a} = 0$ and $\hat{b} = 1$. The square root transformation aims to stabilitise the variance structure so that the assumption of a homogenous error structure is valid.

#2. the distribution of $F_d = \left(\hat{Y}_d - \hat{Y}_d^*\right)\left[SE\left(\hat{Y}_d - \hat{Y}_d^*\right)\right]^{-1}$ will follow a t-distribution.

#3. the percentage of times that $\hat{Y}_d^*$ and $\hat{Y}_d$ are statistically different at the 95% significance level will be close to 5%.

If conditions #1, #2 or #3 do not hold then it suggests that the model M1 is not true. The same set of diagnostics may be used to test the models M2 and M3 underlying Case 2 and Case 3 respectively.

5.2 Survey effect diagnostic

The survey effect diagnostic attempts to identify whether the value of a data item depends upon whether it was collected from Survey A or Survey B, conditional on the measurement model. For Case 1 this diagnostic involves:

#4. pooling data from surveys A and B and regressing r_i, defined as $r_i = y_i$ if i belongs to Survey A and $r_i = y_i^*(x_i)$ if i belongs to Survey B, against:

- a survey indicator that identifies whether unit i was selected in Survey A or Survey B;
- a set of covariates that are common to surveys A and B; and

- a set of design variables that explain the sample selection process for both surveys A and B.

If the coefficient of the survey indicator variable is statistically significant, then it suggests the measurement model for Case 1 does not explain all the differences between x_i and y_i.

Including the design variables in the model ensures that the effects of the sampling process are not confounded with the effects due to the survey indicator. For example, consider if a remoteness index is correlated with employment in the population and that remote areas are over-represented by Survey A. One of the ways to remove the effects of the sampling process is to include a remote indicator as an auxiliary variable in the model. Another way simply involves using a weighted analysis, where the weight is the inverse of the selection probability, thereby ensuring a valid design based interpretation of the model parameters (Chambers and Skinner, 2003).

It is not straightforward to calculate diagnostic #4 for Cases 2 and 3 – this is because x_i and y_i are random variables (i.e. not observed), respectively. While it is beyond the scope of this paper, such analysis could be obtained within a missing data framework (see Little and Rubin, 2002).

5.3 Movement estimate diagnostic

Previously, this paper has considered population estimates for a given point in time. However, measures of change are often of particular interest. This suggests the question: if one of the estimators in this paper is used for multiple time points, what can we say about the quality of the movement estimates?

The sampling error associated with the movement estimates can readily be measured. However, the bias associated with movement estimates is very difficult to measure. Nevertheless, it is useful to consider the nature of the bias for movement estimates, at least theoretically. This means we should address the following question:

#5. What is the bias on movement estimates between two time points if the measurement model is wrong?

We may consider the situation where the measurement model we have specified is wrong and the true measurement model does not change over time. In this situation, the bias on the point-in-time estimates is constant over time and, consequently, the bias on the movements is zero.

5.4 Mean squared error

Another diagnostic involves testing the sensitivity of the Mean Squared Error (MSE) of an estimate obtained by combining surveys A and B to misspecification of the measurement models M1, M2 and M3.

Denote $\hat{Y}^{(p)}$ to be one of the estimators given in Section 4 where $p = 1, 2, 3$. The sensitivity diagnostic involves:

#6. Plotting the distribution of the Mean Squared Error of $\hat{Y}^{(p)}$, given by

$$\text{MSE}\left(\hat{Y}^{(p)}\right) = \text{Var}\left(\hat{Y}^{(p)}\right) + \text{Bias}^2\left(\hat{Y}^{(p)}\right)$$

for a range of different values of $Bias(\hat{Y}^{(p)})$.

We now illustrate how to obtain an expression for $\text{Bias}(\hat{Y}^{(1)})$ for Case 1.

For simplicity, here we assume that surveys A and B have a common scope and so drop the subscript "c". Now consider if the true measurement model for Case 1 is not given by model M1 but is in fact given by $E_{\xi}(y_i^*) = y_i + b$, where b represents the unknown misspecification.

It is easy to show that

$$\text{Bias}_{\xi^s}\left(\hat{Y}^{(1)}\right) = E_{\xi^s}\left(\hat{Y}_C^{(1)}\right) - Y = (1-\alpha)B$$

where $B = \sum_{i \in U} b$. The bias is a function of B and α, where α is given in Section 4.1. The idea is to appreciate the sensitivity of the MSE to B, which is unknown.

Combining the surveys is beneficial only as long as $MSE(\hat{Y}^{(p)})$ is less than the variance of the corresponding estimate obtained from Survey A.

It would be desirable to obtain a direct estimate of the MSE. Elliott and Davis (2005, see p. 605) suggest an estimate of the MSE for small domains. However, this estimate of the MSE can be volatile and requires an *ad hoc* adjustment to ensure it is positive.

6. CASE STUDY

6.1 Introduction

This case study uses the 2004–05 National Aboriginal and Torres Strait Islander Health Survey (NATSIHS) to potentially improve the Labour Force Survey's estimates of employment status for the Indigenous population.

The LFS estimates for the Indigenous population are obtained by pooling 12 months of survey data which amounts to a total sample size of 12,000 Indigenous records. The LFS estimates are published annually. The NATSIHS, with a sample size of 6,325 records, does not publish estimates of employment status. This is because there is concern about the coherence of the LFS and NATSIHS employment estimates. However, NATSIHS employment estimates are provided, upon request, to ABS clients.

There are a range of differences between these surveys in terms of the coverage, weighting, sample design, and the conceptual definition of employment status. As mentioned in Sections 1–3, it is important that the impact of these differences is minimised in order to reliably combine these surveys. See Appendix D for a detailed description of these differences and efforts made to correct for them. For example, the NATSIHS covered the period August 2004 to July 2005; for the purposes of this paper, the LFS data were pooled over the same period.

In the notation of Section 4, the LFS is Survey A and NATSIHS is Survey B and both surveys are assumed to have a common coverage and scope so that $U_A = U_B$. Section 6.2 describes the measurement models M1, M2 and M3 which are used to correct only for conceptual differences between the LFS and NATSIHS definition of employment status. The measurement models do not adjust for a range of non-sampling factors such as non-response bias, interviewer effects, contextual effects, differences in survey coverage, and different enumeration periods. Section 6.3 gives the estimates motivated from models M1, M2 and M3. Section 6.4 discusses the diagnostics for the estimates motivated under model M1.

6.2 Measurement models

Here we consider measurement models M1, M2 and M3 to explain the relationship between the LFS and NATSIHS conceptual definitions of employment status.

The LFS and the NATSIHS both have two forms (four forms in total):

(a) the long form designed for a majority of persons; and

(b) the short form designed for Indigenous people, typically living in communities.

We assume that a respondent who was given a short form in the LFS would have been given the NATSIHS short form, if they were enumerated in the NATSIHS (and *vice versa*). This assumption is reasonable in practice. It also means that we only need two measurement models: one to explain the conceptual difference between the LFS and NATSIHS short forms and another to explain the conceptual differences between the LFS and NATSIHS long forms.

Case 1

The model M1 in Section 4 assumes that there are no conceptual differences between the LFS and NATSIHS's definition of employment status. This means $x_i = y_i$.

Case 2

Model M2 is a model that predicts the probability that a person would have been classified as employed / unemployed / not-in-the-labour force (NILF) by the NATSIHS, conditional on their LFS employment status.

Mapping from the LFS long form to the NATSIHS long form is deterministic (details are complex and are omitted here). This is because the LFS long form collects more detailed information that the NATSIHS long form. For example, for respondents who are currently away from work, employment status is a function of the length of time away for the LFS but not for the NATSIHS.

Table 6.1 shows the mapping from the LFS short form to the NATSIHS short form. This mapping is stochastic (i.e. not deterministic). This is because the LFS short form collects less detailed information than the NATSIHS short form. Namely, if a respondent to the LFS short form has

> "been looking for work in the last four weeks and has taken steps to find work"

then they are classified as unemployed. A respondent to the NATSIHS would need to be asked the additional question

> if [they] had found a job *could* [they] have started work last week?

before they could be classified as either unemployed or NILF. From the NATSIHS data, 81 out of 94 answered 'yes' to this additional question.

There were 151 people who were classified as unemployed by the LFS short form. The estimated probability that a person, classified as unemployed by the LFS, would have been classified as unemployed by the NATSIHS is 81/94.

6.1 Conceptual mapping of the LFS Short Form to the NATSIHS Short Form

LFS		*NATSIHS*		
Status	*Sample count*	*Description*	*Probability of NATSIHS status*	*Status*
Employed	476	Actually worked one hour or more in a job last week OR Usually work one hour or more and had a job that were away from	1	Employed
Unemployed	151	Have looked for work in the last four weeks and have taken steps to find work and if had found a job could have started work last week	81/94*	Unemployed
		Have looked for work in the last four weeks and have taken steps to find work and if had found a job don't know or could not have started work last week	13/94*	NILF
Not in the Labour Force (NILF)	2	Actually worked less than one hour in a job last week	1	Employed
	0	Usually work less than one hour and had a job that were away from	1	Employed
	1,018	Other than above	1	NILF

Case 3

The model M3 in Section 4 gives a stochastic model that predicts the probability that a person would have been classified as employed / unemployed / NILF by the LFS conditional on their NATSIHS employment status.

Mapping from the NATSIHS long form to the LFS long form is stochastic (details are omitted here). This is because, as mentioned above, the LFS long form collects more detailed information than the NATSIHS long form.

Table 6.2 shows the mapping from the NATSIHS short form to the LFS short form. This mapping is deterministic due to the NATSIHS long form collecting more detailed information than the LFS short form.

6.2 Conceptual mapping of the NATSIHS Short Form to the LFS Short Form

NATSIHS		*LFS*		
Status	*Sample count*	*Description*	*Sample count*	*Status*
Employed	754	Actually worked one hour or more in a job last week	682	Employed
		Actually worked less than one hour in a job last week	18	NILF
		Usually work one hour or more and had a job that were away from	53	Employed
		Usually work less than one hour and had a job that were away from	1	NILF
Unemployed	81	Have looked for work in the last four weeks and have taken steps to find work and if had found a job could have started work last week	NA	Unemployed
NILF	610	Have looked for work in the last four weeks and have taken steps to find work and if had found a job don't know or could not have started work last week	13	Unemployed
		Other than above	597	NILF

6.3 Estimates

The Australian level estimates for Indigenous employment status for Case 1, 2 and 3 are given in table 6.3 and the corresponding Relative Standard Error (RSE) are given in table 6.4. Table 6.4 shows that the RSEs for Case 1, 2 and 3 estimators are smaller than the corresponding LFS and NATSIHS RSEs. This highlights the benefits of combining surveys in order to reduce the sample error.

Given the RSEs, we can see from table 6.3 that the differences between the NATSIHS and LFS estimates of employment status are not statistically significant at the 95% level. The RSEs for the estimates for Case 1, 2 and 3 are also very similar. This suggests that the impact of adjusting for conceptual differences between the surveys is small.

6.3 Estimates (proportion) of Employment status*

Status	*LFS*	*NATSIHS*	*Case 1*	*Case 2*	*Case 3*
Employed	47.1%	49.1%	48.6%	48.3%	48.4%
Unemployed	8.9%	8.9%	8.8%	9.1%	9.3%
NILF	44.0%	42.0%	42.6%	42.6%	42.3%

6.4 Relative standard error of estimates of Employment status*

Status	*LFS*	*NATSIHS*	*Case 1*	*Case 2*	*Case 3*
Employed	4.0%	2.4%	2.1%	2.2%	2.1%
Unemployed	7.5%	6.7%	5.1%	5.4%	4.9%
NILF	3.8%	2.6%	2.2%	2.2%	2.2%

* The values of α, γ, and ψ are all calculated at the state level.

The RSEs for Case 1 at the State by Remoteness level are provided in Appendix E. Again, these RSEs are considerably lower than the RSEs from the LFS alone.

6.4 Diagnostics for Case 1

Given the apparent small differences between the estimates for Cases 1, 2 and 3, here we only consider diagnostics for Case 1, where it is assumed that the value of individual's employment status does not depend upon whether they are enumerated by the NATSIHS or the LFS.

Small area estimate diagnostics

The small area diagnostics, listed below, test the assumption that the variables to be combined are equivalent, which is the model M1 in Section 4. No evidence was found against this assumption.

Regression

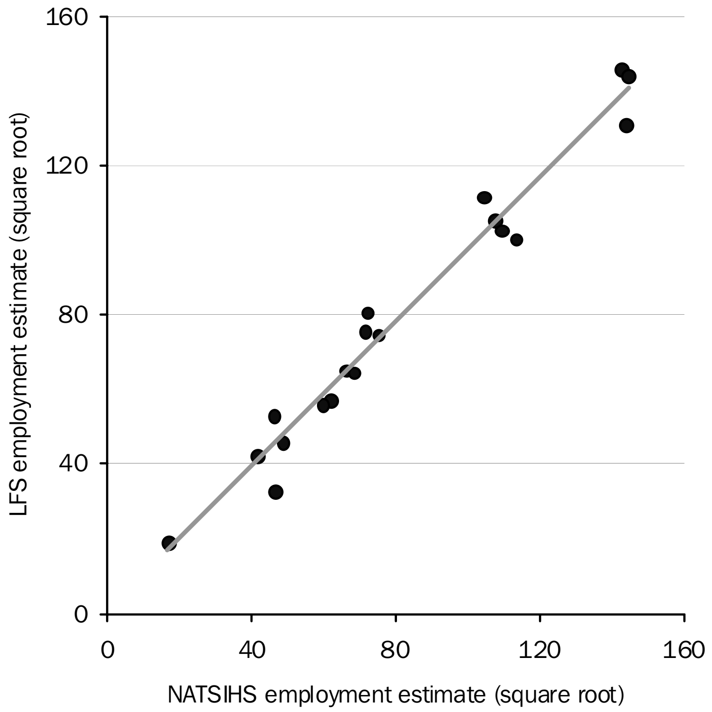
The results of the regression diagnostics are given in table 6.5. They show that there is no evidence to reject the hypothesis that the NATSIHS and LFS estimates are not related by a model with a zero intercept and unit slope. Figure 6.6 plots the LFS' and NATSIHS' estimates of employed persons, after the square root transformation, at the state by remoteness level. The plot shows a strong linear relationship.

6.5 Regression of LFS and NATSIHS estimates

	Intercept		*Slope*	
	Estimate	*p-value*	*Estimate*	*95% CI*
Employed	1.02	0.79	0.97	(0.88 , 1.05)
Unemployed*	6.04	0.09	0.88	(0.74 , 1.01)
NILF	2.62	0.57	0.99	(0.88 , 1.10)

* Points corresponding to estimates with RSEs greater than 35% were excluded from the regression.

6.6 LFS and NATSIHS employment estimates (square root) at the state by remoteness level with line of best fit



Hypothesis testing

There are 19 state by remoteness levels. Consider the null hypothesis that the difference between the NATSIHS and LFS estimates is only due to sample error. Under this hypothesis we would expect about one out of twenty of the estimates at the state by remoteness level to be statistically different at the 95% significance level. We found one out of 19 estimates was statistically different for estimates of employment, unemployment, and NILF. Accordingly, there is little evidence to reject the null hypothesis.

Survey effect diagnostic

There were four different models for the odds of being employed: a model for Long Form Major City, Long Form Regional, Long Form Remote and Short Form Remote. All models include a survey indicator, which takes the value of "1" if the unit is in the LFS and "0" if the unit is in the NATSIHS. Other independent variables

- in the Short Form regression include age, sex, and whether attending school; and
- in the Long Form regression include age, sex, whether attending school, whether married, whether English is the main language at home, and whether currently studying full time.

The results in table 6.7 show, for the odds of being employed, that there is evidence of a survey effect due to the Long form in regional areas and due to the Short form in remote areas. The size of the survey effect has a significant impact on the prediction, at least for some sub-populations. For example, the survey effect model predicts that females who are older than 45 years, not at school and living in remote areas have a 37% and 24% chance of being employed if enumerated by the NATSIHS and LFS respectively.

6.7 Coefficient of the survey effect for the odds of being employed / unemployed

		Employed		*Unemployed*	
Form	*Degree of remoteness*	*Estimate*	*p-value*	*Estimate*	*p-value*
Long	Major City	-0.07	0.630	0.54	0.007
Long	Regional	-0.38	0.004	0.20	0.220
Long	Remote	-0.47	0.160	-0.17	0.590
Short	Remote	-0.63	0.043	-0.15	0.700

* All model parameters are estimated using weights. The corresponding variances are estimated using a Taylor Series technique described in Binder (1983) that allows for the complex sample designs of the LFS and NATSIHS. The method was implemented using PROC SURVEYLOGISTIC in SAS.

Other independent variables could be included to further isolate whether the survey effects apply, for example, to a particular age category. This could be achieved by creating a variable that takes the value "1", if an individual is older than 45 *and* enumerated by the LFS, and "0" otherwise.

Movement estimate diagnostic

Change in the employment characteristics of the Indigenous population, as measured by the LFS, is of strong interest. The difference between Case 1 estimates at two time points will be an unbiased estimate of change as long as:

- α defined in Section 4.1 is held constant at both time points; and
- The true measurement model is the same at both time points.

This will be the case even if working measurement model, used to combine the surveys, is wrong. This suggests it is worthwhile keeping α constant when calculating a series of Case 1 estimates.

Mean squared error

Define $Ratio_{sr}$ as the ratio of the Mean Squared Error, defined in Section 5.4, for the Case 1 estimates and the variance of the LFS estimates, for state s and remoteness r. If $Ratio_{sr} < 1$ then the Case 1 estimate, which combines the NATSIHS and LFS, is more accurate than the estimate based on the LFS alone. Table 6.7 gives the median, tenth percentile and 90-th percentile, for $Ratio_{sr}$ across the 19 state by remoteness levels.

As the definition of MSE requires knowledge of the unknown bias, defined here as the difference in the population totals for the NATSIHS and LFS (i.e. $B = X - Y$), we consider various values of the bias. The bias in table 6.8 is measured in percentage terms: a 1% bias means the NATSIHS and LFS population totals are different by 1%.

Table 6.8 shows that the bias for employment and unemployment needs to be greater than 15% and 30% (from the median) for the Case 1 estimator to be less accurate than the corresponding estimator based on the LFS alone. This means that even a moderate bias is off-set by the substantial reductions in the sample error.

6.8 Ratio of the MSE for the Case 1 estimator and the variance of the LFS estimator

		$Ratio_{sr}$		
	Bias (%)	*10%*	*Median (50%)*	*90%*
Employment	0%	0.16	0.44	0.68
	5%	0.17	0.52	0.80
	10%	0.21	0.70	1.25
	15%	0.26	1.00	2.00
Unemployment	0%	0.15	0.36	1.00
	10%	0.16	0.44	1.06
	20%	0.19	0.61	1.32
	30%	0.24	0.93	1.82

6.5 Summary

There is evidence that the value of employment status depends upon whether it is collected from the NATSIHS or LFS (see the *survey effect diagnostic*) for some sub-populations. However, there is evidence that these differences are likely to be small (see *small area diagnostics*). Moreover, even if these differences are moderate there is still substantial benefit in combining the surveys (see *MSE diagnostic*).

Publishing the estimates for Case 1 would mean an associated MSE would also need to be provided. Calculating the MSE requires an estimate of the bias, which is difficult to estimate. One approach is to calculate an estimate of the bias at a high level and assume it applies at the finer levels. Using this approach, the bias at the Australian level would be 4% for employment estimates and 0.7% for unemployment.

The diagnostics to assess the quality of the estimates obtained in Case 2 and 3 will be completed soon.

7. IMPLICATIONS FOR DATA COLLECTION

A key question is: how could we change ABS survey designs to improve the quality of the estimates obtained from combining surveys? The broad approaches, in decreasing order of desirability, are

I. Collect the data items, in the surveys to be combined, in the same way. This means the conceptual definitions, the form design and interviewer procedures are effectively the same across the surveys. This is not a realistic general solution, but this may be practical in some situations.

II. Allow sample overlap between the surveys that are to be combined, even if this is small, as it removes the need for measurement models and model based assumptions. Chipperfield and Steel (2009) suggest an approach to designing the overlap between the surveys to be combined. This is an application of However, this does not rule out contextual effects (i.e. a response may depend upon whether a respondent is in the overlapping sample). The issue of contextual effect would need to be addressed, but there are many examples in the ABS where contextual effects are assumed to be negligible.

III. Conduct a small survey to obtain an empirical estimate of the measurement model. This small survey would need to collect the data items in the measurement model from the same individual. The measurement model can be assumed to be fixed and apply into the future. This is a less desirable version of (II) because it relies on model assumptions.

IV. Design survey questions so that the conceptual mapping between the surveys' data items is straightforward to obtain. This approach will not always work because the difference between surveys' data items cannot be completely explained by their conceptual definitions alone. For example, the context in which a question is asked may have a significant influence on the response, especially for sensitive questions.

V. Include a core set of variables for some, if not all, surveys and use it as a "first phase". This was essentially a suggestion from the *Ivan King review*.

NATSIHS / LFS

We now consider two practical changes that would improve the reliability of estimates obtained from combining the NATSIHS and the LFS.

First, Indigenous people who have responded to the LFS, and are also due to be rotated out of sample next month, could be also given the SSS employment module. There is no risk of the SSS module affecting the response to the LFS module if the former is sequenced after the latter, during the interview. This change may be practical for Indigenous people who are enumerated with the long form, but may not be practical for Indigenous people who are enumerated with the short form.

We consider an application of approach (IV) to improve the mapping from the LFS to the NATSIHS (referred to as Case 2). As mentioned in Section 6, the LFS long form currently collects more detailed information than the NATSIHS long form, which means the mapping from the former to the latter is deterministic. For the mapping from the LFS short form to the NATSIHS short form (see Case 2, Section 6.2) to be deterministic, the following additional information would need to be collected from the LFS:

- If have looked for work in the last four weeks and have taken steps to find work:

  If you had found a job, could you have started work last week?

8. CONCLUSIONS

This paper develops a framework for combining surveys when the survey data items are related via a measurement model. Measurement models are best obtained when the data items in the model are observed on the same individual. This situation rarely occurs in the ABS since household surveys are non-overlapping by design (i.e. an individual is usually selected in only one survey at a time). As a result, the measurement models considered in this paper's case study only explain conceptual differences between the survey data items.

The case study illustrated an application of the framework by combining the Labour Force Survey and the National Aboriginal and Torres Strait Islander Health Survey to estimate employment characteristics about the Indigenous population. A key concern was misspecification in the measurement error model. This concern was somewhat alleviated through a set of diagnostics which showed that, while there was some misspecification in the measurement model, combining the surveys was worthwhile.

This paper argues that, with relatively small changes to the sample designs, small overlap between surveys can be achieved. This overlapping sample could be used to develop an improved measurement model, in that it incorporates all aspects of the measurement process (i.e. not just conceptual differences), or it could be used to remove the need for a measurement model at all, through the application of standard design based estimation.

REFERENCES

Australian Bureau of Statistics (2004) *Measuring Crime Victimisation, Australia: The Impact of Different Collection Methodologies*, Information Paper, cat. no. 4522.0.55.001, ABS, Canberra.

—— (2006) *Ivan King Review*, Internal Discussion Paper.

Chipperfield, J.O. and Steel, D.G. (2009) "Design and Estimation for Split Questionnaire Surveys", *Journal of Official Statistics*, 25(2), pp. 227–244.

Kumar, A. (2008) *Data Pooling Trial*, Internal discussion paper, Australian Bureau of Statistics, Canberra.

Bell, P. (2001) "Comparison of Alternative Labour Force Survey Estimators", *Survey Methodology*, 27(1), pp. 15–30.

Binder, D.A. (1983) "On the Variances of Asymptotically Normal Estimators from Complex Surveys", *International Statistical Review*, 51, pp. 279–292.

Brown, G.; Chambers, R.; Heady, P. and Heasman, D. (2001) "Evaluation of Small Area Estimation Methods – An Application to Unemployment Estimates from the UK LFS", *Statistics Canada International Symposium Series – Proceedings*, cat. no. 11-522-XIE.

Chambers, R.L. and Skinner, C.J. (2003) *Analysis of Survey Data*, John Wiley & Sons.

Elliott, M.R. and Davis, W.W. (2005) "Obtaining Cancer Risk Factor Prevalence Estimates in Small Areas: Combining Data from Two Surveys", *Journal of the Royal Statistical Society (Series C): Applied Statistics*, 54(3), pp. 595–609.

Godbout, S. and Grondin C. (2005) "Getting around a Conceptual Difference between Two Data Sources when Producing Estimates", *Statistics Canada International Symposium Series – Proceedings*, cat. no. 11-522-XIE.

Hartley, H.O. (1962) "Multiple Frame Surveys", Proceedings of the Social Statistics Section, American Statistical Association, pp. 203–206.

Hidiroglou, M.A. (2001) "Double Sampling", *Survey Methodology*, 27(2), pp. 143–154.

Houbiers M.; Knottnerus P.; Kroese A.H.; Renssen R.H. and Snijders V.S. (2003) "Estimating Consistent Table Sets: Position Paper on Repeated Weighting" Discussion paper 03005, Statistics Netherlands, Voorburg/Heerlen.

Little, R.J.A. and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*, Second edition, John Wiley, New York.

Lohr, S.L. and Rao, J.N.K. (2000) "Inference from Dual Frames", *Journal of the American Statistical Association*, 95(449), pp. 271–280.

Merkouris, T. (2004) "Independent Regression Estimators from Multiple Surveys", *Journal of the American Statistical Association*, 99(468), pp. 1131–1139.

Office for National Statistics (2004) *Consultation Paper: Proposals for a Continuous Population Survey*.

Rao, J.N.K. and Wu, C.F.J. (1988) "Resampling Inference with Complex Survey Data", *Journal of the American Statistical Association*, 83(401), pp. 231–241.

Renssen, R.H. and Nieuwenbroek, N.J. (1997) "Aligning Estimates for Common Variables in Two or More Sample Surveys", *Journal of the American Statistical Association*, 92(437), pp. 368–374.

Särndal, C.E. (1992) "Methods for Estimating the Precision of Survey Estimators when Imputation has been Used", Survey Methodology, 18, pp. 241–252.

Särndal, C.E.; Swensson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*, SpringerVerlag.

Schenker, N. and Raghunathan, T.E. (2007) "Combining Information from Multiple Surveys to Enhance Estimation of Measures of Health", *Statistics in Medicine*, 26 pp. 1802–1811.

Shao, J. and Wu, D. (1995) *The Jackknife and Bootstrap*, SpringerVerlag, New York.

Thomsen, I. and Holmfly, A.M.K. (1998) "Combining Data from Surveys and Administrative Record Systems: The Norwegian Experience", *International Statistical Review*, 66, pp. 201–221.

APPENDIXES

A. SPECIAL SOCIAL SURVEYS PROGRAM 2008–09

Survey	*Enumeration period*
National Aboriginal and Torres Strait Islander Social Survey (NATSISS) 2008	July 2008–December 2008
Survey of Disability and Carers	April 2009–December 2009
Survey of Education and Training	March 2009–June 2009

B. MONTHLY SUPPLEMENTARY SURVEYS PROGRAM 2008–09

Supplementary survey	*Month*
Job Search Experience (JSE)	July
Employee Earnings, Benefits and Trade Union Membership (EEBTUM)	August
Persons Not in Labour Force (PNILF)	September
Under-employed Workers (UEW)	
State supplementaries	October
New South Wales – Household and Workplace Mobility and Implications for Travel	
Western Australia – Labour Mobility and Intentions	
Forms of Employment (FoE)	November
Contract Work	
Locations of Work (LoW)	
Labour Force Experience (LFE)	February
Environment: Waste Management, Transport and Motor Vehicle Usage	March
Children's Participation in Culture and Leisure Activities (CPCLA)	April
New South Wales Crime and Safety Survey (NSW C&SS)	
Survey of Education and Work (SEW)	May

C. HOUSEHOLD FORM DATA ITEMS

The mandatory part of the household form covers the variables Age, Sex, Relationship in household, Social marital status, Family composition, Household composition and Full-time/part-time student status.

The optional part of the household form covers Country of birth of person, Year of arrival in Australia, Indigenous status, Month and year left school, and Registered marital status.

Other possible household form data items, currently supported by an ABS standards framework, relate to Language, Occupation, Employment Status, Housing, Education, Disability, Cultural diversity.

D. DIFFERENCES IN THE SURVEYS USED IN THE CASE STUDY (NATSIHS AND LFS)

Source of difference	*Action taken to remove difference*
	Coverage
Special dwellings (SDs)	People in special dwellings are in coverage of the LFS but out of coverage of the NATSIHS. From the LFS, 1.3% of the Indigenous population live in special dwellings. Action: None.
Persons under 15 years of age	Persons under 15 years of age are in coverage of the NATSIHS but out of coverage in LFS. Action: Exclude persons under 15 from NATSIHS.
Military personnel	Military personnel are in coverage of NATSIHS but out of coverage of the LFS. Military personnel in NATSIHS cannot be identified or removed. According to 2001 Census, the proportion of Indigenous people in the defence forces is around 1%, suggesting the impact of this mismatch in coverage of military personel is small. Action: None
Visitors to private dwellings	Visitors to private dwellings are in coverage of the LFS but out of coverage of the NATSIHS. Approximately 1% of Indigenous respondents to the LFS records were visitors to private dwellings. Action: None
Reference period	Indigenous estimates from the LFS are obtained by pooling sample from January to December. The 2004–05 NATSIHS covered the period August 2004 to July 2005. Action: Pool the LFS over the period August 2004 to July 2005.
	Sample design
Torres Strait Islanders were over-sampled in NATSIHS	TSI status was used as a benchmark category in NATSIHS, so TSI respondents should be representatively weighted. Action: None.
Sample design in non-community areas	NATSIHS selects first stage units (Collection Districts) with probability proportional to the expected no. of Indigenous. The LFS selects first stage units (Collection Districts) with probability proportional to the expected no. of people in coverage of the LFS (see discussion of coverage above). Action: None.

	Enumeration
Screening	LFS asks an Indigenous identification question as part of the survey. NATSIHS uses a screening process to select Indigenous dwellings. There was some concern that Indigenous people may not self-identify during the NATSIHS screening, if they do not wish to participate in the survey. Action: An explicit adjustment was incorporated into the initial weight for NATSIHS sample to ensure areas with low screening rates were not under-represented in the NATSIHS. No such adjustment was made for the LFS.
Non-response	The LFS response rate is about 96%, compared with the NATSIHS response rates of about 84%. Action: None
Forms	The LFS and the NATSIHS both have two forms: – the long form designed for a majority of persons – the short form is a designed for Indigenous people, typically living in communities, to whom the concepts in the long form would be unfamiliar. The LFS and the NATSIHS long and short forms are all different, which means there are four forms in total. Action: None
Conceptual definition of employment status	The LFS's long form requires more detail than the NATSIHS's long form to determine the employment status, particularly for unpaid voluntary workers, people away from work and people about to start work. The LFS short form requires less detail than the NATSIHS short form particularly for people looking for work. Action: Remove conceptual differences between the LFS and the NATSIHS long forms and between the LFS and the NATSIHS short forms.
	Weighting
Benchmark population counts	The LFS benchmark counts for the Indigenous population are at the State, by Sex, by Remoteness (three groups), by Age (three groups). The NATSIHS benchmark counts are at the State, by Remoteness (five groups), by Age (seven groups), by Sex, by ATSIC region (two groups), by TSI status level. Action: Use the LFS benchmark counts for the NATSIHS.
Initial weight	The initial weight is calibrated to the benchmark Population Counts (see above). Calibration occurs separately for the LFS and NATSIHS. The initial weight for an Indigenous person in the LFS is constant within state, reflecting the fact that the LFS is a geographically representative sample. The initial weight for the NATSIHS's estimates in this paper is obtained by calibrating the inverse of the selection probability to the NATSIHS benchmark counts. This ensures that, for example, Torres Strait Islanders are not over-represented in the weighted data set.

E. RELATIVE STANDARD ERRORS OF CASE 1 ESTIMATES

	NSW	*Vic.*	*Qld*	*SA*	*WA*	*Tas.*	*NT*	*ACT*
LABOUR FORCE SURVEY								
Major Cities								
Employed	0.084	0.090	0.079	0.122	0.102			0.064
NILF	0.115	0.152	0.106	0.108	0.123			0.152
Unemployed	0.223	0.494	0.230	0.262	0.190			0.168
Regional								
Employed	0.128	0.137	0.079	0.147	0.099	0.072	0.283	
NILF	0.096	0.114	0.083	0.197	0.078	0.102	0.165	
Unemployed	0.208	0.293	0.122	0.291	0.251	0.179	0.387	
Remote								
Employed	0.984		0.292	0.324	0.114	0.558	0.180	
NILF	0.084		0.422	0.274	0.090	0.567	0.099	
Unemployed	0.644		0.651	0.878	0.931	0.805	0.287	
NATIONAL ATSI HEALTH SURVEY								
Major Cities								
Employed	0.077	0.090	0.088	0.087	0.106			0.061
NILF	0.097	0.140	0.120	0.090	0.087			0.143
Unemployed	0.196	0.312	0.282	0.272	0.303			0.326
Regional								
Employed	0.072	0.117	0.090	0.165	0.124	0.062	0.103	
NILF	0.071	0.130	0.100	0.114	0.143	0.066	0.101	
Unemployed	0.185	0.192	0.195	0.384	0.382	0.152	0.326	
Remote								
Employed	0.087		0.063	0.076	0.064	0.892	0.082	
NILF	0.056		0.070	0.105	0.101	0.000	0.067	
Unemployed	0.773		0.169	0.504	0.208	1.010	0.252	
CASE 1								
Major Cities								
Employed	0.058	0.065	0.063	0.072	0.074			0.045
NILF	0.075	0.104	0.084	0.071	0.071			0.106
Unemployed	0.147	0.292	0.193	0.203	0.193			0.167
Regional								
Employed	0.064	0.089	0.063	0.120	0.086	0.047	0.104	
NILF	0.058	0.092	0.069	0.099	0.089	0.056	0.086	
Unemployed	0.141	0.173	0.135	0.252	0.251	0.116	0.249	
Remote								
Employed	0.339		0.127	0.082	0.058	0.524	0.075	
NILF	0.047		0.165	0.122	0.069	0.567	0.055	
Unemployed	0.522		0.183	0.487	0.318	0.639	0.195	

FOR MORE INFORMATION . . .

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

INFORMATION AND REFERRAL SERVICE

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

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

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

FREE ACCESS TO STATISTICS

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

WEB ADDRESS www.abs.gov.au