Australian Bureau of Statistics

METHODOLOGICAL NEWS

A QUARTERLY INFORMATION BULLETIN FROM THE METHODOLOGY AND DATA MANAGEMENT DIVISION

March 2010

Directions for Household Survey Sampling Framework

The current broad sampling framework underpinning ABS household surveys has been in use for roughly 40 years. Household Survey Methodology (HSM) are currently investigating possible significant changes to the household survey sampling framework as an extension of the routine 5-yearly Monthly Population Survey (MPS) sample redesign hinging off the 2011 Census of Population and Housing. This timing allows for implementation of a new framework in the 2012/13 financial year. The changes being considered are driven by factors including:

- changes to the ABS geography classification structure, including the availability of a significantly smaller finest level of geography (mesh block) than current (Census collection district);
- relatively recent developments in sampling methodology and algorithms, i.e. balanced sampling via the cube method which shows potential non-negligible savings for MPS sample redesign;
- increasing demand for social statistics that require more sophisticated sample designs;
- a rebuild of sample management systems;
- increasing availability of new technology including mapping and geocoding software; and
- the need to find bookable efficiency gains for the organisation.

Currently the MPS, which consists of the Labour Force Survey (LFS) and its supplements, and Special Social Surveys (SSSs) are run off a single sampling framework. In this framework the Primary Sampling Unit (PSU), which is most often a Census Collection District (CD), is selected on the basis of MPS sample design but supports both MPS and SSSs. A block from each selected CD is used for MPS and another block, called the "parallel block", is used for SSSs. The sample design for MPS is area-based multi-stage, redesigned every five years focused on optimising design parameters towards LFS requirements. For both MPS and most SSSs, designs are Equal Probability of Selection (EPS) within state, where clusters of dwellings from within blocks and CDs are systematically selected.

The current approach yields efficiencies in sample preparation activities (because the MPS and parallel blocks are set up at the same time), is easy to manage from an interviewer allocation perspective and controls effectively for overlap between collections. However, the approach has limited flexibility from a SSS sample design perspective, and given that the cost model underpinning the MPS sample design bears little resemblance to SSS enumeration costs, is probably sub-optimal for SSSs. In addition, recent SSS sample designs aimed at targeting specific subpopulations of interest have had to select sample independently of the parallel block framework. Recent investigations have also uncovered potential savings from balanced sampling for MPS sample design but these savings are not present, or at least relatively smaller, for SSSs.

As a result, two new sampling frameworks are being proposed as part of HSM's work, namely:

- An MPS master sampling framework (as it is currently), with an independent SSS sampling framework that will cover the majority of 'standard' SSS sample designs (i.e. for surveys of some specific subpopulations such as Indigenous people or pensioners, there will be no option but to have an independent sample design); and
- An MPS master sampling framework (as it is currently), and independent sample design and selection for all other surveys.

The first option allows greater flexibility in the design of SSSs, however it incurs additional sample preparation cost compared to the current method (through decoupling of sampled areas and hence increase in sample preparation activities) and may not be flexible enough for some surveys. Overlap control is relatively easy to manage in this scenario, though will be more complicated than the current approach. The second option allows maximum flexibility for sample design activities for all surveys, but incurs the greatest additional sample preparation cost and presents the most challenging overlap control scenario. However this approach could yield greater overall efficiency if interviewer field costs greatly outweigh sample preparation costs, and the sample design could be more appropriately tailored to meet the needs of individual surveys.

Potential savings exist in reduced sample preparation activities such as "in-office" address coders for listing dwellings in selected areas and the new geography structure providing small "block type" units (ie mesh blocks) that are pre-defined, thus removing the need to divide a selected area into smaller groups of dwellings. Savings also exist in the potential use of balanced sampling for MPS selections as mentioned above (but not for detailed discussion in this article).

Broadly, HSM will be directing their immediate effort in the following areas:

- Establishing main advantages and disadvantages of the different household survey sampling frameworks proposed above;
- Reviewing other National Statistical Institute's (NSI's) approaches to household survey sampling frameworks;
- Examining potential cost modelling approaches and the data required for those cost models; and
- Establishing current and future demand for the ABS household survey program that will impact on sample design objectives.

For more information, please contact Justin Lokhorst on (08) 8237 7476 or justin.lokhorst@abs.gov.au.

Update Strategy for Derived Employment Size on Business Survey Frames

In August 2008, the Derived Employment Size (DES), a modelled size variable based on number of payees data, replaced the Derived Size Benchmark (DSB) as the indicator of size on the frames for business surveys. From June 2010, two DES variables will be made available on the frames: a Stratification DES, and a Latest DES which will hold the most up-to-date information and is typically used in estimation. A strategy has been developed for the Stratification DES to allow sizing information for stratification to be updated, while minimising adverse effects on sample rotation.

Because most business surveys stratify by size, changes to the stratification variable can lead to units changing strata and a high likelihood of sampled units being rotated out. Payees information is available to the ABS from the ATO, and for most businesses, this value is updated once a year. This means the DES assigned to most businesses on the frame will be updated once a year. The update strategy for Stratification DES tries to minimise this impact on surveys by spreading the updates over four quarters, and by using "sticky stratification".

Each unit on the ABS Business Register is randomly assigned to an update quarter, and while the Latest DES for each unit will be updated as soon as new information becomes available, a unit is only eligible to have its Stratification DES updated in its update quarter. This way, in each quarter, at most only 25 per cent of the units on the Register could possibly have their stratification size value changed, thus controlling the number of units which would change size groups at any one time.

In addition, a standard set of stratification size groups has been defined. The groups are finer than the stratification size bounds used in ABS economic surveys so that the stratification needs of the surveys can be met. For each stratification size group, upper and lower "sticky boundaries" have been defined. In a unit's update quarter, the unit's Latest DES is compared to its Stratification DES, and only if the Latest DES has moved beyond the sticky boundaries will the unit's Stratification DES be changed. For example, if the 5-19 size group has an upper sticky boundary of 26, then a unit would not move to the 20-49 size group until its Latest DES is greater than 26. This minimises stratum changes of units whose DES fluctuates around a size boundary.

By implementing this strategy, a balance will be achieved between keeping the stratification size information up-to-date, and managing sample rotation.

For more information, please contact Carmen Kong (02) 6252 5944 or carmen.kong@abs.gov.au.

Wheat Use Survey Design

The development of the Wheat Use Survey (WUS) provides a good example of the challenges in the design of Agricultural surveys. WUS is one component of a program of user-funded surveys aimed at meeting requirements of the Department of Agriculture, Fisheries and Forestry (DAFF) for information on the storage and use of wheat to determine the levels of uncommitted grain in Australia. The program consists of:

- a wheat use survey;
- a wheat export sales survey; and
- a survey of grain handlers' stocks

The first two of these surveys have been conducted monthly since the start of the program in October 2008, the grain handlers' survey, initially an annual, has been run monthly since March 2009. The program will has funding to September 2011.

The main data items collected are the amount of wheat used during the reference period and the amount stored at the end of the period. A significant issue in the development of the program was the determination of the target population and the assembly of a frame. Initially, the main users and/or storers of wheat were considered to be: manufacturers of foods, animal feed and organic chemicals, beef feedlot operators and pig and poultry farmers. In the initial month a large sampling fraction was used (3,036 from a population of 6,253) since it was expected that there would be significant numbers of dead or out-of scope units on the frame. The data provided by this first survey allowed refinement of the frame and design with a reduction of the sample size to about 1,101 from a population of 5,143.

During the first year of the program it became apparent that there is significant use or storage of wheat by industries not in scope of the existing collection. In particular, significant amounts of wheat appeared to be stored on farm by growers and the dairy industry also uses wheat as stock feed. It was therefore decided to broaden the scope of WUS to include these sectors. A supplement would be selected from the population of dairy and wheat farmers and added to the existing WUS sample. Before this was done a coverage survey was

undertaken which provided design data for the allocation of the supplementary sample.

Design of this supplementary sample illustrates another challenge encountered in the design of agricultural surveys - variables available on the frame for allocating units to size groups generally have only a poor correlation with the design variables of interest. In other economic surveys, frame employment or BAS turnover are used to size units. For agricultural businesses, with their extensive use of mechanisation and contract labour, frame employment is not a good indication of agricultural production. Since many businesses with agricultural operations also derive income from non-agricultural activities, BAS turnover also is not a good measure. For this reason another measure of size has been developed for agricultural units - the estimated value of agricultural operations (EVAO). This is a composite index based on the production of each commodity and the land area allocated to that commodity as reported by the unit in the last Agricultural Census. Three-year averages of commodity prices are used to assign a value to the production of that commodity and the associated land area. The EVAO of the unit is then the sum of these values over all commodities it produces.

EVAO is a good measure of the overall production of a farm, but the specific set of commodities of interest in a particular survey may account for only a part of its production. Furthermore, data items other than commodities may also be of interest. Hence it is possible that EVAO may still not correlate well with major survey design variables. For example, in the case of wheat growers, the annual production of wheat may not be a good indication of the amount stored at the end of any particular reference month (the variable of interest here). Similarly, the milk production from a dairy farm is not necessarily well correlated with the amount of wheat used by the farm since other stock feeds may be used. The consequence of this poor correlation is that strata may be quite heterogeneous with respect to the design variable(s). This means that designs will not usually be efficient. In the case of the WUS supplement, sample sizes, set by budgetary constraints, corresponded to overall sampling fractions of about 1 in 40 for dairy and 1 in 50 for wheat. The design RSEs achievable with these sample sizes were 20% at the national level and 40% at state level, which is in marked contrast to the design performance for other business surveys.

Although the impact of some high estimation weights has raised some issues with the BSC the first two months' estimates from the expanded collections have yielded positive results - with the expanded data adding to the industry's ability to more accurately determine the available national wheat supply.

For more information, please contact Jos Beunen on (02) 9268 4647 or jos.beunen@abs.gov.au.

Small Area Estimation Using a Multinomial Logit Mixed Model with Category Specific Random Effects

The Analytical Services Branch (ASB) recently published a research paper entitled "Small Area Estimation Using a Multinomial Logit Mixed Model with Category Specific Random Effects". Authored by Janice Scealy, the paper describes a model based approach to producing small area estimates of counts for different categories of Australian labour force based status (employed, unemployed and not in the labour force (NILF)), on a multinomial logit mixed model with category-specific random effects. The term 'category-specific' means that within each small area there are two correlated random effects, one associated with the employed category and the other associated with the unemployed category.

In this application the Multinomial Logit Mixed Model gave similar estimates and mean squared errors (MSEs) to that of the Binomial Logit Mixed Model fitted separately to each labour force status variable. However, the main advantage of the multinomial approach in the general case is that it has the capability to produce more accurate small area estimates where there are reasonably strong correlations between the categories, such as the employed and the unemployed. Another advantage is that once the explanatory variables have been selected for each category response variable, model estimation can be carried out simultaneously for all categories. A third advantage is that the estimates of proportion for each category are guaranteed to add to one, which is not assured when applying separate binomial models to each category.

In the study undertaken by Scealy, estimates of the model parameters were produced using penalised quasi-likelihood combined with approximated restricted maximum likelihood estimation and using these, estimated counts were then produced for each small area. MSE estimates, which measure the statistical accuracy of the estimated counts, were approximated using two methods: the parametric bootstrap and analytical approximations. The performance of these methods was then compared. Using a parametric bootstrap we also examine the properties of the combined penalized quasi-likelihood and restricted maximum likelihood estimators and discuss model goodness of fit measures and diagnostics.

For a copy of the paper, visit the ABS website and search for Catalogue Number 1351.0.55.029 - Research Paper: Small Area Estimation Using a Multinomial Logit Mixed Model with Category Specific Random Effects, Jan 2010. For further information on the analysis, contact Janice Scealy on (02) 6252 5764 or janice.scealy@abs.gov.au.

NatStats 2010 Conference

The NatStats 2010 Conference will be held at the Sydney Convention and Exhibition Centre, Darling Harbour, from 15 - 17 September 2010.

NatStats 2010 will bring together an array of leaders and high profile commentators, researchers and policy makers from all levels of government, academia, community and business. NatStats was initiated to assist in the development of a collaborative approach to national statistics. It focuses on various issues surrounding quantitative and qualitative research, which have been highlighted by key government initiatives.

The conference will build on the success of the first NatStats conference held in 2008, where around 480 delegates attended. A major outcome of the 2008 conference was the tabling of a statistical declaration to guide the development of a national statistical strategy for Australia in the 21st century.

NatStats 2010 will build on the enthusiasm and passion generated by delegates at NatStats08 and will offer delegates an opportunity to:

- Explore the impact of expected demographic, environmental and global economic change and their implications for governance, policy, productivity and the well-being of Australia's urban and rural communities.
- Identify the improvements needed in our national statistical system to understand and respond to these challenges.
- Identify improvements needed in the current statistical frameworks for measuring progress and development to ensure Australians' wellbeing.

An exciting program is being developed and will address a range of issues regarding national statistics. If you would like to hear more about NatStats 2010 or have any suggestions, please contact Annette Hants on (02) 6252 6936 or email natstats@nss.gov.au.

How to Contact Us and Subscriber Emailing List

The Methodological Newsletter features articles and developments in relation to work done within the ABS Methodology and Data Management Division. By its nature, the work of the Division brings it into contact with virtually every other area of the ABS. Because of this, the newsletter is a way of letting all areas of the ABS know of some of the issues we are working on and help information flow. We hope the Methodological Newsletter is useful and we welcome comments.

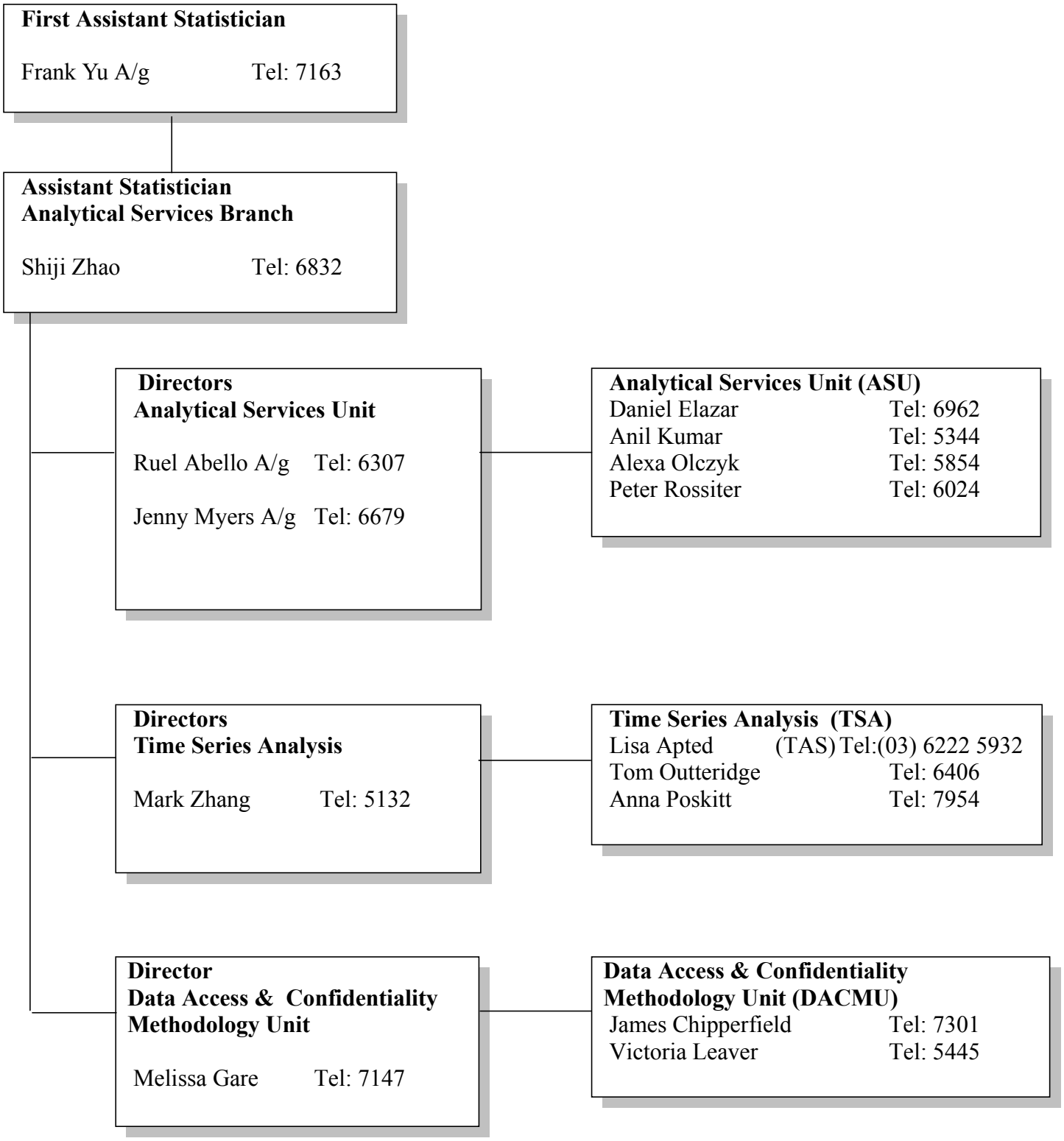
If you would like to be placed on our electronic mailing list, please contact:

Jayne McQualter
Methodology & Data Management Division
Australian Bureau of Statistics
Locked Bag No. 10
BELCONNEN ACT 2617
Tel: (02) 6252 7320
Email: methodology@abs.gov.au

Click on the following links to view the ABS Privacy Statement and Disclaimer
Privacy Statement | Disclaimer

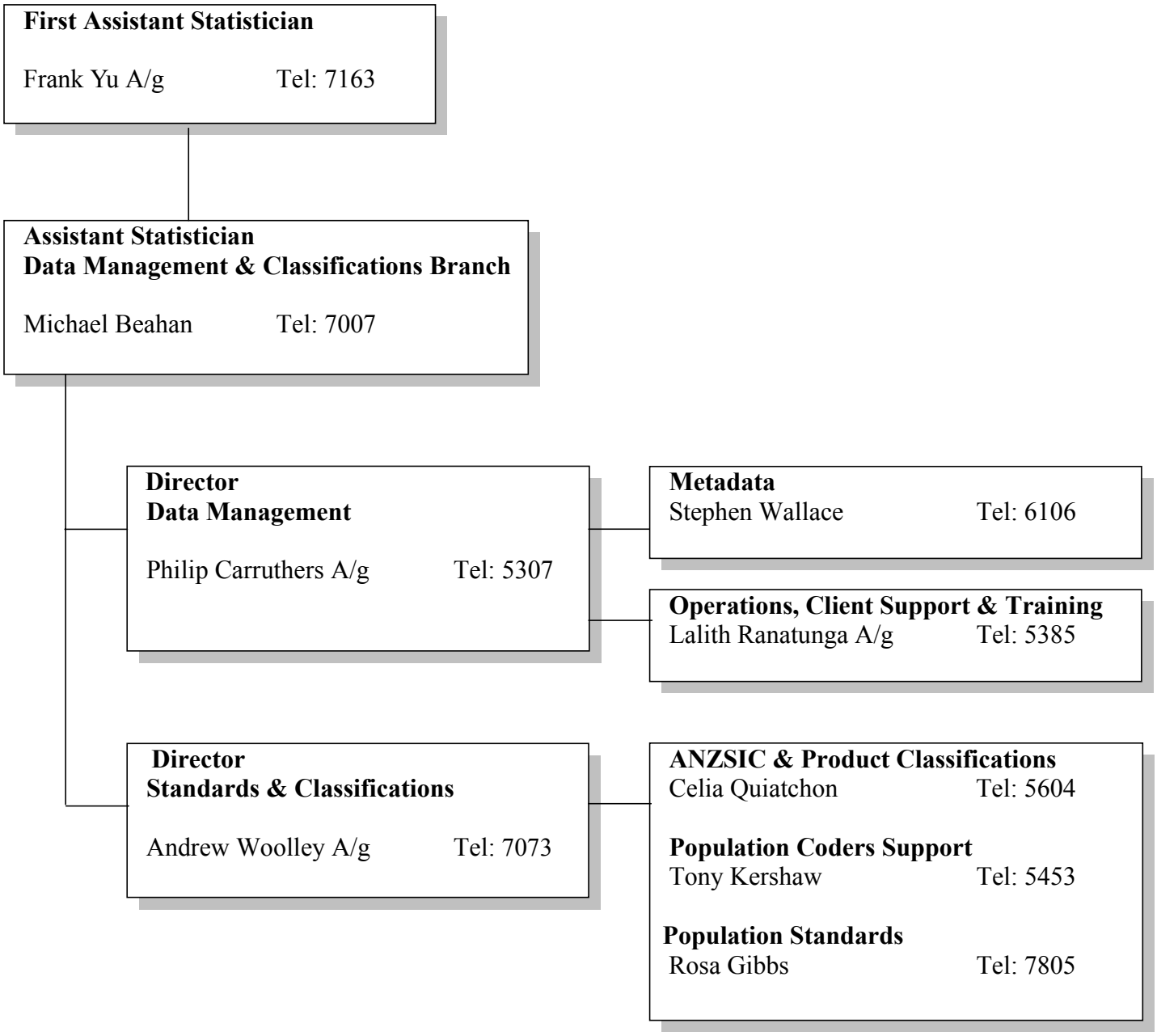
Methodology & Data Management Division
Management Structure

Current at March 2010



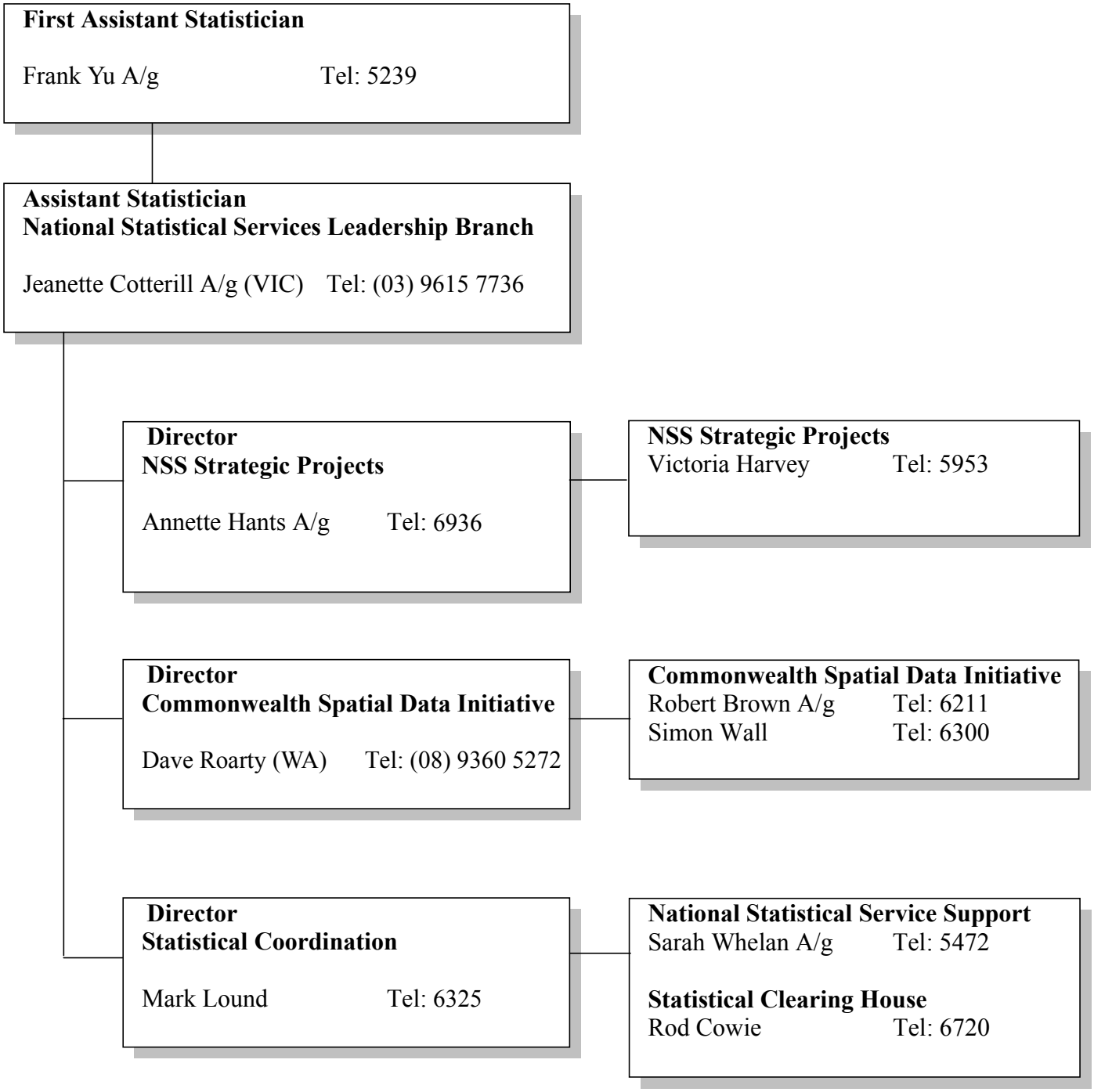
Methodology & Data Management Division
Management Structure

Current at March 2010



Methodology & Data Management Division
Management Structure

Current at March 2010



Methodology & Data Management Division
Management Structure

Current at March 2010

First Assistant Statistician
Frank Yu A/g Tel: 7163

Assistant Statistician
Statistical Services Branch
Paul Schubert A/g Tel: 6591

Director
Household Survey Methodology

Alistair Rogers Tel: 7334

Household Survey Methodology (HSM)
John Martin Tel: 7006
Ross Watmuff Tel: 7084
Justin Lokhorst (SA) Tel: (08) 8237 7476

Director
Methodology Development Unit

Bill Gross Tel: 6302

Methodology Development Unit (MDU)
Carmen Kong Tel: 5944
Philip Bell (SA) Tel: (08) 8237 7304
Carl Mackin (WA) Tel: (08) 9360 5250
John Preston (QLD) Tel: (07) 3222 6229

Directors
Business Survey Methodology

Justin Farrow A/g Tel: 5795

Business Survey Methodology (BSM)
Simon Strickland A/g Tel: 5696
Keith Farwell (TAS) Tel: (03) 6222 5889
Brett Frazer (QLD) Tel: (07) 3222 6028
Irina Pribil (VIC) Tel: (03) 9615 7566

Director
Data Collection Methodology

Emma Farrell A/g Tel: 7316

Data Collection Methodology (DCM)
Narrisa Gilbert Tel: 5283
Kettie Hewett (VIC) Tel: (03) 9615 7843

Director
Operations Research and Process Improvement Unit

Rebecca Farrow A/g Tel: 6022

Operations Research and Process Improvement Unit (ORPI)
Melanie Black A/g Tel: 7241
Rosslyn Starick (VIC) Tel: (03) 9615 7055

Divisional Support Unit (DSU)
Alan Herning Tel: 5350