Australian Bureau of Statistics

METHODOLOGICAL NEWS

A QUARTERLY INFORMATION BULLETIN FROM THE METHODOLOGY AND DATA MANAGEMENT DIVISION

December 2008

Improved Estimation Methodology for LFS Family Estimates

Since the 1970s, family estimates have been produced as a by-product of the Labour Force Survey (LFS). Family type data are derived from questions which establish the relationships between members of the selected households. The family estimates produced from LFS data classify families by type (e.g. couples without dependents, single parent families with children aged under 15), as well as by the labour force characteristics of the family (e.g. number of employed family members).

Recent improvements in the capture of information about the relationship between persons in sample households have enabled improvements to the estimation methodology for family data. Two distinct changes were made to the estimation methodology. First, more households selected in the LFS now contribute to family estimates, and second, a different method is used to calculate the sample weight of the contributing families.

Under the previous estimation methodology, the family weight was calculated as the harmonic mean of the LFS weights of each family member. A family weight was only produced for those families in which all members aged 15+ in their household contributed to the person-level LFS estimates. As a result of this, families in which one or more persons aged 15+ were out of scope (such as members of the Australian Defence Force) or did not respond to the LFS, did not contribute. Since the estimation method did not use independent population estimates (benchmarks) to compensate for these exclusions, family estimates were lower than they would have been if all families were included.

The new estimation method assigns a weight to all families in which the necessary relationship data are collected. The family weights are computed using a Generalised Regression (GREG) Estimator which uses demographic data about all individuals (including children) in the contributing sample families. Benchmarks for both the number of persons and households are used, so the weights are referred to as being 'integrated' at the person and household levels. Each household (and hence family) is assigned a weight such that the sum of the weights in each household benchmark class equals the household benchmark count, and when the household weight is assigned to each of its members, the weighted sum of individuals in each person benchmark class equals the person benchmark count.

The use of benchmarks enables the weighted sample to better represent the Australian population and ensures estimates are more comparable with family estimates from other ABS collections. Including more families in the estimates has also reduced bias. The sample error of estimates has been significantly reduced, with reductions in standard error of more than 40 per cent for those estimates which are a high proportion of the total number of families.

The first set of family estimates produced under the new estimation method used September LFS data and was released in October 2008. Also released at this time were an information paper on the new method (cat. no 6224.0.55.002) and estimates produced under the new method applied to historical data dating back to August 2004. Datacubes containing family estimates will now be released each month (cat. no. 6224.0.55.001) and can be downloaded from the ABS Website.

For further information, please contact Julian Whiting on (08) 8237 7362.

New QBIS Sample Design

Since its inception in 2001, the Quarterly Business Indicators Survey (QBIS) has been the main vehicle for providing economy-wide data to National Accounts Branch (NAB) to support the preparation of the quarterly national accounts. The survey collects values of national sales, wages, inventories and profits and of state sales and wages as well. All industries other than Agriculture are surveyed, though the data collected vary slightly with industry and size of business.

QBIS is currently undergoing a substantial re-design for the twofold purpose of implementing the new ANZSIC06 industry classification and to achieve the goals of the first stage of the QEWS Phase 3 Project given its current formulation as approved in October 2007.

The first stage of this project was originally planned to consist of:

- the expansion of the scope to include non-employers;
- the introduction of a BAS turnover cut-off below which non-employers will be excluded from the sample. Estimates for this part of the population will be sourced from BAS data; and
- the use of BAS turnover to assign units to size groups for the purposes of stratification.

The goals of these changes are, first, to improve the coverage of the economy. Currently, QBIS surveys employers only and so its estimates fail to capture the non-employer contribution to the economy, estimated to be about 15 per cent as measured by Business Activity Statement (BAS) turnover. The introduction of a non-sampled stratum below a turnover cut-off is intended to enable this expansion of scope without significant increase in sample size. Second, the changes will bring the survey into closer alignment with the flagship annual survey, the Annual Integrated Collection (AIC) with respect to scope and coverage. Such closer alignment will facilitate confrontation of corresponding data items from the two surveys and will also assist NAB in the process of preparing the annual national accounts.

Implementation of such a significant change presents a number of challenges, not the least of which is to determine the impact on both levels and movements. The ANZSIC change and the inclusion of non-employers both affect the level of the series, but the seasonality is also likely to change, since the seasonal behaviour of the non-employer contribution is expected to be different from that of employers. To capture both impacts, a two-stage implementation strategy is being planned.

A parallel run will be conducted in March and June quarters of 2009 in which two samples will be produced - one from the current design and one from the new design. The sample selected on the current design will be used to produce current outputs - ANZSIC93 employer-only estimates. The sample from the new design will also produce employer-only estimates but on an ANZSIC06 basis. The difference between the two provides a measure of the impact of the ANZSIC change. ANZSIC06 estimates will be published for the first time for the September 2009 quarter.

Expansion of scope without a corresponding increase in sample size always presents a challenge for sample design and the present case was no exception, especially in view of the requirement that the design be capable of delivering acceptable estimates for the employing sub-population as well as for the total (employing plus non-employing) population. In principle, these requirements could be met by stratifying employers and non-employers separately. However, this would produce too many small strata and thus too large a sample. A compromise has therefore been reached in which there are separate strata for non-employers in those industry-by-size groups in which they are a significant proportion of the population. In all other cases (the majority), employers and non-employers are stratified together.

For further information, please contact Jos Beunen on (02) 9268 4647.

Framework for Household Survey Costs and Fitness for Purpose Research

The Population Statistics Operations (PSO) area uses trained interviewers (also referred to as agents) to collect survey information for the Household Survey Program. A Household Field Cost Model (hereafter referred to as the model) has been developed to predict payment to agent costs up to four years in advance of survey enumeration. Increasingly, however, Household Survey areas are also asking questions such as:

- What if I conducted part of my survey as a telephone interview and part of it face-to-face?
- How would my costs and quality of output change if I were to increase or decrease my target response rate?
- How should I run my survey to achieve the results that I want for the least cost?

Questions such as these are not readily answerable from the model as they involve the interrelationships of a number of variables. Respondents are not all the same and neither are interviewers and these factors play a significant part in the tension between cost and survey outcomes such as response rates and contribution to estimates. The Operations Research Unit has commenced a project to develop a framework for managing the trade-off between field costs and data quality to answer questions such as those raised above.

The first phase of the project was to review the current model to determine its ability to accurately predict costs across the range of surveys in the Household Survey Program. This phase identified some quick modifications to the model to improve the accuracy of household survey cost predictions in the short term.

The next phase of the project, which is currently underway, is aimed at developing a mathematical model for costs which is transparent and accurately reflects costs of operational procedures involved in data collection. The model will be able to assess 'what if ' scenarios, specifically including different field procedures; and will be able to be readily maintained and used for monitoring costs. Work is also underway to develop measures of fitness for purpose for time series and cross sectional estimates, and to understand the relationship between operational procedures and fitness for purpose.

For more information on this project, please contact Louise Gates on (02) 6252 6540.

Temporal Aggregation and Seasonal Adjustment

Due to user demand, the Australian Bureau of Statistics (ABS), in some instances, publishes original, seasonally adjusted and trend estimates at different observation frequencies for the same indicator. Hence sometimes at the quarterly level, original time series estimates are simply a temporal aggregate of their monthly counterpart. Suppose a time series of quarterly seasonally adjusted estimates is desired from such an equivalent time series pair. These estimates can be obtained via two approaches: either by (1) seasonally adjusting the quarterly original time series directly or by (2) seasonally adjusting the monthly original time series and then temporally aggregating to the quarterly level

(referred to as the temporal aggregation approach hereafter). The ABS currently uses the first approach for seasonally adjusting equivalent time series pairs. This causes quality and consistency issues along with duplicate work.

The ABS knows from previous research that estimating calendar-related effects is more accurate when performed at the monthly level and then applied to the quarterly case. This idea of using a monthly time series to estimate a component of its quarterly equivalent is taken further by the temporal aggregation approach. The aim becomes to completely derive the quarterly seasonally adjusted series from its monthly seasonally adjusted counterpart. Hence significant improvements in quality and consistency are expected to be made.

For the Census X11 method, literature suggests that seasonal adjustment first and temporal aggregation second is the more efficient approach in terms of mean squared error and forecast performance. However, the impact of temporal aggregation on current end revisions had not been assessed for the mixed X11/ARIMA forecasting method utilised by the ABS.

The ABS hence conducted case studies to compare the quarterly seasonally adjusted estimates obtained via the two approaches using the ABS X11/ARIMA forecasting method, in terms of their relative efficiency, revisability and consistency. These included analyses on both simulated data and real ABS Balance of Payments estimates. The background, methodological basis and results from these studies were summarised and presented in an ABS Methodological Advisory Committee paper in June 2008 ("Temporal Aggregation and Seasonal Adjustment", cat. no. 1352.0.55.095). The conclusion from the paper was favourable for the temporal aggregation approach because this research indicated that consistency and productivity gains are made and that the quality of seasonal adjustment is not compromised. The ABS now plans to conduct further empirical studies to explore the temporal aggregation approach with ABS Agriculture Survey estimates where it is hoped a more diverse range of series can be explored along with the impact of sampling error effects.

For more information, please contact Lisa Apted on (03) 6222 5932, or Mark Zhang on (02) 6252 5132.

Feasibility of Data Pooling in the ABS

The ABS is always under competing pressures. On the one hand, there is pressure for ABS to produce more precise estimates for small sub-populations. On the other hand, there is pressure on budgets so that we increase the utility of existing collections. With these considerations in mind, Analytical Services Branch (ASB) is currently exploring whether improved estimates can be created by combining (i.e. pooling) data from multiple ABS collections. If successful, data pooling will allow ABS to better use the data it already has and to analyse more of the Australian population in greater depth.

The main aims of the investigation being conducted by ASB are to explore and understand the issues involved in pooling data from multiple sources; to develop a set of criteria to evaluate, on a case by case basis, whether gains can be obtained through pooling; and, to propose techniques for effectively pooling data in common situations, under various assumptions.

The primary benefit of data pooling is increased sample size, which may allow key estimates to be produced with reduced sampling error. It may also be possible to use a pooled dataset to produce estimates for small populations, whose sampling errors were initially too high for publication.

However, inconsistencies between collections may introduce additional non-sampling error. This increase in non-sampling error must be weighed against the reduction in sampling error, to decide whether pooling is beneficial. Possible differences between collections to consider include:

- differences in scope and/or coverage of the collections;
- differences in enumeration periods;
- differences in sample design and/or weighting procedures;
- differences in questionnaires; and
- differences in non-response.

ASB plans to explore the impact of each of the sources of non-sampling error, when using a pooled dataset to create parameter estimates and variance estimates. It will conduct a number of case studies, using collections from the ABS Household Survey Program, to highlight some of the key issues. The first case study, which is underway, looks at Indigenous labour force estimates, by combining Labour Force Survey (LFS) data with data from the National Aboriginal and Torres Strait Islander Health Survey (NATSIHS). Currently, annual Indigenous labour force estimates are produced from the LFS by pooling Indigenous respondents from 12 months (ABS cat. no 6287.0). This pooled sample allows broad aggregates of labour force characteristics to be published at the State and Territory level. However, high standard errors are still problematic, for example for States and Territories with smaller Indigenous populations, and for remote areas. By introducing the NATSIHS sample into the pooled dataset, ASB will investigate what gains are possible in terms of reduction in sampling error and more disaggregated estimates.

However, there are a number of differences between LFS and NATSIHS, which may lead to the introduction of non-sampling error when the datasets are pooled. For example, one key difference is in the questionnaires: LFS uses a much more detailed set of questions to determine labour force status than NATSIHS. ASB will attempt to quantify the effect of this inconsistency, and investigate methods for taking questionnaire differences (i.e. measurement error) into account when pooling the data.

For more information, please contact Russell Lim on (02) 6252 7346.

Recent Activity in the Census Data Enhancement Project

All the linkage work conducted by Analytical Services Branch for four quality studies as part of the Census Data Enhancement project was completed in late October 2007. Since then, the team has been writing reports of what was done, preparing a data linking manual and performing a suite of analyses using the linked data sets. All this work is now coming to fruition with presentations at conferences and the first papers of results being released.

Glenys Bishop and Tenniel Guiver from Analytical Services Branch attended the Second National Symposium on Data Linkage Research held in Adelaide in October 2008. Their presentations on 'Sampling based clerical review methods in probabilistic data linking' and 'Determining the quality of longitudinally linked Census data' were well received for their rigorous and comprehensive approach. Sybille McKeown also spoke about some of the challenges in the Indigenous Mortality Quality Study. Glenys Bishop presented an overview of the Census Data Enhancement project at a Record Linkage Workshop conducted in conjunction with the International Association of Cancer Registries Annual Scientific Meeting in November. Glenys Bishop, Tenniel Guiver and Jeff Wright gave talks about methodological issues of data linking at the Australian Statistical Conference in July.

Two papers using results of the Indigenous Mortality Quality Study were released in November. A discussion paper on the assessment of methods for developing life tables for Aboriginal and Torres Strait Islander Australians used the results of the study to obtain a direct method for estimating Indigenous deaths. An information paper outlined the findings of the study, particularly with respect to assessing the undercoverage of Indigenous deaths in death registration records and identifying factors that may be contributing to undercoverage of Indigenous deaths in death registrations.

James Chipperfield presented a paper to the November Methodology Advisory Committee meeting on generalised linear models with probabilistically linked data. In this piece of work, James and other team members devised a way of implementing a method to adjust regression coefficients when fitting models to inexactly linked data. They were able to show their method worked but that missing links altogether had a much larger effect on the model than incorrect linkages. This paper will be published as a research paper in the near future.

For further information please contact Glenys Bishop on (02) 6252 5140.

New Version of SEASABS Released

The Time Series Analysis area (TSA) has been busy over the last few months with the testing and release of the latest version of the ABS seasonal adjustment software SEASABS. This has been a team effort between TSA, its clients and technology application staff. The main new features can be summarised as those that impact our clients via TSUpdate and those that enhance SEASABS functionality as an analysis tool for TSA. Client impacts include: (1) the removal from TSUpdate of rounding options; (2) the population of combined seasonal factors for "never adjusted" series by TSUpdate on the ABSIW; (3) the streamlining of TSUpdate log message presentation; and (4) the inclusion of an hourglass and counter facility for the download application.

SEASABS Version 2.7 has the following enhancements as an analysis tool: an improved interface to X-12-ARIMA for REGARIMA modelling, additional diagnosis for seasonal factor stability testing, improved ability to manage series which are "conceptually related" to improve consistency of analyses, enhanced Chain Volume Measure (CVM) capabilities, an increased range of spectral analyses, improved presentation of aggregation structures, a SEASABS generated queries template, the ability to copy any SEASABS table into a spreadsheet or Notes table, and increased abilities to compare graphs. These new functionalities will streamline TSA practices and increase productivity.

For more information regarding the release, please contact Lisa Apted on (03) 6222 5932, or Mark Zhang on (02) 6252 5132.

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 December 2008

First Assistant Statistician

Geoff Lee Tel: 5239

Assistant Statistician
Analytical Services Branch

Jill Charker Tel: 7290

Directors
Analytical Services Unit

Ruel Abello A/g Tel: 6307

Glenys Bishop Tel: 5140

Jonathon Khoo Tel: 5506

Analytical Services Unit (ASU)

Nick Biddle	Tel: 5404
James Chipperfield	Tel: 7301
Tetteh Dugbaza	Tel: 7221
Daniel Elazar	Tel: 6962
Tenniel Guiver	Tel: 7310
Anil Kumar	Tel: 5344
Charity Liaw A/g	Tel: 5578
Alexa Olczyk	Tel: 5854
Peter Rossiter	Tel: 6024
Richard Solon	Tel: 5917

Directors
Time Series Analysis

Mark Zhang Tel: 5132

Time Series Analysis (TSA)

Lisa Apted (TAS)	Tel:(03) 6222 5932
Sean Buttsworth A/g	Tel: 5174
Tom Outteridge	Tel: 6406
Anna Poskitt	Tel: 7954
Nick Von Sanden	Tel: 5727

Director
Data Access & Confidentiality
Methodology Unit

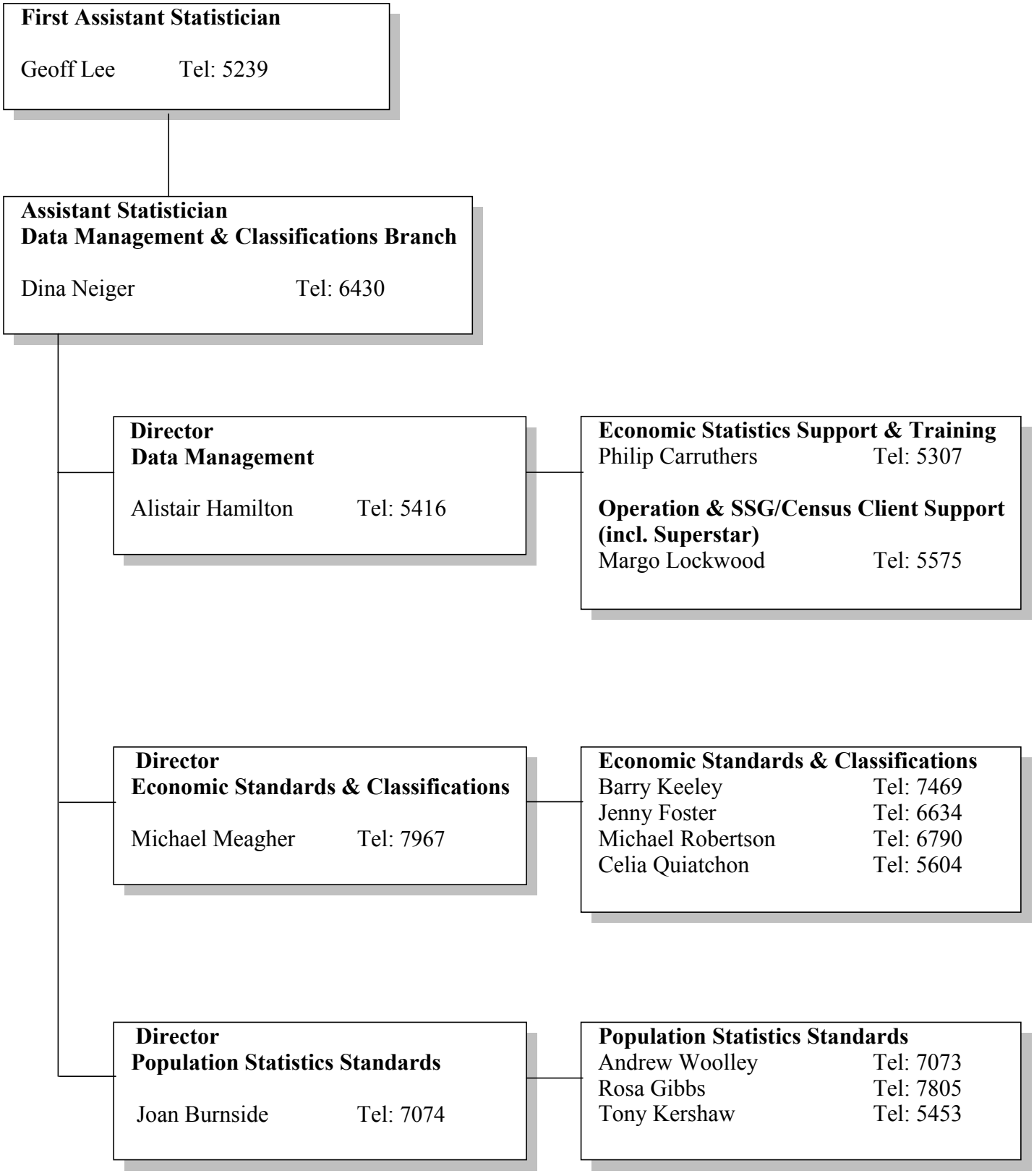
Melissa Gare Tel: 7147

Data Access & Confidentiality
Methodology Unit (DACMU)

Narrisa Gilbert	Tel: 5283
Wendy Howe	Tel: 7508
Victoria Leaver	Tel: 5445

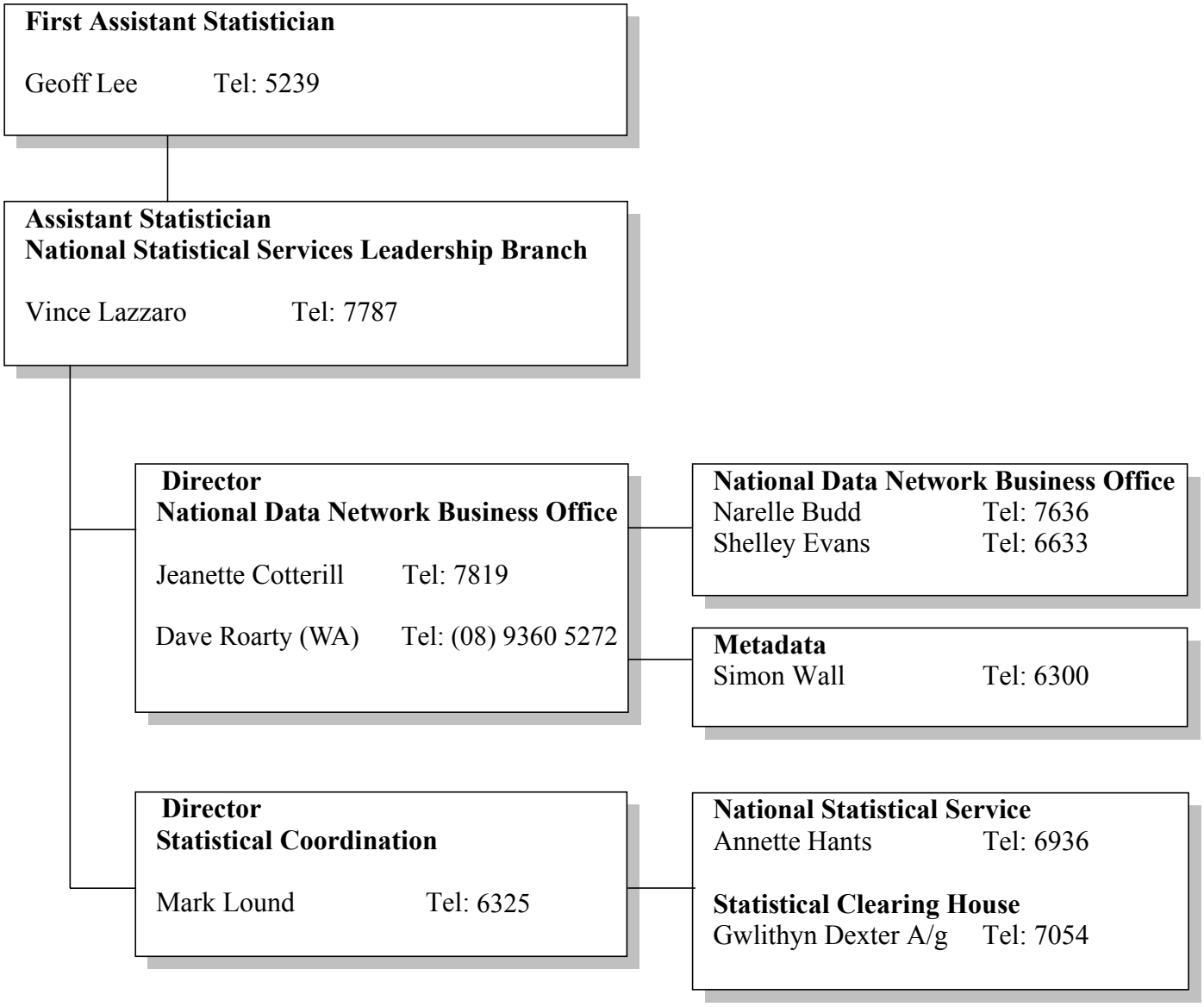
Methodology & Data Management Division
Management Structure

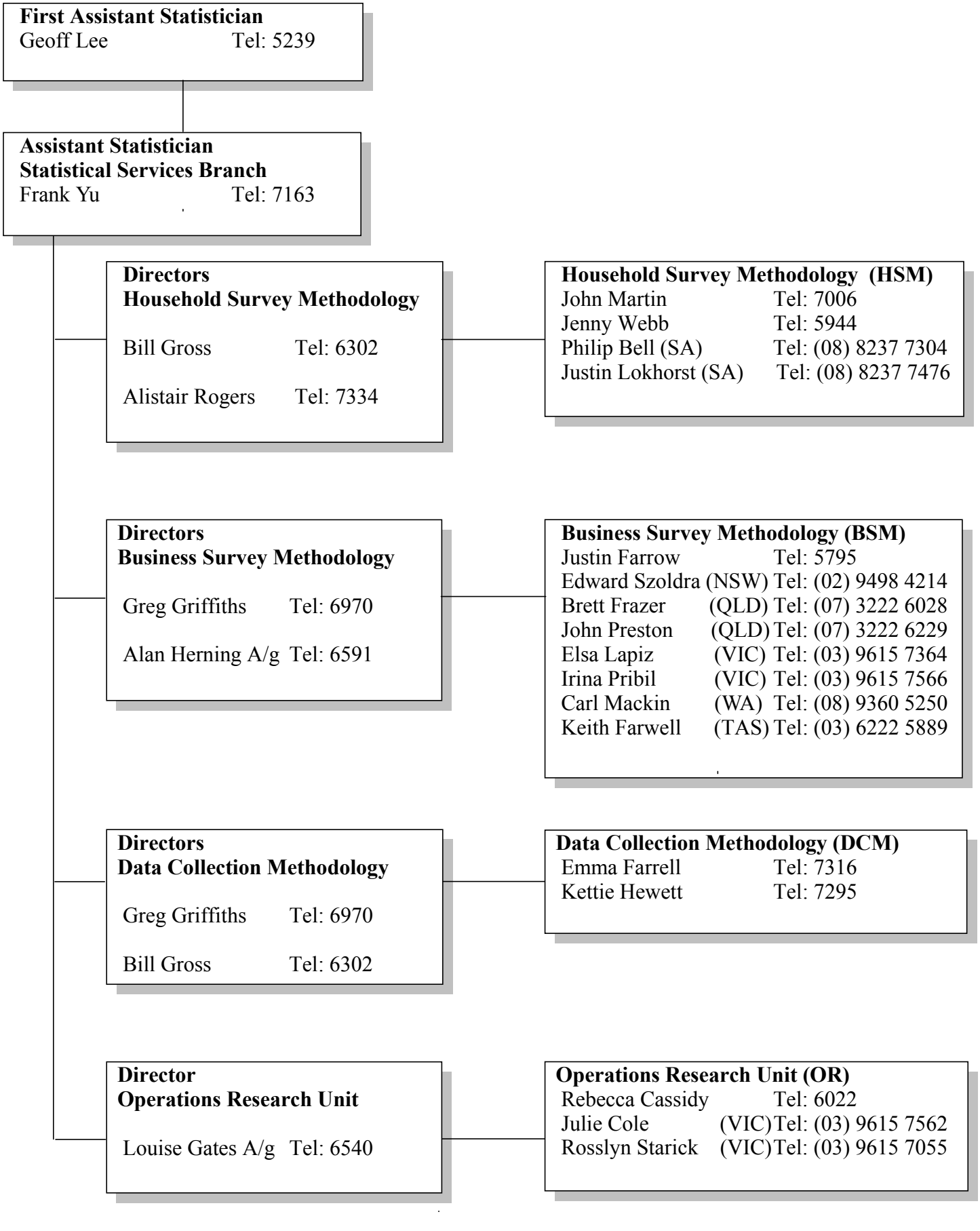
Current at December 2008



Methodology & Data Management Division
Management Structure

Current at December 2008



Methodology & Data Management Division
Management Structure

Current at December 2008