Australian Bureau of Statistics

1352.0.55.132

Research Paper

Robust Modelling of Design Effects for Household Survey Design

New Issue

Research Paper

Robust Modelling of Design Effects for Household Survey Design

Ross Watmuff, Ryan Defina and Jack Steel

Statistical Services Branch

Methodology Advisory Committee

7 June 2013, Canberra

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) FRI 7 FEB 2014

ABS Catalogue no. 1352.0.55.132

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 Ross Watmuff, Statistical Services Branch on Canberra (02) 6252 7084 or email <statistical.services@abs.gov.au>.

ROBUST MODELLING OF DESIGN EFFECTS FOR HOUSEHOLD SURVEY DESIGN

Ross Watmuff, Ryan Defina and Jack Steel
Statistical Services Branch

QUESTIONS FOR THE COMMITTEE

1. Does the analytic method of modelling design effects (DEFFs) account for changes in cluster size as appropriately as direct simulation?
2. How should we interpret and present DEFFs for appropriate, objective decision making when there are a large range of outputs of interest?
3. Are our approaches taking too much stock in our fine level intraclass correlations / DEFFs and should they be averaged or calculated at higher levels?
4. How can we evaluate and best utilise the standard error of the intraclass correlation?
5. Does the variance components model implemented (Section 3) appear methodologically sound and valid for the context of modelling design effects?
6. Can the method of developing summary parameters be more closely linked to the underlying model, rather than using a percentile approach?
7. Do the committee think the Bayesian framework has the potential to provide gains in future household sample designs?

ABBREVIATIONS

ABS	Australian Bureau of Statistics
AIC	Akaike Information Criterion
BFU	Base Frame Unit
BIC	Bayesian Information Criterion
DEFF	Design Effect
DEFT	Design Effect (square root)
FRHH	Fully Responding Households
HSS	Health Services Survey
LFS	Labour Force Survey
MAC	Methodology Advisory Committee
MCMC	Markov Chain Monte Carlo simulation
ML	Medicare Local region
MPHS	Multi-Purpose Household Survey suite
NHPA	National Health Performance Authority
PEx	Patient Experience Survey
POS	Part Of State
SE	Standard Error
SIH	Survey of Income and Housing
SRSWOR	Simple Random Sample Without Replacement
VC	Variance Components (Model)

CONTENTS

ROBUST MODELLING OF DESIGN EFFECTS FOR HOUSEHOLD SURVEY DESIGN

Ross Watmuff, Ryan Defina and Jack Steel
Statistical Services Branch

ABSTRACT

To date the Australian Bureau of Statistics (ABS) lacks a consistent and reliable set of methods to model design effects from past survey data in order to assess the sampling error properties of proposed household surveys. This paper articulates the key challenges in modelling robust design effects for household surveys and examines general solutions in the context of two recent sample design applications. Advice is sought from the Methodology Advisory Committee as to the suitability of these solutions as standard methods for ongoing and future application.

1. INTRODUCTION

1.1 DEFFs and ABS household surveys

The design effect (DEFF) is defined as a multiplicative factor which accounts for the change in variance under some proposed design when compared with that of a simple random sample (Kish, 1965):

$$\mathrm{DEFF} = \frac{Var_{\mathrm{Design}}}{Var_{\mathrm{SRSWOR}}} \; .$$

It is a useful simplifying quantity to work with when investigating sampling error properties of complex survey sample designs, as the overall variance of an estimate under a proposed design can then be expressed as the product of a complex survey DEFF and the design variance under a simple random sample of the same size (n), where the latter has a simple, closed, analytic form approximately expressed in terms of n and the population variance s^2 :

$$Var_{\mathrm{Design}} = \mathrm{DEFF} \times \frac{s^2}{n} \; .$$

There are a number of common design features that give rise to the DEFF being more or less than 1. These include stratification (which tends to lower the DEFF), geographic sample clustering (which tends to increase the DEFF), various sample selection methods such as systematic or cubic sampling and estimation techniques such as post-stratification which are assumed in the sample design.

ABS household surveys continue to utilise an area-based sampling framework and are hence characterised by geographically clustered samples. There are two main reasons for this. Firstly, face to face contact remains the preferred mode of collection for many of these surveys. The associated travel costs imposed by this constraint can be dramatically reduced by clustering the sample. Secondly, the ABS continues to operate in the absence of a list of dwellings or persons of sufficient quality to form the basis of a list-based sampling framework. Rather, areas are first sampled followed by the construction of dwellings lists of sufficient quality for these areas.

In this context, the DEFF component associated with the clustering feature of ABS household survey designs remains of key interest in developing efficient samples which appropriately trade off cost and sampling error.

1.2 Use of survey data to derive DEFFs for ABS household surveys

This paper concerns the appropriate use of historic survey data to model DEFFs for future surveys. Ideally one would have access to relevant population information (e.g. the Census) from which to model DEFFs for survey sample design. Although there are some cases where Census data can be used, for example Census labour force status for designing the monthly Labour Force Survey, it is generally unable to inform on the properties of the many statistical data items produced by ABS household surveys.

Interestingly, however, survey data has not been widely used in the ABS to derive DEFFs for household survey design. Part of the reason for this is that most population surveys have utilised the sampling infrastructure developed for the Labour Force Survey. Most of these have either been extensions of the LFS survey questionnaire (supplementary surveys) or used a 'shadow' sample within the areas selected for LFS, referred to within the ABS as the 'parallel block sample'. This approach is superficially cost efficient but results in these surveys largely inheriting the design characteristics of the LFS, somewhat negating the imperative to utilise the design information from past survey cycles that would be sought if designing these from scratch.

The 2011 Household Sample Redesign was historic in that the master sample used for Special Social Surveys was 'decoupled' from the LFS/supplementary survey master sample for the first time since the vehicle was pioneered (in the 1970s). This allows significantly greater freedom for designing samples more optimal to the specific survey of interest and encourages greater use of past survey data for design, leading to the need for standardised and defensible approaches in doing this.

1.3 Challenges in deriving robust DEFFs for sample design

The paper addresses three significant challenges encountered by the ABS in using survey data to model robust household survey DEFFs.

I. As discussed above, a key reason for deriving DEFFs in this context is to quantify the impact of area sample clustering on the design variance. Given design data from a past survey with a particular realised level of clustering, how does one appropriately model DEFFs under different cluster sizes?

II. DEFFs estimated directly from sample data can be subject to high volatility depending on the sample available in the domain of interest. Imprecision in these estimated quantities may strongly undermine the process of optimising cost and sample error which underpins various proposed designs.

III. Although some population surveys such as the Labour Force Survey have a relatively narrow set of well-defined objectives, the more common situation is to have many output objectives where outstanding priorities are not easily identified. The National Health Survey, for example, supports the calculation of population prevalences for hundreds of health variables over hundreds of output levels defined by geography, demography and time period constraints. A key challenge is to model DEFFs that enable outcomes for this large set of objectives to be in some sense optimised.

Challenges II and III are highly interrelated as having a large set of output objectives (Challenge III) can lead to the sample data being split finely into small subsets, which can result in high volatility of population quantities and variances estimated (Challenge II).

1.4 General approaches and sample design applications discussed

As a means to examining general solutions to the challenges above, the paper presents recent sample design applications for two surveys with significantly varying objectives: the Survey of Income and Housing (SIH), a well-established ABS household survey, and a new Health Services Survey (HSS) with design properties being modelled for the very first time.

Challenge I has often been tackled using a direct simulation approach by constructing datasets of varying cluster sizes. The SIH design exercise primarily focusses on Challenge I and contrasts a more analytic approach in favour over direct simulation.

Challenges II and III are often simultaneously addressed by essentially smoothing across a large number fine level DEFFs corresponding to survey outputs to achieve a broader, summary DEFFs. The SIH application used typical, primitive methods for this smoothing exercise whilst the HSS application pioneers a more sophisticated, generalisable solution using Mixed General Linear Modelling.

2. MODELLING OF DEFFS TO EVALUATE OPTIMAL CLUSTER SIZE

This section of the paper addresses Challenge I discussed in the introduction (Section 1.3), that is, how design effects from survey data have been calculated to assist in determining the optimal cluster size for a survey. Cost considerations of different cluster sizes are not presented but would need to be explored when determining the optimal cluster size. An application that was used in the sample design for the Survey of Income and Housing (SIH) 2013/14 is presented. This application while primarily focussing on Challenge I also addressed Challenges II and III briefly in a primitive manner.

2.1 Background to the Special Social Surveys Master Sample

The Special Social Surveys Master Sample is a two stage design where the first stage of selection comprises a pool of base frame units (BFUs) spread across Australia. At the second stage of selection each BFU is split into several clusters of the same size (number of dwellings) where each cluster is geographically spread throughout a BFU, i.e. it is a representative sample of the BFU (see Appendix A). Typically one cluster is then selected to be enumerated at the second stage. Note that in the Master Sample the first stage of selection is probability proportional to size (PPS) and the second stage is a Simple Random Sample of a cluster. Therefore, we have a self-weighting design which can be thought of as a single stage sample of clusters.

2.2 Simulation approach to calculating DEFFs

It is quite simple to calculate DEFFs from survey data as we can easily estimate the variance resulting from the sample design and divide this by the variance from a simple random sample. However, we want to estimate the impact on variance from changing the cluster size from a survey's previous cycle. To do this we need to know the DEFF under the previous design, as well as, those estimated under different cluster sizes. These DEFFs are then divided by the DEFF from the existing design to produce a relative design effect, which is used to evaluate the change in survey variance due to a change in the cluster size from a previous design, typically the cluster size defined on the Master Sample.

Our initial method of calculating DEFFs under different cluster sizes was a simulation approach where we simulated different cluster sizes from the survey data and estimated the associated survey variance. This was done for the Work, Life and Family Survey where we believed the variables of interest to be clustered and, consequently, that it would be desirable to have a small cluster size and enumerate many clusters.

The simulation method involved splitting the original clusters from the survey data into smaller clusters. For example, if we had a cluster of six dwellings and wanted to evaluate the impact of clusters half that size we would split the cluster into two new clusters consisting of three dwellings. This method has several disadvantages.

Firstly, by artificially constructing clusters in the manner described, simulated clusters from the same origin cluster would have very similar characteristics. In reality when reducing the cluster size the number of BFUs enumerated would be increased to maintain the original sample size. This means the simulation method is not capturing the extra between cluster variance which would actually be experienced by the inclusion of more BFUs. This means the DEFFs produced would be underestimated.

The other major drawback is we can only simulate clusters of a size smaller than the original cluster and a reasonable method to simulate larger cluster sizes was not apparent.

Another method was conceived to simulate cluster sizes, again smaller than the original. This method involves simulating clusters of different sizes by removing/dropping dwellings from existing clusters. For example, if we had a cluster of six dwellings and wanted to simulate a cluster of half that size, we would select three dwellings to be removed from the data. The remaining three dwellings would comprise the simulated half cluster.

This method is superior to the previously presented method of simulating clusters as multiple clusters are not being formed from an original cluster, which are then assumed to be different despite all originating from the same cluster.

Unfortunately, this method had its own limitations such as it leading to a reduction in sample (as some sample is discarded) and also not increasing the number of BFUs, as would occur in reality when decreasing the cluster size. Also, areas that originally had small cluster sizes were not ideal for this dropping methodology as there were issues with how best to drop dwellings from clusters consisting of one or two dwellings only.

This brings us to a general disadvantage of simulations methods, they are time consuming to conceive and code up. Also, to generalise the process for multiple surveys can also be quite difficult.

Despite these disadvantages simulations methods can be attractive as they are solely reliant upon the data itself and make few assumptions compared with other methods such as those described in Section 2.3.

Due to the issues with simulation methods a more robust approach was sought to evaluate DEFFs under different cluster sizes.

2.3 Analytical approach to calculating DEFFs

There is a well-known result equating DEFF to the average cluster size and intraclass correlation (Hansen *et al.*, 1953). This result is seen by first observing that the DEFF is given by:

$$\text{DEFF} = \frac{Var_{\text{Design}}(y)}{Var_{\text{SRSWOR}}(y)}.$$

This can then be simplified to the following:

$$\text{DEFF} = \frac{s_{\text{CT}}^2}{\overline{n} \times s_y^2}$$

where:

s_{CT}^2 is the between cluster total variance given by

$$s_{\text{CT}}^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(y_i - \frac{y}{m}\right)^2$$

y_i is the i-th cluster total,

m is the number of clusters,

y is the sum of y_i across all clusters,

y/m is the mean total per cluster; and

$\overline{n}$ is the average cluster size.

If we assume that m, the number of clusters, is large the DEFF can then be approximated to the following,

$$\text{DEFF} = 1 + (\overline{n} - 1)\delta$$

where δ is the intraclass correlation, which is a measure of the homogeneity of (or correlation between) units within a cluster. It is given by

$$\delta = 1 - \frac{s_w^2}{s_y^2},$$

s_w^2 is the within cluster variance given by

$$s_w^2 = \frac{1}{n-m}\sum_{i}^{m}\sum_{j}^{n_i}\left(y_{ij} - \overline{y}_i\right)^2,$$

and $\overline{y}_i$ is the cluster mean of the i-th cluster, i.e. $\overline{y}_i = y_i / n_i$ where n_i is the i-th cluster size.

This is a very simple formula for the DEFF which requires only the average cluster size and an estimate of intraclass correlation to determine the DEFF. Consequently, as the DEFF is so dependent on intraclass correlation a thorough understanding of the intraclass correlation is needed.

2.4 Intraclass correlation

As stated in the previous section, the intraclass correlation is a measure of the homogeneity of (or correlation between) units within a cluster, where the larger the intraclass correlation the more similar units are within a cluster, with a value of 1 meaning the units in a cluster are identical. Depending on the form of intraclass correlation used, the intraclass correlation can be slightly negative. A value near zero implies that the cluster sample is behaving like a simple random sample, i.e. the clustering effect is small. This in turn implies that units in a cluster are quite heterogeneous with little variation between cluster (small between cluster variance and high within cluster variance).

Typically, the larger size and more geographically spread a BFU the wider variety of characteristics will be exhibited in the BFU and, consequently, there will generally be a smaller intraclass correlation. As outlined in Section 2.1, a cluster is a sample of a BFU and, as a result, the within BFU variance can be estimated by the within cluster variance. As clusters are a representative sample of a BFU, clusters of any size should have the same underlying within cluster variance as the BFU it comes from. Hence, the intraclass correlation estimated from clusters should be the same as the intraclass correlation of BFUs and the size of these clusters will not change the intraclass correlation. Of course, as a cluster is a sample of a BFU an intraclass correlation estimated from this sample data will have variance associated with it, and the variance in the intraclass correlation would be larger for a smaller cluster size.

The attractiveness of the above simple formula is that we can assume the intraclass correlation is constant across all cluster sizes, negating the need for a simulation method. To use the formula, we simply need to estimate the intraclass correlation from past sample data and apply it with different cluster size options to find the resulting DEFFs. However, the ability to use this formula to evaluate the impact on variance of different cluster sizes is heavily dependent on the assumption that the intraclass correlation is constant over different cluster sizes. In the next section we analytically show the assumption holds through simulation.

2.5 Empirical evidence intraclass correlation is constant over cluster size

Using past SIH survey data and the dropping dwellings method of simulating cluster sizes (see Section 2.2) a set of 500 simulated datasets were produced by removing different dwellings from the data under each simulation. The intraclass correlation for some key SIH variables were calculated from each simulation and simulations were run for cluster sizes of three quarter, as well as a half of the original cluster size.

Table 2.1 shows results for South Australia ex-Metropolitan areas, where we have averaged the intraclass correlation over 500 simulations of three quarters and half cluster sizes (the original cluster size does not require an average as we did not need to simulate that cluster size).

2.1 Intraclass correlation – South Australia ex-Metropolitan areas

	Cluster size (cluster fraction & dwellings per cluster)		
Variable	*Original (9.6)*	*Three-quarters (7.15)*	*Half (4.82)*
Weekly household income from all sources	12.4%	12.4%	12.4%
Weekly household income from unincorporated businesses	3.5%	3.2%	2.7%
Weekly household income from wages and salaries	9.9%	9.6%	10.1%
Weekly household income from government pensions	6.1%	6.1%	5.9%
Weekly housing costs	7.8%	7.9%	8.1%

In table 2.1, Weekly household income from unincorporated businesses has the largest variation in intraclass correlation across the cluster sizes. This variable had a very low prevalence in the survey data with on average less than one dwelling with this characteristic per cluster. This led to high variation in the intraclass correlations that were estimated from the simulations and is likely why it differs the most over the simulated cluster sizes. For the remaining variables the intraclass correlation is constant across each cluster size, with only minor differences between the estimated values for each size. These results give us empirical confidence that the intraclass correlation is constant across different cluster sizes.

Note that the intraclass correlation estimated from a design with a small cluster size may have a large variance and therefore extrapolating this intraclass correlation for larger cluster sizes may not be ideal. The Standard Error of intraclass correlation is discussed in Section 2.9.

2.6 SIH survey background

SIH is a biennial survey of approximately 20,000 households across Australia, with a typical response rate of around 85% and sample loss of 14%, resulting in approximately 15,000 fully responding households. SIH produces estimates of income at many different output levels such as the broad levels of National, State and Part of State (POS), as well as at finer levels such as State by POS and by further demographic categorical variables such as household composition.

Estimates at these levels were all identified as being of interest for the SIH 2013/14 sample design, with a particular interest in estimates at State by POS by demographic categories. This poses difficulties in achieving a design which balances desirable outcomes for estimates at both broad and fine levels. For example, estimates at fine levels, such as income at State by POS by age and household composition, will exhibit little or no clustering as on average they will have one or no dwellings with that characteristic per cluster. Conversely, estimates at broader levels, such as income at the State by POS level, may be highly clustered as we would expect dwellings in a cluster to have a tendency to exhibit similar income characteristics.

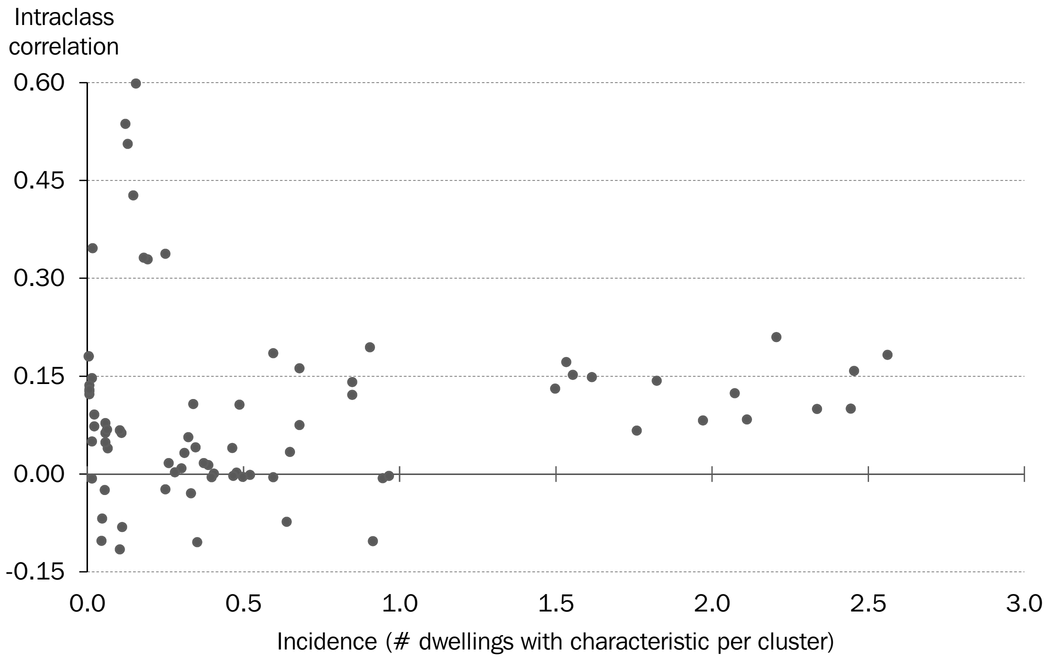
This means that different cluster sizes may be ideal for broad level estimates versus finer level estimates. This leads us to Challenge III, discussed in Section 1.3, in that estimating such a large number of intraclass correlations (see Appendix B) and consequently DEFFs, makes it quite difficult to interpret results and to recommend an optimal cluster size.

2.7 Intraclass correlation results for SIH 2013/14 design

A previous cycle of SIH sample data was used to estimate intraclass correlations at each of the key output levels for an income and housing cost variable. Figure 2.2 displays, for New South Wales ex-Metropolitan areas, the estimated intraclass correlation for each variable by output level estimate, and the incidence per cluster of the estimate, expressed as the average number of dwellings per cluster with the characteristic of the estimate.

This plot highlights Challenge II outlined in the introduction (Section 1.3) in that there is a great amount of volatility in intraclass correlation between those variable by output level estimates with an incidence per cluster less than one as they are based on few observations (e.g. an incidence per cluster of 10% equates to 40 observations of that variable). These intraclass correlations are centred close to zero, when the extreme values are ignored. This is to be expected as the within cluster variation will be very small for estimates that have an incidence per cluster less than one. Those with an incidence greater than one vary to a much smaller degree and appear to be centred around approximately 15%.

2.2 Intraclass correlation by incidence per cluster – New South Wales ex-Metropolitan areas



As noted earlier, the estimated intraclass correlation has variance associated with it. We do not know how much of the difference between intraclass correlations is due to a real difference compared to simply random variation in estimates. For example, there are several estimates with an intraclass correlation that is larger than 30%, which is unusually large. These variables are income and households cost for:

- lone households with persons greater than 64 years,
- dwelling type of flats, units, apartments,
- dwelling type of semi-detached households.

There are two reasons which may explain why the intraclass correlations are larger for these estimates. They may have a large variance associated with them due to the small sample size they are based upon. However, the large difference between them and the other intraclass correlations also suggests there may be a real difference between them. Perhaps these characteristics are highly clustered in areas, which seems plausible for these variables. For example, many apartment blocks may be constructed nearby to each other, or elderly people may live nearby to each other in retirement communities. Care needs to be taken as it can be easy to falsely convince oneself that there is a real world reason for one estimate to be more clustered than another. Consequently, an objective manner of determining a real difference between intraclass correlation should be used such as a measure of Standard Error for hypothesis testing (see Section 2.9)

These results are similar to those seen by Burden *et al.* (2011). They found that for estimates with an incidence per cluster less than one the DEFF was around one, with a high degree of variability between the DEFFs. Although we have focussed on intraclass correlation instead of DEFF, similar results were observed for SIH as small intraclass correlations result in DEFFs around one (assuming a moderate cluster size).

Intraclass correlation was produced each State in ex-Metropolitan areas for a suite of income and housing cost variables at different demographic categorical variable output levels. A plot of these results can be seen in Appendix C.1.

The plot shows that, not only is there variability in intraclass correlation across different estimates within an area, but also across areas for a given estimate. This highlights the real difficulty in interpreting such a multitude of intraclass correlation estimates. The assumptions that have been made to estimate intraclass correlation, such as those highlighted in previous sections, may be of little consequence given the variance associated with estimated intraclass correlation and difficulty in interpreting the multiple estimates of intraclass correlation produced. As a result, some effort was given determining how the results should be presented to make informed sample design decisions.

2.8 Smoothing of intraclass correlation variability

The large variability across the estimated intraclass correlations motivated the development of a crude process of smoothing where extreme values of intraclass correlation were brought into the typical value range. This helped to lessen the impact on design decisions of extreme intraclass correlations.

The smoothing process was applied separately to Metropolitan and ex-Metropolitan areas. The process involved identifying, for each variable by demographic categorical variable estimate, the median intraclass correlation across the States and grouping variables with similar medians together. The interquartile range of intraclass correlations was found for each group and those outside the interquartile range were identified as outliers and smoothed. Smoothing was performed by reducing the amount an outlier was outside the interquartile range by an arbitrarily set 80%. In Appendix C.2 the intraclass correlations that resulted from the smoothing process are displayed.

Comparing the smoothed intraclass correlations with those pre-smoothing (C.1 *vs* C.2) a much smaller amount of variability in the intraclass correlations is observed. By smoothing the intraclass correlations we have preserved some of what may be real world difference between States or variables at different output levels but reduced the variability, which hopefully reflects reducing the volatility in the estimation of intraclass correlation and subsequently DEFF. Advice was also given to survey areas that when interpreting results care should be given to look at overall patterns of results rather than focussing on any individual result for a particular estimate.

The method of smoothing estimated intraclass correlation for SIH was extremely crude, nonetheless, this was considered fit for purpose for the sample design given timing and resource pressures at the time. A more sophisticated approach would have ideally been developed and one such example is presented in Section 3.

2.9 Standard error of intraclass correlation

An estimate of the variation of each intraclass correlation could be used to determine whether differences between them are statistically significant. This would be greatly beneficial in interpreting intraclass correlation and presenting results.

We attempted to estimate the variance of the intraclass correlation but were not successful and had limited time to investigate further. The investigation made use of the simulated datasets that were produced to confirm the intraclass correlation was constant across different cluster sizes (see Section 2.5). These datasets were used as they had multiple simulated samples of the same cluster size. This meant we could estimate the intraclass correlation for each sample and find the variance across them. Unfortunately, this only meant we were estimating the variance of intraclass correlation for subsamples of our cluster sample, not the variance from selecting different samples. For example, we found that the variance of intraclass correlation was approaching zero the closer the cluster size to the original on the data, when in fact there would still be variance associated with sampling different clusters.

There is a result for the Standard Error (SE) of the intraclass correlation (Donner, 1986) where it is given by

$$SE(\delta) = (1-\delta)\left[1+(\bar{n}-1)\delta\right]\sqrt{\frac{2}{\bar{n}(\bar{n}-1)(m-1)}} \; .$$

This result assumes a design with equal cluster sizes. Donner also presents a result for unequal cluster sizes, but this result is very complex and depends on more parameters. The simplicity of the equal cluster size result is very attractive as it relies upon a few known or easily produced qualities and hence should be easy to calculate. Unfortunately, these results for the SE of intraclass correlation were discovered after the sample design work for SIH was completed and to date has not been explored in great detail for their appropriateness and usefulness.

3. MODELLING OF DEFFS WITH INSUFFICIENT SURVEY DATA

This section of the paper addresses Challenges II and III discussed in the introduction (Section 1.3). These are aligned with instances where:

- Design data is not deemed to be fit for the purpose of deriving robust design effects, due to insufficient data being available at required design levels. This is largely driven by a desire to implement a relatively fine level of stratification where design data is in some cases very scant or non-existent.
- Quite a large suite of design constraints are required to be satisfied. Such constraints highlight priority data items as well as detail the level of dissemination in terms of geography. Considerations need to be made to ensure that even the most optimistic constraint is likely to be satisfied.

It is asserted that a parametric modelling approach will provide a framework for which these challenges can be explicitly addressed. Such an approach is defined, within the confines of this section, as the modelling of design effects under the assumption that cluster size and intraclass correlation can be treated as fixed. Hence Challenge I is not addressed. DEFFs are instead treated as unique and informative measures without a need to partition into components.

3.1 An overview of the Health Services Survey (HSS)

The ABS has been contracted to run a new survey addressing the insufficient data available on patient experience within the health services industry. It acts as a top-up of the Patient Experience (PEx) survey, simply allowing for more accurate estimates to be formed over multiple small areas and design variables.

Medicare Local (ML) regions are seen as the desired level of geography required in terms of outputs. In light of these areas presenting as quite small (in both area and population), the survey is being asked to produce fine level estimates that traditionally would be not be derived through a direct collection method.

More specifically, the key output requirements of HSS are sub-domain estimates within an ML region subject to a common stratum-level constraint dependent on relative domain size.

There are 61 ML regions and 52 design variables in total, for which each combination is required to satisfy the above constraints, dependent on their estimated relative domain size.

3.2 Challenges of modelling suitable DEFFs for small area domains

Ideally a sample design requires Census data or at worst, sample data of high precision, in order to form estimates of the stratum-level population characteristics. This point is magnified for the case of estimating DEFFs for small area domains. With HSS, two years of PEx data was readily available for use in design. The nature of a top-up arrangement ensured that the questions asked previously in PEx lined up well with the questions intended to be asked in HSS. Unfortunately the down side was that PEx was designed at a state level, and hence, many ML regions had undesirably low sample data available from which to infer about the relevant population.

The problem is also one of high dimension. PEx design data has 52 variables, matched to the 61 ML regions, across two years, i.e. roughly 6000 records. Potential demographic indicators for use in explaining DEFF variation (candidates for model covariates) within the data are also limited.

The broad objective of the research investigation conducted was to settle upon smooth and reliable estimates of DEFF at the ML level. This leads to a robust sample allocation. Approaches applied therefore focussed on smoothing out the volatility in the small sample estimates derived from design data. A reduction of residual noise via parametric models was seen as an appropriate starting point.

Given insufficient *a priori* information to suggest that population variances within these ML regions differ, an equal allocation was settled upon as a starting point for design.

A rough skirmish, using a constant design effect of 1.44 and conservative prevalence of 50% yielded an equal allocation of approximately 400 Fully Responding Households (FRHH) per ML region. As such, at an aggregate level, the final allocation should align well with this approximation – costings were conducted using these figures. The DEFF of 1.44 was determined through naively averaging design effects across three historical data items, and also across ML regions, extracted from readily available PEx data. However this level of analysis would not suffice given the unique properties of each ML region by design variable combination, and hence differing clustering characteristics.

The following sections will detail a means to ensure an adequate trade-off between robustness and responsiveness is achieved. This aims to produce smooth and reliable DEFFs, all while retaining useful ML level properties.

3.3 A linear random effects model for DEFFs

A model is required in order to be able to 'predict' design effects based on some explanatory covariates. The approach taken within the following analysis selects the root design effect (or DEFT = $\sqrt{\text{DEFF}}$) as the response variable. These DEFTs are calculated for each design variable by ML combination, pooling across the two years of design data. Consequently, the outputs from the chosen model will be predictions for each of these DEFTs. One asserts that the predictions should be substantially less volatile than that of directly calculated DEFTs which are based on minimal design data at these finer levels.

A key consideration behind choosing suites of models was the interest in capturing correlation that existed between design data consisting of a common ML or design variable. Hence the family of linear random effects models was prioritised due to its flexibility and ability to explicitly capture such properties, while retaining a sense of simplicity in the analysis. Skinner, Holt and Smith (1989) identify regression models as appropriate for modelling DEFFs in certain cases, opening up the scope for further discussion in this paper.

Formally the model is specified by the following equation:

$$y = X\beta + Z\gamma + \varepsilon$$

where

y is the dependent variable (root design effects),

X is a design matrix of the fixed effects (intercept, median age, median age by remoteness interaction),

β is a vector of coefficients for the fixed effects,

Z is a design matrix of random effects (Medicare Local and design variables),

γ is a vector of coefficients for the random effects, and

ε is a vector of residual errors.

3.4 Deriving DEFF inputs for the model

In order to calculate initial DEFFs upon which to construct the model , jack-knife variance estimation was first applied to the design data (PEx), with the resulting quantity forming an approximation to the clustered variance. DEFFs were then directly calculated by dividing the SRSWOR (or Horwitz–Thompson) variance estimate defined as:

$$
\begin{aligned}
SE_{\text{HT,SRS}}\left(\hat{p}_{d\pi}\right) &= \left[\frac{1}{n_d}\left(1-f_d\right)\frac{n_d}{n_d-1}\,p_d\left(1-p_d\right)\right]^{1/2} \\
&\cong \left[\frac{1}{n_d}\,p_d\left(1-p_d\right)\right]^{1/2}
\end{aligned}
$$

where

p_d is the proportion of the sample in the domain *d,*

n_d is the sample size in domain *d,*

f_d is the sampling fraction in domain *d.*

This can be conservatively estimated through maximising the $p_d\left(1-p_d\right)$ term at p_d equal to a half, essentially assuming a worst case scenario:

$$
SE_{\text{CONSERVATIVE}}\left(\hat{p}_{d\pi}\right) = \frac{1}{2\sqrt{n_d}} \; .
$$

The ratio of the clustered variance to the SRSWOR variance forms the DEFF estimator.

It's important to understand that a key assumption has been made. Here one has assumed that the simulation volatility makes up a relatively small component of the clustered variance estimator. A large number of replicate groups were run in an attempt to minimise the impact of simulation volatility. Risk is also present in that the variance estimator is thought to be downward biased and therefore form an inappropriately optimistic estimator. This is a result of the inability to factor in the systematic skip selection mechanism within a replication framework.

3.5 Selection of model covariates

A preliminary set of model covariates was selected through the use of a simple linear regression model, incorporating a naïve independence assumption. This was used as more of a guide rather than a formal test under the assumption that such a simplistic model would be insufficient for our purposes. Median age within the ML region, Remoteness classification (based on relative access to services), Socio Economic Index For Areas (SEIFA) and State by Area type (Metropolitan / ex-Metropolitan split) were the variables found to be significant predictors of the PEx design effects.

Seen as a small increase in the complexity of underlying covariance matrix representation, a variance components (VC) model was then applied including the covariate combination aforementioned. Note that the variance components chosen are at the ML and design variable level. This is due to the existence of multiple PEx records available for each of these two components.

In order to further refine the choice of model covariates, an information criterion was applied, as opposed to the most obvious alternative – backward/forward selection. Details of the covariate comparison process are tabled below (for only the best performing combinations), suggesting Median Age alone was the best option. The Bayesian Information Criterion (BIC) was seen as the best criterion given that it assigns preference towards a more parsimonious model than alternatives such as the Akaike Information Criterion (AIC).

3.1 Covariate comparison process

	Included covariates				
Model no.	*Interaction*	*Age*	*Remote*	*BIC*	*BIC rank (lower is better)*
1	YES	YES	YES	2,366	6
2	YES	NO	YES	2,361	3
3	NO	YES	YES	2,353	2
4	YES	YES	NO	2,366	6
5	NO	NO	YES	2,368	7
6	YES	NO	NO	2,366	6
7	NO	YES	NO	2,348	1

The next step involved an assessment of whether an alternative covariance matrix assumption would drive gains in the modelling, conditioned on the covariate suite already chosen. Options tested included the compound symmetric and the independent, which are both simpler models with less capability of capturing correlation (zero correlation for the latter). More complex variance assumptions were also tested, see table 3.2 below. The VC model was retained as it exhibited the lowest BIC.

3.2 Variance assumptions

Model type (covariance structure)	*{i,j}th element*	*BIC*	*BIC rank (lower is better)*
Variance components	$\sigma_{ij} = \sigma_k^2 \, 1(i = j)$	2,348	1
Compound symmetric	$\sigma_{ij} = \sigma_1 + \sigma^2 \, 1(i = j)$	2,349	2
Autoregressive – AR(1)	$\sigma_{ij} = \sigma^2 \rho^{\lvert i-j \rvert}$	2,350	3
Independent	$\sigma_{ij} = \sigma^2 \, 1(i = j)$	2,476	4
Unstructured	σ_{ij}	***	N/A

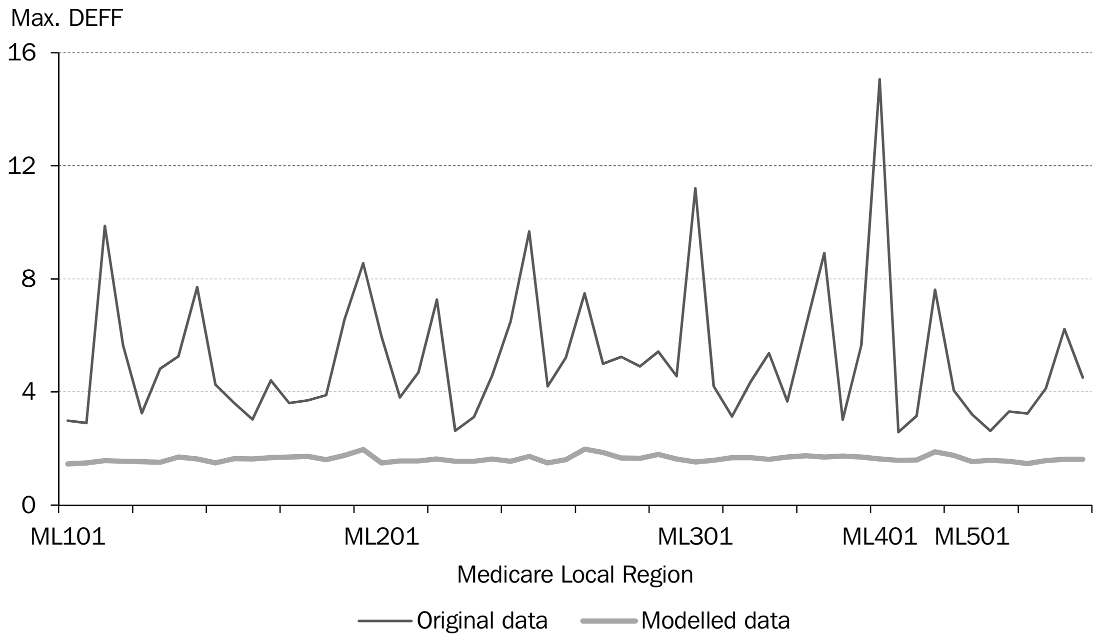
*** The unstructured case was unable to fit stable estimates due to large quantity of unknowns in the covariance matrix.

3.6 Developing summary parameters

A clear drawback from this parametric modelling approach is that one is provided with a predicted design effect for each record of design data i.e. multiple values per ML region. The key outcome is to derive one single figure per ML region to use in allocating sample. Indeed design effects are only a useful metric at some informative aggregate level, often aligning closely with the stratification in the context of sample design. This paper will refer to the output from such a dimensionality reduction process as a suite of summary parameters.

Given that constraints occur for all variables at the ML level, a conservative approach was once again taken in the development of summary parameters. Therefore the maximum modelled (or predicted) DEFF at the ML level was extracted – this is more formally the predicted DEFF output from the variance components model. Upon plotting against the raw DEFF, purely based on the design data with zero modelling, clear improvements were observed in terms of volatility being smoothed both within and between ML regions.

3.3 Maximum design effects per Medicare Local region for modelled vs original data*



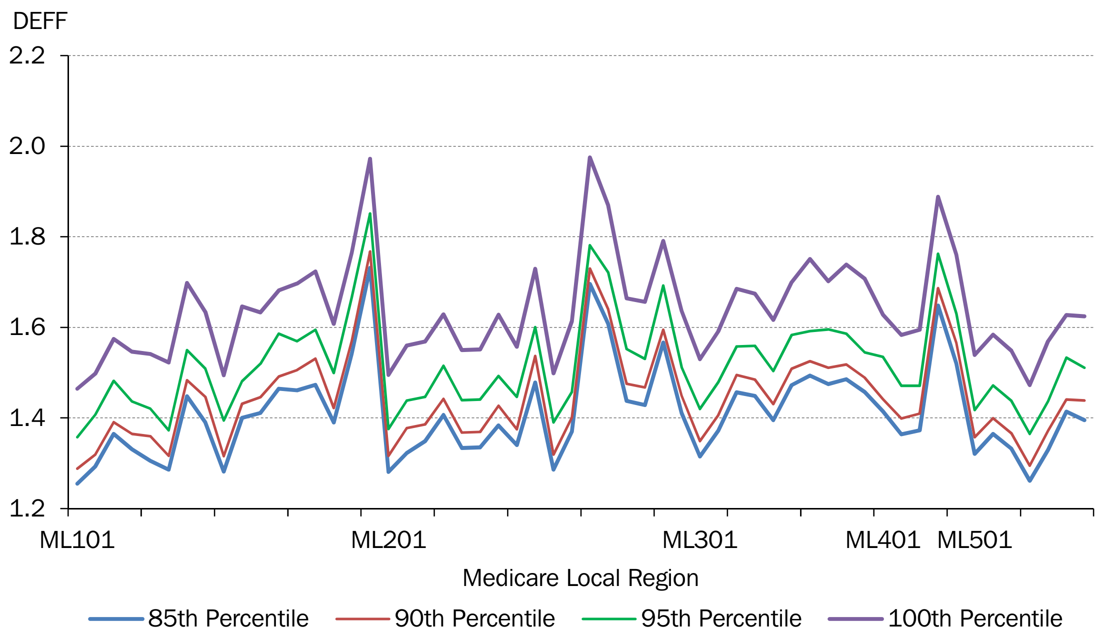
* **Modelled Data** refers to the maximum predicted design effect as specified by the variance components at the ML level. **Original Data** refers to the maximum design effect based purely on the design data at the ML level.

Reasons why the maximum modelled values are a better option than the raw DEFFs for use in sample allocation include:

1. A reduction in volatile spiking meaning they're more likely to align with the true underlying population design effects at the ML level.
2. Robustness against small amounts of design data sample causing volatile estimates.
3. Robustness against unusual / outlier observations in the sample data.
4. When fed into allocation algorithms, we will not see drastic changes in sample among ML regions.

Consideration was also given to alternative summary parameters. The following graph compares the maximum modelled value (100th percentile) against the 95th, 90th and 85th percentile of modelled design effect values. Clearly the further you move from the 100th percentile, the less likely you are to encounter an unusual estimate in the upper tail of the modelled values' distribution. This also infers that the 85th percentile is much more robust to volatile sample data than the 100th percentile.

3.4 Percentiles of predicted design effects from variance components model (at ML level)



In deciding which percentile to use as the summary parameter, the followings issues were prioritised;

1. DEFFs in the vicinity of 2.0 were judged as being excessive, and thus guarded against (even 1.8 was considered somewhat high).
2. Sufficient smoothness is present to warrant approximate robustness in the resulting sample allocation.
3. Differences between ML level modelled DEFFs still needed to be present in order to capture ML-specific differences in clustering properties.
4. After converting to expected FRHHs, the HSS and PEx sample combined should approximate total 400 FRHH (from the initial skirmish analysis).
5. After converting, the total expected FRHH should align with agreed costings, to be therefore financially viable.
6. DEFFs should typically fluctuate around the design effect calculated in the initial skirmish process, in this case, within the vicinity of 1.44.

The 90th percentile was viewed as an appropriate trade-off between each of the above methodological and financial/operation issues.

It's important to highlight here that the processes implemented have ensured an equal allocation will not be settled upon for the design. This is seen a positive move away from a somewhat naïve preliminary assumption, to a more optimised design that draws robust conclusions from the design data. Nonetheless, there is an acknowledgement that the skirmish design was useful in articulating design requirements and hence proved a useful starting point.

3.7 Comments on model suitability

It must be stated that the method applied in this section of the paper was highly tailored to the case of finer-level stratification with minimal design data at these levels. One could easily argue that such a method is excessively complex for instances where reliable design data is available and/or excessive volatility in design effect calculation is not present.

The methods implemented were able to effectively smooth design effects in order to produce robust estimates for sample design. Smoothing is achieved via two mechanisms.

Mechanism A: Application of the variance components model produces predicted values, which reduce the volatility at the unit level. This is due to parsimony in the model, and therefore, a tendency to tend towards to the overall mean (conditioned on observed covariate values).

Mechanism B: The approach of taking the 90th percentile of maximum predicted values, smoothing the design effects further, this time at the ML level.

Overall, Mechanism A completes a vast bulk of the smoothing, bringing volatility of estimated design effects down to usable levels. Mechanism B purely acts as a fine-tuning exercise to ensure useful properties and aggregates are retained.

4. FUTURE DIRECTIONS

4.1 A small area estimation treatment

It is possible to consider the analysis of design effects with insufficient sample as a small area estimation (SAE) problem with both spatial and non-spatial elements. One can trivially identify an opportunity to formally specify and interrogate correlations existing between ML regions in close proximity. Standard SAE methods can cater for this through generalised linear mixed models, with allowance for complexity in the underlying variance-covariance matrices. Such a treatment would not hinder the analyst's ability to include fixed components and other random components that are non-spatial.

Such a treatment goes well beyond the scope of this paper, given the relatively parsimonious treatment mentioned previously. Clearly one must consider these cases on an individual basis, ensuring that generalisation does not over-parameterise a problem that can be handled by methods implemented in the past.

4.2 Standard error of intraclass correlation / DEFF

As discussed a measure of Standard Error (SE) for intraclass correlation, or DEFF, would be of great use. It could be used simply for hypothesis testing of the differences between intraclass correlations or in a more sophisticated manner such as in the modelling of design effects, as outlined in Section 3. Consequently, future research should be undertaken on how to estimate the SE of the intraclass correlation/DEFF and the best uses of these estimates.

A method of analytically estimating the SE of the intraclass correlation should explored further and compared against the SE estimated from the result given by Donner in Section 2.9. The SE could perhaps be analytically produced by a replicate variance method, i.e. a jackknife type procedure. If the analytical method produces similar estimates of the SE to the result given by Donner then we could be confident of using this simple formula in the future as best practice for calculating the SE of the intraclass correlation.

Discussion from the committee on methods to calculate the standard error of the intraclass correlation or DEFF would be valuable, as well as suggestions on the best uses of the standard error.

4.3 Further research on modelling DEFFs to evaluate optimal cluster size

A goal of this paper is to move towards a standard, best practice approach to calculating DEFFs under different cluster sizes to evaluate optimal cluster size (as discussed in Section 2). As a result, we are interested in the Methodology Advisory Committee's thoughts of the methodology used, particularly ways to improve our methodology.

In our optimal cluster size investigation that was performed for SIH, Challenges II and III were only addressed briefly through a crude smoothing methodology. Section 3 overcomes this problem through a sophisticated modelling approach. In future research this approach could be explored in the context of evaluating DEFFs under different cluster sizes. Does the Committee think this would be beneficial? Is there an alternative method to account for Challenges II and III, whilst still making use of the analytical approach to calculating DEFFs (Section 2.3)?

ACKNOWLEDGEMENTS

The authors wish to thank the following individuals:

- Bill Gross for his helpful comments and suggestions regarding the methodology discussed in the paper.
- Professor David Steel, whose recommendations throughout the sample design process for SIH 2013/14 and HSS motivated much of the discussion reflected in the paper.
- Benedict Cusack, Chris McKendry and Tatiana Surzhina for their significant contribution to the development of subject matter in Section 3 of this paper as well as the derivation of many crucial analytical results for the HSS sample design.
- Robert Poole for his guidance on the SIH sample design investigation, as well as for the development of the subject matter for this paper.

REFERENCES

Burden, S.; Probst, Y.; Steel, D. and Tapsell, L. (2011) "The Impact of Complex Survey Design on Prevalence Estimates of Intake of Food Groups in the Australian National Children's Nutrition and Physical Activity Survey", *Public Health Nutrition*, 15(8), pp. 1362–1372.

Donner, A. (1986) "A Review of Inference Procedures for the Intraclass Correlations Coefficient in the One-Way Random Effects Model", *International Statistical Review*, 54(1), pp. 67–82.

Hansen, M.H.; Hurwitz, W.N. and Madow, W.G. (1963) *Sample Survey Methods and Theory, Volumes 1 and 2*, John Wiley, New York.

Kish, L. (1965) *Survey Sampling*, John Wiley, New York.

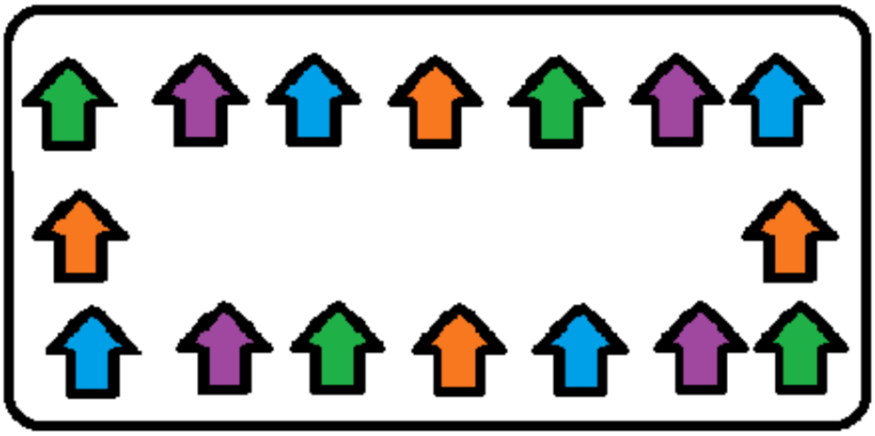
Skinner, C.; Holt, D. and Smith, T. (1989) *Analysis of Complex Surveys*, John Wiley & Sons Ltd.

APPENDIXES

A. BASE FRAME UNIT AND CLUSTERS DIAGRAM

The diagram below displays how a Base Frame Unit (BFU) is constructed. Each coloured house comprises a cluster within a BFU (the rectangle). This BFU contains four clusters (green, purple, blue and orange) each comprising of four dwellings. Each of the clusters are geographically spread across the BFU.

A.1 Construction of a Base Frame Unit



B. SURVEY OF INCOME AND HOUSING 2013/14 SAMPLE DESIGN

B.1 List of variable by category estimates of key interest for SIH 2013/14 sample design

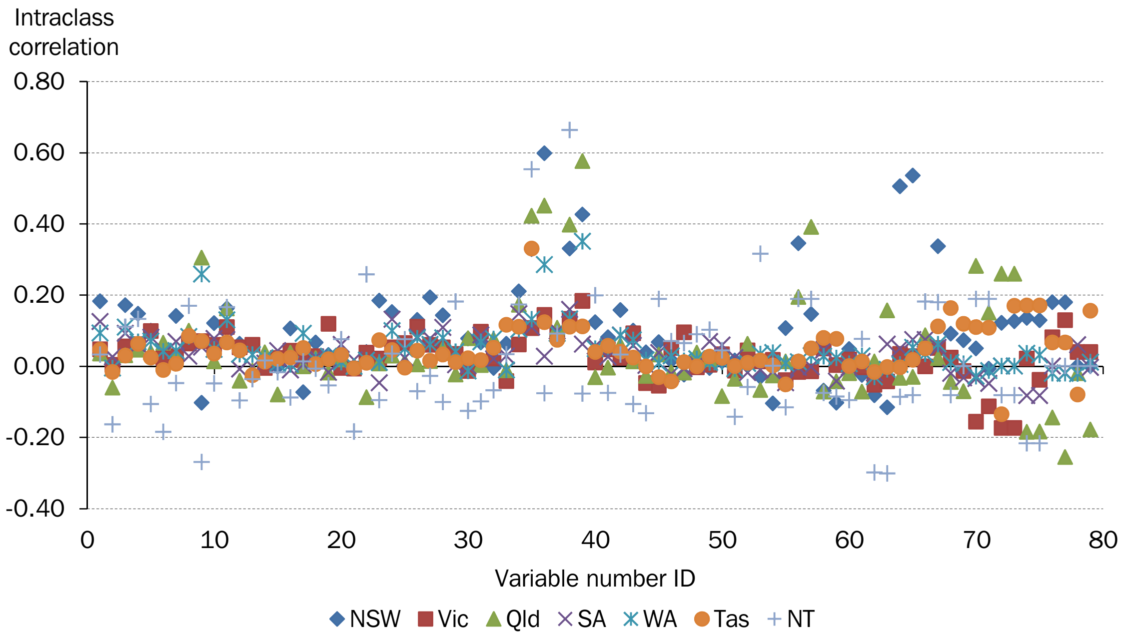
Variable ID	*Variable*	*Categorical variable*	*Category*
1	Weekly HH Income from All Sources	none	none
2	Weekly HH Income from Unincorporated Businesses	none	none
3	Weekly HH Income from Wages & Salaries	none	none
4	Weekly HH Income from Government Pensions	none	none
5	Weekly Housing Costs	none	none
6	Weekly HH Income from All Sources	Tenure Type	1. Owner without a mortgage
7	Weekly HH Income from All Sources	Tenure Type	2. Owner with a mortgage
8	Weekly HH Income from All Sources	Tenure Type	5. Renter
9	Weekly Housing Costs	Tenure Type	1. Owner without a mortgage
10	Weekly Housing Costs	Tenure Type	2. Owner with a mortgage
11	Weekly Housing Costs	Tenure Type	5. Renter
12	Weekly HH Income from All Sources	Age Group	<25 years
13	Weekly HH Income from All Sources	Age Group	25–34 years
14	Weekly HH Income from All Sources	Age Group	35–44 years
15	Weekly HH Income from All Sources	Age Group	45–54 years
16	Weekly HH Income from All Sources	Age Group	55–64 years
17	Weekly HH Income from All Sources	Age Group	> 64 years
18	Weekly Housing Costs	Age Group	<25 years
19	Weekly Housing Costs	Age Group	25–34 years
20	Weekly Housing Costs	Age Group	35–44 years
21	Weekly Housing Costs	Age Group	45–54 years
22	Weekly Housing Costs	Age Group	55–64 years
23	Weekly Housing Costs	Age Group	> 64 years
24	Weekly HH Income from All Sources	Labour Force Status	1. Employed
25	Weekly HH Income from All Sources	Labour Force Status	3. Not in the labour force
26	Weekly Housing Costs	Labour Force Status	1. Employed
27	Weekly Housing Costs	Labour Force Status	3. Not in the labour force
28	Weekly HH Income from All Sources	Household Composition	11. One family household with only family members present
29	Weekly HH Income from All Sources	Household Composition	31. Lone person household
30	Weekly HH Income from All Sources	Household Composition	32. Group household
31	Weekly Housing Costs	Household Composition	11. One family household with only family members present
32	Weekly Housing Costs	Household Composition	31. Lone person household
33	Weekly Housing Costs	Household Composition	32. Group household
34	Weekly HH Income from All Sources	Dwelling Type	11. Separate house
35	Weekly HH Income from All Sources	Dwelling Type	Semi Detached
36	Weekly HH Income from All Sources	Dwelling Type	Flat, Units, Apartments
37	Weekly Housing Costs	Dwelling Type	11. Separate house
38	Weekly Housing Costs	Dwelling Type	Semi Detached
39	Weekly Housing Costs	Dwelling Type	Flat, Units, Apartments

B.1 List of variable by category estimates of key interest for SIH 2013/14 sample design (continued)

Variable ID	*Variable*	*Categorical variable*	*Category*
40	Weekly HH Income from All Sources	Oldest Dependent Child	< 11 years
41	Weekly Housing Costs	Oldest Dependent Child	< 11 years
42	Weekly HH Income from All Sources	Youngest Dependent Child	< 16 years
43	Weekly Housing Costs	Youngest Dependent Child	< 16 years
44	Weekly HH Income from All Sources	HH composition by age group	11. One family household, <25 years
45	Weekly Housing Costs	HH composition by age group	11. One family household, <25 years
46	Weekly HH Income from All Sources	HH composition by age group	11. One family household, 25–34 years
47	Weekly Housing Costs	HH composition by age group	11. One family household, 25–34 years
48	Weekly HH Income from All Sources	HH composition by age group	11. One family household, 35–44 years
49	Weekly Housing Costs	HH composition by age group	11. One family household, 35–44 years
50	Weekly HH Income from All Sources	HH composition by age group	11. One family household, 45– 54 years
51	Weekly Housing Costs	HH composition by age group	11. One family household, 45– 54 years
52	Weekly HH Income from All Sources	HH composition by age group	11. One family household, 55–64 years
53	Weekly Housing Costs	HH composition by age group	11. One family household, 55–64 years
54	Weekly HH Income from All Sources	HH composition by age group	11. One family household, > 64 years
55	Weekly Housing Costs	HH composition by age group	11. One family household, > 64 years
56	Weekly HH Income from All Sources	HH composition by age group	31. Lone person household, <25 years
57	Weekly Housing Costs	HH composition by age group	31. Lone person household, <25 years
58	Weekly HH Income from All Sources	HH composition by age group	31. Lone person household, 25–34 years
59	Weekly Housing Costs	HH composition by age group	31. Lone person household, 25–34 years
60	Weekly HH Income from All Sources	HH composition by age group	31. Lone person household, 35–44 years
61	Weekly Housing Costs	HH composition by age group	31. Lone person household, 35–44 years
62	Weekly HH Income from All Sources	HH composition by age group	31. Lone person household, 45–54 years
63	Weekly Housing Costs	HH composition by age group	31. Lone person household, 45–54 years
64	Weekly HH Income from All Sources	HH composition by age group	31. Lone person household, 55–64 years
65	Weekly Housing Costs	HH composition by age group	31. Lone person household, 55–64 years
66	Weekly HH Income from All Sources	HH composition by age group	31. Lone person household, > 64 years
67	Weekly Housing Costs	HH composition by age group	31. Lone person household, > 64 years
68	Weekly HH Income from All Sources	HH composition by age group	31. Group household, <25 years
69	Weekly Housing Costs	HH composition by age group	31. Group household, <25 years
70	Weekly HH Income from All Sources	HH composition by age group	31. Group household, 25–34 years
71	Weekly Housing Costs	HH composition by age group	31. Group household, 25–34 years
72	Weekly HH Income from All Sources	HH composition by age group	31. Group household, 35–44 years
73	Weekly Housing Costs	HH composition by age group	31. Group household, 35–44 years
74	Weekly HH Income from All Sources	HH composition by age group	31. Group household, 45–54 years
75	Weekly Housing Costs	HH composition by age group	31. Group household, 45–54 years
76	Weekly HH Income from All Sources	HH composition by age group	31. Group household, 55–64 years
77	Weekly Housing Costs	HH composition by age group	31. Group household, 55–64 years
78	Weekly HH Income from All Sources	HH composition by age group	31. Group household, > 64 years
79	Weekly Housing Costs	HH composition by age group	31. Group household, > 64 years

C. INTRACLASS CORRELATIONS

C.1 Intraclass correlation for each State in ex-Metropolitan areas for each variable by output level estimate



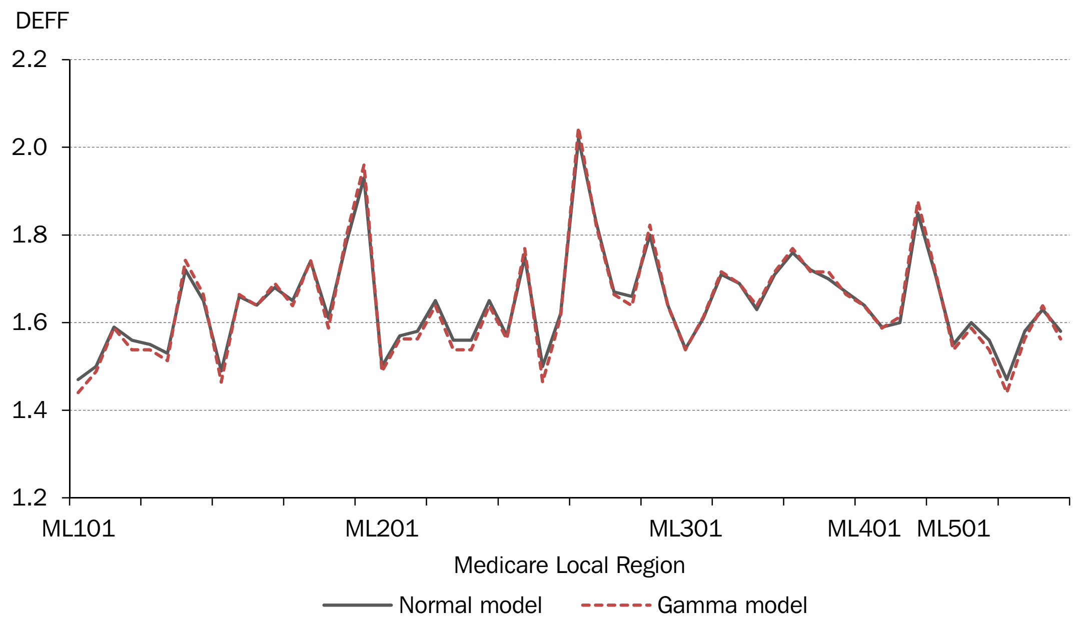
C.2 Smoothed intraclass correlation for each State in ex-Metropolitan areas for each variable by output level estimate

D. A COMMENT ON NON-NORMAL MODELS

As a test of the suitability of linearity in the mixed model assumptions applied in Section 3, comparison to a Gamma response distribution was made. A consistent treatment was ensured through specifying the same parameterisation (except response distribution) as the chosen linear model. The Gamma distribution seems appropriate as it exhibits right skew properties often encountered in design effects.

The diagram in the graph below highlights minimal differences in the maximum modelled design effects under what is a more complex model. As is a common theme in Section 3 of this paper, parsimony was given preference, yielding the Gamma option undesirable. Formal tests of fit were avoided as likelihood measures, and hence information criterion, are not available. A pseudo-likelihood approach was used for the parameter estimation in the Gamma model.

D.1 Maximum modelled DEFFs for Normal vs Gamma models



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.
LIBRARY	A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

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, or purchase a hard copy publication. 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