Australian Bureau of Statistics

1351.0.55.006

Research Paper

Small Area Estimation of Disability in Australia

100 years of STATISTICS

www.abs.gov.au

New Issue

Research Paper

Small Area Estimation of Disability in Australia

Daniel Elazar and Lewis Conn

Analytical Services Branch

Australian Statistical Conference

11–16 July 2004, Cairns, Australia

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) THURS 13 OCT 2005

ABS Catalogue no. 1351.0.55.006
ISBN 0 642 48090 7

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Daniel Elazar, Analytical Services Branch on Canberra (02) 6252 6962 or email <daniel.elazar@abs.gov.au>.

CONTENTS

SMALL AREA ESTIMATION OF DISABILITY IN AUSTRALIA

Daniel Elazar and Lewis Conn
Analytical Services Branch

ABSTRACT

In Australia, as in many countries, there has been a rapidly growing demand from policy makers in both regional and national jurisdictions for social and economic small area data to satisfy expanding decision-making requirements. To date, the Australian Bureau of Statistics (ABS) has attempted to meet this demand using simple synthetic estimation, occasionally using more sophisticated small area models. The increasing user demand for small area estimates, together with practical difficulties in increasing survey sample sizes, has motivated the need to identify ways of finding reliable and defensible methods for producing quality small area estimates.

A project has commenced at the ABS to produce a series of manuals on the theory, application and processes for producing small area statistics in the Australian context. As part of this project, an empirical study has commenced into alternative approaches for producing small area estimates of disability. This builds on ABS experience in producing synthetic estimates from previous surveys of disability. The empirical study will assist in writing the small area estimation manual by developing practical knowledge and understanding of available small area methods with the ABS. In addition to this, we hope the results will go some way towards satisfying user demand for small area data on this topic.

The main purpose of this paper is to discuss methodological approaches undertaken in the application of existing small area methods to the topic of disability. This paper firstly briefly discusses the context of small area estimation in Australia. The paper then details the various small area models applied to the disability data. We use a simple demographic synthetic estimator in addition to a multivariate Poisson GLMM and a Bernoulli GLM. The latter two models are estimated using hierarchical Bayesian methods. We then present a set of diagnostic measures, found in the literature, which we used to assess the quality of the small area estimates produced by these models. Finally a comparison is drawn of the performance of these models against the diagnostic measures used. The results presented here are still preliminary and for this reason we also discuss the priorities for future work to improve on these models.

1. THE AUSTRALIAN CONTEXT

Over the past fifteen years there has been a rapidly growing demand for small area estimates in Australia. There are a variety of socio-political factors that have also stimulated this growth. These are as follows:

1. Growth in demand for small area statistics has largely coincided with an increased emphasis on evidence-based decision-making by government. Government agencies are now subject to greater accountability in providing a more efficient, effective and coordinated approach to the delivery of program services to regions with greatest need.
2. Local governments are taking a more pro-active role in the economic and social development of their jurisdictions.
3. There has been an increased focus of government policy making, both at national and state levels, on addressing the increased levels of economic and social disadvantage faced by communities residing in outer regional and remote areas of Australia. While these areas are often geographically very large, the populations they contain are usually very small. A significant proportion of Australia's indigenous people, who are the subject of a number of health, cultural, community development and housing programs, also reside in these remote areas.
4. Non-government organization service-providers increasingly require data for funding submissions and planning.
5. There is a rapidly growing statistical sophistication among economists in the use of more complex models, that combine both micro and macro economic dynamics, in forecasting economic trends and relationships. In the statistical realm, a flourish of new small area estimation methods, combined with unprecedented increases in computing capabilities, have meant that small area problems that were once intractable are becoming feasible.

ABS survey collections are designed to produce reliable estimates only at broad geographic levels such as at national and state levels. Practical issues have meant that there is little prospect of increasing sample allocations at the regional level in order to produce small area estimates of useful quality. The most feasible option, therefore, for satisfying the demand for small area data is the development of appropriate small area techniques to make use of existing survey data sources, along with suitable auxiliary data sources, including those available from other government or private agencies.

A number of small area estimation projects have been undertaken in the ABS over the last decade. These include a study of the estimation of labour force status for chosen small area regions (Bell and Carolan, 1998) using a time series model that takes account of autocorrelation between sample overlap groups. Several projects have also been undertaken into producing small area estimates of disability using census data.

The ideal approach to small area estimation, in the context of a national statistical service, would be to meet the following goals: meet decision-making objectives of users; cost-effectiveness; produce output of sufficient reliability for intended uses; be appropriate to the context of the problem; involve defensible methodologies; and provide results that can be readily interpreted and explained to users. In practice it may be difficult to meet all of these goals simultaneously, so that judgement needs to be exercised in determining the relative priorities of these objectives.

In response to this need, the ABS is developing a manual (or more likely a suite of manuals) on small area estimation methods and processes that will present a consistent approach to be used in the ABS, as a step towards assuring a standard quality of small area output. The manuals will be tailored to areas of the ABS involved in the production of small area estimates, whether they are involved in servicing client requests, implementing complex methodologies or validating, clearing and releasing output data.

The goals of the Small Area Estimation Practice Manuals Project, as it is known, are to increase the ABS' capability in satisfying the growing demand for small area statistics and to standardize and focus the ABS' approach to meeting this need. The manuals will help bridge the gap between the theoretical knowledge of small area estimation techniques and the practical application of such techniques. The manuals are intended to not only be a practical and methodological resource on how to go about producing small area data, but also a repository for capturing the growing experience of small area methods and processes. Within the context of a national statistical office, these goals are constrained by cost considerations, ease of implementation and interpretability of models and output for both statisticians and clients.

Brackstone (2002) gives a very good account of the issues facing national statistical offices in producing small area statistics. Pfeffermann (2002a) gives a concise summary of the key issues being confronted in small area estimation. Pfeffermann (2002b) and Trewin (1999) give an outline of the future directions for the application of small area estimation methods in the production of official statistics. They conclude that the use of more sophisticated model based methods is inevitable for improving the reliability of small area estimates.

2. EMPIRICAL STUDY OF DISABILITY

An empirical study into small area estimates of disability is currently being undertaken as part of the Small Area Estimation Practice Manuals Project. The response variables for this empirical study are obtained from the *Survey of Disability, Ageing and Carers* (SDAC) (Australian Bureau of Statistics, 1998). The main purpose of this empirical study is to develop within the ABS, knowledge and understanding of the implementation and effectiveness of small area techniques. The empirical study will assist in providing answers to the following questions:

1. What is the gain in quality in using sophisticated small area techniques over simpler ones (for example, Poisson generalized linear mixed models over linear or synthetic methods)?
2. What contribution does the quality of auxiliary data sources have towards the overall quality of small area estimates and what is the minimum level of quality for auxiliary data? By quality we mean not just the accuracy of the auxiliary data but also it's relevance to and correlation with the response variables.
3. What is the relative efficiency gain of using a unit (person) level model compared with a corresponding area level model?
4. At how fine a level can viable small area estimates be produced before suffering from model breakdown? By fine we mean either geographical size or the cross-classification of small area estimates by other variables such as severity of disability, age and sex.
5. What are the best approaches to validating output small area estimates in practice? How can user knowledge or preconceptions be best utilised in validating small area output? For example, disability administrators will be able to identify the districts in which the demand for services out-strips supply and whether the small area statistics reflect this. Disability administrators will also be useful in assessing the validity of small areas with extreme values.
6. What are the most efficient and appropriate measures of accuracy for modelled small area estimates? (Trewin, 1999) Are reliable and comprehensive measures available that are similar in effectiveness to those used for measuring the sampling error of direct sample estimates?

This paper will cover the models fitted to the data, the estimation method using hierarchical Bayes (HB), an assessment of the performance of the various models using various diagnostics, and finally gives a prioritised list of areas for further work. The reader interested in a detailed treatment of the survey and auxiliary data used, including associated quality issues, plus the quality of direct survey estimates, is referred to Elazar (2004).

3. SMALL AREA MODELS

In this section we discuss the small area models we intend applying to the disability empirical study. One of the main objectives behind this empirical study is to apply a range of appropriate models, ranging from simple to more sophisticated, to assess which performs best in terms of accuracy, ease of implementation, production costs and model interpretability. In a production context, the choice of model will depend intrinsically upon client quality requirements and the availability and quality of auxiliary data. The understanding gained from this empirical study will be a starting point for applications to other small area problems.

The models we will be using for this empirical study will all be extensions of the basic Fay–Herriot model. Models of the Fay–Herriot form are widely used in the small area estimation literature. One reason for using Fay–Herriot models is that they incorporate, as a special case, synthetic estimation, which has been used in previous small area estimation work on disability at the ABS. It is therefore convenient that synthetic model estimates can be easily obtained from the Fay–Herriot model by removing the random effects term (Rao, 2003). Another attractive feature is that the Best Linear Unbiased Predictor (BLUP) under the Fay–Herriot model can be shown to take the form of a composite estimator (Pfeffermann, 2002a). The composite estimator is a weighted average of the direct survey estimator and a synthetic estimate based on a generalised linear model fitted to the observed data at a broader area.

In all the small area models proposed, we consider the multivariate case wherever possible, subject to model identifiability. There are a number of response variables we wish to predict at the small area level, these mainly being impairment type (intellectual, physical, psychological/psychiatric, sensory, and head injury/brain damage) by level of severity (profound, severe, moderate, mild). It is important to take account of the variance covariance structure of these categories in order to improve model efficiency via multivariate models. A secondary reason is that the model fitting process is simpler than fitting separate models for each category.

There are three key dimensions to the small area models we wish to consider for the disability empirical study, which together form a framework. These dimensions are linear versus nonlinear, area versus unit level and mixed effects versus synthetic. We wish to consider a range of models across these dimensions to identify the best trade-off in competing models between simplicity, reliability, accuracy and interpretability. Previous work done in producing small area estimates of disability at the ABS, have used unit level models so one of the aims of the empirical study is to identify how much gain or loss of efficiency is afforded by unit as opposed to area level models. We present below a range of models that cover the more promising elements of this framework but not necessarily all.

3.1 Demographic synthetic method

This approach has in the past been commonly used for statistical consultancies carried out by the ABS. Direct estimates of proportion for impairment types were calculated at a broader area, in this case at the level of state, where direct estimates possess sufficient statistical reliability. These proportions are calculated at the age by sex level and then multiplied by demographic population counts (again at the age by sex level), which were obtained at the small area level, in our case the SSD. Initially we chose the statistical division (SD) as our broad area, which is finer than state and allowed for more geographical variation in impairment rates. Unfortunately, the use of SD as the broad area caused instability in the estimates and their estimates of variance. We therefore resorted to using state as the broad area.

An intrinsic assumption being made in this and all synthetic approaches is that all small areas within the broader region have the same relationship between the predictor variable and explanatory variables. In this particular case, impairment rates at the age by sex level are assumed to remain constant throughout the broader area. If this assumption fails to hold, the resulting estimates will suffer from bias, which in practice is difficult to measure.

Mathematically, letting i denote the i-th SSD and S denote the state in which the i-th SSD is located, the demographic synthetic estimate of the number of persons in the r-th impairment type in the i-th SSD is given by:

$$\hat{y}_i^{(r)} = \sum_g \frac{N_{i,g}}{N_{S,g}} \hat{y}_{S,g}^{(r)}$$

where

$N_{i,g}$ = population count of persons in the g-th age by sex cell in a particular SSD,

$N_{S,g}$ = $\sum_{i \in S} N_{i,g}$ = population count of persons in the g-th age by sex cell in state S

$\hat{y}_{S,g}^{(r)}$ = direct survey estimate of the number of persons in the g-th age by sex cell with the r-th impairment type in state S.

The efficiency of the demographic synthetic estimator $\hat{y}_i^{(r)}$ is reflected in its sampling error, which can be measured using a replicate variance technique such as the jackknife. The bias of this method is not so easy to quantitatively measure without the benefit of actual or simulated population data. However the diagnostic comparing the distribution of small area predictions against that of the direct estimates (see Section 4.1.6) gives a reasonable indication of the presence of a pronounced bias.

3.2 Poisson model

The demographic method is quite simple but apart from the assumption already mentioned of constant impairment rate within broad areas, suffers from several additional drawbacks that may compromise the accuracy of the resulting estimates. Firstly it cannot borrow strength from correlations between the different types of impairment. Secondly, it doesn't take account of additional auxiliary information that may help improve the reliability of predicted small area estimates. Thirdly, at the small area level the disability data is count data, for which a Poisson generalised linear model (GLM) would be better suited.

For these reasons, plus the fact that we are dealing with impairment type count data, we chose as a second model, the Poisson model with log transform. The Poisson model is commonly regarded as the "benchmark model for count data" (Cameron and Trivedi, 1998). The Poisson model we fitted was at the small area level (not person level) and incorporated both fixed and random effects terms.

The Poisson model is taken from the generalised exponential family of models discussed by Ghosh, et al. (1998). We firstly present the notation to be used. Let

$$\hat{\boldsymbol{\theta}}_{\mathrm{i}} = \left(\hat{\theta}_{i1}, \ldots, \hat{\theta}_{ir} \right)^{\mathrm{T}}$$

be an $r \times 1$ vector of SDAC direct estimates of r impairment type variables, for small area $i, (i=1,...,m)$. Assume that the elements $\hat{\theta}_{ir}$ of $\hat{\boldsymbol{\theta}}_{\mathrm{i}}$ are distributed $Poisson(\mu_{ir})$ with mean parameters μ_{ir} obeying

$$\boldsymbol{\gamma}_{\mathrm{i}} = \mathrm{X}_{\mathrm{i}}\boldsymbol{\beta} + \boldsymbol{\nu}_{\mathrm{i}}$$

where $$\boldsymbol{\gamma}_{\mathrm{i}} = \left[\ln\left(\mu_{i1} \right), \ldots, \ln\left(\mu_{ir} \right) \right]^{\mathrm{T}}$$

X_{i} is an $r \times rp$ matrix of p auxiliary variables for each of the r impairment type variables, and

$\boldsymbol{\beta}$ is an $rp \times 1$ vector of regression coefficients.

The $\boldsymbol{\nu}_{\mathrm{i}}$ are the small area level random effects, independently distributed with multivariate normal $\mathrm{N}_r(0, \boldsymbol{\Sigma}_{\boldsymbol{\nu}})$.

The above model was estimated using the WinBUGS software. For reasons of brevity a hierarchical Bayes formulation of the Poisson model is not presented here but is available on request.

Over-dispersion is a common problem when fitting a Poisson model and this is likely with the model we have fitted. At the time of writing, further work is being undertaken to incorporate more effective ways of taking account of the over-dispersion problem.

3.3 Bernoulli model

The Poisson model incorporates auxiliary information at the small area level. One of the aims of the disability empirical study was to test the performance of an area level model against that of a unit level model or at least in this case a combined unit/area level model. As a third model, a Bernoulli model with underlying logistic transform was chosen. Like the previous model, the intention was to fit this model with multivariate error structure and random effects. However at the time of writing, we have only been successful in fitting a multivariate model with univariate error structure and fixed effects only, using the WinBUGS software. Further work will shortly be undertaken to try and fit the model with a multivariate error structure and random effects.

An outline specification for the Bernoulli model is as follows. We wish to model the r-th impairment type y_{ij}^{r} of person j ($j = 1,\ldots, n_i$) within small area i ($i = 1,\ldots, m$) and then use this model to predict impairment types for non-sampled units, thereby producing estimates of disability $\hat{\boldsymbol{\theta}}_{\mathrm{i}}$ (as defined in Section 3.2). We assume the y_{ij}^{r} to be independent Bernoulli (p_{ij}^{r}) variables with conditional probability density function given by:

$$f\left(y_{ij}^{r} = 1 \mid p_{ij}^{r} \right) = p_{ij}^{r}$$

$$f\left(y_{ij}^{r} = 0 \mid p_{ij}^{r} \right) = 1 - p_{ij}^{r}$$

$$\boldsymbol{\lambda}_{\mathrm{ij}} = \left[\operatorname{logit}\left(p_{ij}^{1} \right),\ldots, \operatorname{logit}\left(p_{ij}^{r} \right) \right]^{\mathrm{T}} \qquad i=1,\ldots,m; j=1,\ldots,n_i$$

Then $\boldsymbol{\lambda}_{\mathrm{ij}}$ can be modelled thus:

$$\boldsymbol{\lambda}_{\mathrm{ij}} = \mathrm{X}_{\mathrm{ij}}^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{v}_{\mathrm{i}} \qquad i=1,\ldots,m; j=1,\ldots,n_i$$

where the area level random effects $\mathbf{v}_{\mathrm{i}}$ are distributed $\mathrm{N}_r(0,\boldsymbol{\Sigma}_{\mathbf{v}})$.

Although the matrix of auxiliary variables, $\mathbf{X}_{\mathrm{ij}}$ is subscripted to indicate both person and small area level variables, the only person level variables it will include will be age and sex, as determined from the survey. All other covariates will actually be at the small area level.

A multinomial logistic model was considered as a multivariate extension to the Bernoulli model, however time limitations have so far prevented us from pursuing this option.

3.3.1 Generating small area estimates from the Bernoulli model

Predictions of small area estimates need to be formed from the models that use unit level response variables. This can be done for each impairment type by summing un-weighted response values for variable r, from the sample s_i in small area i and then adding to that the sum of the predicted proportions $\tilde{p}_{ij}^r$ across the sample complement s_i^c. The $\tilde{p}_{ij}^r$ are predicted (Rao, 2003) by estimating $\boldsymbol{\beta}$ and generating a realisation of v_i^r from its underlying distribution. We then have:

$$\hat{y}_i^r = \sum_{j \in s_i} y_{ij}^r + \sum_{j \in s_i^c} \tilde{p}_{ij}^r$$

where $\hat{y}_i^r$ is the predicted count estimate for the r-th impairment type in the i-th small area, $\hat{y}_{ij}^r$ is the sample response for the r-th impairment type from the j-th person in the i-th small area.

An important issue is how do we ensure that the sum of the modelled count estimates across disability categories, $\sum_r \hat{y}_i^r$, agrees with the population benchmark for the i-th small area.

4. MODEL RESULTS

The three models referred to in Section 3 were fitted to the SDAC data using the auxiliary information described in Elazar (2004). Due to the drawbacks in the scope and nature of the main auxiliary datasets, there was only limited success in using the Disability Support Pension and the Commonwealth State/Territory Disability Agreement (CSTDA) datasets (for details, see Elazar, 2004). Auxiliary variable parameters were significant only for a restricted number of impairment type variables. Age and sex proved to be more reliable predictors across all impairment types.

The philosophical approach taken was to choose auxiliary variables for the models that should be closely related to and therefore be good predictors of impairment. It may also be desirable to throw in a wide range of demographic, social and economic variables in the hope of obtaining a good predictive model. Such data might be obtained for example from either population census data or other administrative data sources. There are numerous person level variables such as income level, marital status, labour force status and family structure that were collected in the SDAC questionnaire and are also available for the nearest population census.

The main purpose of the model is in this case to predict small area estimates of disability. While it is desirable to maintain parsimony, it may be wise to consider including a wider range of auxiliary variables. Sometimes statistical relationships can be found between the response variable and certain auxiliary variables that one would not have anticipated. However caution should be exercised in avoiding over-fitting the model while trying to minimise prediction errors. Biases resulting from the presence of model misspecification caused by the use of spurious covariates or poor model choice won't be accounted for by measures of model prediction error.

4.1 Diagnostics for predicted small area estimates

Several diagnostics have been used to assess the quality of small area model predictions. Most of these have been obtained from Brown, et al. (2001). These diagnostics have been invaluable in assessing the relative performance of competing small area model predictions. Subjective judgment has to be exercised in using some of these diagnostics. Some models may perform better on some diagnostics but not others, but this may be reversed for other models. Clearly there is a need to assess the relative importance of the diagnostics.

4.1.1 Measure of bias

The first diagnostic measures a tendency for bias to be present in the small area model predictions. Although small area direct survey estimates may have high sampling error, they are largely unbiased. If the model predicted small area estimates are also unbiased, then a plot of these estimates against the direct survey estimates would show the scatter plot evenly centred around the unit line (y=x). There is an important caveat to be aware of when using this diagnostic test. Where the direct survey estimates are derived from a non-linear estimator (e.g. ratio estimator), the estimator will to some extent be biased and this bias will be most significant when the small area sample size is small. This will usually affect the lower left hand side of the scatter plot where population estimates are smallest. Therefore some degree of caution needs to be exercised when interpreting this kind of diagnostic.

The diagnostic measure is carried out in the following manner. After plotting model predicted estimates against the direct survey estimates, a regression line is fitted to the scatterplot. If the confidence interval calculated for the estimated regression slope does not contain the value of 1, then the hypothesis that the model predictions are unbiased is rejected. How far the value 1 is outside the confidence interval gives at least a qualitative indication of the extent of the bias.

Figure 4.1 below shows an example of this diagnostic for the intellectual impairment type under the Bernoulli model. The slope of the regression line (the lower dashed line) is not statistically different from the unity, despite the large variation in the residuals.

4.1 Bias diagnostic: Bernoulli model – Intellectual impairment type

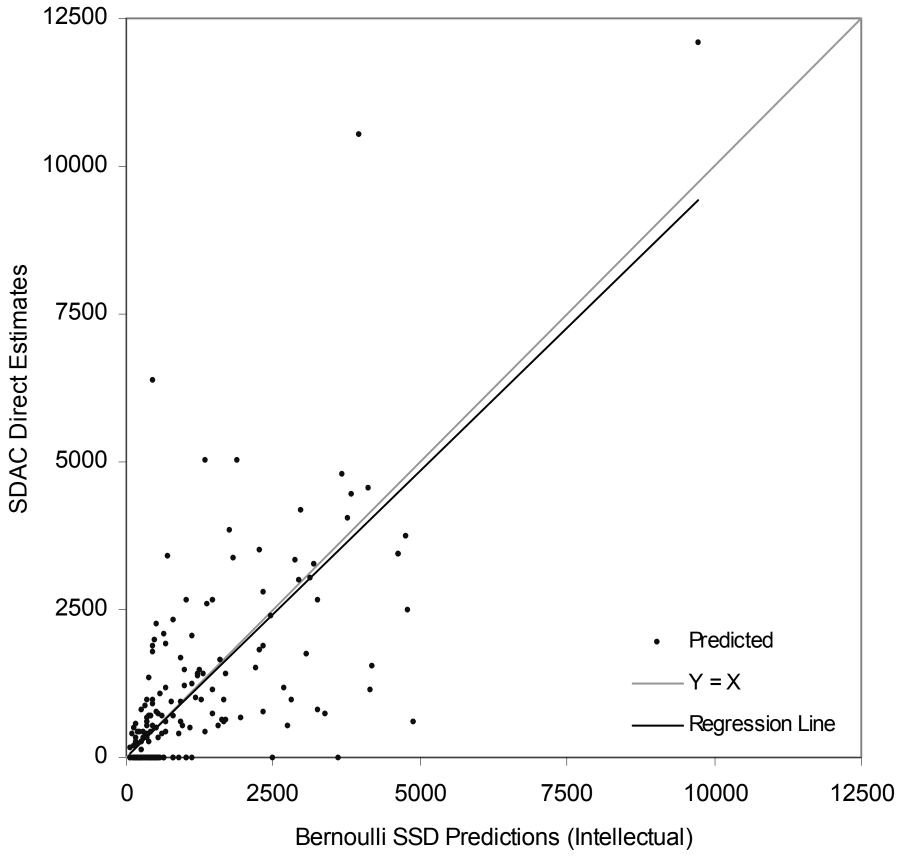
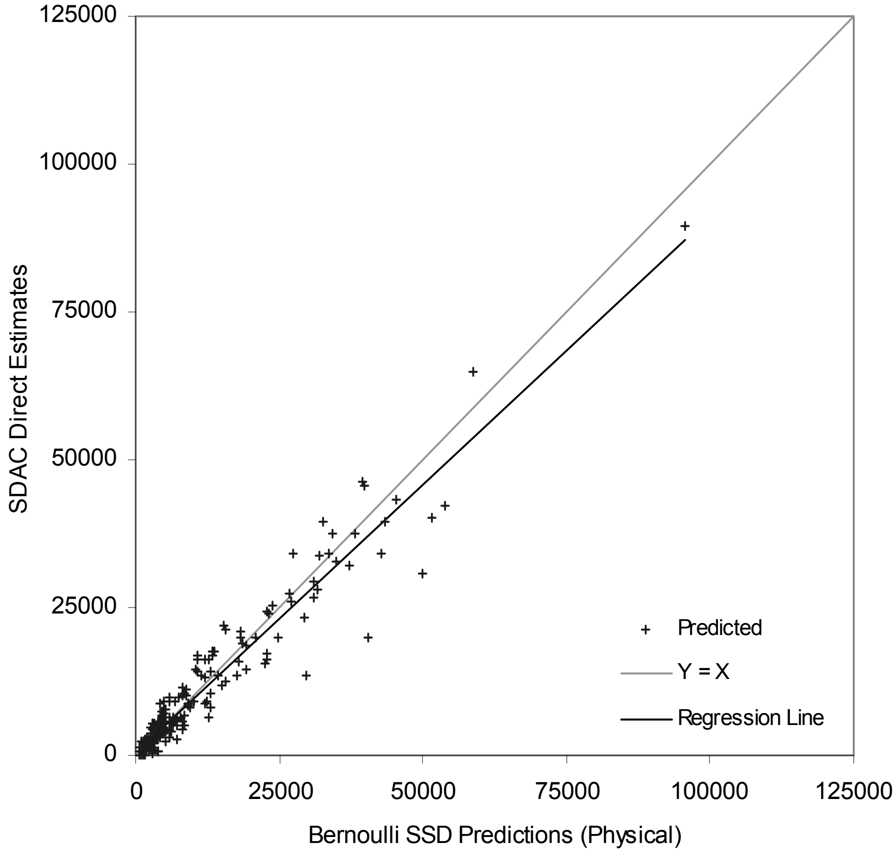


Figure 4.2 below shows the same diagnostic for the physical impairment type. Here the regression slope (again the lower dashed line) is statistically different to 1. The reader may recall that the direct ratio estimator used for the vertical axis and as the response variable in the models, can suffer from bias in small areas where the sample size is quite small. However as the regression line intercepts are not statistically different from zero, it appears that any bias arising from the use of the post-stratified ratio estimator for areas with a small population is not significant.

4.2 Bias diagnostic: Bernoulli model – Physical impairment type



A summary of how the various models perform for each impairment type is shown at table 4.3. For each impairment type and small area model, the fitted slope is shown with the 95% minimum and maximum confidence interval bounds. Confidence intervals for regression slopes shown in bold do contain the unit slope.

4.3 Bias diagnostic: Significance of regression slopes

	Synthetic			*Poisson*			*Bernoulli*		
	Slope	*CI min*	*CI max*	*Slope*	*CI min*	*CI max*	*Slope*	*CI min*	*CI max*
Physical	0.95	0.91	0.98	0.78	0.73	0.82	0.90	0.87	0.94
Intellectual	1.02	0.92	1.13	2.96	2.46	3.46	0.97	0.83	1.11
Sensory	0.96	0.92	1.01	0.80	0.74	0.85	0.92	0.86	0.97
Psychological	0.92	0.82	1.02	1.88	1.61	2.15	0.94	0.81	1.06
Head injury	0.98	0.91	1.06	1.29	1.16	1.41	0.98	0.89	1.06
None	1.01	1.00	1.02	1.03	1.02	1.04	1.02	1.01	1.03

Table 4.3 shows that, for almost all impairment types, we would reject the hypothesis that the synthetic demographic modelled estimates are biased, at the 5% significance level. The synthetic demographic modelled estimate for the physical impairment type is only borderline significant in bias. The Poisson model indicates significant bias for all impairment types which suggests that the model, at least in its current form, is not performing well. There is no significant bias for intellectual, psychological and other impairments under the Bernoulli model, although the remaining impairment types are not as biased as under the Poisson model.

4.1.2 Coverage statistic for confidence intervals

RSE's can readily be calculated for direct survey estimates. Likewise a measure of prediction error, such as the root relative MSE can be calculated for predicted small area estimates. Based on these respective measures of error, confidence intervals for each small area can be constructed for the direct estimate and the predicted modelled estimate. We expect each of these confidence intervals to include the "true" value 95% of the time. These confidence intervals are adjusted so that they have a 95% chance of overlapping. The number of times (across small areas) they do actually overlap is then counted and compared to the Binomial distribution to give a non-parametric significance test of the bias of model estimates relative to their precision. (Brown, et al., 2001)

Brown, et al. (2001) show that if $z(\alpha)$ is the probability that a standard normal variable takes values greater than $z(\alpha)$, then the two intervals $X \pm z(\beta)\sigma_x$ and $Y \pm z(\beta)\sigma_x$ do not overlap with probability α if $z(\beta)$ is chosen as:

$$z(\beta) = z(\alpha)\left(1 + \frac{\sigma_X}{\sigma_Y}\right)^{-1}\sqrt{1 + \frac{\sigma_X^2}{\sigma_Y^2}}$$

Table 4.4 below shows the results of this diagnostic for the three models used in this empirical study and each impairment type.

4.4 Coverage diagnostic: Percentage non-overlap between confidence intervals of direct and model predicted estimates

	Synthetic	*Poisson*	*Bernoulli*
Physical	1.1%	18.0%	9.8%
Intellectual	1.6%	19.1%	6.0%
Sensory	1.1%	9.8%	3.8%
Psychological	0.0%	6.0%	0.5%
Head injury	0.0%	6.6%	1.6%
None	2.2%	14.2%	15.8%

The simple synthetic method performs quite well for most impairment types on this diagnostic with 2% or less non-overlap in the confidence intervals. Surprisingly, for the Poisson and Bernoulli models, the level of non-overlap is statistically high for the no impairment category, and at present we do not have any firm explanation for why this might be the case. A possible explanation is that, being around 80% of the population, the RSEs for this impairment type are quite low but the predicted model estimates don't perform as relatively well. Hence the level of non-overlap is higher than for any other impairment type.

The Poisson model has an acceptable level of non-overlap for the psychological and head injury impairment types, but not for the other impairment types. Confidence intervals for the Bernoulli model predictors overlap sufficiently well with those of the direct estimates for most variables except for the physical and no impairment types. Brown, et al. (2001) suggest that a lack of sufficient overlap is indicative of random effects not being taken account of in the model. This is certainly true of the current Bernoulli model. The Poisson model, although it does include random effects terms, may be suffering from untreated over-dispersion, which we currently postulate is the cause of its poor performance under most of the diagnostics.

4.1.3 Measures of accuracy for model predictions – Posterior root relative variances

It is important to assess the quality of a set of small area predictors on the basis of their performance across all the diagnostics. However most small area practitioners, particularly those within a national statistical service, desire a single summary measure of the precision and accuracy of small area estimates. Such a measure is easier to explain to clients and better conveys the reliability and fitness for purpose of the small area estimates.

In this context, measures of model prediction error are important for this purpose. Many small area practitioners, operating within a frequentist paradigm, calculate mean square errors (MSEs) for predicted small area estimates. For a good introduction to the calculation of MSEs in the case of linear mixed models, see Section 4 of Saei and Chambers (2003).

A number of commentators on small area techniques, whether they be frequentists or Bayesians, state that measures of prediction error are more important than goodness of fit measures on the small area model used to make those predictions. In other words it is the prediction that one is more interested in, and a slight degree of model over-fitting is not so much of a concern if the resulting predictions are of a superior precision. However, supporters of this position would agree, that an important assumption is that there is no serious model misspecification present. In particular if the small area model is considerably over-fitted, the possibility of spurious covariates is more likely and the model is also unlikely to be applicable to other datasets.

Given the complex matrix manipulations that need to be undertaken in calculating MSEs and the fact that we used hierarchical Bayes (HB) to estimate model parameters, we chose to calculate measures of predictive accuracy directly from the conditional posterior distribution produced as output from WinBUGS. We used the available WinBUGS diagnostic tests to ensure that the simulation chains had fully converged and removed the burn-in (pre-convergence) part of the chain before calculating posterior variances. Dual Markov Chain Monte Carlo (MCMC) chains, rather than a single chain, were used to facilitate the diagnosis of convergence.

Rao (2003) gives a formula for deriving the general posterior variance conditional on the data, using a single MCMC chain. Using similar notation as in Sections 3.2 and 3.3, the corresponding single chain formula for the Bernoulli model is as follows,

$$V(\theta_{ir} \mid \mathrm{y_s}) =$$
$$D^{-1}\sum_{k=d}^{d+D}\sum_{j\in s_i^c}\left[p_{jr}^{(k)}\left(p_{jr}^{(k)}-1\right)\right]+D^{-1}\sum_{k=d}^{d+D}\left(\sum_{j\in s_i^c}p_{jr}^{(k)}\right)^2-D^{-2}\left(\sum_{k=d}^{d+D}\sum_{j\in s_i^c}p_{jr}^{(k)}\right)^2$$

where

θ_{ir} = total predicted estimate of impairment type r for small area i,

$\mathrm{y_s}$ = vector of sample data,

d = length of burn-in period for MCMC run (chain),

D = length of full MCMC chain, and

$p_{jr}^{(k)}$ = the k-th MCMC chain sample value for Bernoulli rate parameter (p) for r-th response variable for person j.

The corresponding formula for the Poisson model is

$$V(\theta_{ir} \mid \mathrm{y_s}) = D^{-1}\sum_{k=d}^{d+D}\lambda_{ir}^{(k)} + D^{-1}\sum_{k=d}^{d+D}\left(\lambda_{ir}^{(k)}\right)^2 - D^{-2}\left(\sum_{k=d}^{d+D}\lambda_{ir}^{(k)}\right)^2$$

where λ_{ir} is the corresponding Poisson parameter.

The above formulae apply to a MCMC simulation using a single chain, however these formulae can be readily adapted for the case of multiple chains.

Other measures to assess the fit of the models have been used in the literature. Nandram, et al. (1999) used three different measures to assess the fit of four Poisson generalised linear models. These were the posterior expected predicted deviance (EPD), the posterior predicted p-value, and standardised residuals. EPD is used to rank proposed models, while the p-value is used as a goodness of fit measure. The standardised residuals are used to evaluate the alternative models. For more details and references see Dey, et al. (2000, pp. 101–2).

Figure 4.5 shows non-parametric splines fitted to the root relative posterior variances (RRPV) for the Poisson and Bernoulli models for the physical impairment as well as RSEs for the direct and demographic synthetic estimators.

4.5 Root relative posterior variance diagnostic: Physical impairment

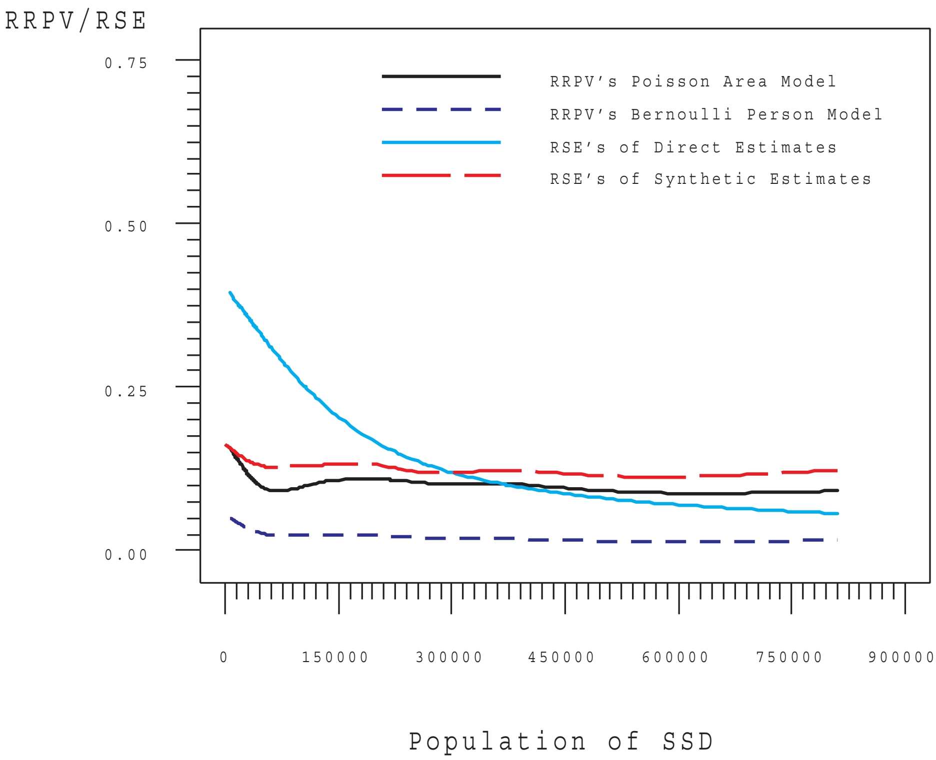
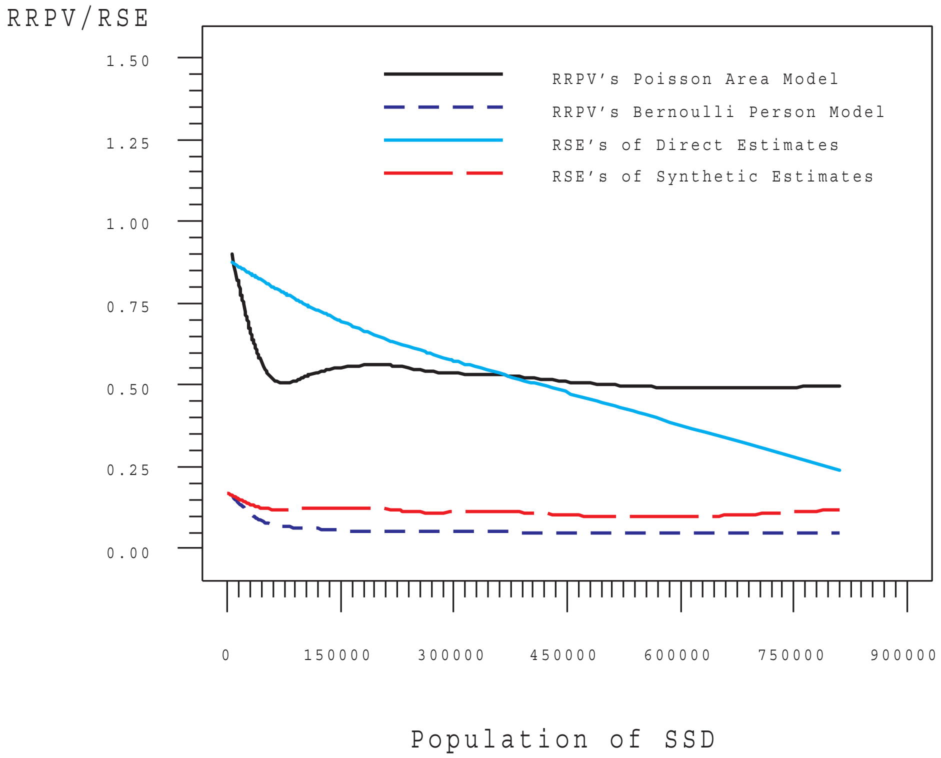


Figure 4.6 shows the same diagnostic for the intellectual impairment. Both graphs show that the Bernoulli RRPV's are much lower than the corresponding measure of accuracy of all other small area estimates. The astute reader may regard this with a fair degree of suspicion, and as we shall see this appears to be indicative of the substantial degree of over-shrinkage exhibited by the Bernoulli small area predictions. This

serves as a good example of how relying solely upon RRPV's (or root MSEs) as a measure of prediction accuracy can in some situations be misleading.

4.6 Root relative posterior variance diagnostic: Intellectual impairment

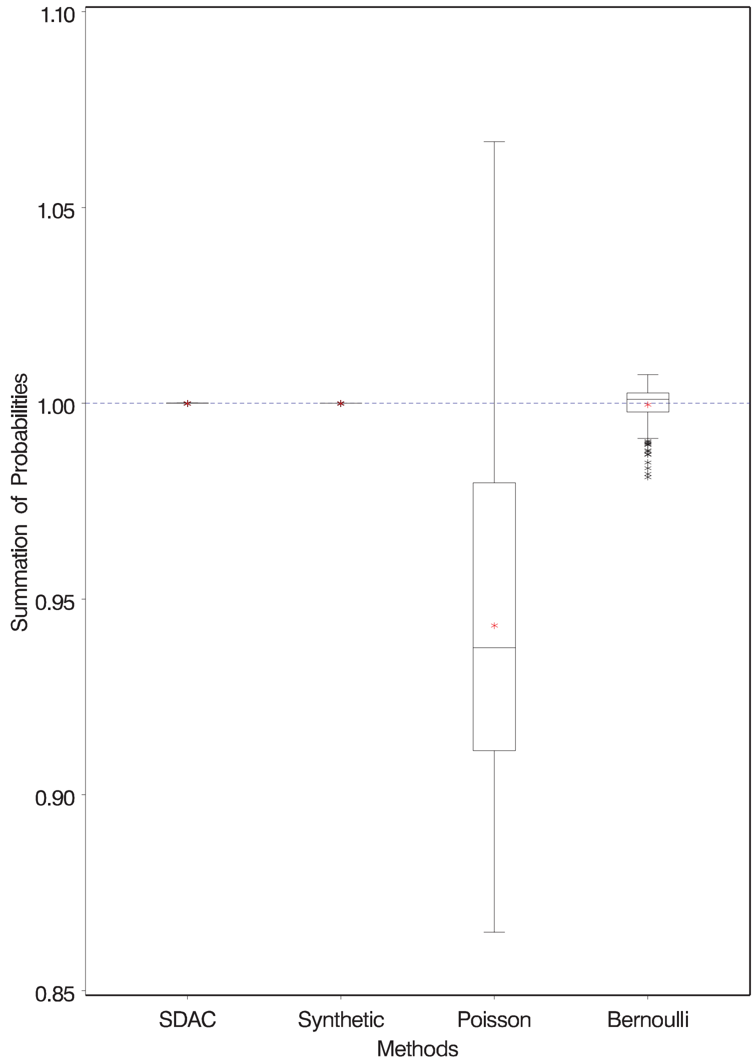


4.1.4 Additivity of proportions to one

In the case of disability, we are dealing with six impairment types, these being physical, sensory, intellectual, psychiatric, head injury / brain damage and no impairment. It is clearly desirable for predicted impairment rates to sum to one in each small area. Models that preserve this additivity are clearly more appropriate than those that don't.

Figure 4.7 below shows how well the direct SDAC estimates at small area level and the predictions from each of the three small area models preserve additivity to one.

4.7 Additivity of proportions to one



The SDAC and synthetic estimates ensure additivity of proportions to one exactly, whilst the Bernoulli model predictions preserves additivity very close to one in all small areas. For those small areas where proportions don't sum to one, they are out by at most about 2% in the case of the Bernoulli model. Even though it doesn't have a multivariate error structure, the Bernoulli model still preserves additivity to one quite well because of its general multivariate form. The Poisson model is quite disappointing even though it is of a multivariate form. At present we don't have a precise explanation as to why this might be but we are keen to find out whether this problem is resolved when Poisson over-dispersion is taken account of.

4.1.5 Additivity of model predictions to broad area direct estimates

The fifth diagnostic we've utilised is to compare the sum of small area model predictions within a broad region to the direct estimate for that same broad region. The broad regions should ideally be chosen as the minimum size region that affords direct survey estimates with sufficiently low sampling error. We have chosen state /

territory as our broad region as it is an important administrative region and usually possesses reliable survey estimates. However the direct estimates of the smallest state/territories such as the Northern Territory (NT), the Australian Capital Territory (ACT) and Tasmania, have higher levels of sampling error compared with the other states. We have not aggregated these smaller states and territories, as they are still important administrative units.

A model that produces small area predictions that, when aggregated, agree closely with broad region direct estimates is more preferable to alternative models that give a lesser degree of agreement. Small area predictions can of course be calibrated to ensure they do sum to the direct estimate at the state/territory level. This can be done either by using a technique such as iterative proportional fitting on the predictions output from the model, or by using a constrained regression approach to incorporate additivity as a constraint in the model estimation. Brown, et al. (2001) recommend first deriving small area predictions without applying either kind of calibration so that the results of the intrinsic model against this diagnostic test can be assessed. Either calibration technique can then be applied, if necessary, to the model that performs the best against all the diagnostic tests.

Figure 4.8 below shows the performance of the Bernoulli model in terms of additivity to state for each impairment type. The vertical axis shows the percentage relative difference between the state direct estimate and the sum of the predicted small area estimates. As is to be expected, the graph shows that the relative difference is smaller for the impairments types with larger populations. The worst case is for the intellectual impairment type, which varies from –15% to +25% in the larger states. In the smaller states the relative difference is more often negative than positive, because direct estimates are commonly zero, due to small sample sizes, but the predicted small area estimates are rarely is ever zero.

4.8 Additivity of state – Bernoulli model, All impairment types

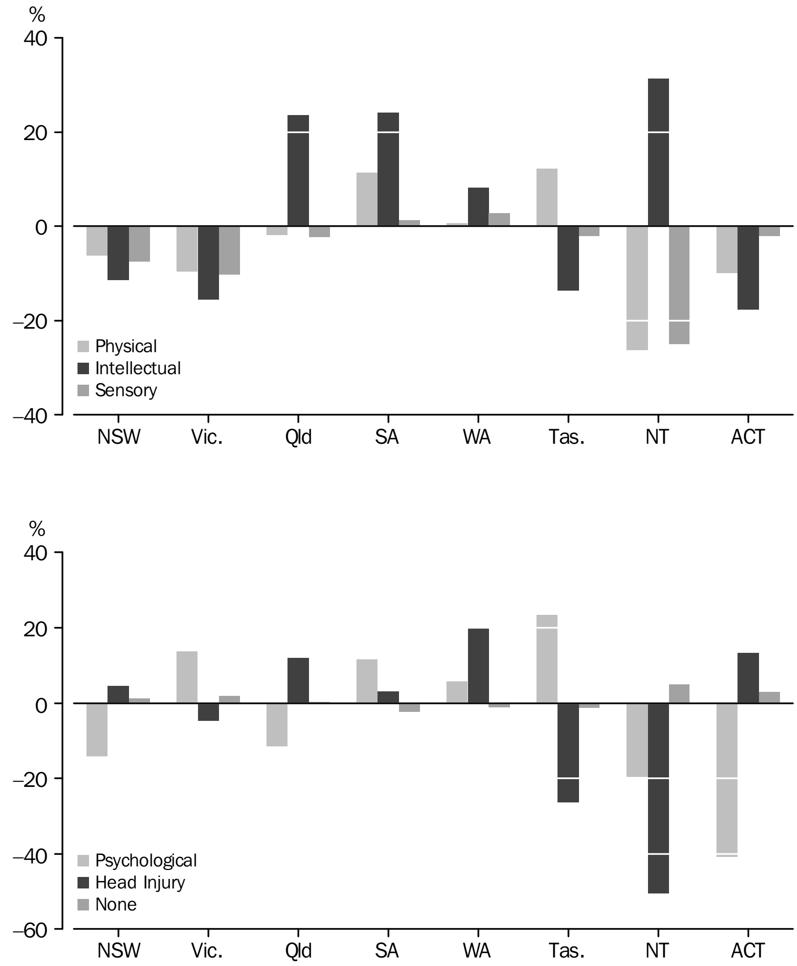


Figure 4.9 shows the results of additivity to state in the three models used for the intellectual impairment. The demographic synthetic approach performs best, followed by the Bernoulli model. The Poisson model performs very poorly and has been excluded from this diagnostic graph. We currently have no explanation for this phenomenon.

4.9 Additivity of state – Intellectual impairment type

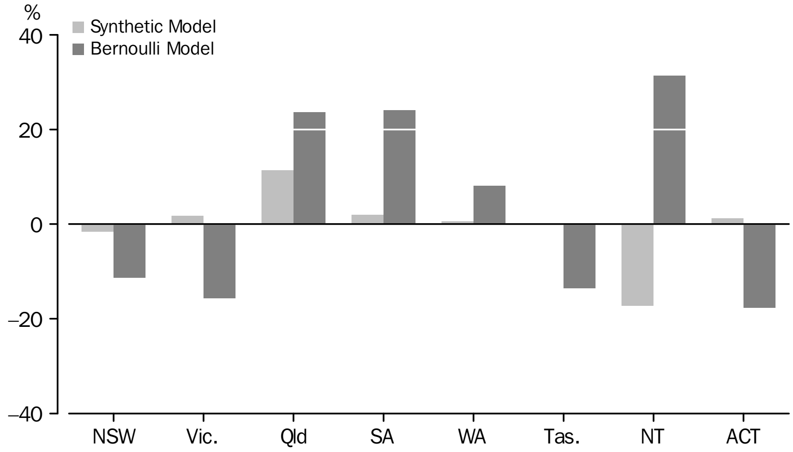
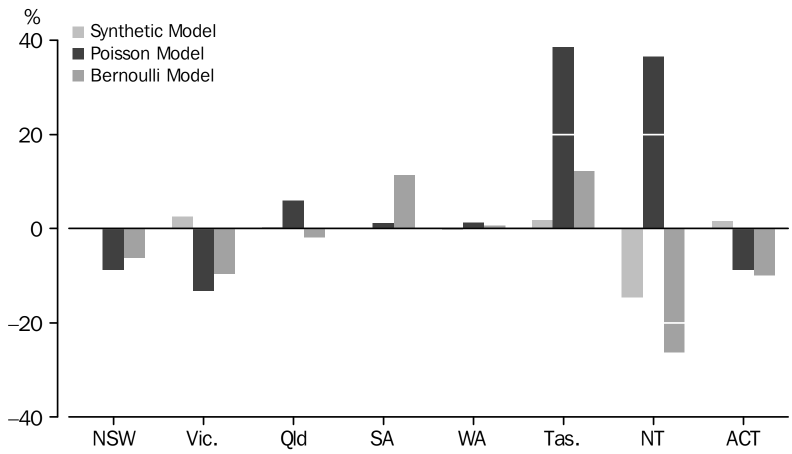


Figure 4.10 shows the results of additivity to state in three models used to predict small area estimates of physical impairment. As for intellectual impairment, the demographic synthetic approach performs best, followed by the Bernoulli model, which ranges from –10% to +10% in the larger states. For physical impairment, the Poisson model performs only marginally worse than the Bernoulli model.

4.10 Additivity of state – Physical impairment type



4.1.6 Distribution of predicted small area estimates

The final diagnostic we have used involves plotting box plots of survey and model small area estimates, for each impairment type, in order to compare distributions across models.

Figure 4.11 shows these box plots alongside a box plot for the direct survey estimates. The dashed, horizontal line passes through the mean of the direct survey small area estimates. Figure 4.11 shows that the SDAC direct survey estimates, demographic synthetic and Bernoulli models all have very similar means. The demographic synthetic exhibits less variation in small area estimates than the direct survey estimates, while the Bernoulli model predictions show the least variation by far. The distribution of Poisson model predictions has slightly more variation than the Bernoulli model but what is more worrying, it is the most biased when compared with the other three distributions. This is most likely due to over-dispersion in the Poisson model, which has yet to be taken care of. To adjust for the over-dispersion we propose to use the negative binomial regression model approach of Lawless (1987), as described in Chapter 6 of Dey, et al. (2000, page 94).

4.11 Distribution of small area estimates by model – Intellectual impairment type

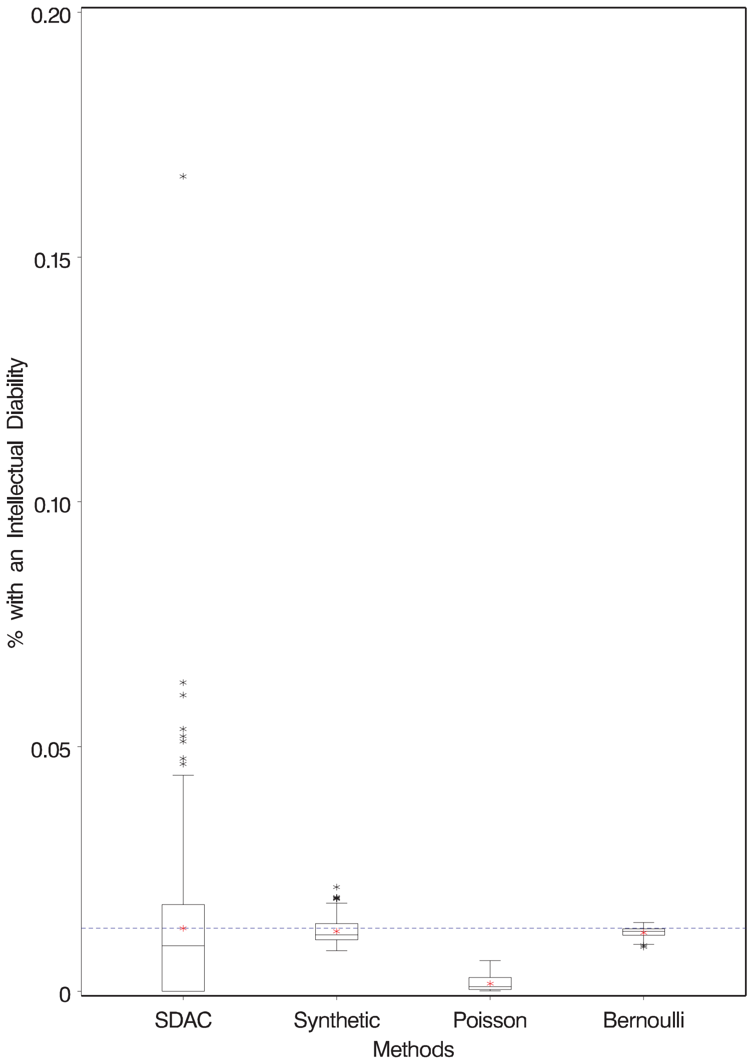
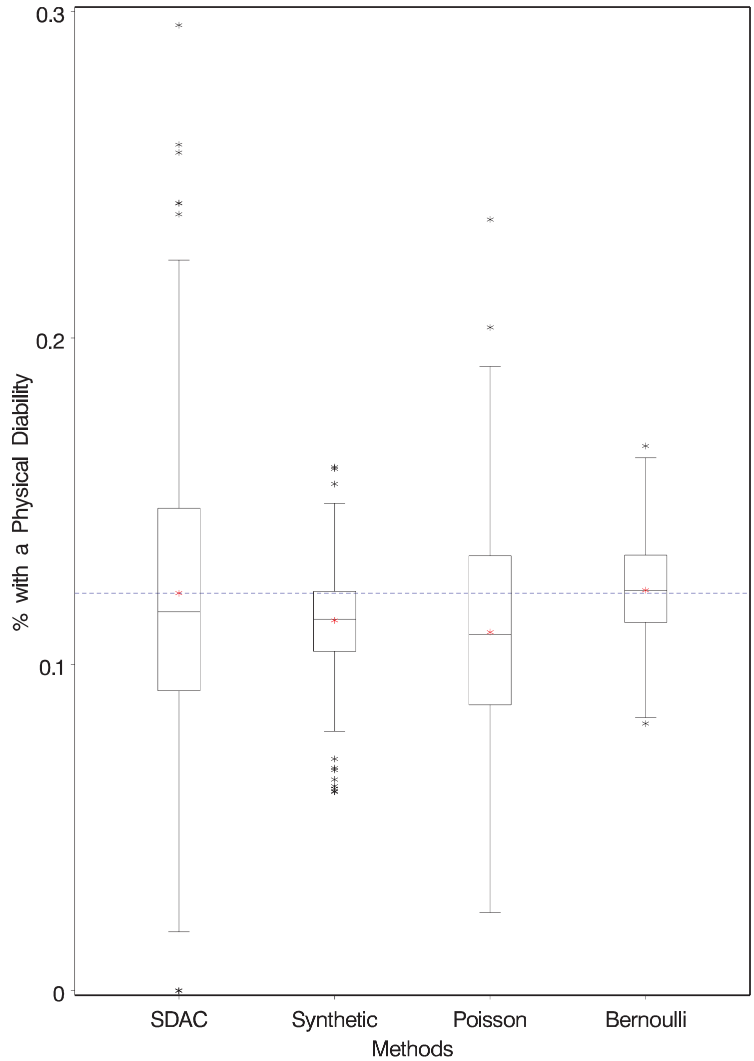


Figure 4.12 below shows the corresponding comparison of distributions for the physical impairment. The story for this impairment type is very similar to that of the intellectual impairment. The SDAC direct estimates and predicted small area estimates from the demographic synthetic and Bernoulli models all share a very similar mean, with the Poisson estimates appearing to be biased. The demographic synthetic estimates are slightly less variable than the direct SDAC estimates while the Bernoulli estimates have the least variation.

4.12 Distribution of small area estimates by model – Physical impairment type



An overly narrow distribution of model predicted small area estimates is referred to in the literature as over-shrinkage and is not a desirable feature as the resulting distribution of small area estimates under represents the true population distribution. It is highly likely that the distribution of Bernoulli model small area predictions is exhibiting over-shrinkage. Possible reasons for this are firstly that the Bernoulli model, in its present form, does not include random effects terms, and therefore does not allow for variation between small areas, not explained by the auxiliary variables. Secondly we are currently using only a limited number of auxiliary variables, which may not be accounting for all the differences between small areas. Work is currently underway to fit the Bernoulli model with random effects terms and multivariate error structure.

5. CONCLUSIONS AND FURTHER WORK

The disability empirical study has explored three small area estimation models (demographic synthetic, Poisson (multivariate error structure with random effects) and Bernoulli (multivariate with univariate error structure, no random effects) models) to assess their relative performance against six diagnostic measures. These diagnostics are: bias, overlap of confidence intervals, posterior root relative variances (PRRV), additivity of proportions to one, additivity to state and the distribution of modelled small area estimates.

The Poisson model has performed poorly on just about all the diagnostics. We believe this has been due to one or both of two possible reasons. Firstly that most of the variation in disability characteristics occurs at the person level. Hence a model such as the Poisson, which uses auxiliary data only at the area level, cannot improve much upon the precision of direct small area estimates. Secondly, there is a significant problem of over-dispersion with the current Poisson model, which is impacting adversely upon the reliability of the small area predictions. We propose to undertake further work to implement a suitable method for accounting for this problem of over-dispersion.

Although the Bernoulli model predictions have the lowest measures of prediction error (PRRV), this is more likely to be due to over-shrinkage in the estimator rather than an indication of high quality of the estimates. Extreme caution therefore needs to be exercised when interpreting measures of model prediction error and these measures need to be carefully assessed in the light of results from other diagnostic measures such as those discussed in Section 4. Work is currently underway to further improve this model by including a random effects term and a multivariate error structure. We also intend carrying out further work firstly into comparing our measures of PRRV against the frequentist estimates of mean square error and secondly into Bayesian methods for model checking such as those referred to in Section 4.1.3 and in Dey, et al. (2000, page 402).

There are a number of other statistical issues that could be addressed in future work. The first of these is the inclusion of additional auxiliary variables, preferably at the person level, to help further improve the fit of the models. Census data, which includes variables such as income levels, marital status, family structure, dwelling type, is the most readily available source of auxiliary data even though it is not strictly contemporaneous. Another area for further work is the extension of these models to take account of spatial autocorrelations of disability using approaches taken in the field of disease mapping (Wakefield and Elliot, 1999; Pascutto, et al., 2000 and Elliot, et al., 2000).

REFERENCES

Australian Bureau of Statistics (1998) *Disability and Long Term Health Conditions, Australia*, cat. no. 4433.0, ABS, Canberra.

Bell, P.A. and Carolan, A.M. (1998) *Trend Estimation for Small Areas from a Continuing Survey with Controlled Sample Overlap*, Working Paper No. 98/1, cat. no. 1351.0, ABS, Canberra.

Brackstone, G. J. (2002) "Strategies and Approaches for Small Area Statistics", *Survey Methodology*, 28(2), pp. 117–123.

Brodie–Reed, I., Black, K., Chubb, P. and Ng, C. (2002) *Socioeconomic Factors and Health (Data from the 2002 General Social Survey)*, Internal ABS Report.

Brown, G., Chambers, R., Heady, P. and Heasman, D. (2001) "Evaluation of Small Area Estimation Methods – An Application to Unemployment Estimates from the UK LFS", *Proceedings of Statistics Canada Symposium 2001, Achieving Data Quality in a Statistical Agency: A Methodological Perspective*.

Cameron, A.C. and Trivedi, P.K. (1998) *Regression Analysis of Count Data*, Cambridge: Cambridge University Press.

Dey, D.K., Ghosh, S.K. and Mallick, B.K. (eds.) (2000) *Generalised Linear Models: A Bayesian Perspective*, New York: Marcel Dekker.

Elazar, D.N. (2004) "Small Area Estimation of Disability in Australia", *Statistics in Transition*, 6(5), pp. 667–684.

Elliot, P., Wakefield, J.C., Best, N.G. and Briggs, D.J. (2000) *Spatial Epidemiology, Methods and Applications*, Oxford University Press.

Ghosh, M., Natarajan, K., Stroud, T.W.F. and Carlin, B.P. (1998) "Generalised Linear Models for Small-Area Estimation", *Journal of the American Statistical Association*, 93, pp. 273–282.

Lawless, J.F. (1987) "Negative Binomial and Mixed Poisson Regression", *Canadian Journal of Statistics*, 15, pp. 209–225.

Mellor, J.M. and Milyo, J. (2002) "Income Inequality and Health Status in the United States: Evidence from the Current Population Survey", *Journal of Human Resources*, 37(3), pp. 510–539.

Pascutto, C., Wakefield, J.C., Best, N.G., Richardson, S., Bernardinelli, L., Staines, A. and Elliott, P. (2000) "Statistical Issues in the Analysis of Disease Mapping Data", *Statistics in Medicine*, 19, pp. 2493–2519.

Pfeffermann, D. and Bleuer, S.R. (1993) "Robust Joint Modelling of Labour Force Series of Small Areas", *Survey Methodology*, 19, pp. 149–163.

Pfeffermann, D. and Sverchkov, M. (1999) "Parametric and Semi-Parametric Estimation of Regression Models Fitted to Survey Data", *Sankhya*, 61, pp. 166–186.

Pfeffermann, D., Moura, F. and Silva, P.N. (2001) *Multi-Level Modelling Under Informative Probability Sampling*, 53rd session of the International Statistical Institute, Seoul, Korea, 2001.

Pfeffermann, D. (2002a) "Small Area Estimation – New Developments and Directions", *International Statistical Review*, 70, pp. 125–143.

Pfeffermann, D. (2002b) "The Riga Conference: Summing Up Remarks", *Statistics in Transition*, 4(5), pp. 799–801.

Pfeffermann, D. and Sverchkov, M. (2003), "Fitting Generalised Linear Models under Informative Sampling", Chapter 12 of Chambers, R.L. and Skinner, C.J. (eds.) *Analysis of Survey Data*, Wiley.

Rao, J.N.K. (2003) *Small Area Estimation*, Hoboken, NJ: Wiley.

Saei, A. and Chambers, R. (2003) *Small Area Estimation: A Review of Methods Based on the Application of Mixed Models*, Methodology Working Paper M03/16, Southampton Statistical Sciences Research Institute.

Trewin, D. (1999) "Small Area Statistics Conference", *Survey Statistician*, 41, pp. 8–9.

Wakefield, J. and Elliott, P. (1999) "Issues in the Statistical Analysis of Small Area Health Data", *Statistics in Medicine*, 18, pp. 2377–2399.

World Health Organisation (2001) *International Classification of Functioning, Disability and Health*, Geneva. <URL: http://www3.who.int/icf/icftemplate.cfm> last viewed 20 April 2004.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the cooperation and kind assistance from the Australian Institute of Health and Welfare in providing the ABS with tabulated data compiled from the Commonwealth State/Territory Disability Agreement – Minimum Datasets. We are also grateful to Centrelink and the Department of Family and Community Services for providing the Disability Support Pension data used in this study. We would also like to thank Marion McEwin, Philip Bell and Ken Black for their comments on earlier drafts of the paper. Dr David Steel and Dr Borek Puza also gave invaluable suggestions on statistical approaches to undertaking the empirical study.

FOR MORE INFORMATION...

INTERNET **www.abs.gov.au** the ABS web site is the best place to start for access to summary data from our latest publications, information about the ABS, advice about upcoming releases, our catalogue, and Australia Now—a statistical profile.

LIBRARY A range of ABS publications is 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 web site for a list of libraries.

CPI INFOLINE For current and historical Consumer Price Index data, call 1902 981 074 (call cost 77c per minute).

DIAL-A-STATISTIC This service now provides only current Consumer Price Index statistics call 1900 986 400 (call cost 77c per minute).

INFORMATION SERVICE

Data which have been published and can be provided within five minutes are free of charge. Our information consultants can also help you to access the full range of ABS information—ABS user-pays services can be tailored to your needs, time frame and budget. Publications may be purchased. 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 2001

FREE ACCESS TO PUBLICATIONS

All ABS publications can be downloaded free of charge from the ABS web site.

WEB ADDRESS www.abs.gov.au

2000001511824
ISBN 0 642 48090 7

Produced by the Australian Bureau of Statistics