Small domain population estimation based on an administrative list subject to under and over-coverage

Patrick Graham*
Statistics New Zealand, Christchurch, New Zealand - patrick.graham.br@gmail.com

Anna Lin
Statistics New Zealand, Christchurch, New Zealand -anna.lin@stats.govt.nz

Abstract

We outline a Bayesian approach to population estimation based on an administrative list that is subject to both under and over-coverage with respect to a target population. We assume a sample survey of the target population has been conducted and linked to the list without error. Given the sample-list union we jointly model under and over-coverage at a small domain level. We use under and over-coverage estimates to obtain the posterior predictive distribution of a coverage-corrected list. Inference for small domain population counts follows straightforwardly. Application to an initial simulated example produced encouraging results.

Keywords: Population estimation; Administrative data; Bayesian Inference; Missing data.

1. Introduction Statistical agencies in several countries are investigating methods for replacing traditional census based population estimation systems with approaches based on administrative data (see, for example, Bycroft, (2015)). Administrative lists may fail to include some people who are in fact in the target population and also include people who are no longer in the target population, due, for example, undetected out-migration. Relative to a traditional census, the latter problem (over-coverage) may be a more significant issue for population estimation based on administrative data.

In this paper we outline a Bayesian approach to population estimation from an administrative list that adjusts the list for both over and under-coverage. By population estimation we mean, not just the total size of the population, but also the distribution of population across categories of key demographic variables such as age, sex, ethnic group and area. We assume that it is possible to conduct a highly quality survey of the target population and that this sample can be linked to the list without error. We assume no other fieldwork. In particular, the methodology outlined does not require any sampling from the list. Our approach can therefore be viewed as a Bayesian version of Zhang (2015) with the important differences that we assume access to only a single administrative list and our focus is on small domain estimation rather than estimating the population total.

Although administrative data is prone to measurement error and the methodology can be extended to adjust for such errors, in order to concentrate on the main issue of adjustment for under and over-coverage we assume no measurement error or misclassification on the list. Similarly, although it is possible to extend the methodology to accommodate errors in the linkage of the population sample to the list we do not deal with that issue here.

2.Basic set-up. To establish basic concepts, suppose a target population (e.g usually resident population of New Zealand) could be cross-tabulated with an administrative list that is thought to overlap the target population. The resulting table would have the structure shown in Table 1.

The only directly observable quantity in Table 1 is the total number of people on the list, N_L. An unknown number, n_{01}, of individuals on the list are not in the target population. These n_{01} people constitute "over-coverage" of the list with respect to the target population. If we had an indicator for inclusion or otherwise in the target population it would be straightforward to exclude people not in the target population from population estimation. However, we assume no such indicator and identifying the n_{01} people included on

Table 1: Basic structure for population estimation from an administrative list

		List		
		1	0	
Target	1	n_{11}	n_{10}	N_T
	0	n_{01}	0	
		N_L		

Table 2: Underlying cell-probabilities for population-list union at some setting $\mathbf{x}$ of covariates

		List	
		1	0
Target	1	$\phi_{11}(\mathbf{x})$	$\phi_{10}(\mathbf{x})$
	0	$\phi_{01}(\mathbf{x})$	0

the list but not in the target population is therefore a missing data problem. The missing data absent from the list is an indicator for inclusion in the target population. If we could determine the over-coverage, n_{01}, then since the list total, N_L, is directly observed we would immediately be able to determine the number of people both in the target population and on the list since $n_{11} = N_L - n_{01}$. On the other hand, an unknown n_{10} individuals are in the target population but not on the list. This group represents the "under-coverage" of the list with respect to the target population. If we could estimate n_{11}, then given an estimate of n_{10} and using $\hat{}$ to indicate estimates we could obtain an estimate of the the target population total N_T as

$$\hat{N}_T = \hat{n}_{11} + \hat{n}_{10} = \hat{N}_L - \hat{n}_{01} + \hat{n}_{10}$$

Ideally we would like to estimate not just the total population size N_T but the number of people in the target population by characteristics such as age, sex, ethnic group and area. Therefore we assume a structure such as Table 1 for each combination of these variables. We let $\mathbf{X}$ denote the covariates of interest and $\mathbf{X} = \mathbf{x}$ a particular combination of these variables.

Allowing for dependence on the covariates, Table 2, describes a probability model underpinning the the cross-tabulation of the target population and list. The probabilities for the three occupied cells in Table 2 sum to one. Under this model, an individual in the target population - list union, with covariates $\mathbf{x}$ is allocated to one of the three possible cells with the probabilities given in Table 2. Thus, we posit a multinomial model (with one trial) at the level of individuals.

Given the cell probabilities from Table 2 we can define the under-coverage probability, Pr(not on list |in Target,$\mathbf{X} = \mathbf{x}$) as $\phi^{under}(\mathbf{x}) = \phi_{10}(\mathbf{x})/(\phi_{11}(\mathbf{x}) + \phi_{10}(\mathbf{x}))$ and the over coverage probability for the list, Pr(not in Target|on list, $\mathbf{X} = \mathbf{x}$) as $\phi^{over}(\mathbf{x}) = \phi_{01}(\mathbf{x})/(\phi_{11}(\mathbf{x}) + \phi_{01}(\mathbf{x}))$. Since $\phi_{11}(\mathbf{x}) + \phi_{10}(\mathbf{x}) + \phi_{01}(\mathbf{x}) = 1$ we need specify only two of the cell probabilities to fully specify the multinomial implied by Table 2. A convenient approach to modelling is to model $\phi^{under}(\mathbf{x})$ and $\phi_{01}(\mathbf{x})$. The remaining cell probabilities can then be obtained as $\phi_{11}(\mathbf{x}) = (1 - \phi_{01}(\mathbf{x}))(1 - \phi^{under}(\mathbf{x}))$, $\phi_{10}(\mathbf{x}) = (1 - \phi_{01}(\mathbf{x}))\phi^{under}(\mathbf{x})$.

Notice that the number of people in the (0,0) cell in Table 1, corresponding to "not in the target population and not on the list" is assumed to be 0. In fact most of the world's population falls in this cell! However, we are not interested in the estimating the population of the world but of some specific target population such as the usually resident population of New Zealand and we are seeking to use an administrative list for this purpose. For this problem only people in the target population or on the list or in both are relevant. That is, our conceptual starting point for estimation is the union of the target population and the list (cf Zhang (2015)).

Table 3: Cell-probabilities for the sample-list union at setting $\mathbf{x}$ of the covariates

		List	
		1	0
Sample	1	$\lambda(x)\phi_{11}(\mathbf{x})$	$\lambda(\mathbf{x})\phi_{10}(\mathbf{x})$
	0	$(1-\lambda(\mathbf{x}))\phi_{11}(x)+\phi_{01}(\mathbf{x})$	$(1-\lambda(\mathbf{x}))\phi_{10}(\mathbf{x})$

If a sample has been drawn from the target population with sample inclusion probabilities, $\lambda(\mathbf{x})$, independently of list inclusion, and the sample is linked to the list without error, cross-tabulation of sample and list inclusion indicators produces a 2 x 2 table (at each setting of $\mathbf{X}$) underpinned by the probabilities shown in Table 3. For simplicity we regard the $\lambda(\mathbf{x})$ as known. In practice these sample inclusion probabilities may need to be estimated. From Table 3 it can be seen the sampling process transfers some people from the $(1,0)$ cell in the target population-list union to the $(0,0)$ cell in the sample-list joint distribution. This cell is, in reality, not observable; we do not see the people that are neither on the list nor in the sample, $\mathbf{S}$. An important point is that Table 3 does not represent a traditional capture-recapture population estimation problem. Whereas the latter involves two or more samplings from a target population we have a single sampling from the population which is linked to a list that overlaps the target population.

3. Inference We base inference on the posterior predictive distribution of a corrected list from which individuals not in the target population have been removed and target population members missed by the list have been added. If we can generate, corrected lists from this distribution, then for each draw we could obtain population counts for all cells of interest by simple tabulation. The tabulations obtained by repeating this for each simulated corrected list represent a sample from the joint posterior distribution of the cell counts. Summaries of this distribution such as the mean, median, other quantiles, and approximate credible intervals, obtained in the case of a 95% credible interval, by locating the 2.5th and 97.5th percentiles of the distribution can be obtained straightforwardly.
Introducing the notation Y to denote the cell-location for an individual in the target population - list union and $\tilde{Y}$ to denote the cell location in the sample-list union, assuming the covariate values in the population-list union are drawn from some distribution $G(\boldsymbol{\theta})$ and letting $\boldsymbol{\xi}=(\boldsymbol{\phi},\boldsymbol{\theta})$ we have the model

$$\begin{aligned}
[\mathbf{X}|\boldsymbol{\xi}] &\overset{\text{indep}}{\sim} G(\boldsymbol{\theta}) \\
[Y|\mathbf{X},\boldsymbol{\xi}] &\overset{\text{indep}}{\sim} \text{Multinomial}(1,\boldsymbol{\phi}(\mathbf{X})) \\
[\tilde{Y}|Y,\mathbf{X},\boldsymbol{\xi},\boldsymbol{\lambda}] &\overset{\text{indep}}{\sim} H_Y(\boldsymbol{\lambda},\mathbf{X}) \qquad (1)
\end{aligned}$$

where if $Y=(1,1)$ $H_{(1,1)}\boldsymbol{\lambda},\mathbf{X})$ is the Bernoulli distribution with possible values $(1,1)$ and $(0,1)$ with $\Pr(\tilde{Y}=(1,1)|Y=(1,1),\mathbf{X},\boldsymbol{\lambda})=\lambda(\mathbf{X})$; if $Y=(1,0)$ $H_{(1,0)}(\boldsymbol{\lambda},\mathbf{X})$ is the Bernoulli distribution with possible values $(1,0)$ and $(0,0)$ with $\Pr(\tilde{Y}=(1,0)|Y=(1,0),\mathbf{X},\boldsymbol{\lambda})=\lambda(\mathbf{X})$; if $Y=(0,1)$ $H_{(0,1)}(\boldsymbol{\lambda},\mathbf{X})$ is the degenerate distribution with $\Pr(\tilde{Y}=(0,1)|Y=(0,1),\mathbf{X},\boldsymbol{\lambda})=1$. Sampling of the target population has no impact on the group that is on the list but not in the target population.
The observed data is $\mathbf{D}^{obs}=\{X_i,\tilde{Y}_i;i:\tilde{Y}_i\neq(0,0)\}$. This is the sample-list union. The extra information required to obtain the complete population-list union can be characterised as $\mathbf{D}^{mis}=(\mathbf{Y}^{mis},N_T,\mathbf{X}^{mis})$ where $\mathbf{Y}^{mis}$ are the unobserved population-list cell locations for individuals not in the sample but on the list $(\tilde{Y}=(0,1))$, N_T is the total size of the target population, and $\mathbf{X}^{mis}$ represents the covariate values for people not on the list, in the population but not selected into the sample and hence not observed. Note that people in the $\tilde{Y}=(0,1)$ group are a mix of those on the list but not in the target population $(Y=(0,1))$ and people on the list and in the target population $(Y=(1,1))$ but not selected into the population sample. In contrast $\tilde{Y}=(1,1)\Rightarrow Y=(1,1)$ and $\tilde{Y}=(1,0)\Rightarrow Y=(1,0)$. Given $\mathbf{D}^{mis}$ we could obtain the target population

by first forming $\mathbf{D}^{full} = (\mathbf{D}^{mis}, \mathbf{D}^{obs})$ and then dropping records with $Y = (0,1)$. Cross tabulation of the remaining records provides population counts by categories of $\mathbf{X}$.
Our primary inferential task is therefore to obtain $p(\mathbf{D}^{mis}|\mathbf{D}^{obs})$:

$$p(\mathbf{D}^{mis}|\mathbf{D}^{obs}) = \int p(\mathbf{D}^{mis}|\mathbf{D}^{obs}, \boldsymbol{\xi}) p(\boldsymbol{\xi}|\mathbf{D}^{obs}) \mathrm{d}\boldsymbol{\xi} \tag{2}$$

$$= \int p((\mathbf{D}^{mis}, \boldsymbol{\xi}|\mathbf{D}^{obs}) \mathrm{d}\boldsymbol{\xi} \tag{3}$$

One approach to computing the posterior predictive distribution for $\mathbf{D}^{mis}$ is to use a Gibbs sampler to simulate the joint distribution $p(\mathbf{D}^{mis}, \boldsymbol{\xi}|\mathbf{D}^{obs})$. The generated draws of $\mathbf{D}^{mis}$ can then be used in conjunction with $\mathbf{D}^{obs}$ to produce a Monte Carlo representation of the posterior distribution for the population counts. The Gibbs sampler alternates between sampling from the following full conditional distributions: (i) $p(\boldsymbol{\theta}|\boldsymbol{\phi}, \mathbf{D}^{mis}, \mathbf{D}^{obs})$; (ii) $p(\boldsymbol{\phi}|\boldsymbol{\theta}, \mathbf{D}^{mis}, \mathbf{D}^{obs})$; (iii) $p(\mathbf{D}^{mis}|\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{D}^{obs})$. Steps (i) and (ii) amount to reasonably standard Bayesian computations since they are conditional on the full data. For example step (ii) amounts to drawing from the posterior distribution for the parameters of a multinomial logistic regression model. However, step (iii) is more interesting. The components of $\mathbf{D}^{mis}$ can be simulated sequentially using the following decomposition

$$p(\mathbf{D}^{mis}|\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{D}^{obs}) = p(\mathbf{Y}^{mis}|\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{D}^{obs}) p(N_T|\mathbf{Y}^{mis}, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{D}^{obs}) p(\mathbf{X}^{mis}|N_T, \mathbf{Y}^{mis}, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{D}^{obs}). \tag{4}$$

We can simulate the unobserved cell locations $Y = (1,1)$ or $Y = (0,1)$ for individuals with $\tilde{Y} = (0,1)$ as Bernoulli probabilities with

$$\Pr(Y = (1,1)|\tilde{Y} = (0,1), \mathbf{X}, \boldsymbol{\phi}) = \frac{(1 - \lambda(\mathbf{X}))\phi_{11}(\mathbf{X})}{(1 - \lambda(\mathbf{X}))\phi_{11}(\mathbf{X}) + \phi_{01}(\mathbf{X})}$$

Having generated the $\mathbf{Y}^{mis}$ values we can can count the number of list members with observed or imputed $Y = (1,1)$ (recalling that $\tilde{Y} = (1,1) \Rightarrow Y = (1,1)$). Denoting this count as n_{11}^{sim}, which will of course vary with each iteration of the sampler, and letting $\tilde{n}_{10}$, denote the number of people in the sample-list union with $\tilde{Y} = (1,0)$), the conditional posterior for the total population size can be shown to be

$$p(N_T|\mathbf{Y}^{mis}, \boldsymbol{\xi}, \mathbf{D}^{obs}) \propto p(N_T) \frac{N_T!}{(N_T - n_{11}^{sim} - \tilde{n}_{10})!} \left(\int (1 - \lambda(\mathbf{X})) \frac{\phi_{10}(\mathbf{X})}{\phi_{11}(\mathbf{X}) + \phi_{10}(\mathbf{X})} p(\mathbf{X}|Y \neq (0,1), \boldsymbol{\theta}) \mathrm{d}\mathbf{X} \right)^{N_T - n_{11}^{sim} - \tilde{n}_{10}} \tag{5}$$

(see Fienberg et al (1999) for an analogous formulation in a related, though not identical, setting).
Given a total population size of N_T we know there are $\tilde{n}_{00} = N_T - n_{11}^{sim} - \tilde{n}_{10}$ people missing from the (1,0) cell (i.e in the observed (0,0) cell) and so in order to complete the target population we generate $\tilde{n}_{00}$ draws from

$$p(\mathbf{X}|\tilde{Y} = (0,0), \boldsymbol{\xi}) \propto (1 - \lambda(\mathbf{X}))\phi_{10}(\mathbf{X}) p(\mathbf{X}|\boldsymbol{\theta})$$

$$\propto (1 - \lambda(\mathbf{X})) \frac{\phi_{10}(\mathbf{X})}{\phi_{11}(\mathbf{X}) + \phi_{10}(\mathbf{X})} p(\mathbf{X}|\boldsymbol{\xi}, Y \neq (0,1)) \tag{6}$$

This latter step gives the $p(\mathbf{X}^{mis}|N_T, \mathbf{Y}^{mis}, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{D}^{obs})$ term in (4).
A disadvantage of the Gibbs Sampler when working with large datasets is that computation times can be long. Consequently, we consider an approximation to the posterior predictive distribution of the missing data based on (2). Firstly we note that (2) implies

$$p(\mathbf{D}^{mis}|\mathbf{D}^{obs}) = \int p(\mathbf{D}^{mis}|\mathbf{D}^{obs}, \boldsymbol{\phi}, \boldsymbol{\theta}) p(\boldsymbol{\phi}|\mathbf{D}^{obs}) p(\boldsymbol{\theta}|\boldsymbol{\phi}, \mathbf{D}^{obs}) \, \mathrm{d}\boldsymbol{\phi} \mathrm{d}\boldsymbol{\theta}. \tag{7}$$

We have just seen how to sample from $p(\mathbf{D}^{mis}|\mathbf{D}^{obs}, \boldsymbol{\phi}, \boldsymbol{\theta})$ which is just step (iii) of the Gibbs sampler. The posterior for the coverage parameters $p(\boldsymbol{\phi}|\mathrm{D}^{obs})$ can be obtained directly by combining a prior, $p(\boldsymbol{\phi})$, with a

likelihood, $L^{cond}(\boldsymbol{\phi})$ constructed from the probabilities given in Table 3, after conditioning on being recorded in the observed data ($\tilde{Y} \neq (0,0)$):

$$L^{cond}(\boldsymbol{\phi}) = p(\tilde{\mathbf{Y}}^{obs}|(\tilde{\mathbf{Y}} \neq (0,0), \mathbf{X}^{obs}, \boldsymbol{\phi})$$
$$= \prod_{i:\tilde{Y}_i=(1,1)} \frac{\lambda(\mathbf{x}_i)\phi_{11}(\mathbf{x}_i)}{1-(1-\lambda(\mathbf{x}_i))\phi_{10}(\mathbf{x}_i)} \prod_{i:\tilde{Y}_i=(1,0)} \frac{\lambda(\mathbf{x}_i)\phi_{10}(\mathbf{x}_i)}{1-(1-\lambda(\mathbf{x}_i))\phi_{10}(\mathbf{x}_i)} \prod_{i:\tilde{Y}_i=(0,1)} \frac{\phi_{01}(\mathbf{x}_i)+(1-\lambda(\mathbf{x}_i))\phi_{11}(\mathbf{x}_i)}{1-(1-\lambda(\mathbf{x}_i))\phi_{10}(\mathbf{x}_i)} \tag{8}$$

Obtaining the posterior for $\boldsymbol{\theta}$ is more difficult, in part because of the subtle dependence on the coverage parameters induced by the implicit conditioning on $\tilde{Y} \neq (0,0)$ inherent in the observed data. However, considering the steps required for computing $p(\mathbf{D}^{mis}|\mathbf{D}^{obs}, \boldsymbol{\phi}, \boldsymbol{\theta})$ it can be seen, that since only covariate values for the population sub-group under-covered by the list need to be imputed, there is a sense in which only the covariate distribution for the population, rather than for the population-list union needs to be modelled. Further the observed data includes a sample from the target population which could be used for inference concerning parameters of the population covariate distribution. Formalising these ideas let $\boldsymbol{\theta}_T$ denote the parameters of the population covariate distribution. Strictly speaking $\boldsymbol{\theta}_T$ should satisfy

$$p(\mathbf{X}|\boldsymbol{\theta}_T, Y \neq (0,1)) = \frac{1-\phi_{01}(\mathbf{X})p(\mathbf{X}|\boldsymbol{\theta})}{\int(1-\phi_{01})(\mathbf{X})p(\mathbf{X}|\boldsymbol{\theta}),\, \mathrm{d}\mathbf{X}} \tag{9}$$

for all values of $\mathbf{X}$. However if we are willing to ignore (9), replace $p(\boldsymbol{\theta}|\boldsymbol{\phi}, \mathbf{D}^{obs})$ with $p(\boldsymbol{\theta}_T|\mathbf{S})$, in (7) where $\mathbf{S}$ is the data from the population sample and replace $p(\mathbf{X}|\boldsymbol{\xi}, Y \neq (0,1)$ with $p(\mathbf{X}|\boldsymbol{\theta}_T, Y \neq (0,1))$ in (6) we can approximate the posterior predictive distribution of $\mathbf{D}^{mis}$ by sampling the components of the right hand side of (7), proceeding from right to left.

Using $p(\boldsymbol{\theta}_T|\mathbf{S})$, for inference for the target population covariate distribution involves throwing away information about this distribution in the observed $(0,1)$ cell. Since this cell contains a mix of target population individuals not selected into the population sample and people not in the target population it seems, intuitively, that the cell must contain some information on the target population distribution. However it is difficult to access this information with first stripping away the list over-coverage, as accomplished by the conditioning on the completed population - list union in step (i) of the Gibbs sampler.

4. Application In order to provide an initial test of the methods described above we constructed a small target population example by drawing a simple random sample of 5000 records from the synthetic version of the 2011 New Zealand Household Income Survey and treating this as a target population. From this small target population we then drew a subsample of 4,414 records in a manner which ensured higher income groups and younger age-groups were under-represented. This subsample was included on the simulated list and constitutes the true (1,1) cell for this exercise. The 586 records not included in the subsample constituted the under-coverage, i.e the (1,0) cell. To simulate the over-coverage we drew 1000 records from the set of just over 13,000 Household Income Survey records that were not included in the target population sample of 5000. Records with lower incomes were over-represented in the over-coverage sample. The over-coverage sample was appended to the subsample of 4,414 records drawn from the target population to constitute the observed list for this exercise, comprising a total of 5,414 records. Finally, we simulated a coverage survey by drawing a simple random sample of 1000 records from the simulated target population. The simulated coverage survey sample included 846 records that were also included in on the simulated list and these 846 records therefore constitute the observed (1,1) cell for this exercise.

For modelling and estimation we re-parameterised the multinomial model in terms of the under-coverage, ϕ^{under} and the over-coverage-parameter ϕ_{01} and specified logistic regression models for these probabilities. We adopted weakly informative priors for the logistic regression parameters (Gelman et al, 2008). We used the approximate method describe in Section 3 for directly obtaining the posterior predictive distribution of the corrected list rather than the Gibbs Sampler. We adopted a restricted uniform prior for the total population size, N_T, and a Bayesian bootstrap model (Rubin, 1981) for the covariate distribution in the target population, based on the values of $\mathbf{X}$ represented in the coverage survey sample.

The posterior median for the total population size was 4,981 and the 95% credible interval based on the 2.5th and 97.5th percentiles was (4,711 to 5,290). These estimates are in close agreement with the true

population size of 5,000 and the methodology appears to have successfully dealt with the net over-coverage in the simulated list of 5,414 records. A comparison of posterior median and 95% credible intervals by sex and age (in ten year groups) with the true counts and counts obtained directly from the observed list are shown in Figure 1. The estimates generally track the true counts well, although the posterior median for females aged 35-45 was actually further than from the true count than the count obtained from the observed lists. Most likely this is just a quirk of the single simulation we have run. However we will investigate the issue by conducting a full simulation study.

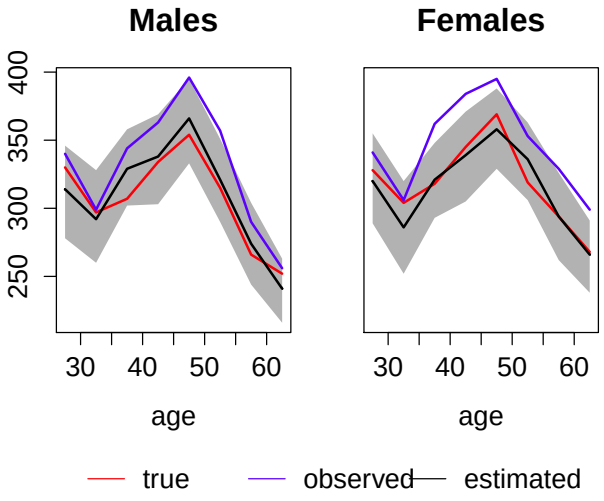

Figure 1: Plot of observed and estimated population counts with 95% credible intervals compared to true counts.

5. Conclusions We have outlined a Bayesian approach to estimating the size and distribution of a population using an administrative list in conjunction with a coverage survey sample drawn from the target population and linked to the list. In our single simulated example our methodology showed encouraging results. However, much more extensive evaluation of the methodology, involving larger scale problems and full simulation evaluation of the Bayesian estimates is required. Our approximate approach to obtaining the posterior predictive distribution of the corrected list runs fast enough to make simulation evaluation possible. However, comparisons of our approximate method with the Gibbs sampling approach would be useful to understand the impact, if any, of ignoring information in observed $(0, 1)$ cell concerning the target population covariate distribution. In order to apply the methodology outlined here to realistic situations it needs to be extended to accommodate more complex survey designs, survey non-response, misclassification of list variables and potentially error in the linkage of the survey to the list. We are currently investigating all these issues.

References

Bycroft, C. (2015) Census Transformation in New Zealand: Using administrative data without a population register. Statistical Journal of the IAOS, 31(3), 401-411.

Fienberg S., Johnson, M.S. & Junker B.W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. Journal of the Royal Statistical Society, A, 162(3), 383-405.

Gelman, A., Jakulin, A., Pittau, M.G., & Su, Y-S. (2008). A weakly informative default prior distribution for logistc and other regression models. Annals of Applied Statistics, 2(4), 1360-1383.

Rubin, D.B. (1981). The Bayesian bootstrap. Annals of Statistics, 9(1), 130-134.

Zhang, L.C. (2015). On modelling register coverage errors. Journal of Official Statistics, 31(3), 381-396.