Methods and results for small area estimation using smoking data from the 2008 National Health Interview Survey

* Neung Soo Ha[†] Partha Lahiri[‡] Van Parsons[§]

"The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention."

Abstract

The National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics (NCHS), is designed to provide reliable design-based estimates for national and four major geographical regions of the United States for a wide range of health related variables. However, state or sub-state level estimates are likely to be unreliable since they are based on small sample sizes. In this paper, we compare the efficiency of different area level models in estimating smoking prevalence for the fifty U.S. states and the District of Columbia using the 2008 NHIS survey data in conjunction with a a number of potentially related auxiliary variables obtained from the Area Resource File (ARF), a large county level database compiled from different U.S. federal agencies. A major portion of this study is devoted to the investigation of various methods to estimate the sampling variances needed to implement an area level model. In our data analysis, the hierarchical Bayes method based on the random sampling variance model appears to be have an edge over other area level models considered in the paper.

Key Words: Hierarchical Bayesian modeling, MCMC, Small area estimation, National Health Interview Survey, Generalized mixed effect model.

1. Introduction

National sample surveys are generally designed to provide reliable design-based direct estimates for targeted large areas or domains of a survey population. However, direct design-based methods tend to be unreliable for different subpopulations with small sample sizes such as subnational geographical regions (e.g., states, counties, etc.) or cells formed by cross-classifying different demographic variables. As an alternative to design-based direct estimators, a wide range of methods that borrow strength from supplementary databases using explicit models, such as linear mixed effects models (LMM) or generalized linear mixed models (GLMM) have been proposed in the literature. For a review of different small area methods, we refer to Rao (2003) and Jiang and Lahiri (2006).

In section 2, we describe databases used in this paper. In section 3, we first point out a few shortcomings of the well-known Fay-Herriot area level model (Fay and Herriot (1979)) and then discuss a few alternative area level modeling approaches proposed in the literature. In section 4, we introduce five methods for stabilizing the variance estimates for small areas. In sections 5 and 6, we examine model-based estimates and the model fit via posterior distributions, and we discuss results. We conclude in section 7 with the summary and provide future research areas.

*List of authors is in alphabetical order of their last names

[†]National Center for Health Statistics , 3311 Toledo Rd, Hyattsville, MD 20782

[‡]Joint Program in Survey Methodology, 1218 LeFrak Hall College Park, MD 20742

[§]National Center for Health Statistics , 3311 Toledo Rd, Hyattsville, MD 20782

2. Data Description and Design Characteristics

2.1 The NHIS description

The NHIS is an annual survey with a state-stratified multi-stage complex design. The survey design includes primary sampling units, (PSUs), which are individual counties or contiguous groups of counties, and they are taken without replacement with probability of proportional to their estimated sizes. Within each PSU, the survey uses an area frame for sampling and it further divides into sub-strata and clusters. We use in-house data and so we are able to include survey design features unavailable to the public. In particular, we use clusters of blocks of housing units for variance estimation.

In addition to detailed survey design variables, the data include various geographical identifiers, demographic and health related variables. The survey is designed to produce reliable survey-weighted direct estimates and their associated design-based standard errors for the nation and four major census regions, (Region 1: Northeast, Region 2: Midwest, Region 3: South, and Region 4: West). For more details about the survey, see (`http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm#2008_NHIS`).

2.2 The Area Resource File

For our area-specific auxiliary variables, we have used data from the Area Resource File (ARF). The ARF is a database containing more than 6,000 variables for each of the nation's counties. It contains health related information, measures of resource scarcity, economic activity, health training programs, and socioeconomic and environmental characteristics. It also contains geographic codes that enables various quantities of interests to be aggregated into desired groupings. The data sets are from many different government agencies, such as Bureau of Labor Statistics, the Census, and the U.S. Department of Agriculture. More information about the ARF, see (`http://arf.hrsa.gov/faqs.htm`).

Some of the variables derived from ARF could be sample survey based and hence could be subject to sampling errors, an aspect we did not address in the paper. But, if the county specific sample sizes are large (e.g., statistics computed from the American Community Survey), one could ignore sampling variances for such estimates as an approximation.

3. Different Area Level Models and Statistical Inference

3.1 Notation and model description

Let y_{ij} be a binary response for the characteristic of interest for unit j in area i and N_i be the population size in area i ($i = 1, \ldots, m$, $j = 1, \ldots, N_i$). The parameter of interest is the small area proportion, $\theta_i = \sum_{j=1}^{N_i} y_{ij}/N_i$. We can estimate θ_i by $\hat{\theta}_i = \frac{\sum_{j=1}^{n_i} w_{ij} y_{ij}}{\sum_{j=1}^{n_i} w_{ij}}$, direct survey-weighted estimate of θ_i; n_i is the sample size for area i; w_{ij} is the survey-weight associated with unit j in area i ($i = 1, \ldots, m$, $j = 1, \ldots, N_i$).

A widely used small area model is the Fay-Herriot model (Fay and Herriot (1979)) described as:

Model 1: The Fay-Herriot Model
For $i = 1, \ldots, m$,

$$\begin{aligned} &Level1\ (Sampling\ model) && : && \hat{\theta}_i|\theta_i \overset{\text{ind}}{\sim} \mathcal{N}(\theta_i, \psi_i) \\ &Level2\ (Linking\ model) && : && \theta_i|\boldsymbol{\beta}, A \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, A), \end{aligned} \tag{1}$$

where $\mathbf{x}_i$ is a vector of known covariates for area i; the sampling variance ψ_i is assumed to be known; the hyper-parameters A and $\boldsymbol{\beta}$, are unknown.

The Fay-Herriot model is an example of a matched model, Rao (2003), since the two levels can be combined into a single linear mixed model:

$$\hat{\theta}_i = \mathbf{x}_i'\boldsymbol{\beta} + \nu_i + \epsilon_i, \tag{2}$$

where $\{\nu_i\}$ and $\{\epsilon_i\}$ are all independent with $\nu_i \sim \mathcal{N}(0, A)$ and $\epsilon_i \sim \mathcal{N}(0, \psi_i)$.

The assumption of normality for both sampling and the linking models could be problematic as they do not guarantee positive support between 0 and 1 for the posterior distribution of θ_i; see Liu et al. (2007) and Lahiri (2011). Normality of the sampling distribution as well as the linking distribution can be handled using models given in Jiang and Lahiri (2001), Liu et al. (2007) and others. As argued by Lahiri and Rao (1995), the assumption of normality of the linking model may be more severe than that of the sampling distribution since there may be certain central limit effect on the survey-weighted proportions. Thus, for the sake of simplicity, we consider normality for the sampling distribution but a logistic model for the linking distribution. Such a model may be called a unmatched model, Rao (2003). In this paper, we consider the following two unmatched models:

Model 2: The Normal-Logistic Model
For $i = 1, \ldots, m$,

$$\begin{aligned} &Level1\ (Sampling\ model) && : && \hat{\theta}_i|\theta_i \overset{\text{ind}}{\sim} \mathcal{N}(\theta_i, \psi_i) \\ &Level2\ (Linking\ model) && : && \text{logit}(\theta_i)|\boldsymbol{\beta}, A \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, A). \end{aligned} \tag{3}$$

Model 3: The Normal-Logistic Random Sampling Variance Model
For $i = 1, \ldots, m$,

$$\begin{aligned} &Level1\ (Sampling\ model) && : && \hat{\theta}_i|\theta_i \overset{\text{ind}}{\sim} \mathcal{N}\Big(\theta_i, \psi_i = \frac{\theta_i(1-\theta_i)}{n_i}\text{DEFF}_i\Big) \\ &Level2\ (Linking\ model) && : && \text{logit}(\theta_i)|\boldsymbol{\beta}, A \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, A), \end{aligned} \tag{4}$$

where DEFF_i is the true design effect; that is, the ratio of the true sampling variance of the survey-weighted proportion under the complex design to the true sampling variance of the unweighted sample proportion under simple random sampling.

Both Model 2 and Model 3 were considered by Liu et al. (2007). The idea of capturing a part of uncertainty due to the estimation of small area sampling variances using a random sampling variance model can be traced back to Arora and Lahiri

(1997). For implementation of the three models, we need to estimate ψ for Model 1 and Model 2 and DEFF for Model 3. You and Rao (2002) considered an alternative decomposition of the sampling variances as $\psi_i = \theta_i^2 \mathrm{CV}_i^2$, where CV_i is the coefficient of variation of the survey-weighted proportion $\hat{\theta}_i$. However, unknown θ_i seems to have more influence on CV_i than DEFF_i (see Lahiri 2011 for discussion) and thus we prefer to choose the decomposition considered by Liu et al. (2007).

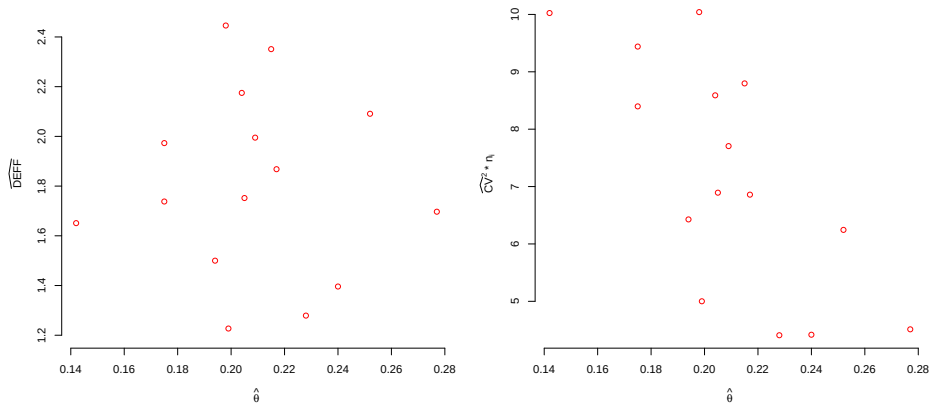

Figure 1: Comparison btw estimates of $\widehat{\mathrm{DEFF}}_i$ and $\widehat{\mathrm{CV}}^2$ vs. $\hat{\theta}$

Figure 1 shows how the direct design estimates $\hat{\theta}$ are related to $\widehat{\mathrm{DEFF}}$ and $n_i \times \widehat{\mathrm{CV}}^2$ for the 15 largest U.S. states. It seems that $\widehat{\mathrm{CV}}$ has strong dependency with direct design-based estimates with adjusted $R^2 = 0.479$ when $n_i \times \widehat{\mathrm{CV}}_i^2$ estimates are regressed against $\hat{\theta}_i$. However, $\widehat{\mathrm{DEFF}}$ estimates have no clear relationship with direct design-based estimates with adjusted $R^2 = -0.076$ when $\widehat{\mathrm{DEFF}}_i$ estimates are regressed against $\hat{\theta}_i$.

3.2 Estimation of ψ and DEFF_i

To use the hierarchical models discussed in section 3.1, an estimate of the within-area sampling variance, ψ_i, must be available. In the 2008 NHIS data, while all states have sample, some low population states have few sampled clusters, thus their design-based variance estimates could lead to bias and instabilities. Treating such estimators as accurate inputs may reduce the effectiveness of the hierarchical models. Here, we propose five methods for estimating ψ_i to more accurately represent state proportion estimates and their properties.

A Generalized Variance Function, (GVF), is a model that describes the relationship between a statistic and its corresponding variance, and traditionally it has been used for variance estimates, Otto and Bell (1995), Wolter (1985). Our first method uses a combination of GVF var_i^{gvf} and direct sampling variance estimate var_i for each state i. The GVF model in use was defined as $var_i = \alpha + \beta/\hat{\theta}_i$, and we have used this model to obtain fitted values for states with small number of clusters. The variance estimate, $\hat{\psi}_i$, for the method 1 is defined as:

Method 1

$$\hat{\psi}_i = \begin{cases} max(var_i^{gvf}, var_i) & \text{, if number of clusters for } i\text{th state} < 100 \\ var_i & \text{, if number of clusters for } i\text{th state} > 100. \end{cases} \quad (5)$$

We have perceived that using the maximum between the GVF fitted values and var_i seems as a conservative approach, and also recognized that direct design-based variance estimates for states with large number of clusters are stable (the number 100 for cluster size for each state is arbitrary).

In the next four methods, we have applied methods considered by Liu et al. (2007), where the estimator $\hat{\psi}_i$ is decomposed into estimates of DEFF and the true sampling variance of the unweighted sample proportion under simple random sampling, (SRS).

For the second method, we have replaced true sampling variance for sampling proportion under SRS with its estimate, $\frac{\hat{\theta}_i(1-\hat{\theta}_1)}{n_i}$, and parameter DEFF is replaced with $deff_i^{adj}$, which is defined below.

Method 2

$$\hat{\psi}_i = \frac{\hat{\theta}_i(1-\hat{\theta}_i)}{n_i} \cdot deff_i^{adj}, \quad deff_i^{adj} = \begin{cases} deff_i & \text{if } .99 < deff_i < 3.5 \\ \overline{deff_i} & otherwise, \end{cases} \quad (6)$$

where $deff_i$ is the design-based estimate of DEFF$_i$; $\overline{deff_i} = \frac{1}{10}\sum_{j=1}^{10} deff_{i(j)}$ and $deff_{i(j)}$ $j = 1, \ldots, 10$ are design effects of 10 largest states.

For the third method, we have estimated ψ_i by using quantities from the four census regions. For $j = 1, \ldots, 4$, we have:

Method 3

$$\hat{\psi}_i = \frac{\hat{\theta}_{j,region}(1-\hat{\theta}_{j,region})}{n_i} \cdot \text{deff}_{j,region}, \quad (7)$$

where $\hat{\theta}_{j,region}$ and $deff_{j,region}$ are the direct survey-weighted proportion and the direct design-based design effect estimates for the region in which the small area i lies.

For the fourth method, instead of using the direct design-based estimate $\hat{\theta}_i$, we first estimate $\hat{\theta}_i$ by fitting a logistic regression on 20 largest states and then predict $\hat{\theta}_i$ for other states by applying the same function. The underlying idea is that when n_i is small in area i, the direct design-based estimate, $\hat{\theta}_i$, is not very reliable, thus by fitting a regression line, the fitted estimates for ith area "borrow strength" from other areas with similarities. Thus, $\hat{\psi}_i$ is defined as:

Method 4

$$\hat{\psi}_i = \frac{\hat{\theta}_i^{synth} \cdot (1-\hat{\theta}_i^{synth})}{n_i} \cdot deff_i^{adj}, \quad (8)$$

where $\hat{\theta}_i^{synth}$, is the fitted value of θ_i from the logistic regression.

For the fifth method, (the random sampling variance model), we have used a close approximation of ψ_i suggested from the model 3 in section 3. Liu et al. (2007) have shown that true ψ_i is approximately:

$$\psi_i \approx \frac{\theta_i(1-\theta_i)}{n_i} \cdot deff_i$$

, where $deff_i$ is a reasonable design effect estimate. We have tried both $\text{deff}_{j,region}$ and deff_i^{adj}. In this paper, we present results for deff_i^{adj} only, so that $\hat{\psi}_i$ is defined as:

Method 5

$$\psi_i \approx \frac{\theta_i \cdot (1-\theta_i)}{n_i} \cdot deff_i^{adj}. \tag{9}$$

3.3 Inference on Small Area Estimation

For inference about parameters in small area estimation models, the hierarchical Bayesian (HB) approach is explored. With the HB method, making an appropriate choice of prior distribution for hyper-parameters is important in calculating the posterior distributions. We choose non-informative prior distributions for A and $\boldsymbol{\beta}$ (uniform in a finite interval with large length) so the the issue of impropriety does not arise.

The WinBUGS package, (`http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml`), is used to implement the HB models, and the Bayesian inference is carried out by the Markov Chain Monte Carlo (MCMC) techniques. For convergence diagnostics of MCMC chains, we have followed the suggestion by Gelman et al. (2004); we have calculated the Gelman and Rubin $\hat{R}$ statistic using three chains, and for each parameter, we examined its auto-correlation function (ACF) plot, Albert (2007). See Robert and Casella (2004) and Gelman et al. (2004) for more details about stochastic sampling methods, such as Metropolis Hastings and Gibbs sampling methods.

4. Model Selection

The task for model selection is quite complex for our project. We focused on three aspects of model selections: (i) auxiliary variables, (ii) choice of the two level models and (iii) the choice of the design effects method. We considered two basic steps to choose the final model. First we focused on selection of auxiliary variables in subsection 4.1 and then given the set of selected auxiliary variables we compared five different methods in subsection 4.2 and 4.3.

4.1 Selection of auxiliary variables

There are several county specific auxiliary variables available from ARF. First, we obtained state specific auxiliary variables from the ARF by appropriate aggregation. In order to select a few useful auxiliary variables for the area level models, we

applied standard regression model selection techniques with $\text{logit}(\hat{\theta})$ as the dependent variable using data for the 15 largest states. Basically, we are assuming that for these 15 largest states sampling variability in the survey-weighted estimates can be ignored so we can take advantage of standard regression tools.

The following state specific auxiliary variables emerge as potential auxiliary variables for our area level models:

- Percent of minority population
- Poverty rate
- Percent of population without high school diploma
- Percent of population with age 65 and above

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.3820     0.3594  -1.063  0.31280
min_per      -0.9493     0.5372  -1.767  0.10764
pov_per       4.1615     1.2538   3.319  0.00776 **
edu9_per    -10.0193     1.9895  -5.036  0.00051 ***
p.agegp3     -3.3229     1.3313  -2.496  0.03167 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.0916 on 10 degrees of freedom
Multiple R-squared: 0.8544,        Adjusted R-squared: 0.7961
F-statistic: 14.67 on 4 and 10 DF,  p-value: 0.0003453
```

We have ran a simple linear regression model with these covariates against the logit of direct design-based estimates for the 15 largest states. Under the model, even though the minority percentage was not significant by a slight margin, including this covariate improved the overall model fit by adjusted-R squared value.

4.2 Comparison of the five methods

Given the set of four auxiliary variables selected in subsection 4.2, we compared five different methods. Methods 1-4 represent the normal-logistic model with the four different estimators for ψ. Method 5 represents the normal-logistic random sampling variance model where DEFF is estimated by Method 2. We now present two different ways to compare these five methods.

4.2.1 Model selection using benchmarking criterion

Without the benchmarking, one way to analyze the performance of the posterior estimates for each state is to compare their aggregates at the census region against the regional design estimators. (The 2008 NHIS is constructed to give precise estimators at the census region.) So, we examined the following relative error (RE):

$$\text{RE}_j = \left| \frac{\sum_{i \in j} w_i \hat{\theta}_i^{ps} - \hat{\theta}_{j,region}}{\hat{\theta}_{j,region}} \right|, j = 1, \ldots, 4,$$

where $\hat{\theta}_{j,region}$ is the census regional design-based direct estimate, w_i is survey weight for state i, and $\hat{\theta}_i^{ps}$ is posterior mean from the area-level HB model at state i. This ratio would provide a criterion for comparing the five different methods.

Cen. Rgn.		Method 1	Method 2	Method 3	Method 4	Method 5
Rgn 1:	NE	0.0763	0.0562	0.0467	0.0539	0.0420
Rgn 2:	MW	0.0752	0.0480	0.0431	0.0396	0.0379
Rgn 3:	S	0.0869	0.0549	0.0493	0.0472	0.0428
Rgn 4:	W	0.0316	0.0090	0.0062	0.0008	0.0012

Table 1: Relative errors for five methods

From Table 1, we see that, in the census region 4, all methods performed well. Overall, method 5 performs the best except in region 4 where Method 4 is marginally better. Method 1 consistently performed the worst in all regions.

4.2.2 Bayesian p-value

Even though the previous section describes an assessment of model performance for different methods at the census regional level, an additional appraisal is still needed at the state level. For this purpose, we use the Bayesian p-value (Gelman et al., 2004). Similar data analysis was carried out in a number of papers (see, e.g., Datta et al. (1999) and Rao (2003)). Their main idea is that if the model fits, then replicated data generated under the model should look similar to observed data. That is, the observed data should look plausible under the posterior predictive distribution. Thus, the technique for checking the model fit is to draw simulated samples from the posterior predictive distribution and compare these to the observed data.

Let y_{obs} denote observed data and y_{new} be predicted data from a distribution, $f(y|\theta)$. Let functions $f(d(y_{obs},\theta)|y_{obs})$ and $f(d(y_{new},\theta)|y_{obs})$ be the posterior (predictive) distributions of $d(y_{obs},\theta)$, and $d(y_{new},\theta)$, where $d(y,\theta)$ is a χ^2-type discrepancy measure, Gelman et al. (2004), defined as:

$$d(y,\theta) = \sum_{i=1}^{50} (\sigma_i^2)^{-1}(y_i - \theta_i)^2.$$

Parameter σ_i^2 is the true variance for area i, and we implement this measure for our analysis with replacing σ_i^2 by different $\hat{\psi}_i$.

We generate parameters, $\theta^{(l)}$, from the posterior distribution, $f(\theta|y_{obs})$ and new data $y^{(l)}$ from $f(y|\theta^{(l)})$, $l = 1,\ldots,B$, where $B(= 3,500)$ is the total number of iterations. Then, we have generated two sets of samples, $d(y_{obs},\theta^{(l)})$ and $d(y^{(l)},\theta^{(l)})$. These are used to approximate the Bayesian p-value, Gelman et al. (2004), by the test quantity $P\{d(y_{new},\theta) \geq d(y_{obs},\theta)|y_{obs}\}$, which is approximated by,

$$p_B \approx B^{-1}\sum_{l=1}^{B} \mathcal{I}\{d(y^{(l)},\theta^{(l)}) \geq d(y_{obs},\theta^{(l)})\}, \tag{10}$$

where $\mathcal{I}(\cdot)$ is an indicator function.

An extreme value (near 0 or 1) of the Bayesian p-value approximate (10) indicates lack of fit of a given model, whereas for an adequate model, this measure will be close to 0.5.

Methods	measure
1	0.2552
2	0.3922
3	0.3716
4	0.4025
5	0.4019

Table 2: Bayesian p-values for different methods.

According to this diagnostic, method 4 outperforms method 5 by a slight margin, and method 2 performs equally as method 4 and 5. This diagnostic further strengthens the claim that method 1 performs the worst; it is still outperformed by all other methods.

5. Estimation

Since from the previous data analysis, methods 4 and 5 emerge as the best two, we compare these methods further.

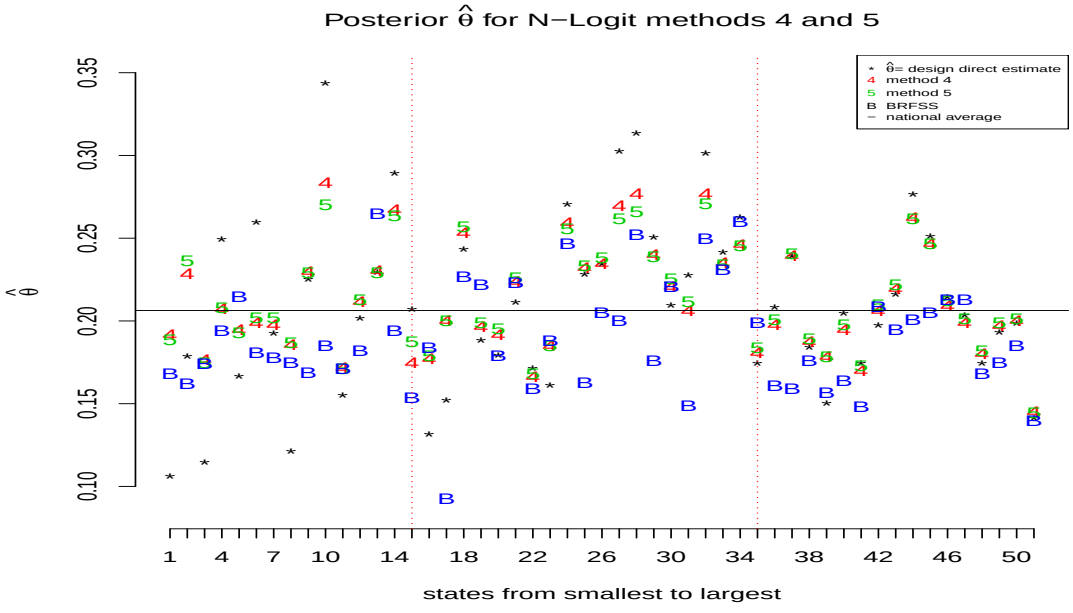

Figure 2: Estimates by states

Figure 2 shows the hierarchical Bayes estimates obtained using method 4 and 5 for each state by its sample size. For confidentiality reasons, actual state names are omitted. States are partitioned into three groups by their sample size, (shown by vertical dotted lines). The symbol "B" represents estimates from another large

survey, the Behavioral Risk Factor Surveillance System, (BRFSS). The BRFSS is a state-based system of health surveys that collects information on health related subjects via telephone interviews. For more information about the BRFSS, see (`http://www.cdc.gov/BRFSS/about.htm`). The national smoking average of 2008 is represented by the solid line at 20.6%.

Note that the direct estimates are bouncing around the national estimates with more variability among the estimates for the group of states with smaller sample size. The variability among the direct estimates reduces as we move from left to right side of the graph. For the largest states direct estimates are very similar to the two hierarchical Bayes estimates. As shown in figure 3, we see a pattern that among smaller states, model-based estimates of both methods are pulled towards the national average. The BRFSS estimates for the large states are mostly below the national line, which is in contrast with the other estimates.

Let us now turn our attention to compare the posterior standard errors from method 5 with the design-based standard error of the survey-weighted proportions. We only consider method 5 since it is likely to report a more accurate standard error estimates than method 4 as method 5, unlike method 4, attempts to explain a part of variability in estimating the sampling variances.

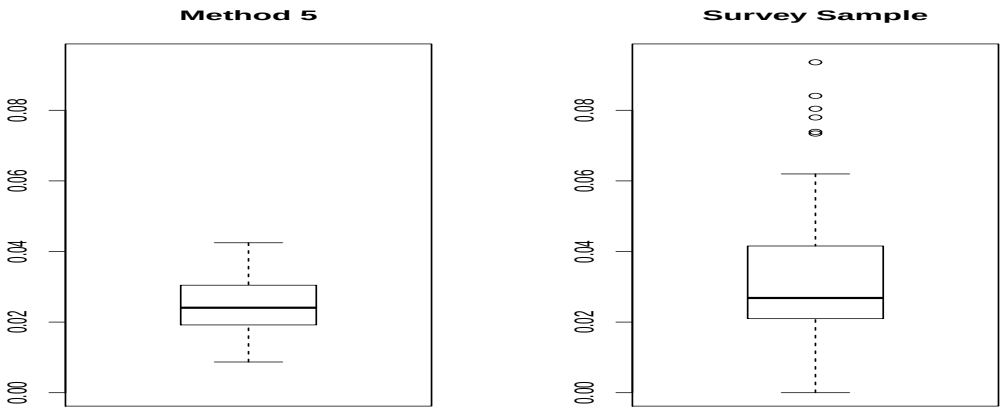

Figure 3: Box plots of std. errors

The box plots of the reported standard errors for the design-based method and method 5 are displayed in Figure 3. The design-based standard errors are much more variable than the posterior standard errors. Also, on the average the posterior standard errors seem to be smaller than the design-based standard errors. While Figure 3 is indicating superiority of method 5 over the design based method in terms of their respective reported standard errors, it is not a confirmatory test since design-based standard errors are not strictly comparable with the posterior standard errors.

6. Summary

This study provides estimates of smoking proportion for each state by using the 2008 NHIS data and small area estimation models. The area-level models require the user to know ψ_i, the true value of area-level variance, but in practice, it must be estimated. We have shown five methods for estimating the area-level variance and provided their assessment at the census regional level and the state level. Our results have shown that although all methods have reasonable outcomes, the most widely used method performed worse than others. We also have concluded that the estimation of ψ_i in the smaller states results in larger differences compared to their design estimate counter parts.

In future research, we would like to explore other models for sub-state estimates using the NHIS data and analyze their performance. We hope that this study has provided some insights about small area estimation modeling.

References

Albert, J. (2007). *Bayesian Computation with R*. Springer, New York, NY.

Arora, V. and Lahiri, P. (1997). On the superiority of the bayesian method over the BLUP in small area estimation problems. *Statistical Sinica*, 7:1053–1063.

Datta, G., Lahiri, P., Maiti, T., and Lu, K. (1999). Hierarchical bayes estimation of unemplyment rates for the states of the U.S. *Journal of the American Statistical Association*, 94:1074–1082.

Fay, R. E. and Herriot, R. A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, 74:269–277.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall/CRC, New York, NY.

Jiang, J. and Lahiri, P. (2001). Empirical best prediction for small are inference with binary data. *Annals of Institute of Statistical Mathematics*, 53:217–243.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15:111–999.

Lahiri, P. and Rao, J. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association*, 82:758–766.

Liu, B., Lahiri, P., and Kalton, G. (2007). Hierarchical bayes modeling of survey-weighted small area proportions. In *Proceedings of the Survey Research Methods Section, ASA*.

Otto, M. and Bell, W. (1995). Sampling error modelling of poverty and income statistics for states. In *Proceedings of the Government Statics Section, ASA*.

Rao, J. (2003). Wiley Series in Survey Methodology, Hoboken, NJ.

Robert, P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, NY.

Wolter, K. (1985). *Introduction to Variance Estimation*. Springer, New York, NY.

You, Y. and Rao, J. (2002). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, 30:3–15.