Comparing Direct Survey and Small Area Estimates of Health Care Coverage in New York

Jeniffer Iriondo-Perez, Amang Sukasih, and Rachel Harter
RTI International, P.O. Box 12194, Research Triangle Park, NC 27709-2194

Abstract
The Behavioral Risk Factor Surveillance Survey (BRFSS) is designed to produce estimates for states and large metropolitan areas. For some variables, county-level BRFSS estimates have been produced by others using small area estimation methodology. We generated county-level estimates of the proportion of adults with health insurance coverage (insured rate) using the New York BRFSS data. The goal was to demonstrate the small area estimation (SAE) technique using the readily-available *R* package `BayesSAE`. As a validation, we compared the results from `BayesSAE` with those from `OpenBUGS`, a well-established software for Bayesian computation. We also compared the model-based estimates with direct design-based estimates.

Keywords: SAE, area level model, Fay-Herriot model, hierarchical Bayes, `BayesSAE`, `OpenBUGS`, Behavioral Risk Factor Surveillance Survey (BRFSS)

1. Background

State and local government agencies need actionable data to effectively manage their programs. State and local agencies are particularly interested in health/welfare, education, labor, transportation, public safety, and special issues such as the opioid epidemic. Such agencies rarely have funds to conduct surveys on a regular basis, although they might have related administrative data for their own jurisdictions. Federal programs, on the other hand, regularly produce data that inform federal policy, but it is typically not their mission to provide data at state and local levels. Even so, federal agencies struggling to justify their existence encourage state and local governments to use their data. One solution to leverage the federal data and fill the data gap for subnational areas is small area estimation (SAE).

SAE is intended for situations in which a sample suitably designed for one purpose is too small for estimating subdomains or sub-geographies needing estimates. For example, a national sample may not be enough for providing state or county estimates. The key to successful subdomain estimation is to identify auxiliary data that are correlated with the outcomes of interest and available for the entire population. The relationships between the survey outcome variables and auxiliary variables are modeled, and the models are used to generate predictions, either at the unit level or at the domain/area level. By pooling many small areas together in the model, assuming they follow the same model, the estimation of the model parameters is strengthened.

There are challenges in applying SAE methods for production in statistical systems. Czajka, Sukasih, and Queen (2013) identified a few factors including availability of auxiliary variables, computational challenges (software availability and programming

capability), comparative evaluations of alternative small area methods, validation, interpretation, and communication of the results. This paper specifically addresses software and programming capability, demonstrates the work with available software, determines appropriate models, and presents results in a compelling and useful way so that the estimates provide meaningful information to the customers. The goal is to demonstrate the capabilities of SAE for state and local agencies that may be new to SAE.

2. Methodology and Data

2.1. Small Area Estimation Method

Small Area Estimation uses a model or a set of models (either implicit or explicit) to link related areas/domains, time periods, and/or auxiliary data to the variables of interest. A recent, comprehensive guide to SAE is Rao and Molina (2015). Compared to direct estimation[1] with survey data, SAE requires extra steps such as modeling (including variable and model selection), checking distributional assumptions for random effects in the model, and performing parameter estimation that involves iterative computation. Availability of good auxiliary data and determination of suitable model relationships are crucial factors to success with SAE techniques. When good auxiliary data are available with suitable model relationships, SAE techniques confer several advantages over direct estimation, such as the capability of making estimates for small domains with better precision due to "borrowing strength" from neighboring or related areas/domains, time periods, or auxiliary information. One well- known disadvantage is that the estimate is often biased; however, due to better precision/variance, the mean square error (defined as $(bias)^2 + variance$) is often smaller than that of the direct estimate. That is, reduction in the MSE is the main reason for using SAE.

There are many variations of SAE techniques, and they can be grouped into implicit and explicit modeling (Rao and Molina 2015). While modern techniques of small area estimation are usually based on explicit modeling (linear mixed model, hierarchical model, etc.), traditional (implicit) linking models are still used to produce small domain estimates. Estimation with implicit modeling includes:

- synthetic estimation, i.e., the indirect estimator for a small area is derived from a direct estimator for a larger area covering several small areas;

- composite estimation, i.e., the indirect estimator for a small area is computed as a weighted average of an indirect estimator for the small area (can be a synthetic estimator) and a direct estimator for the small area.

Estimation with explicit modeling can be done using one of two model types, depending on the availability of data:

- area level models, i.e., models that relate small area estimates to area-specific auxiliary variables;

- unit level models, i.e., models that relate unit (element) level data to unit-specific auxiliary variables.

---

[1] Direct design-based estimation is the common approach where the statistics of interest are directly computed as weighted aggregations of individual raw data across responding samples.

For each type of explicit models, there is a wide range of model extensions depending on the type of study variables, estimation techniques, assumptions on parameters, etc. Rao and Molina (2015) provide a complete discussion of these models.

Areal-level Fay-Herriot Model for Proportions

As mentioned above, there are two types of SAE models; area-level models and unit-level models. Suppose an analyst is constructing an SAE model using data from a survey. When access to microdata at the individual observational unit level is possible, and when the variable indicating the "small area" is available to the data user, one can utilize these data to build SAE unit-level models. However, when access to microdata is limited to public use data files, usually the variable identifying the small area is protected (suppressed or dropped from the data) due to disclosure avoidance, so constructing unit-level models may not be possible. In addition, limitations on the auxiliary variables available from a particular survey discourage users from building unit-level models. Alternatively, one can build SAE area-level models where auxiliary data/covariates at the area-level can be used as predictors. The use of area-level models expands opportunities to include more auxiliary variables/covariates, which can come from the particular survey at hand or from other external surveys, as long as the covariates can be constructed at the small area level.

A popular area-level model in the application of SAE is the Fay-Herriot (F-H) model (Rao and Molina 2015; Fay and Herriot 1979) where the analyst can model the estimation of an area-level parameter of interest through a regression model using area-level covariates as predictors. The Fay-Herriot model assumes that the variance in the model estimates comes from two sources: sampling variability and error in the model. The models are defined as the combination of a sampling model and linking model as follows:

Sampling model: $\hat{\theta}_i = \theta_i + e_i$

Linking model: $\theta_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + v_i$

for small area $i = 1, \cdots, m$. The term e_i is the sampling error, and the term v_i is the model error. When the sampling and linking models are said to be matched, we can combine the two into a single linear mixed model:

$$\hat{\theta}_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + e_i + v_i.$$

In this paper we demonstrate the SAE F-H model to estimate a proportion π_i for county i using data designed for state-level estimation. Therefore, the small area is defined as a county. The two variance sources are described with the following models. Because the value of the proportion has values between zero and one, the linking model is transformed into logit form:

Sampling model: $p_i = \pi_i + e_i$

Linking model: $logit(\pi_i) = log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + v_i$

where $X_1, X_2, \cdots, X_k$ are the county-level covariates. Because the linking model for π_i is not linear, the sampling model and the linking model are said to be unmatched, so that the

two models cannot be merged into a linear mixed model. The sampling errors and model error terms are all independent.

Our approach used a Hierarchical Bayes model that put a prior distribution on model parameters. The sampling error e_i may be assumed to follow a normal distribution with mean 0 and variance $\varphi_i = 1/S_i$. That is,

$$e_i \sim N(0, \varphi_i)$$

and φ_i is empirically estimated by the direct-estimate of the sampling error variance.

The model error v_i (area-specific random effect) is also assumed to be distributed as normal with mean 0 and variance $\sigma_v^2 = 1/S2$, where we specify a prior distribution on hyperparameter $S2$. That is,

$$v_i \sim N(0, \sigma_v^2)$$

$$\sigma_v^2 = 1/S2$$

$$S2 \sim U(0,20)$$

In addition, we also specify a distribution on the model parameters $\beta_0, \beta_1, \cdots, \beta_k$ as $\boldsymbol{\beta} \sim \boldsymbol{N(0, 1000)}$. These prior distributions were chosen as diffuse proper prior distributions, as this application for proportions with unmatched sampling-linking models would not allow improper priors to generate a proper posterior distribution.

The Hierarchical Bayesian method uses the sample data to iteratively update the model parameters $\Omega = (\beta, v, \sigma_v^2)$.[2] Bayesian inference on π_i and model parameters Ω is based on integrating the joint posterior distribution of $f(\pi_i, \Omega | y)$ for data y. However, this integration may not be easily done. Therefore, we used simulation; i.e., random draws from the posterior distribution to generate R replicates of Markov Chain Monte Carlo (MCMC) samples of $\beta_0^{(r)}, \beta_1^{(r)}, \cdots, \beta_k^{(r)}, v_i^{(r)}$, where $r = 1, \cdots, R$. Then we substituted $\beta_0^{(r)}, \beta_1^{(r)}, \cdots, \beta_k^{(r)}, v_i^{(r)}$ into the linking model above and solved for $\pi_i^{(r)}$. The model-based estimate of π_i^{SAE} is then the (posterior) mean of $\pi_i^{(r)}$ across R replicates.

Due to unmatched sampling and linking models in the application of the Fay-Herriot estimator, the final estimates are calculated as a combination of the direct survey estimate and the model-based prediction, where the importance placed on each component is based on the variance associated with each component. That is, if the sample size is relatively large so that the direct estimate has smaller variance, more of the weight is on the direct

---

[2] In traditional statistical inference, the parameter of interest π is considered as an unknown but fixed quantity. In Bayesian analysis, the parameter of interest π is considered as a quantity that has a variation and is modeled by a probability distribution (called a *prior distribution*). Then a sample is taken, and the data are observed, say Y. The prior distribution is then updated with these data, and it is then called the *posterior distribution*., which is the conditional distribution of π given Y. The Bayesian estimation of π can be done by making statements about π from the posterior distribution. For example, the point estimate of π is computed as the mean/expected value from the posterior distribution.

estimate, but if the model prediction has smaller variance, more of the weight is on the model prediction. If fact, if a county had no sample data, all the weight would be on the model prediction.

2.2. Data

Survey Outcomes

In this paper, we demonstrated this Hierarchical Bayes SAE method for estimating the county-level proportion of adults with health care insurance (insured rates) in the state of New York using data from the New York Behavioral Risk Factor Surveillance Survey (NY-BRFSS).[3] The BRFSS is an annual system of statewide random telephone surveillance surveys designed by the Centers for Disease Control and Prevention. The BRFSS data provide valuable information in support of health care policy decisions, and the BRFSS surveys are intended for estimates at the state and large metropolitan area levels. When direct design-based estimation is used to compute statistics at small area levels such as counties, the statistics may not be sufficiently reliable because the sample was not designed for those levels. Our paper created model-based estimates to provide a more reliable prediction of adult insured rates for NY counties by making use of direct estimates and covariates from other sources with larger sample sizes.

Under a data user agreement with the New York State Department of Health (NYSDOH), we obtained the 2016 NY-BRFSS restricted use file. This data file has a county identification variable and an analysis weight that can be used for direct design-based estimation at the county level. Any analyses, interpretations, or conclusions based on these data are to the responsibility of the authors rather than NYSDOH or BRFSS.

Auxiliary Variables

Graven and Turner (2011), and Bauder et al. (2017) have tested and used many auxiliary variables in small area estimation models of health care coverage rates. It is clear from the literature that health care coverage, or lack thereof, is related to race/ethnicity, age, sex, education level, employment status, usage of public assistance, and various measures of income or poverty. Some of the variables such as sex that are important for predicting insurance coverage for an individual are less meaningful at the county level. These covariates are publicly available, or derived from public sources, most at no cost to the user. We started with the covariates used in those two papers as candidates for our insured rate model. The initial set of auxiliary variables for consideration as model covariates and the source of data (in brackets) are listed below:

X_1 = total adult population (ACS[4] 5-year)
X_2 = percent Hispanic (ACS 5-year)
X_3 = percent AIAN (ACS 5-year)
X_4 = percent household 65 and over (ACS 5-year)
X_5 = percent non-citizen (ACS 5-year)
X_6 = percent adults employed by non-retail firms (County Business Patterns)
X_7 = percent of housing units that are rural (2010 Census)
X_8 = percent of housing units that are resident-owned (ACS 5-year)

---

[3] https://www.health.ny.gov/statistics/brfss/
[4] ACS = American Community Survey (https://www.census.gov/programs-surveys/acs/)

X_9 = percent adults with less than high school (ACS 5-year)
X_{10} = average unemployment rate (Bureau Labor Statistics)
X_{11} = median AGI or median HH income (SAIPE[5] covariate)
X_{12} = ACS-Like Poverty Universe, all age (SAIPE covariate)
X_{13} = SNAP[6] benefit recipients (SAIPE covariate)
X_{14} = per capita income (Bureau of Economic Analysis)

We implemented variable selection methods to select a subset of these variables for the model. With potential correlations among these covariates and the limited number of observations (counties) used to build the model, we selected four variables as predictors as follows: percent non-citizen (X_5), percent of housing units that are resident-owned (X_8), percent adults with less than high school education (X_9), and average unemployment rate (X_{10}).

2.3. Software

For SAE calculations, there are several statistical software packages available, either commercially or as freeware. For our demonstration on Hierarchical Bayes models, we used *R* packages that are freeware.

R Packages

Several packages for SAE are available in *R*, including `Sae`, `Sae2`, `Rsae`, `JoSAE`, `saeRobust`, `hbsae`, `saery`, `mme`, `smallarea`, `saeSim`, `maSAE`, and `BayesSAE`. However, some packages were developed for specific models only. For estimating small area proportions (unmatched sampling and linking models), we used the `BayesSAE` package (Shi 2018) that can handle Hierarchical Bayes models, runs MCMC samples, and has an option for logit transformation (`tran="logit"` option). The code to run the SAE model is very simple ("one line") and does not require high level programming skill. We fit the model by using the function `BayesSAE()` using the default prior distribution. To evaluate this package, we ran a comparison using another freeware called `OpenBUGS`, a popular software package for Bayesian analysis.

OpenBUGS

`OpenBUGS` (http://www.openbugs.net/w/FrontPage) is an open source statistical software package for Bayesian analysis of complex statistical models using MCMC techniques (Robert and Casella 1999). `OpenBUGS` programs were run within *R* using the `R2OpenBUGS` package (Sturtz, Liggesy, and Gelman 2017). Because the use of the Hierarchical Bayes method with MCMC simulation produces replicates of variance components, as well, the software also produces estimates of mean squared errors. `OpenBUGS` also allows more than one chain of simulated data.

The `R2OpenBUGS` package allowed us to run `OpenBUGS` from *R*. the associated coda package contains a suite of functions that can be used to summarize, plot, and diagnose convergence from MCMC samples. It also provides a set of convergence diagnostics such

---

[5] SAIPE = Small Area Income and Poverty Estimates (https://www.census.gov/programs-surveys/saipe.html)

[6] SNAP = Supplemental Nutrition Assistance Program
(https://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program-snap)

as Gelman & Rubin's potential scale reduction factor (Gelman and Rubin 1992). However, the user needs to have substantial knowledge of the BUGS language to specify the models and distributions.

We used 10,000 replicates in MCMC with 5,000 iterations of burn-in. In addition, to reduce the potential autocorrelation in the replicate samples, we sampled from every kth iteration, where $k = 5$.

3. Results

Table 1 shows the comparison of the model-based estimates of proportions of insured adults computed using `BayesSAE` and `OpenBUGS` for all counties in New York, as well as their variance (posterior variance for the Bayesian estimates). For comparison, we also included the direct design-based estimates of proportions of insured adults.

Table 1. Design-based and model-based estimates for adult insured rates for New York counties, 2016

County	Point estimate			Variance*		
	Bayes SAE	Open BUGS	Design-based	Bayes SAE	Open BUGS	Design-based
Albany County	0.928	0.928	0.932	0.00019	0.00018	0.00023
Allegany County	0.924	0.923	0.928	0.00026	0.00027	0.00037
Bronx County	0.825	0.825	0.830	0.00047	0.00046	0.00046
Broome County	0.927	0.926	0.934	0.00031	0.00030	0.00047
Cattaraugus County	0.918	0.918	0.923	0.00028	0.00027	0.00038
Cayuga County	0.900	0.900	0.905	0.00038	0.00038	0.00054
Chautauqua County	0.905	0.905	0.904	0.00023	0.00022	0.00028
Chemung County	0.948	0.948	0.964	0.00017	0.00018	0.00014
Chenango County	0.937	0.937	0.960	0.00021	0.00022	0.00013
Clinton County	0.914	0.914	0.923	0.00022	0.00023	0.00025
Columbia County	0.882	0.884	0.886	0.00049	0.00048	0.00064
Cortland County	0.933	0.932	0.943	0.00034	0.00034	0.00051
Delaware County	0.912	0.912	0.919	0.00032	0.00032	0.00041
Dutchess County	0.908	0.908	0.912	0.00030	0.00030	0.00038
Erie County	0.897	0.897	0.889	0.00031	0.00031	0.00047
Essex County	0.947	0.947	0.965	0.00015	0.00016	0.00011
Franklin County	0.880	0.879	0.879	0.00067	0.00067	0.00115
Fulton County	0.880	0.880	0.869	0.00058	0.00058	0.00106
Genesee County	0.924	0.924	0.935	0.00029	0.00029	0.00036
Greene County	0.857	0.857	0.844	0.00094	0.00097	0.00191
Hamilton County	0.900	0.898	0.909	0.00082	0.00085	0.00164
Herkimer County	0.922	0.921	0.924	0.00027	0.00026	0.00039
Jefferson County	0.913	0.912	0.898	0.00036	0.00036	0.00070
Kings County	0.838	0.838	0.836	0.00021	0.00021	0.00023
Lewis County	0.901	0.900	0.878	0.00057	0.00061	0.00133
Livingston County	0.893	0.894	0.889	0.00064	0.00060	0.00128
Madison County	0.906	0.907	0.900	0.00063	0.00064	0.00170
Monroe County	0.923	0.923	0.929	0.00020	0.00022	0.00025
Montgomery County	0.873	0.873	0.868	0.00068	0.00068	0.00115
Nassau County	0.900	0.900	0.904	0.00029	0.00029	0.00034
New York County	0.870	0.870	0.864	0.00022	0.00023	0.00026
Niagara County	0.914	0.914	0.903	0.00045	0.00046	0.00104
Oneida County	0.902	0.902	0.908	0.00040	0.00042	0.00059
Onondaga County	0.891	0.891	0.886	0.00028	0.00028	0.00037
Ontario County	0.909	0.910	0.905	0.00042	0.00041	0.00078

County	Point estimate			Variance*		
	Bayes SAE	Open BUGS	Design-based	Bayes SAE	Open BUGS	Design-based
Orange County	0.901	0.901	0.904	0.00026	0.00026	0.00033
Orleans County	0.854	0.853	0.827	0.00082	0.00089	0.00171
Oswego County	0.910	0.910	0.914	0.00038	0.00037	0.00058
Otsego County	0.885	0.884	0.867	0.00057	0.00057	0.00112
Putnam County	0.897	0.897	0.893	0.00045	0.00044	0.00069
Queens County	0.861	0.860	0.865	0.00023	0.00023	0.00023
Rensselaer County	0.947	0.948	0.967	0.00017	0.00017	0.00012
Richmond County	0.894	0.893	0.888	0.00046	0.00047	0.00075
Rockland County	0.899	0.898	0.906	0.00026	0.00026	0.00028
St. Lawrence County	0.948	0.948	0.972	0.00022	0.00021	0.00013
Saratoga County	0.931	0.932	0.938	0.00020	0.00020	0.00025
Schenectady County	0.912	0.912	0.917	0.00029	0.00028	0.00037
Schoharie County	0.893	0.893	0.882	0.00070	0.00067	0.00164
Schuyler County	0.886	0.885	0.844	0.00077	0.00076	0.00226
Seneca County	0.861	0.861	0.859	0.00063	0.00063	0.00090
Steuben County	0.914	0.913	0.911	0.00037	0.00040	0.00067
Suffolk County	0.872	0.872	0.866	0.00045	0.00045	0.00064
Sullivan County	0.890	0.890	0.899	0.00050	0.00050	0.00066
Tioga County	0.918	0.918	0.927	0.00038	0.00038	0.00057
Tompkins County	0.911	0.911	0.893	0.00045	0.00045	0.00098
Ulster County	0.913	0.914	0.918	0.00022	0.00021	0.00025
Warren County	0.927	0.927	0.931	0.00019	0.00020	0.00026
Washington County	0.918	0.918	0.938	0.00038	0.00037	0.00041
Wayne County	0.887	0.887	0.877	0.00046	0.00047	0.00078
Westchester County	0.880	0.880	0.882	0.00038	0.00039	0.00048
Wyoming County	0.902	0.902	0.906	0.00043	0.00043	0.00065
Yates County	0.648	0.648	0.584	0.00151	0.00151	0.00179

*Posterior variance for Bayes estimates

`BayesSAE` produced results that are very close to those from `OpenBUGS`. The differences of estimates of proportions are trivial (the largest difference is 0.002). Furthermore, both `BayesSAE` and `OpenBUGS` produced the same or very close posterior variances. We are pleased with these results because computational coding in `BayesSAE` is much simpler than in `OpenBUGS`.

Comparing model-based estimates with the direct design-based estimates, in Figure 1 we plotted the ratio of the direct point estimates over the `BayesSAE` model-based point estimates[7] against the sample size as another way to evaluate the model. The ratio should converge to one, as large counties do not need to borrow strength. The plot shows the convergence of the estimates and the gradual decrease in borrowing strength as the sample size gets larger. The same pattern is observed for the model-based estimates using the `OpenBUGS` package.

---

[7] Given that the estimates of proportions and posterior variances obtained through `OpenBUGS` and `BayesSAE` are almost identical, the comparison of the model-based estimates with the direct design-based estimates were done using the results from `BayesSAE`.
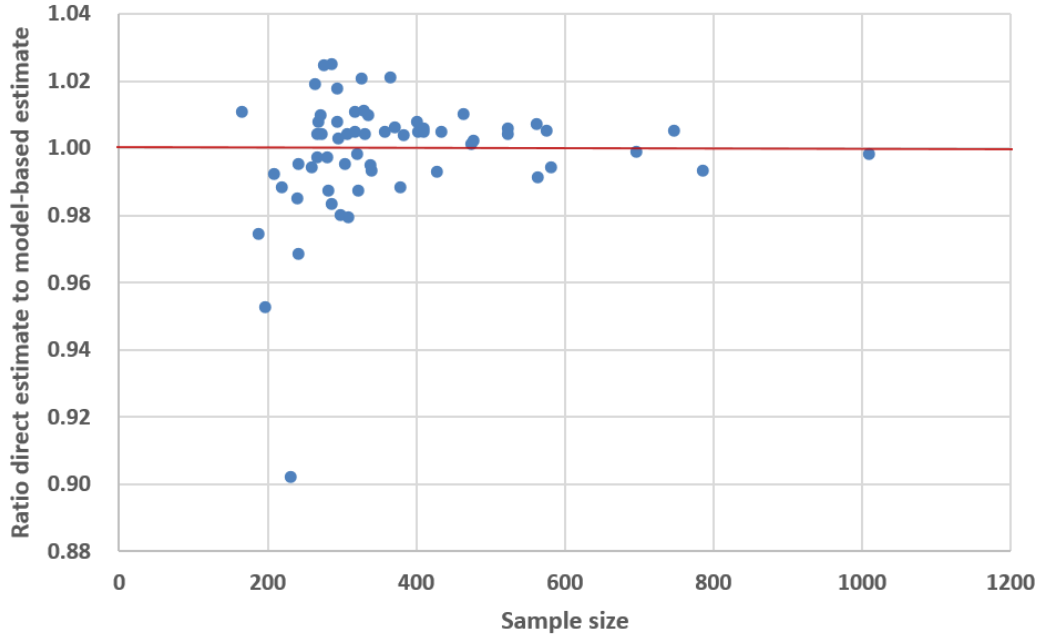
Figure 1. Ratio of design-based to model-based point estimates for county insured rates in New York, 2016.
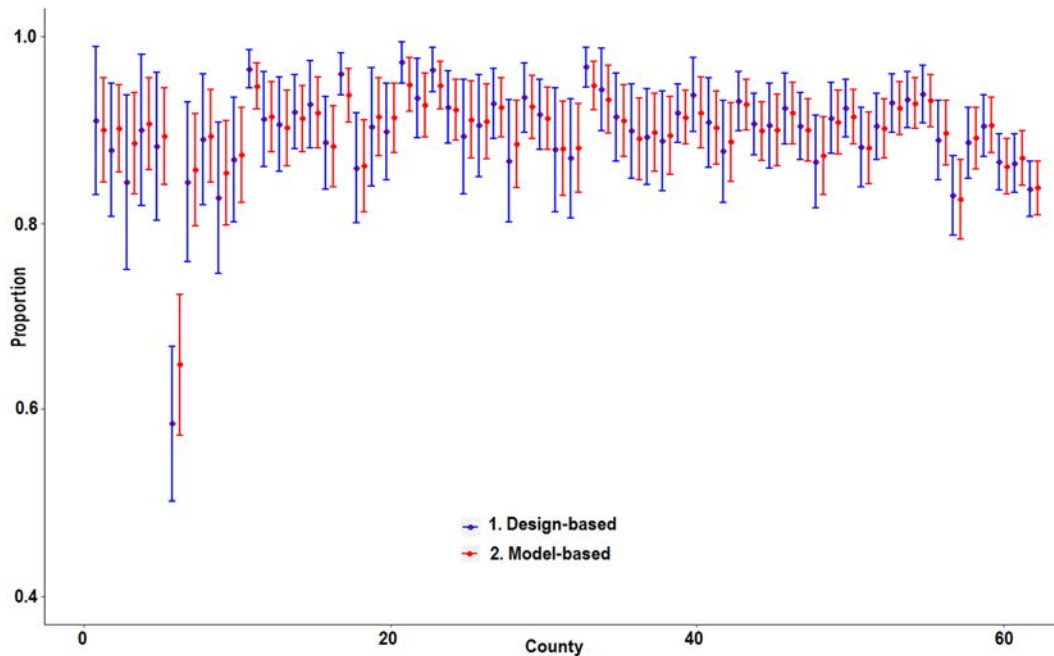

Figure 2. `BayesSAE` model-based vs. design-based 95% confidence intervals for county-level adult insured rates in New York, 2016

Figure 2 shows the 95% confidence intervals and Bayesian credible intervals for estimated insured rates for each county in New York, with the estimates given on the Y-axis and counties ordered by increasing sample size on the X-axis. The solid blue lines show confidence intervals for direct estimates while the red lines show credible intervals for

model-based estimates using the `BayesSAE` function for the corresponding counties. Because the adult insured rates were transformed using the logit transformation, the confidence intervals for the model-based estimates were calculated on the logit scale and transformed back to the original scale.

We observed that the model-based SAE and design-based estimates did not differ greatly (except for Yates County).[8] We noticed that the sample sizes in all these counties are not small, ranging from 166 to 1,009 with an average of 369 cases. With the design-based point estimates ranging from 0.58 to 0.97, and with the coefficients of variation (CVs) of the design-based estimates all less than 10%, the counties may not be considered as "small areas". However, we noticed from Figure 2 that when the sample size is smaller, there is a reduction in variance when comparing design-based vs. model-based estimates.

Figure 3 shows the ratio of variances of proportion based on the direct design-based estimation to those based on the model-based estimation against the sample size on the X-axis. For most of the counties, especially for counties with smaller sample sizes, the model-based estimates have an advantage of decreasing the variance over the direct-design-based estimates, even though the actual magnitude of the reduction may not be large. We can observe that as the sample size becomes larger, such advantage diminishes.
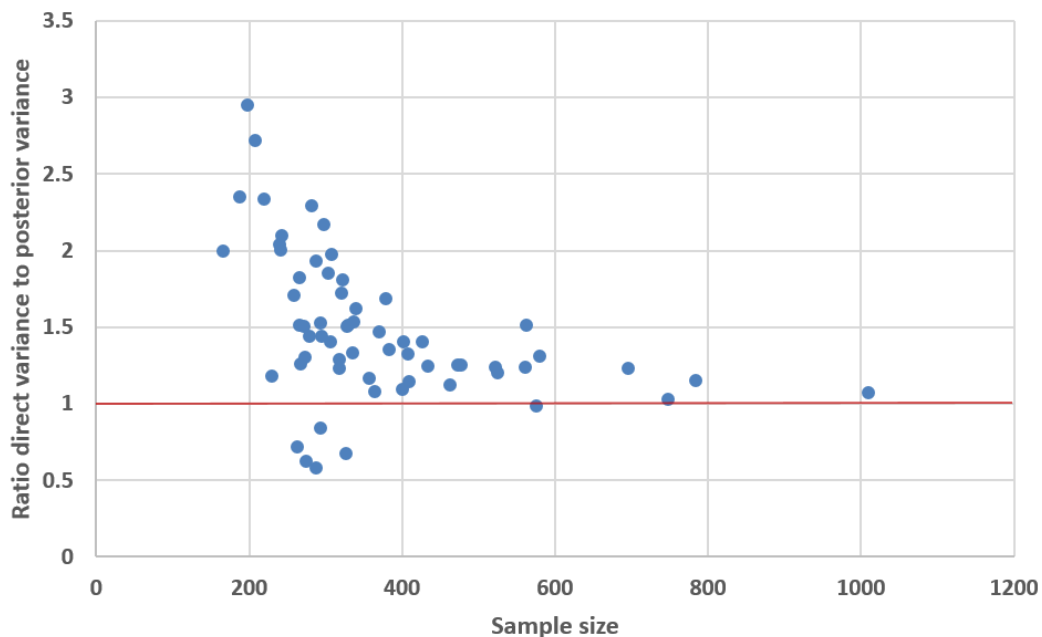

Figure 3. Ratio of design-based to model-based variance estimates for county insured rates in New York, 2016.

As part of our demonstration for state and local agencies, we summarized the county insured rates in an interactive map. Figure 4 shows a static version of one such map in which the county insured rates are highlighted in ranges.

---

[8] We did not investigate further data in Yates county. However, our estimates of adult insured rate in Yates county are lower than those from other data sources (the American Community Survey and the Small Area Health Insurance Estimates Program).
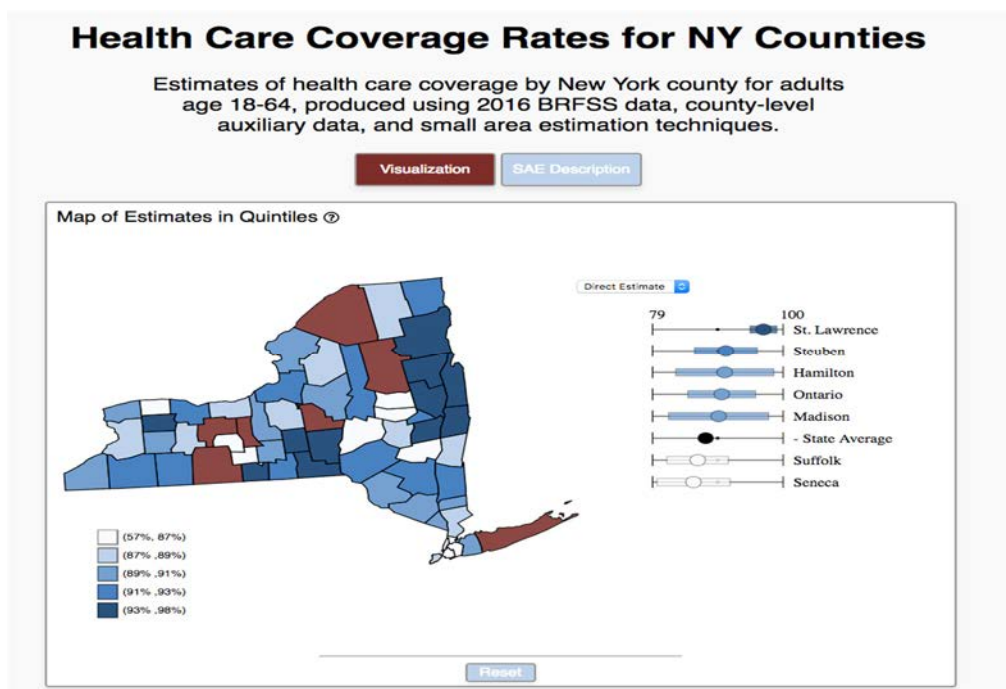
Figure 4. Visualization of the county insured rates in New York, 2016.

4. Conclusion and Discussion

In this paper, we explored the performance and capabilities of the *R* package `BayesSAE` and `OpenBUGS` to perform Bayesian estimation using the Fay-Herriot Hierarchical Bayes approach for small area estimation of proportions where the sampling and linking models are unmatched. We implemented these methods in estimating county-level estimates of adult insured rates in the state of New York. Both packages `BayesSAE` and `OpenBUGS` performed similarly in terms of point estimates and posterior variances. The differences were trivial.

The *R* package `BayesSAE` makes the application of SAE methods almost as simple as fitting a linear regression model, reducing the gap between theoretical advances and their application. One advantage about this package is that it allows for transformations to be taken on the response variable in the linking model. However, the package lacks evaluation and diagnostic tools and has limited prior distribution choices.

In this demonstration, from looking at the lengths of the Bayes credible intervals and the design-based confidence intervals, we may not see an obvious advantage of SAE methods for county-level estimation of insured rates using BRFSS data in NY counties. The relatively large BRFSS sample sizes may account for the strong showing of the direct estimates. Even so, the small area estimates performed well and demonstrated an option for estimation where sample sizes are not so robust.

Note that this paper was limited only to comparisons of estimates of proportions as the statistics of interest, and to area-level models. This work may be extended to other types of statistics and to unit-level modeling. For other types of statistics, we may also explore

and compare other available software packages such as JAGS, an additional off-the-shelf package for small area estimation.

References

Bauder, M., Luery, D., and Szelepka, S. (2017). Small area estimation of health insurance coverage in 2010-2015, *Small Area Methods Branch, Social, Economic, and Housing Statistics Division*, U.S. Census Bureau.

Czajka, J.L., Sukasih, A., and Queen, S. (2013). Small area estimation: New developments and directions for Health and Human Services data. *Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference*. Washington, DC: FCSM.

Fay III, 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*. 74:366a, 269-277.

Gelman, A., and D. B. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 7: 457–511.

Graven, P., and Turner, J. (2011). Modeling health insurance coverage estimates for Minnesota counties. *Proceeding of Joint Statistical Meeting, Section on Survey Research Methods*. pp. 5264-5277.

Liu, B. and Gilary, A. (2014). Small Area Estimation for the Tobacco-Use Supplement to the Current Population Survey. *Proceeding of Joint Statistical Meeting, Section on Survey Research Methods*. pp. 2542-2552.

Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation*. Second Edition. Hoboken, NJ: John Wiley & Sons.

Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.

Shi, C. (2018). Package 'BayesSAE'. Downloaded from:
https://cran.r-project.org/web/packages/BayesSAE/BayesSAE.pdf

Sturtz, S., Liggesy, U., Gelman, A. (2017). R2OpenBUGS: A package for running OpenBUGS from R. Downloaded from: https://cran.r-project.org/web/packages/R2OpenBUGS/vignettes/R2OpenBUGS.pdf