INCORPORATING DESIGN WEIGHTS AND HISTORICAL DATA INTO MODEL-BASED SMALL-AREA ESTIMATION

Hui Xie[1*], Lawrence E. Barker[2] and Deborah B. Rolka[1]

[1] *Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Diabetes Translation, Atlanta, Georgia, USA.*

[2] *Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office of the Director, Atlanta, Georgia, USA.*

ABSTRACT

Bayesian hierarchical regression (BHR) is often used in small area estimation (SAE). BHR conditions on the samples. Therefore, when data are from a complex sample survey, neither survey sampling design nor survey weights are used. This can introduce bias and/or cause large variance. Further, if non-informative priors are used, BHR often requires the combination of multiple years of data to produce sample sizes that yield adequate precision; this can result in poor timeliness and can obscure trends. To address bias and variance, we propose a design assisted model-based approach for SAE by integrating adjusted sample weights. To address timeliness, we use historical data to define informative priors (power prior); this allows estimates to be derived from a single year of data. Using American Community Survey data for validation, we applied the proposed method to Behavioral Risk Factor Surveillance System data. We estimated the prevalence of disability for all U.S. counties. We show that our method can produce estimates that are both more timely than those arising from widely-used alternatives and are closer to ACS' direct estimates, particularly for low-data counties. Our method can be generalized to estimate the county-level prevalence of other health related measurements.

* Corresponding author
email: yum0@cdc.gov

1. Introduction

National and state-level surveys are crucial to public health surveillance in the United States. These are typically designed to provide direct estimates. However, direct estimates are, due to small sample sizes, often impractical at the county level (a unit of local government in the United States). Some U.S. states use other names for these units of government, such as 'parish'. Since they function like counties, we will call them all counties.

Small area estimation (SAE), through modeling, provides estimates for counties that do not have large enough samples for direct estimation. In brief, by borrowing "strength" from the entire domain (i.e., whole counties across state-wide or/and nation-wide) as well as auxiliary variables from other survey studies (Erciulescu, A., 2019), county level estimates can often be derived (Ghosh, M. and Rao, J.N.K., 1994).

Frequentist model-based methods, such as those of Das, K. et al., (2004) and Pierannunzi C. et al., (2016) can provide SAE-based estimates for more counties than can direct estimation. However, such methods can also fail for counties with small sample sizes. Bayesian hierarchical regression (BHR) models can provide estimates for counties with small sample sizes (Planck, N. R. V. et al., 2017). However, BHR methods typically condition on the samples and the parameters of interest. That is, the sampling design and survey weights are not used (Pfeffermann, D., 2013). Thus, BHR can be vulnerable to model misspecification, resulting in both bias and large variance (Kish, L. and Frankel, M., 1974; Hansen, M. et al., 1983).

Additionally, BHR requires specification of a prior. Specifying an informative prior often requires more information than is readily available. Further, informative priors are inherently subjective. Non-informative priors avoid this difficulty. However, for small sample size counties, the posterior distributions are heavily influenced by the subjective priors. Avic (2017) suggested obtaining an informative prior, when sample sizes are small, via meta-analysis of multiple published data sets.

The power prior (Ibrahim, J. G. et al., 2015) is a compromise between informative and uninformative priors – it uses historical data (if available) to modify an uninformative prior to make it more 'informative'. In brief, the power prior is derived by raising the historical likelihood to a power $\alpha_0 \in (0,1)$ and combining it with the uninformative 'flat' prior. This results in a proper posterior distribution. Many studies (Neelon, B. et al., 2010; De Santis, F., 2007; Congdon, P., 2008) have shown that the power prior can improve the precision and accuracy of posterior estimates.

Here, we propose an approach, which we call the power prior log-weights estimates (PLOW). The PLOW incorporates adjusted design-based sampling weights and uses a power prior. As an example, we apply the PLOW to obtain estimates of county level prevalence of impaired vision from 2015 Behavioral Risk Factor Surveillance System (BRFSS) data. We provide estimates for all 3142 counties in the United States.We validate the estimates using both simulations and data from the American Community Survey (ACS), a survey large enough to provide direct estimates for many counties.

2. Survey Data

Behavioral Risk Factor Surveillance System (BRFSS) is an annual state-level telephone surveillance system conducted by the Centers for Disease Control Prevention (CDC). It collects data on risk behaviors, preventive health practices and health-related conditions in the non-institutionalized adult household population with ages 18 years and older. Details on BRFSS have been previously published (Cadwell, B.L. et al., 2010). ACS, conducted by the US Census Bureau, is also an ongoing annual survey, using internet, mail, telephone, and person visits to collect data (Gettens, J. et al., 2015). ACS has more than eight times as many respondents as BRFSS.

In 2013, five survey questions concerning disability were used in common by both BRFSS and ACS. Table 1 displays those disabilities and the definitions used in ACS and survey questions in BRFSS. Possible responses for both surveys were yes, no, and no answer.

Table 1. The descriptions of five disabilities listed in ACS and BRFSS

Disability	ACS Disability Definitions	BRFSS Disability Questions
Vision	Blind or having serious difficulty seeing, even when wearing glasses	Are you blind or do you have serious difficulty seeing, even when wearing glasses?
Cognitive	Because of a physical, mental, or emotional problem, having difficulty remembering, concentrating, or making decisions	Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions?
Ambulatory	Having serious difficulty walking or climbing stairs	Do you have serious difficulty walking or climbing stairs?
Self-care	Having difficulty bathing or dressing	Do you have difficulty dressing or bathing?
Independent	Difficulty because of a physical, mental, or emotional problem, having difficulty doing errands alone such as visiting a doctor's office or shopping	Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a doctor's office or shopping?

3. Methods

3.1 Notation

Auxiliary variables are variables other than the variable of interest used to construct models for SAE. We use the county-level covariates age, sex and race/ethnicity as auxiliary variables. Similar to Barker, L. E. et al. (2013), sampled persons in each county are cross-classified by age-groups (20-44, 45-64, 65 years plus), sex (male and female) and race/ethnicity (non-Hispanic white and all others; in some counties, numbers of people other than non-Hispanic whites were too small to stratify further). This resulted in 12 classes.

Let y_{ijk} be the binary disability outcome of survey participant kth in the county i (k=1,…, n_{ij}, i =1 to 3142) and class j (k =1 to 12), where n_{ij} is the sample size of county i and class j.

We denote $y_{ij} = \sum_{j=1}^{n_i} y_{ijk}$ as the total case counts and N_{ij} as the total population size in county i and class j, respectively. The BRFSS uses design weights and raking weights. We use w_{ijk} to denote the sampling raking weight attached to kth sample in county i and class j. County-level data y_{ijk} and w_{ijk} were derived from 2015 BRFSS while N_{ij} were from Census Bureau Center county-level population projections (Barker, L. E. et al., 2013).

3.2 Adjustment of sampling weights

The most common design-based inference for the SAE is the Horvitz-Thompson (HT) estimator (Horvitz, D.G. and Thompson, D.J., 1952). Let $\hat{p}_{ij}$ be the direct estimate probability in county i and class j. The HT estimate is $\hat{p}_{ij} = \sum_{k=1}^{n_{ij}} w_{ijk} y_{ijk} / \sum_{k=1}^{n_{ij}} w_{ijk}$. The HT estimate is impossible in counties with no survey respondents, and yield large variance in counties with small sample sizes. If the sampling weights have a distribution with a heavy right tail, some estimates can have very large variances (Beaumont, J. and Rivest, L., 2009). As can be seen in Table 2, BRFSS sampling weights have a heavy right tail (the 'adjusted' sampling weights are explained later).

Table 2. The distribution of BRFSS sampling weights and the adjusted sampling weights (N = 426218). GF= $\max W_{ijk} / \min W_{ijk}$ is the Gelman Factor, a measure of the dispersion of sampling weights.

Weighs	min	max	median	skewness	GF
Raw	1.18	36700	231.28	5.74	31102
Adjusted	1.24	1612	68.16	2.49	1300

Meng, X. L. et al. (2010) presented a power transformation to deal with weights with right tails. However, their approach did not work well with the most extreme weights. The trade-off of bias-variance relied on correlation between variable and weights (Chen, C. N. et al., 2006). Instead of what was done earlier, we propose a "log-weight" transformed method to adjust the sampling weights. Here, we use τ as an index of transformation in the range 0 to 1, in other words, $\tau \in [0.1]$, is a tuning parameter; in Tukey's ladder of transformations (Tukey, J. W., 1977), a logarithmic transformation corresponds to an asymptotically zero exponent. We let $\hat{p}_{ij}^{\tau}$ denote the weighted direct estimate probability. This modifies the HT estimator as:

$$\widehat{p}_{ij}^{\tau} = \begin{cases} \dfrac{\sum y_{ijk}}{N_{ij}} & if\ \tau = 0 \\ \dfrac{\sum (log(w_{ijk})^{\tau}) y_{ijk}}{\sum log(w_{ijk})^{\tau}} & if\ 0 < \tau \leq 1 \end{cases} \tag{3.1}$$

In particular, τ=0 corresponds to the unweighted adjustment while τ=1 to the fully-weighted adjustment. To target the "optimal" τ value, we calculated the mean squared error (MSE) between $\hat{p}_{ij}$ and $\hat{p}_{ij}^{\tau}$ as follows (Cox, B. G. and McGrath, D. S., 1981):

$$\widehat{MSE}(\hat{p}_{ij}, \hat{p}_{ij}^\tau) = (\hat{p}_{ij} - \hat{p}_{ij}^\tau)^2 + var(\hat{p}_{ij}^\tau)$$
$$= (\hat{p}_{ij} - \hat{p}_{ij}^\tau)^2 + \frac{1}{n_{ij}}(1 - \frac{n_{ij}}{N_{ij}})\frac{1}{n_{ij}-1}\sum w_{ij}^2(y_{ijk} - \hat{p}_{ij}^\tau)^2$$

In general, the bias $(\hat{p}_{ij} - \hat{p}_{ij}^\tau)^2$ increases and $var(\hat{p}_{ij}^\tau)$ decreases as τ increases, and vice versa (Appendix. Figure 1). The "optimal" τ is then selected by the cutoff associated with minimal MSE value (Potter, F.J., 1988). Using the "optimal" cutoff of τ, we can calculate the "effective" number of cases, named as y_{ij}^e , as the product of raw sampling cases and the weighted probability (3.1):

$$y_{ij}^e = y_{ij} \times \hat{p}_{ij}^\tau \tag{3.2}$$

3.3 Power prior using historical survey data

When historical data are available, it is possible to 'borrow' strength from it. Let $\alpha_0 \in (0,1)$ be the power parameter (defined later) which controls how much the historical data impacts the prior. Let $\pi_0(p_{ij})$ be the prior for p_{ij} from before the historical data are observed (Ibrahim, J. G. and Chen, M.H., 2000).We denote the historical data Y_0 and the likelihood function of Y_0 is L(p|Y_0). The power prior for p_{ij} is then defined as:

$$\pi(p_{ij}|Y_0) \propto L(p_{ij}|Y_0)^{\alpha_0}\pi_0(p_{ij})$$

From Bayes' theorem, the posterior distribution of $p_{ij}|y_i^e, Y_0$ can be re-written as:

$$f(p_{ij}|y_i^e, Y_0) = \frac{L(y_i^e|p_{ij})\pi(p_{ij}|Y_0)}{\int L(y_i^e|p_{ij})\pi(p_{ij}|Y_0)d(p_{ij}|Y_0)}$$

$$\propto L(y_i^e|p_{ij})\pi(p_{ij}|Y_0) = L(y_i^e|p_{ij})\frac{L(Y_0|p_{ij})^{\alpha_0}\pi_0(p_{ij})}{\int L(Y_0|p_{ij})^{\alpha_0}\pi_0(p_{ij})d(p_{ij})}$$

$$\propto L(y_i^e|p_{ij})L(Y_0|p_{ij})^{\alpha_0}\pi_0(p_{ij}) \tag{3.3}$$

For notational convenience, we combine $L(y_i^e|p_{ij})L(Y_0|p_{ij})^{\alpha_0}$ into $L'(y_i^e, Y_0|p_{ij})$ and call it the "power likelihood".

3.4 Estimating disability prevalence using the PLOW approach

We employ the BHR model for county-level disability prevalence estimates. Let p_{ij} be the prevalence of disability in county i and class j. The "effective" case count y_{ij}^e (3.2) is assumed to follow a binomial distribution:

$$y_{ij}^e|p_i \sim Binomial(n_{ij}, p_{ij}).$$

Let α_{ij} be the overall mean effect for county i and class j; X^T be the vector of auxiliary variables; and β_{ij} be the associated vector of fixed effect in county i and class j. Under this framework, $U_i|\sigma_i^2 \sim (0, \sigma_i^2)$ and $V_{s(i)}|\sigma_{s(i)}^2 \sim (0, \sigma_{s(i)}^2)$ are county-specific and state-specific random effects, respectively. We assume U_i and $V_{s(i)}$ are independent. Borrowing "strength" (Rao, J. N. K. and Molina I., 2015) from county and state variabilities, the logit function of p_{ij} in county i and class j can be modeled as:

$$logit(p_{ij}) = \alpha_i + X^T\beta_{ij} + U_i + V_{s(i)}.$$

We apply the power prior, as described earlier. For example, to estimate the prevalence of disability in 2015, we use 2015 BRFSS data (most current available) and 2013 and 2014 BRFSS data (historical). We integrate them with current data through use of the power prior (3.3). Let p_i be the model-based estimated prevalence of disability in county i. Then the estimated prevalence in county i is:

$$p_i = \frac{\sum_{j=1}^{12} p_{ij} N_{ij}}{\sum_{j=1}^{12} N_{ij}}.$$

3.5 Validation of PLOW estimates

The ACS releases single-year disability data for the 835 large counties (population>65000). For those counties, validation is done via direct comparison of HT (hereafter, direct estimates), PLOW estimates, and the unweighted, non-informative prior estimates of Cadwell et al. (2010) (hereafter, 'Cadwell estimates').

For counties for which single year data are not released, direct comparison is not possible. For those counties, we validate through a simulation. We create fictional small sample size 'pseudo-counties' by sampling from actual counties for which ACS data are available. That is, we:

(1) Randomly select 200 counties from 835 counties covered by ACS 1-year data
(2) Randomly select 1%~5% survey samples from each selected county to form pseudo-counties; some pseudo-counties may have no data
(3) Apply PLOW to estimate the prevalence of disability of each pseudo-county
(4) Repeat steps 1 to 3 for 500 times and average the estimates
(5) Validate simulation results with the real ACS 1-year results for each county

4. Results

4.1 Shrinkage of sampling weights

Weightings heavily influence the estimates, particularly in small population counties where the weights can be large. The weights tend to have a highly right-skewed distribution. To limit this effect, we rescale weights by a logarithmic transformation. Table 2 describes the distribution of sampling weights before and after transformation.   The transformation

substantially reduced the skewness of the weights. The Gelman Factor (Burgard, et al. 2014) is the ratio of the largest to smallest sampling weights, which reflects the dispersion of weights. In general, the Gelman Factor of survey design should not exceed 1000 (Meng, X. L. et al., 2010). With the adjustment, the Gelman factor is reduced by over 95%.

4.2 Validation with ACS for large counties

Table 3 presents the MSE of the disability questions for direct estimates, Cadwell estimates, and PLOW estimates. For all of these, the MSE of the Cadwell estimate is smaller than it is for the direct estimate. Similarly, the MSE of the PLOW estimate is smaller (sometimes much smaller) than that of the Cadwell estimate.

Table 3. Mean squared error (MSE) of HT direct estimates, unweighted model estimates and new approach (weighted model estimates) for estimation of disabilities (Vision, Cognitive, Ambulatory, Self-care and Independent living) prevalence

Method	Mean Square Error (MSE)$\times 10^4$ (cases/10000 persons)				
	Vision Difficulty	Cognitive Difficulty	Ambulatory Difficulty	Self-care Difficulty	Independent living
HT direct estimates	7.12	34.65	43.53	3.36	6.32
unweighted model estimates (Cadwell estimates)	3.52	26.17	34.80	1.52	3.25
weighted model estimates (PLOW estimates)	1.02	2.61	4.05	0.88	2.07

Validation, expressed as scatter plots, appears in Figure 1. There, we plot ACS direct estimates (x-axis) vs BRFSS direct estimates, Cadwell estimates, and PLOW estimates (y-axes). Were all estimates unbiased, one would expect to see random scatter around the diagonals. A systematic departure indicates bias. Figure 1 suggests that direct and Cadwell estimates both have a positive bias, compared with ACS estimates. In contrast, the PLOW estimates are generally similar to ACS estimates. This finding indicates that inappropriately using sampling weights (direct estimate) or not using sampling weights (Cadwell estimates) can introduce bias.
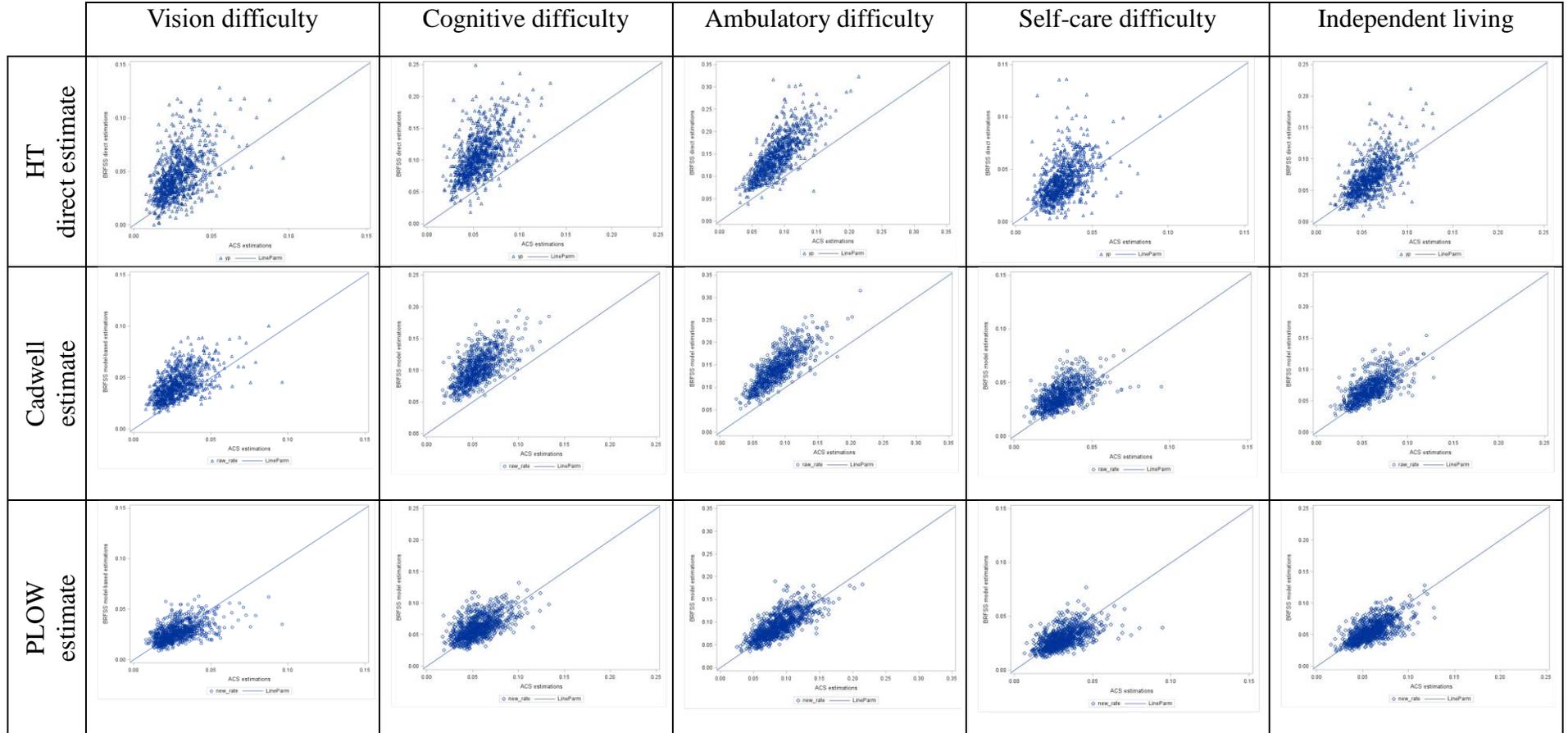
Figure1. Scatter plots showing validation of the disability (Vision, Cognitive, Ambulatory, Self-care and Independent living) prevalence estimates through HT direct estimate, Cadwell estimate and PLOW estimate using ACS 1-year results for "large" counties

4.3 Validation with ACS 5-year county-level and ACS state-level data

Next, we plot PLOW results against the ACS 5-year county-level data. ACS 5-year county-level data were collected in the range of five years, e.g., survey data from 2011 to 2015 were aggregated as the 2015 reports for all 3142 counties. Based on the Census Bureau guideline, we classify 3142 counties into three groups: large size with population>65000, ("large", hereafter), medium size with population>20000 and <65000 ("medium", hereafter) and small size with population<20000 ("small", hereafter). Using Cognitive Difficulty for the example, Figure 2 shows the agreement of results between PLOW and ACS 5-year decreases from "large" to "small" counties.
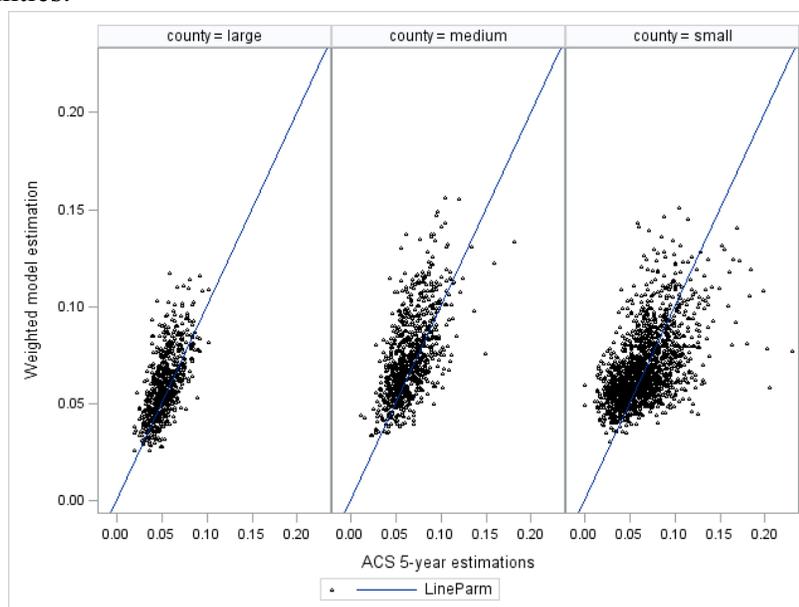

Figure 2. Scatter plots of prevalence estimates of Cognitive Difficulty using PLOW against ACS 5-year county-level results by county size ("large", "medium" and "small")

Additionally, we compared the results of the three methods at the state-level. We aggregated the county-level estimations into state-level and compared them with ACS state-level results. Figure 2 (suppl.) presents the state-level Cognitive Difficulty prevalence with PLOW and the other two methods. PLOW estimates are closer to ACS estimates than are direct and Cadwell estimates.

4.4 Validation for small counties

Figure 3 displays estimates in pseudo-counties of prevalence of cognitive difficulty using PLOW (red triangle), BRFSS direct (black dot) and Cadwell estimates (open square); results for other disabilities are similar and appear in Figure 3 (suppl.) for reference. The PLOW estimates have less dispersion and lower bias.
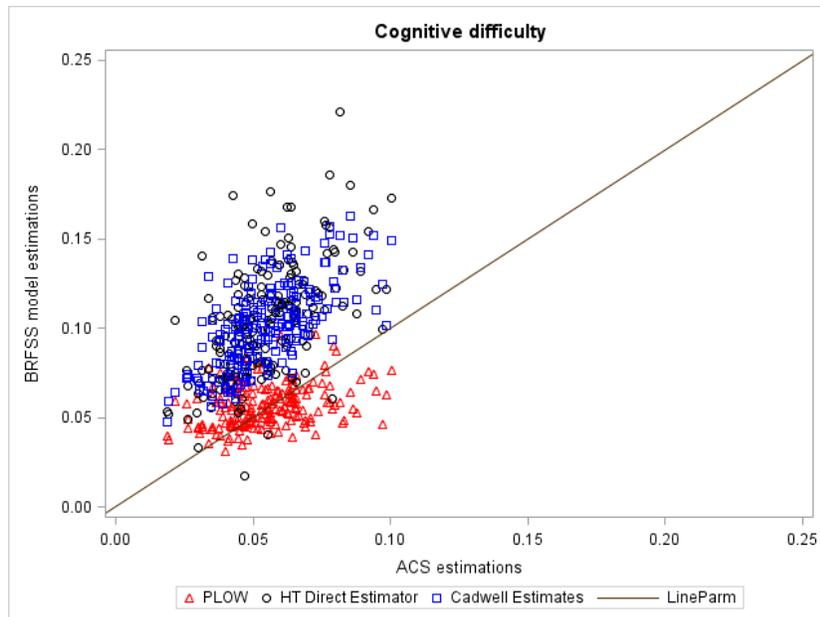
Figure 3. Validation plot for comparison of results of ACS 1-year data with results of HT direct estimator (black circle), Cadwell estimates (black square) and PLOW (red triangle) for the small sizes counties using "pseudo-county" data. The diagonal reference line (dash line) represents the estimates results are same to ACS 1-year county-level results

4.5 Power prior vs flat prior

To compare the power prior with flat prior implemented in the Bayesian hierarchical model, we applied PLOW to estimate prevalence of Vision Difficulty with these two types of prior, respectively, using BRFSS 2015 data. The difference is that we use BRFSS 2013 and BRFSS 2014 data as "historical" data in power prior modeling, whereas we chose the inverse Gamma distribution as the non-informative prior in the flat prior modeling. Figure 4 describes the distribution of prevalence of 3142 counties modeling with power and flat prior, respectively. The flat prior tends to generate larger estimates, with more outliers, than the power prior in each quantile, accordingly: 0.023 vs. 0.016 (1st quartile), 0.027 vs. 0.019 (median) and 0.033 vs. 0.023 (3rd quartile).
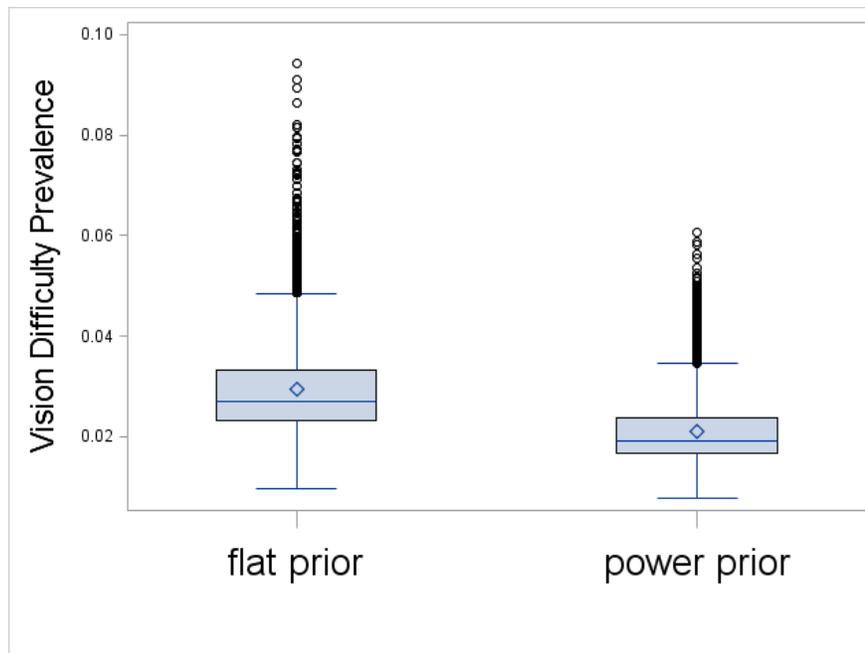
Figure 4. Box-plot comparison of Bayesian hierarchical model implemented with power prior and flat prior to estimate the Vision Difficulty prevalence

5. Discussion

Previous work has been done on SAE for county-level prevalence estimates using BRFSS survey data. Cadwell, et al. (2010) used area-level Bayesian hierarchical regression model to estimate the diabetes prevalence rates in 3141 counties. Zhang, et al. (2011) estimated county-level obesity prevalence in Mississippi using multilevel logistic model and synthetic estimation techniques. The model borrowed the "strength" from the county-specific demographic variables, geographic and socioeconomic variables and auxiliary random effect from counties. Other examples are found in Xie, et al. (2007), Goodman (2010), Olives (2013) and Schneider (2009).

We have introduced PLOW, a design assisted model-based approach, and used it to estimate county-level disabilities prevalence. This approach has several advantages over model based estimates that do not use weights and design based estimates that are directly derived from weights. We demonstrate that PLOW can produce estimates with less bias and variance than other approaches.

First, PLOW calculates "effective" disability case counts using the adjusted sampling weights before applying the Bayesian hierarchical model. Smaller skewness for the weights reduces bias and variance in the downstream model-based estimates (Beaumont, J. and Rivest, L., 2009). Secondly, PLOW uses a power prior, which shows promise in counties with little or no data. Estimates based on non-informative priors tend to be close to each other and, for examples consider here, were much higher than ACS reports. Inverse gamma is widely used as a flat prior for the unknown variance of normal distribution. Gelmen, A. (2006) demonstrated the inverse gamma distribution with small shape and scale produced large

reductions of parameter space, especially for a domain with few observations. Informative priors require prior elicitation and are inherently subjective. Power priors, while remaining objective, avoid prior elicitation by borrowing strength from historical data.

Perhaps most importantly, PLOW makes possible annual county-level prevalence estimates for all counties, using single-year data. Other methods are typically limited to counties with larger samples sizes or combine multiple years of data. Combining multiple years of data obscures secular trends and makes timely detection of rapid changes difficult or impossible. For example, Cadwell estimates, used by CDC for estimating county level prevalence of diabetes, obesity, and regular physical activity (Gregg, E.W., et al., 2009), combines three years of data. This introduces bias if there are secular trends in prevalence. ACS states "single-year and multiple-year estimates are not expected to be the same" because in counties with smaller population, small changes in numbers of people who are members of a given demographic class can result in large percentage changes.

In short, PLOW, when applied to survey data for which historical data are available, can provide prevalence estimates that are both more useful and more timely. While we have provided estimates for disability prevalence, the PLOW method does not depend on any characteristic that is unique to disability. It can be used to provide county-level estimates of any measure with historical data.

References

[1]   Avci, E. (2017). Using informative prior from meta-analysis in Bayesian approach. Journal of Data Science, 15(4), 575-588.

[2]   Barker, L. E., Thompson, T. J., Kirtland, K. A., Boyle, J. P., Geiss, L. S., McCauley, M. M. and Albright, A. L. (2013). Bayesian small area estimates of diabetes incidence by United States county. Journal of Data Science, 11(2), 249-267.

[3]   Beaumont, J. and Rivest, L. (2009). Dealing with outliers in survey data. Handbook of Statistics, Sample surveys: design, methods and applications, 29A, 247-279.

[4]   Burgard, P. J., M¨unnich, R. and Zimmermann, T. (2012). Small Area Modelling Under Complex Survey Designs for Business Data. Proceedings of the Fourth International Conference of Establishment Surveys.

[5]   Cadwell, B. L., Thompson, T. J., Boyle, J. P. and Barker, L. E. (2010). Bayesian small area estimates of diabetes prevalence by U.S. county, 2005. Journal of Data Science, 8(1), 173-188.

[6]   Chen, C. H., Duan, N., Meng, X. L. and Margarita, A. (2006). Power-Shrinkage and Trimming: Two Ways to Mitigate Excessive Weights. ASA Proceedings of the Joint Statistical Meetings.

[7]   Chen, M. and Ibrahim, J. G. (2000). Power prior distributions for regression models.

Statistical Science, 15(1), 46-60.

[8]   Congdon, P. (2008). Indirect Area Estimates of Disease Prevalence: Bayesian Evidence Synthesis with an Application to Coronary Heart Disease. Journal of Data Science, 6(1), 15-32.

[9]   Cox, B. and McGrath, D. (1981). An Examination of the Effect of Sample Weight Truncation on the Mean Square Error of Survey Estimates. Proceedings of Biometric Society ENAR Meeting.

[10] Das, K., Jiang, J. and Rao, J. N. K. (2004). Mean squared error of empirical predictor. The Annals of Statistics, 32(2), 818-840.

[11] De Santis, F. (2007). Using historical data for Bayesian sample size determination. Journal of the Royal Statistical Society: Series A (Statistics in Society), 170(1), 95-113.

[12] Erciulescu, A., Cruze, N. B., and Balgobin, N. (2019). Model-based county level crop estimates incorporating auxiliary sources of information. Journal of the Royal Statistical Society: Series A (Statistics in Society), 182(1), 283-303.

[13] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Analysis, 1(3), 515-533.

[14] Gelman, A. (2007). Rejoinder: Struggles with survey weighting and regression modeling. Statistical Science, 22(2), 184-188.

[15] Gettens, J., Lei, P-P. and Henry, A. D. (2015). Using American Community Survey Disability Data to Improve the Behavioral Risk Factor Surveillance System Accuracy. Mathematica Policy Research, DRC Brief, 2015-05.

[16] Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: an appraisal. With comments and a rejoinder by the authors. Statistical Science, 9(1), 55-93.

[17] Goodman, M. S. (2010). Comparison of small-area analysis techniques for estimating prevalence by race. Preventing Chronic Disease 7(2), A33.

[18] Gregg, E. W., Kirtland, K. A., Cadwell, B. L., Burrows, R. N., Barker, L. E., Thompson, T. J. and Geiss, L. S. (2009). Estimated county-level prevalence of diabetes and obesity— United States, 2007. Morbidity and Mortality Weekly Report, 58(45), 1259-1263.

[19] Hansen, M. H., Madow, W. G. and Tepping, B. J. (1983). An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys. Journal of the American Statistical Association, 78, 776-793.

[20] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47,

663-685.

[21] Ibrahim, J. G. and Chen, M.H. (2000). Power Prior Distributions for Regression Models. Statistical Science, 15, 46-60.

[22] Ibrahim, J G., Chen, M-H, Gwon, Y. and Chen, F. (2015). The power prior: theory and applications. Statistics in Medicine, 34(28), 3724-3749.

[23] Kish, L. and Frankel, M. (1974). Inference from complex samples (with discussion). Journal of the Royal Statistical Society: Series B, 36, 1-37.

[24] Meng X. L., Chen C., Duan N. and Alegria M. (2010). Power-shrinkage: An alternative method for dealing with excessive weights. Presentation at Joint Statistical Meetings.

[25] http://andrewgelman.com/movabletype/mlm/meng_JSM_presentation_20090802_8am.pdf

[26] Neelon, B. and O'Malley and A. J. (2010). Bayesian analysis using power priors with application to pediatric quality of care. Journal of Biometrics & Biostatistics, 1, 1-9.

[27] Olives, C., Myerson, R., Mokdad, A. H., Murray, C. J. and Lim, S. S. (2013). Prevalence, Awareness, Treatment, and Control of Hypertension in United States Counties, 2001-2009. PLoS ONE, 8(4).

[28] Pierannunzi, C., Xu, F., Wallace, R. C., Garvin, W., Greenlund, K. J., Bartoli, W., Ford, D., Eke, P. and Town, G. M. (2016). A Methodological Approach to Small Area Estimation for the Behavioral Risk. Preventing Chronic Disease, 14(13), E91.

[29] Pfeffermann, D. (2013). New important developments in small area estimation. Statistical Science, 28(1), 40-68.

[30] Planck, N. R. V., Andrew, F and Silver, H, E. (2017). Hierarchical Bayesian models for small area estimation of county-level private forest landowner population. Canadian Journal of Forest Research, 47(12), 1577-1589.

[31] Potter, F.J. (1988). Survey of Procedures to Control Extreme Sampling Weights. Proceedings of the Survey Research Methods Section of the American Statistical Association.

[32] Rao, J. N. K. and Molina I. (2015). Small Area Estimation. Second Edition. Wiley, New York.

[33] Schneider, K. L., Lapane, K. L., Clark, M. A. and Rakowski, W. (2009). Using Small-Area Estimation to Describe County-Level Disparities in Mammography. Preventing Chronic Disease, 6(4), A125.

[34] Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley, Reading, MA.

[35] Xie, D., Raghunathan, T. E. and Lepkowski, J. M. (2007). Estimation of the proportion of overweight individuals in small areas-a robust extension of the fay-herriot model. Statistics in Medicine, 26(13), 2699-2715.

[36] Zhang, Z., Zhang, L., Penman, A. and May, W. (2011). Using small-area estimation method to calculate county-level prevalence of obesity in Mississippi, 2007-2009. Preventing Chronic Disease, 8(4), A85.

APPENDIX

Optimal "Shrinkage" Factors

The weights "shrinkage" index τ can be optimized by the MSE$\left(\hat{p}_{ij} , \hat{p}_{ij}^{\tau} \right)$ which consists of two parts: $(\hat{p}_{ij} - \hat{p}_{ij}^{\tau})^2$ and $\text{var}(\hat{p}_{ij}^{t})$. Suppl. Figure 2 shows the relationship between τ and MSE $\hat{p}_{ij,\hat{p}_{ij}^{\tau}}$. Two dash lines represent $(\hat{p}_{ij} - \hat{p}_{ij}^{\tau})^2$ and $\text{var}(\hat{p}_{ij}^{\tau})$ changing across τ , respectively. The solid represents mean square error which is the sum of those two parts. The joint of these two dish lines points to the minimal of MSE which corresponds to the optimal τ. Since the MSE is relied on the sample sizes and case counts, the optimal τ varies: Self-care Difficulty has the smallest τ=0.81 while Ambulatory Difficulty has the largest τ=0.89.
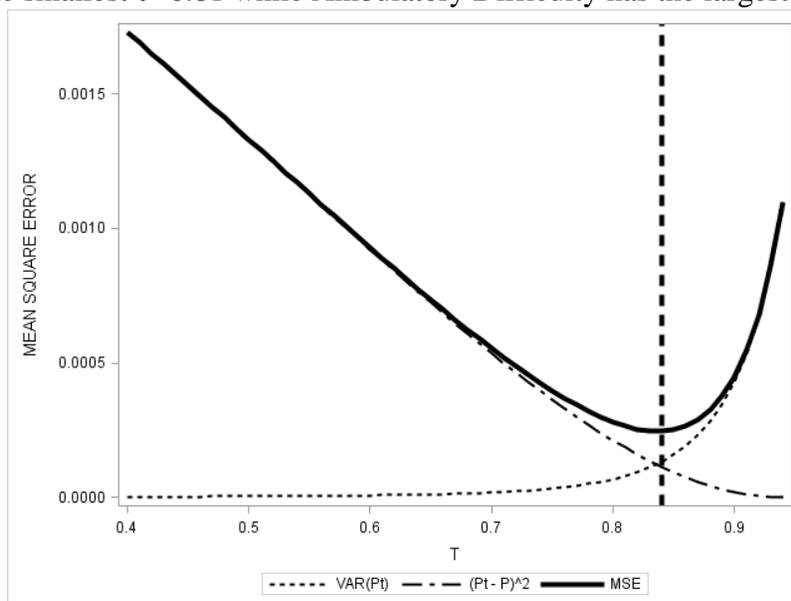

Figure 1 (supple.) MSE$\left(\hat{p}_{ij} , \hat{p}_{ij}^{\tau} \right)$ changes across the weights transformed index τ.
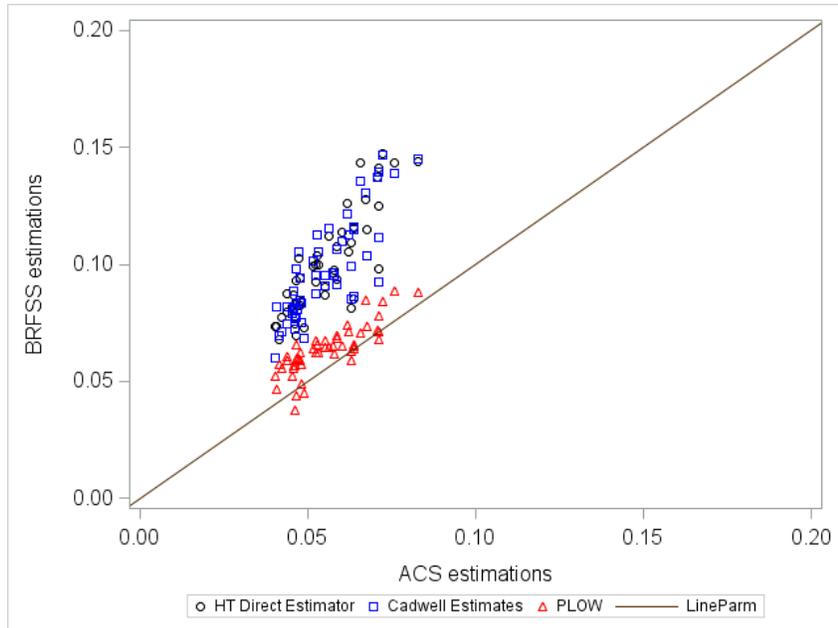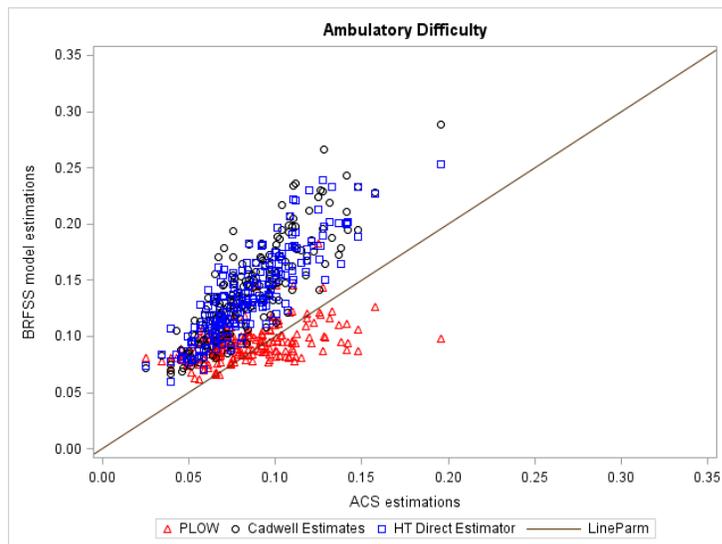
Figure 2 (suppl.) Scatter plot for comparisons of state-level prevalence of Cognitive Difficulty of ACS with estimates using HT direct estimator (black circle), Cawdell estimates (blue square) and PLOW (red triangle). The diagonal reference line (dash line) represents the estimates are same to ACS 1-year state-level results
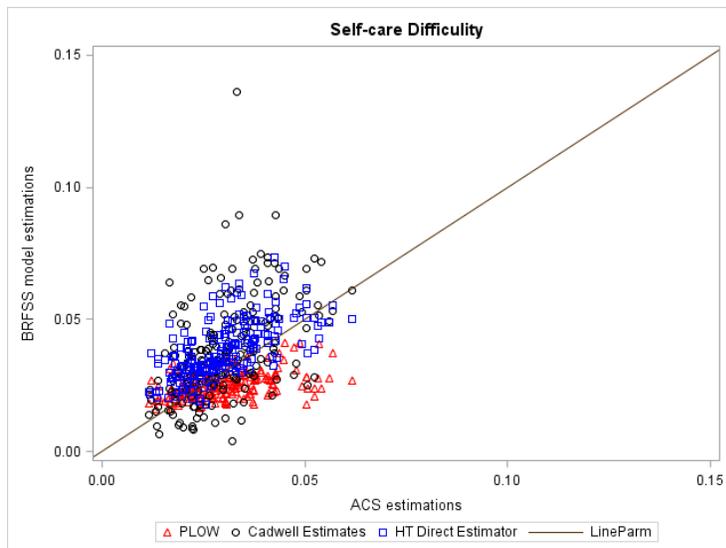
Figure 3 (suppl.) Validation plots for comparison of results of ACS 1-year data with results of HT direct estimator (black circle), Cadwell estimates (black square) and PLOW (red triangle) for the Ambulatory and Self-care Difficulties using "pseudo-county" data. The diagonal reference line (dash line) represents the estimates results are same to ACS 1-year county-level results