Proceedings of Statistics Canada Symposium 2002
Modelling Survey Data for Social and Economic Research

OBTAINING CANCER RISK FACTOR PREVALENCE ESTIMATES IN SMALL AREAS: COMBINING DATA FROM THE BEHAVIORAL RISK FACTOR SURVEILLANCE SURVEY AND THE NATIONAL HEALTH INTERVIEW SURVEY

Michael Elliott[1]

ABSTRACT

The National Health Interview Survey (NHIS) and the Behavioral Risk Factors Surveillance Survey (BRFSS) contain information about behavioral risk factors, but neither provides ideal small area prevalence estimates. The NHIS is a nationally representative, face-to-face survey with a high response rate; however, small area identifiers are unavailable to the public. The BRFSS is a state-level telephone survey that excludes non-telephone households and has a lower response rate, but does provide county identifiers. We combine information from the BRFSS and NHIS using calibration estimators, allowing the complementary strengths of one survey compensate for the weakness of the other.

KEY WORDS: Calibration Estimation, Generalized Raking Estimators, Small-Area Estimation, Cigarette Smoking.

1. INTRODUCTION

1.1 Description

Cancer surveillance research requires accurate estimates of risk factors, including behavioral risk factors, at the small area level. Behavioral risk factors of interest include life style characteristics (e.g., smoking, dietary habits, physical activity, and obesity), economic status (e.g., education and income) and health care utilization (e.g., insurance characteristics and cancer screening practices). These estimates have been used as inputs into regressions equations using estimate cancer incidence at the US county level (Pickle *et al*. (1996) and Nandram *et al.* (2000)), or as leading indicators of changes in the cancer outcomes that have policy implications at both the national and the small area level. For example, differential rates of cancer screening use in the U.S. by age, race, education, and income have been well documented (Potosky *et al.*, 1998). Although successful strategies to increase cancer screening among underutilizing populations have recently been reported, there has been an unevenness in the targeting of intervention research resulting in gaps in coverage. These gaps are evident geographically, which may warrant further investigation for the need for tailoring intervention research as well (Legler *et al.*, 2002).

Risk factor prevalence estimates are often obtained from surveys such as the Current Population Survey (CPS), the National Health Interview Survey (NHIS), or the Behavioral Risk Factors Surveillance Survey (BRFSS). Unfortunately, none of these surveys is an ideal vehicle to provide small-area estimates. The NHIS is a high-response rate, nationally representative area sample of households that collects information based on a face-to-face interview. However, the NHIS is designed only to construct national level estimates of prevalence, not state- or sub-state-level estimates. On the other hand, BRFSS is a state level telephone survey that provides reasonable coverage across all states and many counties, but only of the households that have a telephone, and it suffers from a considerably lower response rate than the NHIS.

[1]Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, 612 Blockley Hall, 423 Guardian Drive, Philadelphia, PA USA 19104 (melliott@cceb.upenn.edu)

One strategy to obtain small-area estimates, then, is to combine information from both surveys. Although a variety of methods have been considered to combine information from multiple surveis (dual frame estimation [Hartley, 1962, 1974; Lohr and Rao 2000] and statistical matching [Rodgers 1984; Moriarity and Scheuren 2001] among others), the problem in this context has two unique issues. First, the NHIS lacks small-area identifiers that can be used. Second, the BRFSS is believed to suffer from selection and frame bias relative to the NHIS. Hence we propose a new "model-assisted" approach that has the advantage of not requiring small-area identifiers for the survey dataset providing the bias calibration, yet is able to take advantage of such small area identifiers if they are available. The basic concept is to make the BRFSS cases "look more like" the NHIS cases by recalibrating the existing survey case weights in the BRFSS so that the weighted estimates from the BRFSS more closely reflect the weighted estimates of the NHIS. We accomplish this reweighting using calibration estimation. We apply this approach to the estimation of 1997 prevalence of cigarette smoking among males 18 and older in the US at the county level.

2. MODEL ASSISTED ESTIMATION: RECALIBRATING THE BRFSS USING THE NHIS

In the calibration estimation approach, the NHIS provides a vector of "control totals" X^s for a set of key demographic measures available in both the NHIS and BRFSS over a "large" area *S*. For example, X^s might consist of a vector of sums of various demographic and health-related measures. We then consider estimators of the form

$$\hat{Y}_{CAL} = \sum w_i y_i$$

where the w_i satisfy the constraint

$$\sum_{i \in S} w_i \mathbf{x}_i = \mathbf{X}^S. \tag{2}$$

subject to minimizing a distance criterion between the w_i and the sampling weights a_i given by $D(w_i, a_i)$.

Deville and Sarndal (1992) describe a range of options for $D(w_i, a_i)$. If $D(w_i, a_i) = (1/2)\sum (w_i - a_i)^2 / a_i$, we obtain the Generalized Regression Estimator or GREG estimator (Cassel, Sarndal, and Wretman 1976; Isaki and Fuller 1982). For the GREG estimator, closed-form solutions for (2) under the constraint $D(w_i, a_i)$ are possible:

$$w_{i,GREG} = a_i \left\{ 1 + \left(\mathbf{X} - \hat{\mathbf{X}} \right) \mathbf{T}^{-1} \mathbf{x}_i \right\}$$

where $\mathbf{T} = \sum_{i \in S} a_i \mathbf{x}_i \mathbf{x}_i^T / c_i$ and $\hat{\mathbf{X}} = \sum_{i \in S} a_i \mathbf{x}_i$, where c_i is a specified constant. If there are *p* categorical control variables each of dimension I_k, and $\mathbf{X}^s$ is given by the $I_1 \times \cdots \times I_p$ cell counts from a complete contingency table, then $\hat{Y}_{GREG}$ is simply the BRFSS estimator further poststratified to the NHIS population total estimates.

More generally, an iterative algorithm can be employed to solve (2) under the constraint $D(w_i, a_i)$. The minimizing distance criterion between the w_i and the BRFSS sampling weights a_i yields

$$d_i(w_i, a_i) - \mathbf{x}_i^T \boldsymbol{\lambda} = 0 \tag{3}$$

where $d_i(w_i, a_i) = \frac{\partial D(w,a)}{w_i}$ and $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers induced by the constraint (2). When $D(w,a) = \sum \left[w_i \log(w_i / a_i) - w_i + a_i \right]$ ("generalized raking estimator"), (3) yields

$$\log\left[\frac{w_i}{a_i}\right] - \mathbf{x}_i^T \boldsymbol{\lambda} = 0$$

or

$$w_{i,RAKE} = a_i e^{\mathbf{x}_i^T \boldsymbol{\lambda}} \qquad (4)$$

where (2) requires

$$\sum_{i \in s} a_i e^{\sum_j \lambda_j x_{ji}} x_{ji} - X_j = 0 \text{ for all } j = 1, \ldots J \ . \qquad (5)$$

We then solve (5) for λ_j using a Newton-Raphson algorithm.

One difficulty with GREG estimation is that the resulting w_i are not constrained to be positive. Negative weights are generally not usable by standard software packages, and can cause other problems (e.g., negative small-area direct prevalence estimates). Generalized raking weights also tend to be more stable, in keeping with the "main-effect only" interpretation of raking as opposed to the "interaction" interpretation of poststratification. Thus the analysis below uses the generalized raking estimator.

2.1 Calibration region *S* and calibration variables $\mathbf{X}^S$.

The area *S* over which the calibration "distances" are computed could range from the entire United States to the smallest areas for which any NHIS cases exist. The choice of *S* involves a bias-variance tradeoff, where the smallest area will make the BRFSS "look most like" the extant NHIS cases, but the small amount of NHIS calibration information will make the resulting estimators potentially unstable; likewise larger areas will assume greater exchangeability over a set of small areas of estimation but reduce the associated variability.

For the public-only NHIS data, the smallest calibration area *S* is a four region-by-seven MSA size categorization. The regions are the four US Census regions: Northeast: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania; Midwest: Ohio, Indiana, Illinois, Michigan, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, and Kansas; South: Delaware, Maryland, the District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Texas, and Oklahoma; West: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, and Hawaii. The MSAs are seven categories defined by: metropolitan areas >5 million, 2.5-5 million, 1-2.5 million, 500 thousand-1 million, 250-500 thousand, less than 250 thousand; and non-metropolitan areas. This is the set of calibration regions used in this analysis.

Calibration variables to be included in $\mathbf{X}^S$ should clearly include variables likely to be sensitive to any response bias difference between the BRFSS and the NHIS. The length *J* of the calibration vector again involves a bias-variance tradeoff, with a large *J* inducing the largest degree of recalibration and hence bias reduction, but also creating larger values of w_i that may inflate the variability of the small-area prevalence estimates. Because the number of recalibration variables is limited and the direct estimators of prevalence are unstable, we will generally use some sort of smoothing technique to obtain small-area prevalence estimates (see Section 2.2 below). Our preliminary analyses indicated that use of this smoothing allows potential bias correction to be maximized by including the largest set of control totals available. This included the common risk behavior of interest (here smoking behavior), common demographic measures – education, household income, marital status, race/ethnicity, and gender, and common health measures: health insurance status, age, and body-mass index (BMI).

2.2 Smoothing Small-Area Estimates.

Our primary small area of interest *s* will be US counties, which currently number 3115 (not including the Alaskan Census Areas). We exclude Kalawao County in Hawaii (population 147) since we had very limited demographic data available, and report results on 3114 US counties.

Direct estimates for small areas are simply given by $\hat{\theta}_{ws}^{DIR} = \sum_{i\in s} w_i y_i \Big/ \sum_{i\in s} w_i$. However, these direct estimators can be quite unstable if the number of subjects in county is small. An alternative is to use a weighted logistic regression estimator to make point estimates for the individual counties; that is, we solve the score equation (Binder 1983)

$$U_w(\boldsymbol{\beta}) = \sum_i w_i \mathbf{z}_i \left(y_i - \text{expit}\left(\mathbf{z}_i^T \boldsymbol{\beta}\right)\right) \qquad (7)$$

where $\text{expit}(x) = \exp(x)\big/\left(1+\exp(x)\right)$ and $\mathbf{z}_i \equiv \mathbf{z}_s$ for all $i \in s$ is a vector of county-level covariates. Solving the score equation we obtain an estimate satisfying $U_w(\hat{\boldsymbol{\beta}}_w) = 0$, then for each county the calibrated regression estimator is given by

$$\hat{\theta}_{ws}^{REG} = \text{expit}\left(\mathbf{z}_s^T \hat{\boldsymbol{\beta}}_w\right) \quad . \qquad (8)$$

For the analyses reported below, $\mathbf{z}_s$ consists of percentage of the 1996 county population that was black, percentage of the 1996 county population that was Hispanic (of any race), percentage aged 25 and older that graduated from high school (1990), percent 25 and older that graduated from college (1990), per capita property taxes (1992), per capita federal funds and grants (1997), per capita social security benefits (1996), percent of persons below poverty (1993), number of serious crimes known to police per capita (1995), the civilian labor force unemployment rate (1994), number of social service establishments per capital (1995), per capita wages and salaries by county of residence (1996), Monday-Friday newspaper readership rate (1997), population per square mile (1990), median effective buying income index (2000), per household total retail, eating and drinking sales (2000), blue collar occupations per capita (2000), and total population (1997-2000).

For comparison with a BRFSS-only analysis, the direct estimator $\hat{\theta}_{as}^{DIR} = \sum_{i\in s} a_i y_i \Big/ \sum_{i\in s} a_i$ can be obtained. Similarly, the "uncalibrated" regression estimator $\hat{\theta}_{as}^{REG}$ that does not make use of the NHIS data can be obtained by replacing $\hat{\boldsymbol{\beta}}_w$ in (3) with $\hat{\boldsymbol{\beta}}_a$, where $\hat{\boldsymbol{\beta}}_a$ is obtained by replacing the calibration weights w_i in (7) with the sampling weights a_i .

2.3 Inference

Like most techniques that reduce bias, the proposed calibration weights will tend to increase variance in the resulting prevalence estimates. This is due to the increase in the variability of the weights themselves, as well as in the uncertainty in the NHIS calibration totals as well. More specifically, consider the estimator $\hat{\boldsymbol{\beta}}_w$ in (8) as a function of the calibration weights w_i , which themselves have variability to due uncertainty in the NHIS-derived calibration totals $\mathbf{X}^S$. Conditioning on the vector of weights w_i we obtain

$$Var(\hat{\boldsymbol{\beta}}_w) = E\left[Var(\hat{\boldsymbol{\beta}}_w) \mid w\right] + Var\left[E(\hat{\boldsymbol{\beta}}_w) \mid w\right] \qquad (9)$$

We estimate $E\left[Var(\hat{\boldsymbol{\beta}}_w) \mid w\right]$ using a jackknife procedure (Korn and Graubard 1999, ch. 2) that accounts for the state-level stratification, and the case weights, ignoring the minor effects of clustering due to any PSU-level clustering due to Waksberg random digit dial sample design. A parametric bootstrap (Davidson and Hickley 1997, ch. 2) is then be used to estimate $Var\left[E(\hat{\boldsymbol{\beta}}_w)\,|\,w\right]$: The covariance of $\mathbf{X}^S$ is estimated using a jackknife technique that accounts for the stratification, clustering, and case weights of the NHIS sample design. Simulations from a multivariate normal with mean $\mathbf{X}^S$ and $Cov(\mathbf{X}^S)$ are then made, and the calibration weights w_i recomputed using the replicated draws of $\mathbf{X}^S$. $\hat{\boldsymbol{\beta}}_w$ is then recomputed using (8) and the recomputed/replicated w_i, and the resulting variability of $\hat{\boldsymbol{\beta}}_w^{rep}$ used to assess $Var\left[E(\hat{\boldsymbol{\beta}}_w)\,|\,w\right]$. The Central Limit Theorem suggests that the multivariate normal assumption for $\mathbf{X}^S$ is reasonable since the components of $\mathbf{X}^S$ are sums from hundreds or thousands of respondents.

Once $Var(\hat{\boldsymbol{\beta}}_w)$ is estimated, the variance of the regression estimator of prevalence $\hat{\theta}_{ws}^{REG}$ is then estimated via the Delta Method.

The variance of $\hat{\theta}_{ws}^{DIR}$, v_{ws}^{DIR}, is given by $E\left[Var(\hat{\theta}_{ws}^{DIR})\,|\,w\right]+Var\left[E(\hat{\theta}_{ws}^{DIR})\,|\,w\right]$, where the first component is estimated by a standard jackknife procedure (now typically a leave-one-out jackknife, since state-level stratification is no longer an issue) and the second component estimated using the parametric bootstrap procedure outlined in the preceding paragraph.

The sampling weights a_i do not incorporate sampling variability of the NHIS and thus the variance of the BRFSS-only estimators $\hat{\theta}_{as}^{REG}$ and $\hat{\theta}_{as}^{DIR}$ can be estimated using standard jackknife or Taylor Series approximations.

2.4 Missing Data

The analysis below was conducted under a missing completely at random (MCAR) assumption (Little and Rubin 1987, ch. 1,5). Under MCAR, missingness is assumed to be completely independent of the data, Hence the only problem that missing data may introduce is recalibration resulting solely from differences in missingness at the large area level. To prevent this from happening, separate NHIS weights were generated for each variable, set to 0 if the NHIS subject was missing that data element and adjusted upward so that the total of the non-missing adjusted weights equaled the total weights in the large area *S* (Korn and Graubard 1999, ch. 4):

$$a_{ji}^{NHIS^{(2)}} = \begin{cases} 0 \text{ if } x_{ji} \text{ is missing} \\ \dfrac{\sum_{i\in s} a_{ji}^{NHIS}}{\sum_{i\in s} a_{ji}^{NHIS} I(x_{ji} \text{ observed})} a_i^{NHIS} \text{ if } x_{ji} \text{ is observed} \end{cases}$$

These weights were then used to compute the NHIS calibration totals $\mathbf{X}^S = \left(\sum_{i\in s} a_{1i}^{NHIS^{(2)}} x_{1i}^{NHIS}, \ldots, \sum_{i\in s} a_{Ji}^{NHIS^{(2)}} x_{Ji}^{NHIS}\right)^T$.

3. RESULTS

Using the methods described above, we determine county-level prevalence estimates of current smoking prevalence among males 18 and older using data from the 1997 NHIS and BRFSS.

Table 1 compares the 1997-2000 nationwide NHIS and BRFSS estimates of the smoking behaviors, together with other selected nationwide estimates. The NHIS estimates somewhat higher smoking prevalences. The NHIS also appears to reach a larger proportion of persons of the lowest and highest socioeconomic status, and somewhat more minorities. This is consistent with the inclusion of non-telephone households and the higher response rate in the NHIS.

	1997	
	BRFSS	NHIS
Current Smoker	23.1%	24.7%
Ever Smoked	46.9%	47.6%
Education: Less than HS	13.7%	19.4%
College or more	26.2%	21.8%
Income: <$10K	6.3%	9.5%
>$75K	13.0%	18.2%
African-American	9.8%	11.0%
Hispanic	9.8%	9.6%
Has Health Insurance	85.8%	85.2%
Male	48.0%	48.0%
Mean Age (years)	45.0	44.6
Mean BMI	25.9	26.2

Table 1: Selected nationwide estimates of demographic and health status indicators from the 1997 BRFSS and NHIS.

Figures 1 and 2 show the uncalibrated and calibrated regression estimators of 1997 smoking prevalence among males given by $\hat{\theta}_{as}^{REG}$ and $\hat{\theta}_{ws}^{REG}$ for all 3114 counties in the United States. The basic pattern is similar in both maps, with the highest rates in two bands stretching from the South through eastern Ohio, and across the Rocky Mountains. Metropolitan centers tend to have lower rates than surrounding rural areas.

As Figures 1 and 2 suggest, the correlation between the two regression estimators is fairly high (r=.76). Figure 3 plots difference between the calibrated (NHIS-adjusted) and uncalibrated (BFRSS-only) regression estimators for males and females against the average of the two estimators – “Bland and Altman” plots (Bland and Altman 1986). These plot show that there are non-trivial differences between the two estimators, and, more particularly, that on average the NHIS-adjusted estimates give higher estimates of smoking rates, and that counties with the highest estimated prevalences tended to have the greatest increases under the recalibrated weights, while those with the lowest prevalences tended to be somewhat reduced. In general rates increase under the NHIS-calibrated estimates, consistent with the higher prevalence estimates under the NHIS: the 2.5, 50, and 97.5 percentiles of 1997 county-level adult male smoking prevalences are 21.2%, 27.9%, and 36.2% among the uncalibrated regression estimators, and 21.7%, 30.7%, and 40.2% among the calibrated regression estimators. Rural counties tended to have higher NHIS-calibrated estimates, while urban counties had equal or lower NHIS-calibrated estimates. Regressing the relative change in male smoke prevalence between the uncalibrated and calibrated regression estimators with the county covariates and examining the resulting Type III sum of squares shows that per capita social security benefits – a proxy primarily for age -- had the largest coefficent of partial determination (.89), followed by per cent Hispanic (.87), poverty rate (.81), and educational attainment (.77). These changes are consistent with bias correction for non-telephone households and for “middle-class” bias in response rates (Goyder et al. 2002). (This latter assumption is somewhat contentious: see Groves and Couper 1998.)

Figure 4 gives the standard errors of all 3114 US county estimates for the calibrated regression estimators of male smoking prevalence. The smoothing technique employed tends to lead to large variances only in counties with extreme county covariate predictors. The standard errors of the NHIS-calibrated 1997 county-level smoking prevalence estimates averaged 1.56 times greater among males and 2.29 times greater among females, due to both the uncertainty in the recalibrated weights due to sampling variability in the NHIS and to the increased variability in the weights.

4. DISCUSSION

This manuscript considers whether publicly-available data from the National Health Interview Survey and Behavioral Risk Factor Surveillance Survey can be combined to produce improved small-area estimates of the prevalence of behaviors related to cancer risk, specifically smoking and cancer screening. Because the NHIS includes non-telephone households in its sampling frame and has a considerably higher response rate than the BRFSS, we might assumes its estimates are less biased than equivalent estimates obtained from the BRFSS. Because the public NHIS does not contain small-area identifiers, we consider using calibration techniques to adjusting the sampling weights of the BRFSS so that the weighted estimates from the BRFSS more closely reflect the weighted estimates of the NHIS.

In general, logistic regression estimation procedures that used NHIS-calibrated case weights produced higher estimates of smoking prevalence than estimates derived using BRFSS case weights only. Counties with the highest estimated smoking prevalences tended to have the greatest increases under the recalibrated weights, while those with the lowest smoking prevalences tended to be somewhat reduced. Age, race/ethnicity, poverty rate, and education attainment were the most important predictors of the the difference between the uncalibrated and NHIS-calibrated estimators, consistent with the “middle-class bias” that selection bias and frame non-coverage is hypothesized to create in low-response-rate, telephone-based surveys.

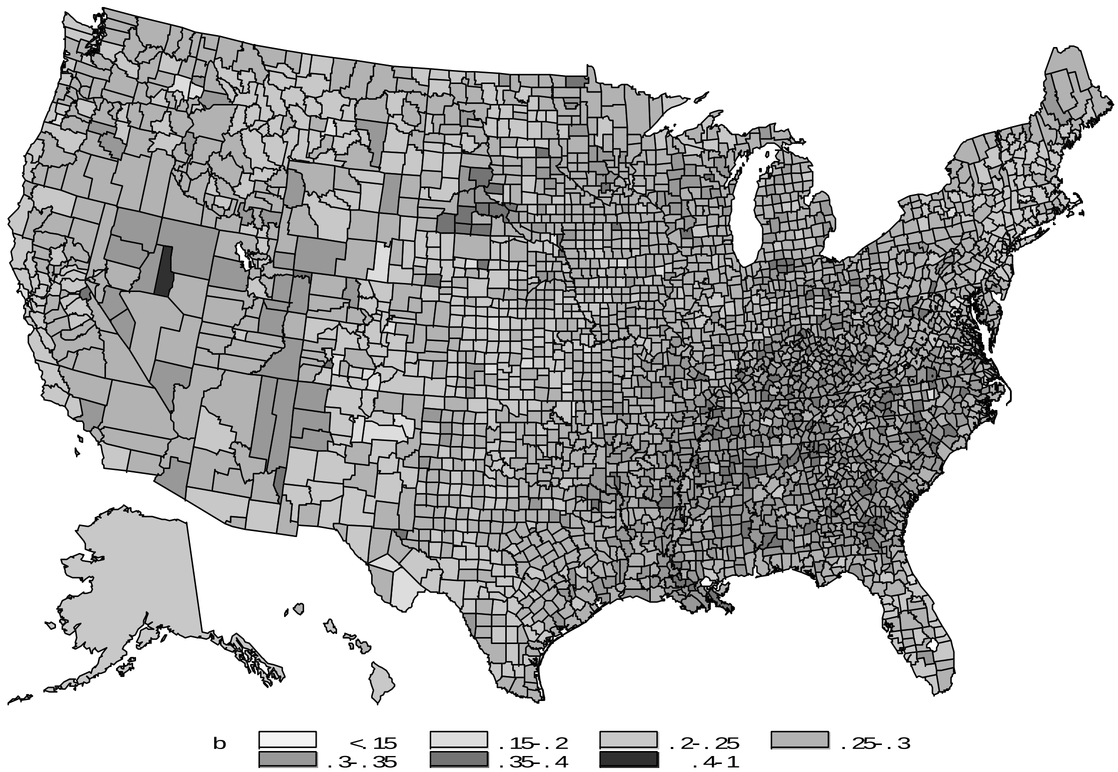

Figure 1: County level estimates of 1997 current smoking prevalence among males 18 and older using uncalibrated (BRFSS-only) regression estimates.

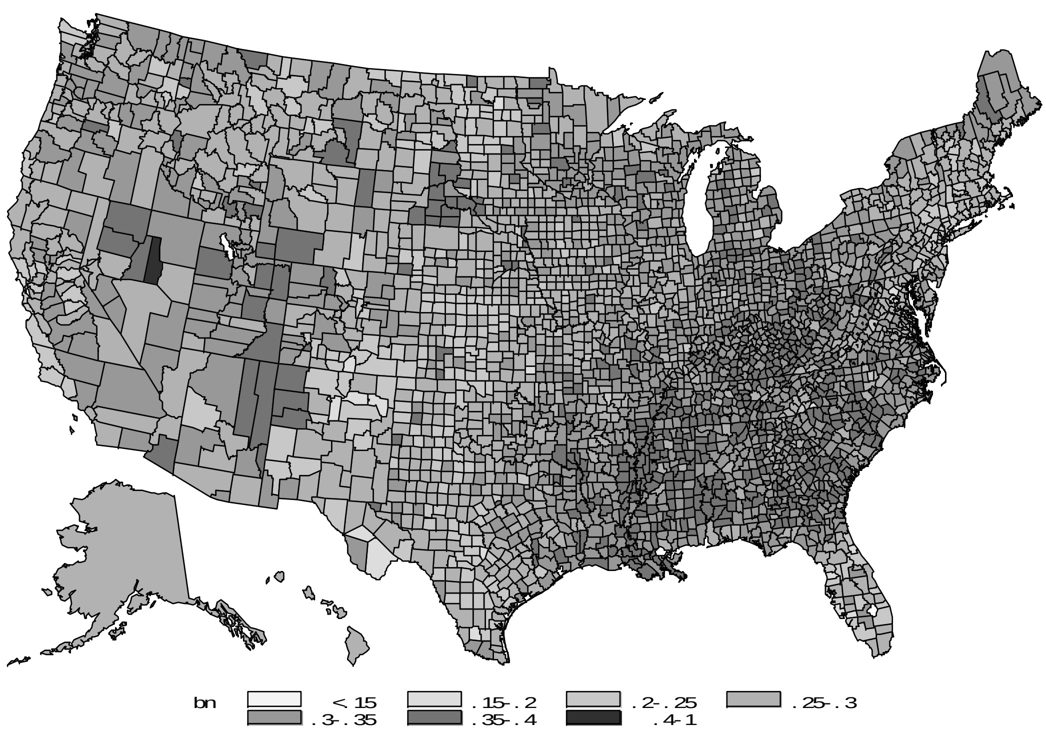

Figure 2: County level estimates of 1997 current smoking prevalence among males 18 and older using NHIS-calibrated regression estimates.

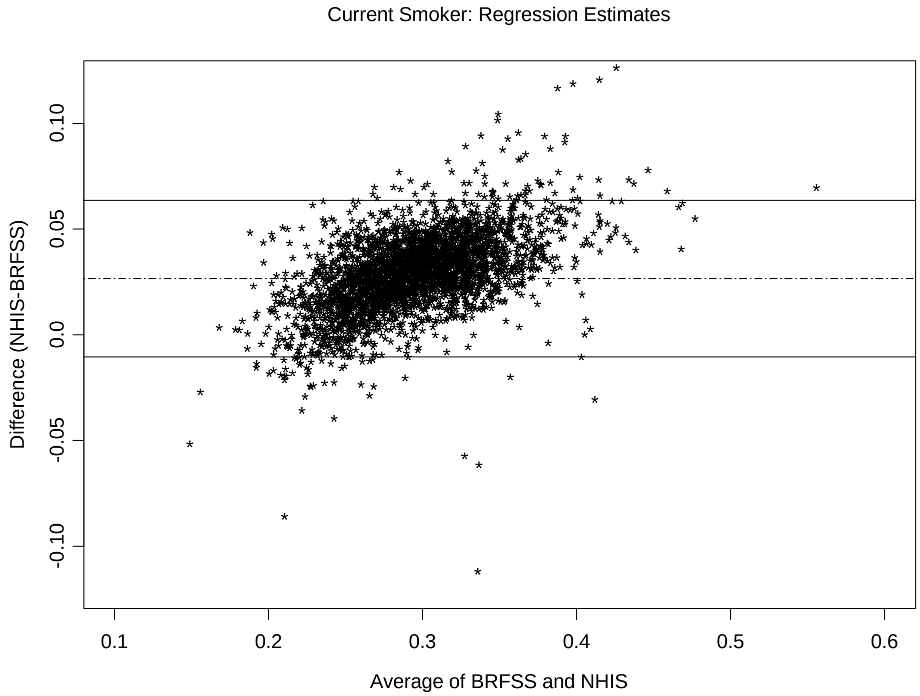

Figure 3: Plot of the difference between NHIS-calibrated and uncalibrated (BRFSS-only) regression estimators of current smoking prevalence among males 18 and older in 1997, versus average of calibrated and uncalibrated estimators, in all 3,114 US counties. (–·–·–) mean difference; (—) mean +/- 2xs standard deviation of difference.

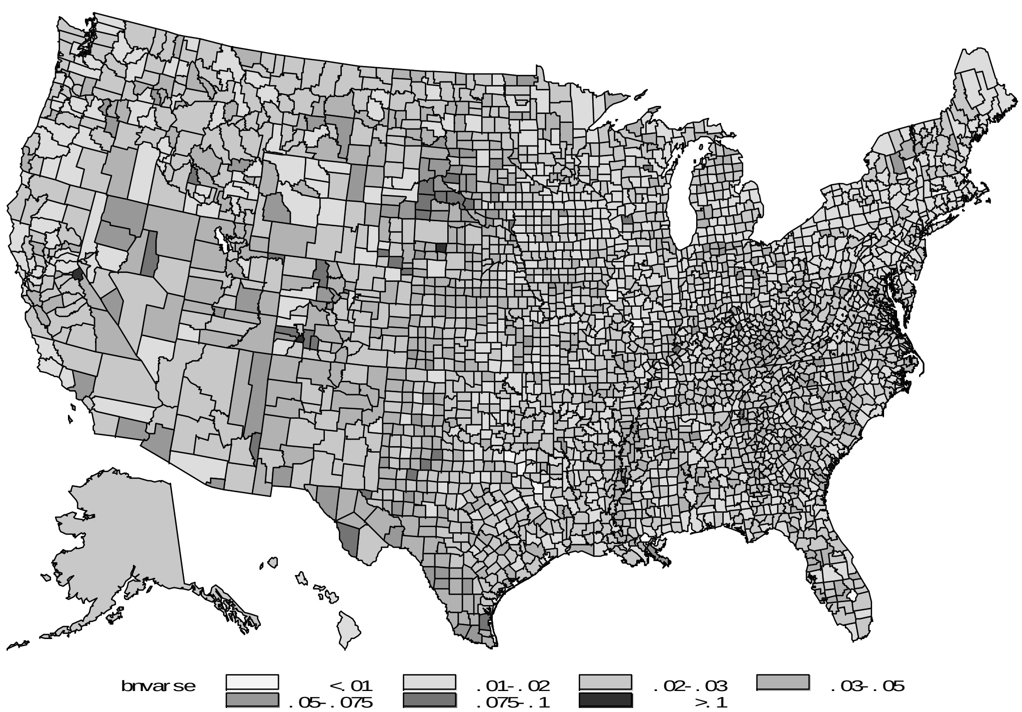

Figure 4: Standard errors of county level estimates of 1997 current smoking prevalence among males 18 and older using NHIS-calibrated regression estimates.

REFERENCES

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-292.

Bland, J.M. Altman, D.G. (1986). Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet*, 327, 307-310.

Cassel CM, Sarndal CE, Wretman JH. (1976), "Some results on generalized difference estimation and generalized regression estimation for finite populations," *Biometrika* 63, pp. 615-620.

Deville JC, Sarndal C-E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association* 87, pp. 376-382.

Davidson, AC, Hinckley, DV (1997), *Bootstrap Methods and their Applications*. Cambridge: Cambridge University Press.

Goyer, J., Warriner, K., Miller, S. (2002), "Evaluating Socio-economic Status (SES) Bias in Survey Nonresponse," *Journal of Official Statistics*, 18, pp. 1-12.

Groves, R.M., Couper, M.P. (1998), *Nonresponse in Household Interview Surveys*. New York: John Wiley and Sons.

Hartley, H.O. (1962), "Multiple Frame Surveys," *Proceedings of the Social Statistics Section, American Statistical Association*, 245-249.

——— (1974), "Multiple Frame Methodology and Selected Applications," *Sankhya*, C36, pp. 99-118.

Isaki C, Fuller W. (1982), "Survey Design under the Regression Superpopulation Model," *Journal of the American Statistical Association* 77, pp. 89-96.

Korn, E.L., Graubard, B.I. (1999), *Analysis of Health Surveys*. New York: John Wiley and Sons.

Legler J., Meissner H., Breen N., Malec D. (2002), "Mammography and the underserved: a geographical perspective on the congruency of national needs and intervention research," draft manuscript, Washington, DC: National Cancer Institute.

Little R.J.A., Rubin D.B. (1987), *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.

Lohr, S.L., Rao, J.N.K. (2000), "Inference from Dual Frame Surveys," *Journal of the American Statistical Association*, 95, pp. 271-280.

Moriarity, C., Scheuren, F. (2001), "Statistical Matching: A Paradigm for Assessing the Uncertainty in the Procedure," *Journal of Official Statistics*, 17, pp. 407-422.

Nandram, B., Sedransk, J., Pickle, L.W. (2000), "Bayesian Analysis and Mapping of Mortality Rates for Chronic Obstructive Pulmonary Disease," *Journal of the American Statistical Association*, 95, pp. 1110-1118.

Pickle, L.W., Mungiole, M., Jones, G.K., White, A.A. (1996), *Atlas of United States Mortality*. Hyattsville, MD: National Center for Health Statistics.

Potosky A.L., Breen N., Graubard B.I., Parsons P.E. (1998), "The association between health care coverage and the use of cancer screening tests: Results from the 1992 National Health Interview Survey." *Medical Care*, 36, pp. 257-70.

Rodgers, W.E. (1984), "An Evaluation of Statistical Matching," *Journal of Business and Economic Statistics*, 2, pp. 91-102.