Assessment of Small-Area Estimates from a Complex Survey Cancer Surveillance Project

Van Parsons[1], Nathaniel Schenker[1], Trivellore E. Raghunathan[2], Dawei Xie[3], William Davis[4]
National Center for Health Statistics, CDC[1]
University of Michigan[2]
University of Pennsylvania[3]
National Cancer Institute[4]

Keywords: Hierarchical Bayes, NHIS, BRFSS, combine surveys, cancer risk factor

1. Introduction[a]

The National Cancer Institute (NCI) recently funded a cancer surveillance research project that required the accurate estimation of cancer risk factors on a periodic basis for most of the 3000+ counties in the United States. The behavioral risk indicators of current smoker status and lifetime smoking status along with the female screening practices of having a mammography within the last two years and having a PAP smear within the past three years were of major interest, but there was no one survey that could produce the magnitude of accurate estimates that were needed. Available were two major survey systems, the National Heath Interview Survey (NHIS) and Behavioral Risk Factors Surveillance Survey (BRFSS). Both surveys used the same or similar questions for the topics of interest and covered the years 1997-2003 completely for smoking variables, but covered the screening questions with planned omissions in some years. While direct design-based estimates could be obtained for larger counties, the small-area focus required some degree of modeling that would "borrow strength" from the existing survey data and externally available (e.g., from Census) socio-economic-demographic county covariates to create estimates over areas with insufficient survey data. This small-area project became a collaborative effort among NCI, the National Center for Health Statistics (NCHS) and the University of Michigan. The resulting research led to the development of a Bayesian hierarchical model which is fully discussed in Raghunathan, et.al (2007). This current paper, based on a poster session, is intended to provide complementary material to this referenced work. Space limitations require omission of details and emphasis on a few tables and maps.

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

2. NHIS and BRFSS Comparisons

The NHIS and BRFSS are two independently managed and sampled surveys. Table 1 provides some design comparisons.

Table 1: Survey Comparisons

	BRFSS	NHIS
Type	State, Telephone only	National, Face-to-face
Sample size per year	150-250 K Households	30-40 K Households
Cost/response	Low	High
Organization	CDC/States	NCHS/Census
Response rate	Lower	Higher
Coverage	Landline Telephone Residential Households, Almost all counties	Dwelling Units for Civilian Noninstitutionalized population, Sample 800+ counties
Available Geographical Information	State (public) County (special request needed)	4 Regions (public) State/County (restricted) Research Data Center

Table 2: National Estimates over Telephone Households

Prevalence(%)	Year	BRFSS	NHIS	Diff
current smoking males	1997	26	27	-1
	1998	25	25	0
	1999	25	25	0
	2000	24	25	-1
mammogram	1997-1998	74	67	7
	1999-2000	72	67	5

For this paper we will focus on two examples of prevalence: men, age 18+, who are current smokers and women, age 40+, who have had a mammography in the past two years. At the national level for telephone households, Table 2 shows a mammography status difference between the two surveys. We attribute this difference to bias, possibly due to mode effect, response rate, or questionnaire variation. A similar comparison was done at the county level over the common counties of both samples with the data aggregated for survey years 1997-1999. Figure 1 displays the *standardized* difference of the estimated prevalence, (p_{NHIS} - p_{BRFSS}), between the two surveys. For those counties with large effective sample sizes there is a tendency for the BRFSS estimate to be larger

(suggesting bias) than the NHIS estimate. The smoking variable showed no apparent bias. Figure 1 also exhibits a large number of counties with BRFSS effective sample sizes less than 25. Such counties may have low precision.

3. Model-Based County-Level Estimation

For county-level estimation the BRFSS survey was considered to be the major data component, but to be supplemented with NHIS data and a statistical model. A hierarchical Bayesian model based on the MCMC method was developed. The major hypothesis imposed was that the NHIS had biases of smaller magnitude than the BRFSS. Ultimately, the model would compensate for small county samples, correct for households without telephones, and correct for possible BRFSS system biases (e.g., nonresponse and mode effects). BRFSS county-based prevalence rates for current smoking status and mammography screening are displayed before and after the modeling in Figures 2 to 5.

The effect of the estimation procedure on detecting change in prevalence over time is also of interest. If p_t is a prevalence estimator at time period t, we considered inferential change from time period t_1 = 1997-1999 to time period t_2 = 2000-2003 for the three estimators, direct BRFSS, direct NHIS, and our model. Consider $t = (p_{t2} - p_{t1}) / \text{stderr}(p_{t2} - p_{t1})$ as an exploratory statistic at the county-level, with the magnitude $| t_{\text{method}} | > 2$ treated as an indication of significance. These t values for the different estimation methods are summarized in Table 3. Both the BRFSS and the model tests for trend are compared for about 3000+ counties, but the model-based statistics appear less sensitive to finding change. At this time we have limited measures of the "truth" at the county level, and the results are still being investigated

4. Poster Conclusions/Comments

- Model provides a method to combine BRFSS + NHIS county estimates
- Model tends to drive estimates in the NHIS direction
- Model Bias at the county level is unknown
- Model may be more adapted to estimating prevalence at time periods rather than estimating change over time
- More evaluations are necessary

Table 3: Status of Simple Tests for Trend 1997-1999 vs. 2000-2003 by Type of Estimator

	Current Smoker		Mammography	
Type of Estimator Used for Testing	Decr[1]	Incr[2]	Decr[1]	Incr[2]
BRFSS	58	73	17	166
Model	8	0	1	59
NHIS[3]	29	27	13	53
Both Model & BRFSS	1	0	1	26
Both Model & NHIS	8	0	1	27

[1] Number of Counties with Significant Decreasing Trend
[2] Number of Counties with Significant Increasing Trend
[3] NHIS uses 800+ counties for comparison

5. References

Raghunathan, T. E., Xie, D., Schenker, N., Parsons, V. L., Davis, W. W., Dodd, K. W., Feuer, E. J., (2007) Combining Information from Two Surveys to Estimate County-Level Prevalence Rates of Cancer Risk Factors and Screening, *Journal of the American Statistical Association* (to appear)

Figure 1: % Mammography, Female 40+ in households with telephones Standardized Z difference (NHIS − BRFSS) for sampled NHIS counties ,1997−1999

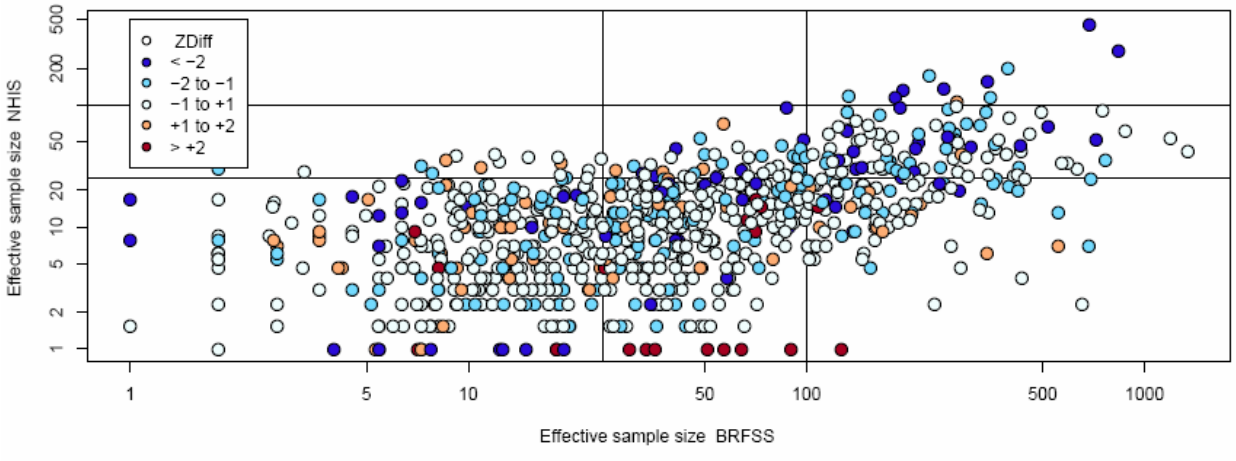
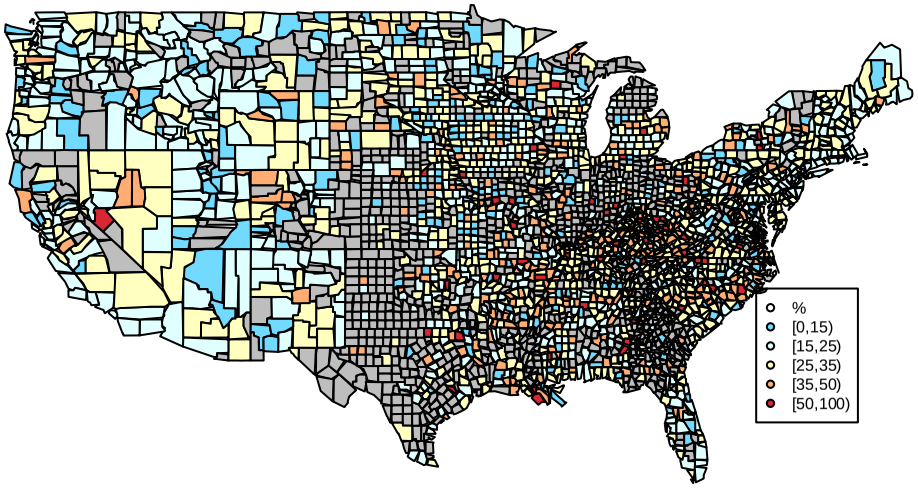
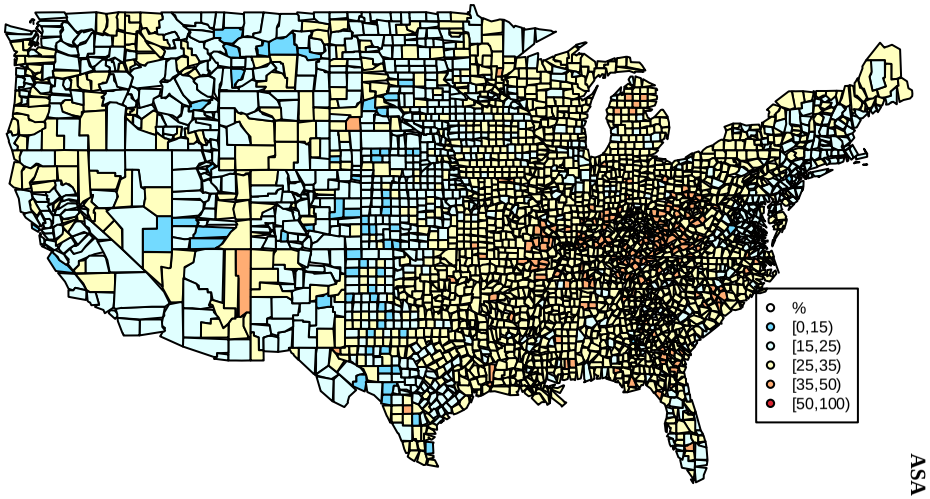


Figure 2: BRFSS % Current smoker, Male 18+ ,1997−1999
(Counties with BRFSS sample 10+)



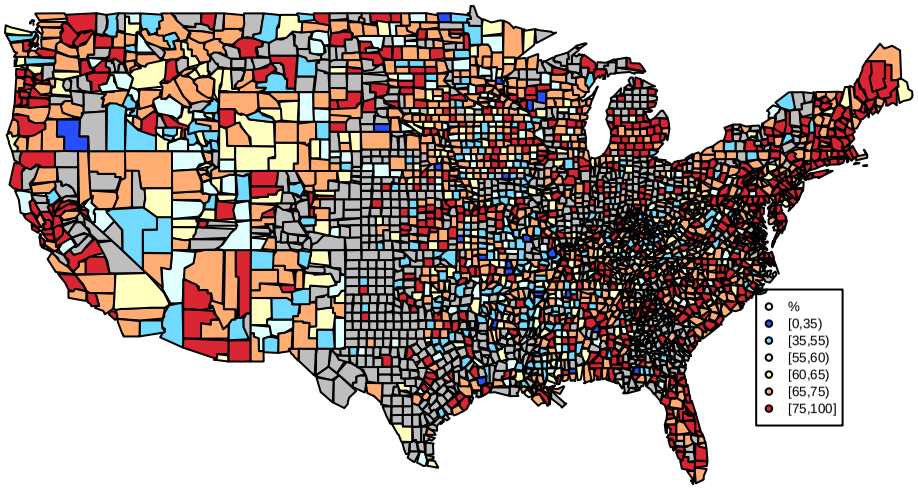
County Quartiles : 18% , 26% , 34%, range: 0% − 100%

Figure 3: Model−based % Current smoker, Male 18+, 1997−1999
(Model−based sample sizes are equivalent to100+ for almost all counties)



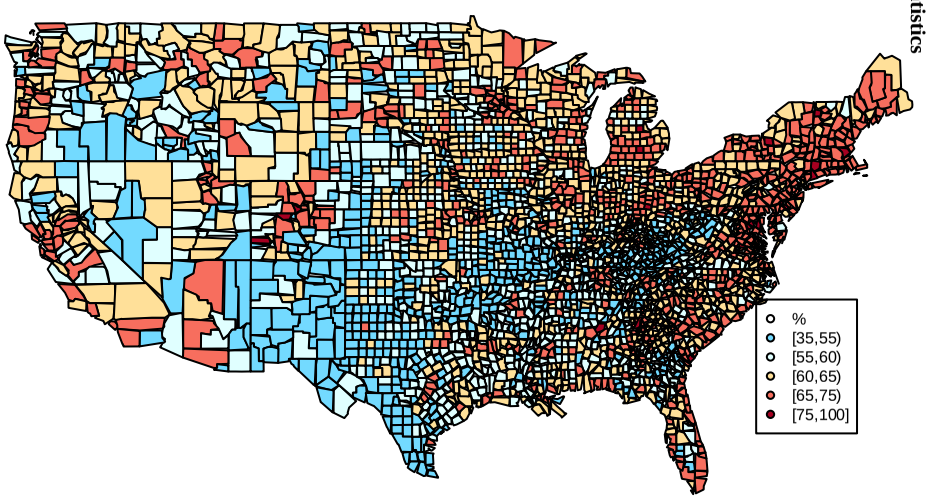
County Quartiles : 23%, 27%, 31%, range: 6% − 47%

Figure 4: BRFSS % of Women 40+ having Mammography
(Counties with BRFSS sample 10+)



County Quartiles : 58% ,69% , 78%, range: 0% − 100%

Figure 5: Model % of Women 40+ having Mammography
(Model−based sample sizes are equivalent to 100+ for almost all counties)



County Quartiles : 56% ,61% , 65%, range: 35% − 82%