

# Comparison of Small-Area Analysis Techniques for Estimating County-Level Outcomes

Haomiao Jia, PhD, Peter Muennig, MD, MPH, Elaine Borawski, PhD

**Background:** Since many health data are unavailable at the county level, policymakers sometimes rely on state-level datasets to understand the health needs of their communities. This can be accomplished using small-area estimation techniques. However, it is unknown which small-area technique produces the most valid and precise results.

**Methods:** The reliability and accuracy of three methods used in small-area analyses were examined, including the synthetic method, spatial smoothing, and regression. To do this, severe work disability measures were first validated by comparing the 2000 Behavioral Risk Factor Surveillance System (BRFSS) and Census 2000 measures (used as the gold standard). The three small-area analysis methods were then applied to 2000 BRFSS data to examine how well each technique predicted county-level disability prevalence.

**Results:** The regression method produces the most valid and precise estimates of county-level disability prevalence over a large number of counties when a single year of data is used.

**Conclusions:** Local health departments and policymakers who need to track trends in behavioral risk factors and health status within their counties should utilize the regression method unless their county is large enough for direct estimation of the outcome of interest.  
(*Am J Prev Med* 2004;26(5):453–460) © 2004 American Journal of Preventive Medicine

## Introduction

State and local health departments, community-based organizations, and policymakers need valid and precise local data for program planning, program evaluation, and resource allocation.<sup>1,2</sup> However, while there are rich and varied sources of national health data, few local health data are available for most counties, and local health departments and community-based organizations often lack the resources to collect data on their own. The Behavioral Risk Factor Surveillance System (BRFSS), which was designed to help improve local decision making by providing state-level data, provides an array of annual health behavior and health status information.<sup>3,4</sup> However, in the BRFSS, the sample size at the county level is often too small to estimate local outcomes.<sup>4–6</sup>

Local health agencies commonly rely on temporal (i.e., combine several years of data for one county) or spatial (i.e., combine several counties together) data aggregation to increase sample size and therefore in-

crease precision. However, estimates based on temporally aggregated data cannot show time-trend differences for smaller counties, and area differences require spatial delineation. Many statistical procedures, which fall under the rubric of “small-area analyses,” have been developed to help fill this void.<sup>7,8</sup> These include the synthetic method, spatial data smoothing, and regression analysis. The synthetic method applies statistics for the nation as a whole to local areas based on each area’s demographic characteristics.<sup>9</sup> Spatial data smoothing uses data from neighboring communities and the area of interest to calculate a weighted moving average value of neighboring areas.<sup>10</sup> Finally, the local area outcomes can be estimated through multivariate regression with area-specific data as predictors.<sup>8,11</sup> To our knowledge, the validity and precision of the BRFSS for county-level analyses have not been examined across these methods.

One useful statistic for local planners is the prevalence of severe work disability. Severe work disability is defined as inability to work due to physical or mental conditions.<sup>12,13</sup> Since 1993, the BRFSS has included a number of questions that measure respondents’ perceived health status and activity limitation.<sup>14</sup> A measure of severe work disability based on these survey instruments has been developed and has shown good to excellent validity.<sup>15</sup> The Centers for Disease Control and Prevention (CDC) subsequently promoted the BRFSS as a potential source of annual disability data.<sup>16</sup> Likewise, the decennial census contains a measure of

---

From the Department of Community Medicine, Mercer University School of Medicine (Jia), Macon, Georgia; Department of Health Policy and Management, Mailman School of Public Health, Columbia University (Muennig), New York, New York; and Department of Epidemiology and Biostatistics, Case Western Reserve University (Borawski), Cleveland, Ohio

Address correspondence to: Haomiao Jia, PhD, Department of Community Medicine, Mercer University School of Medicine, 1550 College Street, Macon GA 31207. E-mail: haomia@yahoo.com.

severe work disability.<sup>17</sup> While slightly different from the measure contained in the BRFSS, the census measure can be used to validate county-level BRFSS disability data, since it was designed with a sampling frame large enough to provide precise and reliable estimates for geographic regions much smaller than a county.<sup>15,17</sup>

The main objective of this study was to investigate various methods for estimating the annual prevalence of county-level severe work disability using the BRFSS (referred to henceforth as “county-level disability prevalence”). Disability is used as a proxy measure to examine the validity and precision of three small-area analysis techniques for other measures or variables as well. The challenge has been to obtain stable estimates of patterns of disease or risk factors for disease, and to allow for relative comparisons with the county of interest.

## Materials and Methods

County-level disability prevalence was examined using 2000 BRFSS data and Census 2000 disability data validated the estimates. The performance of three small-area estimation methods—the synthetic method, spatial data smoothing, and regression analysis—was then compared. A working regression procedure was provided that may help local health officials and other planners understand the needs of the communities they serve, track their progress from year to year, and make comparisons between their counties and neighboring counties.

## Data Sources

The 2000 BRFSS data were used for individual-level classification of severe work disability.<sup>15</sup> The BRFSS is an annual telephone interview survey coordinated by the CDC, with data collected by individual state health departments.<sup>3,4</sup> The BRFSS uses random digit dialing with a multistage cluster design to generate probability-based samples of the adult civilian noninstitutional population of each of the 50 states and the District of Columbia. The BRFSS data contain weighting factors to adjust for poststratification as well as selection probabilities.

Severe work disability was defined as self-reported inability to work, or not working and self-rated poor health, or reported 14 or more days of activity limitation in the past month among adults aged 18 to 64 years.<sup>15</sup> In addition, the following BRFSS variables were used in this analyses: respondent’s age, race, gender, and county of residence.

County-specific data were collected from different sources and merged into the BRFSS survey data by matching the county of the survey respondents’ residence across datasets. The following county-specific data were used: (1) sociodemographic descriptors, socioeconomic statistics, and work disability data from Cen-

sus 2000<sup>17</sup>; (2) the Area Resource File<sup>18</sup>; and (3) 1996–1998 age-adjusted mortality from the CDC.<sup>19</sup>

## Data Analysis

The SAS software package (SAS Institute Inc., Cary NC, 1999) was used for data analysis. Standard error was calculated using SUDAAN (Research Triangle Institute, Research Triangle Park NC, 1997) to adjust for complex sampling design in the BRFSS. County-level disability prevalence was estimated by applying the same data to each small-area analysis method under study.

## The Synthetic Method

Considering the notation  $y$  for the dichotomous presence or absence of severe work disability, suppose that  $y_{ij}$  is a Bernoulli variable with probability  $p_{ij}$  for observations in the demographic category  $j$  and county  $i$ . Let  $\hat{p}_j$  be the estimated three-way age-race-gender tabulated rates at the state level for demographic category  $j$ . Suppose that the population in demographic group  $j$  and county  $i$  is  $n_{ij}$ , and that the total population in county  $i$ ,  $n_i = \sum_j n_{ij}$ , are known (from the Census 2000). Then the synthetic estimate of the rate for county  $i$  is the summary of  $\hat{p}_j$  over all demographic groups, weighted by the proportion of the county population in each demographic category:

$$\hat{p}_i = \sum_j \frac{n_{ij}}{n_i} \hat{p}_j$$

## Spatial Data Smoothing

Because of the small sample size of county-level data in the BRFSS, direct calculation of county-level disability prevalence produces a scattered pattern of data points. The prevalence rates were smoothed using a spatial data smoothing algorithm, called “weighted head-banging.” This stabilizes estimates and better reveals the underlying pattern of prevalence.<sup>10,20</sup>

Let  $\hat{p}_i$  be the estimated prevalence based on the direct calculation of severe work disability prevalence rates for county  $i$ . Each county had up to 27 neighboring counties. The median estimated disability rate was calculated for neighboring counties ( $u_i$ ). These neighboring counties were then grouped according to whether their estimated disability rates fell above or below  $u_i$ . The following two quantities were defined as:

high screen for county  $i$  = weighted median prevalence of neighboring counties  $\geq u_i$

low screen for county  $i$  = weighted median prevalence of neighboring counties  $< u_i$

with weights based on the county population. If  $\hat{p}_i$  was between the high and low screens, its value was not

changed. If  $\hat{p}_i$  was greater than its high screen, then its value was changed to the high screen and vice versa. The smoothing method was applied to all counties, and then the above process was repeated ten times in order to generate a moving average.<sup>10</sup>

## Regression Analysis

Considering a vector of  $t$  covariates,  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijt})'$  for each respondent, and corresponding fixed effects  $\beta = (\beta_1, \dots, \beta_t)'$ , a multilevel logistic regression model was used with a county-level random effect  $\alpha_i$ :

$$\text{logit}(p_{ij}) = \mathbf{x}'_{ij}\beta + \alpha_i.$$

The random effect  $\alpha_i$  is assumed to be normally distributed with a mean of zero and a variance equal to  $\sigma^2$ . The covariates  $\mathbf{x}_{ij}$  include data measured at two levels: the BRFSS survey respondents' demographics and the county socioeconomic and health-related variables. Once the regression parameters were estimated, we calculated  $\hat{p}_i$ , the prevalence rate for demographic group  $j$  in county  $i$ , and then the county-level prevalence rate:

$$\hat{p}_i = \sum_j \frac{n_{ij}}{n_i} \hat{p}_{ij}.$$

In most applications, the random effect term  $\alpha_i$  is too small to have an impact on the accuracy of  $\hat{p}_i$  (<0.001%). To simplify the analysis, values of random effects were not estimated and  $\alpha_i$  was assumed to be 0. However, the random effect term was included in the model in order to improve estimation for fixed effect  $\beta$  to ensure the correct selection of variables for the model.<sup>21,22</sup>

Demographic variables (i.e., age, gender, race/ethnicity) were first added into the model, and county-level variables that were statistically significant were identified using variables known to be associated with work disability. Because some of these variables were moderately to highly correlated, a factor analysis technique was applied to include as many county characteristics as possible.<sup>23</sup> Four factors were identified from these county-level variables: (1) socioeconomic characteristics (unemployment rate, poverty rate, per capita income, median household income, and the proportion of adults with <12 years of education); (2) mortality (age-adjusted mortality by gender and race); (3) demographic characteristics (proportion of persons who live in rural areas and proportion of adults married); and (4) health care service resources and utilization (hospitals, hospital beds, and hospital admissions per 100,000 population).

These four county-level factors, as well as individual demographics, were fitted into the logistic models to obtain regression-based estimates of county-level disability prevalence. The regression analysis was conducted separately for the nine census divisions: New England, Middle Atlantic, East North Central, West

North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific.

## Validations

County estimates were validated by comparing them to county-level disability prevalence data from Census 2000. The census was chosen to serve as the "gold standard" for severe work disability because the census data were designed to provide a precise estimation of severe work disability rates at the county level.

Discrepancies between the census ( $r_i$ ) and the BRFSS ( $p_i$ ) rates were examined by:

1. Scatter plots of the BRFSS estimates versus the census rates
2. Pearson ( $\rho$ ) and Spearman ( $s$ ) correlation coefficients
3. Mean square error (MSE):

$$\frac{1}{N} \sum_{i=1}^N (p_i - r_i)^2;$$

4. Mean absolute difference (MAD):

$$\frac{1}{N} \sum_{i=1}^N |p_i - r_i|;$$

5. Mean relative absolute difference (MRAD):

$$\frac{1}{N} \sum_{i=1}^N \frac{|p_i - r_i|}{r_i}; \text{ and}$$

6. Rank statistics<sup>24</sup>:

$$\frac{\sqrt{12}}{N} \sum_{i=1}^N \left( \frac{\text{rank}(p_i - r_i)}{N + 1} - \frac{1}{2} \right) (p_i - r_i),$$

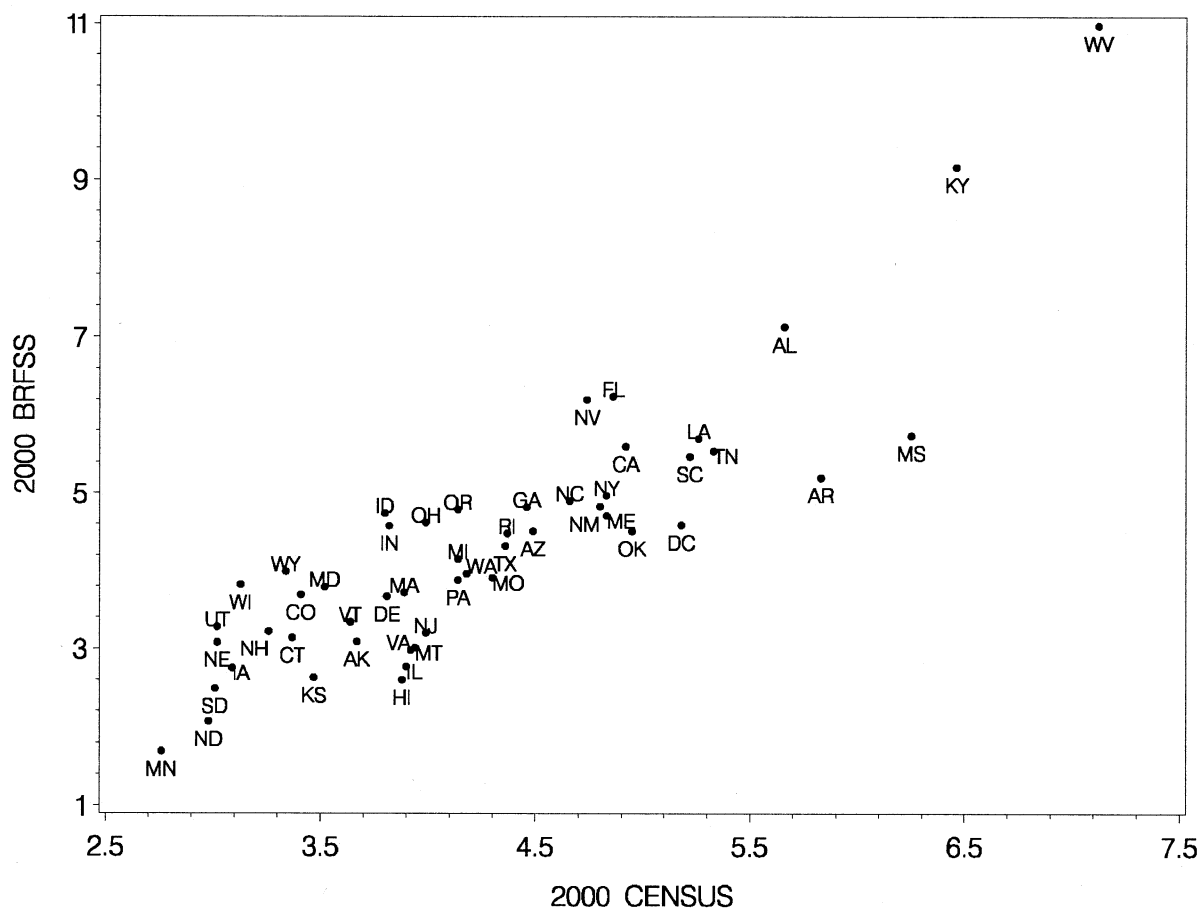
where  $N$  is the number of counties.

## Results

### Nationwide and State-Level Estimates of Disability

In 2000, the BRFSS sampled a total of 146,018 adults aged 18 to 64 years. The overall self-reported severe work disability prevalence was 4.57% (95% confidence interval [CI]=4.35%–4.79%), which was very close to the estimated prevalence of disability obtained from census data (4.43%). Individuals aged 55 to 64 years were approximately six times more likely to report being disabled than those aged 18 to 24 (12.37% vs 2.07%,  $p < 0.001$ ). The increase in risk with age only occurred for persons aged  $\geq 35$  years. This positive relationship between age and disability was present in all race/ethnicity–gender subgroups.

Women were approximately 74% more likely to report severe work disability than men (5.92% vs



**Figure 1.** Plots of state-level severe work disability rates estimated from Behavioral Risk Factor Surveillance System (BRFSS) data versus rates calculated from the 2000 Census. (X and Y values represent percent prevalence.) States are represented by postal code abbreviations.

3.41%), and these gender differences were also significant in all age-gender subgroups. Nonwhites had a higher severe work disability rate than whites (5.82% vs 4.26%), but white and nonwhite women aged 18 to 24 had a similar risk (2.85% vs 2.87%).

State-level prevalence data were calculated directly from the BRFSS. Figure 1 presents plots of state-level, severe work disability prevalence rates estimated from the BRFSS data versus rates calculated from the census. The correlation coefficient between BRFSS and census data was 0.89 and 43 of the 51 (84%) of the BRFSS estimates were within 1% of the census data.

### Comparison of Small Area Techniques

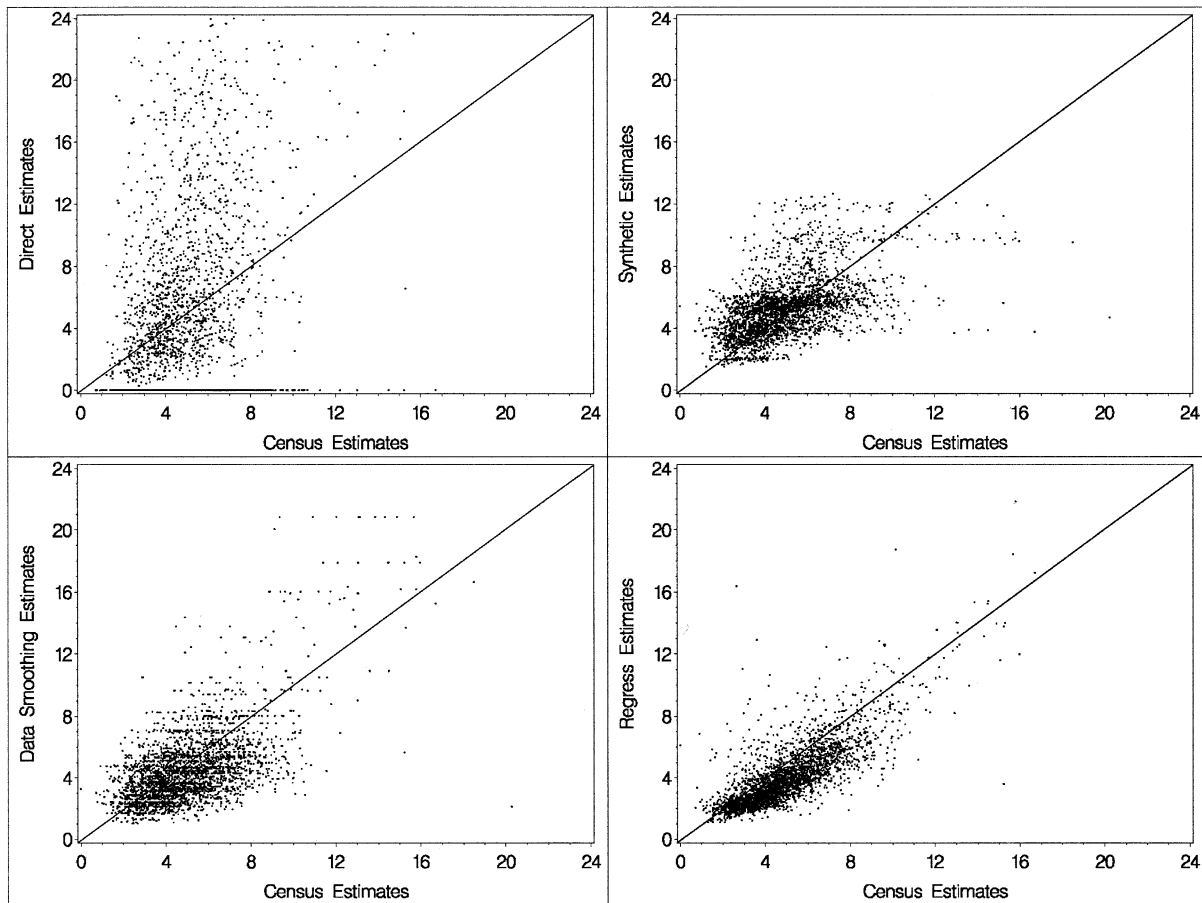
Most counties did not have enough data to directly calculate the prevalence of disability. Nearly 84% of the counties (2615 of 3117) had <50 observations, the minimum number of subjects that the CDC accepts for reporting prevalence.<sup>6</sup> Furthermore, estimations were obtained with acceptable accuracy, that is, coefficient of variation <30%, for only 66 counties (2.1%).

To validate each small-area analysis method, estimates obtained from each technique were compared to

severe work disability rates obtained from the census (Figure 2). In this figure, estimates obtained from each of the three small-area estimation methods and direct estimates obtained from the BRFSS are plotted against estimates from the census. As expected, the direct estimates were unstable. The map of direct estimates (not shown here) was scattered, with nearly half (46%) of estimates either equal to 0% or 100%.

The synthetic method improved county-level estimates greatly over direct estimation. All discrepancy statistics showed an acceptable match between synthetic estimates and census prevalence. For example, the correlation coefficient with the census rate was 0.55 relative to 0.26 for direct estimation (see Table 1). The synthetic method greatly improved MSE, MAD, MRAD, and ranked statistics values over direct estimation (Table 1; small values are better). However, as suggested by other investigators,<sup>25</sup> the synthetic method aggregated estimates to state means (Figure 2), limiting the ability of this method to show geographic patterns at the substate level.

The data smoothing method better defined within-state variation and appeared to stabilize the county-



**Figure 2.** Comparison of estimates obtained from each technique to severe work disability rates obtained from the census. (X and Y values represent percent prevalence.)

level estimates (see Figure 2). In general, the data smoothing method performed better than the synthetic method in county level estimation. For instance, its correlation with the census data increased from a coefficient of 0.55 to 0.63. However, estimates based on this method were still unstable for counties from regions with sparse BRFSS data. Because some of these unstable estimates had large errors, the smoothing method had larger MSE and MAD values than the synthetic method (Table 1).

The regression method demonstrated a significant improvement over the synthetic method and data smoothing by further defining the within-state variation and providing more stable estimates than those based on other methods. Regression estimates were much closer to census rates with less bias (Figure 2). The Pearson and Spearman rank correlation coefficients both improved to 0.80, and all discrepancy statistics showed much lower values of the MAD, MSE, MRAD, and ranked statistics.

There are some counties in the BRFSS with over 400 observations, a sample size regarded by some to be large enough to produce precise direct survey estimates.<sup>13</sup> Therefore, final estimates were further im-

proved by using a composite method,<sup>8</sup> which involved obtaining a weighted average of regression estimates and direct estimates.

The results of the composite estimate of county-level disability prevalence for all 3117 counties are presented as a map in Figure 3. The regions with the highest rates were found in the Appalachian region and the Mississippi Delta. Other regions with high prevalence rates were Northern Maine, the Southeast Coast, Upper Michigan and Northern Wisconsin, Northern California and Oregon, and some areas in the Southwest. The Midwest had the lowest severe work disability rates in the nation.

## Discussion

The regression method was found to be superior to the synthetic method and data smoothing with respect to correlation with census data and other discrepancy statistics. We also demonstrated that it can be easily deployed.

To date, the synthetic method has been the most frequently employed technique in the field of public health, perhaps due to its ease of use.<sup>9,25</sup> However, this



**Table 1.** Discrepancy statistics between 2000 BRFSS estimates and Census 2000

Discrepancy statistics	Direct	Synthetic	Smoothing	Regression
Pearson correlation coefficient	0.26	0.55	0.63	0.80
Spearman correlation coefficient	0.23	0.58	0.54	0.80
Mean Square Error (MSE): $\frac{1}{N} \sum_{i=1}^N (p_i - r_i)^2$	99.99	3.75	4.05	2.56
Mean absolute different (MAD): $\frac{1}{N} \sum_{i=1}^N  p_i - r_i $	5.63%	1.43%	1.54%	1.27%
Mean relative absolute different (MRAD): $\frac{1}{N} \sum_{i=1}^N \frac{ p_i - r_i }{r_i}$	118%	33.7%	32.6%	26.5%
Ranked statistics: $\frac{\sqrt{12}}{N} \sum_{i=1}^N \left( \frac{\text{rank}(p_i - r_i)}{N+1} - \frac{1}{2} \right) (p_i - r_i)$	7.05	1.78	1.83	1.14

BRFSS, Behavioral Risk Factor Surveillance System

method is biased and lacks specificity for tracking changes over time or examining local patterns of diseases because it applies statistics from larger areas to local areas based on demographic characteristics.

The spatial data smoothing method removes some of the background noise<sup>10,20</sup> and appears to produce more useful local area estimates, particularly for counties within regions with large sample sizes. Estimates based on smoothing have more accuracy and validity than synthetic estimates. However, when a single year of BRFSS data is used, it may not stabilize all county-level estimates for regions with a small sample size.<sup>10</sup> While aggregating data over many years improves the predictive power of such estimates for most counties, it does not permit policymakers to follow health trends or evaluate the effects of public health interventions such as public communications campaigns. Another weakness of the smoothing method is that estimates at U.S. border regions do not perform well since there are fewer neighboring counties.

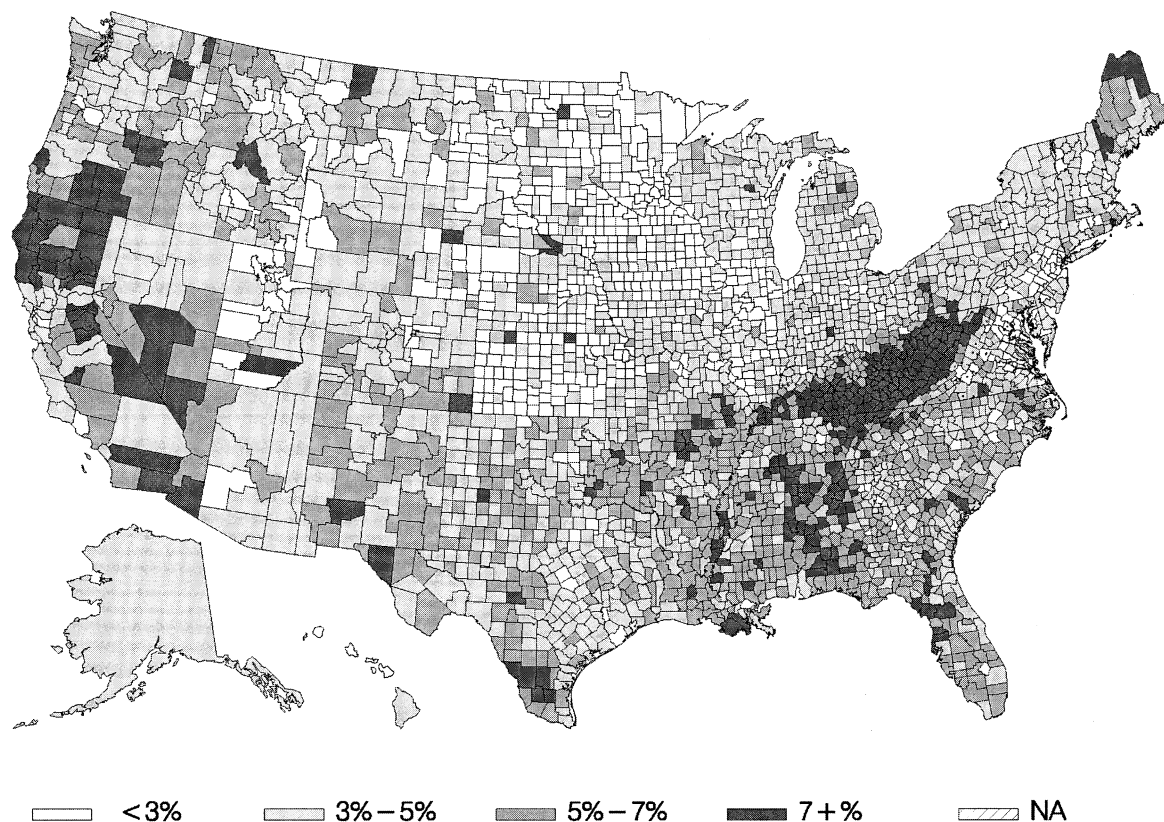
In this analysis, a multilevel model was used with one county-level random effect similar to the generalized linear mixed model by Jiang et al.,<sup>26</sup> modifying previous approaches.<sup>8,22,27</sup> Although a test of variance component rejected the null hypothesis of  $\sigma^2=0$ , estimates for random effects  $\alpha_i$  were all approximately zero and had very little impact on their corresponding estimates of county prevalence  $p_i$ . Therefore, the model inference was simplified by not estimating random effect terms and assuming them to all be equal to zero. The random effect was included in this model to obtain a more accurate estimation of variance for fixed effects.<sup>21,22,26</sup>

The main limitation of the regression model is that it requires area-level demographic data, which might be not available in all circumstances; in some areas, this requires additional demographic work during nondecennial years. The accuracy of regression estimates may also be compromised in some counties by poor prediction. Some have suggested including a spatial smoothing procedure for regression estimates given that similar measures are not typically available in larger datasets for cross validation.<sup>10</sup>

Previous analyses have shown that the prevalence of severe work disability has been relatively stable for decades in the United States.<sup>28</sup> The current results were similar to those recorded in health and demographic datasets over the past 3 decades. However, rates in local areas are much more variable. Therefore, state and local health officials need valid and relatively precise data in order to plan accordingly for rehabilitation, job training, and other programs for the disabled, to evaluate progress in meeting the needs of disabled people, for identification and prevention of causes of disability, and to allocate federal and state funds.

Model building takes time and effort. It would therefore be logical for the BRFSS group to either coordinate outside efforts examining other variables, or to conduct such analyses internally. This would ensure that efforts are not duplicated and are maximally cost-effective.

In summary, this regression-based small-area estimation procedure is relatively computationally simple and provides a significant improvement over the often used synthetic method. This study provided evidence that



**Figure 3.** Percent prevalence of severe work disability rates obtained using the regression method.

the BRFSS can be used as a valid source of annual county-level health data. Another significant contribution of this study was to demonstrate that the estimated severe work disability prevalence, at national, state, and county levels, based on the 2000 BRFSS data were similar to Census 2000 disability data. The regression method simplified using a mixed model can produce versatile and valid estimates of local health outcome measures. We anticipate that this will improve planning among local health agencies with limited resources.

We are grateful to Jiming Jiang, PhD, Joseph Sedransk, PhD, David Moriarty, and Matthew Zack, MD, for their contribution to this research. The study was supported in part by an Association of Teachers of Preventive Medicine/Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry Cooperative Agreement (TS 221-12/12), with funding provided by the National Center for Environmental Health, Office on Disability and Health.

## References

- Douglas MJ, Conway L, Gorman D, Gavin S, Hanlon P. Achieving better health through health impact assessment. *Health Bull* 2001;59:300-5.
- Paul-Shaheen P, Clark JD, Williams D. Small area analysis: a review and analysis of the North American literature. *J Health Policy Law* 1987;12:741-97.
- Remington PL, Smith MY, Willianson DF. Design, characteristics, and usefulness of state-based Behavioral Risk Factor Surveillance: 1981-87. *Public Health Rep* 1988;103:366-75.
- Frazier EL, Franks AL, Sanderson LM. Using Behavioral Risk Factor Surveillance data. In: *Using chronic disease data: a handbook for public health practitioners*. Atlanta GA: Centers for Disease Control and Prevention, 1992:4.1-4.17.
- Kim I, Keppel KG. Priority data needs: sources of national, state and local-level data and data collection systems. In: *Healthy People 2000 Statistical Notes*. Washington DC: National Center for Health Statistics, 1997.
- Centers for Disease Control and Prevention. *Surveillance summaries*. *MMWR Surveill Summ* 1993;4.
- Purcell NJ, Kish L. Estimation for small domains. *Biometrics* 1979;35:365-84.
- Ghosh M, Rao JNK. Small area estimation: an appraisal (with comments). *Stat Sci* 1994;9:55-93.
- National Center for Health Statistics. *Synthetic estimation of state health characteristics based on the Health Interview Survey*. Washington DC: U.S. Government Printing Office, 1977.
- Pickle LW, Su Y. Within-state geographic patterns of health insurance coverage and health risk factors in the United States. *Am J Prev Med* 2002;22:75-83.
- Ericksen EP. A regression method for estimating population changes of local areas. *J Am Stat Assoc* 1974;69:867-75.
- Domzal C. *Disability demographics and definitions*. Washington DC: National Institute on Disability and Rehabilitation Research, 1996.
- LaPlante MP. *State estimates of disability in America*. Washington DC: National Institute on Disability and Rehabilitation Research, 1993.
- Hennessy CH, Moriarty DG, Zack MM, Scherr PA, Brackbill R. Measuring health-related quality of life for public health surveillance. *Public Health Rep* 1994;109:665-72.
- Case Western Reserve University. Borawski E, Jia H, Wu G. The use of the Behavioral Risk Factor Surveillance System (BRFSS) to estimate the prevalence of state and substate disability. Atlanta GA: Centers for Disease Control and Prevention, 1999.
- Centers for Disease Control and Prevention. *Measuring healthy days*. Atlanta GA: Centers for Disease Control and Prevention, November 2000.

17. U.S. Census Bureau. Census 2000 data, prepared by the U.S. Census Bureau, 2002 Available at: [www.census.gov/main/cen2000.html](http://www.census.gov/main/cen2000.html). Accessed on August 19, 2003.
18. Bureau of Health Professions Office of Data Analysis and Management. Technical documentation with field numbers for the Area Resource File. Rockville MD: U.S. Department of Health and Human Services, 1991.
19. Centers for Disease Control and Prevention. Mortality file summary 1996–1998. Available at: <http://wonder.cdc.gov/mortSQL.shtml>. Accessed on August 19, 2003.
20. Mungiole M, Pickle LW, Simonson KH. Application of a weighted head-banging algorithm to mortality data maps. *Stat Med* 1999;18:3201–9.
21. Binder DA. On the variances of asymptotically normal estimators from complex surveys. *Int Stat Rev* 1983;51:279–92.
22. Jiang J. Consistent estimates in generalized linear mixed models. *J Am Stat Assoc* 1998;93:720–9.
23. Johnson RA, Wichern DW. Applied multivariate statistical analysis. Upper Saddle River NJ: Prentice Hall, 1998.
24. Hettmansperger T. Statistical inference based on ranks. New York: Wiley, 1984.
25. Levy PS, French DK. Synthetic estimation of state health characteristics based on the Health Interview Survey: vital and health statistics. Washington DC: U.S. Government Printing Office, 1977.
26. Jiang J, Jia J, Chen H. Maximum posterior estimation of random effects in generalized linear mixed models. *Stat Sinica* 2001;11:97–120.
27. Malec D, Sedransk J, Moriaty C, LeClere F. Small area inference for binary variables in the National Health Survey. *J Am Stat Assoc* 1997;92:815–26.
28. Ycas MA. Trends in the incidence and prevalence of work disability. In: Thompson-Hoffman S, Storck IF, eds. Disability in the United States: a portrait from national data. New York: Springer, 1991;161–83.