

**Small Area Estimates of Diabetes and Smoking Prevalence in North Carolina  
Counties: 1996-2002 Behavioral Risk Factor Surveillance System**

**Akhil K. Vaish\*, Neeraja Sathe, Ralph E. Folsom**  
**Research Triangle Institute, 3040 Cornwallis Road**  
**Research Triangle Park, NC 27709**  
 \*avaish@rti.org

**Key Words:** Survey Weighted Hierarchical Bayes, Pseudo Hierarchical Bayes, MCMC, Diabetes, Smoking, Behavioral Risk Factor Surveillance System (BRFSS), and North Carolina County-Level Small Area Estimates.

**1. Introduction**

This research paper summarizes the results and methodology used in producing age-adjusted small area estimates (SAEs) and associated prediction intervals (PIs) for all 100 counties in North Carolina using the 1996-2002 Behavioral Risk Factor Surveillance System (BRFSS) survey data. SAEs for two outcome variables, Lifetime diabetes and Current smoking (henceforth referred to as Diabetes and Smoking), are produced and presented in this paper. Also presented in this paper are county-level estimates of change in Diabetes and Smoking prevalence rates from 1996-1999 to 2000-2002. The county-level SAEs can be used to identify high-risk areas (counties or groups of counties) in North Carolina at which prevention and other health intervention programs can be aimed. For this purpose, data from seven consecutive (1996 to 2002) BRFSS surveys was used. A short description of the methodology used for the study is also presented here.

The county-level sample sizes are not large enough to produce design-based (direct survey weighted) estimates with acceptable precision. For example, Tyrrell County had 13 respondents in the pooled 7 years of BRFSS data. Clearly, any inference solely based on 13 observations will be highly unreliable. To obtain reliable estimates, we use small area estimation (SAE) methodology. The term “small” in SAE refers to the case when the number of area-specific sample observations is small, and the term “area” in SAE refers to any subpopulation or domain of interest such as state, county, or age by gender subgroups. A commonly used strategy is to borrow strength from other areas by postulating a model (mixed linear or nonlinear) with common fixed parameters ( $\beta$ ) and uncommon area-specific random effects (or parameters,  $\eta$ ) to pick up the area-specific differences. The model relates the outcome variable to important predictor variables (covariates). There are various techniques to estimate the model parameters. We use the Survey-Weighted Hierarchical Bayes (SWHB) methodology developed by Folsom, Shah, and Vaish (1999). Once the model parameters are estimated, SAEs can be obtained by first applying the model to the population data set containing all the predictors that are defined for all the members in the population and then aggregating the person-level estimates to the desired small-area level, using appropriate population count projections. The two outcome variables for which SAEs are produced are described below.

**2. Outcome Variables**

We have produced North Carolina county-level SAEs for Diabetes and Smoking. We chose these outcomes for two reasons. First, we assumed that the associated county-level estimates would be of interest to state policy makers. Secondly, for these two outcome variables, BRFSS data was available for all 7 years. Since the states may not cover all modules of the BRFSS every year, data for some outcomes like Past Month Use of Alcohol were not available for each year, rendering pooled estimates unattainable. To create Diabetes and Smoking outcome variables, we used the 2000 codebook of the BRFSS survey (<http://www.cdc.gov/brfss/ti-surveydata2000.htm#survey>) as a reference.

The BRFSS is administered and supported by the Behavioral Surveillance Branch of the Centers for Disease Control and Prevention (CDC). It is an ongoing data collection program designed to measure behavioral risk factors for health and related outcomes, such as Diabetes, in the adult population 18 years of age or older living in households. Data are collected from a random sample of adults (one per household) through a telephone survey. Questions can be developed by participating States and added to the core questionnaire. The creation of the binary outcome variables, Diabetes and Smoking, is described below:

**DIABETES:** We used the response for the following BRFSS question to create Diabetes:

Have you ever been told by a doctor that you have diabetes? In BRFSS data, the response was coded as follows:

Value	Label
1	Yes
2	Yes, female told only during pregnancy
3	No
7	Don't know/Not sure
9	Refused

Using the above response values, we defined Diabetes as:

Diabetes = 1 if value is 1,  
 = 0 if value is 2 or 3, and  
 = . (missing) otherwise.

**SMOKING:** To create Smoking, the following derived (calculated) variable from the BRFSS survey was used:

Smoking Status: _Smoker2	
Value	Label
1	Current-now smoke everyday
2	Current-now smoke some days
3	Former smoker
4	Never smoked
9	Don't know/Not sure or Refused

Then Smoking was defined as:

Smoking = 1 if value is 1 or 2,  
 = 0 if value is 3 or 4, and  
 = . (missing) otherwise.

We separately ratio adjusted the final BRFSS weight (\_finalwt) to account for missing values of the Diabetes and Smoking outcome variables. The weight adjustments were done within every non-missing county by age group (18-34, 35-44, 45-59 and 60 or older) by gender by race (white and non white) domain. Since we produced SAEs based on the pooled 1996-1999 and 2000-2002 BRFSS data, we divided the adjusted BRFSS weight for the pooled 1996-1999 data by 4 and for the pooled 2000-2002 pooled data by 3. We also obtained population count projections for each county by age group by gender by race domain from Claritas, Inc. for 1999 and 2002. These population count projections were used to make the final weight adjustment on the two data sets. To produce age-adjusted county-level SAEs, we developed mixed logistic regression models using the SWHB methodology. For this purpose, we used North Carolina county-level predictors that were obtained from the following sources.

### 3. Predictor Variables

The list of data sources and potential predictors used in the modeling are provided below. We refer to Exhibit 1 for a complete list of predictors.

- *Claritas Inc.*: The demographic data package called *Building Block Basic; Age by Race* for 1999 with projections to 2004 was used to model the 1996-1999 BRFSS data. *Building Block Basic; Age by Race* for 2002 with projections to 2007 was used to model the 2000-2002 BRFSS data.
- *Census Bureau*: 1990 Census (demographic and socioeconomic variables) data was used to model the 1996-1999 BRFSS data and the 2000 census data was used to model the 2000-2002 BRFSS data.
- *Health Resources and Services Administration*: Some variables related to income and employment obtained from the Area Resource File (ARF) February 2001 and February 2002 release from the Bureau of Health Professions, Office of Research and Planning, were used.
- *National Center for Health Statistics (NCHS)*: Mortality data using International Classification of Diseases, 9<sup>th</sup> revision (ICD-9), 1992-1997 and 1993-1998 were used in the 1996-1999 and 2000-2002 BRFSS data respectively. The ICD-9 death rate data was obtained from NCHS at the CDC.

All the predictors in Exhibit 1 were first appropriately transformed before using them in the modeling process described in the next section.

### Exhibit 1. List of Predictors

% Population aged 0-18 in county	% Population who dropped out of high school
% Population aged 19-24 in county	% Housing units built in 1940-1949
% Population aged 25-34 in county	% Persons 16-64 with a work disability
% Population aged 35-44 in county	% Hispanics who are Cuban
% Population aged 45-54 in county	% Females 16 years or older in labor force
% Population aged 55-64 in county	% Females never married
% Population aged 65+ in county	% Females separated/divorced/widowed/other
% Blacks in county	% One-person households
% Hispanics in county	% Female head of household, no spouse, children less than 18
% Other race in county	% Males 16 years or older in labor force
% Whites in county	% Males never married
% Males in county	% Males separated/divorced/widowed/other
% Females in county	% Housing units built in 1939 or earlier
Alcohol death rate, direct cause	Average persons per room
Alcohol death rate, indirect cause	% Families below poverty level
Cigarettes death rate, direct cause	% Households with public assistance income
Cigarettes death rate, indirect cause	% Housing units rented
Drug death rate, direct cause	% Population 9-12 years of school, no high school diploma
Drug death rate, indirect cause	% Population 0-8 years of school
% Families below poverty level	% Population with associate's degree
Unemployment rate	% Population some college and no degree
Per capita income (in thousands)	% Population with bachelor's, graduate, and professional degree
Food stamp participation rate	Median rents for rental units
Categorical variable =1 if MSA with greater than 1 million, =0 otherwise	Median household income
Categorical variable =1 if MSA with less than 1 million, =0 otherwise	Median value of owner-occupied housing units

### 4. Selection of Significant Predictor Variables

The continuous county-level predictors given above were first categorized into five-level quintile scores. These quintile scores were then transformed into linear, quadratic, and cubic trend (orthogonal) polynomials. These trend polynomials, along with the race and gender variables, were used as input to the variable selection methodology described below.

#### Step 1: Selection of Main Effects

For both the outcome variables, we independently developed four models corresponding to the four age groups. In each model, linear trend polynomials corresponding to all the predictors listed above along with race and gender were allowed to enter the model via the SAS stepwise logistic regression procedure with SLE=0.20. All predictors that were significant at the 10% level were retained and allowed to enter in the following step.

#### Step 2: Selection of Higher Order Terms

All the predictors that were significant in Step 1, their corresponding higher order polynomials (quadratic and cubic), gender, race, and the interaction of gender with race were then entered into a SAS stepwise logistic regression procedure with SLE=0.20. The predictors that were significant at the 10% level were used in the next step.

**Step 3: SUDAAN Selection**

All the predictor variables that were significant in Step 2 were then input to a SUDAAN logistic regression model. All variables that remained significant at the 5% level were then used in the SWHB methodology.

Exhibits 2 and 3 list the final set of predictors used in the SWHB modeling of Diabetes and Smoking, respectively. The “x” under the age-group indicates that the variable was significant for that particular age-group. Even if race and gender were not significant, they were forced into all our SWHB models. The SWHB methodology is described in the next section.

**Exhibit 2. List of Significant Predictors Used in SWHB Modeling of Diabetes**

Predictor	18-34	35-44	45-59	60+
Gender indicator, =1 if person is male, =0 if female	x	x	x	x
Race indicator, =1 if person is white, =0 if non-white	x	x	x	x
% Housing units built in 1940-1949	x		x	
Categorical variable, =1 if MSA with <1 million people, =0 otherwise	x			
% Population aged 65+ in county		x	x	
% Families below poverty level		x	x	
% Males 16 years or older in labor force		x		
Unemployment rate		x	x	
% Females in county			x	
% Females separated, divorced, widowed or other			x	
% Female head of household with no spouse, children less than age 18			x	
Alcohol death rate, direct cause				x
% Population aged 0-18 in county				x
% Population aged 25-34 in county				x
% Population aged 45-54 in county				x

**5. Proposed Methodology**

The SWHB methodology is being used to produce age group-specific state-level SAEs using data from various surveys including the National Survey on Drug Use and Health (NSDUH) and the National Household Travel Survey. The SWHB methodology has the following benefits over other commercially available software, such as MLwiN® and BUGS®:

1. The SAEs for large sample areas are close to their design-based analogs; hence, they are robust against model misspecification.
2. The national aggregates (aggregates of a lower level, say state-level SAEs) of the SAEs are design-consistent and,

therefore, approximately self-calibrated to the robust design-based national estimates.

3. The SAEs for large sample areas converge to a design-consistent regression-type estimator that is more precise than the corresponding survey-weighted prevalence estimator. Hence, the posterior predictive density intervals for large sample areas are typically somewhat narrower than the associated design-based confidence intervals.
4. The SWHB methodology allows the use of person or unit-level predictors in the model as opposed to the aggregate level predictors used by the popular Fay-Herriot (1979) solution. As a result, SAEs based on the SWHB method are internally consistent and also more precise than solutions obtained from aggregate-level small area models.

Recently You and Rao (2003), developed a pseudo-Hierarchical Bayes (PHB) SAE methodology. The PHB and SWHB methodologies are similar for the linear mixed model case. The SWHB methodology is a nonlinear extension of the PHB methodology. A brief model description is provided below.

**Exhibit 3. List of Significant Predictors Used in SWHB Modeling of Smoking**

Predictor	18-34	35-44	45-59	60+
Gender indicator, =1 if person is male, =0 if female	x	x	x	x
Race indicator, =1 if person is white, =0 if non-white	x	x	x	x
Interaction of gender and race	x	x	x	x
Alcohol death rate, direct cause	x			
Cigarettes death rate, direct cause	x			
% Population aged 35-44 in county	x			
% Population aged 25-34 in county		x		x
% Population aged 45-54 in county		x		
Categorical variable, =1 if MSA with >1 million people, =0 otherwise	x			
Categorical variable, =1 if MSA with <1 million, =0 otherwise		x		
% Population with bachelor's, graduate, professional degree	x		x	x
% Males never married		x		
% Males separated, divorced, widowed or other			x	
Unemployment rate			x	

**6. Model Description**

We used the following mixed logistic regression model to produce SAEs for all the counties in North Carolina:

$$\log \left[ \frac{\mu_{aijk}}{1 - \mu_{aijk}} \right] = x_{aijk} \beta_a + \eta_{ai} + \nu_{ij}$$

where  $\mu_{aijk}$  denotes the probability of engaging in the behavior of interest (e.g., smoked cigarettes recently) for age group- $a$  ( $a=1$  to 4), survey respondent- $k$  residing in county- $j$  ( $j=1$  to 100) of

region- $i$  ( $i=1$  to 11). The above model has the following characteristics:

- We assume that the person-level dichotomous (1/0) outcome variable has a Bernoulli distribution with probability  $(y_{aijk}=1) = \mu_{aijk}$ .
- Note that the vector  $x_{aijk}$  of fixed predictors is allowed to vary across the four age groups. Also, the corresponding regression coefficients  $\beta_a$  vary with age group.
- The region-level random effects  $\eta_{ai}$  are also age group specific and are assumed to have a four variate normal distribution with a general covariance matrix  $D_\eta$ .
- The county-level random effects  $v_{ij}$  do not vary by age group since the county-level sample sizes were not large enough to estimate age-specific county-level random effects. The county-level random effects were assumed to be normally distributed with common variance  $\sigma_v^2$ .

The advantage of a mixed model over a fixed effect model for SAE is that the mixed model with region- and county-level random effects allows the age-group effects to vary across regions and provides for separate county-level intercepts.

### 7. Estimation of SWHB Model Parameters

Estimation of the model parameters  $(\beta_a, \eta_{ai}, v_{ij}, D_\eta, \sigma_v^2)$  is not straightforward. A series of iterative steps is employed to generate posterior sample values of the desired fixed and random effect parameters from their underlying joint posterior distribution using the GIBBS sampling algorithm. We generated 10,000 Markov Chain Monte Carlo (MCMC) samples for each of the model parameters (fixed effects, random effects, and the

<sup>1</sup> Note: 11 regions were created for modeling purposes comprising 2 to 28 counties:

- Region 1:** Buncombe, Gaston, and Mecklenburg counties.
- Region 2:** Cumberland, New Hanover, and Onslow counties.
- Region 3:** Durham and Wake counties.
- Region 4:** Forsyth and Guilford counties.
- Region 5:** Cherokee, Clay, Graham, Haywood, Henderson, Jackson, Macon, Madison, Polk, Swain, and Transylvania counties.
- Region 6:** Alleghany, Ashe, Avery, Burke, Caldwell, McDowell, Mitchell, Rutherford, Surry, Watauga, Wiles, and Yancey counties.
- Region 7:** Chatham, Franklin, Granville, Lee, Orange, Person, Vance, and Warren counties.
- Region 8:** Alexander, Cabarrus, Catawba, Cleveland, Iredell, Lincoln, Rowan, Stanly, and Union counties.
- Region 9:** Alamance, Anson, Caswell, Davidson, Davie, Montgomery, Moore, Randolph, Richmond, Rockingham, Stokes, and Yadkin counties.
- Region 10:** Bladen, Brunswick, Columbus, Harnett, Hoke, Johnston, Pender, Robeson, Sampson, and Scotland counties.
- Region 11:** Beaufort, Bertie, Camden, Carteret, Chowan, Craven, Currituck, Dare, Duplin, Edgecombe, Gates, Greene, Halifax, Hertford, Hyde, Jones, Lenoir, Martin, Nash, Northampton, Pamlico, Pasquotank, Perquimans, Pitt, Tyrell, Washington, Wayne, and Wilson counties.

associated variance-covariance matrices) after deleting a sufficient number of burn-in samples. Every 8<sup>th</sup> replicate was selected, yielding a total of 1,250 independent samples from the joint posterior distribution of the parameters. The selected set of 1,250 parameter vectors was then used to produce predicted prevalence estimates for every North Carolina county.

### 8. Production of County by Age Group Small Area Estimates

For age group- $a$ , let  $x_{a1}, x_{a2}, \dots, x_{aq}$  denote the  $q$  fixed predictors and  $\beta_{a1r}, \beta_{a2r}, \dots, \beta_{aqr}$  the associated parameter estimates from the  $r^{\text{th}}$  MCMC sample. Now, let  $\eta_{air}$  and  $v_{ijr}$  denote the region- and county-level random-effect estimates from the  $r^{\text{th}}$  MCMC sample, respectively. Let  $n_{aijeg}$  denote the population count projection for county- $j$  of region- $i$  for age group- $a$ , race- $e$  ( $e=1,2$ ), and gender- $g$  ( $g=1,2$ ). The prevalence estimate  $p_{aijegr}$  for persons in race by gender domain- $eg$  of age group- $a$  of county- $j$  of region- $i$  for the  $r^{\text{th}}$  MCMC sample is thus given by  $p_{aijegr} = [1 + \exp(-\lambda_{aijegr})]^{-1}$ , where  $\lambda_{aijegr} = \left[ \sum_{m=1}^q x_{am}(ijeg) \beta_{amr} + \eta_{air} + v_{ijr} \right]$  with  $x_{am}(ijeg)$  denoting the value of the  $m^{\text{th}}$  fixed predictor for persons in domain- $eg$  of age group- $a$  in county- $j$  of region- $i$ . The SWHB estimator for  $\mu_{aij}$  can be formed from the MCMC replicate- $r$  estimates, as given below:

$$p_{aijr} = \frac{\sum_{e=1}^2 \sum_{g=1}^2 p_{aijegr} n_{aijeg}}{\sum_{e=1}^2 \sum_{g=1}^2 n_{aijeg}}$$

To find the SWHB prevalence estimates for county- $j$  of region- $i$  and age group- $a$ , we calculate the average of  $p_{aijr}$  over the chosen MCMC replicates. After running a burn-in of 500 cycles, we thinned our Gibbs sample of 10,000 MCMC replicates to 1,250 by taking every 8<sup>th</sup> replicate. The corresponding 95% prediction or Bayes credible intervals, based on 1,250 MCMC replicates, are obtained in the following manner. Let

$$l_{aijr} = \log \left( \frac{p_{aijr}}{1 - p_{aijr}} \right), \quad \bar{l}_{aij} = \frac{1}{1,250} \sum_{r=1}^{1,250} l_{aijr}, \quad \text{and}$$

$$\sigma_{aij}^2 = \frac{1}{1,250} \sum_{r=1}^{1,250} (l_{aijr} - \bar{l}_{aij})^2;$$

then the lower bound  $L$  and the upper bound  $U$  of the 95% prediction interval (PI) are given by

$$L = [1 + \exp\{-\bar{l}_{aij} - 1.96 \sigma_{aij}\}]^{-1} \quad \text{and}$$

$$U = [1 + \exp\{-\bar{l}_{aij} + 1.96 \sigma_{aij}\}]^{-1}.$$

To improve the precision of county estimates and associated measures of change over time, 3 to 4 consecutive years of BRFSS data were pooled. We pooled the 1996-1999 BRFSS data to produce one set of SAEs ( $p(96\_99)$ ) and pooled 2000-2002

BRFSS data to produce the other set of SAEs ( $p(00\_02)$ ). Estimates of change (odds ratios) between the two periods are also presented here.

## 9. Estimation of Change

The change between the  $p(96\_99)$  and  $p(00\_02)$  is defined in terms of the log-odds-ratio (*lor*) as opposed to the simple difference since the *lor* has better asymptotic Gaussian distributional properties than the simple difference. The *lor* is defined as

$$lor = \log \left[ \frac{p(00\_02)/[1-p(00\_02)]}{p(96\_99)/[1-p(96\_99)]} \right].$$

The SAEs  $p(96\_99)$ ,  $p(00\_02)$ , and *lor* are obtained by simultaneously fitting the 1996-1999 and 2000-2002 data to produce the required number of MCMC samples from the joint posterior distribution of  $p(96\_99)$  and  $p(00\_02)$ . For this simultaneous model, there are 8 (4 age groups by 2 time periods) subpopulation-specific models with their own sets of fixed and random effects. Note that the survey-weighted Bernoulli-type log likelihood employed in the SWHB methodology is appropriate for this simultaneous model because the 8 age group by time period subpopulations are non-overlapping. The 95% PI for the odds ratio change measure is computed as  $\left[ e^{lor-1.96*\sqrt{\text{var}(lor)}}, e^{lor+1.96*\sqrt{\text{var}(lor)}} \right]$ , where *lor* is calculated using the means of the posterior distributions of  $p(96\_99)$  and  $p(00\_02)$ , and  $\text{var}(lor)$  is obtained by using MCMC samples from the posterior distribution of *lor*.

## 10. Age-Adjusted Small Area Estimates

We combined the age group-specific county-level SAEs using the 1999 state-level North Carolina (NC) population counts to produce the 18 or older age-adjusted SAEs for Diabetes and Smoking prevalence rates. We also produced change estimates between the 1996-1999 and 2000-2002 age-adjusted prevalence rates for Diabetes and Smoking. We used the 1999 state-level age distribution of NC population counts as a reference for comparison purposes since 1999 was the mid point for the 1996-2002 time period.

Table 1 presents 1996-1999, 2000-2002 SAEs for Diabetes and the corresponding change estimates in terms of the odds ratios. Table 2 presents the similar information for Smoking. The numbers in the parenthesis show the 95% PIs.

Age-adjusted county-level SAEs can best be represented on a North Carolina map. Figures 1-4 are pictorial representations of the 1996-1999 and 2000-2002 SAEs for Diabetes and Smoking. For pictorial representation, all the counties are ranked using their prevalence rates and five quintiles (five groups each containing approximately 20 counties) are formed. The top quintile (1<sup>st</sup> quintile) represents counties with the highest prevalence rates and is shown in red on the North Carolina map.

The bottom quintile (5<sup>th</sup> quintile) represents counties with the lowest prevalence rates and is shown in white on the North Carolina map.

## 11. Discussion of Findings

### 11.a Diabetes

The age-adjusted NC Diabetes rate (SAE) has increased from 5.57% in 1996-1999 to 6.94% in 2000-2002 (see Table 1). The corresponding change estimate in terms of odds ratio was 1.26 and it was significant at the 5% level of significance. In 1996-1999, Edgecombe County had the highest prevalence rate (10.46%) for Diabetes. In 2000-2002 its prevalence rate rose to 12.29% but its ranking fell to the second place behind Gates County (12.47%). Henderson reported the smallest prevalence rate for Diabetes both in 1996-1999 (3.38%) and 2000-2002 (3.88%).

Among the top-quintile counties in 1996-1999<sup>2</sup> and 2000-2002<sup>3</sup>, twelve counties (Bertie, Duplin, Edgecombe, Gates, Halifax, Hertford, Hoke, Martin, Mitchell, Northampton, Robeson, and Wilson) were common. Out of these twelve counties, eleven are in the Eastern region. Most of the Eastern region counties in 1996-1999 and 2000-2002 were in either red or orange colors (see Figures 1 and 2), indicating higher prevalence rates of Diabetes in the Eastern region. Most of the counties with low prevalence rates (shown in either white or beige in Figures 1 and 2) for Diabetes in both the time periods belonged to the Piedmont and Western regions. The Eastern, Piedmont, and Western regions are defined in Gizlice et al. (2003). Among the bottom-quintile counties in 1996-1999<sup>4</sup> and 2000-2002<sup>5</sup>, ten counties (Buncombe, Chatham, Durham, Forsyth, Henderson, Iredell, Macon, Moore, Transylvania, and Union) were common.

While the state-level Diabetes prevalence rate rose significantly, most of the county-level rates did not change significantly between the two periods. In Table 1, most of the 95% PIs for odds ratio overlap 1, indicating that there was no significant change between the 1996-1999 and 2000-2002 Diabetes prevalence rates. Only eleven counties (Chowan, Davie, Gaston, Guilford, Harnett, Johnston, Lee, Pitt, Sampson, Scotland, and Watauga) showed a significant increase (lower bounds of PIs were greater than 1) in Diabetes prevalence rates between 1996-

<sup>2</sup> Top-quintile counties for Diabetes in 1996-1999: Edgecombe, Gates, Montgomery, Robeson, Northampton, Graham, Halifax, Mitchell, Hertford, Bertie, Camden, Hoke, Hyde, Wilson, Caswell, Yancey, Dare, Martin, Anson, and Duplin.

<sup>3</sup> Top-quintile counties for Diabetes in 2000-2002: Gates, Edgecombe, Robeson, Hoke, Hertford, Pitt, Bertie, Wilson, Scotland, Duplin, Northampton, Gaston, Halifax, Greene, Martin, Mitchell, Washington, Vance, Alleghany, and Cabarrus.

<sup>4</sup> Bottom-quintile counties for Diabetes in 1996-1999: Currituck, Rowan, Orange, Chowan, Buncombe, Durham, Davie, Macon, Mecklenburg, Forsyth, Chatham, Lee, Johnston, Union, Transylvania, Moore, Watauga, Polk, Iredell, and Henderson.

<sup>5</sup> Bottom-quintile counties for Diabetes in 2000-2002: Alexander, Rockingham, Rutherford, Ashe, Wake, Perquimans, Forsyth, Catawba, Buncombe, Dare, Iredell, Chatham, Durham, Union, Transylvania, Jackson, Haywood, Moore, Macon, and Henderson.

1999 and 2000-2002. Almost all of these eleven counties (except Guilford) belonged to the top 20 group of counties with the highest odds ratio values. Diabetes rates for Lee County increased the most (odds ratio=1.83), from 4.11% in 1996-1999 to 7.26% in 2000-2002. Though Dare County had the lowest odds ratio (0.67), indicating a decrease in prevalence rate from 7.61% in 1996-1999 to 5.26% in 2000-2002, the change was not statistically significant.

**11.b Smoking**

There was a slight increase in the Smoking rate in NC from 25.27% in 1996-1999 to 26.11% in 2000-2002, yielding an odds ratio increase of 1.05 (see Table 2). However, this increase was not statistically significant since the 95% PI of the odds ratio included 1.

Referring to the Table 2, we see that there were ten counties (Anson, Cherokee, Montgomery, Onslow, Richmond, Rockingham, Stokes, Surry, Wayne, and Yadkin) that remained in the top quintile in 1996-1999<sup>6</sup> and 2000-2002<sup>7</sup>, whereas eleven counties (Camden, Chatham, Durham, Halifax, Iredell, Lenoir, Mecklenburg, Nash, Orange, Pasquotank, and Wake) remained in the bottom quintile in 1996-1999<sup>8</sup> and 2000-2002<sup>9</sup> (Figures 3 and 4). Notably, Durham, Mecklenburg, Orange, and Wake counties had some of the lowest Smoking rates in 1996-1999 and 2000-2002. Surry County had the highest Smoking rate (32.06%) in 1996-1999, whereas Robeson County had the highest Smoking rate (34.45%) in 2000-2002. Robeson County was one of the three counties for which the Smoking rate increased significantly (odds ratio=1.34) from 1996-1999 to 2000-2002 (see Table 2). The other two counties were Brunswick (odds ratio=1.34) and Craven (odds ratio=1.27). Granville County's Smoking rate decreased the most (odds ratio=0.77), but the decrease was not statistically significant at the 5% level.

Though most of the changes in Smoking rates were not statistically significant, it is interesting to note that most of the high odds-ratio counties belonged to the Eastern and Western regions indicating an increase in smoking rates in those regions. Counties in the Piedmont region tended to show decreases in smoking rates.

<sup>6</sup> Top-quintile counties for Smoking in 1996-1999: Surry, Stokes, Warren, Gaston, Rowan, Anson, Richmond, Davidson, Rockingham, Wayne, Yadkin, Onslow, Randolph, Person, Montgomery, Alleghany, Vance, Cherokee, and Franklin.

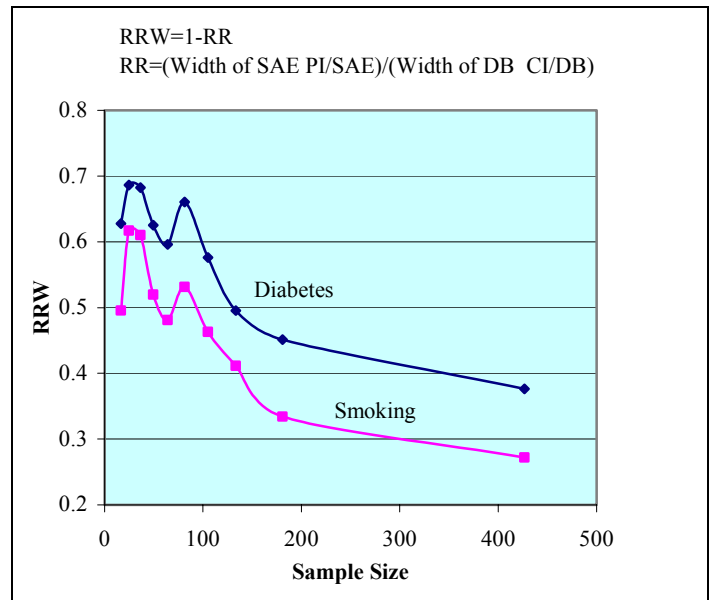
<sup>7</sup> Top-quintile counties for Smoking in 2000-2002: Robeson, Cherokee, Wayne, Hoke, Brunswick, Stokes, Graham, Yadkin, Rockingham, Onslow, Surry, Sampson, Montgomery, Alamance, Alexander, Polk, Yancey, Richmond, Clay, and Anson.

<sup>8</sup> Bottom-quintile counties for Smoking in 1996-1999: Catawba, Dare, Harnett, Iredell, Henderson, Halifax, Watauga, Camden, Lenoir, Carteret, Pasquotank, Nash, Chatham, Wilson, Pitt, Craven, Orange, Wake, Mecklenburg, and Durham.

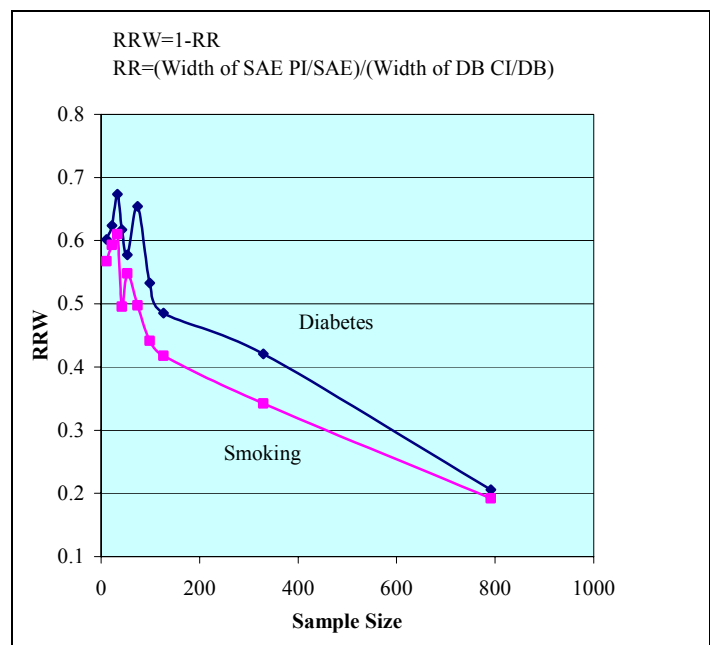
<sup>9</sup> Bottom-quintile counties for Smoking in 2000-2002: Franklin, Bertie, Cabarrus, Lincoln, New Hanover, Forsyth, Nash, Avery, Guilford, Halifax, Lenoir, Chatham, Camden, Pasquotank, Iredell, Orange, Durham, Granville, Wake, and Mecklenburg.

The following Exhibits 4 and 5 compare the relative widths (RRWs) of SAE PIs with the design based confidence intervals (CIs) for Diabetes and Smoking prevalence rates for 1996-1999 and 2000-2002 time periods. For this purpose reduction in relative width (RRW) (defined below) was calculated for all the 100 counties. We then formed 10 groups using county-level sample sizes and for each group we calculated the average sample size and average RRW. For small sample sizes (<50-100), design based estimates are highly unreliable and this is depicted by the ups and downs in the following exhibits. For moderately large sample sizes (around 300), SAE PIs for Diabetes and Smoking are approximately 45% and 35% narrower than the design based CIs, respectively.

**Exhibit 4: Reduction in Relative Width (RRW) : 1996-1999**



**Exhibit 5: Reduction in Relative Width (RRW) : 2000-2002**



## 12. Conclusion and Future Research

We have shown that small area estimation can be a useful and effective technique in producing reliable county or sub-state-level estimates. For over-sampled counties and for a low prevalence outcome variable such as Diabetes, the SAE PI relative widths were, on average, 45% narrower than the corresponding design-based CI widths, whereas for Smoking, the reduction was around 35%. Also, this research does not address any issues with the survey data quality such as self-reporting, telephone survey coverage, etc. A summary of the important issues related to this study and future enhancements is given below:

- **Pooling of BRFSS Data:** Even though BRFSS sample sizes are increasing steadily, the demand for more local-area (such as city/county) information is also increasing. From our past experiences, we need at least 275 respondents to produce reliable SAEs. There are around 3,000 counties in the United States. This means that we need 825,000 respondents to produce reliable county-level SAEs. Clearly, due to cost and other considerations, this is not a viable option. To circumvent this problem, 3-4 years of BRFSS data can be pooled to produce reliable county-level SAEs.
- **Outcome Variables:** Even though data can be pooled across years, some of the modules such as alcohol consumption are not included in the core section of the BRFSS questionnaire every year. This makes it impossible to produce reliable county-level estimates for some of the important outcome variables such as past month use of alcohol. Across seven years of NC BRFSS data from 1996-2002, Diabetes and Smoking were the only two important health-related outcome variables of interest that were uniformly available.
- **Predictors:** The main emphasis was to develop a model with high prediction power. Less attention was paid to the substantive nature of the predictors. In an ideal situation, an initial set of predictors would be chosen in consultation with an epidemiologist with substantive expertise. In a preliminary study, conducted earlier using the 1996-2000 BRFSS data, the initial set of predictors also included the local area predictors obtained from the North Carolina State Center for Health Statistics (LINC Data). However, at the end of the variable selection process, almost all of the local area predictors had dropped out in the presence of predictors obtained from the national data sources mentioned in Section 3.
- **Validation:** Gizlice et al. (2003) obtained synthetic and age-adjusted design-based estimates for Diabetes and Smoking prevalence rates using the 2001 NC BRFSS data for the 13 large sample areas (10 over-sampled counties and three regions consisting of the rest of the 90 counties). It is well known that synthetic estimates (a subclass of model-based estimates) are substantially biased and have unjustifiably low variances, whereas design-based estimates are almost unbiased but have large variances compared to the synthetic estimates. The SAEs are composite estimates (i.e., an optimal combination of model-based and design-based estimates) which have lower biases than synthetic estimates and also have lower variances than design-based estimates. Under the

assumed model, SAEs are even unbiased. Though SAEs (point estimates) for the 10 over-sampled counties compared reasonably well with the estimates given in Gizlice et al. (2003), there is a need for a comprehensive validation study to judge the quality of SAEs for small sample counties. This is not an easy task due to the unavailability of reliable benchmarked county-level estimates. One way to overcome this problem is to pool 10 or more years of data and produce design-based estimates for 5-10 large sample counties. The design-based estimates for these counties can then be used for benchmarking purposes. One can take those benchmarked counties and randomly partition their sample into 8-10 small pseudo counties mimicking the annual BRFSS design for small sample counties. The SAEs for these small sample pseudo counties can now be obtained and compared with the corresponding benchmarked estimates and a measure of reliability can be constructed. This reliability measure can be used to judge the quality of SAEs for small sample counties.

**Acknowledgement:** Research supported by a CDC (NCHS) grant to the UNC Center for Health Statistics Research, Chapel Hill, NC. Work involved in preparing this paper was partially funded by contract number 200-98-0103 T.O. #3 (awarded by the Behavioral Sciences Branch of the Centers for Disease Control and Prevention, Department of Health and Human Services).

### References:

- Fay, R. E and R. A. Herriot (1979). Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, 74, 269-277.
- Folsom, R. E., B. Shah, and A. Vaish (1999). Substance Abuse in States: A Methodological Report on Model Based Estimates from the 1994-1996 National Household Surveys on Drug Abuse, *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 371-375.
- Gizlice, Z., H. Herrick, and K. Fultz-Butts (2003). Geographic Variation in Health Indicators across North Carolina Counties and Regions: A Comparison of Direct and Synthetic Estimates from the North Carolina BRFSS, SCHS Studies No. 137, State Center for Health Statistics, Raleigh, NC.
- You, Y. and J. N. K. Rao (2003). Pseudo Hierarchical Bayes Small Area Estimation Combining Unit Level Models and Survey Weights, *Journal of Statistical Planning and Inference*, 111, 197-208.

ASA Section on Survey Research Methods

**Table 1. SAEs, Odds-Ratio, and Corresponding 95% PIs for Diabetes**

County	Small Area Estimates		Change (Odds-Ratio)	County	Small Area Estimates		Change (Odds-Ratio)
	1996-1999	2000-2002			1996-1999	2000-2002	
North Carolina *	5.57 (5.11-6.08)	6.94 (6.43-7.47)	1.26 (1.12-1.43)				
Alamance	6.41 (4.65-8.78)	6.49 (4.69-8.92)	1.01 (0.70-1.46)	Johnston *	4.06 (2.77-5.90)	6.79 (4.77-9.58)	1.72 (1.13-2.62)
Alexander	7.12 (4.75-10.53)	5.83 (3.88-8.67)	0.81 (0.51-1.28)	Jones	6.91 (4.68-10.08)	8.08 (5.64-11.46)	1.18 (0.85-1.64)
Alleghany	6.05 (3.92-9.23)	9.18 (6.01-13.78)	1.57 (0.96-2.56)	Lee *	4.11 (2.70-6.20)	7.26 (4.81-10.80)	1.83 (1.16-2.86)
Anson	7.42 (4.98-10.93)	8.43 (5.64-12.40)	1.15 (0.75-1.76)	Lenoir	6.64 (4.80-9.12)	8.01 (5.77-11.03)	1.22 (0.90-1.67)
Ashe	4.99 (3.25-7.60)	5.77 (3.76-8.73)	1.16 (0.75-1.82)	Lincoln	6.11 (4.13-8.96)	7.41 (5.07-10.72)	1.23 (0.78-1.94)
Avery	5.79 (3.75-8.84)	6.98 (4.53-10.62)	1.22 (0.74-2.01)	Macon	4.41 (2.92-6.61)	4.02 (2.69-5.95)	0.91 (0.60-1.38)
Beaufort	6.2 (4.41-8.65)	7.94 (5.64-11.06)	1.3 (0.93-1.83)	Madison	6.11 (3.87-9.54)	5.84 (3.83-8.80)	0.95 (0.58-1.56)
Bertie	8.56 (6.08-11.93)	10.31 (7.39-14.19)	1.23 (0.89-1.70)	Martin	7.43 (5.34-10.24)	9.48 (6.86-12.97)	1.31 (0.98-1.75)
Bladen	6.86 (4.78-9.76)	8.5 (5.93-12.03)	1.26 (0.87-1.82)	McDowell	6.32 (4.26-9.30)	7.52 (5.26-10.65)	1.21 (0.80-1.82)
Brunswick	5.56 (3.75-8.17)	6.82 (4.58-10.05)	1.24 (0.76-2.02)	Mecklenburg	4.36 (3.31-5.72)	5.94 (4.70-7.49)	1.39 (0.98-1.96)
Buncombe	4.63 (3.28-6.48)	5.31 (3.91-7.19)	1.16 (0.77-1.74)	Mitchell	8.6 (5.44-13.35)	9.4 (6.44-13.51)	1.1 (0.68-1.80)
Burke	6.67 (4.60-9.59)	7.8 (5.63-10.70)	1.18 (0.79-1.78)	Montgomery	9.39 (6.41-13.55)	8.68 (5.86-12.68)	0.92 (0.60-1.40)
Cabarrus	6.62 (4.73-9.20)	9.18 (6.62-12.60)	1.43 (0.96-2.11)	Moore	3.86 (2.48-5.97)	4.17 (2.66-6.49)	1.08 (0.65-1.80)
Caldwell	7.26 (4.98-10.47)	8.25 (5.74-11.72)	1.15 (0.71-1.84)	Nash	6.59 (4.87-8.87)	8.5 (6.27-11.41)	1.32 (0.97-1.78)
Camden	8.27 (5.73-11.81)	7.68 (5.25-11.08)	0.92 (0.66-1.30)	New Hanover	5.82 (4.18-8.04)	6.24 (4.45-8.68)	1.08 (0.72-1.62)
Carteret	4.95 (3.45-7.06)	6.25 (4.15-9.32)	1.28 (0.84-1.96)	Northampton	9 (6.33-12.64)	9.94 (7.03-13.87)	1.12 (0.80-1.57)
Caswell	7.67 (5.16-11.26)	7.31 (4.80-10.98)	0.95 (0.60-1.49)	Onslow	5.43 (3.56-8.19)	6.37 (4.34-9.26)	1.19 (0.73-1.92)
Catawba	5.38 (3.77-6.71)	5.41 (3.79-7.67)	1.01 (0.68-1.49)	Orange	4.67 (3.15-6.88)	6.29 (4.11-9.49)	1.37 (0.84-2.23)
Chatham	4.3 (2.79-6.58)	5.13 (3.23-8.05)	1.2 (0.72-2.00)	Pamlico	5.56 (3.61-8.47)	7.31 (4.81-10.96)	1.34 (0.86-2.08)
Cherokee	6.77 (4.35-10.39)	6.77 (4.36-10.36)	1 (0.61-1.64)	Pasquotank	6 (4.09-8.73)	7.88 (5.36-11.43)	1.34 (0.92-1.95)
Chowan *	4.65 (3.03-7.07)	7.86 (5.49-11.15)	1.75 (1.20-2.56)	Pender	5.46 (3.75-7.89)	6.8 (4.66-9.82)	1.26 (0.85-1.88)
Clay	6.91 (4.18-11.21)	6.03 (3.88-9.27)	0.86 (0.50-1.49)	Perquimans	5.01 (3.24-7.67)	5.62 (3.75-8.35)	1.13 (0.73-1.74)
Cleveland	6.18 (4.38-8.64)	5.99 (4.14-8.61)	0.97 (0.65-1.44)	Person	6.2 (4.08-9.31)	6.54 (4.29-9.84)	1.06 (0.68-1.66)
Columbus	6.99 (4.85-9.99)	8 (5.50-11.50)	1.16 (0.77-1.74)	Pitt *	7.17 (5.15-9.91)	10.41 (7.76-13.83)	1.5 (1.03-2.20)
Craven	6.61 (4.73-9.16)	6.74 (4.68-9.62)	1.02 (0.71-1.47)	Polk	3.82 (2.37-6.12)	5.9 (3.79-9.07)	1.58 (0.94-2.63)
Cumberland	5.96 (4.36-8.11)	7.8 (5.86-10.32)	1.33 (0.90-1.98)	Randolph	5.49 (4.00-7.48)	6.56 (4.71-9.05)	1.21 (0.85-1.71)
Currituck	4.8 (3.32-6.89)	6.2 (4.23-8.98)	1.31 (0.94-1.82)	Richmond	6.33 (4.22-9.40)	8.26 (5.61-12.02)	1.33 (0.84-2.11)
Dare	7.61 (4.78-11.90)	5.26 (3.31-8.26)	0.67 (0.39-1.18)	Robeson	9.15 (6.67-12.42)	11.95 (8.98-15.74)	1.35 (0.92-1.98)
Davidson	5.53 (4.00-7.60)	7.2 (5.23-9.84)	1.33 (0.92-1.91)	Rockingham	6.92 (4.93-9.62)	5.82 (4.10-8.20)	0.83 (0.57-1.23)
Davie *	4.42 (2.97-6.52)	7.11 (4.53-10.98)	1.66 (1.08-2.53)	Rowan	4.68 (3.18-6.86)	6.99 (4.92-9.84)	1.53 (1.00-2.35)
Duplin	7.36 (5.13-10.45)	10.03 (7.10-13.98)	1.4 (0.98-2.00)	Rutherford	5.95 (3.95-8.86)	5.79 (3.92-8.48)	0.97 (0.62-1.52)
Durham	4.55 (3.13-6.57)	5.11 (3.56-7.27)	1.13 (0.72-1.77)	Sampson *	5.65 (3.92-8.06)	8.74 (6.27-12.06)	1.6 (1.12-2.30)
Edgecombe	10.46 (7.52-14.37)	12.29 (8.92-16.70)	1.2 (0.85-1.69)	Scotland *	6.67 (4.50-9.78)	10.27 (7.18-14.50)	1.6 (1.05-2.45)
Forsyth	4.33 (3.27-5.72)	5.44 (4.21-7.02)	1.27 (0.90-1.79)	Stanly	6.35 (4.33-9.23)	7.75 (5.31-11.17)	1.24 (0.80-1.92)
Franklin	5.55 (3.72-8.20)	7.94 (5.39-11.53)	1.47 (0.94-2.30)	Stokes	6.11 (4.16-8.88)	9.17 (6.17-13.42)	1.55 (1.00-2.41)
Gaston *	6.23 (4.61-8.38)	9.81 (7.45-12.82)	1.64 (1.16-2.32)	Surry	6.05 (4.13-8.79)	8.37 (6.00-11.57)	1.42 (0.91-2.20)
Gates	10.43 (6.63-16.03)	12.47 (8.45-18.04)	1.22 (0.73-2.05)	Swain	6.35 (4.16-9.59)	8.9 (5.95-13.10)	1.44 (0.91-2.28)
Graham	8.9 (5.80-13.44)	8.65 (5.56-13.20)	0.97 (0.57-1.64)	Transylvania	3.95 (2.47-6.24)	4.89 (3.06-7.71)	1.25 (0.75-2.09)
Granville	5.74 (3.65-8.92)	8.21 (5.26-12.60)	1.47 (0.88-2.46)	Tyrrell	6.08 (4.06-9.00)	7.08 (4.57-10.82)	1.18 (0.78-1.78)
Greene	7.15 (4.79-10.52)	9.58 (6.67-13.56)	1.38 (0.95-2.00)	Union	4.05 (2.58-6.29)	5.08 (3.47-7.38)	1.27 (0.80-2.00)
Guilford *	5.28 (4.04-6.88)	7.43 (5.95-9.23)	1.44 (1.04-1.99)	Vance	6.94 (4.61-10.34)	9.3 (6.33-13.47)	1.37 (0.87-2.18)
Halifax	8.63 (6.30-11.71)	9.78 (7.10-13.31)	1.15 (0.83-1.59)	Wake	5.03 (3.77-6.67)	5.69 (4.39-7.34)	1.14 (0.79-1.65)
Harnett *	5.4 (3.77-7.68)	9.07 (6.44-12.64)	1.75 (1.18-2.59)	Warren	5.57 (3.49-8.77)	7.35 (4.66-11.40)	1.35 (0.77-2.34)
Haywood	5.76 (3.89-8.45)	4.57 (3.13-6.61)	0.78 (0.51-1.21)	Washington	6.83 (4.55-10.14)	9.33 (6.32-13.55)	1.4 (0.94-2.08)
Henderson	3.38 (2.17-5.22)	3.88 (2.54-5.89)	1.15 (0.69-1.94)	Watauga *	3.83 (2.43-5.99)	6.35 (4.12-9.65)	1.7 (1.03-2.81)
Hertford	8.57 (5.95-12.19)	10.59 (7.64-14.50)	1.26 (0.90-1.77)	Wayne	6.84 (4.91-9.47)	8.88 (6.56-11.92)	1.33 (0.96-1.84)
Hoke	7.93 (5.29-11.73)	11.4 (7.91-16.17)	1.49 (0.98-2.27)	Wilkes	4.82 (3.22-7.17)	7.06 (5.03-9.83)	1.5 (0.98-2.29)
Hyde	7.93 (5.42-11.45)	7.41 (4.93-10.99)	0.93 (0.62-1.40)	Wilson	7.79 (5.79-10.41)	10.29 (7.60-13.80)	1.36 (1.00-1.85)
Iredell	3.47 (2.25-5.33)	5.26 (3.57-7.69)	1.54 (1.00-2.39)	Yadkin	5.44 (3.50-8.36)	7.45 (4.97-11.04)	1.4 (0.87-2.26)
Jackson	5.71 (3.70-8.69)	4.76 (3.22-6.98)	0.83 (0.53-1.29)	Yancey	7.64 (4.92-11.70)	9.07 (6.29-12.90)	1.21 (0.75-1.93)

\* Change was statistically significant at the 5% level of significance.

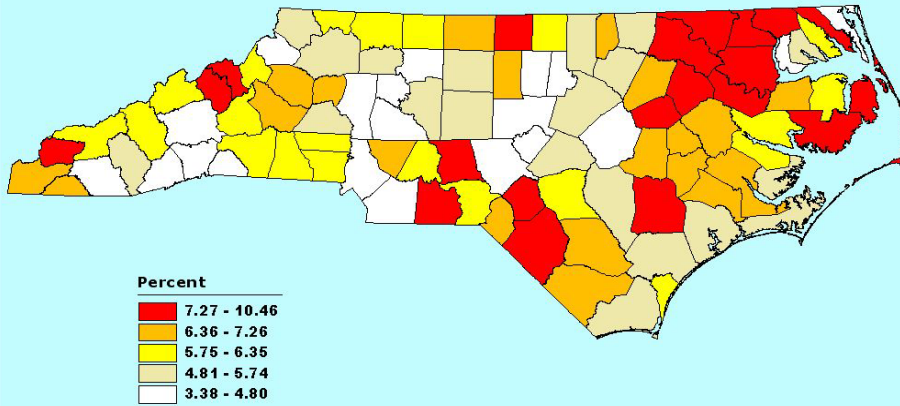
ASA Section on Survey Research Methods

**Table 2. SAEs, Odds-Ratio, and Corresponding 95% PIs for Smoking**

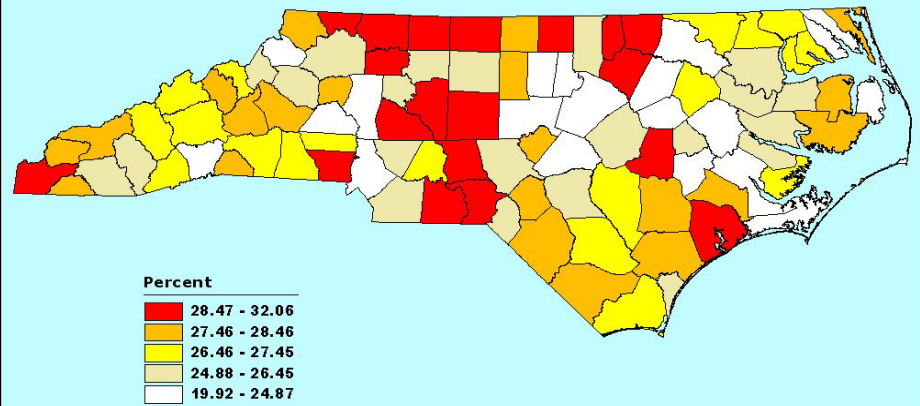
County	Small Area Estimates		Change (Odds-Ratio)	County	Small Area Estimates		Change (Odds-Ratio)
	1996-1999	2000-2002			1996-1999	2000-2002	
North Carolina	25.27 (24.20-26.36)	26.11 (25.03-27.22)	1.05 (0.96-1.14)				
Alamance	28.43 (24.02-33.31)	31.47 (26.69-36.68)	1.16 (0.92-1.45)	Johnston	25.52 (20.86-30.82)	27.85 (22.68-33.69)	1.13 (0.86-1.48)
Alexander	28.24 (22.44-34.87)	31.39 (25.04-38.53)	1.16 (0.88-1.53)	Jones	27.9 (21.94-34.76)	28.46 (22.35-35.47)	1.03 (0.81-1.30)
Alleghany	28.72 (22.68-35.63)	29.87 (23.52-37.10)	1.06 (0.83-1.35)	Lee	27.54 (21.79-34.13)	26.65 (20.71-33.56)	0.96 (0.70-1.30)
Anson	30.01 (24.03-36.76)	30.73 (24.59-37.65)	1.03 (0.81-1.32)	Lenoir	23.61 (19.02-28.91)	23.51 (18.75-29.04)	0.99 (0.80-1.23)
Ashe	28.25 (22.34-35.03)	28.99 (22.93-35.91)	1.04 (0.79-1.36)	Lincoln	27.13 (22.25-32.63)	25.38 (20.39-31.12)	0.91 (0.70-1.19)
Avery	26.34 (20.76-32.79)	24.69 (19.27-31.05)	0.92 (0.71-1.19)	Macon	26.45 (21.01-32.72)	29.74 (23.84-36.39)	1.18 (0.88-1.58)
Beaufort	25.04 (20.17-30.64)	27.98 (22.70-33.94)	1.16 (0.92-1.47)	Madison	26.95 (20.89-34.03)	29.45 (23.05-36.78)	1.13 (0.83-1.55)
Bertie	25.59 (20.34-31.66)	25.54 (20.24-31.67)	1 (0.80-1.25)	Martin	26.27 (20.88-32.48)	27.33 (21.83-33.63)	1.06 (0.84-1.33)
Bladen	26.49 (21.13-32.64)	29.2 (23.14-36.10)	1.14 (0.88-1.49)	McDowell	28.14 (22.79-34.20)	28.24 (22.76-34.46)	1.01 (0.79-1.27)
Brunswick*	26.82 (21.76-32.56)	32.97 (26.98-39.56)	1.34 (1.02-1.77)	Mecklenburg	20.47 (17.80-23.42)	18.99 (16.53-21.73)	0.91 (0.73-1.14)
Buncombe	27.05 (23.00-31.53)	30.39 (25.91-35.28)	1.18 (0.93-1.49)	Mitchell	27.41 (21.50-34.24)	28.35 (22.12-35.52)	1.05 (0.81-1.36)
Burke	28.06 (23.03-33.72)	28.78 (23.85-34.27)	1.04 (0.80-1.34)	Montgomery	29.01 (23.42-35.33)	31.69 (25.39-38.74)	1.13 (0.90-1.43)
Cabarrus	26.07 (21.74-30.91)	25.49 (20.84-30.76)	0.97 (0.77-1.23)	Moore	25.32 (19.81-31.76)	27.05 (21.32-33.67)	1.09 (0.80-1.50)
Caldwell	25.95 (21.24-31.29)	29.33 (24.09-35.18)	1.18 (0.91-1.54)	Nash	23.29 (19.04-28.15)	25 (20.45-30.17)	1.1 (0.87-1.39)
Camden	23.85 (18.41-30.30)	23.36 (17.86-29.95)	0.97 (0.76-1.25)	New Hanover	24.88 (20.67-29.37)	25.27 (20.97-30.12)	1.02 (0.78-1.34)
Carteret	23.54 (18.93-28.87)	26.61 (21.66-32.22)	1.18 (0.94-1.48)	Northampton	27.25 (21.50-33.88)	27.41 (21.63-34.06)	1.01 (0.80-1.26)
Caswell	27.73 (21.79-34.58)	29.95 (23.61-37.17)	1.11 (0.86-1.44)	Onslow	29.17 (23.96-34.98)	32.38 (27.03-38.24)	1.16 (0.87-1.55)
Catawba	24.87 (20.59-29.71)	25.89 (21.38-30.98)	1.06 (0.83-1.35)	Orange	21.13 (16.56-26.55)	21.81 (17.10-27.40)	1.04 (0.73-1.48)
Chatham	23.21 (17.76-29.72)	23.39 (17.92-29.92)	1.01 (0.71-1.42)	Pamlico	26.73 (20.93-33.46)	27.15 (21.08-34.20)	1.02 (0.81-1.29)
Cherokee	28.61 (22.43-35.72)	33.62 (26.97-41.00)	1.26 (0.94-1.70)	Pasquotank	23.42 (18.29-29.48)	23.05 (18.00-29.01)	0.98 (0.76-1.27)
Chowan	26.51 (20.82-33.11)	26.87 (21.06-33.60)	1.02 (0.80-1.29)	Pender	28.46 (22.92-34.73)	27.61 (21.95-34.10)	0.96 (0.73-1.25)
Clay	27.61 (21.24-35.05)	30.73 (23.87-38.57)	1.16 (0.87-1.56)	Perquimans	26.78 (21.05-33.40)	29.28 (23.13-36.30)	1.13 (0.88-1.45)
Cleveland	27.45 (22.62-32.87)	30.6 (25.21-36.58)	1.17 (0.90-1.50)	Person	29.04 (22.65-36.37)	26.92 (20.78-34.09)	0.9 (0.64-1.26)
Columbus	28.46 (22.86-34.83)	28.8 (22.90-35.53)	1.02 (0.77-1.35)	Pitt	23.13 (19.26-27.51)	26.43 (21.93-31.48)	1.19 (0.94-1.51)
Craven *	22.89 (18.56-27.88)	27.35 (22.41-32.91)	1.27 (1.01-1.60)	Polk	27.55 (20.96-35.30)	31.27 (24.47-38.99)	1.2 (0.85-1.68)
Cumberland	24.94 (21.24-29.05)	26.59 (22.80-30.75)	1.09 (0.86-1.38)	Randolph	29.16 (24.44-34.38)	29.62 (24.58-35.21)	1.02 (0.80-1.31)
Currituck	27.58 (21.71-34.35)	27.34 (21.41-34.20)	0.99 (0.76-1.28)	Richmond	29.94 (24.51-36.61)	31.05 (25.10-37.72)	1.05 (0.83-1.34)
Dare	24.63 (18.97-31.34)	26.75 (20.77-33.72)	1.12 (0.81-1.53)	Robeson *	28.21 (23.18-33.86)	34.45 (29.19-40.11)	1.34 (1.01-1.78)
Davidson	29.87 (25.19-35.00)	30.31 (25.28-35.86)	1.02 (0.81-1.29)	Rockingham	29.68 (24.68-35.23)	32.42 (27.01-38.34)	1.14 (0.90-1.44)
Davie	25.23 (19.65-31.75)	26.86 (20.74-34.02)	1.09 (0.84-1.41)	Rowan	30.02 (25.29-35.23)	27.17 (22.38-32.56)	0.87 (0.67-1.13)
Duplin	27.85 (22.47-33.95)	30.24 (24.52-36.65)	1.12 (0.90-1.40)	Rutherford	26.9 (21.66-32.88)	28.05 (22.67-34.14)	1.06 (0.82-1.37)
Durham	19.92 (16.37-24.02)	20.81 (17.26-24.88)	1.06 (0.83-1.34)	Sampson	27.32 (22.09-33.26)	32.31 (26.24-39.04)	1.27 (0.99-1.63)
Edgecombe	26.98 (21.79-32.88)	27.11 (21.68-33.33)	1.01 (0.80-1.27)	Scotland	25.89 (20.52-32.10)	28.52 (22.36-35.59)	1.14 (0.87-1.50)
Forsyth	25.71 (22.39-29.33)	25.17 (21.72-28.98)	0.97 (0.79-1.20)	Stanly	27.04 (21.88-32.91)	30.17 (24.33-36.74)	1.17 (0.91-1.49)
Franklin	28.55 (22.30-35.75)	25.75 (19.78-32.77)	0.87 (0.62-1.21)	Stokes	31.21 (25.46-37.60)	32.66 (26.44-39.56)	1.07 (0.83-1.38)
Gaston	30.24 (25.68-35.22)	27.15 (22.77-32.03)	0.86 (0.68-1.09)	Surry	32.06 (26.62-38.04)	32.34 (26.83-38.40)	1.01 (0.80-1.29)
Gates	26.83 (21.03-33.55)	27.98 (21.91-34.98)	1.06 (0.83-1.35)	Swain	27.51 (21.17-34.91)	29.2 (22.66-36.73)	1.09 (0.81-1.46)
Graham	27.51 (21.24-34.80)	32.64 (25.72-40.41)	1.28 (0.95-1.72)	Sylvania	26.73 (20.86-33.55)	28.85 (22.83-35.73)	1.11 (0.82-1.50)
Granville	25.01 (19.34-31.69)	20.39 (15.30-26.65)	0.77 (0.55-1.07)	Tyrrell	27.76 (21.67-34.79)	27.84 (21.62-35.05)	1 (0.78-1.29)
Greene	26.36 (20.51-33.17)	27.09 (21.27-33.81)	1.04 (0.82-1.32)	Union	25.67 (20.73-31.31)	26.74 (21.38-32.89)	1.06 (0.80-1.39)
Guilford	25.85 (22.71-29.26)	24.44 (21.49-27.65)	0.93 (0.75-1.14)	Vance	28.63 (22.52-35.65)	26.47 (20.55-33.38)	0.9 (0.65-1.24)
Halifax	24.05 (19.03-29.91)	24.28 (19.29-30.07)	1.01 (0.80-1.28)	Wake	20.81 (18.12-23.78)	19.71 (17.27-22.40)	0.93 (0.76-1.16)
Harnett	24.62 (19.53-30.53)	27.27 (21.97-33.29)	1.15 (0.86-1.53)	Warren	30.81 (24.28-38.21)	28.89 (22.10-36.78)	0.91 (0.66-1.26)
Haywood	27.19 (21.96-33.13)	30.33 (24.79-36.50)	1.17 (0.87-1.57)	Washington	25.96 (19.84-33.18)	26.61 (20.60-33.64)	1.03 (0.77-1.39)
Henderson	24.28 (19.62-29.64)	27.65 (22.66-33.26)	1.19 (0.89-1.59)	Watauga	23.99 (18.60-30.35)	25.85 (20.17-32.49)	1.1 (0.81-1.51)
Hertford	27.07 (21.39-33.63)	27.04 (21.61-33.24)	1 (0.79-1.26)	Wayne	29.5 (24.08-35.57)	33.29 (27.68-39.41)	1.19 (0.91-1.56)
Hoke	28.44 (22.35-35.44)	33.07 (26.32-40.59)	1.24 (0.95-1.63)	Wilkes	25.2 (20.32-30.80)	29.89 (24.41-36.01)	1.27 (0.98-1.64)
Hyde	27.53 (21.47-34.54)	27.76 (21.61-34.89)	1.01 (0.79-1.29)	Wilson	23.19 (18.60-28.52)	26.07 (21.16-31.67)	1.17 (0.91-1.50)
Iredell	24.35 (19.63-29.79)	22.9 (18.21-28.39)	0.92 (0.70-1.22)	Yadkin	29.26 (23.30-36.03)	32.43 (25.93-39.68)	1.16 (0.87-1.54)
Jackson	26.45 (20.40-33.54)	29.51 (23.31-36.57)	1.16 (0.87-1.55)	Yancey	27.56 (22.00-33.92)	31.1 (24.97-37.98)	1.19 (0.92-1.52)

\* Change was statistically significant at the 5% level of significance.

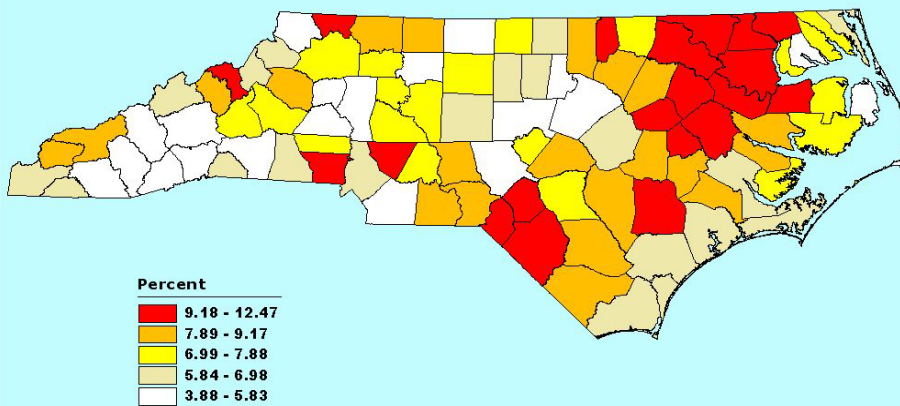
**Figure 1. Age-Adjusted Prevalence Rates of Diabetes (Percent) in North Carolina for Pooled 1996-1999 BRFSS Data**



**Figure 3. Age-Adjusted Prevalence Rates of Smoking (Percent) in North Carolina for Pooled 1996-1999 BRFSS Data**



**Figure 2. Age-Adjusted Prevalence Rates of Diabetes (Percent) in North Carolina for Pooled 2000-2002 BRFSS Data**



**Figure 4. Age-Adjusted Prevalence Rates of Smoking (Percent) in North Carolina for Pooled 2000-2002 BRFSS Data**

