STATISTICS IN MEDICINE
Statist. Med. **18**, 3189–3200 (1999)

MODEL-BASED SMALL AREA ESTIMATES OF OVERWEIGHT PREVALENCE USING SAMPLE SELECTION ADJUSTMENT

DONALD MALEC[1]*, WILLIAM W. DAVIS[2] AND XIN CAO[3]

[1]*U.S. Bureau of the Census, Statistical Research Division, room 3132-4, Washington, D.C. 20233, U.S.A.*
[2]*Westat, 1650 Research Blvd., Rockville, MD 20850-3129, U.S.A.*
[3]*IndexCo, LLC, One Corporate Center, Hartford, CT 06103, U.S.A.*

SUMMARY

Using a hierarchical model with an adjustment for sample selection, we estimate the overweight prevalence for adults, by states, using data from the Third National Health and Nutrition Examination Survey (NHANES III). A two-stage hierarchical model was selected to account for geographic variability of outcomes and to model possible overdispersion of estimates due to cluster sampling. We compare our model-based estimates with design-based estimates at the national level and obtain excellent agreement. We also provide a check of our model at the state level by comparing estimates with design-based and synthetic estimates. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

There is a continuing need to assess health status, health practices and health resources at both the national and subnational level. Estimates of these health items help determine the demand for health care and the access individuals have to it. Although the NCHS personal interview surveys can provide much of this information at the national level, little can be provided for states and counties because of excessive field costs. Making design-based state estimates from the current NHANES III is problematic for several reasons. The main reason is that no sample, at all, is selected in many states. For states with a large sample size, design-based state estimates can still be of low quality due to the high degree of geographic clustering of the sample into primary sampling units (PSUs). The need for subnational health statistics exists, however, because health and health care characteristics are known to vary geographically. Also, health care planning often takes place at the state and local level.

One alternative approach for producing subnational estimates has been to, effectively, increase the sample size by utilizing models defined across the subnational areas.[1] A challenge has been to use models realistic enough to produce accurate estimates. Towards this end, hierarchical models (models which include geographic variation among rates and can account for overdispersion due

* Correspondence to: Donald Malec, U.S. Bureau of the Census, Statistical Research Division, room 3132-4, Washington, D.C. 20233, U.S.A. E-mail: donald.j.malec@ccmail.census.gov

CCC 0277–6715/99/233189–12$17.50

to cluster sampling) have been adapted to small area estimation.[2] With the availability of Markov chain Monte Carlo (MCMC) methods,[3] estimates (and precision estimates) can be made that account for all model errors. Given current resources, model-based estimates can be made for subnational levels. At a minimum, a measure of the geographic variability of health characteristics can be determined and used to make decisions about which health characteristics to measure in future surveys of small areas.

In this paper we present a methodology for making subnational estimates which extends the hierarchical model of Malec *et al.*[2] by including an oversampling (that is, unequal selection probability) component in the likelihood. We illustrate the methodology by estimating the adult overweight prevalence by state using data from the Third National Health and Nutrition Examination Survey (NHANES III). The methodology is general and is especially useful for producing subnational estimates which, at a national level, should closely agree with design-based estimates.

1.1. NHANES III: Survey Design

NHANES III is a stratified, multi-level, clustered, personal interview survey of households[4] that was conducted in two phases: during the years 1988–1991 and 1991–1994. Sampled persons provide health and dietary information through a questionnaire and also through a physical exam. Persons were selected to represent the civilian, non-institutional population of the United States and provide national characteristics and nutrition status for the entire population and age, race and ethnic subgroups. NHANES III was designed to oversample the two largest minority groups of the U.S. population, Blacks and Mexican Americans.

1.2. Overweight prevalence in U.S.

Overweight is associated with a number of adverse health outcomes including mortality[5] and has become an increasing problem for adults in the United States.[6,7] Overweight is typically defined in terms of body mass index (BMI) which is defined by

$$\text{BMI} = \text{weight}/\text{height}^2. \tag{1}$$

Expressing BMI in the units kg/m^2, overweight is defined as $\geqslant 27.8$ for adult men and $\geqslant 27.3$ for adult women. These are the gender-specific 85th percentiles of BMI for men and women aged 20 to 29 from NHANES II (1976–1980). We refer to an adult with a BMI value below the gender specific threshold as 'normal'.

1.3. Description of the Finite Population

Let Y_{tidj} denote the overweight status for the jth individual, in demographic category d, in county i, during phase t of sampling. In particular, $Y_{tidj} = 1$ denotes overweight status as determined by BMI and $Y_{tidj} = 0$ denotes normal. Let N_{id} be the total number of individuals in demographic group d and county i as measured in the 1990 census. Here, a demographic group d describes a type of person. Specifically, d is defined as a classification (g, r, a), where g denotes gender, r denotes race/ethnicity (non-Hispanic White, non-Hispanic Black, and Mexican-American) and 'a' denotes an age category (20–24, 25–29, ..., 75–79, 80+). The six cross-classifications defined by crossing gender with race/ethnicity are important later, and each will be denoted by $c = (g, r)$.

Of interest are estimates of the finite population mean for individual characteristics defined by groupings of 'd' for local areas defined by county groupings. That is

$$\theta_{LD} = \frac{\sum_{i\in L}\sum_{d\in D}\sum_{j=1}^{N_{id}} Y_{idj}}{\sum_{i\in L}\sum_{d\in D} N_{id}} \tag{2}$$

where L indexes a particular collection of counties (for example, all counties in a specific state), and D is the set of specific subgroups of interest (for example, all females regardless of age or race). Here, Y_{idj} denotes overweight status at census day 1990.

2. ESTIMATION METHODOLOGY WITHOUT OVERSAMPLING

It is well known that ignoring the sample design can cause a selection bias and lead to erroneous conclusions.[8] A detailed discussion of the use of models to eliminate selection bias (of any kind) can be found in chapter seven of Gelman *et al.*[9] As described therein, a non-ignorable design is a design that results in selection bias while an ignorable design does not. A non-ignorable design can be made ignorable by adding appropriate design variables to the model (that is, the selection bias can be eliminated by including variables in the model that account for it). Realistically, incorporating appropriate sample design information into a model can either minimize or eliminate problems associated with selection bias.

This basic approach was adhered to by Malec *et al.*[2] in estimating small areas using the National Health Interview Survey. There, a two-stage hierarchical model was employed to account for overdispersion due to correlation within primary sampling units and a stepwise variable selection procedure was employed on the socio-economic variables used as stratifiers in the design. In addition, missed components of variance in the model were checked using cross-validation methods.

In Section 2.1, we specify a general method for estimating prevalence at a subnational level, similar to Malec *et al.*[2] This method is based on a population model that is appropriate when the sample selection is ignorable, given the two-stage model including covariates. In Section 3.1 we extend this model and estimation method to include a sample design that is non-ignorable, given the two-stage model including covariates. In Section 5 we employ data-based methods to check the adequacy of using this method.

2.1. The Population Model

A two-stage hierarchical model is used to describe individual and county variation. Conditional on the parameters p_{tid}, the Y_{tidj} are assumed to be independent Bernoulli random variables with

$$\Pr(Y_{tidj}\,|\,p_{tid}) = p_{tid}^{Y_{tidj}}(1-p_{tid})^{1-Y_{tidj}}. \tag{3}$$

Although we allowed county covariates related to the stratifiers to be candidates for model selection, these covariates did not explain much of the variation. We found the following specification of the model adequate for overweight status:

$$\text{logit}\{p_{tid}\} = \alpha_{td} + \beta_{ic} \tag{4}$$

where the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_i$ denote the fixed and random effects, respectively. The components of $\boldsymbol{\beta}_i^{\mathrm{T}} = (\beta_{i1}, \ldots, \beta_{i6})$ represent the six race/gender groups. The density of $\boldsymbol{\beta}_i$, $f(\boldsymbol{\beta}_i\,|\,\Gamma)$, is a

multivariate Gaussian density with mean vector zero and dispersion matrix, Γ, that is

$$f(\boldsymbol{\beta}_i \mid \Gamma) \propto |\Gamma|^{-1/2} \mathrm{e}^{-1/2\boldsymbol{\beta}_i^{\mathrm{T}}\Gamma^{-1}\boldsymbol{\beta}_i}. \tag{5}$$

The likelihood L_{I} of $\boldsymbol{\alpha}$, $\{\boldsymbol{\beta}_i\}$, and Γ is proportional to the product of terms in (3) and (5) corresponding to sampled individuals and counties:

$$\begin{aligned} L_I(\{\boldsymbol{\beta}_i\}_{i\in s}, \boldsymbol{\alpha}, \Gamma) &\propto \prod_{tidj\in s} \Pr(Y_{tidj} \mid p_{tid}) \times \prod_{i\in s} f(\boldsymbol{\beta}_i \mid \Gamma) \\ &= \prod_{i\in s}\left[\prod_t \prod_d p_{tid}^{m_{tid}}(1-p_{tid})^{n_{tid}-m_{tid}}\right] |\Gamma|^{-1/2} \mathrm{e}^{-1/2\boldsymbol{\beta}_i^{\mathrm{T}}\Gamma^{-1}\boldsymbol{\beta}_i} \end{aligned} \tag{6}$$

where $tidj \in s$ denotes the set of all individuals in sample and $i \in s$ denotes the set of counties that contain sampled individuals. Also, n_{tid} and m_{tid} denote the number of individuals in demographic group d, county i and phase t who are in sample and who are overweight, respectively.

2.2. Estimation

We use a Bayesian approach to make inference about θ_{LD}. Bayesian inference is performed conditionally on the sampled individuals, their responses are known and need not be estimated. However, at our reference time of census day 1990, we do not know where any of our sampled individuals reside (except for the few that might have been interviewed on that day). Hence, we estimate overweight prevalence for the entire population (sampled and unsampled). Since our model includes a phase effect, we make predictions using $\Pr(Y_{idj} = 1 \mid p_{1id}, p_{2id}) = (p_{1id} + p_{2id})/2$ (that is, using proportion averaged over phase). After specifying a prior distribution for $\boldsymbol{\alpha}$ and Γ, we estimate the posterior mean and variance of θ_{LD} using MCMC methods.[3]

3. ESTIMATION METHODOLOGY WITH SAMPLE SELECTION ADJUSTMENT

Ignoring the way the sample was selected may produce erroneous inferences.[9] By including sufficient design information in a model, any design can become ignorable. Often, however, the design information needed for modelling and prediction is difficult or impossible to obtain. This is particularly true at the lower levels of sampling, where sampling frames are only constructed for the higher-level units in sample. Here, we do not attempt to incorporate within-PSU design characteristics in our model. Instead, for simplicity, we utilize non-ignorable design methodology.[9]

3.1. Sample Selection Model

Based on previous extensive data analysis, we conclude that the design is ignorable above the PSU-level so that the model, defined in (4) and (5), is not affected by the PSU-level design characteristics. Hence, we use non-ignorable design methodology within counties (PSUs), only.

Let π_{tidj} denote the selection probability of individual $tidj$, as specified in the sample design. The resulting 'empirical Bayes' likelihood L_{NI} based on a non-ignorable design is

$$L_{NI}(\{\boldsymbol{\beta}_i\}_{i\in s}, \boldsymbol{\alpha}, \Gamma) = \prod_{i\in s}\left[\prod_t \prod_d \frac{p_{tid}^{m_{tid}}(1-p_{tid})^{n_{tid}-m_{tid}}}{(p_{tid}/\bar{w}_{1td} + (1-p_{tid})/\bar{w}_{0td})^{n_{tid}}}\right] |\Gamma|^{-1/2} \mathrm{e}^{-1/2\boldsymbol{\beta}_i^{\mathrm{T}}\Gamma^{-1}\boldsymbol{\beta}_i} \tag{7}$$

where $\bar{w}_{1td}$ and $\bar{w}_{0td}$ are the sampling weights (inverse selection probabilities) for demographic group d in phase t, averaged over overweight and normal persons, respectively. That is

$$\bar{w}_{1td} = \sum_{(i,k)\in s_{td}} Y_{tidk}\pi_{tidk}^{-1} \Big/ \sum_{(i,k)\in s_{td}} Y_{tidk}$$

and

$$\bar{w}_{0td} = \sum_{(i,k)\in s_{td}} (1 - Y_{tidk})\pi_{tidk}^{-1} \Big/ \sum_{(i,k)\in s_{td}} (1 - Y_{tidk})$$

where $(i, k)\in s_{td}$ denotes all sampled persons in demographic group d in phase t. The sample adjusted likelihood in (7) differs from (6) only in the denominator which adjusts for oversampling of both overweight and normal persons. A brief description on how the denominator of (7) was derived is presented in the Appendix. Alternatively, the adjustment in the denominator of (7) can be viewed as conditioning on the sample selection (for example, see Jewell[10]).

A full Bayesian analysis includes a model for the distribution of the π_{tidj}'s as part of the likelihood, instead of substituting in their corresponding maximum likelihood estimates (MLEs), as we have done (in equation (7)). Our empirical Bayes analysis, although only an approximation to the full Bayesian analysis, was chosen for its relative simplicity. The likelihood in (7) when combined with an appropriate prior distribution enables one to make estimates for θ_{LD}.

4. INFERENTIAL METHODOLOGY

A Bayesian analysis requires the specification of a prior distribution for $(\boldsymbol{\alpha}, \Gamma)$. To ensure that the sample information dominates the inference, we used an overdispersed prior distribution. In particular, we choose the conditional density of $\boldsymbol{\alpha}\,|\,\Gamma$ to be constant and an inverse Wishart distribution for Γ with one degree of freedom and mean $= vI_{6\times 6}$. This prior (with $v = 10^4$) is dominated by the data but seems to avoid problems with the use of vague priors in hierarchical models.[11]

Since the posterior moments of θ_{LD} are non-linear functions, and the posterior distribution cannot be expressed in a simple form, numerical evaluation is needed. We used MCMC methodology to generate a random set of parameters which converges to a stationary distribution that is the posterior distribution. This is achieved by successively generating parameter subsets from their conditional distributions. We used graphics and formal statistical tests to determine when convergence to the stationary distribution was attained. The parameters after this point were treated as a sample from the posterior distribution and used to estimate posterior moments. More specifically, we used the block-at-a-time Metropolis–Hastings algorithm[12] to generate one long run of the chain. Since the explicit posterior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are unknown, the modes and Hessians were searched at each iteration to determine the candidate-generating Gaussian densities. Conditionally Γ was sampled directly from its inverse Wishart distribution. We also used CODA software[13] to perform the output analysis and convergence diagnosis for the chain. Within CODA, we used the Heidelberger and Welch[14] test to determine the number of iterations to discard and to determine if the Markov process was indeed stationary.

5. ESTIMATION RESULTS

We illustrate the calculations using two choices of L and D from (2). In Section 5.1 we show the estimates made at the national level for demographic subgroups. We compare our model-based

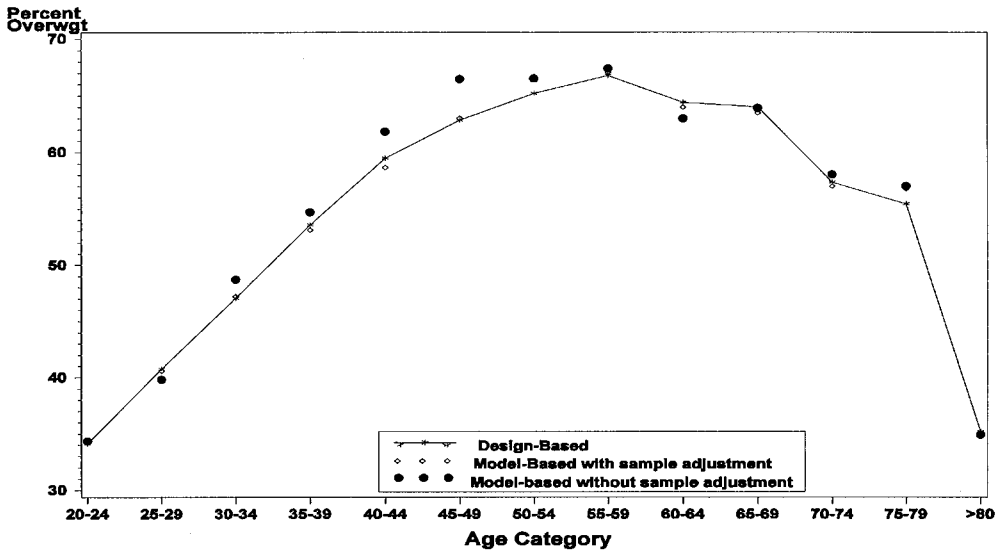

Figure 1. Comparison of estimates of overweight for non-Hispanic Black females at the national level: design-based versus model-based (with and without sample adjustment)

estimates with design-based estimates to justify our claim that the results may coincide at a national level. In Section 5.2 we provide comparison of our state level estimates with design-based and synthetic estimates. In Section 5.3 we show our estimates for all adults within the 50 states and D.C.

5.1. Evaluation of National Estimates by Demographic Subgroups

In this section we compare model- and design-based estimates for demographic categories. For estimation, we used the 16,523 BMI values for all adults (20 and over) who were examined in a mobile examination centre (MEC). We used standard expansion estimators to estimate the overweight prevalence for all demographic categories using the MEC examination weights.[15] Overweight prevalence is highest for ethnic (non-Hispanic Black and Mexican-American) females.[6]

For the model-based estimates, we approximated the selection probabilities of Section 3.1 by the inverse of the MEC examination weights after post-survey adjustment. We used SAS IML for the calculations and 1200 iterations of the Gibbs sampler. The estimates are based on the final 1000 iterations since the Heidelberger and Welch test indicates that the chain had converged by then. The values shown were obtained for the prior distribution with $v = 10^4$. We used sensitivity analysis to insure that our prior was overdispersed.

In Figure 1, we compare design-based, model-based without adjustment for sample selection, and model-based with adjustment for sample selection. Since the conclusions are similar for all race/ethnicity/gender categories, we illustrate our methodology using Non-Hispanic Black females. With adjustment for sample selection (specified in Section 3) the model-based estimates tracked the design-based estimates well for all ages. Without the sample selection adjustment, the

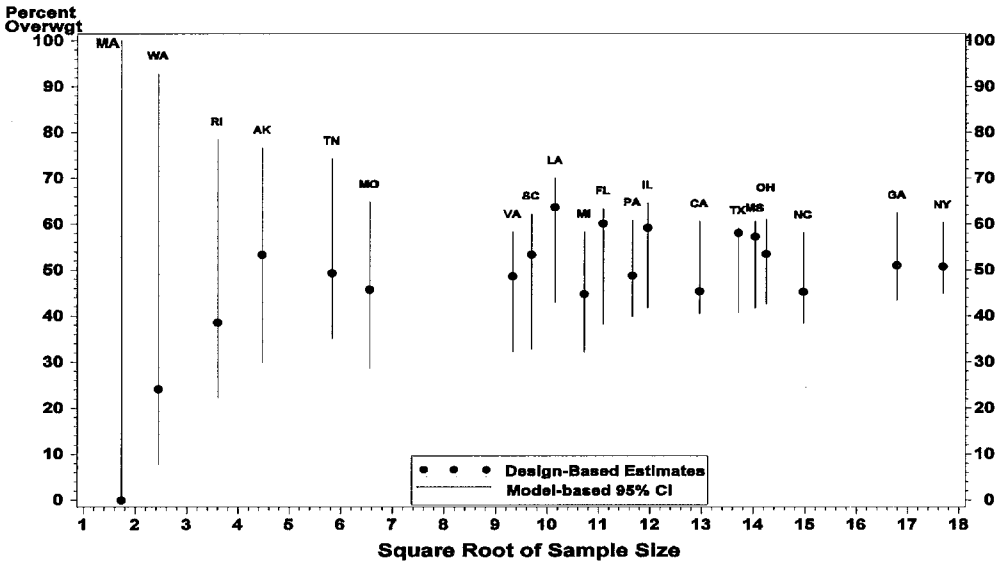

Figure 2. Comparison of state design-based estimates and model-based 95 per cent credible intervals (CIs) for overweight for non-Hispanic Black females

differences between the model-based and the design-based estimates were more pronounced (up to 5 percentage points difference). We take these results as verification that the sample selection adjustment was necessary.

5.2. Evaluation at the State Level

It is well known that hierarchical models, like the one used here, tend to 'smooth' estimates. Some smoothing is desirable because state estimates using only state data will contain more error and, hence, be more variable than estimates that 'borrow strength' from data outside of the state. However, there is concern that the model being used could inadvertently be smoothing the actual population values, not just removing error. If our model under represents the variability between counties, we will oversmooth. To check our model for oversmoothing, we evaluate the variability inherent in our model against the variability of the raw data as follows:

1. We first produce state design-based estimates for each state that contains at least one sampled PSU. These estimates use only data collected within the state and, hence, are not smoothed by averaging over data collected in other states.
2. Using our model with sample selection adjustment, we produce new values of overweight status for each sampled person via their posterior predictive distribution (with sample adjustment). We make corresponding state design-based estimates from these new outcomes. Since the new outcomes are from the posterior, estimates based on them will be smoothed.

Figure 2 compares the design-based state estimates with 95 per cent credible intervals[11] from our new outcomes for non-Hispanic Black females. Each raw design-based estimate falls within the credible interval formed using our model. If our model had oversmoothed, our generated

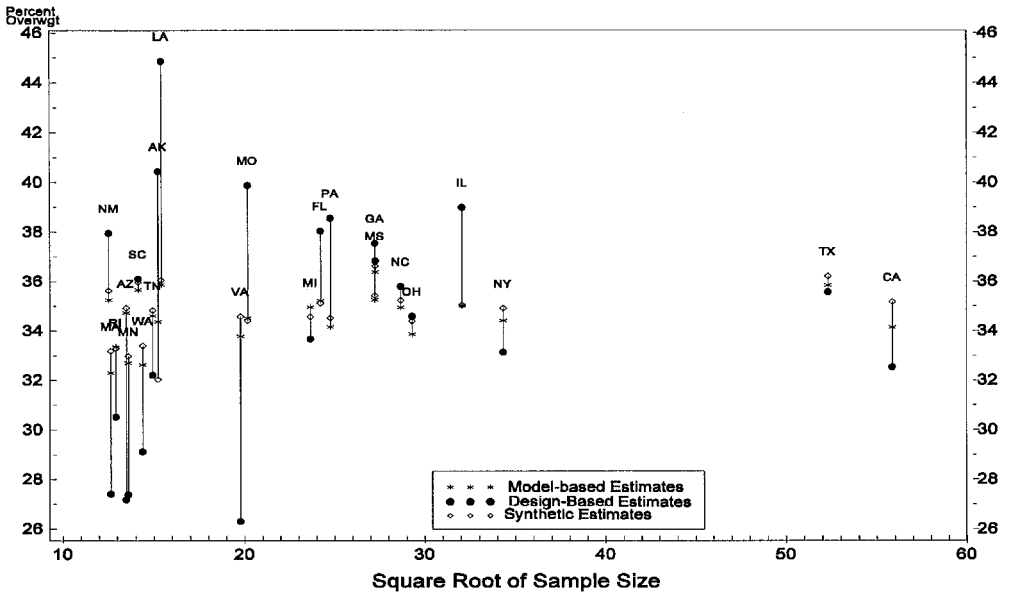

Figure 3. Comparison of model-based design-based and synthetic estimates of overweight for adults in sampled states

outcomes would tend to underestimate the data variability and would have resulted in poor coverage of the design-based estimates. We obtain this conclusion in the other race/gender categories, also.

Figure 3 compares the following three estimates for states that have a sample: model-based, design-based, and synthetic (defined by $\hat{\theta}_{LD} = \sum_{i\in L}\sum_{d\in D} N_{id}\hat{r}_d / \sum_{i\in L}\sum_{d\in D} N_{id}$, where $\hat{r}_d$ is the design-based estimate of the national prevalence rate for domain d).

The figure shows that, for states with a large sample, the model-based estimate is closer to the design-based estimate than is the synthetic estimate. In addition, the model-based posterior variance (not shown), decreases with state sample size. These two observations illustrate that the model-based approach preferentially uses state data. In general, the synthetic estimates are close to our estimates, suggesting that they are adequate in this case. However, without using the hierarchical model to account for between-county variation, we could not have made this conclusion. In addition, our method provides estimates of precision including county variation.

5.3. State Estimates

We computed the overweight prevalence estimate by state and show the results in Figure 4. The figure shows a relatively small range (0.32 to 0.40) and a north/south difference (reflecting the difference in minority population).

6. CONCLUSIONS

With the aim of producing accurate state estimates of overweight status, a hierarchical model with random county variation was employed. This model provides information on the

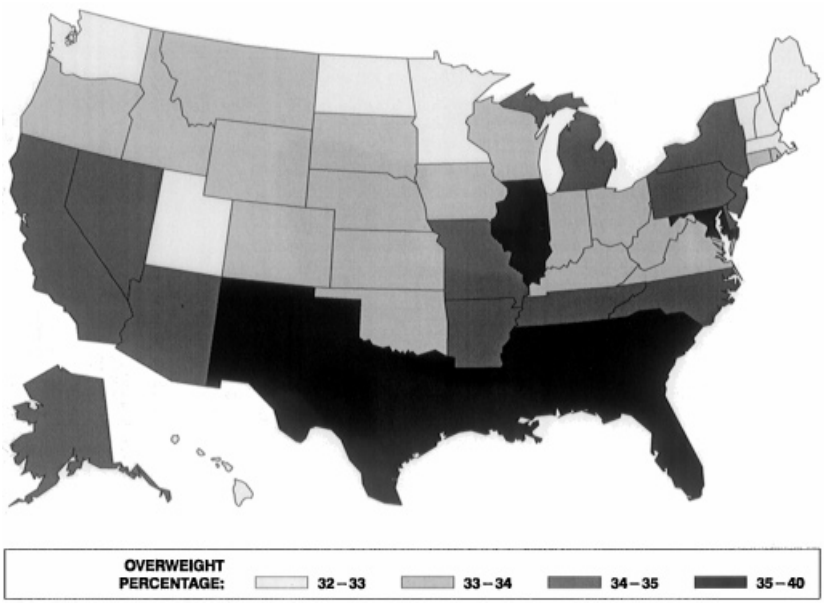

Figure 4. Overweight prevalence for adults by state

geographic variation of county prevalence rates and produces state estimates that preferentially use state data. The effects due to sample selection bias were minimized in three ways. First, county variables related to the NHANES III stratification variables were candidates in the stepwise variable selection procedure. For overweight status, however, these variables do not account for much of the overall data variability and were omitted from the model. Second, the county random effects component used to account for geographic variation also accounts for possible overdispersion due to the clustered sample. Third, the sample selection model was included in the likelihood to account for oversampling within PSUs.

Model misspecification was evaluated in two ways. First, we compared our estimates utilizing sample selection adjustment, with design-based estimates at the national level and obtained excellent agreement for all demographic groups. Second, we produced model-based credible intervals for the design-based state estimates and observed that these estimates fell within the intervals, indicating that the variability of these estimates had been adequately modelled. An unresolvable problem, however, in evaluating small area estimates is that accurate comparisons are not available at the small area level. In this application, for example, the design-based estimates at the state level are of poor precision and cannot be viewed as the 'gold standard'. Based on the model evaluations that we can make, we conclude that these estimates provide useful information on overweight prevalence for the states and D.C.

APPENDIX: SAMPLE SELECTION DERIVATION

We assume that the population of selection probabilities within a PSU, π_{tidj}, are identically distributed within demographic group, phase and overweight status. We specify an approximate

distribution for the selection probabilities by first assuming that they have support consisting only of the unique observed values $\pi_1^*, \ldots, \pi_U^*$. In particular, given that $Y_{tidj} = y$ $(y = 0, 1)$, π_{tidj} is modelled as a multinomial random variable with

$$\Pr(\pi_{tidj} = \pi_u^* \mid \theta, Y_{tidj} = y, p_{tid}) = \theta_{tdyu}. \tag{8}$$

Note that (8) makes few assumptions about the distribution of the selection probabilities. We have, however, assumed that the distribution of selection probabilities is identical between counties, within a demographic group and phase. We first show how the likelihood of overweight status is derived when $\boldsymbol{\theta}$ is fixed and then show how an estimate of $\boldsymbol{\theta}$ is derived to obtain (7).

We apply the general methodology for handling non-ignorable designs[9] to this case. Since our inference is conditional on county demographic groups and the NHANES III uses extensive implicit stratification, we assume each sampled person can be viewed as a sample of one person per substratum consisting of people in the same demographic group and phase. For sampled person j, we assume that the substratum size N_{tidj} is unknown. Dropping the extraneous subscripts (*tid*), for each person k in substratum j we let δ_{kj} be a Bernoulli random variable with sample selection probability π_{kj}. Without loss of generality, we label the sampled person as the first person in substratum j. The joint distribution of the one sampled person and the remaining unsampled persons in substratum j is

$$\begin{aligned}\Pr(\delta_{1j} = 1, Y_{1j}, \pi_{1j}, \{\delta_{kj} = 0, Y_{kj}, \pi_{kj}\}_{k=2,\ldots,N_j} | N_j) \\ = \pi_{1j}\Pr(\pi_{1j} | Y_{1j})\Pr(Y_{1j}) \prod_{k=2}^{N_j} (1 - \pi_{kj})\Pr(\pi_{kj} | Y_{kj})\Pr(Y_{kj}).\end{aligned} \tag{9}$$

The marginal distribution of only the observed quantities can be obtained by summing (9) over the unobserved components, giving

$$\begin{aligned}\Pr(\delta_{1j} = 1, Y_{1j}, \pi_{1j}, \{\delta_{kj} = 0\}_{k=2,\ldots,N_j} | N_j) \\ = \pi_{1j}\Pr(\pi_{1j} | Y_{1j})\Pr(Y_{1j}) \left(1 - \sum_{y=0}^{1} \sum_u \pi_u^* \Pr(\pi_u^* | y)\Pr(y)\right)^{N_j - 1}.\end{aligned} \tag{10}$$

Utilizing a non-informative prior, $\Pr(N_j) \propto$ constant, N_j can be removed from (10) giving

$$\begin{aligned}\Pr(\delta_{1j} = 1, Y_{1j}, \pi_{1j}, \{\delta_{kj} = 0\}_{k \neq 1}) &\propto \sum_{N_j=1}^{\infty} \pi_{1j}\Pr(\pi_{1j} | Y_{1j})\Pr(Y_{1j}) \left(1 - \sum_{y=0}^{1} \sum_u \pi_u^* \Pr(\pi_u^* | y)\Pr(y)\right)^{N_j - 1} \\ &\propto \pi_{1j}\Pr(\pi_{1j} | Y_{1j})\Pr(Y_{1j}) \Big/ \sum_{y=0}^{1} \sum_u \pi_u^* \Pr(\pi_u^* | y)\Pr(y).\end{aligned} \tag{11}$$

Equation (11) provides the likelihood component for one sampled person and replaces the specification in (3). Substituting the full subscript notation, for sample person j, into (11), one has

$$\begin{aligned}\Pr(\delta_{tidj} = 1, Y_{tidj} = y, \pi_{tidj} = \pi_u^*, \{\delta_{tidk} = 0\}_{k \neq j} | p_{tid}) \\ = \pi_u^* \theta_{tdyu} p_{tid}^y (1 - p_{tid})^{1-y} \Big/ \sum_{y=0}^{1} \sum_u \pi_u^* \theta_{tdyu} p_{tid}^y (1 - p_{tid})^{1-y} \\ \propto p_{tid}^y (1 - p_{tid})^{1-y} \Big/ \sum_{y=0}^{1} \sum_u \pi_u^* \theta_{tdyu} p_{tid}^y (1 - p_{tid})^{1-y}.\end{aligned} \tag{12}$$

By including the distribution of the selection probabilities specified in (8) as part of the likelihood, we could perform a complete Bayesian analysis. For simplicity, we substitute the MLE of $\boldsymbol{\theta}$ (based on conditioning on overweight status and the selection probability) into (12). For sampled individual *tidj*, the component of the conditional likelihood is

$$\Pr(\pi_{tidj} = \pi_u^* \mid \delta_{tidj} = 1, Y_{tidj} = y, \boldsymbol{\theta}) = \frac{\Pr(\delta_{tidj} = 1 \mid \pi_{tidj} = \pi_u^*, Y_{tidj} = y, \boldsymbol{\theta}) \Pr(\pi_{tidj} = \pi_u^* \mid Y_{tidj} = y, \boldsymbol{\theta})}{\Pr(\delta_{tidj} = 1 \mid Y_{tidj} = y, \boldsymbol{\theta})}$$

$$\propto \frac{\theta_{tdyu}}{\sum_u \pi_u^* \theta_{tdyu}}. \tag{13}$$

Multiplying all the terms of (13) together, the conditional likelihood is $\prod_{tdyu} (\theta_{tdyu} / \sum_u \pi_u^* \theta_{tdyu})^{\tau_{tdyu}}$ where τ_{tdyu} denotes the sample frequency of π_u^* within demographic group d, phase t and overweight status y. It can be shown that the MLEs of the θ_{tdyu} are

$$\hat{\theta}_{tdyu} = \frac{\tau_{tdyu}/\pi_u^*}{\sum_u \tau_{tdyu}/\pi_u^*}. \tag{14}$$

Plugging the estimates from (14) into (12) completes the within-PSU likelihood component in (7).

ACKNOWLEDGEMENTS

We want to thank Kurt Maurer for helpful comments and interest in the NHANES III small areas estimation project, Joe Sedransk for providing helpful comments on an earlier draft, and the referees and editor for providing comments that greatly improved the presentation of results.

REFERENCES

1. Ghosh, M. and Rao, J. N. K. 'Small area estimation: an appraisal', *Statistical Science*, **9**, 55–93 (1994).
2. Malec, D., Sedransk, J., Moriarity, C. and LeClere, F. 'Small area inference for binary variables in the National Health Interview Survey', *Journal of the American Statistical Association*, **92**, 815–826 (1997).
3. Gilks, W. R., Richardson, S. and Spielgelhalter, D. (eds). *Practical Markov Chain Monte Carlo*, Chapman and Hall, New York, 1996.
4. National Center for Health Statistics. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988–94, *Vital Health Statistics*, **1**, (1994).
5. Troiano, R. P., Frongillo, E. A. Jr., Sobal, J. and Levitsky, D. A. 'The relationship between body weight and mortality: a quantitative analysis combining information from existing studies', *International Journal of Obesity*, **20**, 63–75 (1996).
6. Kuczmarski, R. J., Flegal, K. M., Campbell, S. M. and Johnson, C. L. 'Increasing prevalence of overweight among US adults: the National Health and Examination Surveys, 1960 to 1991', *Journal of the American Medical Association*, **272**, 205–211 (1994).
7. Galuska, D. A., Serdula, M., Panuk, E., Siegel, P. Z. and Byers, T. 'Trends in overweight among US Adults from 1987 to 1993: a multistate telephone survey', *American Journal of Public Health*, **86**, 1729–1735 (1996).
8. Scott, A. J. 'On the problem of randomization in survey sampling', *Sankhya*, **39**, Series C, Pt.1, 1–9 (1977).
9. Gelman, A., Carlin, B. B., Stern, H. S. and Rubin, D. B. *Bayesian Data Analysis*, Chapman and Hall, New York, 1995, Chapter 7.
10. Jewell, N. P. 'Least squares regression with data arising from stratified samples of the dependent variable', *Biometrika*, **72**, 11–22 (1985).
11. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*, 2nd edn, Springer-Verlag, New York, 1986, Chapter 4.

12. Chib, S. and Greenberg, E. 'Understanding the Metropolis-Hastings algorithm', *American Statistician*, **49**, 327–335 (1995).
13. Best, N., Cowles, M. K. and Vines, K. *CODA*: *Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output*, Version 0.30, MRC Biostatistics Unit, Cambridge, 1995.
14. Heidelberger, P. and Welch, P., 'Simulation run length control in the presence of an initial transient', *Operations Research*, **31**, 1109–1144 (1983).
15. Mohadjer, L., Montaquila, J., Waksberg, J., Bell, B., James, P., Flores-Cervantes, I. and Montes, M. 'National Health and Nutrition Examination Survey III: Weighting and Estimation Methodology', Westat Inc., Rockville, MD, 1996.

