

USING MIXED-EFFECTS MODELING TO AID IN THE SAMPLE DESIGN PROCESS

Donald Malec, National Center for Health Statistics (NCHS), David Judkins, and Huseyin A. Goksel, Westat, Inc., Keith Hoffman and Iris Shimizu, NCHS, and Michael Monsour, National Institutes of Health

Donald Malec, National Center for Health Statistics, Rm. 915, 6525 Belcrest Rd., Hyattsville, MD 20782

Key Words: Gibbs Sampler, MCMC

1. Introduction and background

The National Home and Hospice Care Survey (NHHCS) is conducted by the National Center for Health Statistics (NCHS) to collect data on the characteristics of home health care agencies and hospices and their clients. The universe includes those hospices and home health care agencies which are either certified by the U.S. Health Care Financing Administration or licensed by a state. The survey was first implemented in 1992-94 with a three-stage sample in which geographic areas were sampled at the first stage, home care agencies and hospices were sampled at the second stage and clients were sampled at the third stage. NCHS plans to redesign the survey for 1996.

The sampling frame for the survey consists of the home health care agencies and hospices listed in the 1991 National Health Provider Inventory and new agencies listed by the Agency Reporting System as of April 30, 1992. These lists include agency characteristics used to design the survey. Two general designs are being considered for 1996. One is a three-stage design, similar to the design currently in use. The other is a stratified two-stage design, in which clients are sampled within agencies which are, in turn, selected from strata. As part of the design selection, an optimal allocation of sampling units will also be determined. In order to evaluate designs and to determine optimal sample allocations, population quantities, such as population variance components, are needed.

There are a number of estimates from the NHHCS for which precision is needed. The final design will be one which yields acceptable precision for estimates of ten client characteristics. The ten characteristics are:

- 1) Widowed current clients receiving home care
- 2) Black current clients receiving home care
- 3) Current clients with injury and poisoning receiving home care
- 4) Current clients receiving hospice care
- 5) Current clients receiving hospice care certified by Medicare/Medicaid
- 6) Current clients over 64 years old receiving hospice care
- 7) Current clients with neoplasms receiving hospice care
- 8) Female current clients receiving hospice care
- 9) Discharged clients 65-69 years old who received home care

- 10) Discharged clients who received hospice care.

The final design selection will be made on an ad hoc basis by evaluating the designs on each of these ten characteristics. Unfortunately, population values are not available for any of the ten characteristics, precluding a straightforward evaluation of variances and variance components. Fortunately, 1992 sample data is available from the current design and will be used to produce estimates of the desired variances and variance components.

Procedures for estimating variance components from relatively simple multi-stage samples are provided in most sampling texts. However, the design for the 1992 NHHCS was quite complex. Specifically, the three stages of selection were: PSUs (metropolitan areas, counties or groups of counties), agencies and clients. Two PSUs were selected per stratum without replacement according to the Brewer-Durbin method. Agencies were selected PPS systematic in some PSUs and with certainty in other PSUs. Systematic samples of 6 current and 6 discharged clients were selected within each agency.

A number of difficulties in the design-based approach to variance component estimation were perceived. First, there is a concern that unbiased estimates of the variance components will be negative. This is a common problem. The authors are familiar with similar experiences on a number of health and demographic surveys. Although a more reasonable estimate of a variance component is zero in this case, using zero only acknowledges the likelihood that the true value is small.

Second, although standard sampling reference texts like Hansen, Hurwitz, and Madow (1953) and Sukhatme and Sukhatme (1970) give techniques for estimating the components of variances for three-stage samples, these methods are for simpler designs. In other surveys, the components of variance are sometimes studied by ignoring one or more facets of the sample design so that the simpler formulae may be used. Relying solely on this practice was thought to be problematic for the NHHCS since the impact of all stages of sampling needs to be determined.

Third, the derivation of estimates will be further complicated since data selected from one design will be used to estimate the variance components from another design. Although possible, design-based estimates will likely become less efficient the more different the two designs are. A model-based approach can more easily incorporate additional covariates in the estimation.

Lastly, the overall accuracy of both the estimated variance components and their estimated precision would be helpful in

deciding just how important the selection of a particular design is. This is rarely done as it requires estimating the variance of the estimated variance components. It is straightforward using this approach.

At the same time that we were contemplating the complexity and instability of purely design-based analysis of variance components for NHHCS, the developments in random effects modeling were beginning to demonstrate that it is possible to develop realistic models of a population and the developments in Monte Carlo estimation, via Gibbs Sampling, enabled the straightforward production of estimates of any function of the population. Malec, Sedransk and Tompkins (1993) used it for small area estimation. Schafer, et al (1993) used it in imputation for NHANES III.

These influences coalesced in late 1993 into the idea that one could simulate the entire universe of home health care and hospice clients in order to be able to calculate the variances and components of variances for various sample designs exactly. In other words, rather than approximating the variance for various sample designs directly from a sample, we would first blow the sample up into an entire universe and then obtain the desired statistics exactly from the estimated universe. This approach alleviates the design-based problems outlined above but raises the new concern that everything relies upon how well the underlying population can be modeled. Clearly, an inferior model will yield inferior projections of the effects of various sample designs.

Section two outlines the general methodology for using a population model to assess a design. Section three details the process that was used to arrive at a final population model for one of the ten variables under consideration: the number of clients who are female and receive hospice care. Section four presents the uses of the model for selecting a design. Section five contains some concluding remarks.

2. Design evaluation using population models

Define \hat{X}_D to be a specific estimator based on design "D", \underline{X} to be a vector containing the population values and $Var(\hat{X}_D|\underline{X})$ to be the variance of \hat{X}_D due to repeated selections of elements of \underline{X} according to design "D". Note that if \underline{X} is known, then $Var(\hat{X}_D|\underline{X})$ can always be determined given that the design clearly specifies how to draw a sample.

Given that the population follows a specific model, an estimate of the population can be made by selecting a typical population randomly from the model. Denote this estimated population as $\tilde{\underline{X}}$. An estimate of the variance can be made by substituting the estimated population into the population variance formula, i.e. $Var(\hat{X}_D|\underline{X})=Var(\hat{X}_D|\tilde{\underline{X}})$. Using numerical methods, a stable estimate of both the variance and its variance can be made by drawing, say, R independent

populations $\tilde{\underline{X}}^{(1)}, \tilde{\underline{X}}^{(2)}, \dots, \tilde{\underline{X}}^{(R)}$ and evaluating the mean and variance of $Var(\hat{X}_D|\tilde{\underline{X}}^{(1)}), Var(\hat{X}_D|\tilde{\underline{X}}^{(2)}), \dots, Var(\hat{X}_D|\tilde{\underline{X}}^{(R)})$. Estimates for other functions of the population, such as ratios of variances from competing designs, can similarly be made from the same set of populations.

We have limited our design problem to include only the use of an expansion estimator and two- and three- stage sample designs.

3. Modeling the population

The models developed will be used to characterize the effect of a sample design on specific populations. To achieve this end, design variables were allowed to be included as parameters of the model. Other covariates, available for all agencies, that may aid in the prediction were also allowed to be included in the model. Finally, random effects were included in the model to help account for geographical local variation.

In this approach, the number of clients with a specific characteristic needs to be simulated for each agency. This is achieved by first independently modeling the proportion of clients in an agency with the characteristic and modeling the total number of clients in each agency. There are actually two pools of clients in each agency that are of interest: current clients and discharged clients. The size of both pools are modeled and predicted.

3.1 Fixed effect modeling

A fixed effect model was used to select the variables to use in the random effects model. The covariates examined included agency-level variables, county-level variables, and sample design variables.

In order to predict characteristics for clients in all agencies, covariates in the model need to be available for all agencies. The agency characteristics available from the sampling frame include the agency type (home health agency, hospice); ownership type (profit, nonprofit, government); agency certification status (certified for Medicare or Medicaid -including pending-, not certified); agency current client size; agency discharged client size.

County level variables were obtained from the Area Resource File (ARF), U.S. Department of Health and Human Services, September 1993. The ARF contains county information from a number of resources. The file is updated periodically, e.g., the MSA definitions in the file are based on the 1990 Census. We did not construct PSU level statistics from the county-level ARF data because the counties are more homogeneous than the PSU's and county level data were directly available from the ARF.

The county-level characteristics included county age and race/ethnicity composition, county socio-economic

characteristics, health care services and utilization and mortality statistics.

In addition to these county characteristics, first-stage sample design variables, census region, PSU type (self-representing, non-self-representing MSA, and nonMSA) variables were also included in model development. For a complete listing of all variables considered in the model selection see Judkins and Goksel (1995).

We used SAS PROC LOGISTIC software to select variables. As in Malec, Sedransk and Tompkins (1993), an R^2 , for logistic models, was used to evaluate the fit of the model. The R^2 for a particular model is a comparison between the log-likelihoods of the model and that of a model with a fixed effect for each agency in sample. (Agencies with samples of individuals either all having the

characteristic or all not having the characteristic do not have an MLE on the logistic scale. However, the MLE is one or zero, respectively, on the original probability scale and this value was used to calculate the likelihood.)

In order to assess possible nonlinear terms in the logistic model, residuals for binary data (Landwehr, et al. (1984) were evaluated. In addition, the Schwarz criterion (provided in SAS) was used to arrive at a final set of covariates. After the initial set of predictors was selected, all possible first-order interactions of these variables were included and the stepwise regression procedure was rerun. However, the inclusion of the interaction terms did not improve the model for any of the ten characteristic. It was decided to use a main effects only model.

The final fixed effects covariates selected for each model are listed in table 1.

Table 1. Final covariates selected for fixed effects model.

Dependent variable	Final covariates*
Widowed current clients receiving home care	A_TYPE, PCTMEX1 and LPNFT
Black current clients receiving home care	PCTBLKNH, A_TYPE and OWNER3
Current clients with injury and poisoning receiving home care	A_TYPE and MEDFMINC
Current clients receiving hospice care	A_TYPE, OWNER2, OWNER3 and MSA_ARF1
Current clients receiving hospice care certified by Medicare/Medicaid	A_TYPE, CERTIF, OWNER2 and OWNER3
Current clients over 64 years old receiving hospice care	A_TYPE, OWNER2, MSA_ARF1, REGION34 and NUHOMRES
Current clients with neoplasms receiving hospice care	A_TYPE, OWNER2 and OWNER3
Female current clients receiving hospice care	A_TYPE and OWNER2
Discharged clients 65-69 years old who received home care	A_TYPE
Discharged clients who received hospice care	A_TYPE, OWNER2 and OWNER3

*These covariates are represented by the following:

A_TYPE - indicates if agency is home health agency, *CERTIF* - indicates if agency is medicare/medicaid certified, *OWNER2* - indicates nonprofit agency, *OWNER3* - indicates government agency, *MSA_ARF1* - indicates if MSA county, *PCTMEX1* - indicates if county Mexican origin population exceeds 15 percent, *LPNFT* - number of full time LPN/LVN in short term general hospitals per 100,000 population, *PCTBLKNH* - county percent black non-Hispanic population, *MEDFMINC* - county median family income, *NUHOMRES* - 1986 number of nursing home residents per 100,000 persons, *REGION34* - indicates if county in South or West Census region

3.2 Mixed effect modeling

Using the covariates selected for the fixed effects model, random effects were added to help account for local geographic variation that could impact on the variance of a clustered sample. A probit mixed model was used, due to the availability of a fast, numerically exact, Gibbs sampler (Albert and Chib 1993). It was thought that the variables selected using the logistic model would differ little if redone using a probit fixed model, since the link functions are similar.

Specifically, the model includes the covariates selected in the fixed effect modeling process (sec. 3.1). In addition, an agency-level and a county-level random effect were also included to account for correlation between clients within an agency and between agencies in a county. The values of the parameters in this model are determined from the sample data and the underlying model via the Gibbs sampler. For a general description of the Gibbs sampler and a description of posterior inference for parameters, see Smith and Roberts (1993).

Let i index a county, j index an agency within county i and k index a client within agency ij . First a unique agency effect is specified by assuming that the probability of having a characteristic is constant within an agency. That is,

$Pr(X_{ijk}=1|p_{ij})=p_{ij}$, independent, and

$$p_{ij} = \int_{-\infty}^{\mu_{ij}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

The agency effect, μ_{ij} , is specified as a function of a county-level random effect, α_i , a fixed effect, $Z_{ij}\beta$, plus random error; i.e.,

$$\mu_{ij} \sim N(\alpha_i + Z_{ij}\beta, \sigma^2)$$

Lastly, a unique county effect is specified as the random effect:

$$\alpha_i \sim N(0, \gamma^2)$$

A noninformative prior, $p(\beta, \sigma^2, \gamma^2) \propto c$, was also used, allowing the resulting estimates to be influenced by the sample data.

Taken together, the model/prior specification implies a unique posterior for all parameters.

The Gibbs sampler requires a "burn-in" time in which iterates begin to reach their stationary distribution. Values of the parameters were kept at each iteration and visually assessed for stability. The variance of the variance estimates outlined above require independent iterates. It is well known

that Gibbs sampler iterates are correlated, with nearby iterates being more correlated. In order to use simple estimates based on independent iterates, we determined a subsequence of iterates far enough apart so that the population predictions were nearly independent (see Geyer, 1991).

Table 2 presents a summary of the posterior distribution of β , σ^2 and γ^2 for the percent female current clients receiving hospice care. The covariates selected for this variable were two agency-level indicators: whether the agency is a home-health agency, and whether the agency is government operated. No county-level covariates appeared to be related to this variable. These moments were determined numerically using the Gibbs sampler. As can be seen, a negative home health effect and a positive nonprofit agency effect indicates that women receiving hospice care are relatively less likely to use home health agencies and relatively more likely to use a nonprofit agency. The within county agency variance is by far the major source of variation. The between county variance is practically zero. There are 287 counties that contain more than one sampled agency, so this latter estimate is likely accurate. Of course conclusions about two- versus three-stage sampling cannot be drawn directly from these values since the finite population variance components are not directly comparable to the model-based components. These conclusions are in section 4.

Table 2: Posterior Moments of Model Parameters

Parameter	Mean	Standard Error
β_0 :Intercept	-0.65	.11
β_1 :Home Health	-2.39	.09
β_2 :Nonprofit	0.54	.10
σ^2	0.33	.06
γ^2	0.8×10^{-21}	$.2 \times 10^{-20}$

As a further check of the adequacy of the model, we compared the model-based estimate of the number of female current clients receiving hospice care with the published, design-based, estimate from the 1992 sample. The estimate from the model is 35,770. The standard deviation is 4,360. The published estimate that was calculated from the 1992 sample is 28,000 with a standard error of 3,200. The published estimate falls within two standard errors of the mean of the estimates derived from the model, i.e., 28,000 is above the lower bound of 27,050 in the confidence interval for the mean of the model estimates using two standard errors.

3.3 Prediction of the field agency client population sizes for the nonsample and nonresponding agencies

The current client population size that was available from the sampling frame was used as the measure of size in selecting the NHHCS agency sample. However, because of the imperfections in the sampling frame (e.g., the variable had to be imputed prior to sampling usage) and obsolescence of agency information, the client sizes reported in 1992 during NHHCS survey interviews frequently differ from the frame data. In some instances, the measures of size were not available. It was believed important that the estimated variances should reflect this component of change.

For the responding agencies, the 1992 current and discharged client population size data were available from the survey data. Using this information, models were developed to predict the 1992 current and discharged client sizes for the nonsample and nonresponding sample agencies.

Because the sampling frame includes agencies that are not in scope for the NHHCS, the first step was to develop an eligibility model using sample agencies that have known eligibility to predict the eligibility for the nonsample and nonresponding sample agencies. We included as potential predictors of eligibility agency type, agency ownership type, certification status, MSA status, and census region variables that were available from the sampling frame. Instead of the model selection procedure used to develop the fixed-effects model, we used CHAID here, which partitions the data into mutually exclusive and exhaustive subsets that best describes the dependent variable eligibility (Kass, 1980).

Using the eligibility propensity for each segment of data provided by the model we predicted the eligibility for the nonsample and nonresponding agencies.

The 1992 current and discharged client sizes for the nonsample and nonresponding agencies that were imputed as eligible were predicted using a simultaneous equation model. The model was estimated using the data for the respondent eligible agencies. The current and discharged client sizes are determined jointly by the following bivariate model:

$$\begin{aligned} \ln(N_i) &= \alpha_{1i} + \beta_1 \ln(M_i) + \tau_1 \ln(N_i') + \theta_1 z_{1i} + \varepsilon_{1i} \\ \ln(M_i) &= \alpha_{2i} + \beta_2 \ln(N_i) + \tau_2 \ln(M_i') + \theta_2 z_{2i} + \varepsilon_{2i} \end{aligned} \quad (1)$$

where

N_i is the 1992 current client size for the i -th sampled agency,
 M_i is the 1992 discharged client size for the i -th sampled agency,
 N_i' is the current client size for the i -th agency according to the old data in the sampling frame,

M_i' is the discharged client size for the i -th agency according to the old data in the sampling frame,
 Z_{1i} and Z_{2i} are vectors of indicator variables for the i -th agency, including facility characteristics and some sample design variables,
 ε_{1i} and ε_{2i} are independent error terms for the i -th agency.

The parameters of the model are estimated using the full-information maximum likelihood method (FIML). For more details and a comparison of alternative derivations see Judkins and Goksel (1995).

Based on the estimated model of eligibility, an agency was determined to be eligible or not. For eligible agencies, the sizes were estimated using the model in equation (1). The variability due to estimating the total client sizes was not accounted for in the final inference. Although multiple imputations could be used to remedy this, staff time did not allow further work.

4. Using the Simulated Populations to Evaluate Sample Designs

For specific designs, estimates of the resulting variances will be compared. In addition, population variance components will be estimated. Section 4.1 contains an evaluation of the two- versus three-stage design. Section 4.2 contains an evaluation of the within agency sample size. Currently up to six clients of each type (current and discharged) are selected per agency. If the estimated measure of size (i.e., number of current clients reported in the frame) used to select agencies was accurate, this selection of current clients would be self-weighting. This section looks at the effect of selecting clients based on the actual number of current and discharged clients in sampled agencies. Lastly, section 4.3 specifies population variance components within each of the three stages of sample selection, by strata. These variance components are estimated for each of the ten client characteristics that will be used, later, for general design guidelines.

4.1 Comparing two- and three-stage designs

As mentioned in the introduction, the 1992 NHHCS had three stages. A specific two-stage design was also developed and considered for 1992; it is still an option for 1996. Thus, one of the research objectives for the next redesign is to contrast the variances for these two specific designs, based on equal sample size. The comparison between these two designs will include the effect of small design changes on the variance.

The variance for the three-stage design may be written as

$$V_3 = V_{\text{PSU}} + V_{\text{AGENCY3}} + V_{\text{CLIENT3}}$$

The variance for the two-stage design may be written as

$$V_2 = V_{\text{AGENCY2}} + V_{\text{CLIENT2}}$$

For each simulated population, we obtained the values for all five components shown above for the estimate of females receiving hospice care. We calculated V_{PSU} using the standard formula for the variance of linear statistics from a Durbin-Brewer design conditioned on known PSU populations (which we did have as a result of the simulation). We calculated V_{AGENCY3} and V_{AGENCY2} by actually generating all possible samples of agencies and then calculating the variance among them. With this methodology, the effects of the systematic PPS selection are fully reflected for the simulated population. It was quite simple to calculate V_{CLIENT3} and V_{CLIENT2} since simple random sampling without replacement was assumed within sampled agencies.

We note that the software for the exact calculation of V_{PSU} , V_{AGENCY3} and V_{AGENCY2} was rather time-consuming to write and took quite a bit of computer time and memory to run. An approach that we considered and rejected but would perhaps consider again in the future is to approximate V for a given population and sample design by simply drawing a sequence of samples from the population according to the design and then measuring the variance among the resulting estimates. Such an approach would have been easier to program and would probably have required less computer time and memory as well. The impact on computer time is not clear.

Table 3 summarizes the simulations of the ratio, V_3/V_2 , based on generating 135 populations for the characteristic: number of female hospice clients. The estimated variance ratio associated with the three-stage design was 1.55, meaning that the variance for female hospice clients using a three stage design causes the variance to be 55% higher than it would be using the two-stage design, which, it is important to note, has the same number of sample agencies and sample clients. The ratio varied substantially over the populations as evidenced by the measures of dispersion in the table. Although most of the ratios were in the range of 1.0 to 1.6, the distribution has a heavy right tail, with one ratio as large as 2.49. It is also interesting to note that some of the ratios were slightly below 1.0, implying there is a small possibility that a lower variance could be achieved with the three-stage design than with the two-stage design, based on our information.

Table 3: Summary of Distribution of ratios of 3-stage to 2-stage designs: Females with Hospice care

Mean	1.55
Standard Deviation	0.4
Median	1.44
Minimum	.86
Maximum	2.49

Fixed sample, rather than fixed cost, sample designs were used because the relative unit sampling costs, between the two designs, were unavailable. The variance ratios can be used to evaluate potential costs, however. For example, if it is decided that a two-stage design with 1000 agencies (and 6 clients per agency) provides the required precision, then a three-stage design with ~1500 agencies (and 6 clients per agency) would be necessary to provide the same expected precision. If 1500 agencies can be interviewed for lower cost in a three-stage design than 1000 in a two-stage design, then the decision should be for three stages. If, on the other hand, 1500 agencies in a three-stage design are more expensive than 1000 in a two-stage design, then the decision should be for two-stages. It has recently been estimated that \$60,000 is saved by using a three-stage instead of a two-stage design, for the current sample size. Since the total collection costs are around \$1,000,000, annually, a decision to prefer a two-stage to a three-stage design is clear, based on the evidence at hand.

4.2 Evaluating alternate within-agency sampling rules

The 1992 design employs a near constant within-agency sample size. Selecting a sample of up to six current and six discharged clients and abstracting their records takes an interviewer about one day. This procedure is operationally efficient and, if the measure of size for each agency was accurate, would result in a sample selection that is nearly self-weighting. The fact that the agency measures of size are not very accurate leads to problems. For example, if it is believed that an agency in New York City has 6 current clients, it will be given a very small probability of selection. If the data collector finds 300 current clients, then it would be necessary to select all 300 current clients in order to have a sample that is nearly self-weighting. Current procedures would not allow more than 6 current clients to be selected. As a result, those 6 sampled clients would have unbiased weights that are 50 times larger than desired.

Another goal of this study was to quantify how much design effects could be reduced if the rules for selecting clients within sample agencies were changed to select as many (or as few) as are required for a self-weighting sample.

For the simulated populations to show this effect accurately, careful modeling of the relationship between the different client counts (the measure of size used in sampling and the client counts found at the time of interview) is important.

The terms V_{PSU} , $V_{AGENCY3}$ and $V_{AGENCY2}$ do not change as a result of the within-agency sampling rule and thus do not need to be recalculated. The changes in $V_{CLIENT2}$ and $V_{CLIENT3}$ were quite easy to obtain.

Tables 4a and 4b present the percent reduction in variance of the estimated number of females receiving hospice care that can be achieved when the within agency sample sizes are allowed to approach a self-weighting design. Table 4a presents results for the 3-stage design and Table 4b presents results for a 2-stage design. The results quantify the benefit of the self-weighting design.

Table 4a: Summary of percent reduction in variance that can be achieved with a self-weighting design: 3-stage design

Mean	16.1
Standard Deviation	3.0
Median	16.2
Minimum	6.2
Maximum	27.1

Table 4b: Summary of percent reduction in variance that can be achieved with a self-weighting design: 2-stage design

Mean	13.3
Standard Deviation	3.0
Median	13.0
Minimum	6.5
Maximum	23.1

4.3 Components of variance for a simple 3-stage design

Population variance components were desired for research into optimum sample allocations for the NHHCS.

For this study, the PSU universe defined for the 1985-94 National Health Interview Survey (NHIS), also conducted by NCHS, was assumed. It was possible that the redesigned

NHHCS could be limited to the 1985-94 NHIS PSU sample to minimize the expense of recruiting and maintaining new staff for sample PSUs that were not also in the PSU samples for the remaining NCHS establishment surveys.

To simplify design research and computation, twenty-four superstrata were assumed with strata defined by four regions (northeast, midwest, south, and west), establishment type (hospice versus home health care), PSU certainty status [self-representing (SR) versus non-self-representing (NSR)] in the NHIS sample, and PSU MSA status (MSA versus non-MSA). (The MSA for the 1985-94 NHIS universe were the metropolitan statistical areas defined by the U.S. Office of Management and Budget in 1983.) While an actual three stage sample for the redesigned NHHCS would likely use smaller PSU strata, we felt that the variations between and within PSUs of those smaller strata can be reasonably approximated by those for the superstrata.

The population variance components formulas were typical for such variances in stratified clustered populations where the clusters vary in size. [The formulas for self-representing PSUs and Non-self-representing PSUs used are in Hanson, Hurowitz, and Madow (1953), pages 317-318 and 398, respectively.] For those superstrata with two stage populations (with certainty PSUs), variances between agencies and variances between clients within agencies were computed. For those superstrata with three-stage populations (with non-certainty PSUs), variances between PSUs, variances between agencies within PSUs, and variances between clients within agencies were computed.

For the estimated number of females receiving hospice care, the population variance components were calculated for 51 initial population iterations and the means, and standard deviation of those components were then used to estimate the total number of iterations required to meet the precision standard used for the study. That standard was a maximum error of five percent of each estimate for 95 percent confidence intervals. The required number of iterations vary by strata and sampling stage. The numbers of iterations ranged from 5 to 6,613. Since separate component estimates are only used to produce estimates of the final variance, the number of iterations used was calculated based on a weighted average of the variance components. Doing so considerably reduced the number of iterations needed.

5. Conclusions

The use of a hierarchical model with the Gibbs sampler offers a procedure for estimating the variance of estimates needed in sample design. By producing the simulated populations the method offers a flexible approach to evaluate various proposed sampling designs.

This approach was more difficult to implement than we had initially hoped. The final effort in terms of person hours

was at least twice the initial projections. Also, vast amounts of CPU time on large mainframes were consumed. Given what we have learned, it is likely that we could apply a more efficient, improved method in the future. One improvement would be to use more automatic variable selection procedures throughout the modeling. Another would be to obtain the variances due to sample selection via simulation instead of programming specific formulas. Also, we initially began using models that included random regression coefficients but ended with the simpler variance component model, due to lack of variation of the regression coefficients. Some time could have been saved starting with the simple random effects model and, then considering the addition of extra coefficient variation. Finally, the method could become more economical if the model was used for more than one purpose, e.g., in small area estimation and for multiple imputation of nonresponding agencies.

The quality of results rely on the accuracy of the models used to generate the simulated populations. If data used for modeling the populations are incomplete or erroneous, the quality of results is further subject to accuracy of imputation for the missing or erroneous data. Careful model checking and evaluation are always needed.

Short of taking a complete census, what alternatives exist? First, one can adopt simplifying assumptions in the design process and hope that they hold in practice. A second option is to modify easily estimated population parameters for use in evaluating the effects of other designs. This latter option may work if the design is robust to changes in the population. In the future, it would be good to compare the resource requirements and values for variance estimates produced via the Gibbs sampler methods with the resource requirements and values of corresponding estimates produced with more traditional procedures.

In summary, when the effects of all stages of sampling may be important to the design or when design-unbiased estimates are problematic, the method described in this paper provides a doable and systematic approach to design selection.

References:

- Albert, J.H. and Chib, S. (1993) "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, 88, pp. 669-679.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory, Volume I*. Wiley Classics Library Edition, 1993. New York: Wiley and Sons.
- Judkins, D. and Goksel, H. (1995) "Report on task two of the alternative survey design", Westat technical report.
- Geyer, C.J. (1991) "Markov Chain Monte Carlo Maximum Likelihood", *Proceedings of the 23-rd Symposium on the Interface*. ed. Keramidas, E.M., 156-163.
- Kass, G.V. (1980), "An Exploratory Technique for Investigating large Quantities of categorical Data", *Applied Statistics*, 29, No.2, pp. 119-127.
- Landwehr, J.M., Pregiborn, D., and Shoemaker, A.C. (1984), "Graphical Methods for Assessing Logistic Regression Models", *Journal of American Statistical Association*, 79, 61-83.
- Malec, D., Sedransk J. and Tompkins L. (1993) "Bayesian Predictive Inference for Small Areas for Binary Variables in the National Health Interview Survey", In: *Case Studies in Bayesian Statistics*, eds. C. Gatsonis, J.S. Hodges, R.E. Kass and N.D. Singpurwalla. Springer-Verlag.
- Schafer, J.L., Khare, M., Little, J.R.A., and Rubin, D.B. (1993). "Multiple Imputation of NHANES III", *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp 297-302.
- Smith, A.F.M. and Roberts, G.O. (1993) "Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods", *Journal of the Royal Statistical Society, Series B*, 55, 3-23.
- Sukhatme, P.V. and Sukhatme, B.V. (1970). *Sampling Theory of Surveys with Applications*. Second Edition. Ames, IA: Iowa State University Press.