A Bayesian Prediction for Undecided Voters

Balgobin Nandram, Worcester Polytechnic Institute, Woercester, MA balnan@wpi.edu, and
Jai Won Choi, NCHS, 3311 Toledo Road, Hyattsville, MD 20782, jwc7@cdc.gov

1. Introduction

It is a common practice to use two-way categorical tables to present survey data. Our application is to "predict" the winner in an election using tables constructed from a short series of polls. For many surveys, there are missing data and this gives rise to partial classification of the sampled individuals. Thus, for the two-way table there are both item nonresponse (one of the two categories is missing) and unit nonresponse (both categories are missing). One may not know how the data are missing, and a model that includes some difference between the observed data and missing data (i.e., nonignorable missing data) may be preferred. For a general $r \times c$ categorical table we address the issue of estimation of the cell probabilities of the two-way table. This problem is important because, with a substantial number of undecided voters, an election prediction based on only the partially observed data may be misleading.

Essentially there are four two-way tables, one table with all completers and three supplemental tables. Of the three supplemental tables, the first with only row classifications and the second with only column classifications correspond to item nonresponse, and the third without any classification corresponds to unit nonresponse. We have extended the ignorable and nonignorable nonresponse models for two-way categorical tables (see Nandram, Cox and Choi 2003) to accommodate a third category (e.g., time in a sequence of election polls). We have extended these models further to include a time-dependent nonignorable nonresponse structure. The inclusion of the time-dependent structure can provide a more efficient prediction. A Bayesian method permits modeling different patterns of missingness under ignorability, nonignorability, and time-dependent model assumptions.

Our application is in Ohio governor's election, and there are several related problems. They categorize the sampled persons by two types of attributes and analyze the cells of such categorical tables. However, only partial classification of the individuals is available because some individuals are classified by at most one attribute, and others are left unclassified. Specifically, we use tabular data from the Ohio polls to study the relation between a measure of voters' status (likely to vote, unlikely to vote, and undecided) and candidacy (Fisher, Taft, others and undecided) to illustrate our methodology. It is interesting that voters' status are related to candidates. Also, it is desirable to make an adjustment for undecided data because the proportion of undecided voters is usually high.

There are many publications related to our research (Nandram, Han and Choi 2002, Nandram and Choi 2002 a,b, Forster and Smith 1998, Chen and Fienberg 1974, Chen and Stasny 2003, Little and Rubin 1987, Little and Rubin 1987, Little 1985, and Nandram, Cox and Choi 2003). The key difference of our paper from these papers is the introduction of a Bayesian method to analyze data from an $r \times c$ categorical table when there are both item and unit nonresponses, and the missing data mechanism can be nonignorable and time-dependent.

About 73% of the voters are completely classified, 27% have no decision about candidacy, only 1% did not know whether they would vote or not, and only 5 persons completely unclassified among the 648 participants. The data set, used in our study, is presented in Table 1 as a 4×3 categorical table of candidates and voters' status. Our problem is to predict the winning candidate by estimating the proportion of final votes for each candidate.

In Section 2, we describe the methodology to obtain estimates of the cell probabilities incorporating the two types of missing data, and we show how to expand an ignorable nonresponse model into a nonignorable nonresponse model and time-dependent model. We also show how to use Markov chain Monte Carlo methods to fit the nonignorable nonresponse model. In Section 3, we analyze the Ohio election data to demonstrate the versatility of our methods.

2. Methodology

We have constructed a time-dependent nonignorable nonresponse model for the 1998 Ohio Poll data. For comparison we have also considered two other models, an ignorable and a nonignorable nonresponse model. These latter two models are not time dependent because we assume that the three time points come from the same stochastic process. Essentially we start with the ignorable nonresponse model which is expanded into a nonignorable nonresponse model, and we extend the nonignorable nonresponse model to a time-dependent model.

Let $I_{tjk\ell} = 1$ or 0, time $t = 1, \dots, T$, $j = 1, \dots, r$, $k = 1, \dots, c$, $\ell = 1, \dots, L$ denote the characteristic (observed or missing) of an individual in the two-way table (i.e., the row and column the individual belongs to). In our application $T = 3$, $r = 2$ and $c = 3$. Let $\underset{\sim}{J}_{t\ell} = 1$ or 0 (1 if the subject belongs to one of the situations: $1 = (1, 0, 0, 0)$, $2{=}(0, 1, 0, 0)$, $3{=}(0, 0, 1, 0)$, or $4{=}(0, 0, 0, 1)$ of the four tables (one with completers and three supplemental tables). (e.g., $\underset{\sim}{J}_{t\ell} = (1, 0, 0, 0)$ indicates that the individual belongs to the completely observed table).

Let $P(J_\ell = 1 \mid I_{jk\ell} = 1,\ I_{j'k'\ell} = 0,\ j' \neq j, k' \neq k) = \pi_{jk}$. If $\pi_{jk} = \pi$, the model is ignorable (see Rubin 1976 for further explanation); otherwise, the model is nonignorable. We do not know whether an ignorable nonresponse model or a nonignorable nonresponse model is appropriate, but one may have uncertainty about the ignorability of undecided voters in election polls.

Let the cell counts be $y_{tsjk} = \sum_{\ell=1}^{n} I_{tjk\ell} J_{ts\ell}, s = 1, 2, 3, 4$ for the four cases. Here y_{t1jk} are observed and $y_{tsjk}, s = 2, 3, 4,\ t = 1, \dots, T$ are missing (i.e., latent variables). For y_{t1jk} we know that $\sum_{j=1}^{r} \sum_{k=1}^{c} y_{t1jk} = n_{t0}$, the number of individuals with complete data. For y_{t2jk} we know that $\sum_{k=1}^{c} y_{t2jk} = u_{tj}$, where the row margins $u_{tj},\ j = 1, \dots, r$ are observed. For y_{t3jk} we know that $\sum_{j=1}^{r} y_{t3jk} = v_{tk}$, where the column margins $v_{tk},\ k = 1, \dots, c$ are observed. For y_{t4jk} we know that $\sum_{j=1}^{r} \sum_{k=1}^{c} y_{t4jk} = w_t$. In this analysis n_{t0}, $\underset{\sim}{u}_t$, $\underset{\sim}{v}_t$ and w_t are held fixed (i.e., fixed margin analysis) and known. Whenever it is convenient, we will use notations such as

$$\sum_{s,j} y_{tsj} \equiv \sum_{s=1}^{4} \sum_{j=1}^{r} y_{tsj},\ \ \prod_{s,j} \pi_{tsj} \equiv \prod_{s=1}^{4} \prod_{j=1}^{r} \pi_{tsj}$$

and $\underset{\sim}{y}_{t(1)} = (\underset{\sim}{y}_{t2}, \underset{\sim}{y}_{t3}, \underset{\sim}{y}_{t4})$, $\underset{\sim}{y}_{t(2)} = (\underset{\sim}{y}_{t1}, \underset{\sim}{y}_{t3}, \underset{\sim}{y}_{t4})$ etc., where $\underset{\sim}{y}_{ts} = (y_{tsjk}, j = 1, \dots r,\ k = 1, \dots, c),\ t = 1, \dots, T,\ s = 1, 2, 3, 4$. Also, $\sum_{s,j,k}^{4,r,c} y_{tsjk} = n_t$. We will also use $y_{ts\cdot\cdot} = \sum_{j,k} y_{tsjk},\ y_{t\cdot jk} = \sum_{s} y_{tsjk}$ etc. and $\underset{\sim}{y}_t = (\underset{\sim}{y}_{t1}, \underset{\sim}{y}_{t2}, \underset{\sim}{y}_{t3}, \underset{\sim}{y}_{t4})$.

In Section 2.1 we describe the three models, and in Section 2.2 we show how to fit the time-dependent model. The ignorable and nonignorable nonresponse models can be fit in a similar manner.

2.1 Nonresponse Models

Letting $\underset{\sim}{I}_{t\ell} = (I_{tjk\ell}, t = 1, \dots, T,\ j = 1, \dots, r,\ k = 1, \dots, c)$, for all models, we take

$$\underset{\sim}{I}_{t\ell} \mid \underset{\sim}{p}_t \overset{iid}{\sim} \text{Multinomial}\{1,\ \underset{\sim}{p}_t\}, \tag{1}$$

where $\sum_{j=1,k=1}^{r,c} p_{tjk} = 1,\ p_{tjk} \geq 0,\ t = 1, \dots, T,\ j = 1, \dots, r,\ k = 1, \dots, c$.

For the ignorable nonresponse model we take

$$\underset{\sim}{J}_{t\ell} \mid \underset{\sim}{\pi}_t \overset{iid}{\sim} \text{Multinomial}\{1, \underset{\sim}{\pi}_t\}. \tag{2}$$

That is, there is no dependence on the cell status of an individual. For the nonignorable nonresponse models we take

$$\underset{\sim}{J}_{t\ell} \mid \{I_{tjk\ell} = 1, \underset{\sim}{\pi}_{tjk}\} \overset{iid}{\sim} \text{Multinomial}\{1, \underset{\sim}{\pi}_{tjk}\}. \tag{3}$$

Assumption (3) specifies that the probabilities an individual belongs to one of the four tables depend on the two characteristics (i.e., row and column classification) of the individual. In this manner we incorporate the assumption that the missing data is nonignorable.

It follows from (1) and (2) that for the ignorable model

$$g(\underset{\sim}{p}_t, \underset{\sim}{\pi} \mid \underset{\sim}{y}_{t1}) \propto \prod_{t=1}^{T} \left[\prod_{s=1}^{4} \pi_{ts}^{y_{ts\cdot\cdot}} \right] \left[\prod_{j=1}^{r} \prod_{k=1}^{c} p_{tjk}^{y_{t1jk}} \right] \tag{4}$$

subject to $\sum_{k=1}^{c} y_{t2jk} = u_{tj}, j = 1, ..., r$, $\sum_{j=1}^{r} y_{t3jk} = v_{tk}, k = 1, ..., c$, and $\sum_{j=1}^{r} \sum_{k=1}^{c} y_{t4jk} = w_t$. Note that under ignorability the likelihood function in (4) separates into two pieces, one that contains the π_{ts} only and the other the p_{tjk}, and inference about these two parameters are unrelated. Also, it follows from (1) and (3) that for the nonignorable nonresponse models the augmented likelihood function for $\underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{y}_{(1)} \mid \underset{\sim}{y}_1$ is

$$g(\underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{y}_{(1)} \mid \underset{\sim}{y}_1) \ \ \propto \ \ \prod_{t=1}^{T} \left[\prod_{s,j,k}^{4,r,c} \frac{\pi_{sjk}^{y_{tsjk}}}{y_{tsjk}!} \prod_{j,k}^{r,c} p_{tjk}^{y_{t\cdot jk}} \right] \tag{5}$$

subject to $\sum_{k=1}^{c} y_{t2jk} = u_{tj}, j = 1, ..., r$, $\sum_{j=1}^{r} y_{t3jk} = v_{tk}, k = 1, ..., c$, and $\sum_{j=1}^{r}\sum_{k=1}^{c} y_{t4jk} = w_t$. See Nandram, Cox and Choi (2003) for a description of identifiability in a similar situation.

We next describe the stochastic models for the $\underset{\sim}{p}_t$ and $\underset{\sim}{\pi}_{tjk}$.

For the ignorable and nonignorable nonresponse models, we take

$$\underset{\sim}{p}_t \mid \underset{\sim}{\mu}_2, \tau_2 \overset{iid}{\sim} \text{Dirichlet } (\underset{\sim}{\mu}_2\tau_2). \tag{6}$$

The probabilistic structure in (6) permits a "borrowing of strength" across time. For the time-dependent model, we take

$$\underset{\sim}{p}_t \mid \underset{\sim}{p}_{t-1}, \tau_2 \overset{iid}{\sim} \text{Dirichlet } (\underset{\sim}{p}_{t-1}\tau_2),\ t = 1, \ldots, T. \tag{7}$$

Note that

$$E\{\underset{\sim}{p}_t \mid \underset{\sim}{p}_{t-1}, \tau_2\} = \underset{\sim}{p}_{t-1}, t = 1, \ldots, T;$$

so that $\{p_t\}$, a priori, is a martingale vector. Note also that for the latent vectors $\underset{\sim}{p}_0$ and $\underset{\sim}{p}_{T+1}$ we have taken $\underset{\sim}{p}_0,\ \underset{\sim}{p}_{T+1} \overset{iid}{\sim}$ Dirichlet $(\underset{\sim}{1})$.

For the ignorable nonresponse model, we take

$$\underset{\sim}{\pi}_t \mid \underset{\sim}{\mu}_1, \tau_1 \overset{iid}{\sim} \text{Dirichlet } (\underset{\sim}{\mu}_1\tau_1),\ t = 1, \ldots, T \tag{8}$$

and for the nonignorable nonresponse models we take, t=1,…,T, j=1,…,r, k=1,…,c,

$$\underset{\sim}{\pi}_{tjk} \mid \underset{\sim}{\mu}_1, \tau_1 \overset{iid}{\sim} \text{Dirichlet } (\underset{\sim}{\mu}_1\tau_1) \tag{9}$$

Thus, again we have a "borrowing of strength" across time.

Finally, we specify prior densities for the hyperparameters. First, we take

$$\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2 \overset{iid}{\sim} \text{Dirichlet } (\underset{\sim}{1}), \tag{10}$$

essentially noniformative prior densities. Except for τ_1 for the nonignorable nonresponse models, τ_1 and τ_2 are independent and identically distributed random variables from

$$f(x) = 1/(1+x)^2, x \geq 0. \tag{11}$$

Again this is an essentially noniformative prior density.

For the nonignorable nonresponse models we need to be more careful to specify the prior density of τ_1 because $\underset{\sim}{\pi}_{tjk}$ are not identifiable. Here we attempt to "center" the nonignorable nonresponse models on the ignorable nonresponse model. In (9) the parameter τ_1 tells us about the closeness of the nonignorable model to the ignorable model. For example, if τ_1 is small, the $\underset{\sim}{\pi}_{jk}$ will be very different, and if τ_1 is large, the $\underset{\sim}{\pi}_{jk}$ will be very similar. Thus, inference will be sensitive to the choice of τ_1, and one has to be careful in choosing τ_1. We would like to choose a prior density for τ_1 so that the nonignorable nonresponse model is kept close to the ignorable nonresponse model. Thus, we take

$$\tau_1 \sim \text{ Gamma } (\frac{1}{c_0^2}, \frac{1}{\mu_0 c_0^2}) \tag{12}$$

where $E(\tau_1) = \mu_0$ and c_0 is the coefficient of variation of τ_1; both μ_0 and c_0 are to be specified. We will do so using samples from the posterior density of τ_1 under the ignorable nonresponse model.

2.2 Fitting the Time-Dependent Nonignorable Nonresponse Model

Combining the likelihood function in (5) with all the priors (5)-(12) via Bayes' theorem, the joint posterior density of the parameters $\underset{\sim}{\pi}$, $\underset{\sim}{p}$ $\underset{\sim}{\mu}_1$, τ_1 τ_2 and the latent variables $\underset{\sim}{y}_{(1)}$ is $\pi(\underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{\mu}_1, \tau_1, \tau_2 \underset{\sim}{y}_{(1)} \mid \underset{\sim}{y}_1)$,

$$\propto \tau_1^{1/c_0^2 - 1} e^{-\tau_1/\mu_0 c_0^2} \frac{1}{(1+\tau_2)^2} \tag{13}$$

$$\times \prod_{t=1}^{T} \left[\prod_{s,j,k}^{4,r,c} \frac{\pi_{sjk}^{y_{tsjk}}}{y_{tsjk}!} \right] \left[\prod_{j,k}^{r,c} p_{tjk}^{y_{t\cdot jk}} \right]$$

$$\times \prod_{t=1}^{T} \left\{ \frac{\prod_{j=1,k=1}^{r,c} p_{tjk}^{p_{t-1jk}\tau_2 - 1}}{D(\underset{\sim}{p}_{t-1}\tau_2)} \prod_{j=1,k=1}^{r,c} \frac{\prod_{s=1}^{4} \pi_{tsjk}^{\mu_{1s}\tau_1 - 1}}{D(\underset{\sim}{\mu}_1\tau_1)} \right\}$$

subject to $\sum_{k=1}^{c} y_{t2jk} = u_{tj},\ j = 1, ..., r$, $\sum_{j=1}^{r} y_{t3jk} = v_{tk},\ k = 1, ..., c$, and $\sum_{j=1}^{r}\sum_{k=1}^{c} y_{t4jk} = w_t,\ t = 1, \ldots, T$.

The posterior density in (13) is complex, so we will use Markov chain Monte Carlo methods to fit it. However, it is easy to fit the time-dependent model using the griddy Metropolis-Hastings sampler as we will describe. Also, in a similar manner using the griddy Gibbs sampler, it is easy to fit the ignorable and the nonignorable nonresponse models. We obtain a sample from the joint posterior density in order to make inference about the parameters. Specifically, we need to make inference about $\underset{\sim}{p}_t$. To run the Metropolis-Hastings sampler, we need the conditional posterior density of each of the parameters given the others.

3. Data Analysis

In this section we compare our models with those of Chen and Stasny (2003) and the actual outcomes.

We have introduced a new parameter to help predict the outcome of the election. We also study extensively sensitivity of inference with choices of κ_1 and κ_2. Based on our procedure, we have specified the coefficient of variation, $c_0 = 0.65$, and the mean, $\mu_0 = 0.50$, of the prior distribution of τ_1.

In Table 2 we compare inference about the proportion of October votes allocated to the three candidates by our models and those of Chen and Stasny (2003). In this table the results are based on the prior $\tau_1 \sim \text{Gamma}(1/c_0^2,\ 1/\mu_0 c_0^2)$. We also present the actual proportions taken from the Chang and Krosnick (1998). The actual proportions are (.45, .50, .05) for Fisher, Taft, and Other. Using our time-dependent nonresponse model these proportions are estimated to be (.44, .49, .07). These compare favorably with the actual outcomes. The corresponding estimates are (.41, .51, .08) for the ignorable nonresponse model and (.36, .56, .08) for the nonignorable nonresponse model. The best result of Chen and Stasny (2003) is obtained from their Model D, and their estimates are (.42, .51,.07); but there is no measure of the reliability of their estimates. We have provided 95% credible intervals for our estimates, and this is a clear advantage over Chen and Stasny (2003) who have proposed to provide measures of variability in the future.

Although our estimates from the time-dependent model are close to the actual estimates, the 95% credible intervals for p_{311} and p_{312} overlap, thereby making it difficult to predict that Taft is the winner. Although the 95% credible intervals for our other models are shorter, the point estimates are not so good and they still overlap. One weakness in our analysis in Table 2 is that we have ignored the correlation between the two estimates (i.e., we should really study the difference $p_{312} - p_{311}$, the margin of winning).

In Table 3 we present estimates of $p_{312} - p_{311}$, and sensitvity to κ_1 and κ_2 which we have chosen as $\kappa_1 = .01,\ .05,\ .5, 1$ and $\kappa_2 = 1,\ 5,\ 25,\ 100$. Inference is unaffected by these choices (i.e., all the 95% credible intervals - not presented - contain 0). However, there are some changes in the posterior means (PM) and the posterior standard deviation (PSD) of $p_{312} - p_{311}$ but these do not affect the intervals.

It is surprising that the PMs of $p_{312} - p_{311}$ for the time-dependent model are sometimes larger than for the nonignorable nonresponse model (see Table 2). A possible explanation is that the posterior density of τ_2 under the time-dependent nonresponse model is to the right of the one under the nonignorable nonresponse model.

We seek an alternative parameter to help us predict the winner more convincingly. We pose the following question: "What is the probability that the proportion of Taft's voters is larger than that of Fisher's voters?"

Thus, it is pertinent to consider the parameter $\Delta = Pr(p_{312} > p_{311} \mid p_{311} + p_{312} + p_{313}, \underset{\sim}{\alpha})$ where $\alpha_1 = \mu_{11}\tau_2,\ \alpha_2 = \mu_{12}\tau_2,\ \alpha_3 = \mu_{13}\tau_2$ for the ignorable and nonignorable nonresponse models, and $\alpha_1 = p_{211}\tau_2,\ \alpha_2 = p_{212}\tau_2,\ \alpha_3 = p_{213}\tau_2$ for the time-dependent model. In either case, letting $q_1 = p_{311}/p_{31.}$, $q_2 = p_{312}/p_{31.}$, and $q_3 = p_{313}/p_{31.}$ with $p_{31.} = \sum_{k=1}^{3} p_{31k}$ and $\sum_{k=1}^{3} q_k = 1$, we have

$$\begin{aligned} \Delta &= Pr(q_2 > q_1 \mid \underset{\sim}{\alpha}) \\ &= \int_0^1 \int_{q_1}^{1-q_1} \frac{q_1^{\alpha_1 - 1} q_2^{\alpha_2 - 1}(1 - q_1 - q_2)^{\alpha_3 - 1}}{D(\alpha_1, \alpha_2, \alpha_3)} dq_2 dq_1. \end{aligned}$$

Then, it is easy to show that Δ is

$$1 - \int_0^{1/2} F_{\alpha_2,\alpha_3}\{q_1/(1-q_1)\} \frac{q_1^{\alpha_1 - 1}(1 - q_1)^{\alpha_2 + \alpha_3 - 1}}{B(\alpha_1, \alpha_2 + \alpha_3)} dq_1$$

where $F_{\alpha_2,\alpha_3}(a) = \int_0^a \frac{x^{\alpha_2 - 1}(1-x)^{\alpha_3 - 1}}{B(\alpha_2 \alpha_3)} dx$. For each $\alpha_1, \alpha_2, \alpha_3$, Δ can be computed using Monte Carlo integration with $q_1 \sim \text{Beta}(\alpha_1, \alpha_2 + \alpha_3)$ as an importance function. Thus, for each $\underset{\sim}{\alpha}^{(h)}$, $h = 1, \ldots, M$, M=1,000 from the Metropolis-Hastings sampler, we can compute $\Delta^{(h)}$. A posteriori inference about Δ is obtained in the standard empirical manner.

In Table 4 we compare inference about $p_{312} - p_{311}$ and Δ with respect to time and model. Except for January poll under the ignorable nonresponse model, all the 95% intervals for $p_{312} - p_{311}$ contain 0. For both the ignorable and nonignorable nonresponse models the intervals for Δ contain .5 showing difficulty in predicting the winner. It is amazing that our time-dependent model gives a 95% credible interval of $(.68, 1.00)$ for Δ (Δ lies to the right of .5), showing convincingly that Taft is the winner. Our time-dependent model provides posterior inference that are closer to the truth than the ignorable and nonignorable nonresponse models as well as those of Chen and Stasny (2003). Our new parameter shows convincingly that Taft is the winner, as is evident.

References

Chen, Q. L. and Stasny, E. A. (2003), "Handling Undecided Voters: Using Missing Data Methods in Election Forecasting," *Technical Report*, Department of Statistics, The Ohio State University.

Chang, L. C. and Krosnick, J. A. (2001), "Improving Election Forecasting," *Draft Manuscript*, Department of Psychology, The Ohio State University.

Chen, T. and Fienberg, S. E. (1974), "Two-dimensional Contingency Tables with Both Completely and Partially Cross-classified Data," *Biometrics*, 30, 629-642.

Cohen, G. and Duffy, J. C. (2002), "Are Non-respondents to Health Surveys Less Healthy Than Respondents," *Journal of Official Statistics*, 18, 13-23.

Draper, D. (1995), "Assessment and Propagation of Model Uncertainty" (with discussion), *Journal of the Royal Statistical Society*, Series B, 57, 45-97.

Forster, J. J. and Smith, P. W. F. (1998), "Model-based Inference for Categorical Survey Data Subject to Non-ignorable Nonresponse," *Journal of the Royal Statistical Society*, Series B, 60, 57-70.

Kalton, G. and Kasprzyk, D. (1986), "The Treatment of Missing Survey Data," *Survey Methodology*, 12, 1-16.

Little, R. J. (1985), "Nonresponse Adjustments in Longitudinal Surveys: Models for Categorical Data," *Bulletin of the International Statistical Institute*, 15.1, 1-15.

Little, R. J. A. and Rubin D. B. (1987), *Statistical Analysis with Missing Data*, New York: Wiley.

Nandram, B. and Choi, J. W. (2002a), "Hierarchical Bayesian Nonresponse Models for Binary Data from small areas with Uncertainty about Ignorability," *Journal of the American Statistical Association*, 97, 381-388.

Nandram, B. and Choi, J. W. (2002b), "A Bayesian Analysis of a Proportion under Nonignorable Nonresponse," *Statistics in Medicine*, 21, 1189-1212.

Nandram, B., Cox, L. H. and Choi, J. W. (2003), "Bayesian Analysis of Nonignorable Missing Categorical Data: An Application to Bone Mineral Density and Family Income," *Technical Report*, Department of Mathematical Sciences, Worcester Polytechnic Institute.

Nandram, B., G. Han and Choi, J.W. (2002), "A Hierarchical Bayesian Nonignorable Nonresponse Model for Multinomial Data From Small Areas," *Survey Methodology*, 28, 145-156.

Ritter, C. and Tanner, M. A. (1992), "The Gibbs Stopper and the Griddy Gibbs Sampler," *Journal of the American Statistical Association*, 1992; 87: 861-868.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581-592.

Tanner, M. A. (1996), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 3^{rd} Edition, New York: Springer.

Wang, H. (2001), "Two-way contingency tables with marginally and conditionally imputed non-respondents," *Ph.D. Dissertation*, Department of Statistics, University of Wisconsin-Madison.

Table 1: Classification of October 1998 Buckeye State Poll by voting status (likely to vote-L, Not likely to vote-N, Undecided-U) and candidate (Fisher-F, Taft-T, Other-O, Undecided-U, Total-TL)

	Candidate				
Status	F	T	O	U	TL
a. 1-1998					
L	127	183	8	109	427
N	57	94	4	59	214
U	0	2	0	5	7
TL	184	279	12	173	648
b. 4-1998					
L	114	135	1	61	311
N	104	149	3	78	334
U	2	6	0	3	11
TL	220	290	4	142	656
c. 10-1998					
L	112	140	23	61	336
N	96	108	21	73	298
U	7	11	1	4	23
TL	215	259	45	138	657

NOTE: These data are taken from Chang and Krosnick (1998); Chen and Stasny (2003) used a very similar data set.

Table 2: the proportion of likely voters by status for the October 1998 poll for different models (Sample estimate-S, Chen/Stasny models A,B,C-ABC, Chen/Stasny model D-D, Chen/Stasny model E-E, Ignorable model-I, Nonignorable model-N, Time dependent nonignorable model-TN), with actual outcome-A, of the November 1998 election by candidates (Fisher-H, Taft-T, Other-O). CI stands for 95% confidence interval.

	Candidate		
Status	F	T	O
S	.41	.51	.08
A	.45	.50	.05
ABC	.41	.51	.08
D	.42	.51	.07
E	.41	.51	.08
I	.41	.51	.08
CI	(.35, .47)	(.45, .57)	(.06, .12)
N	.36	.56	.08
CI	(.29, .50)	(.42, .63)	(.05, .11)
TN	.44	.49	.07
CI	(.30, .55)	(.38, .63)	(.04, .11)

NOTE: In this table $c_0 = .65$ and $\mu_0 = .50$.

Table 3: Sensitivity of the posterior mean (PM) and the posterior standard deviation (PSD) of $p_{312} - p_{311}$ with respect to changes in κ_1 and κ_2 for models, A for Nonignorable Nonresponse Model, and B for Time-Dependent Model

	κ_2			
κ_1	1	2	3	4
A				
1	.19 (.08)	.11 (.08)	.10 (.07)	.10 (.06)
2	.15 (.12)	.12 (0.09)	.10 (.07)	.10 (.06)
3	.20 (.10)	.15 (.13)	.19 (.12)	.15 (.13)
4	.20 (.10)	.22 (.07)	.16 (.14)	.21 (.09)
B				
1	.09 (.12)	.13 (.08)	.10 (.07)	.10 (.06)
2	.14 (.13)	.11 (.08)	.11 (.07)	.10 (.06)
3	.17 (.13)	.18 (.13)	.19 (.11)	.20 (.09)
4	.05 (.15)	.19 (.13)	.22 (.07)	.20 (.10)

NOTE: We have taken $\tau_1 \sim$ Gamma $(\frac{1}{\kappa_1^2 c_0^2}, \frac{1}{\kappa_2 \mu_0 \kappa_1^2 c_0^2})$ and we study sensitivity with respect to κ_1=$.01, .05, .5, 1.0$, and κ_2=$1, 5, 25, 100$. For the ignorable nonresponse model $p(\tau_1) = 1/(1+\tau_1)^2$, $\tau_1 > 0$ (i.e., τ_1 does not depend on κ_1 and κ_2), and PM and PSD are .10 and .06 respectively.

Table 4: Posterior mean (PM), posterior standard deviation (PSD) and 95% credible interval of $p_{312} - p_{311}$ and Δ by time and model

Time	PM	PSD	CI
a. Ignorable Nonresponse Model			
Jan	.18	.05	(.07, .29)
Apr	.09	.06	(-.03, .21)
Oct	.10	.06	(-.02,.21)
Δ	.69	.16	(.32,.93)
b. Nonignorable Nonresponse Model			
Jan	.30	.13	(-.13,.45)
Apr	.20	.11	(-.12,.35)
Oct	.20	.10	(-.07,.33)
Δ	.63	.17	(.26,.91)
c. Time-Dependent Model			
Jan	.09	.21	(-.18,.42)
Apr	.04	.17	(-.21,.36)
Oct	.05	.15	(-.17,32)
Δ	.88	.10	(.68,1.00)

NOTE: We have taken $\tau_1 \sim$ Gamma $(\frac{1}{\kappa_1^2 c_0^2}, \frac{1}{\kappa_2 \mu_0 \kappa_1^2 c_0^2})$ and we study sensitivity with respect to κ_1 and κ_2.