Structured Measurement Error in Nutritional Epidemiology: Applications in the Pregnancy, Infection, and Nutrition (PIN) Study

Brent A. JOHNSON, Amy H. HERRING, Joseph G. IBRAHIM, and Anna Maria SIEGA-RIZ

Preterm birth, defined as delivery before 37 completed weeks' gestation, is a leading cause of infant morbidity and mortality. Identifying factors related to preterm delivery is an important goal of public health professionals who wish to identify etiologic pathways to target for prevention. Validation studies are often conducted in nutritional epidemiology in order to study measurement error in instruments that are generally less invasive or less expensive than "gold standard" instruments. Data from such studies are then used in adjusting estimates based on the full study sample. However, measurement error in nutritional epidemiology has recently been shown to be complicated by correlated error structures in the study-wide and validation instruments. Investigators of a study of preterm birth and dietary intake designed a validation study to assess measurement error in a food frequency questionnaire (FFQ) administered during pregnancy and with the secondary goal of assessing whether a single administration of the FFQ could be used to describe intake over the relatively short pregnancy period, in which energy intake typically increases. Here, we describe a likelihood-based method via Markov chain Monte Carlo to estimate the regression coefficients in a generalized linear model relating preterm birth to covariates, where one of the covariates is measured with error and the multivariate measurement error model has correlated errors among contemporaneous instruments (i.e., FFQs, 24-hour recalls, and biomarkers). Because of constraints on the covariance parameters in our likelihood, identifiability for all the variance and covariance parameters is not guaranteed, and, therefore, we derive the necessary and sufficient conditions to identify the variance and covariance parameters under our measurement error model and assumptions. We investigate the sensitivity of our likelihood-based model to distributional assumptions placed on the true folate intake by employing semiparametric Bayesian methods through the mixture of Dirichlet process priors framework. We exemplify our methods in a recent prospective cohort study of risk factors for preterm birth. We use long-term folate as our error-prone predictor of interest, the FFQ and 24-hour recall as two biased instruments, and the serum folate biomarker as the unbiased instrument. We found that folate intake, as measured by the FFQ, led to a conservative estimate of the estimated odds ratio of preterm birth (.76) when compared to the odds ratio estimate from our likelihood-based approach, which adjusts for the measurement error (.63). We found that our parametric model led to similar conclusions to the semiparametric Bayesian model.

KEY WORDS: Adaptive rejection sampling; Dirichlet process prior; MCMC; Semiparametric Bayes.

1. INTRODUCTION

Measurement error is a common and well-known challenge in nutritional epidemiology. One only has to glance at a recent issue of any one of the leading epidemiological journals to see this and to verify that there still are many unresolved questions. One of the more intriguing recent developments in nutritional epidemiology concerns the fitness and applicability of traditional error models used to assess the validity and generalizability of estimated risks obtained from studies using the food frequency questionnaire (FFQ).

Despite many documented pitfalls (Block 2001; Byers 2001; Willett 2001), including systematic biases and within- and between-subject variability, the FFQ is a common dietary instrument because of its ease of administration and economy in large nutritional studies. Naive regression methods that use the error-prone FFQ in place of the true long-term dietary intake often attenuate the regression coefficients toward 0 [although the result is not true in general nonlinear models (Fuller 1987; Carroll, Ruppert, and Stefanski 1995)]. Although several statistical methods have been proposed for the analysis of data where covariates are measured with error, regression calibration (Stefanski and Carroll 1985) seems to be the default method in nutrition (Willett 1998). The method is popular because it may be implemented using standard software assuming one has a reliable calibration model (Spiegelman, Carroll, and Kipnis 2001; Spiegelman, Zhao, and Kim 2004). In addition, much money and energy have been spent on validation studies over the past several decades; therefore, bias and variance parameters relating the FFQ to the true, long-term dietary intake can be estimated with some degree of precision. A related problem to the one considered here is the error in covariate misclassification (cf. Holcroft and Spiegelman 1999; Morrissey and Spiegelman 1999; Spiegelman et al. 2001; Zucker and Spiegelman 2004).

The traditional statistical analysis and inference proceeds by first regressing the FFQ on the outcome to obtain a naive estimate of the regression coefficient. Then we regress a reference instrument—that is, an unbiased measure for the true dietary intake—on the FFQ to estimate the attenuation factor. It can be shown that dividing the naive estimated regression coefficient by the estimated attenuation factor leads to a corrected estimate of the desired regression coefficient, that is, one obtained if we could have regressed the outcome on the true long-term dietary intake (Carroll et al. 1995; Kipnis et al. 2001). If the systematic bias or the correlated errors in the FFQ or 24-hour recall is

Brent A. Johnson is Postdoctoral Fellow (E-mail: *brentj@email.unc.edu*), Amy H. Herring is Associate Professor (E-mail: *amy_herring@unc.edu*), and Joseph G. Ibrahim is Alumni Distinguished Professor (E-mail: *ibrahim@bios.unc.edu*), Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599. Anna Maria Siega-Riz is Associate Professor, Departments of Nutrition and Epidemiology, University of North Carolina, Chapel Hill, NC 27599 (E-mail: *am_siegariz@unc.edu*). We thank the associate editor and two anonymous referees for helpful comments and suggestions that led to a much improved manuscript. The research of the first author was supported in part by grants from the National Institute for the Environmental Health Sciences (P30ES10126, T32ES007018). Dr. Herring's research was supported in part by grants from the National Institutes of Child Health and Human Development (1R03HD045780, HD37584, HD39373) and NIEHS (P30ES10126). The PIN study was funded by grants from NICHD (HD28684, HD05798, DK55865, AG09525), UNC Clinical Nutrition Research Center (DK56350), UNC General Clinical Research Resources (RR00046), and funds from the Wake Area Health Education Center in Raleigh, NC. Dr. Ibrahim's research was supported in part by NIH grants #CA 70101, #CA 69222, #GM 070335, and #AI 060373.

Journal of the American Statistical Association
???? 0, Vol. 0, No. 00, Applications and Case Studies
DOI 10.1198/00

ignored, then the attenuation factor will be biased, and subsequently, the "corrected" regression coefficient estimate will no longer be reliable. Although primary interest often lies in estimating this true regression coefficient, epidemiologists are also quite interested in the estimated attenuation factor. Because the power of the study to detect a significant effect is a function of the attenuation factor, epidemiologists use this fact to make post hoc calculations to determine whether a null finding appears, in fact, be to the case or whether it seems to be a result of low power.

Our method uses models that allow for correlation in the errors for contemporaneous instruments as suggested in the literature (Kaaks, Riboli, Esteve, Van Kappel, and Vab Staveren 1994; Kipnis et al. 2001, 2003). Our point and interval estimation method is different from that considered in Kipnis et al. (2001, 2003) in that we use a likelihood-based approach (also called a structural measurement error model), whereas Kipnis et al. (2001) estimated the attenuation coefficient first and then appropriately scaled the naive regression coefficient estimate to obtain the corrected coefficient estimate. Recently, Spiegelman et al. (2004) considered a joint model for all the parameters in the disease (or outcome) model and the measurement error model (as we do in Sec. 3) by "stacking" the estimating equations for all the unknown parameters from both the disease model and the calibration model and forming an M estimator (cf. Stefanski and Boos 2002). Again, this regression calibration approach is different from our likelihood-based approach. We subsequently extend our likelihood-based model through the mixture of Dirichlet processes (MDP) methodology to avoid placing strict parametric assumptions on the latent true dietary intake variables. The remainder of this article is organized as follows: Section 2 describes the Pregnancy, Infection, and Nutrition (PIN) study, from which the data are acquired, and scientific questions of interest; Section 3 describes our statistical model and notation; Section 4 gives an overview of the joint full conditional distribution; Section 5 summarizes a small simulation study; and Section 6 summarizes the results of our analysis; we end with a short discussion on the implications of our findings in Section 7.

2. THE PIN STUDY DATA

The PIN study was a prospective cohort study of risk factors for preterm birth (Savitz et al. 1999). Recruitment occurred between 24 and 29 weeks' gestation, and several questionnaires, including an FFQ to assess dietary intake in the second trimester, were administered at this time as described in Savitz et al. (1999; Siega-Riz et al. 2004). The outcome of interest, preterm birth, was defined as delivery before 37 completed weeks of gestation. Siega-Riz et al. (2004) examined the relationship between maternal folate status and preterm birth, reporting increased risks of preterm birth among women with mean daily folate intake less than 500 μg and among women with serum folate levels less than 16.3 ng/mL. A variety of folate exposure variables, including mean daily dietary intake from the FFQ and two biomarkers, serum and red blood cell folate, were used in separate analyses, with all results reported.

To address FFQ measurement issues, the investigators conducted a validation substudy to determine whether dietary intake changed over the course of pregnancy and to quantify measurement error in the FFQ. Women in the validation study were enrolled in the first trimester and were asked to complete three FFQs over the course of pregnancy, with each FFQ reflecting intake over the past trimester. The purpose of the longitudinal component of the validation substudy was to determine whether one FFQ measurement during the second trimester of pregnancy would be sufficient to characterize intake throughout pregnancy. In addition, three daily in-depth diet interviews (also called "24-hour recalls") were collected proximal to each FFQ, providing a maximum of 12 measurements over three time points. The replicate dietary records were collected in order to help quantify measurement errors in each FFQ.

Finally, we make two additional points regarding the PIN study data. First, one serum folate biomarker was collected on every woman in the study, that is, both in the main study and in the substudy. This feature of the PIN study is not common among dietary studies, where a "typical" study collects biomarkers only on women in the validation substudy. However, we found that the additional biomarker information compensated for a lack of information in the validation substudy (i.e., missing FFQs, 24-hour recalls, or biomarkers). Second, the PIN study collected serum and red-blood cell folate biomarkers, which we use as our reference instruments in our analyses. As pointed out by a referee, these biomarkers are measures of folate concentration and not folate intake. Better measures of the latter are replicate urinary nitrogen or doubly labeled water measurements, neither of which were collected in the PIN study. This important point does not change the validity of the methods or analyses but does have a significant impact on the interpretation of the analysis results and their generalizability to other studies.

3. MODEL AND NOTATION

In this section we describe the proposed model and inference used in many nutritional studies. The outcome is often modeled in two stages, where the first stage models the response as a function of predictors, both latent and observed, and the second stage specifies a measurement error model for the error-prone covariables. Let Y_i, $i = 1, \ldots, m$, be an outcome of interest belonging to the exponential family of distributions (McCullagh and Nelder 1983, p. 28). In the PIN study, Y_i will be the binary outcome preterm birth, where $Y_i = 1$ if a woman delivered preterm and 0 otherwise. Define $\mathbf{T}_i$ as a $p_T \times 1$ vector of error-prone covariates assumed to be related to the outcome of interest (e.g., $\mathbf{T}_i$ may refer to the true long-term dietary intake of several nutrients of interest, or it may refer to a vector of true dietary intakes for a single nutrient over different trimesters), and $\mathbf{Z}_i$ is a $p_Z \times 1$ vector of other covariates assumed to be "error free." The outcome is related to the covariates through the following model:

$$g\{\theta_i(\boldsymbol{\eta})\} = \eta_0 + \boldsymbol{\eta}_T'\mathbf{T}_i + \boldsymbol{\eta}_Z'\mathbf{Z}_i, \tag{1}$$

where $g(\cdot)$ is a known link function, $EY_i = \theta_i(\boldsymbol{\eta})$, and $\boldsymbol{\eta} = (\eta_0, \boldsymbol{\eta}_T', \boldsymbol{\eta}_Z')'$. The two primary instruments used in nutrition studies are the FFQ and 24-hour recall, which we denote by $Q_{ijl_1},\ l_1 = 1, \ldots, k_{ij}^Q$, and $F_{ijl_2},\ l_2 = 1, \ldots, k_{ij}^F$, respectively. In general, it will be convenient to let $\mathbf{Q}_i$ denote the $k_{i1} \times 1$ vector of all the FFQs for the ith subject, where $k_{i1} = \sum_j k_{ij}^Q$, and, similarly, let $\mathbf{F}_i$ be the $k_{i2} \times 1$ vector of the 24-hour recalls,

$k_{i2} = \sum_j k_{ij}^F$. As discussed previously, evidence suggests the following measurement error model (Kipnis et al. 2001, 2003; Spiegelman et al. 2004) relating the observed instruments to the true dietary intake, $\mathbf{T}_i$:

$$Q_{ijl_1} = \mu^Q + \alpha_j^Q + \beta^Q T_{ij} + b_{i1} + U_{ijl_1}^Q, \quad (2)$$

$$F_{ijl_2} = \mu^F + \alpha_j^F + \beta^F T_{ij} + b_{i2} + U_{ijl_2}^F, \quad (3)$$

where μ^Q, μ^F are means for the FFQ and 24-hour recalls, respectively, $(\alpha_1^Q, \ldots, \alpha_3^Q, \alpha_1^F, \ldots, \alpha_3^F)$ are trimester-level fixed effects, (b_{i1}, b_{i2}) are mean-zero random effects describing subject-specific biases, (β^Q, β^F) describe the systematic bias of the instruments, and $(U_{ijl_1}^Q, U_{ij^\dagger l_2}^F)$ is a bivariate perturbation vector assumed to have mean 0, variance (σ_Q^2, σ_F^2), respectively, and covariance $\rho_j \sigma_Q \sigma_F$ when $j = j^\dagger$ and 0 otherwise. To identify trimester-level effects and systematic bias in the FFQ and 24-hour recall, it is necessary to have one instrument that is unbiased for true dietary intake $\mathbf{T}_i$. Let the serum folate biomarker $M_{ijl_3},\ l_3 = 1, \ldots, k_{ij}^M$, which is obtained from a blood draw taken close in time to the FFQ administration, be such an instrument which is assumed to follow the model (Kipnis et al. 2001, 2003; Spiegelman et al. 2004)

$$M_{ijl_3} = T_{ij} + b_{i3} + U_{ijl_3}^M, \quad (4)$$

where, again, b_{i3} is a mean-zero random effect and $U_{ijl_3}^M$ is an independent, instrument-specific measurement error with variance σ_M^2. Again, recent research in nutritional epidemiology (Kipnis et al. 2001) suggests that it may be prudent to consider models where $\text{corr}(U_{ijl_2}^F, U_{ij^\dagger l_3}^M) \neq 0,\ j = j^\dagger$, and similarly for the FFQ. The resulting model is heavily parameterized, and the identifiability of all parameters will only be satisfied with sufficiently rich data, for example, replicate FFQs, 24-hour recalls, and biomarkers in a validation substudy. Because such data may not be observed in any one dataset, one must reduce the complexity of the measurement error model (2)–(4) through simplification or a priori knowledge of some parameters to identify the remaining unknown parameters. In the following paragraph, we discuss details of the PIN study data and its consequences on our measurement error model; we compare our model to one used in a recent analysis of the Medical Research Council (MRC) study data (Kipnis et al. 2001).

In the PIN study, women had at most one FFQ per trimester j ($k_{ij}^Q \leq 1$) and at most three 24-hour recalls ($k_{ij}^F \leq 3$) per trimester. Only one biomarker was collected throughout the study period. In contrast, the MRC study collected one FFQ throughout their study period, but collected eight biomarkers (two per season) and four 24-hour recalls (one per season). For our analysis of the PIN data, we set and model a single error-prone random variable T_i—that is, $T_{ij} = T_i,\ j = 1, 2, 3$, in (1)–(3)—and use the classical measurement error model for the biomarker in (4):

$$M_i = T_i + U_{i3}, \quad (5)$$

with σ_M^2 assumed to be known. Because replicate biomarkers are collected for every season of the MRC study, Kipnis et al. (2001) did not need to simplify the error model in the biomarker (4) as we have done for the PIN study. However, because only one FFQ is observed in the MRC study, identifiability for all the parameters in model (2) becomes problematic. For example, it is not possible to identify $\text{var}(b_{i1})$ and σ_Q^2 separately from one FFQ per subject without additional assumptions. Despite our model simplifications, we use general notation following models (2)–(4) as our methods and subsequent analyses are germane to other measurement error problems with similar data.

To write the likelihood for the observed data, it is convenient to introduce some new notation and assumptions. Let $\mathbf{W}_i = (\mathbf{Q}_i', \mathbf{F}_i', \mathbf{M}_i')'$ be the $k_i \times 1$ vector of all the instruments, where $\mathbf{M}_i$ is a $k_{i3} \times 1$ vector ($k_{i3} = \sum_j k_{ij}^M$) of unbiased reference instruments for the ith subject and $k_i = k_{i1} + k_{i2} + k_{i3}$. Here, we also assume that the random-effect vector $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{D}$ and the measurement error vector $\mathbf{U}_i = (\mathbf{U}_i^Q, \mathbf{U}_i^F, \mathbf{U}_i^M)'$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. The likelihood function of the observed data conditional on $\mathbf{Z}_i$ is $\prod_i L_i(Y_i, \mathbf{W}_i|\mathbf{Z}_i)$, where

$$L_i(Y_i, \mathbf{W}_i|\mathbf{Z}_i) = \int L_i(Y_i|\mathbf{T}_i, \mathbf{Z}_i) L_i(\mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i) L_i(\mathbf{T}_i|\mathbf{Z}_i)\, d\mathbf{T}_i, \quad (6)$$

$L_i(\mathbf{T}_i|\mathbf{Z}_i)$ is the likelihood of the true dietary intake vector $\mathbf{T}_i$ (e.g., Gaussian), $L_i(\mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i)$ is the error distribution conditional on $\mathbf{Z}_i$, and $L_i(Y_i|\mathbf{T}_i, \mathbf{Z}_i)$ is the probability density function from the exponential family with the systematic and random components and link function given in (1). We will assume that $\mathbf{U}_i$ is independent of $\mathbf{Z}_i$ and, therefore, replace $L_i(\mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i)$ with $L_i(\mathbf{W}_i|\mathbf{T}_i)$, which is a multivariate normal distribution defined by the models in (2), (3), and (4). This assumption seems tenable in many applications but would not be reasonable if, for example, the mother's height or weight were somehow related to the error in the instrument. If such an assumption were unjustified, a more complicated error model could be included without any additional difficulty. A detailed description of $L_i(\mathbf{W}_i|\mathbf{T}_i)$ is given in the next section.

3.1 Measurement Error Model

We first consider a simplified version of the model in (6), motivated by data from the PIN study described in Section 2. For simplicity, let $j = 1, \ldots, 3$ and $T_{ij} = T_i$ for all j. Conditional on the random effects $\mathbf{b}_i$ and true dietary folate intake T_i, we have

$$\begin{pmatrix} \mathbf{Q}_i \\ \mathbf{F}_i \\ \mathbf{M}_i \end{pmatrix} \sim \text{N}_{k_i} \left\{ \mathbf{X}_i \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\alpha} \end{pmatrix} + \mathbf{A}_i T_i \begin{pmatrix} \boldsymbol{\beta} \\ 1 \end{pmatrix} + \mathbf{R}_i \mathbf{b}_i, \mathbf{\Sigma}_i \right\},$$

where $\boldsymbol{\mu} = (\mu^Q, \mu^F)',\ \boldsymbol{\alpha} = (\alpha_1^Q, \ldots, \alpha_3^Q, \alpha_1^F, \ldots, \alpha_3^F)',\ \boldsymbol{\beta} = (\beta^Q, \beta^F)'$, and $\mathbf{X}_i$, $\mathbf{A}_i$, and $\mathbf{R}_i$ are *fixed* design matrices linking the instruments/biomarkers to the calibration parameters and random effects, respectively. To continue this illustration, we make another common assumption and subsequent simplification in the measurement error model. In particular, one typically assumes that the measurement errors in the biomarkers for the ith subject are independent of the measurement errors in the FFQs and 24-hour recalls. This assumption seems tenable in the PIN data as the FFQs and 24-hour recalls are both self-reported, whereas the biomarkers are laboratory measured with no a priori knowledge of FFQ or 24-hour recall. If we

partition $\mathbf{\Sigma}_i$ into $\mathbf{\Sigma}_{11i}$, $\mathbf{\Sigma}_{12i}$, $\mathbf{\Sigma}_{21i}$, and $\mathbf{\Sigma}_{22i}$ where $\mathbf{\Sigma}_{11i}$ corresponds to the covariance matrix for the FFQs and 24-hour recalls, $\mathbf{\Sigma}_{12i} = \mathbf{\Sigma}'_{21i}$ corresponds to the covariance between instruments and biomarkers, and $\mathbf{\Sigma}_{22i}$ is the covariance matrix of the biomarkers, then the conditional independence assumption implies $\mathbf{\Sigma}_{12i} = \mathbf{\Sigma}_{21i}^T = 0$. From here, it is useful to treat the biomarkers separately in the model as well as in the likelihood (6). Now, we focus on the error calibration model for the FFQ and 24-hour recall only. Hence, we rewrite this portion of the model as

$$\begin{pmatrix} \mathbf{Q}_i \\ \mathbf{F}_i \end{pmatrix} \sim \mathrm{N}(\mathbf{H}_i \boldsymbol{\gamma} + \mathbf{R}_i \mathbf{b}_i, \mathbf{\Sigma}_{11i}), \tag{7}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_Q, \boldsymbol{\gamma}'_F)'$, $\boldsymbol{\gamma}_r = (\mu^r, \alpha_1^r, \alpha_2^r, \alpha_3^r, \beta^r)'$ for $r = Q, F$, where Q is short-hand for FFQ and F denotes 24-hour recall. Because $\mathbf{H}_i$ is not full rank, it is necessary to constrain some of the parameters to achieve estimability of $\boldsymbol{\gamma}$. We constrain the first trimester-level effect $\alpha_1^r = 0, r = Q, F$, which implies $\boldsymbol{\gamma}_r = (\gamma_1^r, \gamma_2^r, \gamma_3^r, \gamma_T^r)'$ has the following interpretations: $\gamma_1^r = \mu^r + \alpha_1^r$, $\gamma_2^r = \alpha_2^r - \alpha_1^r$, and $\gamma_3^r = \alpha_3^r - \alpha_1^r$ for $r = Q, F$. For consistency, we label $\gamma_T^r = \beta^r$. In (7), we also have that $\mathbf{H}_i$ is block diagonal, that is, $\mathbf{H}_i = \mathrm{diag}\{\mathbf{H}_i^Q, \mathbf{H}_i^F\}$, where

$$\mathbf{H}_i^Q = \left(\mathbf{B}_i^Q \middle| \mathbf{1}_{k_{i1}} \middle| T_i \mathbf{1}_{k_{i1}}\right)$$

and $\mathbf{B}_i^Q = \mathrm{diag}\{\mathbf{1}_{k_{i1}^Q}, \mathbf{1}_{k_{i2}^Q}, \mathbf{1}_{k_{i3}^Q}\}$. $\mathbf{H}_i^F$ is defined similarly. With the $\mathbf{Q}_i$ and $\mathbf{F}_i$ organized as in (7), $\mathbf{\Sigma}_{11}$ (as a function of its parameters) may be written as

$$\mathbf{\Sigma}_{11i}(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \mathbf{G}_i^{1/2}(\boldsymbol{\sigma}) \mathbf{\Gamma}_i(\boldsymbol{\rho}) \mathbf{G}_i^{1/2}(\boldsymbol{\sigma}), \tag{8}$$

where $\boldsymbol{\sigma} = (\sigma_Q, \sigma_F)'$, $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3)'$ for the three trimester correlation parameters, $\mathbf{G}_i(\boldsymbol{\sigma}) = \mathrm{diag}\{\sigma_Q^2 \mathbf{I}_{k_{i1}}, \sigma_F^2 \mathbf{I}_{k_{i2}}\}$ and $\mathbf{\Gamma}_i(\boldsymbol{\rho})$ is a symmetric $k_i \times k_i$ correlation matrix. Assuming a single FFQ in each of three trimesters (i.e., $k_{i1}^Q = k_{i2}^Q = k_{i3}^Q = 1$) and three 24-hour recalls at each of three trimesters (i.e., $k_{i1}^Q = k_{i2}^Q = k_{i3}^Q = 3$), $\mathbf{\Gamma}_i(\boldsymbol{\rho})$ is a correlation matrix with the following structure:

$$\mathbf{\Gamma}_i(\boldsymbol{\rho}) = \left(\begin{array}{ccc|ccc} & \mathbf{I}_3 & & \rho_1 \mathbf{1}'_3 & \mathbf{0} & \mathbf{0} \\ & & & \mathbf{0} & \rho_2 \mathbf{1}'_3 & \mathbf{0} \\ & & & \mathbf{0} & \mathbf{0} & \alpha_3 \mathbf{1}'_3 \\ \hline \rho_1 \mathbf{1}_3 & \mathbf{0} & \mathbf{0} & & & \\ \mathbf{0} & \rho_2 \mathbf{1}_3 & \mathbf{0} & & \mathbf{I}_9 & \\ \mathbf{0} & \mathbf{0} & \rho_3 \mathbf{1}_3 & & & \end{array}\right), \tag{9}$$

where the $\mathbf{1}_r$ is a column vector of 1's of length r and the $\mathbf{0}$'s are vectors with the appropriate implied dimensions. Note that if we have replicate FFQs and 24-hour recalls greater than or equal to 2 at each time, then $\mathbf{1}$ (and analogously the $\mathbf{0}$'s) will no longer refer to vectors, but matrices of 1's (or 0's). So far, we have placed no restrictions on $\boldsymbol{\rho}$. We discuss three correlation models of interest and subsequent restrictions on $\mathbf{\Sigma}_{11i}(\boldsymbol{\sigma}, \boldsymbol{\rho})$ in Section 3.2.

3.2 Correlation Models and Their Implied Constraints

In this section we focus on three correlation models of interest and derive the conditions on $\boldsymbol{\rho}$ that lead to the positive definiteness of $\Sigma_{11i}(\boldsymbol{\sigma}, \boldsymbol{\rho})$ and, therefore, ultimately lead to model identifiability.

The three correlation models (CMs) of interest can be summarized as follows: For every l_1, l_2,

$$\text{CM1:} \quad \mathrm{corr}\left(U_{ijl_1}^Q, U_{ij^\dagger l_2}^F\right) = 0 \quad \text{for every } j, j^\dagger, \tag{10}$$

$$\text{CM2:} \quad \mathrm{corr}\left(U_{ijl_1}^Q, U_{ij^\dagger l_2}^F\right) = \begin{cases} \rho & \text{if } j = j^\dagger \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

$$\text{CM3:} \quad \mathrm{corr}\left(U_{ijl_1}^Q, U_{ij^\dagger l_2}^F\right) = \begin{cases} \rho_j & \text{if } j = j^\dagger \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

for subject i at time j. In words, correlation model 1 (CM1) in (10) assumes that measurement errors between FFQs and 24-hour recalls are mutually independent, whereas CM2 and CM3 assume correlated errors. CM3 assumes measurement errors for different instruments are correlated differently for each measurement time, whereas CM2 assumes the correlation remains the same over time. Both CM2 and CM3 are expected to reflect better the errors in contemporaneous instruments observed in nutritional epidemiological studies (Kipnis et al. 2001; Subar et al. 2003; Carroll 2003; Carroll, Ruppert, Crainiceanu, Tosteson, and Karagas 2004).

Now, we turn our attention to the positive definiteness of $\mathbf{\Sigma}_{11i}$. By definition, $\mathbf{\Sigma}_{11i}$ will be positive definite when the quadratic form $\boldsymbol{\lambda}' \mathbf{\Sigma}_{11i} \boldsymbol{\lambda} = 0$ if and only if $\boldsymbol{\lambda} = 0$. We use a corollary that allows us to check the positivity of the determinants of all the leading minors or, analogously, to check that the eigenvalues are all positive (Searle 1971).

Assuming that there are J measurement times and a constant number of replicate FFQs and 24-hour recalls across trimesters, n_Q and n_F, respectively, the general form of the determinant of $\mathbf{\Sigma}_{11i}$ is

$$|\mathbf{\Sigma}_{11i}| = \sigma_F^{2Jn_F} \sigma_Q^{2Jn_Q} \prod_{j=1}^{J} (1 - n_Q n_F \rho_j^2), \tag{13}$$

and the *unique* eigenvalues of $\mathbf{\Sigma}_{11i}$ are σ_Q^2, σ_F^2, and

$$\frac{1}{2}\left\{\sigma_Q^2 + \sigma_F^2 \pm (\sigma_Q^4 + \sigma_F^4 - 2\sigma_Q^2 \sigma_F^2 + 4 n_F n_Q \rho_j^2 \sigma_Q^2 \sigma_F^2)^{1/2}\right\}$$

for $j = 1, \ldots, J$. It is straightforward to verify that the product of the eigenvalues is indeed the determinant by including the missing replicate eigenvalues, that is, $J - 1$ repeats of σ_Q^2 and σ_F^2. Now, through some straightforward algebra, it is easy to see that the condition that will ensure the positivity of the eigenvalues is

$$|\rho_j| < (n_F n_Q)^{-1/2}, \qquad j = 1, \ldots, J. \tag{14}$$

The condition in (14) is necessary and sufficient for the positive definiteness of $\mathbf{\Sigma}_{11i}$. Furthermore, any prior distribution placed on $\boldsymbol{\rho}$ must have support (14). Note that neither models (11) nor (12) will be able to detect/estimate correlation parameters that are extreme in either direction.

4. PRIOR AND POSTERIOR DISTRIBUTIONS

In this section we discuss the prior specification for all the parameters in the preceding models and the resulting posterior distributions to be used in a Gibbs (Geman and Geman 1984) or Metropolis–Hastings (Metropolis and Ulam 1949; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953; Hastings 1970) sampling algorithm. For now, assume that $\mathbf{T}_1, \ldots, \mathbf{T}_m$ are

independent and identically distributed random vectors from the distribution F_T with mean $\boldsymbol{\mu}_T$ and variance $\boldsymbol{\Sigma}_T$. Define $L_i(Y_i|\mathbf{T}_i, \mathbf{Z}_i; \boldsymbol{\eta})$ in (6) as the ith contribution to the conditional likelihood given $\mathbf{T}_i$ arising from (1); for example, for a Bernoulli response and a logit link function

$$\log L_i(Y_i|\mathbf{T}_i, \mathbf{Z}_i; \boldsymbol{\eta}) = Y_i(\eta_0 + \boldsymbol{\eta}_T'\mathbf{T}_i + \boldsymbol{\eta}_Z'\mathbf{Z}_i) + \log\{1 - \theta_i(\boldsymbol{\eta})\},$$

where $\theta_i(\boldsymbol{\eta})$ was defined in (1). For simplicity, we assume normal prior distributions on the mean parameters $\boldsymbol{\eta}$, the systematic bias parameters $\boldsymbol{\gamma}$ in (7), and mean of the latent dietary intake random variables $\boldsymbol{\mu}_T$ from $L_i(\mathbf{T}_i|\mathbf{Z}_i)$ in (6), that is, $\boldsymbol{\eta} \sim \mathrm{N}(\boldsymbol{\eta}_0, \mathbf{V}_{0,\eta})$ in (1), $\boldsymbol{\gamma} \sim \mathrm{N}(\boldsymbol{\gamma}_0, \mathbf{V}_{0,\gamma})$ in (7), and $\boldsymbol{\mu}_T \sim \mathrm{N}(\boldsymbol{\mu}_{T,0}, \mathbf{V}_{0,\mu_T})$ in $L_i(\mathbf{T}_i|\mathbf{Z}_i)$, and conjugate Wishart priors on $\mathbf{D}^{-1}$ in (7) and $\boldsymbol{\Sigma}_T^{-1}$ in $L_i(\mathbf{T}_i|\mathbf{Z}_i)$ in (6), $\mathbf{D}^{-1} \sim W_q(\nu_D, C_D)$ and $\boldsymbol{\Sigma}_T^{-1} \sim W(\nu_{\boldsymbol{\Sigma}_T}, C_{\boldsymbol{\Sigma}_T})$, respectively. Note that although it is common to assume inverse Gamma priors for $\boldsymbol{\sigma}$, this will not necessarily imply a conjugate prior distribution because of the correlation parameters $\boldsymbol{\rho}$ in $\boldsymbol{\Sigma}_{11i}$. Because our constraints on the correlation parameters do not depend on $\boldsymbol{\sigma}$, we may factor our joint prior $\pi(\boldsymbol{\sigma}, \boldsymbol{\rho})$ into the product $\pi(\boldsymbol{\sigma})\pi(\boldsymbol{\rho})$. Because there are typically more replicate FFQs and 24-hour recalls than biomarkers, we assume flat priors for σ_Q^2 and σ_F^2 but an inverse Gamma (IG) prior for σ_M^2. We define our prior on $\boldsymbol{\sigma}$ as

$$\pi(\boldsymbol{\sigma}) = \sigma_Q^{-2}\sigma_F^{-2}e^{-1/(b_M\sigma_M)} \big/ \sigma_M^{a_M+1},$$

where a_M, b_M are specified hyperparameters. For $\boldsymbol{\rho}$, we specify a uniform prior with support given by the parameter constraints given in Section 3.2, that is, $\pi(\boldsymbol{\rho}) \propto 1$ with $|\rho_j| < (n_F n_Q)^{-1/2}$, $j = 1, \ldots, J$. Finally, we also assume $\mathbf{T}_i$ is normally distributed with mean $\boldsymbol{\mu}_T$ and covariance matrix $\boldsymbol{\Sigma}_T$. Given $\pi(\boldsymbol{\sigma})$ and $\pi(\boldsymbol{\rho})$, and prior variances $\mathbf{V}_{0,\eta}$, $\mathbf{V}_{0,\gamma}$, and $\mathbf{V}_{0,\mu_T}$, the joint posterior of the parameters is given by

$$\begin{aligned}
&p(\boldsymbol{\eta}, \boldsymbol{\gamma}, \mathbf{b}, \mathbf{T}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \mathbf{D}, \boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T|\mathbf{Y}, \mathbf{W}) \\
&\quad \propto |\boldsymbol{\Sigma}_W|^{-1/2}|\mathbf{D}^{-1}|^{(\nu_D+m-q-1)/2}|\boldsymbol{\Sigma}_T^{-1}|^{(\nu_{\boldsymbol{\Sigma}_T}+m-p_T-1)/2} \\
&\qquad \times \exp\Bigg[\sum_{i=1}^{m}\Bigg\{\log L_i(\boldsymbol{\eta}) - \frac{1}{2}(\mathbf{W}_i - \mathbf{H}\boldsymbol{\gamma})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{W}_i - \mathbf{H}\boldsymbol{\gamma}) \\
&\qquad - \frac{1}{2}\mathbf{b}_i'\mathbf{D}^{-1}\mathbf{b}_i - \frac{1}{2}(\mathbf{T}_i - \boldsymbol{\mu}_T)'\boldsymbol{\Sigma}_T^{-1}(\mathbf{T}_i - \boldsymbol{\mu}_T)\Bigg\} \\
&\qquad - \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'\mathbf{V}_{0,\eta}^{-1}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) - \frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)'\mathbf{V}_{0,\gamma}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) \\
&\qquad - \frac{1}{2}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_{T,0})'\mathbf{V}_{0,\mu_T}^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_{T,0}) \\
&\qquad - \frac{1}{2}\mathrm{tr}(\mathbf{C}_D^{-1}\mathbf{D}^{-1}) - \frac{1}{2}\mathrm{tr}\big(\mathbf{C}_{\boldsymbol{\Sigma}_T}^{-1}\boldsymbol{\Sigma}_T^{-1}\big)\Bigg]\pi(\boldsymbol{\sigma})\pi(\boldsymbol{\rho}), \qquad (15)
\end{aligned}$$

where $\boldsymbol{\Sigma}_W = \mathrm{diag}\{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m\}$. Additional details for the full conditional distributions are given in the Appendix.

4.1 Relaxing Distributional Assumptions on $\mathbf{T}_i$

In measurement error problems, $\mathbf{T}_i$ is a latent random vector with distribution F_T. The Bayesian paradigm offers a convenient method for handling latent variables and other incomplete data problems by sampling the latent variable from its full conditional distribution. When F_T is parametric (e.g., Gaussian), the full posterior is given by (15). However, this distributional assumption is difficult to check and a more flexible model is often desirable. One method is to use a scale mixture of normals for $\mathbf{T}_i$. Toward this goal, suppose that we start with a univariate Gaussian distribution with mean μ_T and variance σ_T^2. Then, we may write

$$T_i = \mu_T + \epsilon_i,$$

where $\epsilon_i \sim \mathrm{N}(0, \sigma_T^2)$. A straightforward extension of this model is to assume $\epsilon_i \sim \mathrm{N}(0, \lambda_i\sigma_T^2)$ where the λ_i are subject-specific latent variables and assumed to have Gamma distributions. A second method makes even fewer assumptions about the distribution function F_T, requiring only that F_T be a proper distribution function. We employ the mixture of Dirichlet process (MDP) methodology based on a Polyá urn scheme (Antoniak 1974; Escobar 1994; MacEachern 1994). In addition to using the Dirichlet process prior for parameters, the MDP methodology has been successfully applied to other missing-data problems, such as random effects in mixed models (Kleinman and Ibrahim 1998; Brown and Ibrahim 2003). Less work has been done using the MDP prior in measurement error models. Two exceptions are Mallick, Hoffman, and Carroll (2002) and Müller and Roeder (1997), the latter of which describes an application of the MDP prior methodology to case-control studies. There are at least two differences worth noting between our application here and the one presented in Müller and Roeder (1997). First, there is the fundamental difference in design between the retrospective and prospective study design, where the case-control design has the additional complexity derived from conditioning on the prevalence of cases in the sample, that is, conditioning on $\sum_i Y_i = 1$ (cf. Breslow and Day 1980). Second, our two applications are different in that our model incorporates multiple validation instruments with correlated errors. We expect that in a case-control study with multivariate instruments, as in our application presented here, a combined model using ideas presented here and in Müller and Roeder (1997) could be applied. In the following discussion, we describe how to apply the mixture of Dirichlet process methodology to our measurement error problem.

Assume the random vectors $\mathbf{T}_i$ are drawn from an arbitrary distribution F_T, where F_T has a Dirichlet process prior, denoted by $F_T \sim \mathrm{DP}(\xi F_0)$, $F_0 \sim \mathrm{N}(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T)$, and ξ is an *unknown* scalar confidence parameter. Suppressing parameters other than the error-prone covariate $\mathbf{T}_i$, the full conditional distributions for $\{\mathbf{T}_i,\ i = 1, \ldots, m\}$ are given by (see Kleinman and Ibrahim 1998)

$$\begin{aligned}
&[\mathbf{T}_i|\{\mathbf{T}_i, k \neq i\}, Y_i, \mathbf{W}_i] \\
&\qquad \sim q_0 L_i(Y_i, \mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i) f_0(\mathbf{T}_i|\mathbf{Z}_i) + \sum_{k\neq i}\delta(d\mathbf{T}_i|\mathbf{T}_k), \qquad (16)
\end{aligned}$$

where $f_0(\mathbf{T}_i|\mathbf{Z}_i) \equiv L_i(\mathbf{T}_i|\mathbf{Z}_i)$, and $L_i(Y_i, \mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i)$ was defined in (6). Recall that $L_i(Y_i, \mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i)$ factors into the product $L_i(Y_i|\mathbf{T}_i, \mathbf{Z}_i)L_i(\mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i)$ by the nondifferential measurement error assumption. Also, $\{q_0, q_k,\ k = 1, \ldots, m\}$ are unnormalized selection probabilities where

$$q_0 \propto \xi \int \cdots \int L_i(Y_i, \mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i) f_0(\mathbf{T}_i|\mathbf{Z}_i) \qquad (17)$$

and $q_k \propto L_i(Y_i, \mathbf{W}_i|\mathbf{T}_k^*, \mathbf{Z}_i)$ and $\mathbf{T}_k^*$ are the unique atoms of $f_0(\mathbf{T}|\mathbf{Z})$. Because (17) does not, in general, have a closed-form solution, numerical integration is typically needed. However, it would be possible to find a closed-form solution if, for example, $L_i(Y_i, \mathbf{W}_i|\mathbf{T}_i, \mathbf{Z}_i)$ and F_0 were both multivariate normal. At the next stage, we sample the unique vector $\mathbf{T}_j^*$ from its full conditional distribution $p(\mathbf{T}_j^*|\mathbf{D}_{\text{obs}}, \mathit{rest})$, where $\mathbf{D}_{\text{obs}}$ denotes the observed data, and *rest* is short hand for all remaining parameters. For a fixed confidence parameter ξ, the full conditional distribution of $\mathbf{T}_j^*$ is defined as

$$p(\mathbf{T}_j^*|\mathbf{D}_{\text{obs}},\ \mathit{rest}) \propto \exp\Bigg[\sum_{i\in\mathcal{S}_j}\Big\{Y_i g(\theta_i) + \log(1-\theta_i) - \frac{1}{2}(\mathbf{W}_i - \mathbf{H}_i\boldsymbol{\gamma} - \mathbf{R}_i\mathbf{b}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{W}_i - \mathbf{H}_i\boldsymbol{\gamma} - \mathbf{R}_i\mathbf{b}_i)\Big\} - \frac{1}{2}(\mathbf{T}_j^* - \boldsymbol{\mu}_T)'\boldsymbol{\Sigma}_T^{-1}(\mathbf{T}_j^* - \boldsymbol{\mu}_T)\Bigg] \tag{18}$$

for $\mathcal{S}_j = \{i|\mathbf{T}_i = \mathbf{T}_j^*\}$.

Define I^* as the number of unique clusters of $\mathbf{T}^*$, $I^* \le m$. Then, the confidence parameter ξ influences the tendency of the Markov chain Monte Carlo (MCMC) algorithm to favor large or small I^*, with $\xi \to \infty$ implying large I^*. In this article, we use initially a two-stage data augmentation algorithm to sample ξ (Tanner and Wong 1987) and then conduct sensitivity studies where ξ is fixed. Assume ξ has a Gamma prior with shape r and rate λ, that is, $\xi \sim \text{Gamma}(r, \lambda)$ with $E\xi = r/\lambda$. At the first stage, the augmentation algorithm samples a latent variable c conditional on the current value of ξ and I^*, that is, $[c|\xi, I^*] \sim \text{Beta}(\xi + 1, I^*)$. Next, we sample the confidence parameter ξ from the mixture of two Gamma distributions given the latent variable c and I^*, that is,

$$[\xi|c, I^*] \sim \pi_c\text{Gamma}\big(r + I^*, \lambda - \log(c)\big) + (1-\pi_c)\text{Gamma}\big(r + I^* - 1, \lambda - \log(c)\big),$$

where $\pi_c = z/(z+1)$ and $z = (r + I^* - 1)/[I^*\{\lambda - \log(c)\}]$. Some care is needed in choosing the prior parameters (r, λ) as this strongly influences the tendency of the algorithm to favor the base measure F_0 or collapse on relatively few clusters. We use two priors, Gamma(1, 1) and Gamma(.01, .01), to check the sensitivity of parameter estimates due to the choice of prior on ξ. Both priors have mean 1, but the latter prior has variance 100 and, therefore, puts mass on both large and small values of ξ.

5. SIMULATION STUDIES

Here, we present a small simulation study to provide some empirical validity that the parameters in the complex measurement error model (2)–(3) are estimable. The structure of our simulation study mimics the PIN study data, and, hence, we use the simplified biomarker model (5) with σ_M^2 known. The details of our simulation study follow.

We begin by simulating T_i as iid standard Gaussian random variates, $i = 1, \ldots, 75$, and independently generating subject-specific biases $\mathbf{b}_i$ from a bivariate Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{D}$. Then, for each subject i and visit $j = 1, 2, 3$, we generate a vector of instruments that satisfies the models:

$$Q_{ij} = \gamma_1^Q + \gamma_2^Q + \gamma_3^Q + \gamma_T^Q T_i + b_i^Q + U_{ij}^Q,$$
$$F_{ijl} = \gamma_1^F + \gamma_2^F + \gamma_3^F + \gamma_T^F T_i + b_i^F + U_{ijl}^F, \qquad l = 1, 2, 3,$$

where $\text{corr}(U_{ij}^Q, U_{ijl}^F) = \rho_j$, $l = 1, 2, 3$, and $\rho_2 = .25$ but $\rho_1 = \rho_3 = 0$. Finally, we independently simulate one unbiased biomarker M_i as Gaussian with mean T_i and variance $\sigma_M^2 = .3$. The specific values for the remaining parameters are given in Table 1.

Table 1. Summary of Posterior Means and Credible Sets Over 500 Monte Carlo Datasets

Parameter	Truth	Mean	SD	Coverage
γ_1^Q	3.00	2.99	.18	.93
γ_2^Q	−.75	−.75	.17	.95
γ_3^Q	.75	.75	.17	.93
γ_T^Q	.50	.52	.19	.95
γ_1^F	1.00	1.01	.21	.93
γ_2^F	−.25	−.25	.09	.95
γ_3^F	.25	.24	.10	.94
γ_T^F	.90	.94	.24	.91
D_{11}	1.25	1.26	.29	.95
D_{12}	.25	.24	.27	.94
D_{22}	2.25	2.31	.48	.95
σ_Q^2	1.00	1.04	.12	.95
σ_F^2	1.00	1.01	.06	.94
ρ_1	.00	.00	.07	.97
ρ_2	.25	.24	.08	.94
ρ_3	.00	.00	.06	.95

NOTE: Mean represents the Monte Carlo average posterior mean, SD represents the Monte Carlo standard deviation of posterior means, and Coverage indicates the proportion of datasets in which a 95% credible set includes the true value. γ are systematic bias parameters in the measurement error model (MEM) and the remaining parameters are covariance parameters in the MEM. For each dataset, we drew 2,000 samples from our joint posterior and treated the first 1,000 as burn-in.

In conclusion, we have not proven formally that the parameters in our measurement error model (2)–(3) are identified. At the same time, our simulation studies suggest that one can estimate all parameters in our measurement error model and draw correct inference from the posterior distribution using the correct likelihood specification and noninformative priors distributions.

6. ANALYSIS OF THE PIN STUDY DATA

For purposes of discussion, we split the data into two groups: women who were included in a substudy and women not in the substudy. In addition to the single FFQ, main study participants also provided serum folate measures, which were incorporated into the measurement error model in the analysis. Women in the substudy provided additional dietary information that other women were not requested to give, ideally providing three FFQs and nine 24-hour recalls (1 FFQ and three 24-hour recalls per trimester for all three trimesters) during the pregnancy. For convenience, we split the ith contribution to the likelihood (6) into two pieces through the use of indicator functions, $I(\cdot)$. Suppressing the parameters arising from $\mathbf{T}_i$, we have

$$L_i(Y_i, \mathbf{W}_i) = \{L_{i,\text{sub}}(\mathbf{W}_i; \boldsymbol{\gamma}, \mathbf{D}, \boldsymbol{\sigma}, \boldsymbol{\rho})\}^{I(S_i=1)} \{L_{i,n\text{sub}}(Y_i, \mathbf{W}_i; \boldsymbol{\eta}, \gamma_1^Q, \sigma_F)\}^{I(S_i=0)},$$

where S_i equals 1 if the ith women belongs to the substudy and 0 otherwise. Therefore, the posteriors for $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}$ will have different contributions from women in the substudy versus those not in the substudy. Of course, the posterior for $\mathbf{T}_i$ depends on substudy status as well as each step in the MDP implementation.

Our analysis uses 172 women from the substudy who had at least one of the nine 24-hour recalls and 1,679 women in the main study. Due to the rigorous protocol of the substudy, women did not provide all 12 dietary measures. The 1,679 women in the main study were chosen to have complete data for preterm birth, the three "error-free" covariables in the outcome model—height, body mass index (BMI), and dietary caloric intake (also called "energy" in our analyses below) as measured in the FFQ—and serum folate. The overall preterm birth rate in the combined data was 12.7% (236/1,851). Two covariables, BMI and dietary caloric intake, were transformed using the natural logarithm. All three covariables were standardized by their sample means and standard deviations (2.6, .24, .47, respectively) and all are assumed to be error free. With additional information on the variability in the measurements in these variables, it would be possible to relax this assumption as well. This investigation is, however, beyond the scope of this article and beyond the data available to the authors. The sample variance of the unbiased serum folate biomarker is .40.

Although nonsubstudy women were chosen to have complete data, the same criterion was not used to select women in the substudy because of frequent nonresponse. As shown in Table 2, although many women provided one 24-hour recall at each trimester (82%, 73%, and 67% at visits 1, 2, and 3, respectively), fewer provided all three 24-hour recalls for any given trimester because of the rigorous protocol. Rather than remove these missing observations, we assumed the missing values were missing at random, then used our model and MCMC methods to sample the missing values (cf. Little and Rubin 2002 for a review of Gibbs sampling for missing-data problems). A similar strategy was employed for missing biomarkers (only 25 biomarkers were observed from the 172 substudy women).

We summarize the mean parameters from the outcome model ($\boldsymbol{\eta}$) and the systematic bias parameters ($\boldsymbol{\gamma}$) in Table 3 and variance parameters (σ_Q^2, σ_F^2, $\mathbf{D}$, ρ_j) in Table 4. In Table 3 we include one column of "naive" parameter estimates, which are calculated by fitting two independent regression models with *complete* data: first, the logistic regression model in (1) with the true folate intake replaced by the serum folate biomarker to obtain $\hat{\boldsymbol{\eta}}_{\text{naive}}$, and second, the linear regression of *substudy* FFQs and 24-hour recalls on serum folate biomarkers assuming model (2)–(3) under CM1 ($\rho_j = 0$) and no subject-specific biases ($\mathbf{D} = 0$). In Table 3 we summarize the parameter estimates under CM1 for folate intake following a normal distribution and also our MDP model with $\xi = .83$, reflecting little confidence in the normality assumption. Interestingly, the protective folate effect from the naive analysis appears even stronger after adjusting for measurement error. Also, there appears to be an inverse intra-individual relationship between the FFQ and 24-hour recall ($D_{12} = -.22$), which suggests that women who respond conservatively on the FFQ tend to respond liberally on the 24-hour recall and vice versa. In the validation study, there is some evidence that folate consumption, as measured by the 24-hour recalls, increases throughout pregnancy, though there appears to be no monotone increase when evaluating folate consumption as measured by the FFQ. Though the cost may be prohibitive, future validation studies in pregnancy might consider including serial biomarkers to help determine whether there are substantial pregnancy-related dietary changes throughout the 9-month

Table 2. Sample Mean and Standard Deviation for Two Biased Measures of Folate Intake—Dietary Folate (FFQ) and 24-Hour Recall—From 172 Women in the PIN Substudy

Instrument	Trimester	Rep	N	Mean	SD
FFQ	1	1	97	5.92	.46
	2	1	134	6.00	.35
	3	1	72	6.00	.42
24-hour recall	1	1	141	5.62	.62
	1	2	104	5.61	.68
	1	3	16	5.18	.43
	2	1	125	5.72	.55
	2	2	95	5.84	.48
	2	3	5	5.55	.58
	3	1	116	5.87	.56
	3	2	87	5.91	.51
	3	3	2	6.14	.22

NOTE: Both FFQ and 24-hour recall measurements are reported on the log-scale with the 24-hour recall attempted three times per trimester and FFQ attempted once per trimester.

Table 3. Analysis Results From the PIN Study Making Parametric Assumptions About True Folate and a Common Correlation Parameter Among Contemporaneous Instruments

Parameter	Naive	Normal	MDP ($\xi = .83$)
Intercept (η_0)	$-1.95_{(.08)}$	$-1.99_{(.08)}$	$-1.99_{(.08)}$
Folate (η_T)	$-.27_{(.11)}$	$-.46_{(.15)}$	$-.48_{(.14)}$
Height (η_{Z_1})	$-.07_{(.03)}$	$-.08_{(.03)}$	$-.08_{(.03)}$
BMI (η_{Z_2})	$.68_{(.30)}$	$.65_{(.31)}$	$.63_{(.31)}$
Energy (η_{Z_3})	$-.01_{(.15)}$	$-.01_{(.16)}$	$-.01_{(.15)}$
γ_1^Q	$5.90_{(.16)}$	$5.97_{(.02)}$	$5.97_{(.02)}$
γ_2^Q	$.18_{(.19)}$	$.11_{(.07)}$	$.10_{(.07)}$
γ_3^Q	$.20_{(.23)}$	$-.34_{(.08)}$	$-.35_{(.08)}$
γ_T^Q	$-.13_{(.13)}$	$.13_{(.03)}$	$.13_{(.03)}$
γ_1^F	$5.27_{(.09)}$	$5.31_{(.07)}$	$5.34_{(.06)}$
γ_2^F	$.25_{(.12)}$	$.31_{(.09)}$	$.31_{(.09)}$
γ_3^F	$.59_{(.13)}$	$.56_{(.09)}$	$.55_{(.09)}$
γ_T^F	$-.01_{(.08)}$	$-.02_{(.05)}$	$-.02_{(.04)}$

NOTE: The "naive" analysis refers to two independent, complete-case analyses that replace the true folate random variable with the serum folate biomarker. $\boldsymbol{\gamma}$ refers to systematic parameters in the measurement error model. Posterior means from 6,000 Gibbs samples with the first 4,500 treated as burn-in are reported with standard deviations reported in parentheses.

Table 4. Summary of Variance Component Estimates (with posterior standard deviations in parentheses) From MCMC Analyses Results From the PIN Study Using Normal Prior Distribution on True Folate Concentration

Parameter	CM1*	CM2	CM3
D_{11}	$.54_{(.09)}$	$.51_{(.09)}$	$.46_{(.08)}$
D_{12}	$-.21_{(.03)}$	$-.23_{(.04)}$	$-.24_{(.05)}$
D_{22}	$.08_{(.01)}$	$.10_{(.02)}$	$.13_{(.03)}$
σ_Q^2	$.43_{(.02)}$	$.43_{(.02)}$	$.43_{(.02)}$
σ_F^2	$1.19_{(.05)}$	$1.19_{(.05)}$	$1.18_{(.05)}$
ρ_1	0^*	$-.02_{(.02)}$	$-.11_{(.04)}$
ρ_2	0^*	$-.02_{(.02)}$	$.03_{(.05)}$
ρ_3	0^*	$-.02_{(.02)}$	$.02_{(.04)}$

NOTE: Correlation models (CM1–CM3) refer to different assumptions about the correlation among errors of contemporaneous instruments and are described in Section 3.

*Model 1 sets $\rho_j = 0$, $j = 1, 2, 3$.

Table 5. Model Comparison Using Deviance Information Criterion

Model	$\bar{\Delta}$	p_δ	Δ^*	*DIC*
CM1 + {**D** = 0}	8,841.5	2,893.3	5,948.2	11,734.8
CM1	8,836.9	3,053.2	5,783.7	11,890.1
CM2	8,834.0	3,052.2	5,781.9	11,886.3
CM3	8,836.6	3,055.1	5,781.5	11,891.7

{**D** = 0} implies $D_{11} = D_{12} = D_{22} = 0$, which implies no subject-specific biases (no heterogeneity) in the FFQ or 24-hour recall. $\bar{\Delta}$ is the posterior mean of minus twice the log-likelihood, p_δ is the effective number of parameters, Δ^* is minus twice the log-likelihood evaluated at the posterior means of the parameters, and DIC= $\bar{\Delta} + p_\delta$.

period that would necessitate serial dietary assessments in studies of nutrition during pregnancy. In addition to the parameters in Table 3, we also estimated the odds ratio for an "IQR increase" in folate, that is, an increase from the 25th percentile to the 75th percentile of the folate sample distribution. Hence, we estimated a 27% reduction in the odds [odds ratio (OR) = .73, (.59–.91)] of preterm birth for an IQR increase in latent folate given BMI, mother's height, and energy level. Mother's height and BMI are important predictors of preterm birth both before and after adjusting for measurement error in the folate variable.

The proposed measurement error model (2)–(4) is parameterized richly, and our analyses did not find substantial differences among parameter estimates in models of increasing complexity. Hence, it may be preferable to select the most parsimonious model and eliminate unnecessary complexity in the measurement error model. To facilitate model comparisons, we use the deviance information criterion (DIC; Spiegelhalter, Best, Carlin and van der Linde 2002) and compare the correlation models (CM1–CM3) in addition to one simpler model "CM1 + {**D** = 0}," which allows for no subject-specific biases in the FFQ or 24-hour recall. Our results are displayed in Table 5 using the following additional notation: $\bar{\Delta}$ is the posterior mean of minus twice the log-likelihood, p_δ is the effective number of parameters, Δ^* is minus twice the log likelihood evaluated at the posterior means of all parameters, and DIC $= \bar{\Delta} + p_\delta$. We immediately notice that p_δ is strikingly large, again emphasizing the large number of unknown variables in our model. Recall, that each latent folate variable T_i is regarded as an unknown variable in addition to all missing FFQs, 24-hour recalls, and biomarkers in the validation substudy. Our model comparison suggests that a model with no correlation among contemporaneous measurements and no subject-specific biases is the best model. The effective number of parameters in this simple model is approximately 160 parameters fewer than model CM1 due to the latent subject-specific biases b_{1i}, b_{2i}, which are absent when $\mathbf{D} = 0$. However, when we believe that $\mathbf{D} \neq 0$ and only focus on CM1–CM3, we find that CM2 is the best model among the three, which suggests that a model that considers nonzero correlations among contemporaneous instruments is useful.

In Tables 3 and 4 we presented parameter estimates that we claim are relatively insensitive to the confidence parameter ξ. To investigate further this claim, we ran more than 60 MDP analyses of the PIN study data with confidence parameters ranging from .01 to 10,000. We found that posterior means and standard deviations from an MDP analysis using confidence parameters greater than 50 did not change significantly. In Figure 1 we plot the number of unique clusters of T, that is, I^*, and the posterior means of five folate-related parameters as a function of the confidence parameter ξ and then fit a cubic spline to the points to illustrate the average trend. So, our empirical findings suggest that the parameter estimates do not change significantly once the number of unique clusters of T gets beyond 120 or so, on average, as we see in Figure 1(a). In Figures 1(b)–1(f), we graph the posterior means of five parameters most significantly impacted by choosing ξ sufficiently small. We note that all five parameters tend to decrease as ξ approaches 0. For example, the posterior mean of η_T is approximately $-.48$ when ξ small but $-.46$ for large ξ, the latter of which corresponds to the normality assumption in Table 3. At the same time, we emphasize a word of caution when drawing conclusions from these figures as the variability in posterior means cannot be ignored, particularly for small ξ. Moreover, the average change in posterior means from $\xi \approx .05$ to $\xi \approx 50$ may be extremely small, for example, less than .01 for γ_T^Q and less than .005 for γ_T^F.

7. DISCUSSION

We have presented a Bayesian semiparametric method to estimate parameters from a generalized linear measurement error model with a structured measurement error model and applied the method to an analysis of the PIN data. Our first method assumes that true long-term folate is normally distributed, whereas the second method using the mixture of Dirichlet process priors framework does not. We found that results based on a naive model that replaces true long-term folate by the observed serum folate to be somewhat conservative when compared to results based on our calibrated analysis. Furthermore, the results presented in Tables 3 and 4 appeared to be insensitive to the normality assumption on folate intake when compared to those from the MDP analysis for modest values of ξ.

In general, there has been mixed evidence in the literature about whether the instruments under- or overestimate intake. In the past, FFQs have been shown both to underestimate intake with respect to food records (Brown et al. 1996) and to overestimate intake relative to food records (Suitor, Gardner, and Willet 1989; Greeley, Storbakken, and Magel 1992; Forsythe and Gage 1987; Robinson, Godfrey, Osmond, Cox, and Barker 1996; Erkkola et al. 2001). The PIN raw data show some evidence of underestimation of dietary intake in FFQ versus 24-hour recall in the second and third trimesters, but this was not significant using tests of means. Food records themselves tend to underestimate dietary intake compared to the true gold standard, doubly labeled water (Goldberg et al. 1993), under certain weight-stable conditions. As one anonymous referee pointed out, even doubly labeled water may have additional measurement error with it, although we expect the error associated with it to be much smaller than the error associated with either FFQ or 24-hour recall.

The PIN study data are unique among nutritional epidemiology studies of dietary intake for many reasons, one of which is the collection of an FFQ *and* a biomarker in the main study. Typically, a study will collect the FFQ in the main study and then conduct a validation substudy to determine the relationship between the FFQ and the biomarker. As suggested by an anonymous referee, it would be interesting to see how our analytic results changed once we removed the biomarker in the main study. We conducted these analyses, including the model

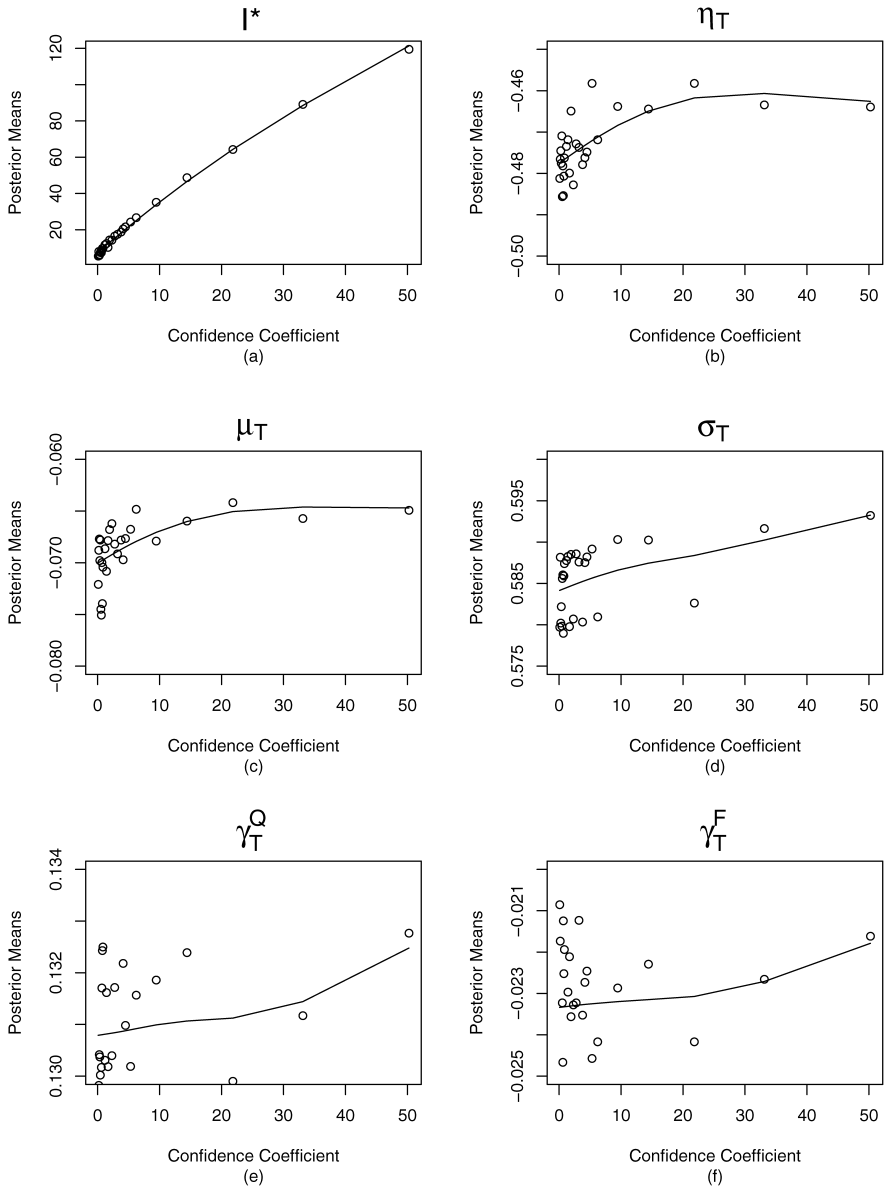

Figure 1. Effect of Confidence Parameter ξ on Latent Folate Concentration Parameters in MDP Analysis of PIN Study Data. I^* is the number of unique values of T; μ_T and σ_T are the mean and standard deviation of T, respectively; η_T is the folate effect on preterm birth; γ_T^Q and γ_T^F are the systematic biases in the FFQ and 24-hour recall, respectively.

comparison in Table 5, and found that our results are sensitive to the removal of these data. First, the measurement error model is too complex for the observed validation data in the PIN substudy. In addition to removing the correlation parameters (ρ_j) and subject-specific biases (i.e., $\mathbf{D} = 0$), a substantial simplification of the trimester-level means (α_j^Q, α_j^F) in (2)–(3) would be necessary. Second, the estimated FFQ–biomarker association using only the PIN substudy is too weak, and, hence, after removing the biomarker in the main study, the posterior means of $\boldsymbol{\eta}$ in the outcome model (1) look more like the "naive" estimates than calibrated or corrected estimates. Thus, the parameter estimates presented earlier do require the biomarker in the main study in an analysis of the PIN study data. In general, however, we conjecture that all parameters in the measurement error model (2)–(4) are estimable given sufficient data in the validation substudy. Therefore, our models and methods are not limited to studies that collect biomarkers in the main study.

Our analysis used serum folate biomarkers as unbiased measures of folate concentration. For the PIN study data, serum biomarkers were analyzed in one of four batches with over 80% of the sample analyzed in the first batch (specifically, the sample proportions were approximately .87, .05, .06, and .02, for batches 1–4, respectively). Siega-Riz et al. (2004) found that batch differences were nonnegligible and should be included in analyses using the serum biomarkers. Our analyses used the first batch as the reference group and placed vague, normal prior distributions on the remaining three batch effects. This additional caveat adds nothing new to the overall measurement error model (2)–(4) and was easily incorporated into our Bayesian framework in Section 4. Finally, while the serum folate biomarker is believed to be free from systematic biases, it is not without drawbacks, which involve individual-specific factors such as personal rates of metabolism. In an ideal experiment, one would use an objective biomarker, such as doubly labeled

water, rather than serum folate. Doubly labeled water is a measure of energy expenditure and intake (under certain weight-stable conditions) and is often regarded among the "best" biomarkers; however, it is not a true biomarker for any particular nutrient.

APPENDIX: FULL CONDITIONAL DISTRIBUTIONS

Let $\mathbf{D}_{\text{obs}}$ denote the observed data and let *rest* be short hand for all remaining parameters. Recall that Y_i is a Bernoulli outcome with canonical link function so that

$$\log L_i(Y_i|\mathbf{T}_i, \mathbf{Z}_i; \boldsymbol{\eta}) = Y_i(\eta_0 + \boldsymbol{\eta}_T'\mathbf{T}_i + \boldsymbol{\eta}_Z'\mathbf{Z}_i) + \log\{1 - \theta_i(\boldsymbol{\eta})\},$$

where $\theta(u) = 1/(1 + e^{-u})$.

1. Sample $[\boldsymbol{\eta}|rest, \mathbf{D}_{\text{obs}}]$ from $p(\boldsymbol{\eta}|rest, \mathbf{D}_{\text{obs}})$ using adaptive rejection sampling (Gilks and Wild 1992), where

$$p(\boldsymbol{\eta}|rest, \mathbf{D}_{\text{obs}}) \propto \exp\left\{\sum_{i=1}^{m} \log L_i(\boldsymbol{\eta}) - \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'\mathbf{V}_{0,\boldsymbol{\eta}}^{-1}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)\right\}.$$

2. Sample $[\boldsymbol{\gamma}|rest, \mathbf{D}_{\text{obs}}]$ from $\mathrm{N}\{\boldsymbol{\Lambda}_{\boldsymbol{\gamma}}\hat{\boldsymbol{\gamma}} + (\mathbf{I} - \boldsymbol{\Lambda}_{\boldsymbol{\gamma}})\boldsymbol{\gamma}_0, \boldsymbol{\Lambda}_{\boldsymbol{\gamma}}(\mathbf{H}' \times \mathbf{H})^{-1}\}$, where $\Lambda_{\boldsymbol{\gamma}} = (\mathbf{H}'\mathbf{H} + \mathbf{V}_{0,\boldsymbol{\gamma}})^{-1}\mathbf{H}'\mathbf{H}$ and $\hat{\boldsymbol{\gamma}} = (\mathbf{H}'\mathbf{H})^{-1} \times \mathbf{H}'(\mathbf{W} - \mathbf{R}\mathbf{b})$.
3. Let $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_m)'$. Sample $[\mathbf{b}|rest, \mathbf{D}_{\text{obs}}]$ from $\mathrm{N}(\boldsymbol{\Lambda}_b\hat{\mathbf{b}}, \boldsymbol{\Sigma}_W^{-1}\boldsymbol{\Lambda}_b(\mathbf{R}'\mathbf{R})^{-1})$, where $\boldsymbol{\Lambda}_b = (\mathbf{R}'\mathbf{R} + \mathbf{I}_m \otimes \mathbf{D}^{-1})^{-1}\mathbf{R}'\mathbf{R}$, $\otimes$ is the Kronecker product, and $\hat{\mathbf{b}} = (\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'(\mathbf{W} - \mathbf{H}\boldsymbol{\gamma})$.
4. Sample $[\mathbf{D}^{-1}|rest, \mathbf{D}_{\text{obs}}] \sim W_q(m + \nu_D, \mathbf{C}_D^{-1} + (\sum_{i=1}^{m} \mathbf{b}_i \times \mathbf{b}_i')^{-1})$.
5. Sample $(\boldsymbol{\sigma}, \boldsymbol{\rho})$ from $p(\boldsymbol{\sigma}, \boldsymbol{\rho}|rest, \mathbf{D}_{\text{obs}})$, where

$$p(\boldsymbol{\sigma}, \boldsymbol{\rho}|rest, \mathbf{D}_{\text{obs}}) \propto |\boldsymbol{\Sigma}_W|^{-1/2} \times \exp\left\{-\frac{1}{2}(\mathbf{W} - \mathbf{H}\boldsymbol{\gamma} - \mathbf{R}\mathbf{b})'\boldsymbol{\Sigma}_W^{-1}(\mathbf{W} - \mathbf{H}\boldsymbol{\gamma} - \mathbf{R}\mathbf{b})\right\} \times \pi(\boldsymbol{\sigma})\pi(\boldsymbol{\rho}).$$

6. Sample the error-prone covariate $[\mathbf{T}_i|rest, \mathbf{D}_{\text{obs}}]$ from $p(\mathbf{T}_i|rest, \mathbf{D}_{\text{obs}})$ for $i = 1, \ldots, m$, where

$$p(\mathbf{T}_i|rest, \mathbf{D}_{\text{obs}}) = \exp\left\{\log L_i(\boldsymbol{\eta}) - \frac{1}{2}(\mathbf{W}_i - \mathbf{H}_i\boldsymbol{\gamma} - \mathbf{R}_i\mathbf{b}_i)'\boldsymbol{\Sigma}_{W_i}^{-1}(\mathbf{W}_i - \mathbf{H}_i\boldsymbol{\gamma} - \mathbf{R}_i\mathbf{b}_i) - \frac{1}{2}(\mathbf{T}_i - \boldsymbol{\mu}_T)'\boldsymbol{\Sigma}_T^{-1}(\mathbf{T}_i - \boldsymbol{\mu}_T)\right\}.$$

7. Sample the missing FFQs and 24-hour recalls in the substudy assuming the observations are missing at random, leading to $[\mathbf{W}_i^{\text{miss}}|rest, \mathbf{D}_{\text{obs}}] \sim \mathrm{N}(\mathbf{H}_i\boldsymbol{\gamma} + \mathbf{R}_i\mathbf{b}_i, \boldsymbol{\Sigma}_{W_i})$.
8. Sample the missing biomarkers from $[\mathbf{M}_i^{\text{miss}}|rest, \mathbf{D}_{\text{obs}}] \sim \mathrm{N}(\mathbf{T}_i + b_{i3}, \sigma_M^2)$, assuming missing observations are missing at random.
9. Sample $[\boldsymbol{\mu}_T|rest, \mathbf{D}_{\text{obs}}] \sim \mathrm{N}(\boldsymbol{\Lambda}_T\bar{\mathbf{T}} + (\mathbf{I} - \boldsymbol{\Lambda}_T)\boldsymbol{\mu}_{0,T}, m^{-1} \times \boldsymbol{\Lambda}_T\boldsymbol{\Sigma}_T)$, where $\boldsymbol{\Lambda}_T = \mathbf{V}_{0,T}(m^{-1}\boldsymbol{\Sigma}_T + \mathbf{V}_{0,T})^{-1}$ where $\bar{\mathbf{T}} = m^{-1}\sum_{i=1}^{m}\mathbf{T}_i$.
10. Sample $[\boldsymbol{\Sigma}_T|rest, \mathbf{D}_{\text{obs}}] \sim W_{p_T}(m + \nu_{\boldsymbol{\Sigma}_T}, \mathbf{C}_{\boldsymbol{\Sigma}_T} + (\sum_{i=1}^{m}\mathbf{T}_i \times \mathbf{T}_i')^{-1})$.

For the MDP implementation, substitute all of Section 4.1 for step 6.

[Received April 2005. Revised January 2006.]

REFERENCES

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Non-Parametric Problems," *The Annals of Statistics*, 2, 1152–1174.

Block, G. (2001), "Invited Commentary: Another Perspective on Food Frequency Questionnaires," *American Journal of Epidemiology*, 154, 1103–1104.

Breslow, N. E., and Day, N. E. (1980), *Statistical Methods in Cancer Research*, Lyon: International Agency for Research on Cancer.

Brown, E. R., and Ibrahim, J. G. (2003), "A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data," *Biometrics*, 59, 221–228.

Brown, J. E., Buzzard, I. M., Jacobs, D. R., Hannan, P. J., Kushi, L. H., Barosso, G. M., and Schmid, L. A. (1996), "A Food Frequency Questionnaire Can Detect Pregnancy-Related Changes in Diet," *Journal of the American Dietetic Association*, 96, 262–266.

Byers, T. (2001), "Food Frequency Dietary Assessment: How Bad Is Good Enough?" *American Journal of Epidemiology*, 154, 1087–1088.

Carlin, B. P., and Louis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall/CRC Press.

Carroll, R. J. (2003), "Variances Are Not Always Nuisance Parameters," *Biometrics*, 59, 211–220.

Carroll, R. J., Ruppert, D., Crainiceanu, C., Tosteson, T., and Karagas, M. (2004), "Nonlinear and Nonparametric Regression and Instrumental Variables," *Journal of the American Statistical Association*, 99, 736–750.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, Boca Raton, FL: Chapman & Hall/CRC Press.

Erkkola, M., Karppinen, M., Javanainen, J., Rasanen, L., Knip, M., and Virtanen, S. M. (2001), "Validity and Reproducibility of a Food Frequency Questionnaire Finnish Women," *American Journal of Epidemiology*, 154, 466–476.

Escobar, M. D. (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268–277.

Forsythe and Gage (1994), **???**.

Fuller (1987), **???**.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gilks and Wild (1992), **???**.

Goldberg, G. R., Prentice, A. M., Coward, W. A., Davies, H. L., Murgatroyd, P. R., Wensing, C., Black, A. E., Harding, M., and Sawyer, M. (1993), "Longitudinal Assessment of Energy Expenditure in Pregnancy by the Doubly Labeled Water Method," *American Journal of Clinical Nutrition*, 57, 494–505.

Greeley, S., Storbakken, L., and Magel, R. (1992), "Use of a Modified Food Frequency Questionnaire During Pregnancy," *Journal of the American College of Nutrition*, 11, 728–734.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Holcroft, C., and Spiegelman, D. (1999), "Design of Validation Studies for Estimating the Odds Ratio of Exposure–Disease Relationships When Exposure Is Misclassified," *Biometrics*, 55, 1193–1201.

Kaaks, R., Riboli, E., Esteve, J., Van Kappel, A., and Vab Staveren, W. (1994), "Estimating the Accuracy of Dietary Questionnaire Assessments: Validation in Terms of Structural Equation Models," *Statistics in Medicine*, 13, 127–142.

Kipnis, V., Midthune, D., Freedman, L. S., Bingham, S., Schatzkin, A., Subar, A., and Carroll, R. J. (2001), "Empirical Evidence of Correlated Biases in Dietary Assessment Instruments and Its Implications," *American Journal of Epidemiology*, 153, 394–403.

Kipnis, V., Subar, A., Midthune, D., Freedman, L. S., Ballard-Barbash, R., Troiano, R., Bingham, S., Schoeller, D. A., Schatzkin, A., and Carroll, R. J. (2003), "The Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study," *American Journal of Epidemiology*, 158, 14–21.

Kleinman, K. P., and Ibrahim, J. G. (1998), "A Semi-Parametric Bayesian Approach to Generalized Linear Mixed Models," *Statistics in Medicine*, 17, 2579–2596.

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

——— (2002), **???**.

MacEachern, S. N. (1994), "Estimating Normal Means With a Conjugate Style Dirichlet Process Prior," *Communications in Statistics: Simulation and Computation*, 23, 727–741.

Mallick, B., Hoffman, F. O., and Carroll, R. J. (2002), "Semiparametric Regression Modeling With Mixtures of Berkson and Classical Error, With Application to Fallout From the Nevada Test Site," *Biometrics*, 58, 13–20.

McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman & Hall/CRC Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1092.

Metropolis, N., and Ulam, S. (1949), "The Monte Carlo Method," *Journal of the American Statistical Association*, 44, 335–341.

Morrissey, M. J., and Spiegelman, D. (1999), "Matrix Methods for Estimating Odds Ratios With Misclassified Exposure Data: Extensions and Comparisons," *Biometrics*, 55, 338–344.

Müler and Roeder (1997), **???**.

Robinson, S., Godfrey, K., Osmond, C., Cox, V., and Barker, D. (1996), "Evaluation of a Food Frequency Questionnaire Used to Assess Nutrient Intakes in Pregnant Women," *European Journal of Clinical Nutrition*, 50, 302–308.

Savitz, D. A., Dole, N., Terry, J. W., Zhou, H., and Thorp, J. M. (2001), "Smoking and Pregnancy Outcome Among African-American and White Women in Central North Carolina," *Epidemiology*, 12, 636–642.

Savitz, D. A., Dole, N., Williams, J., Thorp, J. M., McDonald, T., Carter, C. A., and Eucker, B. (1999), "Study Design and Determinants of Participation in an Epidemiologic Study of Preterm Delivery," *Paediatric and Perinatal Epidemiology*, 13, 114–125.

Savitz, D. A., Henderson, L., Dole, N., Herring, A., Wilkins, D. G., Rollins, D., and Thorp, J. M. (2002), "Indicators of Cocaine Exposure and Preterm Birth," *Obstetrics and Gynecology*, 99, 458–465.

Searle, S. R. (1971), *Linear Models*, New York: Wiley.

Siega-Riz, A. M., Savitz, D. A., Zeisel, S. H., Thorp, J. M., and Herring, A. H. (2004), "Second Trimester Folate Status and Preterm Birth," *American Journal of Obstetrics and Gynecology*, 191, 1851–1857.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 64, 583–639.

Spiegelman, D., Carroll, R. J., and Kipnis, V. (2001), "Efficient Regression Calibration for Logistic Regression in Main Study/Internal Validation Study Designs," *Statistics in Medicine*, 29, 139–161.

Spiegelman, D., Rosner, B., and Logan, R. (2000), "Estimation and Inference for Binary Data With Covariate Measurement Error and Misclassification for Main Study/Validation Study Designs," *Journal of the American Statistical Association*, 95, 51–61.

Spiegelman, D., Zhao, B., and Kim, J. (2004), "Correlated Errors in Biased Surrogates: Study Designs and Methods for Measurement Error Correction," *Statistics in Medicine*, **??**, **??**–**??**.

Stefanski, L. A., and Boos, D. D. (2002), "The Calculus of M-Estimation," *The American Statistician*, 56, 29–38.

Stefanski, L. A., and Carroll, R. J. (1985), "Covariate Measurement Error in Logistic Regression," *The Annals of Statistics*, 13, 1335–1351.

Subar, A. F., Kipnis, V., Troiano, R. P., Midthune, D., Schoeller, D. A., Bingham, S., Sharbaugh, C. O., Trabulsi, J., Runswick, S., Ballard-Barbash, R., Sunshine, J., and Schatzkin, A. (2003), "Using Intake Biomarkers to Evaluate the Extent of Dietary Misreporting in a Large Sample of Adults: The OPEN Study," *American Journal of Epidemiology*, 158, 1–13.

Suitor, C. J., Gardner, J., and Willett, W. C. (1989), "A Comparison of Food Frequency and Diet Recall Methods in Studies of Nutrient Intake of Low-Income Pregnant Women," *Journal of the American Dietary Association*, 89, 1786–1794.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association*, 82, 528–550.

U.S. Department of Agriculture Food and Nutrition Service (1994), *WIC Dietary Assessment Validation Study Executive Summary: Freeman*, San Francisco: Sullivan and Company.

Willett, W. (1998), *Nutritional Epidemiology*, New York: Oxford University Press.

——— (2001), "Invited Commentary: A Further Look at Dietary Questionnaire Validation," *American Journal of Epidemiology*, 154, 1100–1102.

Zucker and Spiegelman (2004), **???**.