Bayesian Inferences on Predictors of Day-Specific Conception Probabilities

David B. Dunson[1,*] **and Joseph B. Stanford**[2]

[1] Biostatistics Branch, National Institute of Environmental Health Sciences, MD A3-03,
P.O. Box 12233, Research Triangle Park, NC 27709

[2] Department of Family and Preventive Medicine, University of Utah

[*] *email*: dunson1@niehs.nih.gov

For couples attempting pregnancy, it is important to identify predictors of the day-specific probabilities of conception in relation to the timing of a single intercourse act. Because most menstrual cycles have multiple days of intercourse, the occurrence of conception represents the aggregation across Bernoulli trials for each intercourse day. Due to this aggregated Bernoulli data structure and to dependency among the multiple cycles from a woman, implementing analyses has proven challenging, particularly when predictors are day-specific. This article proposes a Bayesian approach for addressing this problem, based on a generalization of the Barrett and Marshall model to incorporate a woman-specific frailty and day-specific covariates. The model results in a simple closed form expression for the marginal probability of conception, and has an underlying variables formulation which facilitates efficient posterior computation. A conjugate variable selection prior is proposed, which allows one-sided hypothesis testing, and the methods are applied to simulated and real data examples.

KEY WORDS: Aggregated Bernoulli; Data augmentation algorithm; Gamma frailty; Human fertility; Nonlinear mixed model; Pregnancy; Random effect.

1. Introduction

Identifying predictors of the probabilities of conception in relation to the timing and frequency of intercourse in the menstrual cycle is important for couples attempting pregnancy, users of natural family planning methods, clinicians diagnosing possible causes of infertility, and reproductive epidemiologists. Often interest focuses on one or more day-specific predictors, such as the type of cervical mucus detected by the woman on the day of intercourse or the level of a hormone. Because there may be multiple intercourse acts that occur during the potentially fertile phase of the cycle, the occurrence of conception represents the aggregation across Bernoulli trials for each intercourse day (Zhou and Weinberg, 1996). Due to this aggregated Bernoulli data structure and to the need to account for dependency among the multiple cycles from a woman, statistical analyses can be challenging to implement.

Most analyses of day-specific probabilities of conception have relied on maximum likelihood estimation of the model of Schwartz, MacDonald and Heuchel (1980), which has the form

$$\Pr(Y_{ij} = 1 \,|\, \mathbf{X}_{ij}) = \omega \Big\{ 1 - \prod_k (1 - \lambda_k)^{X_{ijk}} \Big\}, \tag{1}$$

where Y_{ij} is an indicator of conception in cycle j from woman i, $\mathbf{X}_{ij} = (X_{ij1}, \ldots, X_{ijK})'$ is a vector of intercourse indicators for days $1, \ldots, K$ (possibly indexed relative to day of ovulation), ω is the "cycle viability" probability, and λ_k is the probability of conception in a viable cycle with intercourse on only day k. This model does not incorporate covariates or allow for within-couple dependency. In addition, estimates can be highly unstable due to weak identiability between ω and the maximum of the λ_k's.

Generalizations of the Schwartz et al. (1980) model have been proposed to accommodate cycle-specific covariates (Weinberg et al., 1994), within-woman dependency (Zhou et al, 1996; Dunson and Zhou, 2000), and day-specific covariates (Zhou and Weinberg, 1996). Zhou and Weinberg (1996) proposed an EM algorithm for maximum likelihood estimation and used a sandwich estimator of the variance to adjust for within-woman dependency. Unfortunately, estimates of day-specific probabilities are biased downwards and the sandwich estimator is not valid because less fertile women contribute more cycles to the data set. Royston and Ferreira (1999) instead proposed a simple alternative to the Schwartz model, which assumes that in cycles with multiple intercourse acts only the most fertile contributes to the probability of conception. Although this may be a reasonable approximation in some cases, sperm introduced on less optimal days in the fertile window can also compete to fertilize the ovum and should not be ignored.

Dunson (2001) proposed a Bayesian hierarchical modeling approach, which accommodates day-specific covariates and heterogeneity among women. The model allows for differences in both the probability of conception on the most fertile day and the duration of the fertile interval of the cycle. Unfortunately, the highly computationally intensive MCMC algorithm used for posterior computation presents a barrier to routine implementation of data analyses and simulation studies to assess operating characteristics. In addition, the approach does not allow for selection of predictors or for inferences on ordered trends in the day-specific probabilities of conception across levels of a categorical predictor (e.g., type of mucus).

Motivated by these limitations and by data from several recent fecundability studies collecting day-specific predictors (Colombo and Massarotto, 2000; Stanford, Smith and Dunson, 2003; Mion et al., 2004), this article proposes a new approach. A hierarchical model is specified for the probability of conception, which is more stable than existing models, has simplified parameter interpretation, and has a structure that facilitates efficient posterior computation. A variable selection-type prior is then proposed, which allows predictors to be adaptively dropped from the model while also allowing inferences on trends in the probability of conception across levels of a categorical predictor. Posterior computation then proceeds via an auxiliary variables stochastic search MCMC algorithm.

Section 2 proposes the model and prior. Section 3 outlines the algorithm for posterior computation and inferences. Section 4 applies the approach to data from a European fecundability study, and Section 5 discusses the results.

2. Bayesian Modeling of Conception Probabilities

2.1 *Hierarchical model*

Letting $\mathbf{U}_{ij} = [\mathbf{u}'_{ij1}, \ldots, \mathbf{u}'_{ijk}]'$ be a $k \times q$ covariate matrix for cycle j ($j = 1, \ldots, n_i$) from couple i ($i = 1, \ldots, n$), we propose the following hierarchical model for the probability of conception:

$$\begin{aligned} \Pr(Y_{ij} = 1 \,|\, \xi_i, \mathbf{X}_{ij}, \mathbf{U}_{ij}) &= 1 - \prod_{k=1}^{K} (1 - \lambda_{ijk})^{X_{ijk}} \\ \lambda_{ijk} &= 1 - \exp\{-\xi_i \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})\} \\ \xi_i &\sim \mathcal{G}(\phi, \phi), \end{aligned} \tag{2}$$

where λ_{ijk} is the day-specific probability of conception in cycle j from couple i given intercourse only on day k, ξ_i is a fecundability multiplier for couple i, $\mathrm{E}(\xi_i) = 1$, $\mathrm{var}(\xi_i) = \phi^{-1}$ measures the level of couple-to-couple variability, and $\boldsymbol{\beta}$ is a vector of regression coefficients. Because the Schwartz et al. (1980) model (1) and its generalizations tend to be weakly identified and unstable, we exclude the cycle viability multiplier ω and instead include a random effect ξ_i in the day-specific term. In the special case where $\lambda_{ijk} = \lambda_k$ for all i, j, our model simplifies to the form proposed by Barrett and Marshall (1969).

By using expression (2), results tend to be much more stable than for the Schwartz et al. model and its generalizations. In addition, by placing the random effect and covariate effects on the same scale as terms in a regression model for the linear predictor, $\eta_{ijk} = \mathbf{u}'_{ijk}\boldsymbol{\beta} + \log \xi_i$, we avoid the pitfall of nonlinear random effects models mentioned by Cox (1984). In particular, when the random effect and covariate effects are incorporated in different components (e.g., in the ω and λ_k terms) the regression coefficients can be sensitive to the covariance structure, making results difficult to interpret. We assess model stability and frequentist operating characteristics through a simulation study in Section 4.

An additional appealing property of model (2) is that the marginal probability of conception, integrating out the couple-specific frailty ξ_i, has a simple closed form:

$$\begin{aligned}\Pr(Y_{ij} = 1 \,|\, \mathbf{X}_{ij}, \mathbf{U}_{ij}) &= 1 - \int_0^\infty \exp\Big\{ -\xi_i \sum_{k=1}^K X_{ijk} \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})\Big\} \mathcal{G}(\xi_i; \phi, \phi)\, d\xi_i \\ &= 1 - \Big(\frac{\phi}{\phi + \sum_{k=1}^K X_{ijk} \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})}\Big)^{\phi}. \end{aligned} \tag{3}$$

Thus, the marginal day-specific probability of conception in a cycle with intercourse only on day k and predictors $\mathbf{u}$ is

$$\Pr(Y = 1 \,|\, \mathbf{u}) = 1 - \Big(\frac{\phi}{\phi + \exp(\mathbf{u}'\boldsymbol{\beta})}\Big)^{\phi}, \tag{4}$$

which is a logistic regression model for $\phi = 1$. For biomedical journals familiar with logistic regression, one could focus on the simplified model fixing $\phi = 1$ as an approximation.

2.2 *Prior specification*

Our interest focuses on selecting predictors of the day-specific conception probabilities, and quantifying the effects of these predictors. A particular focus is ordered categorical predictors, such as mucus score or age group. To address these goals, we choose variable selection priors for the regression coefficients, $\boldsymbol{\beta}$, which assign point masses at zero values to allow predictors to be dropped from the model. In particular, letting $\gamma_h = \exp(\beta_h)$, we choose priors of the form

$$\pi(\boldsymbol{\gamma}) = \prod_h \delta_1 - \mathcal{G}_{\mathcal{A}_h}(\gamma_h; p_h, a_h, b_h), \tag{5}$$

where $\delta_1 - \mathcal{G}_{\mathcal{A}_h}(p_h, a_h, b_h)$ denotes the density consisting of the mixture of a point mass at one (with probability p_h) and a $\mathcal{G}(a_h, b_h)$ density truncated to the region $\mathcal{A}_h$, with $\mathcal{G}(a_h, b_h)$

denoting the gamma density with mean a_h/b_h and variance a_h/b_h^2. We complete a Bayesian specification of the model with a $\mathcal{G}(c_1, c_2)$ prior for ϕ.

Values of $\gamma_h = 1$ correspond to $\beta_h = 0$ and to the hth predictor in $\mathbf{u}_{ijk}$ being effectively dropped from the model. Hence, under prior (5) the probability of the hth predictor being included is $1 - p_h$. If the predictor is included, then the exponentiated regression coefficient γ_h is constrained to fall in the region $\mathcal{A}_h$, which is typically chosen to be $\Re^+$, $(1, \infty)$, or $(0, 1)$ to correspond to no constraint, positive effect of the predictor on the probability of conception, and negative effect, respectively. By appropriately choosing $\mathcal{A}_h$, one can accommodate *a priori* information on the direction of the association to decrease posterior uncertainty. Because point mass is assigned to the null hypothesis of no association, one or two-sided hypothesis tests can then be conducted.

The prior distribution in expression (5) was carefully chosen to have a conditionally-conjugate form after introducing auxiliary variables, as described in Section 3. In some settings, conjugacy is no longer that important due to the availability of MCMC algorithms, such as Metropolis-Hastings, which do not require closed forms. However, computational efficiency in Bayesian variable selection problems is greatly facilitated by the use of conjugate or conditionally-conjugate priors. For references on Bayesian variable selection using point mass mixture priors and stochastic search algorithms, refer to Geweke (1996), George and McCulloch (1997), and Chipman et al. (2001).

3. Posterior Computation

For posterior computation we propose an efficient auxiliary variables MCMC algorithm. We first reexpress model (2) as follows:

$$Y_{ij} = 1\Big(\sum_{k=1}^{K} X_{ijk} Z_{ijk} > 0\Big),$$

$$Z_{ijk} \overset{ind}{\sim} \text{Poisson}(\xi_i \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})), k = 1, \ldots, K, \tag{6}$$

where $\mathbf{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijK})'$ is a vector of independent Poisson latent variables following log-linear gamma frailty models. To show that expressions (2) and (6) are equivalent, we can integrate out the auxiliary Poisson variables, which are only introduced for computational purposes and have no effect on parameter interpretation. In particular, under (6), $Y_{ij} = 0$ if and only if $Z_{ijk} = 0$ for all k such that $X_{ijk} = 1$ (i.e., all days in the i, jth cycle with intercourse). Because $\Pr(Z_{ijk} = 0 \,|\, \xi_i, \mathbf{u}_{ijk}) = \exp\{-\xi_i \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})\}$ and the Poisson variables are independent, we have

$$\Pr(Y_{ij} = 0 \,|\, \xi_i, \mathbf{X}_{ij}, \mathbf{U}_{ij}) = \prod_{k:X_{ijk}=1} \Pr(Z_{ij} = 0 \,|\, \xi_i, \mathbf{u}_{ijk}) = \prod_{k=1}^{K} \exp\{-\xi_i \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})\}^{X_{ijk}},$$

which is in the form $\prod_k (1 - \lambda_{ijk})^{X_{ijk}}$. Expression (2) follows directly.

Using the auxiliary variables form of the model, the joint posterior distribution is $\propto$

$$\Bigg(\prod_{i=1}^{n} \mathcal{G}(\xi_i; \phi, \phi) \prod_{j=1}^{n_i} \Big\{ 1\Big(\sum_{k=1}^{K} X_{ijk}Z_{ijk} > 0\Big) Y_{ij} + 1\Big(\sum_{k=1}^{K} X_{ijk}Z_{ijk} = 0\Big)(1 - Y_{ij}) \Big\} \times \Bigg[\prod_{k=1}^{K} \frac{\big\{\xi_i \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})\big\}^{Z_{ijk}} \exp\big\{-\xi_i \exp(\mathbf{u}'_{ijk}\boldsymbol{\beta})\big\}}{Z_{ijk}!} \Bigg] \Bigg) \pi(\boldsymbol{\beta})\, \pi(\phi). \tag{7}$$

Following standard algebraic routes, one can then obtain the full conditional posterior distributions for each of the parameters and latent variables. Focusing on the case where each of the predictors is categorical, which is true in the applications motivating this article and in most reproductive epidemiology studies, each of the conditional distributions has a form which is straightforward to sample from directly, with the exception of ϕ. Thus, for posterior computation we recommend a simple hybrid Gibbs sampling and Metropolis algorithm, with a single Metropolis step for updating ϕ. Details are outlined in Appendix A.

This algorithm is a type of stochastic search variable selection algorithm, which samples from the posterior distribution for all of the unknowns including the subset of predictors to be included in the model. Letting $M_h = 1(\beta_h \neq 0)$, for $h = 1, \ldots, q$, the model is indexed by the vector $\mathbf{M} = (M_1, \ldots, M_q)'$. Then, letting $\mathbf{M}^{(s)}$ denote the value of $\mathbf{M}$ at iteration s of the MCMC algorithm starting after a burn-in interval, we can estimate posterior model probabilities using

$$\widehat{\Pr}(\mathbf{M} = \mathbf{m} \,|\, \text{data}) = \frac{1}{S} \sum_{s=1}^{S} 1(\mathbf{M}^{(s)} = \mathbf{m}).$$

In addition to subset selection, interest commonly focuses on assessing whether a given predictor is associated with the probability of conception. For a simple binary predictor, u_{ijkh}, the probability of the null hypothesis of no association, $H_{0h} : \gamma_h = 1$, can be estimated directly from the MCMC output using the Rao-Blackwellized estimator:

$$\widehat{\Pr}(H_{0h} \,|\, \text{data}) = \frac{1}{S} \sum_{s=1}^{S} \widetilde{p}_h^{(s)}, \tag{8}$$

where $\widetilde{p}_h$ is the conditional posterior probability of $\gamma_h = 1$ at iteration s shown in Appendix A.

In many cases (as in the application of Section 4), interest focuses on assessing evidence of an association between an ordered categorical predictor $w_{ijk} \in \{1, \ldots, d\}$ and the occurrence of conception. Parameterizing the model so that the first $d-1$ elements of $\mathbf{u}_{ijk}$ have the structure $[1(w_i \geq 2), \ldots, 1(w_i \geq d)]'$ and the remaining elements consist of other covariates (e.g., indicators of timing in the fertile interval relative to ovulation), the null hypothesis of no association can be expressed as $H_0 : \gamma_1 = \ldots = \gamma_{d-1} = 1$. The prior probability of H_0 is then $\prod_{h=1}^{d-1} p_h$. By choosing constrained $\mathcal{A}_h$, for $h = 1, \ldots, d-1$, one can formulate a one-sided alternative. For example, $\mathcal{A}_h = (1, \infty)$, results in a non-decreasing constraint, with the alternative hypothesis corresponding to at least one increase in the probability of conception across the range of w_{ijk}. This approach is applied in the next Section.

4. Application to European Fertility Study

4.1 *Background and Data*

The European Study of Daily Fecundability (Colombo and Masarotto, 2000) enrolled 782 women recruited from seven centers (Milan, Verona, Lugano, Dusseldorf, Paris, London, Brussels) providing services on fertility awareness and natural family planning. Women enrolled were between 18 and 40 years of age, were not taking hormonal medications or drugs affecting fertility, and had no known impairment of fecundity. The participants kept daily records of basal body temperature and vaginal observations from cervical mucus, and recorded the days during which intercourse and menstrual bleeding occurred. Ovulation days were estimated for each menstrual cycle under study using the last day of hypothermia prior to the post-ovulatory rise in basal body temperature. For a detailed description of the study protocol, refer to Colombo and Masarotto (2000).

Our interest focuses on using data from the European study to relate daily vaginal observations from cervical mucus to the day-specific probabilities of pregnancy from intercourse in relation to the estimated day of ovulation. It is well known that successful sperm transport and survival depends on the presence and amount of estrogenic cervical mucus in the female reproductive tract (Elstein and Moghissi, 1973). In the European study, the fertile potential of the cervical mucus is assessed daily by the woman (after training by a fertility awareness provider) on a four point ordinal scale, ranging from 1=dry to 4=most fertile-type mucus. Focusing on menstrual cycles having complete records of mucus and at least one intercourse day in the six day fertile interval ending on the estimated day of ovulation, there are 1473 menstrual cycles from 516 women, with 343 of these cycles ending in a clinical pregnancy.

We are interested in using the proposed methodology to assess the relationship between daily observations of cervical mucus and the day-specific probabilities of conception. Let w_{ijk} denote the categorical mucus score on day k of cycle j for woman i, with $w_{ijk} = 1$ indicating dry, $w_{ijk} = 2$ a humid or damp feeling, $w_{ijk} = 3$ thick, creamy, or whitish mucus, and $w_{ijk} = 4$ slippery, stretchy or clear mucus. Such vaginal observations by women correlate well with the hydration and related biophysical characteristics of cervical mucus, and a certain level of mucus hydration is necessary for sperm survival and transport (Hilgers and Prebil, 1979; Katz et al., 1997; Odeblad, 1997). Further, changes in mucus hydration increase as ovulation approaches and immediately decrease thereafter (Billings et al., 1972; Dunson and Colombo, 2003).

Based on these biological considerations, it is thought that intercourse on days with $w_{ijk} = 1$ is unlikely to result in conception, that intercourse on days with $w_{ijk} = 4$ is highly likely to result in conception, and that days with $w_{ijk} = 2$ or 3 represent days with intermediate probabilities of conception. The systematic evaluation of this clinical algorithm is of high interest to providers of fertility awareness methods and to couples using daily vaginal observations to predict their likelihood of conception from intercourse on a given day.

4.2 *Model and Prior Specification*

We used the model described in expression (2) to relate the daily observations of cervical mucus and the timing of intercourse in the fertile interval to the probability of conception by letting

$$\mathbf{u}_{ijk} = [1(k = 1), 1(k = 2), \ldots, 1(k = 6), 1(w_{ijk} \geq 2), 1(w_{ijk} \geq 3), 1(w_{ijk} = 4)]',$$

where k indexes the day in the fertile interval, with $k = 1$ five days prior to ovulation and $k = 6$ on the identified day of ovulation. The regression coefficients are divided into two subvectors $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, with $\boldsymbol{\beta}_1$ characterizing the baseline changes in the probability of conception according to timing in the fertile interval, and $\boldsymbol{\beta}_2$ characterizing the changes across levels of the 1-4 mucus score.

To choose a prior for the exponentiated baseline coefficients, $\lambda_k = \exp(\beta_{1k})$ for $k = 1, \ldots, 6$, we first set the point mass probabilities equal to zero, $p_k = 0$, because the probability of conception should be positive within the six day fertile interval ending on the day of ovulation (Wilcox et al, 1995). By instead choosing a wider potential fertile interval and then allowing positive point mass probabilities, one automatically accounts for uncertainty in the choice of fertile interval. In this case, the stochastic search variable selection algorithm produces draws from the posterior distribution for the days in the fertile interval. We do not follow that approach here, because the European study collected mucus data for a relatively narrow interval of days in the cycle, and expanding the fertile interval results in many days with missing observations. We choose a diffuse prior for the values of $\boldsymbol{\lambda}$ by letting $a_{0k} = b_{0k} = 0.1$ for $k = 1, \ldots, 6$.

The exponentiated regression coefficients, $\gamma_h = \exp(\beta_{2h})$ for $h = 1, 2, 3$, measure changes in the probability of conception associated with the mucus score increasing from 1 to 2, 2 to 3, and 3 to 4, respectively. As the mucus score increases, the day-specific probabilities of conception should be non-decreasing, given the well known role of mucus in regulating passage of sperm. This non-decreasing restriction can be incorporated by letting $\mathcal{A}_h = (1, \infty)$ in the prior for γ_h. By incorporating a point mass at $\gamma_h = 1.0$ for $h = 1, 2, 3$, we then assign positive prior probability to flat regions across which increases in the mucus score have no effect. Following the strategy of Westfall, Johnson and Utts (1997), we let $p_h = 0.5^{1/3}$ in order to assign 0.5 prior probability to the null hypothesis of no association between the mucus score and the day-specific conception probabilities. We let $a_h = b_h = 0.1$ to allow a high degree of uncertainty in the values of γ_h under the alternative hypothesis. Finally, we choose $c_1 = 1$ and $c_2 = 2$ to specify a weakly informative prior for the frailty variance ϕ. We will assess the sensitivity of our inferences to this prior choice by conducting a simulation study and by repeating our analysis using reasonable alternative priors.

4.3 *Simulation Study*

We conducted a simulation study to check the robustness of our inferences to the prior specification and choice of model, and to assess the frequentist operating characteristics. We first simulated data under the model of Dunson and Zhou (2000), using previous estimates obtained under that model and assuming no association between the mucus score and the day-specific conception probabilities. Using the intercourse and mucus data from the European study, we generated conception outcome data to obtain 100 simulated data sets, and then analyzed each of these data sets using the proposed procedure. The priors were specified as described in Subsection 4.2, and for each dataset we ran our MCMC algorithm for 1500 iterations, starting at values which differed from the true values and discarding the first 500 iterations as a burn-in. Convergence was rapid and mixing was excellent, and these burn-in and collection intervals were deemed sufficient.

Using the estimated posterior probability of $\gamma_h = 1$ as a Bayesian alternative to the p-value, the proportion of simulated data sets having $\widehat{\Pr}(\gamma_h = 1 \,|\, \text{data}) \leq 0.05$ was 0.05, 0.00, and 0.03 for γ_1, γ_2 and γ_3, respectively. Thus, it appears that our Bayesian procedure yields a valid frequentist test, with no evidence of an inflated type I error rate, even when the model is incorrectly specified. Similar results are obtained when data are simulated under the null hypothesis using model (2) with the parameterization of Subsection 4.2.

To investigate the performance of the method under the alternative hypothesis, we simulated an additional 100 data sets under our proposed model, with $\boldsymbol{\lambda} = (0.04, 0.14, 0.08, 0.34, 0.31, 0.08)$ (values chosen based on Wilcox et al., 1998), $\boldsymbol{\gamma} = (1.5, 1.5, 1.5)'$, and $\phi^{-1} = 0.5$. We then analyzed each data set using the approach described above. The resulting proportion of simulated data sets (out of 100) having $\widehat{\Pr}(\gamma_h = 1 \,|\, \text{data}) \leq 0.05$ was 0.90, 0.90, and 1.00 for γ_1, γ_2, and γ_3, respectively, suggesting high power to detect $\gamma_h > 1$. Table 1 presents averages of the estimated posterior means, empirical 95% intervals, and coverages of the estimated 95% credible intervals for each of the parameters. As expected, there is a small positive bias in the estimated posterior means for the restricted parameters, $\boldsymbol{\gamma}$, as would be the case for any restricted estimation procedure. However, for the unconstrained baseline parameters, $\boldsymbol{\lambda}$, and frailty variance, ϕ, there is no evidence of bias, and the average posterior mean across the simulations is close to the true value for each parameter. In addition, the coverage of the 95% credible intervals is close to the nominal 95% level in each case.

4.4 *Real Data Results*

To obtain results for the real data from the European study, we repeated the MCMC analysis used in the simulation studies, but with a burn-in of 5000 iterations and a collection interval of 45,000 iterations. The prior and posterior densities for the mucus effect parameters, $\gamma_1, \gamma_2, \gamma_3$, are plotted in Figure 1. Although we used a shrinkage prior that assigned a moderately high probability to $\gamma_h = 1.0$ in order to adjust for a possibly inflated type I error rate due to multiple testing, there was clear evidence in the data in favor of $\gamma_h > 1.0$. In particular, the estimated posterior probabilities of $\gamma_h = 1.0$ (i.e., increasing the mucus score from h to $h + 1$ has no effect on the day-specific conception probabilities) were 0.05, 0.03, and < 0.01 for γ_1, γ_2, and γ_3, respectively.

Thus, each unit increase of the mucus score results in a significant increase in the day-specific conception probabilities. Posterior summaries of the parameters are presented in Table 2, and the estimated day-specific conception probabilities stratified by the mucus score are plotted in Figure 2. Surprisingly, it appears that the mucus score does an even better job of predicting the day-specific conception probabilities than the timing of intercourse relative to ovulation. This is a biologically interesting and clinically significant result, to be published in a companion paper by Bigelow et al. (2004).

Because $\phi \approx 1$, the probability of conception in a cycle with intercourse only on day k follows an approximate logistic regression model:

$$\Pr(Y_{ij} = 1 \,|\, X_{ijk} = 1, \mathbf{X}_{ij(-k)} = \mathbf{0}, w_{ijk} = h) \approx \frac{\exp(\beta_{1k} + \beta_{21} + \ldots + \beta_{2,h-1})}{1 + \exp(\beta_{1k} + \beta_{21} + \ldots + \beta_{2,h-1})}, \qquad (9)$$

where $\exp(\beta_{1k})/\{1+\exp(\beta_{1k})\}$ is the (approximate) day-specific probability of conception in a cycle with a mucus score of 1 given intercourse on only day k, and β_{2h} is interpretable as the (approximate) increase in the log odds of conception attributable to a unit increase in the mucus score from h to $h+1$. Because the data are consistent with this simplified model, it is reasonable to focus on this model in presenting results in substantive journals having an audience familiar with logistic regression.

5. Discussion

This article has proposed a Bayesian approach for inferences on predictors of day-specific conception probabilities in the menstrual cycle. A variable selection-type prior is used for the regression coefficients, which allows uncertainty in the predictors to be included in the model, including the specific days in the fertile interval. This prior also allows sign constraints and can be used to construct hypothesis tests comparing null hypotheses of no association to order restricted or unrestricted alternatives. Order restricted alternatives have clear advantages when categorical predictors are of interest and one has *a priori* knowledge of the direction of the effect, as is often the case (e.g., in the data example of Section 4). Routine implementation is greatly facilitated by the simplicity and computational efficiency of the proposed auxiliary variables MCMC algorithm. Due to the excellent efficiency, we were able to run simulation studies of the frequentist operating characteristics, an exercise which is very useful in evaluating Bayesian methods but often not practical. In contrast, previous Bayesian methods for analysis of conception probabilities (Dunson and Zhou, 2000; Dunson, 2001) do not allow variable selection or order restricted inference, and tend to have much slower rates of convergence and mixing.

Although our focus has been studies of conception probabilities, the proposed underlying Poisson variable structure can be easily adapted for other types of data structures. For example, the proposed prior and auxiliary variables MCMC algorithm can also be used for variable selection and order restricted inference in logistic regression models for binary outcomes and log-linear regression models for counts with minor modifications.

Acknowledgements

The authors thank Rene Ecochard for helpful discussions, and the other European investigators for generously providing the data used in the application. In particular, we thank Bernardo Colombo, Guido Masarotto, Guenther Freundl, Christian Gnoth, Petra Frank-Herrmann, Sandro Girotto, Michele Barbato, and Cecilia Pyper.

References

Barrett, J. C. and Marshall, J. (1969). The risk of conception on different days of the menstrual cycle. *Population Studies* **23**, 455-461.

Bigelow, J.L., Dunson, D.B., Stanford, J.B., Ecochard, R., Gnoth, C. and Colombo, B. (2004). Mucus observations in the fertile window: a better predictor of conception than timing of intercourse. *Human Reproduction*, in press.

Billings, E.L., Brown, J.B., Billings, J.J. and Burger, H.G. (1972). Symptoms and hormonal changes accompanying ovulation. *Lancet* **1**, 282-284.

Chen, M.H. and Shao, Q.M. (1998). Monte Carlo methods for Bayesian analysis of constrained parameter problems. *Biometrika* **85**, 73-87.

Chipman, H., George, E.I., and McCulloch, R.E. (2001). The practical implementation of Bayesian model selection. *IMS Lecture Notes - Monograph Series Volume 38.*

Colombo, B. and Masarotto, G. (2000). Daily fecundability: first results from a new data base. *Demographic Research* **3**, 5.

Cox, D.R. (1984). Interaction. *International Statistical Review*, 1-31.

Dunson, D.B. (2001). Bayesian models for distinguishing effects on the level and duration of fertility in the menstrual cycle. *Biometrics* **57**, 1067-1073.

Dunson, D.B. and Colombo, B. (2003). Bayesian modeling of markers of day-specific fertility. *Journal of the American Statistical Association* **98**, 28-37.

Dunson, D.B. and Zhou, H. (2000). Bayesian modeling of fecundability and sterility. *Journal of the American Statistical Association* **95**, 1054-1062.

Elstein, M., Moghissi, K.S., and Borth, R. (1973). *Cervical Mucus in Human Reproduction.* Copenhagen: Scriptor, 128-151.

George, E.I. and McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339-373.

Geweke, J. (1996). Variable selection and model comparison in regression. In *Bayesian Statistics* **5**, J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (eds.). Oxford University Press, 609-620.

Hilgers, T.W., Prebil, A.M. (1979). The ovulation method–vulvar observations as an index of fertility/infertility. *Obstetrics and Gynecology* **53**, 12-22.

Katz, D.F., Slade, D.A., and Nakajima, S.T. (1997). Analysis of pre-ovulatory changes in cervical mucus hydration and sperm penetrability. *Advances in Contraception* **13**, 143-151.

Mion, A., Passarin, K., Scarpa, B., and Colombo, B. (2004). Cervical mucus symptom and daily fecundability: First results from a new data base. *Demographic Research*, under review.

Odeblad E. (1997). Cervical mucus and their functions. *Journal of the Irish Colleges of Physicians and Surgeons* **26**, 27-32.

Royston, P. and Ferreira, A. (1999). A new approach to modeling daily probabilities of conception. *Biometrics* **55**, 1005-1013.

Schwartz, D., MacDonald, P. D. M., and Heuchel, V. (1980). Fecundability, coital frequency and the viability of ova. *Population Studies* **34**, 397-400.

Stanford, J.K., Smith, K., and Dunson, D.B. (2003). Vulvar mucus observations and the probability of pregnancy. *Obstetrics and Gynecology* **6**, 1285-1293.

Weinberg, C. R., Gladen, B. C., and Wilcox, A. J. (1994). Models relating the timing of intercourse to the probability of conception and the sex of the baby. *Biometrics* **50**, 358-367.

Westfall, P.H., Johnson, W.O. and Utts, J.M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika* **84**, 419-427.

Wilcox, A.J., Weinberg, C.R., and Baird, D.D. (1995). Timing of sexual intercourse in relation to ovulation: Effects on the probability of conception, survival of the pregnancy and sex of the baby. *New England Journal of Medicine* **333**, 1517-1521.

Wilcox, A. J., Weinberg, C. R., and Baird, D. D. (1998). Post-ovulatory ageing of the human oocyte and embryo failure. *Human Reproduction* **13**, 394-397.

Zhou, H.B. and Weinberg, C.R. (1996). Modeling conception as an aggregated Bernoulli outcome with latent variables via the EM algorithm. *Biometrics* **52**, 945-954.

Zhou, H.B., Weinberg, C.R., Wilcox, A.J., and Baird, D.D. (1996). A random-effects model for cycle viability in fertility studies. *Journal of the American Statistical Association* **91**, 1413-1422.

Appendix A

Details for Implementing Auxiliary Variables MCMC Algorithm

Step 1. Sample from the full conditional distribution of $\mathbf{Z}_{ij}$ by setting $\mathbf{Z}_{ij} = \mathbf{0}$ if $Y_{ij} = 0$ and otherwise sampling sequentially from

$$\begin{aligned} \pi(Z_{ij} \mid Y_{ij} = 1, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}) &= \text{Poisson}(\mathbf{X}_{ij}'\boldsymbol{\mu}_{ij}) \text{ truncated so that } Z_{ij} > 0, \\ \pi(\mathbf{Z}_{ij} \mid Z_{ij}, Y_{ij} = 1, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}) &= \text{Multinomial}\Big(Z_{ij}; \frac{X_{ij1}\mu_{ij1}}{\mathbf{X}_{ij}'\boldsymbol{\mu}_{ij}}, \ldots, \frac{X_{ijK}\mu_{ijK}}{\mathbf{X}_{ij}'\boldsymbol{\mu}_{ij}}\Big), \end{aligned}$$

where $Z_{ij} = \sum_k X_{ijk}Z_{ijk}$, $\mu_{ijk} = \xi_i \exp(\mathbf{u}_{ijk}'\boldsymbol{\beta})$, and $\boldsymbol{\mu}_{ij} = (\mu_{ij1}, \ldots, \mu_{ijK})'$.

Step 2. Sample from the full conditional distribution of γ_h, which is

$$\begin{aligned} \pi(\gamma_h \mid \boldsymbol{\gamma}_{(-h)}, \phi, \boldsymbol{\xi}, \mathbf{Z}, \text{data}) &\propto p_h 1_{(\gamma_h=1)} \exp\Big(- \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{k=1}^{K} \prod_{l:l\neq h} \xi_i {\gamma_l}^{u_{ijkl}} \Big) + (1-p_h) 1_{(\gamma_h \in \mathcal{A}_h)} \\ &\times \frac{C(a_h, b_h)}{\int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h)\, d\gamma}\, {\gamma_h}^{a_h + \sum_{i,j,k} u_{ijkh} Z_{ijk} - 1} \exp\Big(- \gamma_h \Big\{ b_h + \sum_{i,j,k} \xi_i \prod_{l:l\neq h} \gamma_l^{u_{ijkl}} \Big\} \Big) \\ &\propto p_h 1_{(\gamma_h=1)} \exp\{-(\widetilde{b}_h - b_h)\} + (1-p_h) \frac{C(a_h, b_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; \widetilde{a}_h, \widetilde{b}_h) d\gamma}{C(\widetilde{a}_h, \widetilde{b}_h \int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h) d\gamma} \mathcal{G}_{\mathcal{A}_h}(\gamma_h; \widetilde{a}_h, \widetilde{b}_h) \\ &\propto \widetilde{p}_h 1_{(\gamma_h=1)} + (1 - \widetilde{p}_h) \mathcal{G}_{\mathcal{A}_h}(\gamma_h; \widetilde{a}_h, \widetilde{b}_h), \end{aligned}$$

where $C(a, b) = b^a/\Gamma(a)$, $\widetilde{a}_h = a_h + \sum_{i,j,k} u_{ijkh} Z_{ijk}$, $\widetilde{b}_h = b_h + \sum_{i,j,k} \xi_i \prod_{l:l\neq h} \gamma_l^{u_{ijkl}}$,

$$\widetilde{p}_h = \frac{p_h \exp\{-(\widetilde{b}_h - b_h)\}}{p_h \exp\{-(\widetilde{b}_h - b_h)\} + (1-p_h) \frac{C(a_h,b_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; \widetilde{a}_h, \widetilde{b}_h)\, d\gamma}{C(\widetilde{a}_h, \widetilde{b}_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h)\, d\gamma}},$$

and $\mathcal{G}_{\mathcal{A}_h}(\cdot; a, b)$ is a $\mathcal{G}(\cdot; a, b)$ gamma density truncated to the region $\mathcal{A}_h \subset \Re^+$. Thus, the conditional posterior distribution of γ_h follows the same form as the prior, being a mixture of a point mass at one and a truncated gamma density.

Step 4. Sample ξ_i, for $i = 1, \ldots, n$, from its full conditional distribution, which is

$$\pi(\xi_i \mid \boldsymbol{\beta}, \phi, \mathbf{Z}, \text{data}) = \mathcal{G}\Big(\xi_i; \phi + \sum_{j,k:X_{ijk}=1} Z_{ijk}, \phi + \sum_{j,k:X_{ijk}=1} \exp(\mathbf{u}_{ijk}'\boldsymbol{\beta})\Big),$$

Step 5. Update ϕ using a Metropolis step.

Step 6. Repeat steps 1-5 until apparent convergence and calculate posterior summaries based on a large number of additional iterations.

Table 1. Summary of simulation study for 100 simulated data sets. Results presented include average of the estimated posterior means, empirical 95% intervals, and coverage of the estimated 95% credible interval (CI).

Parameter	True Value	Average Estimate	95% interval	Coverage of 95% CI
λ_1	0.04	0.04	[0.03, 0.05]	0.92
λ_2	0.14	0.13	[0.11, 0.15]	0.94
λ_3	0.08	0.08	[0.06, 0.09]	0.92
λ_4	0.34	0.33	[0.27, 0.37]	0.90
λ_5	0.31	0.31	[0.25, 0.36]	0.91
λ_6	0.08	0.08	[0.07, 0.09]	0.97
γ_1	1.50	1.66	[1.38, 1.84]	0.98
γ_2	1.50	1.62	[1.36, 1.82]	0.95
γ_3	1.50	1.53	[1.39, 1.63]	0.95
ϕ^{-1}	0.50	0.52	[0.44, 0.60]	0.90

Table 2. Posterior summaries of parameters for the analysis of the European fertility data.

Parameter	Mean	Median	SD	95% Credible Interval
λ_1	0.07	0.07	0.02	[0.03, 0.12]
λ_2	0.12	0.12	0.03	[0.06, 0.20]
λ_3	0.16	0.15	0.05	[0.08, 0.26]
λ_4	0.15	0.14	0.04	[0.08, 0.25]
λ_5	0.14	0.13	0.04	[0.07, 0.22]
λ_6	0.08	0.07	0.03	[0.04, 0.14]
γ_1	1.43	1.32	0.40	[1, 2.47]
γ_2	1.47	1.38	0.38	[1, 2.44]
γ_3	1.53	1.50	0.27	[1.09, 2.15]
ϕ^{-1}	0.94	0.93	0.16	[0.65, 1.28]

Figure 1. Prior and posterior densities for (i) γ_1, (ii) γ_2, and (iii) γ_3. The priors are represented with dashed lines, and the posteriors with solid lines, with the posterior probability of $\gamma_h = 1$ represented with a shaded rectangle.

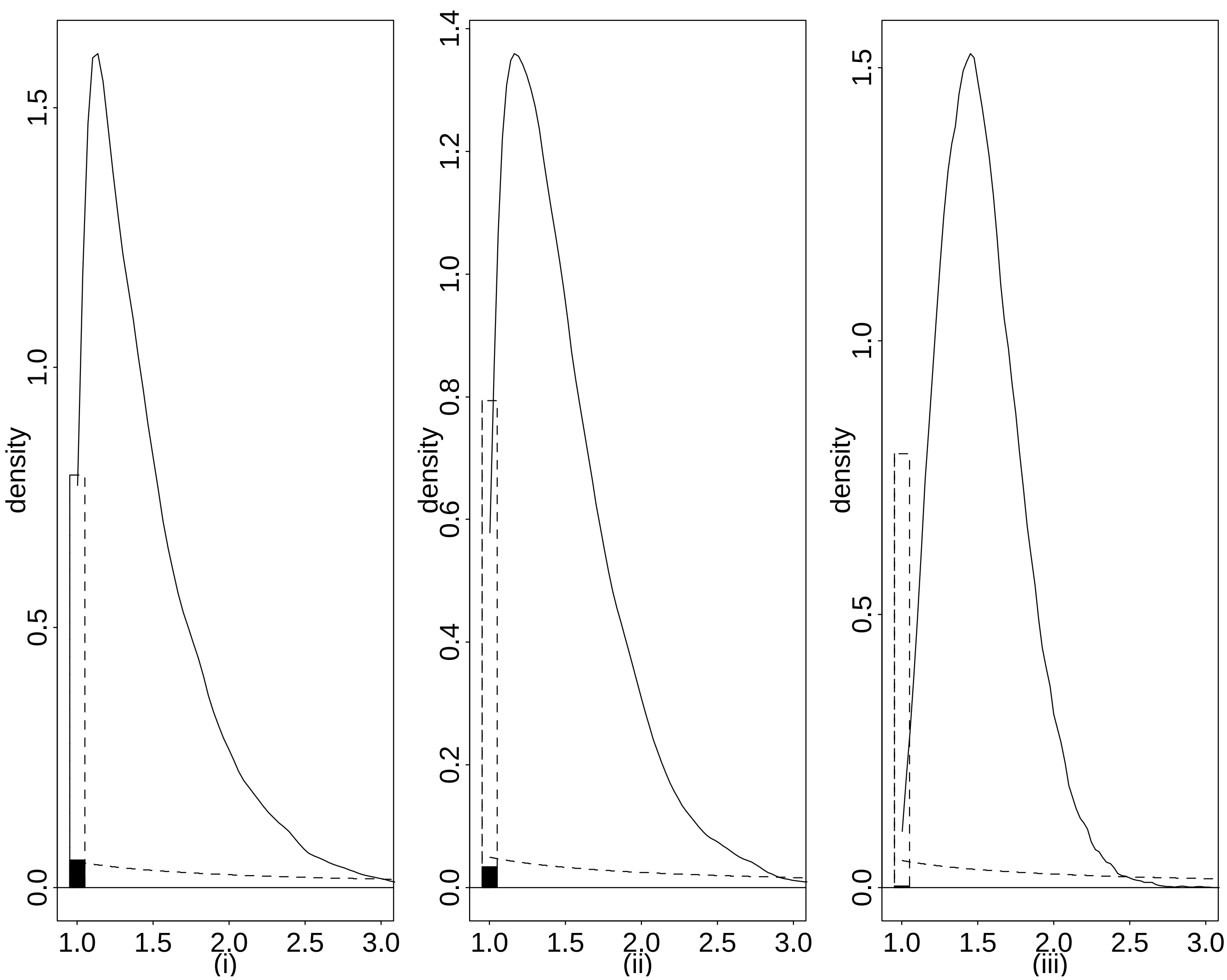

Figure 2. Estimated day-specific probabilities of clinical pregnancy conditional on the mucus score.