

Bayesian Modeling of Markers of Day-Specific Fertility

David B. Dunson^{1,*} and Bernardo Colombo²

¹ Biostatistics Branch, MD A3-03, National Institute of Environmental Health Sciences

P.O. Box 12233, Research Triangle Park, NC 27709, U.S.A

² Department of Statistics, University of Padua, Padua, Italy

* dunson1@niehs.nih.gov

SUMMARY. Cervical mucus hydration increases during the fertile interval prior to ovulation. Since sperm can only penetrate mucus having a high water content, cervical secretions provide a reliable marker of the fertile days of the menstrual cycle. This article develops a Bayesian approach for modeling of daily observations of cervical mucus, and applies the approach to assess heterogeneity among women and cycles from a given woman with respect to the increase in mucus hydration during the fertile interval. The proposed model relates the mucus observations to an underlying normal mucus hydration score, which varies relative to a peak hydration day. Uncertainty in the timing of the peak is accounted for, and a novel weighted mixture model is used to characterize heterogeneity in distinct features of the underlying mean function. Prior information on the mucus hydration trajectory is incorporated, and a Markov chain Monte Carlo approach is developed. Based on data from a study of daily fecundability, there appears to be substantial heterogeneity among women in detected preovulatory increases in mucus hydration but only minimal differences among cycles from a given woman.

KEY WORDS: Fertility awareness; Growth curve; Hierarchical model; Latent variables; Natural family planning; Ovulation method.

1 INTRODUCTION

There is wide interest in predicting the fertile days of the menstrual cycle among couples desiring a pregnancy and those wishing to avoid conception by periodic abstinence. A woman could potentially predict her fertile days on the basis of cycle day alone (c.f., Arévalo, Sinai and Jennings, 2000). However, there is substantial heterogeneity among menstrual cycles in the day of ovulation (Wilcox, Dunson and Baird, 2000), and one can potentially obtain a more accurate prediction by also using daily measurements of basal body temperature, cervical mucus, and other characteristics of the cervix. The widely used Symptothermal Method (Pyper and Knight, 2001) relies on a combination of these indicators, while the Billing's Ovulation Method (Billings, 1983) relies primarily on self-assessment of preovulatory changes in cervical mucus symptoms (e.g., increased hydration). Cervical mucins are associated with motility of human spermatozoa (Eriksen et al., 1998), and sperm can only penetrate mucus having a high water content (Katz, Slade, and Nakajima, 1997). Hence, the occurrence of detectable secretions is potentially an accurate marker of the fertile days (Sinai, Jennings and Arévalo, 1999) and one that relates directly to fertility (Dunson, Sinai and Colombo, 2001). Therefore, it is of substantial interest to assess the magnitude of heterogeneity among women and among cycles from a given woman in the trajectory in detected mucus secretions at different times during an interval of potential fertility defined relative to the time of ovulation.

Ovulation is the key event in the menstrual cycle that determines the fertile interval during which intercourse can potentially result in a pregnancy. Although women typically do not know when they ovulate, one can estimate the cycle day of ovulation retrospectively by using a marker, such as the last day of hypothermia prior to the post-ovulatory rise in basal body temperature, the cervical mucus peak, or the luteinizing hormone surge. Based on studies that collected daily records of intercourse and markers of ovulation for women attempting pregnancy (c.f., Wilcox et al., 1995), the probability that a pregnancy results from a single act of intercourse is small unless intercourse occurs within the interval starting 5 days prior to ovulation and ending on the day of ovulation. Our goal is to study the trajectory in cervical mucus hydration at different times relative to the mucus hydration peak, which is a marker of ovulation day.

Data are drawn from a study of daily fecundability (Colombo and Masarotto, 2000), which collected data prospectively for women enrolled from participating European centers providing fertility awareness services. In this study, women kept daily records of basal body temperature and cervical mucus secretions during one or more menstrual cycles for the purpose of identifying their fertile days. For each cycle day, cervical secretions were categorized on a four point ordinal scale measuring level of mucus hydration. These

measurements ranged from 1=dry to 4=most fertile-type mucus (refer to Colombo and Masarotto, 2000, for a full description of the data). In Table 1, data are provided for 10 menstrual cycles. The categorical observations can be considered as measurements of an underlying continuous mucus hydration score that tends to increase to a peak and then decrease in each cycle. Interest centers on estimating the trajectory through time of the mucus score for a typical cycle from a typical woman and in assessing heterogeneity in trajectories among women and among menstrual cycles. In analyzing these data, we would like to accommodate uncertainty in the peak day, prior knowledge about the shape of the average trajectory (across women) relative to the peak, and heterogeneity among women and menstrual cycles with respect to both the overall level of secretions and the rate of change in secretions close to the peak.

To accommodate these features of the mucus data, we develop a Bayesian hierarchical approach for modeling of nested samples from curves having unknown origins. The mean function for the mucus hydration scores underlying the daily categorical observations is modeled using a flexible mixed-effects model, which is related to self-modeling nonlinear regression (SEMOR) (Lawton, Sylvestre and Maggio, 1972; Lindstrom, 1995; Ke and Wang, 2001). As in SEMOR, our model assumes a common baseline shape for the menstrual cycle-specific response curves. We accommodate systematic variation from the common baseline shape by allowing the timing of a reference point on the curve (e.g., the peak) to vary for the different cycles and by allowing key features of the curve (e.g., the level and flatness) to vary depending on covariates and on woman- and cycle-specific random-effects. To incorporate prior information about the baseline curve, we propose a normal prior density for the second derivative of the curve. This prior is centered on a parametric function representing one’s best guess for the curve. The structure of the prior facilitates efficient posterior computation using a Markov chain Monte Carlo algorithm (MCMC) (Tierney, 1994; Chen, Shao and Ibrahim, 2000).

Although numerous authors have considered semiparametric mixed effects modeling of data sampled from underlying curves (c.f., Brumback and Rice, 1998; Wang, 1998; Zeger and Diggle, 1994; Zhang, Lin, and Sowers, 2000), the incorporation of prior constraints and information regarding the shape of the curve relative to an unknown reference point has received little attention. Our approach to prior elicitation is related to work by Ramgopal, Laud, and Smith (1993) who described a Bayesian approach for incorporating shape restrictions in estimation of potency curves and to work by Ansley, Kohn, and Wong (1993) who proposed a method for incorporating belief that a regression curve is similar in shape to a known parametric curve. The idea of allowing the level and flatness of the curve to vary for different cycles is related to a model

proposed previously for analyzing day-specific conception rates. However, this conception rate model is not applicable to the mucus data and has a fundamentally different structure from the model proposed here.

Section 2 describes the model. Section 3 specifies the prior structure and outlines the MCMC approach for posterior computation. Section 4 applies the approach to data from the Colombo and Masarotto study of daily fecundability, and Section 5 discusses the results.

2 Modeling Daily Cervical Mucus Measurements

2.1 Basic Framework

Suppose that data are collected for N women, with woman i contributing n_i cycles. For cycle j from woman i , let $y_{ij}(t)$ denote an ordered categorical measurement of mucus hydration on day t , with $t = 0$ corresponding to the last day of hypothermia (LDH), and let r_{ij} denote the day of the mucus hydration peak (MHP) relative to the LDH. If $r_{ij} = 0$ then the MHP occurs on the same day as the LDH, which is used as a marker of ovulation day. Measurement error in the LDH marker and variability in the timing of the MHP relative to ovulation can cause $r_{ij} \neq 0$. Ecochard et al. (2001) compared cervical mucus peak and basal body temperature markers of ovulation to direct measures of ovulation obtained using transvaginal ultrasonography for 283 normal menstrual cycles. Motivated by data from their study and from an earlier study by Templeton, Penney and Lees (1982), we assume that $r_{ij} \in [-5, -4, \dots, 4, 5]$ for all i, j .

We link the mucus measurement $y_{ij}(t)$ to r_{ij} and to an underlying continuous mucus hydration score $y_{ij}^*(t - r_{ij})$ through

$$y_{ij}(t) = \{c : \tau_{c-1} < y_{ij}^*(t - r_{ij}) \leq \tau_c, c = 1, \dots, d\}, \quad (1)$$

where $y_{ij}^*(s) \sim N(\mu_{ij}(s), 1)$, s denotes day relative to the MHP, d denotes the number of categories of $y_{ij}(t)$, and $\boldsymbol{\tau} = (\tau_0, \dots, \tau_d)'$ are threshold parameters with $\tau_0 = -\infty$, $0 = \tau_1 < \tau_2 < \dots < \tau_{d-1}$, $\tau_d = \infty$. It follows that the likelihood of $y_{ij}(t)$ is

$$\int \prod_{c=1}^d \left[\Phi\{\tau_c - \mu_{ij}(t - r)\} - \Phi\{\tau_{c-1} - \mu_{ij}(t - r)\} \right]^{1(y_{ij}(t)=c)} dF_{R_{ij}}(r), \quad (2)$$

where $\Phi(z)$ is the standard normal c.d.f., $F_{R_{ij}}$ is the c.d.f. for the day of the mucus hydration peak in cycle i, j relative to the LDH, and $\mu_{ij}(s)$ is an underlying mean function which depends on cycle and on the timing relative to the mucus hydration peak (MHP). Likelihood (2) is a mixture of a multinomial density across the distribution of the unknown MHP reference point.

2.2 Model for the Underlying Mean

Our primary interest is in assessing heterogeneity among women and cycles from a given woman in the underlying mean function μ_{ij} , which defines the expected mucus hydration score at a given time relative to the MHP. To characterize heterogeneity in μ_{ij} , we propose the following weighted mixed-effects model:

$$\mu_{ij}(s) = \alpha(s) + \mathbf{g}\{s; \alpha(s)\} \left(\mathbf{\Gamma} \mathbf{x}_{ij} + \mathbf{\Lambda}_1 \boldsymbol{\eta}_{i1} + \mathbf{\Lambda}_2 \boldsymbol{\eta}_{ij2} \right) \quad \text{for } s \in [s_1, s_M], \quad (3)$$

where $\alpha(s)$ is the expectation of $\mu_{ij}(s)$ across the population of women and cycles i, j with $\mathbf{x}_{ij} = \mathbf{0}$, $\boldsymbol{\eta}_{i1} = \mathbf{0}$, and $\boldsymbol{\eta}_{ij2} = \mathbf{0}$; \mathbf{x}_{ij} is a vector of woman- and cycle-specific covariates;

$$\mathbf{g}\{s; \alpha(s)\} = [1, g_1\{s; \alpha(s)\}, \dots, g_{q-1}\{s; \alpha(s)\}]$$

is a $1 \times q$ vector of known smooth functions of s and $\alpha(s)$; $\mathbf{\Gamma}$ are unknown regression constants; $\boldsymbol{\eta}_{i1} \sim N_q(\mathbf{0}, \mathbf{I}_{q \times q})$ are latent variables measuring deviations for woman i from the typical trajectory in the underlying mucus hydration score; $\boldsymbol{\eta}_{ij2} \sim N_q(\mathbf{0}, \mathbf{I}_{q \times q})$ are latent variables measuring deviations for cycle j from the average trajectory for woman i ; and $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2$ are $q \times q$ factor loadings matrices with certain elements fixed at 0 or 1 for purposes of identifiability. We assume that the latent variables $\boldsymbol{\eta}_{i1}$ and $\boldsymbol{\eta}_{i2} = (\boldsymbol{\eta}_{ij2}, j = 1, \dots, n_i)$ are independent and that the elements of \mathbf{y}_i^* are conditionally independent given $\boldsymbol{\eta}_{i1}$ and $\boldsymbol{\eta}_{i2}$. Expressions (1) - (3) correspond to a generalized probit mixed model for the original polytomous mucus score (c.f., Harville and Mee, 1984).

Scientific interest focuses on the trajectory in the mucus hydration score during a window of potential fertility defined relative the unknown MHP. Previous studies have shown that the fertile days of the menstrual cycle typically begin 5 days prior to ovulation and end on the day of ovulation (Dunson et al., 1999). Although the timing of ovulation is highly variable (Wilcox, Dunson and Baird, 2000), studies have suggested that the fertile interval relative to ovulation is quite stable for women in their reproductive years with no history of infertility (Dunson, Colombo and Baird, 2002). As a conservative estimate of the fertile interval relative to the MHP, we focus on $[s_1, s_M] = [-6, 3]$. Assuming the MHP occurs within ± 5 days of the LDH, we use data on the daily mucus scores in the $[-11, 8]$ interval relative to the LDH in fitting the model. For intervals that are much wider than $[-6, 3]$, one or more days in $[s_1, s_M]$ may fall outside of the cycle, particularly if the follicular or luteal phase lengths are relatively short (the follicular phase begins with menstruation and ends with ovulation, while the luteal phase begins with ovulation and ends with the start of the next cycle). To avoid this problem, one could potentially define $\mu_{ij}(s)$ only for $s \in [s_{ij1}, s_{ijM}] \subset [s_1, s_M]$, where s_{ij1} and s_{ijM} are chosen so that s does not fall outside cycle i, j . Under such an approach, the expected mucus

hydration score on a given day relative to the MHP is interpretable as an average across menstrual cycles containing that day.

In addition to varying the time window for cycles of different lengths one could potentially adjust the time scale. For example, one may hypothesize that the mucus trajectory varies systematically according to the length of the cycle; possibly with the effect of time s in expression (3) inversely proportional to the cycle length. However, such time scale adjustments would lead to complications in interpretation and are not supported by the Colombo and Masarotto (2000) data, based on exploratory analyses. In addition, the cycle length is not available for conception cycles. As a more flexible and interpretable alternative to a time scale adjustment, one could include the follicular phase length as a covariate in \mathbf{x}_{ij} . Unlike the cycle length, the follicular phase length is available for conception and non-conception cycles. In addition, the majority of the variability in cycle length is attributable to variability in the follicular phase length, with the luteal phase length having relatively low variance.

The deviation functions g_1, \dots, g_{q-1} in expression (3) are chosen so that $\int_{s_1}^{s_M} g_b\{s; \alpha(s)\} ds = 0$ for $b = 1, \dots, q-1$. Under this structure, the parameters in the first rows of $\mathbf{\Gamma}$, $\mathbf{\Lambda}_1$, and $\mathbf{\Lambda}_2$ characterize differences in the average level of the trajectory within the interval $[s_1, s_M]$ associated with observed covariates, unobserved woman-specific factors, and unobserved cycle-specific factors, respectively. The parameters in rows 2 through q characterize heterogeneity in different aspects of the shape defined by the deviation functions g_1, \dots, g_{q-1} . A simple and widely useful special case chooses

$$\mathbf{g}\{s; \alpha(s)\} = \{1, \alpha(s) - \bar{\alpha}\}, \quad (4)$$

where $\bar{\alpha} = [\int_{s_1}^{s_M} \alpha(w) dw] / (s_M - s_1)$. In this case, expression (3) allows the peak to be more pronounced for some cycles (and women) and less pronounced for others. If, for example, the baseline trajectory $\alpha(s)$ increases monotonically to a peak and then decreases, then cycles with $\mathbf{\Gamma}_2 \mathbf{x}_{ij} + \mathbf{\Lambda}_{12} \boldsymbol{\eta}_{i1} + \mathbf{\Lambda}_{21} \boldsymbol{\eta}_{ij2} > 0$ ($\mathbf{\Gamma}_2$, $\mathbf{\Lambda}_{12}$, $\mathbf{\Lambda}_{21}$ are the second rows of $\mathbf{\Gamma}$, $\mathbf{\Lambda}_1$, and $\mathbf{\Lambda}_2$, respectively) will have a higher peak and a more rapid decrease away from the peak than cycles with this term equal to zero and with the same average level. Since fertility is highest during the days just prior to the MHP, women with positive values for $\mathbf{\Gamma}_2 \mathbf{x}_{ij} + \mathbf{\Lambda}_{12} \boldsymbol{\eta}_{i1}$ should be better at predicting their fertile days based on observations of mucus hydration than other women who have a less pronounced increase in hydration. The utility of this structure is illustrated in Section 4.

Expression (3) is related to the model proposed by Barry (1995) for growth curve data. Barry's model accounts for nonparametric subject- and time-specific deviations from an average curve under a Gaussian hierarchical structure that penalizes large deviations and rapid changes over time. Expression (3) instead

accommodates deviations from a baseline curve with respect to the overall level and different aspects of the shape through a weighted factor analytic term. In addition to unexplained heterogeneity among subjects, we allow for general covariate effects and heterogeneity among different curves from a given subject.

2.3 Model for the Location of the Reference Point

Within the framework of expressions (1)-(2) it is necessary to specify models for both the underlying mean function μ_{ij} , which characterizes the distribution of the mucus hydration score at a given time relative to the mucus hydration peak, and $F_{R_{ij}}$, the c.d.f. for the timing of the MHP relative to the LDH. Since the MHP and the LDH are integers representing days of the menstrual cycle, the difference in these two quantities must be an integer. In addition, as discussed in Section 2.1, it is reasonable to assume that this difference is bounded (i.e., $R_{ij} \in [-5, \dots, 5]$ for all i, j). Adding the assumption that the distribution of the difference is the same for all women, we have

$$\Pr(R_{ij} = t_l) = F_{R_{ij}}(t_l) - F_{R_{ij}}(t_{l-1}) = \nu_l, \quad \text{for } l = 1, \dots, L, \quad (5)$$

where $\mathbf{t} = (-5, \dots, 5)'$ denotes the vector of possible values for R_{ij} , $F_{R_{ij}}(t_0) = 0$, $\sum_{l=-5}^5 \nu_l = 1$, and $L = 11$. The assumption that the distribution of the MHP reference point is the same for all women could be relaxed by specifying a regression model for ν_l , which accommodates covariates and woman-specific random effects.

3 Bayesian Formulation

3.1 Prior Specification

We complete a Bayesian specification of our model with prior distributions for $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\nu})$ and for the baseline trajectory α . Conditionally conjugate priors are chosen for each of the parameters in $\boldsymbol{\theta}$, with uniform priors for the threshold parameters $\boldsymbol{\tau}$, normal priors for the regression parameters $\boldsymbol{\Gamma}$ and factor loadings $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$, and a Dirichlet prior for the parameters $\boldsymbol{\nu}$ characterizing the distribution for the timing of the MHP relative to the LDH. To enforce the strict order constraint on the elements of $\boldsymbol{\tau}$, we define the uniform prior to have constrained support: $\pi(\boldsymbol{\tau}) \propto 1(\boldsymbol{\tau} \in \Omega)$, where $\Omega = \{\boldsymbol{\tau} : 0 = \tau_1 < \tau_2 < \dots < \tau_{d-1}\}$. In choosing a prior for the baseline trajectory α , it is necessary for identifiability of α and $\boldsymbol{\nu}$ to restrict the timing of the peak. For this reason, we take $s = 0$ to correspond to the MHP.

From known mechanisms of cervical mucus production in the menstrual cycle and from previous

studies that measured cervical mucus secretions (c.f., Fehring, 2002), information is available on the approximate shape of the trajectory in hydration relative to the peak day. Cervical mucus is secreted from cells in the endo-cervix, at least partially in response to circulating estrogen and progesterone levels. Estrogen produced by the developing follicle stimulates the production of mucus. As folliculogenesis proceeds, mucus becomes clearer and more hydrated, with the water content peaking at the time of highest estrogen levels and peak development of the follicle. After rupture of the follicle (i.e., ovulation), estrogen levels decrease, progesterone levels increase, and the hydration level of cervical mucus drops dramatically. Hence, we expect a gradual increase in the mucus hydration score during the follicular phase prior to the peak day and a rapid decrease thereafter.

Our best guess for the baseline trajectory α can be expressed via the parametric function:

$$a_0(s; \mathbf{b}) = b_1 \exp\left[-s^2\{1(s < 0)b_2 + 1(s > 0)b_3\}\right] \quad \text{for } s \in [-6, 3], \quad (6)$$

where we choose $\mathbf{b} = (b_1, b_2, b_3) = (3, 0.05, 0.5)$. There is no standard parametric model for mucus hydration levels in the menstrual cycle. We chose (6) since it (i) follows an umbrella pattern with a single peak at $s = 0$; (ii) has a finite second derivative for all $s \in [-6, 3]$; (iii) can allow changes to occur more rapidly after the peak; and (iv) has only 3 parameters to be elicited. Although we are confident that the mucus hydration trajectory follows a similar pattern to that described by (6), we doubt that the parametric form is exactly appropriate and are somewhat uncertain about how rapidly the curve changes prior to and after the peak. In addition, we have a moderate degree of uncertainty about the overall level of the curve. Our prior for α is motivated by this uncertainty and by computational considerations.

Let A denote a space of curves following a known restriction. In particular, we focus on the case where $A = \{\alpha : \alpha(0) = \sup_{s \in [s_1, s_M]} \alpha(s)\}$. For this choice of A , a prior that assigns $\Pr(\alpha \in A) = 1$ will restrict α to peak at $s = 0$ with probability 1 a posteriori. In order to assign high prior probability to curves that are similar in shape to a_0 , we choose a density that penalizes (i.e., assigns low probability) to large differences between $\alpha''(s)$ and $a_0''(s)$, for $s \in [s_1, s_M]$. The idea of incorporating prior information on a differential equation representation of a curve is related to the l-spline approach described by Ansley, Kohn and Wong (1993) (see also, Wahba, 1990 and Heckman and Ramsey, 2000). The l-spline approach incorporates a differential equation associated with the prior information in the roughness penalty. The regression curve can then be estimated by penalized least squares.

We instead follow a fully Bayesian approach to estimation and inference, with $\alpha''(s) \sim N(a_0''(s), \rho^{-1})$, for $s \in (s_1, s_M)$, where ρ is an investigator-specified precision parameter quantifying prior evidence of

similarity between the shapes of α and a_0 . The choice to use a normal prior density for the second derivative was motivated by the nature of the available prior information and by computational considerations. In particular, the normal prior has appealing conjugacy properties, which are described in Section 3.2. In addition to the prior for the second derivative of $\alpha(s)$ in the interior of $[s_1, s_M]$, we specify normal prior densities for the level of α on the boundaries and we restrict $\alpha \in A$. Our prior for the baseline curve can be summarized as follows:

$$\alpha(s_1) \sim N(\mu_{\alpha_1}, \sigma_{\alpha_1}^2), \quad \alpha''(s) \sim N(a_0''(s), \rho^{-1}) \text{ for } s \in (s_1, s_M) \quad \text{and} \quad \alpha(s_M) \sim N(\mu_{\alpha_M}, \sigma_{\alpha_M}^2), \quad (7)$$

subject to the constraint that $\alpha \in A$. Relatively vague priors can be chosen for the level of the curve by choosing moderate prior variance for $\alpha(s_1)$ and $\alpha(s_M)$. In addition, a moderately informative prior can be chosen for the shape by choosing a low value for ρ^{-1} .

In practice, we do not estimate α at every possible value of $s \in [s_1, s_M]$, but instead rely on a finite approximation to α on a tightly-spaced grid. Let $\mathbf{s} = (s_1, s_2, \dots, s_M)$ denote a grid of equally-spaced times with $s = 0$ corresponding to the reference time and the spacing $s_m - s_{m-1}$ equal to the spacing of the design points $t_l - t_{l-1}$. The data are potentially informative about $\boldsymbol{\alpha} = (\alpha(s_1), \dots, \alpha(s_M))^T$, the values of α on \mathbf{s} , and prior (7) facilitates extrapolation to a finer scale. We will estimate the values of α on the grid $(s_1, s_1 + \delta, \dots, 0, \dots, s_M - \delta, s_M)$ for δ small, with a numerical approximation to $\alpha''(s)$ used in place of the unknown $\alpha''(s)$ in expression (7). Details involved in estimation of a finite approximation to $\alpha(s)$ are provided in an Appendix.

3.2 Posterior Computation

An MCMC algorithm (Tierney, 1994; Chen, Shao and Ibrahim, 2000) is used to obtain posterior summaries of the parameters $\boldsymbol{\theta}$, a finite approximation to the baseline trajectory α , and the latent variables $\{\boldsymbol{\eta}_{i1}, \boldsymbol{\eta}_{ij2}, R_{ij}\}$. Since the number of parameters (counting the latent variables and each element of $\boldsymbol{\alpha}$) is potentially large, efficient posterior computation is a challenging problem. In particular, black box approaches, such as updating the parameters one at a time using Metropolis random walk (Metropolis et al., 1953) and adaptive rejection sampling (Gilks and Wild, 1992) steps, tend to be so inefficient (in terms of number of iterations needed to obtain accurate posterior summaries) that they are practically infeasible. A major advantage of the particular modeling structure and prior formulation chosen in this article is the availability of closed form full conditional densities for subvectors of the parameters. For example, for choices of g that are linear in α (as is the case for the form shown in expression 4), we can sample directly from the full conditional

distributions of $\Gamma, \Lambda_1, \Lambda_2, \boldsymbol{\eta}_{i1}, \boldsymbol{\eta}_{ij2}$ and each of the elements of $\boldsymbol{\alpha}$. The full conditional distributions of R_{ij} and $\boldsymbol{\nu}$ also have simple closed forms. The necessary conditional distributions are outlined in an Appendix, and the MCMC algorithm proceeds as follows:

1. Start with initial values for the parameters and latent variables.
2. Replace $\boldsymbol{\tau}$ by draws from (A.1) with probability (A.2).
3. Impute the underlying data $\mathbf{y}_i^*, i = 1, \dots, n$, by sampling from (A.3).
4. Replace $\Gamma, \Lambda_1, \Lambda_2, \{\boldsymbol{\eta}_{i1}\}$, and $\{\boldsymbol{\eta}_{ij2}\}$ by draws from (A.4)-(A.7).
5. Replace $\{R_{ij}\}$ by draws from (A.8).
6. Replace $\boldsymbol{\nu}$ by draws from (A.9).
7. Replace $\boldsymbol{\alpha}$ by draws from (A.10).
8. Repeat steps 2-6 until the process is judged to have converged, and compute posterior summaries of the quantities of interest based on a large number of additional iterates.

This algorithm is a hybrid procedure involving both Gibbs Sampling (Gelfand and Smith, 1990) and Metropolis steps. Under mild regularity conditions, samples will converge to a target distribution that is the joint posterior distribution.

As is the case for Gibbs sampling algorithms for posterior computation in standard generalized linear mixed models, the efficiency of our MCMC algorithm tends to depend on the choice of prior, and in particular the prior variance. When a moderately informative prior is chosen for the baseline curve α and proper priors are chosen for the factor loadings Λ_1, Λ_2 , the algorithm performs well in the examples we have considered, with low to moderate autocorrelation in the samples. However, when the prior precision $\rho \approx 0$, autocorrelation can be high and slow-mixing is a concern (as would be the case in mixed models with a diffuse prior for the random-effect variance and the intercept). In the next section, the methodology is applied to the Colombo and Masarotto data.

4 Multinational Study of Daily Fecundability

4.1 The Data

We analyzed data on daily observations of cervical mucus hydration for 680 menstrual cycles from 93 women in the Paris center of the Colombo and Masarotto (2000) study. Women enrolled were between 18 and 40 years of age, had at least one menses after the most recent cessation of breastfeeding or delivery, and were not currently taking hormonal medication or drugs affecting fertility. Most participating couples were trying to avoid pregnancy by using natural family planning methods. Individuals with a history of infertility were excluded, as were couples in the habit of mixing unprotected and protected intercourse. The original Paris sample contained data for 787 menstrual cycles and 104 women, which was reduced to 680 cycles and 93 women after excluding cycles with no identified last day of hypothermia (LDH) reference day (e.g., due to missing basal body temperature measurements). The average age in years of the women and men in the Paris center was 29.3 (sd = 4.5) and 31.4 (sd = 5.4), respectively. Out of 104 women, 76 (73.1%) had at least one previous pregnancy and 38 (36.5%) had used hormonal contraception in the past. In addition, out of the 787 menstrual cycles, there were 63 clinically-detected conceptions and 5 miscarriages. The structure of the daily mucus data is described in Section 1 and Table 1, and additional details are available in the Colombo and Masarotto article (www.demographic-research.org/Volumes/Vol3/5/).

4.2 The Analysis

Our focus was on (i) estimating the typical trajectory of mucus secretions in the $[-6, 3]$ interval with origin at the unknown MHP; (ii) estimating the distribution of the MHP relative to the LDH proxy for day of ovulation; and (iii) quantifying heterogeneity among women and menstrual cycles from a given woman in the average mucus hydration score and the steepness of the trajectory in the mucus hydration score around the MHP. Our analysis is based on the model described in Section 2 with g following the form shown in expression (4) and with the underlying mean function equal to

$$\mu_{ij}(s) = \alpha(s) + \lambda_1\eta_{i11} + \lambda_3\eta_{ij21} + \{\alpha(s) - \bar{\alpha}\}(\lambda_2\eta_{i12} + \lambda_4\eta_{ij22}) \text{ for } s \in [-6, 3], \quad (8)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ are constrained to be positive for identifiability; $\lambda_1\eta_{i11}$ and $\lambda_3\eta_{ij21}$ measure deviations for woman i and cycle j from woman i , respectively, with respect to the average cervical mucus hydration score; and $\lambda_2\eta_{i12}$ and $\lambda_4\eta_{ij22}$ quantify deviations for woman i and cycle j with respect to the steepness of the trajectory in the mucus hydration score.

We also considered models that incorporated age as a covariate, but found no evidence of an association between age and the mucus hydration score. As discussed in Section 2.2, the follicular phase length is another potentially important covariate. However, we found no evidence of a relationship between the mucus trajectory and the follicular phase length in exploratory analyses. In future work, we plan to consider other covariates, including pregnancy history and occurrence of conception.

As justified in Section 3.1, to quantify our prior belief about the mucus hydration trajectory relative to the MHP in a typical cycle from a typical woman, we chose the following parametric function as the prior mean for α :

$$a_0(s) = 3 \exp\left[-s^2\{1(s < 0).05 + 1(s > 0).5\}\right] \quad \text{for } s \in [-6, 3].$$

To quantify our uncertainty in this choice, we set $\sigma_{\alpha_1}^2 = 2$, $\rho = .25$ and $\sigma_{\alpha_M}^2 = 2$. This prior allows a moderate to high degree of uncertainty in the baseline trajectory under the restriction that $\alpha(s)$ has its maximum at $s = 0$. Proper priors are chosen for the remaining parameters as follows: $\lambda_1, \lambda_2, \lambda_3$ and λ_4 iid $N(.5, 2)$ truncated below by 0, $\eta_{i11}, \eta_{i12}, \eta_{ij21}$ and η_{ij22} iid $N(0, 1)$ (the variance is fixed at one for identifiability), $\nu(t) = 0$ for $t \notin [-5, 5]$ and $(\nu(-5), \dots, \nu(5)) \sim \text{Dirichlet}(1, \dots, 1)$. The prior densities for the factor loadings parameters, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, were chosen to be somewhat flat within a range of plausible values. We avoid choosing high prior variance, since such priors assign high probability to values that are known to be implausible and additionally can result in slow-mixing of the MCMC algorithm. In implementing the analysis, a discrete approximation is used as described in subsection 3.1 with $\delta = .1$.

4.3 The Results

We used the MCMC algorithm outlined in subsection 3.2 to estimate posterior summaries of the quantities of interest. Briefly, initial values were chosen based on summary statistics for the data and the algorithm was run for 50,000 iterations after discarding the initial 5,000 samples as a burn-in. We assessed convergence by examining plots of the sampled parameters and by using the suite of diagnostics described by Cowles and Carlin (1996), implemented in S-PLUS with the BOA functions (Smith, 2000). All of the parameters appeared to converge by iteration 5,000, with $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ and the baseline parameters α having the slowest convergence and mixing rates. In particular, the lag-10 autocorrelation was 0.28, 0.16, 0.13, and 0.25 for $\lambda_1, \lambda_2, \lambda_3$ and λ_4 , respectively. The posterior correlation among the λ parameters was low, ranging from -0.06 for (λ_2, λ_4) to 0.01 for (λ_2, λ_3) . In addition, the posterior correlation among the α parameters was positive, with the correlation tending to decrease with increasing time separation. For example, the correlation coef-

ficients between $\alpha(0)$ and the individual elements of $[\alpha(-6), \alpha(-5), \alpha(-4), \alpha(-3), \alpha(-2), \alpha(-1), \alpha(1), \alpha(2)]$ were 0.00, 0.04, 0.14, 0.36, 0.54, 0.75, 0.72, 0.38, and 0.07, respectively.

The estimated baseline trajectory α is plotted in Figure 1 along with 95% pointwise credible intervals and the prior mean a_0 . The trajectory in the cervical mucus hydration scores for a typical cycle from a typical woman was slightly flatter than anticipated and the level was somewhat lower, though we had a low degree of confidence in our prior mean a_0 . In addition to estimates of the baseline trajectory, we obtained estimates of the distribution of the mucus hydration peak relative to the last day of hypothermia proxy for ovulation (shown in Figure 2). The first estimate represents the posterior means from our analysis with 95% pointwise credible intervals also shown. The second estimate was based on the traditional approach to estimating the mucus peak, which sets the peak day equal to the last day in a cycle with the most fertile type mucus detected in that cycle (Colombo and Masarotto, 2000).

The traditional estimate of the mucus peak was clearly sensitive to missing data, while our approach can easily accommodate missing observations under the missing-at-random assumption. To clarify, the traditional estimate uses only the data available in a given cycle in estimating the peak, and the peak cannot be assigned to a day with missing data. In contrast, our approach uses information on the average trajectory observed in cycles from other women, along with available data on the characteristics of the woman and the cycle. In addition, the traditional estimate provides no measure of uncertainty and may have a tendency to estimate a peak occurring after the true peak. Our estimate of the distribution was shifted to the left relative to the traditional estimate and had a slightly higher maximum.

The primary goal of our analysis was to quantify heterogeneity in the probability of detected most fertile-type mucus on different days of the cycle relative to the mucus hydration peak, where most fertile-type mucus ($y = 4$) has a high degree of hydration and is characterized by a wet, slippery or smooth feeling with transparent, stretchy and liquid mucus. Figure 3 shows the estimated day-specific probabilities of most fertile-type mucus for women and cycles in different percentiles with respect to the latent factors η_{i11} , η_{i12} , η_{ij21} , and η_{ij22} , respectively, with the latent factors not under consideration set to 0 in each case. Based on this plot and on the posterior summaries presented in Table 2, the majority of heterogeneity can be attributed to differences among women and not to differences between cycles from a given woman. In particular, there were moderate differences among women in the average probability of most fertile-type mucus (refer to Figure 3i) and substantial differences among women in the relationship between time relative to mucus hydration peak and the probability of most fertile-type mucus (refer to Figure 3ii).

For some women most fertile-type mucus was much more likely to occur during the several days leading up to the mucus peak than on days at the edge of the [-6,3] interval. For other women the probability of most fertile type mucus was relatively constant across the [-6,3] interval. Since the day-specific probabilities of pregnancy are highest during the several days leading up to the mucus peak (Colombo and Masarotto, 2000), women in the former group should be better than those in the latter group at predicting their fertile days based on cervical mucus. These conclusions have important implications for users and providers of natural family planning programs that use cervical mucus symptoms to identify the fertile days of the cycle.

4.4 Sensitivity Analyses

It is important to assess the sensitivity of our results to the prior formulation for α and to the choice of prior mean a_0 and prior precision ρ . In particular, we repeated the analysis using independent $N(0, 4)$ prior distributions for $\alpha(-6), \alpha(-5), \dots, \alpha(2), \alpha(3)$, subject to the constraint that $\alpha(0) = \max_{s \in [-6, -5, \dots, 3]} \alpha(s)$. The resulting posterior means (sds) for $\tau_1, \tau_2, \lambda_1, \lambda_2, \lambda_3$ and λ_4 were 0.236 (0.016), 0.992 (0.030), 0.157 (0.026), 0.471 (0.051), 0.032 (0.023) and 0.056 (0.040), respectively. These estimates are similar to those reported in Table 2. In addition, the estimated trajectory in the mucus hydration score relative to the LDH was similar to that produced in the main analysis. As expected, the estimates for α and ν were less robust. The estimated baseline curve had a similar shape and level to that seen in the primary analysis. However, the distribution of the difference between the LDH and MHP peaked at -1 instead of 0 (as in Figure 2) and the estimate of α was shifted by 1 day in the positive direction to compensate for this difference in ν . These differences with the results of the main analysis were not unanticipated due to weak identifiability of the timing of the MHP relative to the LDH. This weak identifiability does not reflect lack of fit of the model but instead limited information in the current data set about the distribution of the MHP relative to the LDH. Had continuous or more detailed mucus score data been available (as is previous smaller studies - e.g., Fehring, 2002) this distribution would have been better identified. Since we have good prior information that the mucus score declines dramatically after the peak day (refer to Section 3.1), it seems reasonable to include this information in estimating α and ν . The question then becomes: are the estimates robust to reasonable changes in the prior for α that are consistent with the available information?

To address this question, we repeated the analysis for alternative choices of a_0 and ρ . We first chose $\rho = 0.15$ instead of $\rho = 0.25$ to assess the effect of increasing the prior variance. The resulting posterior means (sds) for $\tau_1, \tau_2, \lambda_1, \lambda_2, \lambda_3$ and λ_4 were 0.236 (0.014), 1.033 (0.022), 0.158 (0.025), 0.408 (0.047), 0.036

(0.025) and 0.078 (0.052), respectively. These values were similar to those reported in Table 2. The estimates for α and ν were also similar to those obtained in the main analysis, with the largest differences in the peak values, ν_6 (0.35 vs 0.28) and $\alpha(0)$ (1.93 vs 2.20). Finally, we repeated the analysis for a flatter prior mean, with a_0 following the parametric form shown in expression (6), with the parameters $\mathbf{b} = (2.142, 0.044, 0.462)$ chosen to minimize the difference between a_0 and the posterior mean for α from the main analysis. The posterior means (sds) for the parameters in table 2 were 0.238 (0.014), 1.041 (0.023), 0.160 (0.025), 0.368 (0.050), 0.034 (0.024) and 0.082 (0.057), respectively. The estimate for ν was similar overall to the estimate in Figure 2, with the main differences in ν_5 (0.21 vs 0.27) and ν_7 (0.14 vs 0.08). The estimated curve α was similar to the estimate in Figure 1, except in the $[-2,1]$ interval where the curve had a flatter pattern with the largest difference at the peak (1.77 vs 2.20). Hence, although inferences on the magnitude and sources of heterogeneity in the mucus trajectory are quite robust, estimates for ν and α are moderately sensitive to the flatness of the prior mean.

5 Discussion

This article developed a Bayesian approach for modeling of daily categorical observations of cervical mucus hydration in the menstrual cycle, and applied this approach to assess heterogeneity in the level and pattern of days with most fertile type mucus. A distinguishing feature of the model is its flexibility in accounting for heterogeneity in the mucus hydration score trajectory. One characteristic that is allowed to vary from cycle to cycle is the timing of the mucus hydration peak relative to the last day of hypothermia proxy for ovulation day. The approach of allowing for uncertainty in the timing of a distinct feature of a curve can potentially be used in other applications in which patterns in the data are driven by the timing relative to an unobserved event (e.g., ovulation in studies of fertility symptoms in the menstrual cycle). In such cases the distribution of the reference point may depend on both the magnitude of measurement error in a surrogate (from which the data are indexed) and differences between the timing of the reference point and the unobserved event. In addition to allowing uncertainty in the timing of the peak, our model accommodates heterogeneity in the mucus hydration score trajectory through a general weighted mixed effects model. This model incorporates a flexible baseline component and allows variation from this baseline for individual women and cycles within a woman. As was illustrated, this modeling structure allows investigators to assess heterogeneity in biologically important characteristics of the trajectory.

Based on data from the Paris center of the Colombo and Masarotto (2000) study, it appears that

women vary considerably in the timing of occurrence of days with most fertile-type mucus. Mucus hydration is typically high during the fertile days prior to ovulation and sperm are believed to require a high degree of mucus hydration to navigate the female reproductive tract. Therefore, the occurrence of most fertile-type mucus is a marker of the most fertile days of the cycle. For many women, most fertile type mucus seems to be present primarily on the several days leading up to the mucus hydration peak and this occurs consistently for different menstrual cycles. However, for other women there is a much less distinct increase in the probability of most fertile type mucus. This may be due to inadequate training in distinguishing between levels of hydration or to real biological differences in levels of cervical secretions.

We are currently considering applications to a joint analysis of data from the different centers in the Colombo and Masarotto (2000) study. To accommodate possible systematic differences between study centers in measuring the mucus hydration scores, the model can be extended to incorporate center-specific random effects. In the multi-center analysis, we plan to evaluate different covariates. We are also exploring a generalization of the approach to multivariate data on mucus characteristics collected in a recent study of users of the Creighton Model Fertility System. Of ongoing interest is the link between the mucus trajectory in the menstrual cycle and a woman's fertility. This link could potentially be explored by incorporating latent woman-specific factors underlying the mucus trajectory in a model for the day-specific probabilities of conception.

APPENDIX: DETAILS FOR MCMC IMPLEMENTATION

Step 2: Updating the threshold parameters τ

Sample a candidate $\tilde{\tau}$ for τ by letting $\tilde{\tau}_0 = -\infty$, $\tilde{\tau}_1 = 0$,

$$\log(\tilde{\tau}_2), \log(\tilde{\tau}_3 - \tilde{\tau}_2), \dots, \log(\tilde{\tau}_{d-1} - \tilde{\tau}_{d-2}) \sim N\left(\log(\tau_2), \dots, \log(\tau_{d-1} - \tau_{d-2}), \kappa_\tau \mathbf{I}\right), \quad (A.1)$$

and $\tilde{\tau}_d = \infty$. Accept the candidate with Metropolis et al. (1953) probability

$$\min\left(1, \frac{\prod_{i,j,k} \prod_{c=1}^d [\Phi\{\tilde{\tau}_c - \mu_{ij}(t_{ijk} - R_{ij})\} - \Phi\{\tilde{\tau}_{c-1} - \mu_{ij}(t_{ijk} - R_{ij})\}]^{1(y_{ijk}=c)}}{\prod_{i,j,k} \prod_{c=1}^d [\Phi\{\tau_c - \mu_{ij}(t_{ijk} - R_{ij})\} - \Phi\{\tau_{c-1} - \mu_{ij}(t_{ijk} - R_{ij})\}]^{1(y_{ijk}=c)}}\right), \quad (A.2)$$

where $\prod_{i,j,k}$ is shorthand for $\prod_i \prod_j \prod_k$.

Step 3: Updating the Underlying Data

For $l = 1, \dots, M$, sample $y_{ijl}^* = y_{ij}^*(s_l)$ from $N(\mu_{ij}(s_l), \psi^{-1})$ if $s_l + R_{ij} \notin \{t_{ijk}, k = 1, \dots, n_{ij}\}$. Otherwise, if Y is continuous, let $y_{ijl}^* = y_{ij}(s_l + R_{ij})$; and, if Y is ordinal, sample from

$$[y_{ijl}^* | y_{ij}(s_l + R_{ij}) = c, -] \stackrel{d}{=} \text{truncated } N(\mu_{ij}(s_l), 1) \quad \text{by } [\tau_{c-1}, \tau_c]. \quad (A.3)$$

This procedure is equivalent to sampling from the full conditional distribution of y_{ijl}^* .

Step 4: Updating $\Gamma, \Lambda_1, \Lambda_2, \{\eta_{i1}\}, \{\eta_{ij2}\}$

Expression (3) is linear in the parameters $\Gamma, \Lambda_1, \Lambda_2$ and in latent variables η_{i1} and η_{ij2} . Hence, under the prior structure defined in subsection 3.2, each of the full conditional distributions (given the baseline curve α , the underlying data $\{y_{ij}^*\}$, the observed data, and all other parameters and latent variables) follow forms derived previously for normal linear mixed models (e.g., Gilks et al., 1993). Reexpress (3) as $\mu_{ij}(s) = \alpha(s) + \mathbf{z}_{ij}(s)\boldsymbol{\beta}$, where $\mathbf{z}_{ij}(s) = [g_{m-1}\{s; \alpha(s)\}(\mathbf{x}_{ij}^T, \boldsymbol{\eta}_{i1}^T, \boldsymbol{\eta}_{ij2}^T), m = 1, \dots, q]$ is a $1 \times puvq$ row vector of weighted covariates and latent variables, and $\boldsymbol{\beta} = (\gamma_m, \boldsymbol{\lambda}_{1m}, \boldsymbol{\lambda}_{2m}, m = 1, \dots, q)^T \sim puvq \times 1$ is a column vector of regression coefficients and factor loadings with $\gamma_m, \boldsymbol{\lambda}_{1m}$ and $\boldsymbol{\lambda}_{2m}$ denoting the m th rows for Γ, Λ_1 and Λ_2 , respectively. The full conditional of $\boldsymbol{\beta}$ is

$$\text{N}\left(\left(\boldsymbol{\Sigma}_\beta^{-1} + \psi \sum_{i,j} \mathbf{z}_{ij} \mathbf{z}_{ij}^T\right)^{-1} \left(\boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\mu}_\beta + \psi \sum_{i,j} \mathbf{z}_{ij} (\mathbf{y}_{ij}^* - \boldsymbol{\alpha})\right), \left(\boldsymbol{\Sigma}_\beta^{-1} + \psi \sum_{i,j} \mathbf{z}_{ij} \mathbf{z}_{ij}^T\right)^{-1}\right), \quad (\text{A.4})$$

where $\mathbf{Z}_{ij} = (\mathbf{z}_{ij}(s_1)^T, \dots, \mathbf{z}_{ij}(s_M)^T)^T \sim puvq \times m$, $\mathbf{y}_{ij}^* = (y_{ij1}^*, \dots, y_{ijM}^*)^T$, and the prior for $\boldsymbol{\beta}$ is $\text{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$. Letting $\mathbf{g}(\alpha) = [\mathbf{g}\{s_1; \alpha(s_1)\}^T, \mathbf{g}\{s_2; \alpha(s_2)\}^T, \dots, \mathbf{g}\{s_M; \alpha(s_M)\}^T]^T \sim M \times q$ the full conditional densities of η_{i1} and η_{ij2} are respectively,

$$\text{N}\left(\mathbf{V}\boldsymbol{\eta}_{i1} \Lambda_1^T \mathbf{g}(\alpha)^T \left\{ \sum_{j=1}^{n_i} \mathbf{y}_{ij}^* - \mathbf{g}(\alpha) \Gamma \mathbf{x}_{ij} - \mathbf{g}(\alpha) \Lambda_2 \boldsymbol{\eta}_{ij2} \right\}, \mathbf{V}\boldsymbol{\eta}_{i1}\right), \quad (\text{A.5})$$

$$\text{N}\left(\mathbf{V}\boldsymbol{\eta}_{ij2} \Lambda_2^T \mathbf{g}(\alpha)^T \left\{ \mathbf{y}_{ij}^* - \mathbf{g}(\alpha) \Gamma \mathbf{x}_{ij} - \mathbf{g}(\alpha) \Lambda_1 \boldsymbol{\eta}_{i1} \right\}, \mathbf{V}\boldsymbol{\eta}_{ij2}\right), \quad (\text{A.6})$$

where $\mathbf{V}\boldsymbol{\eta}_{i1} = \{\mathbf{I}_{q \times q} + n_i \Lambda_1^T \mathbf{g}(\alpha)^T \mathbf{g}(\alpha) \Lambda_1\}^{-1}$ and $\mathbf{V}\boldsymbol{\eta}_{ij2} = \{\mathbf{I}_{q \times q} + \Lambda_2^T \mathbf{g}(\alpha)^T \mathbf{g}(\alpha) \Lambda_2\}^{-1}$.

Step 5: Updating the location of the reference point $\{R_{ij}\}$

Sample R_{ij} for each i, j from its full conditional distribution, which is multinomial with probabilities $\text{Pr}(R_{ij} =$

$t_l | \mathbf{y}_{ij}, \boldsymbol{\Theta}, \boldsymbol{\eta}_{i1}, \boldsymbol{\eta}_{ij2}) =$

$$\frac{\nu_l \prod_{k=1}^{n_{ij}} \prod_{c=1}^d [\Phi\{\tau_c - \mu_{ij}(t_{ijk} - t_l)\} - \Phi\{\tau_{c-1} - \mu_{ij}(t_{ijk} - t_l)\}]^{1(y_{ijk}=c)}}{\sum_{m=1}^L \nu_m \prod_{k=1}^{n_{ij}} \prod_{c=1}^d [\Phi\{\tau_c - \mu_{ij}(t_{ijk} - t_m)\} - \Phi\{\tau_{c-1} - \mu_{ij}(t_{ijk} - t_m)\}]^{1(y_{ijk}=c)}}, \quad (\text{A.8})$$

for $l = 1, \dots, L$.

Step 6: Updating the p.m.f. of R

Sample $\boldsymbol{\nu} = (\nu_1, \dots, \nu_L)$ from its full conditional distribution, which is

$$\text{Dirichlet}\left(v_1 + \sum_{i=1}^N \sum_{j=1}^{n_i} 1(R_{ij} = t_1), \dots, v_L + \sum_{i=1}^N \sum_{j=1}^{n_i} 1(R_{ij} = t_L)\right). \quad (\text{A.9})$$

Step 7: Updating the baseline curve α

Substituting a numerical approximation to the second derivative of $\alpha(s)$ in expression (7), the prior distribution for $\alpha(s)$ is truncated $N(\mu_{\alpha(s)}, \sigma_{\alpha(s)}^2)$ with

$$\mu_{\alpha(s)} = \begin{cases} \sigma_{\alpha(s)}^2 [\mu_{\alpha_1} \sigma_{\alpha_1}^{-2} + \{2\alpha(s + \delta) - \alpha(s + 2\delta) + \delta^2 a_0''(s + \delta)\} \rho \delta^{-4}] & \text{for } s = s_1 \\ .5\{\alpha(s + \delta) + \alpha(s - \delta) - \delta^2 a_0''(s)\} & \text{for } s \in (s_1, s_M) \\ \sigma_{\alpha(s)}^2 [\mu_{\alpha_M} \sigma_{\alpha_M}^{-2} + \{2\alpha(s + \delta) - \alpha(s + 2\delta) + \delta^2 a_0''(s + \delta)\} \rho \delta^{-4}] & \text{for } s = s_M, \end{cases}$$

$\sigma_{\alpha(s)}^2 = \{\sigma_{\alpha_1}^{-2} + \rho \delta^{-4}\}^{-1}$ for $s = s_1$, $\sigma_{\alpha(s)}^2 = \delta^4 / (4\rho)$ for $s \in (s_1, s_M)$, and $\sigma_{\alpha(s)}^2 = \{\sigma_{\alpha_M}^{-2} + \rho \delta^{-4}\}^{-1}$ for $s = s_M$.

The truncation bounds are chosen to assign zero probability to values of $\alpha(s)$ inconsistent with $\alpha \in A$. For $s \notin \{s_1, \dots, s_M\}$ the full conditional distribution of $\alpha(s)$ is equivalent to the prior, and for $s \in \{s_1, \dots, s_M\}$ the full conditional of $\alpha(s)$ is truncated

$$N\left(V_{\alpha(s)}\left(\mu_{\alpha(s)} \sigma_{\alpha(s)}^{-2} + [\sum_{i,j} y_{ij}(s)^* - \mathbf{g}\{s; \alpha(s)\}(\mathbf{\Gamma} \mathbf{x}_{ij} + \mathbf{\Lambda}_1 \boldsymbol{\eta}_{i1} + \mathbf{\Lambda}_2 \boldsymbol{\eta}_{ij2})]\right), V_{\alpha(s)}\right), \quad (A.10)$$

where $V_{\alpha(s)} = \{\sigma_{\alpha(s)}^{-2} + \sum_{i=1}^N n_i\}^{-1}$, and the density is truncated to ensure $\alpha \in A$.

REFERENCES

- Aitchison, J. and Silvey, S. D. (1957), "The generalization of probit analysis to the case of multiple responses," *Biometrika*, 44, 131-140.
- Albert, J. H. and Chib, S. (1993), "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669-679.
- Ansley, C. F., Kohn, R., and Wong, C. M. (1993), "Nonparametric spline regression with prior information," *Biometrika*, 80, 75-88.
- Arévalo, M., Sinai, I., and Jennings. (2000), "A fixed formula to define the fertile window of the menstrual cycle as the basis of a simple method of natural family planning," *Contraception*, 60, 357-360.
- Barry, D. (1995), "A Bayesian model for growth curve analysis," *Biometrics*, 51, 639-655.
- Billings, J.J. (1983). *The Ovulation Method*, 7th ed., Melbourne: Advocate Press.
- Brumback, B. A. and Rice, J. A. (1998), "Smoothing spline models for the analysis of nested and crossed samples of curves," *Journal of the American Statistical Association*, 93, 961-976.
- Chen, M.-H., Shao, Q.-M., and Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

- Colombo, B. and Masarotto, G. (2000), "Daily fecundability: first results from a new data base," *Demographic Research*, 3, 5.
- Cowles, M. K. and Carlin, B. P. (1996), "Markov chain Monte Carlo convergence diagnostics: a comparative review," *Journal of the American Statistical Association*, 91, 883-904.
- 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., Baird, D.D., Wilcox, A.J., and Weinberg, C.R. (1999), "Day-specific probabilities of clinical pregnancy based on two studies with imperfect measures of ovulation," *Human Reproduction*, 14, 1835-1839.
- Dunson, D. B., Colombo, B., and Baird, D.D. (2002), "Changes with age in the level and duration of fertility in the menstrual cycle," *Human Reproduction*, 17,.
- Dunson, D.B., Sinai, I., and Colombo, B. (2001), "The relationship between cervical secretions and the daily probabilities of pregnancy: Effectiveness of the TwoDay algorithm," *Human Reproduction*, 16, 2278-2282.
- Ecochard, R., Boehringer, H., Rabilloud, M., and Marret, H. (2001), "Chronological aspects of ultrasonic, hormonal, and other indirect indices of ovulation," *British Journal of Obstetrics and Gynecology*, 108, 822-829.
- Eriksen, G.V., Carlstedt, I., Ulbjerg, N. and Ernst, E. (1998), "Cervical mucins affect the motility of human spermatozoa in vitro," *Fertility and Sterility*, 70, 350-354.
- Fehring, R.J. (2002), "Accuracy of the peak day of cervical mucus as a biological marker of fertility,"
- Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398-409.
- Gilks, W. R., Wang, C. C., Yvonnet, B. and Coursaget, P. (1993), "Random-effects models for longitudinal data using Gibbs sampling," *Biometrics*, 49, 441-453.
- Gilks, W.R. and Wild, P. (1992), "Adaptive rejection sampling for Gibbs sampling," *Applied Statistics*, 41, 337-409.

- Harville, D.A. and Mee, R.W. (1984), "A mixed-model procedure for analyzing ordered categorical data," *Biometrics*, 40, 393-408.
- Heckman, N. and Ramsay, J.O. (2000), "Penalized regression with model-based penalties," *Canadian Journal of Statistics*, 28, 241-258.
- Jöreskog, K. G. (1973), "A general method for estimating a linear structural equation system," In A. S. Goldberger and O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences*. New York: Seminar Press, 85-112.
- 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.
- Ke, C. and Wang, Y. (2001), "Semi-parametric nonlinear mixed effects models and their applications," *Journal of the American Statistical Association*, 96, to appear.
- Lawton, W.H., Sylvestre, E.A., and Maggio, M.S. (1972), "Self-modeling nonlinear regression," *Technometrics*, 13, 513-532.
- Lindstrom, M.J. (1995), "Self modeling with random shift and scale parameters and a free-knot spline shape function," *Statistics in Medicine*, 14, 2009-2021.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087-1091.
- Pyper, C.M.M. and Knight, J. (2001), "Fertility awareness methods of family planning: The physiological background, methodology and effectiveness of fertility awareness methods," *The Journal of Family Planning and Reproductive Health Care* **27**, 103-110.
- Ramgopal, P., Laud, P. W., and Smith, A. F. M. (1993), "Nonparametric Bayesian bioassay with prior constraints on the shape of the potency curve," *Biometrika*, 80, 489-498.
- Sinai, I., Jennings, V., and Arévalo, M. (1999), "The TwoDay algorithm: A new algorithm to identify the fertile time of the menstrual cycle," *Contraception*, 60, 65-70.
- Smith, B. J. (2000). *Bayesian Output Analysis Program (BOA) Version 0.5.0 User Manual*, Department of Biostatistics, The University of Iowa College of Public Health.

- Templeton, A.A., Penney, G.C., and Lees, M.M. (1982), "Relation between the luteinizing hormone peak, the nadir of the basal body temperature and the cervical mucus score," *British Journal of Obstetrics and Gynaecology*, 89, 985-988.
- Tierney, L. (1994), "Markov chains for exploring posterior distributions," *Annals of Statistics*, 22, 1701-1762.
- Wahba, G. (1990), *Spline Models for Observational Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59.
- Wang, Y. (1998), "Smoothing spline models with correlated errors," *Journal of the American Statistical Association*, 93, 341-348.
- Wilcox, A.J., Dunson, D.B., and Baird, D.D. (2000), "The timing of the "fertile window" in the menstrual cycle: day-specific estimates from a prospective study," *British Medical Journal*, 321, 1259-1262.
- 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, 517-521.
- Zeger, S. L. and Diggle, P. J. (1994), "Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters," *Biometrics*, 50, 689-699.
- Zhang, D. W., Lin, X. H., and Sowers, M. F. (2000), "Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles," *Biometrics*, 56, 31-39.

Table 1: Cervical mucus data from 10 menstrual cycles.

Woman	Cycle	Cycle day relative to last day of hypothermia (BBT reference)																	
		-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
1	1	?	1	1	1	1	1	3	3	4	4	4	4	4	3	1	1	1	1
1	2	1	3	3	3	3	3	3	4	4	3	1	1	1	1	1	1	1	1
1	3	?	?	?	1	1	1	1	1	1	3	4	4	3	1	1	1	1	1
2	1	3	3	3	1	1	1	3	4	4	4	3	1	1	1	1	1	1	1
3	1	?	?	?	3	3	1	3	4	3	1	4	4	3	1	3	3	1	3
4	1	?	?	?	?	1	1	3	4	4	4	4	3	3	3	1	1	1	1
4	2	?	?	?	1	1	1	3	3	4	4	4	4	1	1	1	1	1	1
4	3	?	?	1	1	1	1	1	3	3	3	1	1	1	1	1	1	1	1
4	4	1	1	3	1	3	3	4	4	4	4	4	4	4	1	1	1	1	1
5	1	1	3	3	3	3	3	3	4	1	4	?	3	3	1	1	1	1	1

BBT reference is estimate of ovulation day based on basal body temperature data (LDH)

? denotes missing data.

Table 2*Posterior summaries of the parameters*

Parameter	Posterior Mean	Median	SD	95% Credible Interval
τ_1	0.254	0.255	0.013	(0.228, 0.281)
τ_2	1.073	1.073	0.021	(1.032, 1.113)
λ_1	0.186	0.184	0.028	(0.135, 0.245)
λ_2	0.467	0.464	0.047	(0.382, 0.565)
λ_3	0.037	0.033	0.026	(0.002, 0.096)
λ_4	0.074	0.070	0.048	(0.003, 0.178)

Figure 1. Posterior means and 95% pointwise credible intervals for the baseline trajectory in the mucus hydration score for a typical cycle from a typical woman. The prior mean a_0 is also shown.

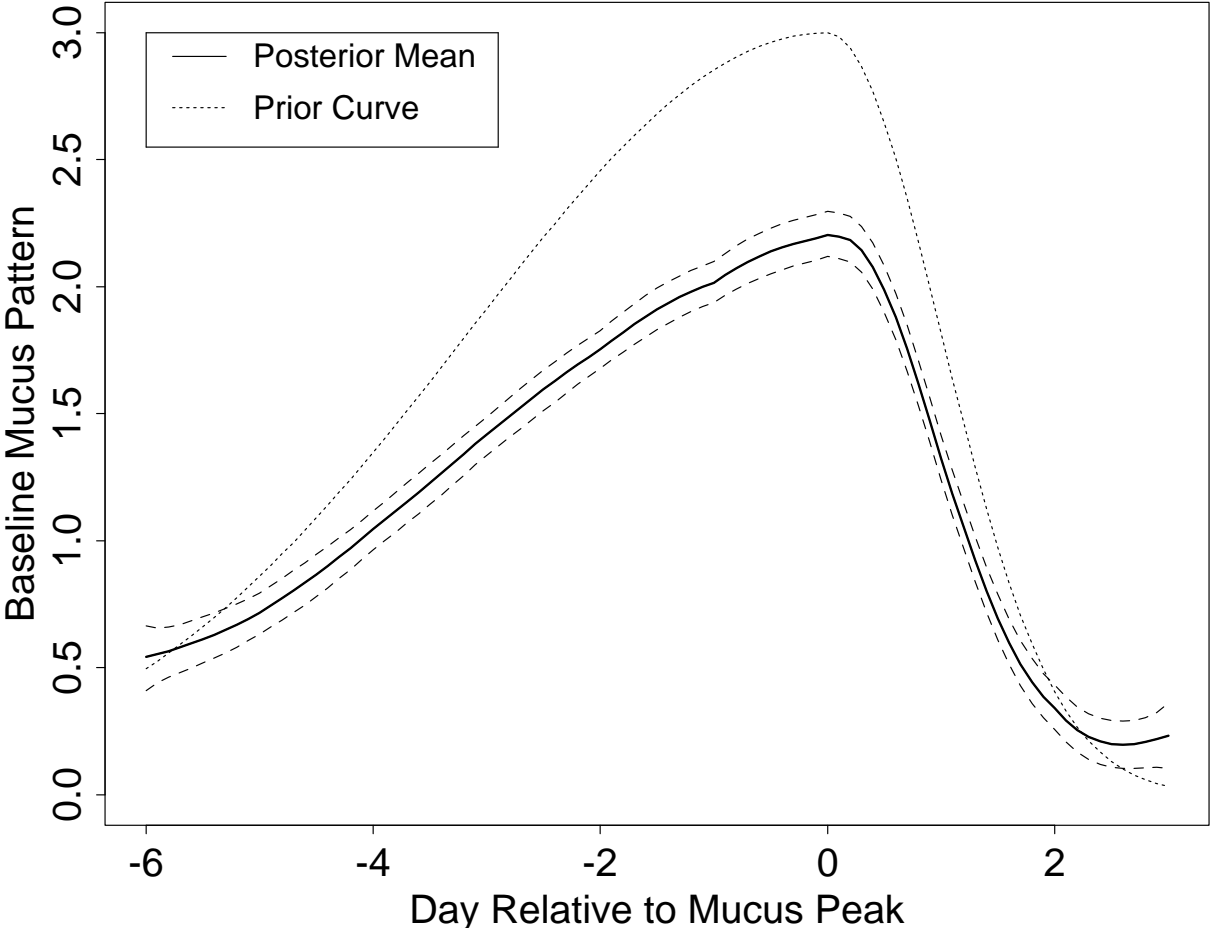


Figure 2. Estimated posterior means and 95% pointwise credible intervals for the distribution of the mucus hydration peak day (MHP) relative to the last day of hypothermia (LDH) reference day. The traditional estimate takes the mucus peak to be the last day in a cycle with the most fertile-type mucus detected in that cycle.

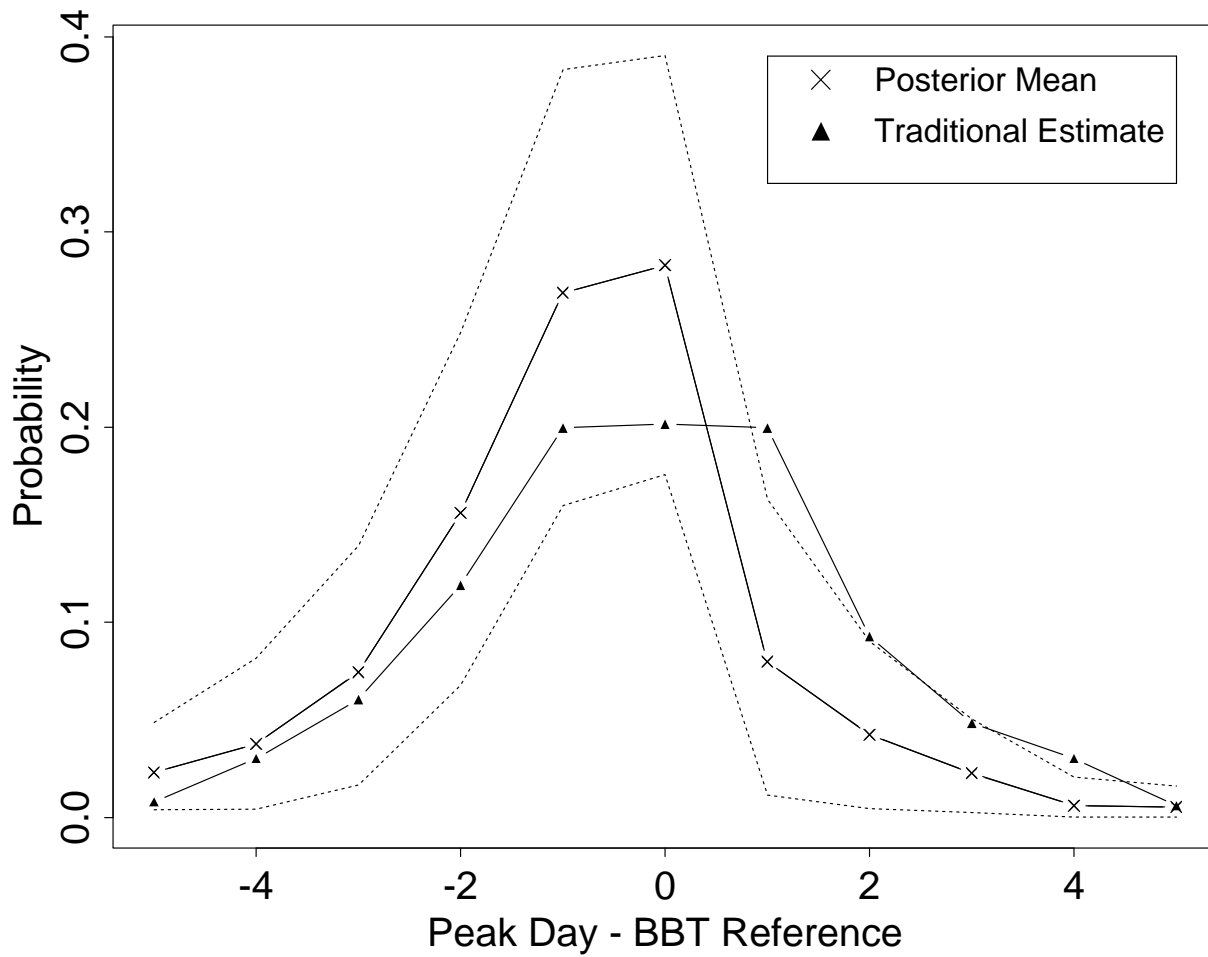


Figure 3. Predictive probabilities of most fertile-type mucus ($y = 4$) on different days of the menstrual cycle relative to the mucus hydration peak (MHP) for cycles in the 5th, 50th, and 95th percentiles with respect to the latent factors measuring (i) woman-specific effects on the level (η_{i11}); (ii) woman-specific effects on the shape (η_{i12}); (iii) cycle-specific effects on the level (η_{ij11}); and (iv) cycle-specific effects on the shape (η_{ij12}).

