Bayesian adaptive regression splines for hierarchical data

Jamie Lynn Bigelow and David B. Dunson

Biostatistics Branch

National Institute of Environmental Health Sciences

Research Triangle Park, NC 27709

April 5, 2005

SUMMARY. This article considers methodology for hierarchical functional data analysis, motivated by studies of reproductive hormone profiles in the menstrual cycle. Current methods standardize the cycle lengths and ignore the timing of ovulation within the cycle, both of which are biologically informative. Methods are needed that avoid standardization, while flexibly incorporating information on covariates and the timing of reference events, such as ovulation and onset of menses. In addition, it is necessary to account for within-woman dependency when data are collected for multiple cycles. We propose an approach based on a hierarchical generalization of Bayesian multivariate adaptive regression splines. Our formulation allows for an unknown set of basis functions characterizing the population-averaged and woman-specific trajectories in relation to covariates. A reversible jump Markov chain Monte Carlo algorithm is developed for posterior computation. Applying the methods to data from the North Carolina Early Pregnancy Study, we investigate differences in progesterone profiles between conception

email: bigelow@niehs.nih.gov

and non-conception cycles.

1. Introduction

In many longitudinal studies, each subject contributes a set of data points that can be considered error-prone realizations of a function of time. Although it is standard practice to model the longitudinal trajectory relative to a single reference point in time, such as birth or the start of treatment, there may be several reference points that are informative about a subject's response at a given time. One example of reference points is disease onset, start of treatment, and death in a longitudinal study of quality of life. The current project uses onset of menses and ovulation as reference points in a study of reproductive hormones.

Our research was motivated by progesterone data from the North Carolina Early Pregnancy Study (NCEPS) (Wilcox et al., 1988). Daily measurements of urinary pregnanediol-3-glucuronide (PdG), a progesterone metabolite, were available for 262 complete menstrual cycles and 199 partial mid-cycle segments from a total of 173 women. It is of special interest to examine the differences in progesterone profiles between conception and non-conception cycles. The onset of menses marks the start of the follicular phase of the menstrual cycle, which ends at ovulation. The luteal phase begins at ovulation and, if no conception occurs, ends at the start of the next menses. In general, progesterone begins to rise in the follicular phase until several days into the luteal phase, when it decreases in preparation for the next cycle or, if conception has occurred, continues to rise. Figure 1 displays log-PdG data from one subject for a non-conception and subsequent conception cycle.

[Figure 1 about here.]

The most common way to examine hormone data within the menstrual cycle is

to restrict attention to a fixed window surrounding ovulation (see Baird et al., 1997; Brumback and Rice, 1998; Massafra et al., 1999; Dunson et al., 2003 for examples). This is desirable for ease of modeling, but fails to use all data by discarding days outside the window. In addition, it ignores cycle length and the relative timing of ovulation within a cycle. Another approach is to standardize all cycles to a common length. Zhang et al. (1998) modeled progesterone with smoothing splines after standardizing cycles to 28 days. Standardization discards biologically important information on the timing of ovulation, obscuring its well known relationship with hormone trajectories. van Zonneveld et al. (2003) indicated that both the onset of menses and the day of ovulation are related to hormone levels within a cycle and implemented separate analyses for windows around each of these reference points. Ideally, a single model would allow the response to vary flexibly relative to multiple reference points, while accommodating covariates and within-woman dependency.

The goal of the analysis is to characterize differences in progesterone profiles between conception and non-conception cycles. When conception occurs, PdG rises in response to implantation of the conceptus, which usually occurs around the eighth or ninth day after ovulation (Baird et al., 1997). We are also interested in differences before implantation because they may predict the fertility of the cycle. Researchers have studied conception differences in midluteal (5-6 days after ovulation) and baseline (preovulatory) PdG. Studies of have shown that conception cycles have elevated midluteal PdG over paired non-conception cycles with well-timed intercourse or artificial insemination (Stewart et al., 1993; Baird et al., 1997), but one study (Lipson and Ellison, 1996) found no difference. None of these three studies found a relationship between baseline PdG and conception. However, limiting analysis to cycles with well-timed exposure to semen is biased to include non-conception cycles of inherently low fertility, failing to represent

the true difference between conception and non-conception cycles. In addition, requiring paired cycles selects against couples of very high or very low fertility and fails to use all data from women with more or less than two cycles. In a previous analysis of the NCEPS data which included cycles without well-timed intercourse, Baird et al. (1999) found that cycles with very low midluteal PdG were unlikely to be conception cycles. Although midluteal PdG did not monotonically affect the odds of conception, increased baseline PdG was associated with decreased odds of conception.

Hormone data are a special case of hierarchical functional data. The daily measurements are subject to assay errors, yielding a noisy realization of the true trajectory of urinary PdG. The hierarchy results from the multiple cycles contributed by each woman. Methods for hierarchical functional data typically require that all curves are observed over or standardized to fall in the same region (Brumback and Rice, 1998; Morris et al., 2003; Brumback and Lindstrom, 2004). To accommodate the dependence structure without cycle standardization, we propose a Bayesian method based on a hierarchical generalization of multivariate adaptive regression splines.

Holmes and Mallick (2001) proposed Bayesian regression with multivariate linear splines to flexibly characterize the relationship between covariates and a scalar response from independent sampling units. The number of knots and their locations are random, and smooth prediction curves are obtained by averaging over MCMC sampled models. A extension of this method yielded a generalized nonlinear regression model for a vector response (Holmes and Mallick, 2003). Our goal is to develop a new hierarchical adaptive regression splines approach to accommodate clustered functional data, potentially having unequal numbers and locations of observations per subject, a common complication in longitudinal studies. We incorporate reference point information by including time relative to each of the reference points as covariates in the regression model.

A popular method for analyzing multivariate response data with spline bases is seemingly unrelated regression (SUR), in which each subject is allowed a unique set of basis functions, but the basis coefficients are common to all subjects (Percy, 1992). We instead use one set of unknown basis functions, allowing the basis coefficients to vary from subject to subject. To estimate the population regression function, we treat the subject-specific basis coefficients as random, centered around the population mean basis coefficients. The resulting model is extremely flexible, and can be used to capture a wide variety of covariate effects and heterogeneity structures.

In section 2, we describe the model, prior structure and a reversible jump Markov chain Monte Carlo (RJMCMC) (Green, 1995) algorithm for posterior computation. In section 3, we illustrate the performance of the approach for a simulation example. Section 4 applies the method to progesterone data from the NC-EPS, and section 5 discusses the results.

2. Methods

2.1 *Prior specification*

Typically, the number and locations of knots in a piecewise linear spline are unknown. By allowing for uncertainty in the knot locations and averaging across the resulting posterior, one can obtain smoothed regression functions. We follow previous authors (Green, 1995; Holmes and Mallick, 2001) in using the RJMCMC algorithm to move among candidate models of varying dimension. Our final predictions are constructed from averages over all sampled models. We assume a priori that all models are equally probable, so our prior on the model space is uniform.

Each piecewise linear model, M, is defined by its basis functions $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)$, where $\boldsymbol{\mu}_l$ is $p \times 1$. Consider y_{ij}, the j^{th} PdG measurement for subject i. Under model M, the true relationship between y_{ij} and its covariates $\mathbf{x}_{ij}' = (1, x_{ij2}, \ldots, x_{ijp})$ can be approxi-

mated by the piecewise linear model:

$$y_{ij} = \sum_{l=1}^{k} b_{il}(\mathbf{x}'_{ij}\boldsymbol{\mu}_l)_+ + \epsilon_{ij}, \tag{1}$$

where $\epsilon_{ij} \overset{iid}{\sim} N(0, \tau^{-1})$. The value of the j^{th} response of subject i is approximated by a linear combination of the positive portion (denoted by the + subscript) of the inner products of the basis functions with the covariate vector, $\mathbf{x}_{ij}$. We require that each model contain an intercept basis, so we define $(\mathbf{x}'_{ij}\boldsymbol{\mu}_1)_+ \equiv 1$ for all i, j. We extend previous methods by allowing the spline coefficients, $\mathbf{b}_i$ to be subject-specific, assuming that observations within subject i are conditionally independent given $\mathbf{b}_i$.

Each piecewise linear model is linear in the basis function transformations of the covariate vectors:

$$\mathbf{y}_i = \boldsymbol{\theta}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{2}$$

where $\mathbf{y}_i$ and $\boldsymbol{\epsilon}_i$ are the $n_i \times 1$ vectors of responses and random errors and $\mathbf{b}_i$ is the $k \times 1$ vector of subject specific basis coefficients for subject i. The $n_i \times k$ design matrix, $\boldsymbol{\theta}_i$, contains the basis function transformations of the covariate vectors for subject i:

$$\boldsymbol{\theta}_i = \begin{pmatrix} 1 & (\mathbf{x}'_{i1}\boldsymbol{\mu}_2)_+ & \dots & (\mathbf{x}'_{i1}\boldsymbol{\mu}_k)_+ \\ 1 & (\mathbf{x}'_{i2}\boldsymbol{\mu}_2)_+ & \dots & (\mathbf{x}'_{i2}\boldsymbol{\mu}_k)_+ \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (\mathbf{x}'_{in_i}\boldsymbol{\mu}_2)_+ & \dots & (\mathbf{x}'_{in_i}\boldsymbol{\mu}_k)_+ \end{pmatrix}$$

Since we use only the positive portion of each linear spline, it is possible that a basis function does not contribute to the model for a given subject (i.e. $\boldsymbol{\theta}_i$ contains a column of zeros, which is non-informative about the corresponding element of $\mathbf{b}_i$). To address this problem, we standardize each column of the population design matrix, $\boldsymbol{\Theta} = (\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_m)'$, to have mean 0 and variance 1. Assuming independent subjects, this model specification yields the likelihood:

$$L(\mathbf{y}|\boldsymbol{b}, \tau, M) \propto \prod_{i=1}^{m} \tau^{\frac{n_i}{2}} exp\big[-\frac{\tau}{2}(\mathbf{y}_i - \boldsymbol{\theta}_i\mathbf{b}_i)'(\mathbf{y}_i - \boldsymbol{\theta}_i\mathbf{b}_i)\big] \tag{3}$$

This likelihood is defined conditionally on the subject-specific basis coefficients, but we wish to make inferences also on population parameters. Treating the subject-specific coefficients as random slopes, we specify a Bayesian random effects model where the subject-specific coefficients are centered around the population coefficients, $\boldsymbol{\beta}$. Under model M of dimension k, the relationship between the population and subject-specific coefficients is specified through the hierarchical structure:

$$\begin{aligned} \mathbf{b}_i|k &\sim N_k(\boldsymbol{\beta}, \tau^{-1}\boldsymbol{\Delta}^{-1}) \quad \forall i \\ \boldsymbol{\beta}|k &\sim N_k(\mathbf{0}, \tau^{-1}\lambda^{-1}\mathbf{I}_k) \end{aligned} \tag{4}$$

To avoid over-parameterization of an already flexible model, we assume independence among the elements of $\mathbf{b}_i$. Thus $\boldsymbol{\Delta} = diag(\boldsymbol{\delta})$, where $\boldsymbol{\delta}$ is a $k \times 1$ vector. The elements of $\boldsymbol{\delta}$ and the scalars λ and τ are given independent gamma priors:

$$\pi(\tau, \lambda, \boldsymbol{\delta}) \propto \tau^{a_\tau - 1} exp(-b_\tau \tau) \lambda^{a_\lambda - 1} exp(-b_\lambda \lambda) \prod_{l=1}^{k} \big(\delta_l^{a_\delta - 1} exp(-b_\delta \delta_l)\big),$$

where $a_\tau,\ b_\tau,\ a_\lambda,\ b_\lambda,\ a_\delta$ and b_δ are pre-specified hyperparameters. Each of the $k-1$ non-intercept basis functions contains a non-zero intercept and linear effect for at least one covariate. Including multiple covariate effects in a single basis allows the covariates to dependently affect the response (i.e. allows for interactions). The number of non-zero covariate effects in a particular basis is called the interaction level of the basis.

Under one piecewise linear model, an observation y with covariates x has the following mean and variance:

$$E(y) = \beta_1 + \sum_{l=2}^{k} \beta_l (\mathbf{x}'\boldsymbol{\mu}_l)_+$$

$$V(y) = \delta_1^{-1} + \sum_{l=2}^{k} \delta_l^{-1} (\mathbf{x}'\boldsymbol{\mu}_l)_+^2 + \tau^{-1}$$

The mean and variance can vary flexibly with the covariates and relative to each other. The elements of $\boldsymbol{\beta}$ can be positive or negative, large or small, and the elements of $\boldsymbol{\delta}$ can also be large or small. A given basis could contribute substantially to the mean and negligibly to the variance (i.e. β_l and δ_l are both large), or vice versa, so that the mean and variance of the response at a given set of covariates are not constrained to vary together.

2.2 *Posterior computation*

At each iteration, we obtain a piecewise linear model for which the parameters can be sampled directly from their full conditionals as derived from the priors and the likelihood following standard algebraic routes. Omitting details, we obtain the following full conditional posterior distributions:

$$\boldsymbol{\beta}|\mathbf{b},\boldsymbol{\delta},\lambda,\tau,\mathbf{D} \sim N_k\Big([\lambda\mathbf{I}_k + m\boldsymbol{\Delta}]^{-1}\boldsymbol{\Delta}\sum_{i=1}^{m}\mathbf{b}_i, \tau^{-1}[\lambda\mathbf{I}_k + m\boldsymbol{\Delta}]^{-1}\Big)$$

$$\mathbf{b}_i|\boldsymbol{\beta},\boldsymbol{\delta},\lambda,\tau \sim N_{k_i}\Big([\boldsymbol{\theta}_i'\boldsymbol{\theta}_i+\boldsymbol{\Delta}]^{-1}[\boldsymbol{\theta}_i'\mathbf{y}_i+\boldsymbol{\Delta}\boldsymbol{\beta}], \tau^{-1}[\boldsymbol{\theta}_i'\boldsymbol{\theta}_i+\boldsymbol{\Delta}]^{-1}\Big) \qquad i = 1,\ldots,m$$

$$\tau|\boldsymbol{\beta},\mathbf{b},\boldsymbol{\delta},\lambda \sim Gamma\Big(a_\tau + \frac{(m+1)k+n}{2},$$
$$b_\tau + \frac{1}{2}\sum_{i=1}^{m}[(\mathbf{b}_i-\boldsymbol{\beta}_i)'\boldsymbol{\Delta}(\mathbf{b}_i-\boldsymbol{\beta}_i) + (\mathbf{y}_i-\boldsymbol{\theta}_i\mathbf{b}_i)'(\mathbf{y}_i-\boldsymbol{\theta}_i\mathbf{b}_i)] + \lambda\boldsymbol{\beta}'\boldsymbol{\beta}\Big)$$

$$\lambda|\boldsymbol{\beta},\mathbf{b},\boldsymbol{\delta},\tau \sim Gamma\Big(a_\lambda+\frac{k}{2}, b_\lambda+\frac{\boldsymbol{\beta}'\boldsymbol{\beta}}{2}\Big)$$

$$\delta_l|\boldsymbol{\beta},\mathbf{b},\boldsymbol{\delta}_{-l},\lambda,\tau \sim Gamma\Big(a_\delta+\frac{m}{2}, b_\delta+\frac{\tau}{2}\sum_{i=1}^{m}(\mathbf{b}_{il}-\beta_l)^2\Big) \qquad l = 0,\ldots,(k-1)$$

where a $Gamma(a,b)$ random variable is parameterized to have expected value a/b and variance a/b^2.

The following is a description of the RJMCMC algorithm we employed:

Step 0: Initialize the model to the intercept-only basis function, where $k = 1$.

Step 1 : Propose with equal probability either to add, alter or remove a basis function. If $k = 1$ in the current model, then we cannot remove or change a basis, so we choose either to add a basis function or to skip to step 2 and redraw the parameters for the intercept basis.

ADD Generate a new basis function as follows: Draw the interaction level of the basis uniformly from $(1, \ldots, p-1)$ and randomly select the corresponding number of covariates. Set basis parameters for all other covariates equal to zero. Sample selected basis function parameters from $N(0,1)$, then normalize to get $(\mu_{l1}, \ldots, \mu_{lp})$, the non-intercept basis parameters. Randomly select one data point, y_{ij}, and let $\mu_{l0} = \mathbf{x}'_{ij,-1}\mu_{l,-1}$. Add the new basis function to the proposed model.

ALTER: Randomly select a basis in the current model. Generate a new basis function as described above. Replace the selected basis function with the new one

REMOVE: Randomly select a basis in the current model. Delete the selected basis from the proposed model.

Step 2: Accept the proposed model with appropriate probability (described below).

Step 3: If a proposal to add or remove has been accepted, the dimension of the model has changed. In order to update the parameters from their full conditionals, all vector parameters must have dimension k^* of the new model. It suffices to adjust the dimension of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, as we can then sample $\{\mathbf{b}_i\}$ from the full conditionals. If we've added a basis, initialize β_{k^*}, the new element of $\boldsymbol{\beta}$, to a pre-determined initial value and initialize δ_{k^*} to the mean of $\boldsymbol{\delta}$ from the previous model. If a basis has been removed, delete the corresponding elements of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.

Step 4: Update $\{\mathbf{b}_i\}$, $\boldsymbol{\beta}$, τ, $\boldsymbol{\delta}$, and λ from their full conditionals.

Repeat steps 1-4 for a large number of iterations, collecting samples after a burn-in to allow convergence.

A challenging aspect of the algorithm is comparing models in the RJMCMC sampler. Our prior assigns equal probability to all piecewise linear models and model proposal is based on generation of discrete random variables. Under this scenario, the probability, Q, of accepting a proposed model, M^*, is the Bayes factor comparing it to the current model, M (Holmes and Mallick, 2003, Denison et al., 2002). The Bayes factor is the ratio of the marginal likelihoods of the data under the two models:

$$Q = min\Big[1, \frac{p(\mathbf{y}|M^*)}{p(\mathbf{y}|M)}\Big].$$

The marginal likelihoods and thus the Bayes factor for this hierarchical model have no closed form. Consider instead the following marginal likelihood under model M.

$$p(\mathbf{y}|M, \boldsymbol{\delta}, \lambda) = \int\int\int L(\mathbf{y}|\mathbf{b}, \tau, \lambda, M) p(\mathbf{b}, \tau, \boldsymbol{\beta}|\boldsymbol{\delta}, \lambda, M)\, d\mathbf{b}\, d\boldsymbol{\beta}\, d\tau,$$

where $p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{b}, \tau, \boldsymbol{\delta}, \lambda, M)$ is the data likelihood under model M, and $p(\mathbf{b}, \tau, \boldsymbol{\beta}|\boldsymbol{\delta}, \lambda, M)$ is the joint prior of $\mathbf{b}$, $\boldsymbol{\beta}$, and τ under model M. This integral has a closed form, so that the likelihood can be written:

$$p(\mathbf{y}|M, \boldsymbol{\delta}, \lambda) = C(\lambda, k)|\mathbf{R}|^{-\frac{1}{2}}(b_\tau + \frac{\alpha}{2})^{-(\frac{n}{2}+a_\tau)} \prod_{l=1}^{k} \delta_l^{\frac{m}{2}} \prod_{i=1}^{m} |\mathbf{U}_i|^{\frac{1}{2}} \tag{5}$$

where

$$\mathbf{U}_i = [\Delta + \boldsymbol{\theta}_i'\boldsymbol{\theta}_i]^{-1}$$

$$\mathbf{R} = \lambda \mathbf{I}_k + m\boldsymbol{\Delta} - \boldsymbol{\Delta}(\sum_{i=1}^{m} \mathbf{U}_i)\boldsymbol{\Delta}$$

$$\alpha = \mathbf{y}'\mathbf{y} - \sum_{i=1}^{m} \mathbf{y}_i'\boldsymbol{\theta}_i\mathbf{U}_i\boldsymbol{\theta}_i'\mathbf{y}_i - (\sum_{i=1}^{m} \mathbf{U}_i\boldsymbol{\theta}_i'\mathbf{y}_i)'\boldsymbol{\Delta}\mathbf{R}^{-1}\boldsymbol{\Delta}(\sum_{i=1}^{m} \mathbf{U}_i\boldsymbol{\theta}_i'\mathbf{y}_i)$$

$$C(\lambda, k) = \frac{{b_\tau}^{a_\tau} \lambda^{\frac{k}{2}} \Gamma(\frac{n}{2} + a_\tau)}{\Gamma(a_\tau)(2\pi)^{\frac{n}{2}}}$$

In a similar fashion, we can write the marginal likelihood for a proposed model M^* of dimension k^*.

$$p(\mathbf{y}|M^*, \boldsymbol{\delta}^*, \lambda^*) = C(\lambda^*, k^*)|\mathbf{R}^*|^{-\frac{1}{2}}(b_\tau + \frac{\alpha^*}{2})^{-(\frac{n}{2}+a_\tau)} \prod_{l=1}^{k*} {\delta_l^*}^{\frac{m}{2}} \prod_{i=1}^{m} |\mathbf{U}^*{}_i|^{\frac{1}{2}} \tag{6}$$

Suppose we propose a move from model M of dimension k to model M^* of dimension k^*. If we let the acceptance probability be the ratio of the two marginal likelihoods, then it depends on λ and $\boldsymbol{\delta}$. It also depends on λ^* and $\boldsymbol{\delta}^*$, for which we do not have estimates. Since we wish to accept or reject a model based only on its set of basis functions, we want to minimize the effects of these variance components on the acceptance probability. Specifically, we assume λ=λ^* at the current sampled value. Since $\boldsymbol{\delta}^*$ and $\boldsymbol{\delta}$ may be of different dimensions, we cannot assume that they are equal. Instead, we assume that they are equal in the elements corresponding to bases common to both models and condition only on those elements.

Consider a proposal to add a basis to the current model. The current model is nested in the proposed model, and the proposed model has exactly one more basis than the current model. The acceptance probability is:

$$Q = min\Big[1, \frac{p(\mathbf{y}|M^*, \lambda, \boldsymbol{\delta})}{p(\mathbf{y}|M, \lambda, \boldsymbol{\delta})}\Big]$$

The denominator has closed form, as we've shown above, and the numerator can be derived as follows, where $\boldsymbol{\delta}^* = (\boldsymbol{\delta}, \delta_{k^*})$.

$$p(\mathbf{y}|M^*, \lambda, \boldsymbol{\delta}) = \int p(\mathbf{y}, \delta_{k^*}|M^*, \boldsymbol{\delta}, \lambda) d\delta_{k^*} = \int p(\mathbf{y}|M^*, \boldsymbol{\delta}^*, \lambda)\pi(\delta_{k^*}) d\delta_{k^*}$$

$$= \frac{C(\lambda, k^*)}{\Gamma(a_\delta)} \prod_{l=1}^{k} \delta_l \int_0^\infty |\mathbf{R}^*|^{-\frac{1}{2}}(b_\tau + \frac{\alpha^*}{2})^{-(\frac{n}{2}+a_\tau)} {\delta_{k^*}}^{a_\delta} exp(-b_\delta \delta_{k^*}) {b_\delta}^{a_\delta} \prod_{i=1}^{m} |\mathbf{U}^*{}_i|^{\frac{1}{2}} d\delta_{k^*} \tag{7}$$

This integral is complicated, and we approximate it using the Laplace method. This involves fitting a scaled normal density to the integrand. Specifically, if we wish to evaluate $\int h(\theta)d\theta$, we assume that $h(\theta) \approx h(\hat{\theta})exp(\frac{-(\theta-\bar{\theta})^2}{2\sigma^2})$, where $\bar{\theta}$ is the mode of $h(\theta)$ and $\hat{\sigma}^2$ is the estimate of the variance of the normal density. A good estimate of the mode, $\hat{\theta}$, can be obtained with a numerical search algorithm. The variance can be estimated by noting that $\frac{h(\hat{\theta})}{h(\hat{\theta}+\epsilon)} \approx exp(\epsilon^2 2\sigma^2)$. We evaluate h at $(\hat{\theta}+\epsilon)$ and $(\hat{\theta}-\epsilon)$ and average the two resulting estimates of σ^2 to get $\hat{\sigma}^2$. The integral is then approximated by $(2\pi)^{\frac{1}{2}}(\hat{\sigma})^{\frac{1}{2}}h(\hat{\theta})$. For additional information on the Laplace method and other methods for Bayes factor approximation, see DiCiccio et al. (1997).

Since the integral we want to approximate is defined over $\Re^+$ and the normal distribution is defined over the entire real line, we will transform δ_{k^*}. Simulations show that this has the added benefit of making the integrand more symmetric. Let $\omega = log(\delta_{k^*})$ and note that the prior on ω_{k^*} is:

$$\pi(\omega_{k^*}) = \frac{exp(a_\delta\omega - b_\delta[exp(\omega)])b_\delta{}^{a_\delta}}{\Gamma(a_\delta)}$$

The integral in (7) can be written:

$$p(\mathbf{y}|M^*, \boldsymbol{\delta}, \lambda) = \int p(\mathbf{y}, \omega|M^*, \boldsymbol{\delta}, \lambda)d\omega = \int_{-\infty}^{\infty} p(\mathbf{y}|M^*, \boldsymbol{\delta}, \omega, \lambda)\pi(\omega)d\omega$$

$$= \frac{C(\lambda, k^*)}{\Gamma(a_\delta)} \prod_{l=1}^{k} \delta_l \int exp(\omega + a_\delta\omega - b_\delta[exp(\omega)])|\mathbf{R}^*|^{-\frac{1}{2}}(b_\tau + \frac{\alpha^*}{2})^{-(\frac{n}{2}+a_\tau)} \prod_{i=1}^{m} |\mathbf{U}_i^*|^{\frac{1}{2}} d\omega$$

Similarly, a basis removal proposal involves integrating out the element of $\boldsymbol{\delta}$ corresponding to the basis proposed for removal. A proposal to alter a basis involves integrating out the element of $\boldsymbol{\delta}$ corresponding to that basis in both the numerator and the denominator.

2.3 *Computation*

In implementing the RJMCMC algorithm described above, we run a burn-in period of several thousand iterations until convergence is apparent. Convergence is evidenced by the stationarity of the distribution of the marginal likelihood in (5) and the distribution of k, the dimension of sampled models. Then the sampler is run for an additional period, during which each selected piecewise linear function is saved. Final estimates of the population regression function are based on averages over all the saved models, and credible intervals for the response can be calculated for any set of covariate values. In addition, the subject-specific coefficients are saved at each step, so that the individual regression function can be estimated and individual credible intervals can be calculated.

The analysis is conducted using Matlab version 7.0.1. The method is computationally intensive, especially for large datasets. However, the rates of convergence and mixing are good enough that it can be practically implemented even in complex settings, such as that described in the data example.

3. Simulated data example

The simulated data do not mimic longitudinal data with reference points. Rather, we illustrate the broad applicability of the method by simulating clustered data with a covariate-dependent random effect. We simulated data for 200 subjects, with each subject contributing 30 observations from the following distribution:

$$(y_{ij}|\mathbf{x}_{ij}) \sim N\Big(x_{1ij} - x_{2ij}^2 + x_{1ij}x_{2ij} + b_i\sqrt{2|x_{1ij}|},\ 2\Big)$$

where the covariates x_{1ij} and x_{2ij} for the j^th observation from subject i are randomly generated integers between -4 and 4, and b_i is a $N(0,1)$ random term for subject i. Note that the random effect varies non-linearly with x_1. We want the method to be able to detect this variation. In addition, the model-estimated population mean, subject-

specific means, and random effects should be consistent with the simulated data.

We ran the RJMCMC algorithm for 50,000 iterations, discarding the first 10,000 as burn-in. In the first chain, the hyperparameters a_τ, b_τ, a_λ, b_λ, a_δ and b_δ were all set to 0.05, yielding vague priors for the variance components. When proposals were accepted, new elements of $\boldsymbol{\beta}$ were initialized to 0. Sensitivity to hyperparameters and initial values was assessed through an additional chain where $a_\tau, a_\lambda, a_\delta = 1$, $b_\tau, b_\lambda, b_\delta = 0.5$, and the new elements of $\boldsymbol{\beta}$ were initialized to 1. The two chains yielded virtually identical results. This suggests that the method is not overly sensitive to specification of initial values and hyperparameters.

We calculated subject-specific estimates for each data point as well as population predictions over the covariate space. Figure 2 illustrates the model's ability to discern features of the data. Figure 2a shows a scatterplot of the population mean values estimated under the algorithm against the true mean values for each covariate combination. This indicates that the model was able to distinguish the underlying population mean structure from the random effects. The empirical estimates of the random effects were calculated by subtracting the model-predicted population mean from the subject-specific posterior mean for each data point. As shown in Figure 2b, the empirical estimates of the random effects were generally accurate estimates of the true values of the random effects, $\{x_1^2 b_i\}$. At each iteration, the estimated variance under the current model for each set of covariate values was calculated:

$$V_e(y|x_1, x_2) = \delta_0^{-1} + \sum_{l=1}^{k-1} \delta_l^{-1} (\mathbf{x}'\boldsymbol{\mu}_l)_+^2 + \tau^{-1}$$

where $\boldsymbol{\delta}$ and τ are the estimates of the variance components under the current k-dimensional model. The empirical variance estimate can be compared to the true

variance:

$$V(y|x_1, x_2) = |x_1| + 2$$

Figure 2c shows the average over all samples of the empirical variance at each covariate pair plotted against the true variance. The model-estimated values pick up the general trend of the true values, but there seems to be a tendency toward slight underestimation.

Figure 2d is a traceplot of the model marginal likelihood (5) over the sampled iterations. The distribution of this quantity, and of the associated predictions, appears to be stationary, so we find no evidence against convergence of the MCMC algorithm.

[Figure 2 about here.]

4. Progesterone example

4.1 *Estimation*

We applied these methods to the progesterone data from the NCEPS described in Section 1 with the goal of assessing differences in PdG profiles between conception and non-conception cycles. We were particularly interested in examining differences prior to implantation, since these may indicate hormonal effects on fecundability and conception probabilities.

We apply the approach described in 2 with three covariates and an intercept. The first covariate is an indicator of whether the cycle during which the measurement was taken resulted in conception. The final two covariates contain the reference point information. They are number of days since cycle start (onset of menses) and number of days relative to ovulation in current cycle. So if response y_{ij} was observed on the third day of a non-conception cycle where ovulation occurred on day 14, then $\mathbf{x}_{ij} = (1, 0, 3, -11)'$.

Vague priors on the variance components were achieved by setting the hyperparameters to 0.05. We collected 40,000 MCMC samples after a 20,000 iteration burn-in.

4.2 *Inference*

We can use model estimates to assess the relationship among progesterone, cycle conception status, and the two reference points. The main analysis goal was to gain a better understanding of the differences in progesterone between conception and non-conceptions cycles. At each iteration, we calculate several summary variables for each cycle based on the trajectories estimated by the subject-specific coefficients. Early follicular PdG was the average over the first 5 days of the cycle, baseline PdG was the average from 6 days until 2 days before ovulation, and midluteal PdG was the average on days 5 and 6 after ovulation. The early luteal PdG rise was the change in PdG from 1 day to 5 days after ovulation. We record the mean of each of these variables for conception and non-conception cycles at each iteration, using these samples to create overall means and credible intervals.

Baird et al. (1999) suggested that conception was less likely in cycles with low midluteal PdG. To test this, we find the 10^{th} percentile of midluteal PdG over all cycles at each iteration and record the proportion of cycles that are conceptions both under and over this threshold. In the process, we obtain posterior means and 95% pointwise credible intervals for the population PdG trajectory for conception and non-conception cycles at any location relative to the reference points.

5. Results

Convergence was deemed adequate, as the distribution of the marginal likelihood appeared to be stationary. In addition, the distribution of the dimension of sampled models was stationary. Sample collection took approximately 72 hours.

Figure 3 displays data from a single subject, the fitted PdG curve, and the predicted population mean log-PdG given the woman's covariates. The subject-specific curve captures the subject's data more closely than the population curve, illustrating

the potential for a shape difference between the population-mean and subject-specific curves.

[Figure 3 about here.]

Table 1 gives the estimated differences in log-PdG between conception and non-conception cycles for the intervals of interest. Early follicular PdG over the first five days of the cycle tended to be higher in non-conception cycles. In addition, non-conception cycles tended to have higher baseline and slightly higher midluteal PdG than conception cycles. There was a larger average early luteal PdG rise in conception cycles, though the 95% credible interval for the difference includes zero.

[Table 1 about here.]

Table 2 summarizes the relationship between conception status and low midluteal progesterone, with 95% credible intervals. Those cycles with estimated midluteal PdG in the lowest 10% were less likely to be conception cycles than those with higher PdG, although in Table 1 we saw that non-conception cycles had higher midluteal PdG on average.

[Table 2 about here.]

These results have been based on the subject-specific basis coefficients only, and we now examine population progesterone curves. Figure 4 displays the predicted population mean log-progesterone for the first 28 days of a conception and non-conception cycle with ovulation on day 14. It is apparent from this figure, and from examination of similar figures with a range of alternative ovulation days, that progesterone rises following ovulation in conception cycles, but peaks and then drops in non-conception

cycles. This result is consistent with the biological role of progesterone and with previous findings of Baird et al. (1997). In addition, these population curves support our findings that conception cycles tend to have lower pre-ovulatory progesterone.

[Figure 4 about here.]

Figure 5 shows the population-average progesterone curves for non-conception cycles when ovulation occurred on the 10^{th} day of the cycle (early) and on the 18^{th} day of the cycle (late). The estimated curves are different, indicating that the model was adequate in discerning the effect of the timing of ovulation on progesterone. The fact that the peak occurs earlier when ovulation occurs earlier is consistent with previous findings about the relationship between progesterone and ovulation (Baird et al., 1997).

[Figure 5 about here.]

Finally, we examined the adequacy of the Laplace approximation to the marginal likelihood. Twenty model proposals were selected at equally-spaced intervals throughout the sampling period, and a plot of each true unnormalized marginal likelihood was compared to the scaled normal approximation. The fit was found to be quite good, although in general the approximation tends to have slightly fatter tails than the true likelihood. Figure 6 displays the likelihood from a birth proposal and its Laplace approximation.

[Figure 6 about here.]

6. Discussion

We proposed Bayesian regression with multivariate linear splines for hierarchical data. This is an extension of the method for independent responses (Holmes and Mallick,

2001) to include subject-specific basis coefficients assumed to be centered around the population coefficients for each of the sampled models.

A different method was developed independently by Verzilli et al. (2005) for predicting the functional consequences of amino acid polymorphisms. Their approach also relies on Bayesian multivariate adaptive regression splines, though they accommodate within-cluster dependency using a simple cluster-specific random intercept. A random intercept is not flexible enough to accommodate the variability in hormone trajectories, motivating our use of a general hierarchical structure for the basis coefficients.

Analysis of the NCEPS data has yielded new insight about the relationship between progesterone and cycle conception status. It has been speculated that very low midluteal PdG may be indicative of a low fertility cycle, and also that signals from the conceptus may promote a pre-implantation increase in progesterone. Our results support both of these hypotheses, as we found evidence for a slightly steeper post-ovulatory PdG rise in conception cycles.

Previous analyses of these and other data (Baird et al., 1997; Stewart et al., 1993) found that non-conception cycles have lower midluteal progesterone than conception cycles, but we found the opposite. However, these previous studies examined non-conception cycles from women of known fertility who were exposed to sperm during a potentially fertile phase of the cycle (either through intercourse or artificial insemination). These non-conception cycles were therefore likely to be of low fertility. A previous analysis of these data using non-conception cycles regardless of intercourse timing found that cycles with low midluteal progesterone were unlikely to be conception cycles (Baird et al., 1999). We found the same, but we also found that, on average, midluteal progesterone was higher in non-conception cycles.

In light of these previous results, our findings suggest that those cycles with very

low progesterone are of low fertility, but that high pre-ovulatory progesterone does not imply an increased probability of conception. Intercourse timing was not used as a covariate here, but it may be informative in future analyses to explicitly differentiate non-conception cycles that were due to lack of intercourse from those that were of low fertility.

The method was applied to longitudinal data, but it could be used in any hierarchical regression problem where the functional form of the relationship between the covariates and the response is unknown. The NCEPS data lends itself readily to a discussion of the incorporation of reference points, but this method is also appropriate for regression when there are no reference points of interest. In this sense, the regression model is widely applicable.

In addition, reference points are not unique to longitudinal data. Brumback and Lindstrom (2004) use reference points to line up features of speech pattern data. Functional data can also occur over space (Morris et al., 2003), in which case the reference points are spatial rather than temporal locations. Rice (2004) discusses the similarities among modeling longitudinal and other types of functional data. Often, the analysis goals are the same, and methods designed for one tend to apply to both. This method is readily applicable to hierarchical functional data such as that studied by Morris et al. (2003).

The Bayesian RJMCMC paradigm allowed estimation of a smooth function based on piecewise-linear models with unknown knots and estimation of the population regression function based on the subject-specific basis coefficients. Although we used piecewise linear splines for their interpretability, this methods could be applied with other basis sets (see Denison et al., 2002 for a discussion).

Acknowledgements

The authors thank Donna Baird, Clarice Weinberg, and Allen Wilcox for their helpful comments and for providing the NC-EPS data.

References

Baird, D., Weinberg, C., Zhou, H., Kamel, F., McConnaughey, D., Kesner, J. and Wilcox, A. (1999). Preimplantation urinary hormone profiles and the probability of conception in healthy women. *Fertility and Sterility* **71**, 40–49.

Baird, D., Wilcox, A., Weinberg, C., Kamel, F., McConnaughey, D., Musey, P. and Collins, D. (1997). Preimplantation hormonal differences between the conception and non-conception menstrual cycles of 32 normal women. *Human Reproduction* **12**, 2607–2613.

Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *JASA* **93**, 961–976.

Brumback, L. and Lindstrom, M. (2004). Self modeling with flexible, random time transformations. *Biometrics* **60**, 461–470.

Denison, D., Holmes, C., Mallick, B. and Smith, A. (2002). *Bayesian methods for nonlinear classification and regression*. John Wiley and Sons, Chichester, West Sussex, England.

DiCiccio, T., Kass, R., Raftery, A. and Wasserman, L. (1997). Computing Bayes factors by combining simulations and asymptotic approximations. *JASA* **92**, 903–915.

Dunson, D., Chen, Z. and Harry, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* **59**, 521–530.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian

model determination. *Biometrika* **82**, 711–732.

Holmes, C. and Mallick, B. (2001). Bayesian regression with multivariate linear splines. *JRSS-B* **63**, 3–17.

Holmes, C. and Mallick, B. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. *JASA* **98**, 352–368.

Lipson, S. and Ellison, P. (1996). Comparison of salivary steroid profiles in naturally occurring conception and non-conception cycles. *Human Reproduction* **11**, 2090–2096.

Massafra, C., De Felice, C., Agnusdei, D., Gioia, D. and Bagnoli, F. (1999). Androgens and osteocalcin during the menstrual cycle. *Journal of Clinical Endocrinology and Metabolism* **84**, 971–974.

Morris, J., Vannucci, M., Brown, P. and Carroll, R. (2003). Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *JASA* **98**, 573–583.

Percy, D. (1992). Prediction for seemingly unrelated regressions. *JRSS-B* **54**, 243–252.

Rice, J. (2004). Functional and longitudinal data analysis: perspectives on smoothing. *Stat. Sinica* **14**, 613–629.

Stewart, D., Overstreet, D., Nakajima, S. and Lasley, B. (1993). Enhanced ovarian and steroid secretion before implantation in early human pregnancy. *J. Clinical Endocrinology and Metabolism* **76**, 1470–1476.

van Zonneveld, P., Scheffer, G., Broekmans, F., Blankenstein, M., de Jong, F., Looman, C., Habbema, J. and te Velde, E. (2003). Do cycle disturbances explain the age-related decline of female fertility? Cycle characteristics of women aged over 40 years compared with a reference population of young women. *Human Reproduction* **18**, 495–501.

Verzilli, C., Whittaker, J., N., S. and D., C. (2005). A hierarchical Bayesian model for

predicting the functional consequences of amino-acid polymorphisms. *JRSS-C* **54**, 191–206.

Wilcox, A., Weinberg, C., O'Connor, J., Baird, D., Schlatterer, J., Canfield, R., Armstrong, E. and Nisula, B. (1988). Incidence of early loss of pregnancy. *New England Journal of Medicine* **319**, 189–194.

Zhang, D., Lin, X., Raz, J. and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *JASA* **93**, 710–719.

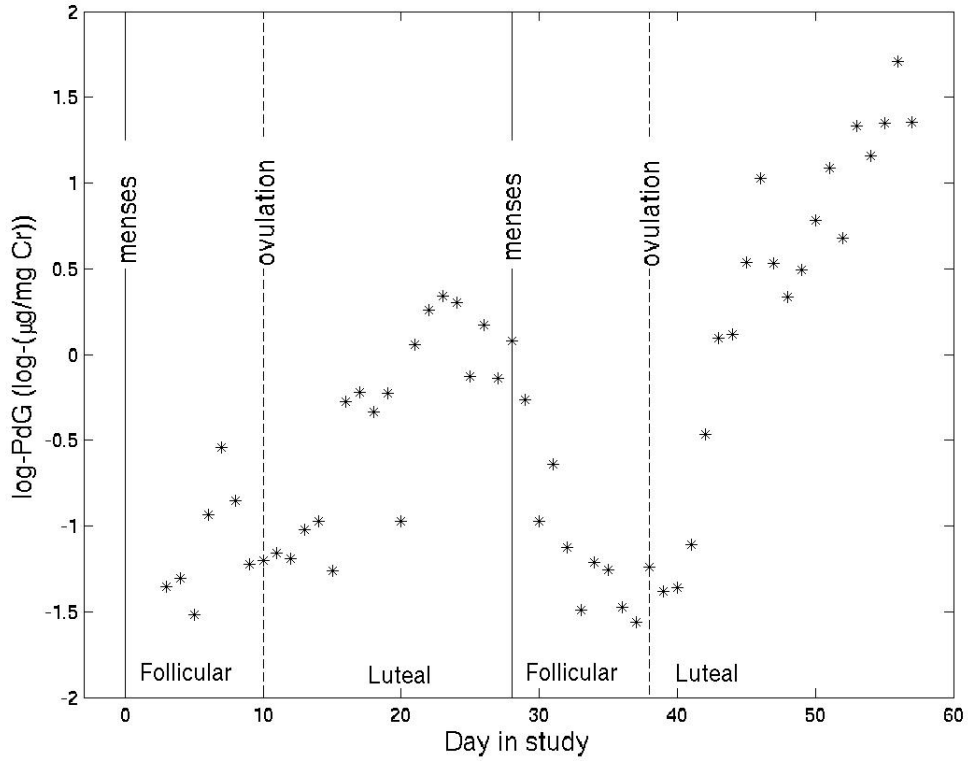

Figure 1. log-PdG for a non-conception followed by a conception cycle from one subject. Solid lines indicate first day of each cycle, and dashed lines indicate ovulation days.

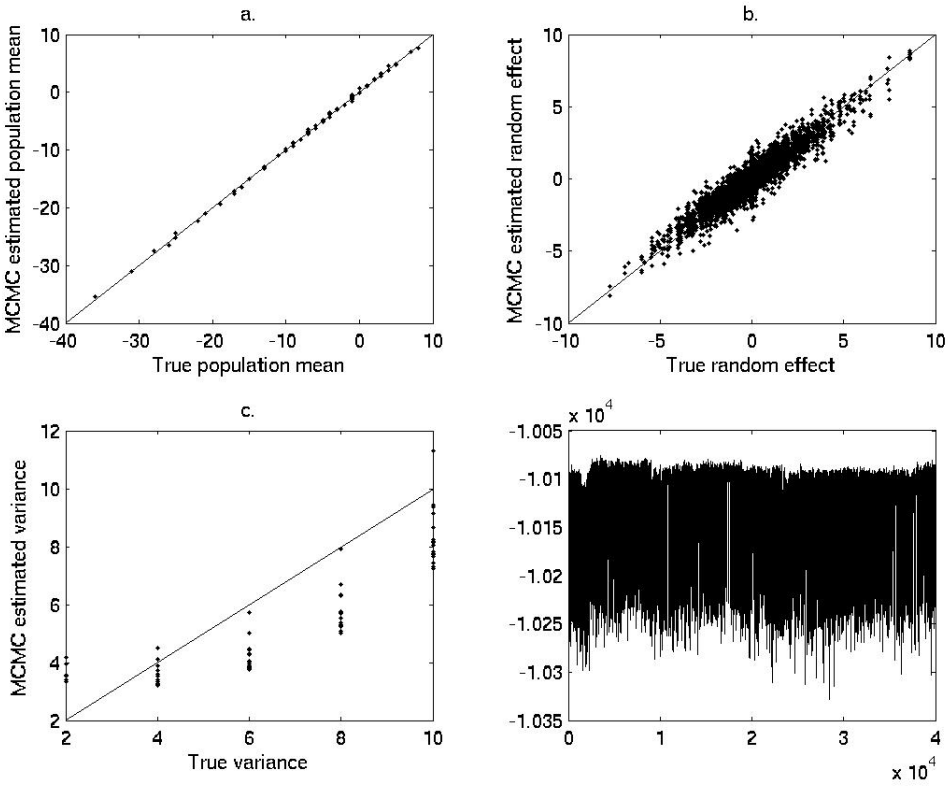

Figure 2. Evaluation of algorithm performance using simulated data

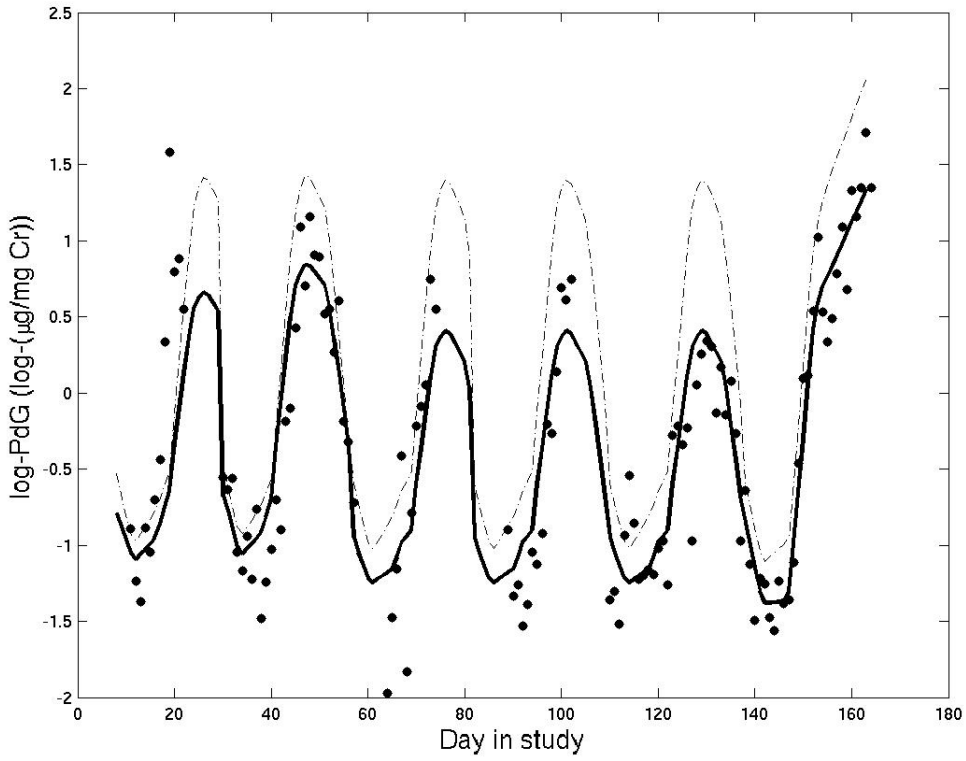

Figure 3. log-PdG data (points) and estimated log-PdG (solid line) for a single woman. The dashed line is the estimated population mean log-PdG given her covariates.

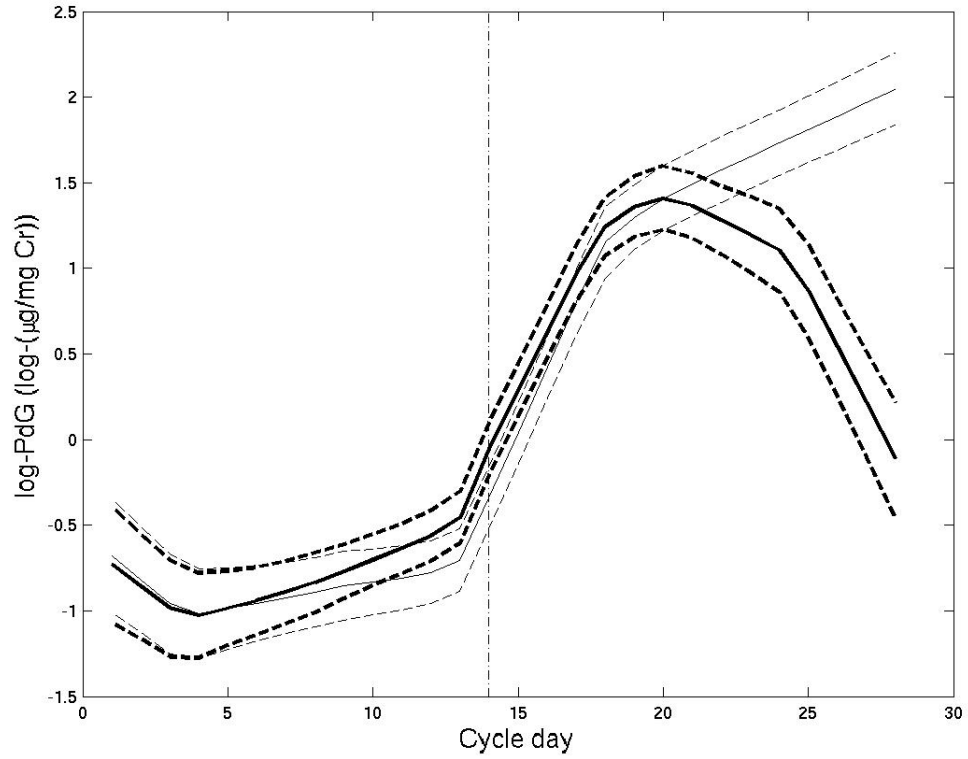

Figure 4. Estimated population mean log-progesterone for a conception (thin line) and non-conception (heavy line) cycle with ovulation on day 14. Pointwise 95% credible intervals are given by the dashed lines.

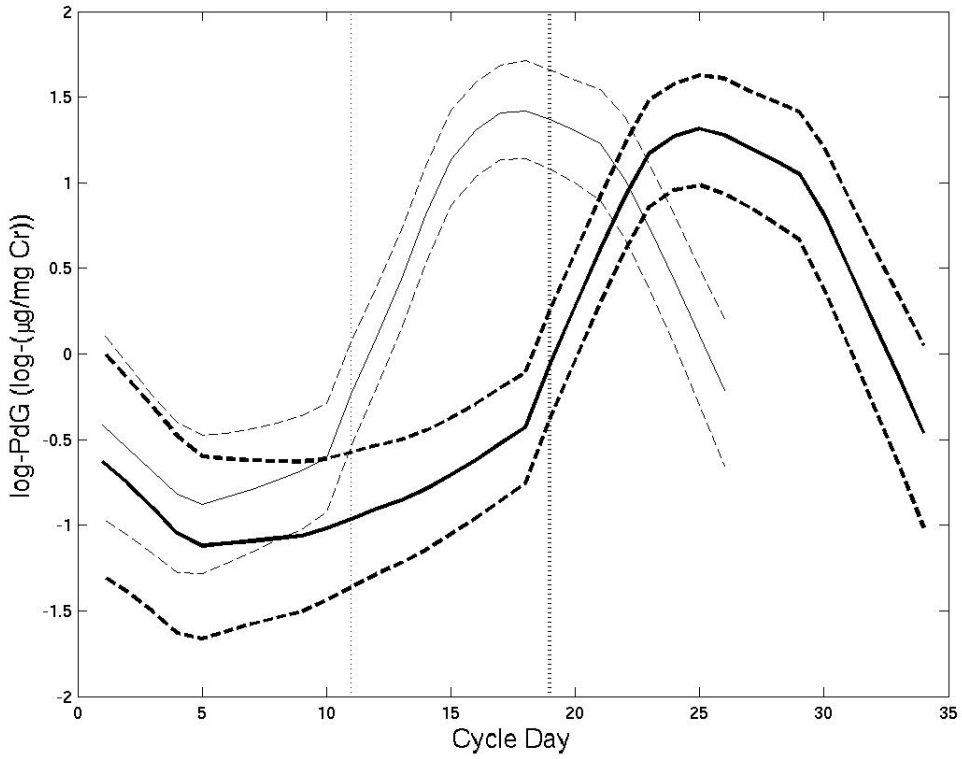

Figure 5. Estimated population mean log-progesterone for non-conception cycles with ovulation on day 10 (thin line) and day 18 (heavy line). Pointwise 95% credible intervals are given by the dashed lines. Vertical lines indicate ovulation days.

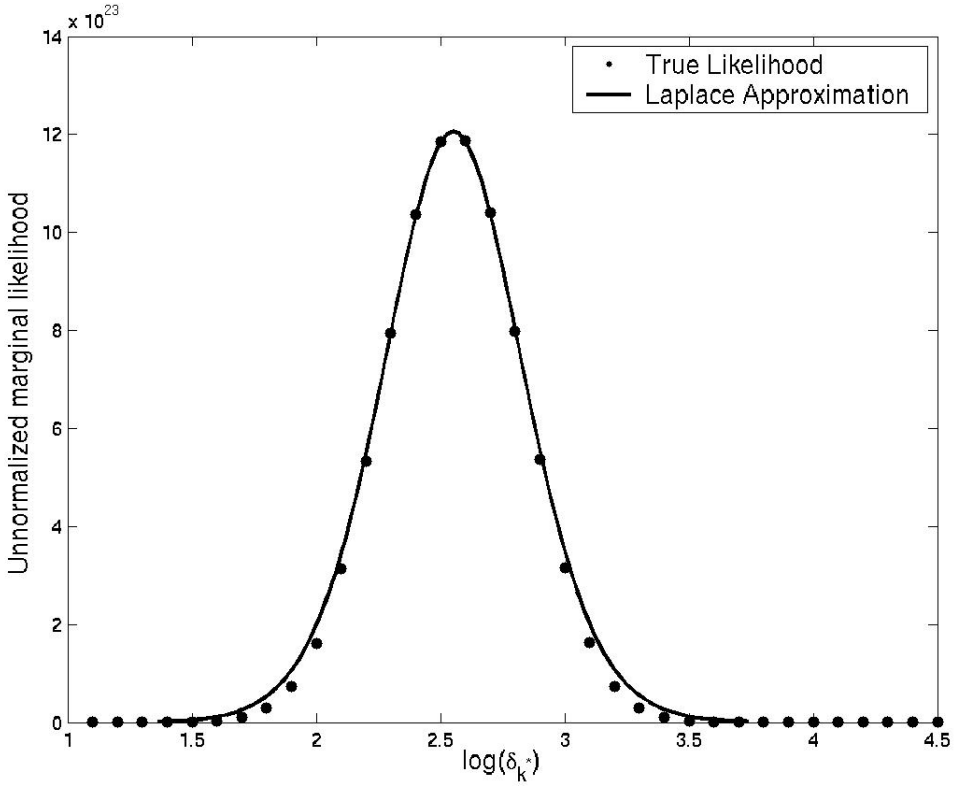

Figure 6. The unnormalized marginal likelihood for a proposed model, $p(\mathbf{y}|\lambda, \boldsymbol{\delta}, M^*)$ and its corresponding Laplace approximation.

Table 1

Summary variables describing the conception vs. non-conception difference in log-PdG across the menstrual cycle with 95% credible intervals. Estimates are based on an average of subject-specific trajectories at each iteration.

	log-PdG		
	Conception	Non-conception	Difference
Early follicular (days 1-5 of cycle)	-0.94 [-0.98, -0.90]	-0.64 [-0.67, -0.61]	-0.30 [-0.35,-0.25]
Baseline (2-6 days before ov.)	-0.94 [-0.97, -0.91]	-0.78 [-0.80, -0.77]	-0.16 [-0.19,-0.12]
Midluteal (5-6 days after ov.)	1.19 [1.14, 1.24]	1.31 [1.28, 1.35]	-0.13 [-0.18,-0.07]
Early luteal rise (days 1-5 after ov.)	1.18 [1.10, 1.26]	1.12 [1.06, 1.18]	0.07 [-0.05,0.15]

Table 2

Probability of conception in cycles with very low vs. normal/high midluteal (days 5-6 after ovulation) PdG

	Estimate, 95% credible interval
Probability of conception, midluteal PdG $< 10^{th}$ percentile	0.144 [0.098,0.195]
Probability of conception, midluteal PdG $\geq 10^{th}$ percentile	0.217 [0.211,0.222]
Difference in conception probabilities	0.073 [0.016,0.124]