Modelling menstrual cycle length and variability at the approach of menopause by using hierarchical change point models

Xiaobi Huang, Michael R. Elliott and Siobán D. Harlow

University of Michigan, Ann Arbor, USA

Summary. As women approach menopause, the patterns of their menstrual cycle lengths change. To study these changes, we need to model jointly both the mean and the variability of cycle length. Our proposed model incorporates separate mean and variance change points for each woman and a hierarchical model to link them together, along with regression components to include predictors of menopausal onset such as age at menarche and parity. Additional complexity arises from the fact that the calendar data have substantial missingness due to hormone use, surgery and failure to report. We integrate multiple imputation and time-to-event modelling in a Bayesian estimation framework to deal with different forms of the missingness. Posterior predictive model checks are applied to evaluate the model fit. Our method successfully models patterns of women's menstrual cycle trajectories throughout their late reproductive life and identifies change points for mean and variability of segment length, providing insight into the menopausal process. More generally, our model points the way towards increasing use of joint mean–variance models to predict health outcomes and to understand disease processes better.

Keywords: Change point model; Final menstrual period; Multiple imputation; Treloar Minnesota data

1. Introduction

Menstrual cycles are the most easily observed markers of ovarian function throughout female reproductive life. The menopausal transition is increasingly recognized to be a critical period in women's lives, as physiologic changes and health practices that are adopted during this period frequently define women's long-term chronic disease risk (Avis *et al.*, 2004; Sowers *et al.*, 2006). Given women's increasing longevity and recognition of the critical interface between ovarian and chronological aging, several proposals for staging reproductive aging have emerged. The Stages of Reproductive Aging Workshop recommendations (Soules *et al.*, 2001), its modifications (Harlow *et al.*, 2007) and several other proposals (Mitchell *et al.*, 2000; Taffe and Dennerstein, 2002a; Mansfield *et al.*, 2004) define criteria primarily by menstrual bleeding characteristics to determine the onset of the transition, as well as the stages within the transition period.

Although many studies have obtained data on menstrual cycle length (Matsumoto *et al.*, 1962; Miolo *et al.*, 1993; Marcus *et al.*, 2000; Colombo *et al.*, 2006), only two menstrual calendar studies have attempted to collect data over a woman's adult reproductive life (Treloar *et al.*, 1967;

Vollman, 1977). Treloar (1981) was the first to define the menopausal transition, estimating age at entry into the transition by visual inspection of menstrual cycle lengths for the 12-year period before the final menstrual period (FMP). He observed that during the menopausal transition longer intervals become mixed with shorter-than-usual intervals, increasing the variability in cycle length. He defined onset of the menopausal transition as the age at which variability in cycle length visually increased, and he estimated the median age of entry into the transition at 45.5 years with a median duration of transition of 4.8 years. Brambilla *et al.* (1994) introduced the term 'late perimenopause' and defined women as being in the late stage of the transition by self-report of 3–9 months of amenorrhoea (lack of menstrual cycles) or menstrual irregularity. Subsequently, investigators from several longitudinal studies (e.g. the Melbourne Women's Midlife Health Project (Dennerstein *et al.*, 1993), the Seattle Midlife Women's Health Study (Mitchell *et al.*, 2000) and Treloar Minnesota (TREMIN) study (Treloar *et al.*, 1967)) proposed various bleeding criteria to define the transition period (Taffe and Dennerstein 2002a; Mitchell *et al.*, 2000; Mansfield *et al.*, 2004; Lisabeth *et al.*, 2004b).

The Stages of Reproductive Aging Workshop divided reproductive life before menopause into the reproductive years (three stages) and the transition years (two stages: early and late transition). Entry to the early transition is characterized by increased variability in menstrual cycle length whereas entry into the late transition is characterized by the occurrence of skipped cycles or amenorrhoea. The multistudy ReSTAGE Collaboration subsequently evaluated bleeding criteria that served as the basis of the Stages of Reproductive Aging Workshop recommendations and documented the extent to which the various proposed criteria identified a similar moment in women's reproductive life and were predictive of the FMP (Harlow *et al.*, 2006, 2007, 2008). All of these proposals attempt to define bleeding criteria that identify a change point in women's menstrual cycle histories. Notably, however, none of the references attempted to model these change points longitudinally. Modelling menstrual cycle data has been an active research area lately. One of the first attempts was Harlow and Zeger (1991), which considered the marginal distribution of the menstrual cycle lengths as consisting of a normally distributed 'standard' cycle and 'non-standard' cycles from an undefined distribution, where non-standard was defined as more than 43 days; Guo *et al.* (2006) extended this to a latent mixture of normal and shifted Weibull distributions. Harlow *et al.* (2000) longitudinally modelled the change in mean cycle length, as well as in between-woman and within-woman variance across the reproductive lifespan and found that within-woman heterogeneity in cycle length was an important source of variation in menstrual patterns, especially after age 40 years. Lisabeth *et al.* (2004b) used generalized estimating equations to model changes in mean cycle length and variance independently of the mean referenced to age at FMP and demonstrated that variance in menstrual cycle lengths increases on average 2–6 years before increases in the mean, depending on age at FMP. Bortot *et al.* (2010) developed a hierarchical state space model that captured skewness in the cycle lengths via an auto-regressive moving average ARMA(1,1) model.

Prior descriptive analyses suggest that there is heterogeneity in women's menstrual trajectories. A prior analysis of the TREMIN data by Wallace *et al.* (1979) reported that women with later menopause had longer mean cycle length and greater variability 2 years before menopause than women with earlier menopause. Lisabeth *et al.* (2004a) in a longitudinal analysis of the same data also reported that longer cycles were associated with a later age of menopause. Another study (Den Tonkelaar *et al.*, 1998) reported that women with a late age at menopause (55–59 years) had a longer mean cycle length in the 9 years before menopause than women with an earlier menopause. Weinstein *et al.* (2003) found that low serial irregularity, which is a measure of the variability of the changes in cycle length, was associated with younger age at FMP, after adjusting for age at menarche, number of births and hormone use.

Our goal is to model how menstrual cycle length and its variability change when women approach menopause. In particular, we wish to estimate specific change points in the mean and variance growth curves that can serve as better well-defined, model-based indicators of the start of early and late menopausal transition than the summary statistics that were previously proposed by the ReSTAGE Collaboration. We assume that there are underlying unknown mean and variance change points for each individual woman and build a hierarchical change point model to estimate distributions of these change points. Furthermore, we impute cycle lengths (bleeding segments) that are missing because of hormone use, gaps in the menstrual calendar and gynaecological surgery, allowing more subjects and information to be included. Most prior reports have censored menstrual calendars when women began to use hormonal contraceptives or hormone therapy (Weinstein *et al.*, 2003; Guo *et al.*, 2006; Harlow *et al.*, 2006, 2008), or limited study participants to natural family planning users (Colombo *et al.*, 2006; Bortot *et al.*, 2010).

Statistically, the objective is to model both the mean and the variance of a set of curves. Several approaches have been proposed for correlated functional data of this type, including the bipartite spline model that was proposed by Harlow *et al.* (2000), which modelled the mean and between-subject variance by a linear random-effect model and used a two-stage log-linear regression to study within-subject variance *versus* age. Crainiceanu *et al.* (2007) proposed penalized splines to model both the mean and the variance by using a set of fixed knots for the splines with structural covariance matrix and random effects to depict the heterogeneity of variance. To model student test achievement, Thum and Bhattacharya (2001) proposed a hierarchical regression model which included a two-phase composite of $y_i \sim N(\beta_{01} + \beta_{11}x_i, \sigma_1^2)$, $i = 1, 2, \ldots, k$, and $y_i \sim N(\beta_{02} + \beta_{12}x_i, \sigma_2^2)$, $i = k+1, k+2, \ldots, n$, where k was the unknown change point. Hall *et al.* (2003) used unknown change points for the splines to capture individual cognitive function over time. These approaches estimated unknown change points for the mean but did not model the variance function over time. Davidian and Carroll (1987) proposed another approach for variance function estimation, which models the variance as proportional to a power of the mean response. This approach builds a separate function to model variance but did not include change points. Here we consider a hierarchical model that estimates individual mean and variance profiles with unknown change points. These change points represent measures of menopausal transition, and, together with intercepts and pre- and post-change point slopes, provide detailed summaries of the menstrual cycle data that can be related to individual level covariates such as age at menarche, parity and secular cohort membership. Our approach contrasts with that of Bortot *et al.* (2010), who did not estimate full individual level parameters for the women, given their more limited follow-up time in their data set (Miolo *et al.*, 1993), and for similar reasons did not develop parameters to focus on the menopausal transition.

Our paper is organized as follows. In Section 2 we describe the TREMIN study data. In section 3 we describe a hierarchical model to study the trajectories of women's menstrual cycle length that estimates unknown change points for both means and variances and allows these change points to be functions of subject level covariates. In addition, we identify different forms of missingness in the data set and incorporate imputation in the Markov chain Monte Carlo (MCMC) sampling that is used to estimate model parameters. In Section 4 we present results from the model fitted to menstrual data, along with Bayesian posterior predictive model checks; we also examine how our model results predict age at FMP, and we discuss what insight this information provides about underlying physiological mechanisms involved in menopausal transition. In Section 5 we discuss how our results compare with and extend previous menstrual cycle staging research, along with possible extensions of our model.

2.   The Treloar Minnesota data set

Our models use TREMIN data, one of only two available data sets providing individual women's menstrual calendar data across nearly all of their adult reproductive life span. The study, which was initiated by Dr Alan Treloar (Treloar *et al.*, 1967), recruited the first cohort of TREMIN, 2350 college-aged women attending the University of Minnesota, between 1934 and 1939.

Definitions recommended by the World Health Organization (Belsey and Farley, 1987) were used to summarize the calendar data. A bleeding segment, which is analogous to the term menstrual cycle, is a period of consecutive bleeding days and the subsequent bleeding-free days. Bleed-free intervals had to consist of at least 3 days; 1–2 bleed-free days between 2 bleeding days were considered part of the bleeding episode. The bleeding segment length is the dependent variable in our study. Age at menopause is determined by the date of the FMP, which is attributed retrospectively after 12 months of amenorrhoea on the calendar cards (World Health Organization, 1996).

We used data from 617 women in the 1935–1939 cohort restricted to those who

- (a) were aged 25 years or less at enrolment,
- (b) used hormones for less than 4 years continuously,
- (c) had at least one observed segment before age 40 years and
- (d) were not censored before age 40 years.

We consider segment lengths beginning at age 35 years. After this left truncation, the data set has a total of 95 246 observed menstrual segment records. Each record consists of woman's age, bleeding segment length and status indicators for pregnancy, hormone use and surgery. Related subject level information including age at menarche and parity are also available.

Pregnancy intervals as well as the first two segments after a birth and the first segment after a spontaneous abortion are coded as non-menstrual intervals. Many women used exogenous hormones at some point during their reproductive lives, mainly as hormonal replacement therapy. When hormones are used, ovarian function is masked during the bleeding segment. Thus, the segment data are considered to be missing when women use hormones. A one-segment washout period after hormone use had ended was also treated as missing. Many studies of menstrual characteristics censor women when they begin hormone use or ignore the time period during which women are using hormones. However, Wegienka and Baird (2003) suggested that these strategies may introduce bias since hormone users are not a random sample of menstruating women. Omitting these women or portions of their data will provide an incomplete description of experiences in the overall population. In our analysis, we consider these data as missing and impute their values for hormone use gaps of up to 4 years. Studies have not found that hormonal use influences menstrual segment length after stopping use and allowing for a washout period (Treloar and Behn, 1971; Nassaralla *et al.*, 2011).

The 617 women who were included in our analysis contributed between 15 and 321 non-missing segments to the analysis. The observed segment lengths vary from 4 to 366 days with a median of 27 days. Final menstruation periods were observed for 313 subjects (50.7%). Only 105 (17.0%) have complete data.

3.   Modelling menstrual cycle data

We construct a hierarchical change point model for the mean and variance of the segment length. Guided by both data and substantive or biological considerations, we create a single change point linear spline for each mean and variance model. Biologically, we understand
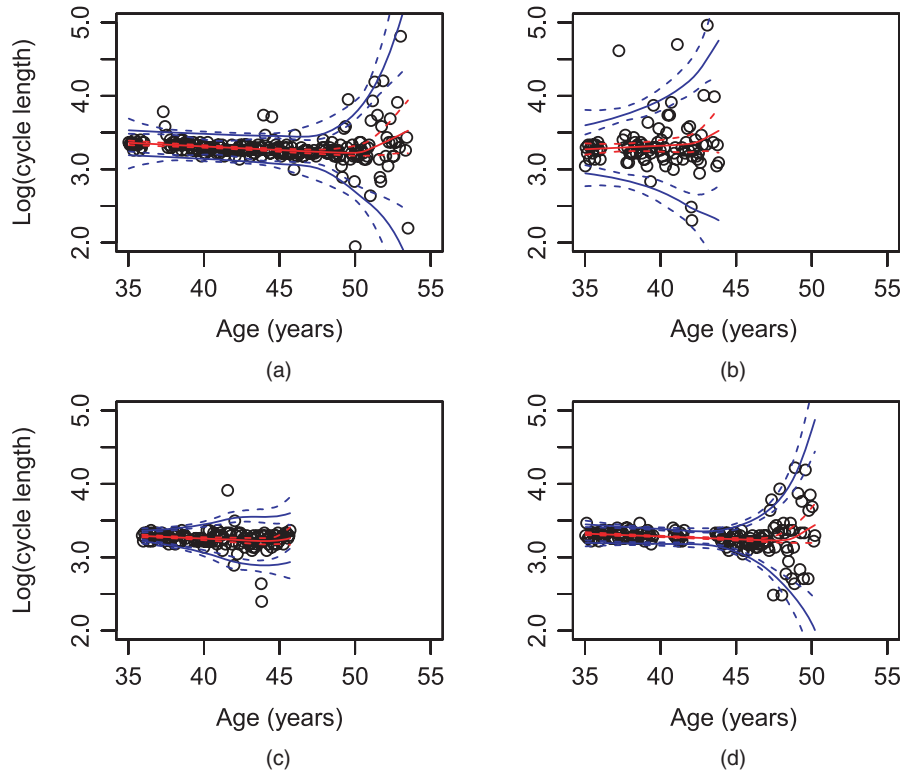
Fig. 1.   Change point model applied to the data for four women (———, posterior mean of the mean segment length; – – – –, associated 95% credible interval; ———, posterior means for the upper and lower 2.5-percentiles for the segment; – – – –, associated 95% credible intervals; ○, logarithm of observed segment lengths): subject (a) has complete data with a single pregnancy gap; subject (b) has missing data because of hormone use from age 36.07 to age 37.24 years; subject (c) was truncated because of a hysterectomy at age 45.78 years and thus has no observed FMP; subject (d) has intermittent missing data at age 39.59 years and from age 41.56 to 43.64 years, with hormone therapy after age 50.21 years and thus no observed FMP

the change in menstrual cycle length to be approximately continuous, rather than subject to dramatic step functions; substantively, our goal is to determine alternative measures of the start of early and late menopausal transition, corresponding to the change points in variance and mean respectively. On the basis of a visual inspection of the data (Fig. 1), such a choice seems reasonable; more formal model fit considerations are employed in Section 4.3.

Let y_{it} denote the tth menstrual segment length of subject i. Let a_{it} denote the age at the beginning of the tth menstrual segment of subject i, where $i = 1, \ldots, N$, $t = 1, \ldots, T_i$ and $N = 617$. To accommodate skewness and changing variance in the cycle length, we consider a heteroscedastic log-normal model with a subject level linear change point for both the mean and the variance:

$$\log(y_{it})|\mu_{it}, \sigma_{it}^2 \sim N(\mu_{it}, \sigma_{it}^2),$$ (1)

$$\mu_{it} = \alpha_{0i} + \alpha_{1i}(a_{it} - 35) + \alpha_{2i}(a_{it} - \alpha_{3i})_+$$

$$\log(\sigma_{it}^2) = \beta_{0i} + \beta_{1i}(a_{it} - 35) + \beta_{2i}(a_{it} - \beta_{3i})_+$$

The function $(x)_+ = x$ if $x \geqslant 0$ and $(x)_+ = 0$ if $x < 0$; α_{3i} and β_{3i} are the unknown change points of the mean and variance for subject i. The intercept for the mean of log-segment-lengths at

age 35 years for subject i is given by α_{0i}, whereas β_{0i} is the intercept for the log-variance of log-segment-lengths at age 35 years for subject i. The annual change rate in mean log-segment-lengths of subject i before the change point is denoted by α_{1i}, whereas β_{1i} denotes the annual change rate in the log-variance of log-segment-lengths of subject i before the change point. The additional annual change rate in mean log-segment-lengths of subject i after the mean change point is given by α_{2i}, whereas β_{2i} gives the additional annual change rate in log-variance of log-segment-lengths of subject i after the variance change point. We denote these eight subject level parameters for each woman as $\Phi_i = (\alpha_{0i}, \alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \beta_{0i}, \beta_{1i}, \beta_{2i}, \beta_{3i})'$.

To link the subject level models, we postulate a multivariate normal prior for the subject level parameters:

$$\Phi_i \overset{\text{ind}}{\sim} N(x_i'\Lambda, \Omega)$$

where x_i are covariates that are associated with subject i. Thus Λ and Ω can also be considered as population level parameters, with Λ as the regression coefficients and $\Omega \otimes I_N$ as the covariance matrix for the regression of Φ_i on x_i.

Finally, we utilize a Bayesian estimation framework for ease of estimation and interpretation. Thus we complete the model specification by postulating a flat hyperprior for Λ $p(\Lambda) \propto 1$ and a weakly informative inverse Wishart hyperprior with 1 degree of freedom and an identity scale matrix for Ω: $p(\Omega) \propto |\Omega|^{-(k+2)/2} \exp -\{\frac{1}{2}\text{tr}(\Omega^{-1})\}$, where k is the dimension of Ω.

3.1. Posterior inference

Let $z_{it} = \log(y_{it})$. The goal of our analysis is to obtain inference on the joint posterior distribution of Φ, Λ and Ω conditionally on the observed data z^{obs}. The posterior based on the complete data z is given by

$$
\begin{aligned}
p(\mathbf{\Phi}, \Lambda, \Omega | \mathbf{z}) &\propto \prod_{i=1}^{N} \left\{ \prod_{t=1}^{T_i} p(z_{it}|\Phi_i)\, p(\Phi_i|\Lambda, \Omega) \right\} p(\Lambda, \Omega) \\
&\propto \left(\prod_{i=1}^{N} \left[\prod_{t=1}^{T_i} \frac{1}{\sigma_{it}} \exp\left\{ -\frac{(z_{it}-\mu_{it})^2}{2\sigma_{it}^2} \right\} \right] |\Omega|^{-1/2} \exp\left\{ -\frac{1}{2}(\Phi_i - x_i'\Lambda)'\Omega^{-1}(\Phi_i - x_i'\Lambda) \right\} \right) \\
&\quad \times |\Omega|^{-(k+2)/2} \exp\left\{ -\frac{1}{2}\text{tr}(\Omega^{-1}) \right\} \\
&= \left(\prod_{i=1}^{N} \prod_{t=1}^{T_i} \sigma_{it}^{-1} \right) |\Omega|^{-(N+k+2)/2} \exp\left[\sum_{i=1}^{N} \left\{ \sum_{t=1}^{T_i} \frac{(z_{it}-\mu_{it})^2}{\sigma_{it}^2} + (\Phi_i - x_i'\Lambda)'\Omega^{-1}(\Phi_i - x_i'\Lambda) \right\} \right. \\
&\quad \left. + \text{tr}(\Omega^{-1}) \right]
\end{aligned}
$$

where $\mu_{it} = \alpha_{0i} + \alpha_{1i}(a_{it} - 35) + \alpha_{2i}(a_{it} - \alpha_{3i})_+$, $\sigma_{it}^2 = \exp\{\beta_{0i} + \beta_{1i}(a_{it} - 35) + \beta_{2i}(a_{it} - \beta_{3i})_+\}$ and $k = \dim(\Omega) = 8$. We sample the parameters via an MCMC algorithm that uses Metropolis-within-Gibbs sampling. Details of the procedure are in Appendix A. Missing data are imputed under a missingness at random assumption that assumes that missingness is random conditional on the observed cycle values (Little and Rubin, 2002) by using a standard selection model. (Although hormone users and women with other forms of missingness might not be a random sample of women, it is more reasonable to assume that, *for a given woman*, the missing cycles are a random subset of all her cycles.) Imputation is embedded within the MCMC algorithm. Details are provided in the next section.

3.2. Imputation of missing data

The majority (512 of the 617 women) have some form of missing data. For 302 women, their segment lengths are censored because of dropout while still menstruating, surgical termination of menstruation due to hysterectomy or bilateral oophorectomy or hormone use that began before the FMP and continued through to the end of the woman's calendar record. For the remaining 207 women, missingness was only intermittent. Intermittent missingness occurred due to sporadic non-reporting (women failing to report an individual segment or series of segments), or to periodic hormone use that stopped before one of the censoring events.

There is concern that missingness, particularly missingness due to hormone use, is not missingness completely at random. To deal with the different types of missingness, we impute the missing data under an assumption of missingness at random. To ensure that the imputation is proper (i.e. fully conditions on the observed data), we need to ensure that the imputed segment lengths sum to the length of the gap between observed segments. In addition, when censoring is present, we need to estimate the age at FMP to terminate the imputation process.

When missingness is intermittent, we ensure that the imputed missing segment lengths sum to the length of the gap by using an importance sampling algorithm. For notational simplicity, we assume that we have a single missing gap of length L_i for subject i, starting after segment y_{ik}. Conditional on Φ_i, the unobserved segment lengths $(y_{i,k+1}, \ldots, y_{i,k+S})' = \tilde{y}_i$ in the gap are independent, subject to the constraint that $\Sigma_{s=1}^{S} y_{i,k+s} = T$. We obtain a draw $\log(y_{i,k+1}^{\text{rep}}) \sim N(\mu_{i,k+1}, \sigma_{i,k+1}^2)$ where $\mu_{i,k+1} = \alpha_{0i} + \alpha_{1i}(a_{i,k+1} - 35)_+ + \alpha_{2i}(a_{i,k+1} - \alpha_{3i})_+$ and $\sigma_{i,k+1}^2 = \exp\{\beta_{0i} + \beta_{1i}(a_{i,k+1} - 35)_+ + \beta_{2i}(a_{i,k+1} - \beta_{3i})_+\}$ and $a_{i,k+1} = a_{ik} + y_{ik}$ is the age of the start of segment $y_{i,k+1}^{\text{rep}}$. A draw of $y_{i,k+2}^{\text{rep}}$ is then obtained as for $y_{i,k+1}^{\text{rep}}$, where now $a_{i,k+2} = a_{i,k+1} + y_{i,k+1}^{\text{rep}}$. This process is repeated until we obtain $y_{i,k+S}^{\text{rep}}$ such that $\Sigma_{s=1}^{S} y_{i,k+s}^{\text{rep}} > L_i$. We then replace $y_{i,k+S}^{\text{rep}}$ with $\tilde{y}_{i,k+S}^{\text{rep}} = L_i - \Sigma_{s=1}^{S-1} y_{i,k+s}^{\text{rep}}$. Let $(y_{i,k+1}^{(t)}, \ldots, y_{i,k+S-1}^{(t)}, \tilde{y}_{i,k+S}^{(t)}) = \tilde{y}_i^{(t)}$ be the tth vector of imputations, $t = 1, \ldots, 50$. Finally, we draw one of the 50 sets with probability $p^t = f(\tilde{y}_i^{(t)} | \Phi_i) / \Sigma_t f(\tilde{y}_i^{(t)} | \Phi_i)$, where

$$f(\tilde{y}_i^{(t)} | \Phi_i) = \prod_{s=1}^{S-1} \phi\left\{ \frac{\log(y_{i,k+s}) - \mu_{i,k+s}}{\sigma_{i,k+s}} \right\} \phi\left\{ \frac{\log(\tilde{y}_{i,k+S}) - \mu_{i,k+S}}{\sigma_{i,k+S}} \right\},$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution. On rare occasions where $y_{i,k+s}^{\text{rep}} < 4$, the imputed values were truncated to be 4; similarly $y_{i,k+s}^{\text{rep}} > 365$ was truncated to 365. These truncation values are set according to the definitions of 'bleeding segment' and 'FMP' recommended by World Health Organization (1996).

When subjects' segment lengths are censored, we need to impute an FMP since it is unobserved. As the FMP is analogous to 'event time' in survival analysis, we model the age at FMP Q_i as a piecewise exponential distribution with hazard $h_i(t) = \eta_k$ for $A_{k-1} \leqslant t < A_k$ for knots $k = 1, \ldots, K$. On the basis of the distribution of observed FMP values, knots are set at every 0.5–3 years between age 40 and 60 years, depending on the density of the observed FMPs; details are provided in Appendix B. Assuming a prior of the form $\eta_k \sim \text{gamma}(a, b)$, we obtain a draw from

$$p(\eta_k | Q) \sim \text{gamma}\left\{ \sum_i I(A_{k-1} \leqslant Q_i \leqslant A_k) + a, \sum_i I(Q_i \geqslant A_{k-1}) + b \right\}$$

for $k = 1, \ldots, K$, where Q includes both the observed FMP and those imputed at the previous iteration of the Gibbs sampler (see Appendix B). As in the intermittent missing data setting, we then obtain a draw of $\log(y_{i,T_i^{\text{rep}}+1}^{\text{rep}}) \sim N(\mu_{i,T_i^{\text{rep}}}, \sigma_{i,T_i^{\text{rep}}}^2)$ where

$$\mu_{i,T_i^{\text{rep}}} = \alpha_{0i} + \alpha_{1i}(a_{i,T_i^{\text{rep}}} - 35)_+ + \alpha_{2i}(a_{i,T_i^{\text{rep}}} - \alpha_{3i})_+,$$

$$\sigma^2_{i,T_i^{\text{rep}}} = \exp\{\beta_{0i} + \beta_{1i}(a_{i,T_i^{\text{rep}}} - 35)_+ + \beta_{2i}(a_{i,T_i^{\text{rep}}} - \beta_{3i})_+\},$$

$a_{i,T_i^{\text{rep}}} = a_{iT_i^{\text{rep}}-1} + y_{iT_i^{\text{rep}}}$ is the age of the start of segment $y^{\text{rep}}_{i,T_i^{\text{rep}}}$ and T_i^{rep} is the number of observed segments plus the number of imputed segments in any intermittent missing gaps. Let $W_{i,1}$ be an indicator for whether this first imputed segment is the FMP; we then obtain a draw $W_{i,1}$ from a Bernoulli distribution with probability (see Appendix B)

$$P\{a_{i,T_i^{\text{rep}}} \leqslant Q_i \leqslant a_{i,T_i^{\text{rep}}+1} | Q_i > \max(a_{i,T_i^{\text{rep}}}, \alpha_{3i}, \beta_{3i})\}$$

$$= \begin{cases} 1 - \exp\{-\eta_k(a_{i,T_i^{\text{rep}}+1} - a_{i,T_i^{\text{rep}}})\} & \text{if } A_{k-1} \leqslant a_{i,T_i^{\text{rep}}} < a_{i,T_i^{\text{rep}}+1} \leqslant A_k, \\ 1 - \exp\{\eta_{k-1}(a_{i,T_i^{\text{rep}}} - A_{k-1}) - \eta_k(a_{i,T_i^{\text{rep}}+1} - A_{k-1})\} \\ \qquad\qquad \text{if } A_{k-2} \leqslant a_{i,T_i^{\text{rep}}} \leqslant A_{k-1} \leqslant a_{i,T_i^{\text{rep}}+1} \leqslant A_k, \\ 1 - \exp\{\eta_{k-1}(a_{i,T_i^{\text{rep}}} - A_{k-1}) - \eta_k(A_k - A_{k-1}) - \eta_{k+1}(a_{i,T_i^{\text{rep}}+1} - A_k)\} \\ \qquad\qquad \text{if } A_{k-2} \leqslant a_{i,T_i^{\text{rep}}} \leqslant A_{k-1} < A_k \leqslant a_{i,T_i^{\text{rep}}+1} \leqslant A_{k+1}. \end{cases}$$

Note that the FMP must occur after both the last observed segment and the latent mean and variance change points; also, since none of our knots are less than 6 months apart, a segment can cover a maximum of three intervals. If $W_{i,1} = 1$, $y^{\text{rep}}_{i,T_i^{\text{rep}}+1}$ is the length of the final FMP. If $W_{i,1} = 0$, we draw $\log(y^{\text{rep}}_{i,T_i^{\text{rep}}+2}) \sim N(\mu_{i,T_i^{\text{rep}}+1}, \sigma^2_{i,T_i^{\text{rep}}+1})$ and repeat the process s times until one of the following occurs: $W_{i,s} = 1$, $y^{\text{rep}}_{i,T_i^{\text{rep}}+1} > 365$ or $a^{\text{rep}}_{i,T_i^{\text{rep}}+s} \geqslant 60$. For the vast majority of imputations, the imputed FMP indicator triggered the end of the segment imputation.

4. Results

We use the methodology that was described in Section 3 to analyse the TREMIN data by using MATLAB software. To assess convergence, we ran two MCMC chains for 10 000 iterations each after discarding the first 10 000 draws as 'burn-in'. We assessed convergence by using the Gelman and Rubin statistic (Gelman and Rubin, 1992), with a thinning interval of five segments. All of the population and 98% of the individual level parameters had a value of less than 1.2, indicating reasonable convergence.

4.1. Individual level parameters

To assess model fit at the individual level visually, Fig. 1 plots the logarithm of the observed segment lengths, the predicted means and the upper and lower 2.5-percentiles of the segment distribution for four typical women in the TREMIN data set. The model appears to capture the trajectories well, with approximately 5% of segment lengths excluded from the 95% predictive intervals. The uncertainty in the position of the variance change point is highlighted for subjects B and C.

Fig. 2 plots the posterior means and 90% credible intervals (CIs) of the mean and variance change points for 25 randomly selected women. As noted by Treloar (1981) and Lisabeth *et al.* (2004a), variability generally begins to increase before mean length. Subjects with earlier change points averaged 4–5 years between mean and variance change points, whereas subjects with later change points averaged only 1–2 years between mean and variance change points, which are consistent with the findings of Harlow *et al.* (2008). Uncertainty in the variance change points was on average greater than in the mean change points, although this was not uniformly true across women.
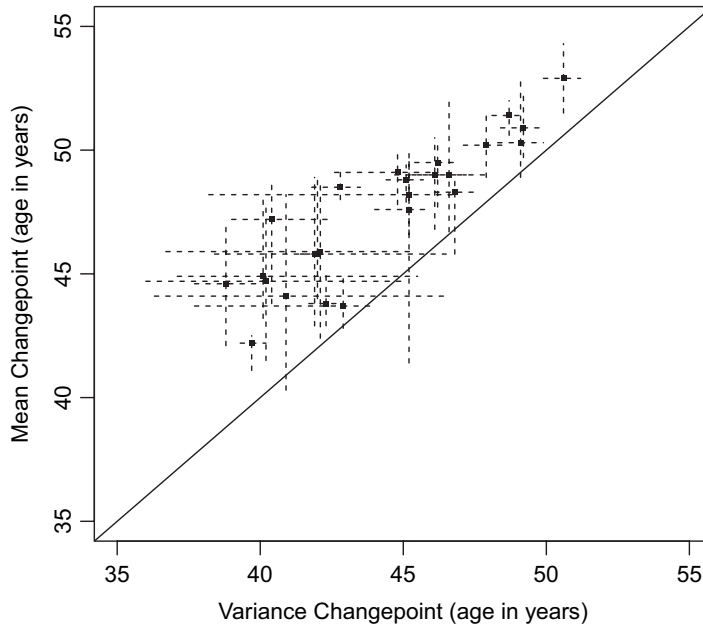
Fig. 2. Posterior means and 90% posterior predictive intervals for mean change points and variance change points (25 randomly selected women)

4.2. Population level parameters

Table 1 summarizes the posterior means and associated 95% CIs for the population level segment length mean and variance regression parameters. The population mean age at the change point for segment length means is 46.23 years (95% CI 45.91–46.55 years), older than the population mean age at change points for segment length variability, which is 42.84 years (95% CI 42.49–43.17 years); thus variability in segment length is predicted to begin increasing 3.39 years earlier (95% CI 3.07–3.74 years) in the population than the mean segment length itself. Mean segment length declined about 1% per year before the change point and increased about 15% per year afterwards. The variability of log-segment-length was stable before the change point and increased by 81% per year after the change point.

Table 2 presents the posterior mean and associated 95% posterior predictive interval for the correlation matrix corresponding to the covariance matrix Ω. The 95% CIs of correlations that exclude 0s are denoted in italics. We note the following results.

(a) Later change points for the variance are highly associated with later change points for the mean.

(b) Later change points for both the mean and the variance are also correlated with longer and more variable segment lengths, and more rapid increases in mean and variance after the change point; consequently mean and variance slopes after change points are positively correlated.

(c) Greater mean length at age 35 years is associated with greater declines in variability before the variance change point and greater increases in variability after.

(d) Larger segment variability is associated with longer mean segment length.

(e) Larger segment variability is highly associated with more rapid declines in variability

Table 1. Posterior mean of population level regression coefficients (Λ) estimates and associated 95% posterior predictive intervals†

Parameter	Results for log-normal model			Results for gamma model, MAR
	MAR	*Censored, FMP late*	*Censored, FMP early*	
Mean intercept (α_0)	3.313 (3.306,3.322)	3.313 (3.305,3.321)	3.313 (3.305,3.321)	3.310 (3.300,3.321)
Mean slope before change point (α_1)	−0.007 (−0.011,−0.003)	−0.007 (−0.010,−0.003)	−0.006 (−0.010,0.003)	0.004 (−0.001,0.009)
Mean slope after change point (α_2)	0.144 (0.128,0.162)	0.146 (0.130,0.163)	0.146 (0.130,0.162)	0.272 (0.240,0.306)
Segment length mean change point (α_3)	46.23 (45.91,46.55)	46.22 (45.91,46.51)	46.22 (45.93,46.51)	44.88 (44.50,45.27)
Log-variance intercept (β_0)	−4.795 (−4.907,−4.678)	−4.808 (−4.928,−4.695)	−4.811 (−4.920,−4.702)	−4.815 (−4.941,−4.691)
Log-variance slope before change point (β_1)	0.009 (−0.022,0.041)	0.020 (−0.009,0.049)	0.018 (−0.010,0.047)	−0.017 (−0.073,0.036)
Log-variance slope after change point (β_2)	0.595 (0.540,0.648)	0.575 (0.522,0.626)	0.588 (0.535,0.641)	0.532 (0.459,0.603)
Segment length variance change point (β_3)	42.84 (42.49,43.17)	42.84 (42.52,43.17)	42.90 (42.60,43.20)	42.21 (41.90,42.54)

†MAR, missingness at random, which assumes that age at FMP follows a common piecewise exponential distribution for all women; FMP late assumes that women with censored FMPs have a hazard rate for FMP that is reduced by 10% relative to women with observed FMPs; FMP early assumes that women with censored FMPs have a hazard rate for FMP that is increased by 10% relative to women with observed FMPs.

Table 2. Posterior means and 95% CIs of correlations of population level regression coefficients (corresponding to covariance matrix Ω)†

Parameter	Mean intercept	Mean slope before change point	Mean slope after change point	Mean change point	Log-variance intercept	Log-variance slope before change point	Log-variance slope after change point	Variance change point
Mean intercept	1.00 —	−0.13 (−0.21,−0.04)	−0.01 (−0.12,0.11)	0.29 (0.20,0.37)	0.17 (0.08,0.26)	−0.14 (−0.24,−0.05)	0.17 (0.08,0.26)	0.27 (0.18,0.35)
Mean slope before change point		1.00 —	−0.02 (−0.11,0.06)	0.00 (−0.08,0.09)	−0.07 (−0.15,0.02)	0.03 (−0.06,0.12)	−0.00 (−0.09,0.09)	0.01 (−0.08,0.09)
Mean slope after change point			1.00 —	0.25 (0.12,0.36)	0.08 (−0.04,0.21)	−0.08 (−0.21,0.03)	0.33 (0.20,0.44)	0.24 (0.09,0.37)
Mean change point				1.00 —	0.15 (0.05,0.24)	−0.25 (−0.37,−0.13)	0.43 (0.32,0.53)	0.79 (0.74,0.83)
Log-variance intercept					1.00 —	−0.69 (−0.74,−0.64)	0.44 (0.37,0.51)	0.02 (−0.07,0.10)
Log-variance slope before change point						1.00 —	−0.74 (−0.78,−0.70)	0.09 (−0.01,0.19)
Log-variance slope after change point							1.00 —	0.34 (0.25,0.42)
Variance change point								1.00 —

†CIs that exclude 0 are in italics.

before but larger increases in variability after the variance change point: thus the change in variability before and after the variance change point is negatively correlated.

We also fitted a two-covariate model, including parity and age at menarche. As these covariates showed no significant relationships with the eight parameters describing the menopausal transition, we do not show the results here.

4.3. Posterior predictive model check

We used posterior predictive distribution (PPD) checks (Gelman *et al.*, 1996) to assess model fit. The PPD '*p*-value' represents the probability that an observed statistic (which can be a function of both the data y and the parameter θ) is more extreme than the replicated statistic, conditional on the observed data: $P\{T(y, \theta) \leqslant T(y^{\mathrm{rep}}, \theta)|y\}$, where y^{rep} is drawn from the PPD

$$f(y^{\mathrm{rep}}|y) = \int f(y^{\mathrm{rep}}|\theta, y)\, p(\theta|y)\, \mathrm{d}\theta.$$

In contrast with frequentist *p*-values under the null hypothesis, PPD *p*-values do not have a uniform distribution when the model is correct; rather, they represent the degree to which data generated from the posterior distribution conflict with the observed data (Gelman, 2003, 2007). Thus values close to 0.5 are preferred. To assess the model fit for the segment length, we considered a χ^2 discrepancy statistic for observed segment lengths of each individual woman given by $T_1(y_i, \mu_i, \sigma_i^2) = \Sigma_t (e_{it}^*)^2$, where $e_{it}^* = \{\log(y_{it}) - \mu_{it}\}/\sigma_{it}$ is the standardized residual. $T_1(y_i^{\mathrm{rep}}, \mu_i, \sigma_i^2)|y$ has a $\chi^2_{T_i^{\mathrm{obs}}}$ distribution where T_i^{obs} is the total number of observed segments for the ith woman, so to obtain the PPD *p*-value we need to compute only $L^{-1}\Sigma_{l=1}^{L} I[T\{y_i, \mu_i^{(l)},$ $(\sigma^2)_i^{(l)}\} \leqslant (X^2)^{(l)}]$, where $\mu_i^{(l)}$ and $(\sigma^2)_i^{(l)}$ are drawn from the posterior distribution and $(X^2)^{(l)}$ is drawn from a $\chi^2_{T^{\mathrm{obs}}}$-distribution. On the basis of the $L = 500$ replications, no subjects had a posterior predictive *p*-value greater than 0.90 and only seven subjects had a posterior predictive *p*-value smaller than 0.10 (Fig. 3(a)). A review of subjects with low posterior predictive *p*-values showed that they contain one or two sporadic very short or very long segments well before the onset of the increase in variability, suggesting that these subjects contain outlying segment lengths rather than indicating more general model failure. Subjects with high posterior predictive *p*-values generally had relatively few observations with little variability—the variance estimates were smoothed back towards larger values, yielding small χ^2 discrepancy statistics.

To assess whether the proposed hierarchical model is sufficient to remove temporal correlation, we compute

$$T_2(\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \frac{\sum\limits_{i} \sum\limits_{t=1}^{T_i-1} \sum\limits_{s=t+1}^{T_i} e_{it}^* e_{is}^*}{\sqrt{\left\{ \sum\limits_{i} \left(\sum\limits_{t=1}^{T_i-1} \sum\limits_{s=t+1}^{T_i} e_{it}^* e_{is}^* \right)^2 \right\}}},$$

corresponding to the Wooldridge statistic for testing general residual serial correlation in panel data (Wooldridge, 2002). Here the PPD *p*-value for overall fit based on 500 replications is 0.28. Looking at the PPD *p*-values of individual level statistics $T_2(y_i, \mu_i, \sigma_i^2)$, we find that no subjects had a posterior predictive *p*-value greater than 0.90 and only one subject has a PPD *p*-value smaller than 0.10. Examination of the posterior distribution of semivariograms (Diggle, 1990) for the subjects with the most extreme PPD *p*-values shows no strong or consistent evidence of auto-correlation at any particular lagged value (the plots are not shown).

Finally, to consider the appropriateness of the final menstrual period modelling, we plot
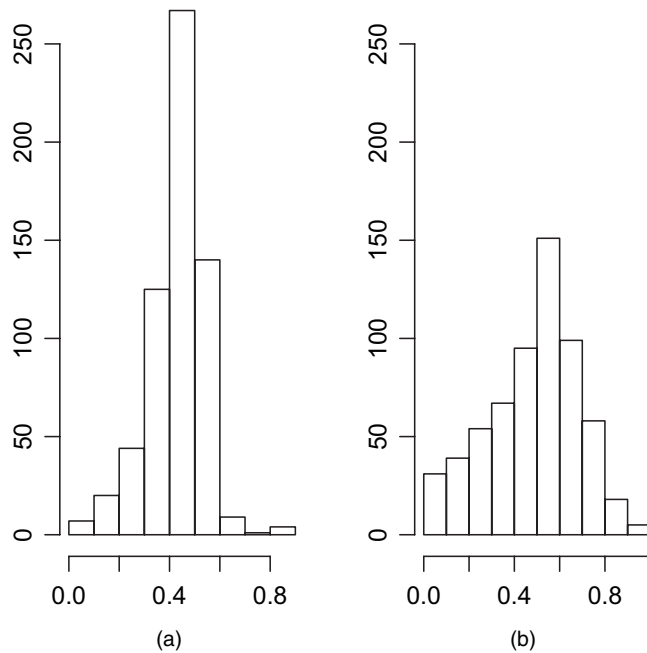
Fig. 3. Subject level PPD *p*-values for the χ^2 discrepancy statistic (note that values close to 0.5 are preferred): (a) log-normal model; (b) gamma model

the median and 95% posterior predictive interval for 50 women with censored cycle calendars against their censoring age for 50 randomly selected women with censored FMPs in Fig. 4. The method for estimating FMP when not observed appears to have worked reasonably well. The median of the predicted FMPs corresponds closely to the observed FMPs (observed median FMP age, 50.3 years; interquartile range IQR $=3.7$) with wide intervals when the censoring age is relatively early and little information is usually available to predict the FMP, whereas subjects with older censoring ages have later predicted medians and narrower intervals.

4.4. Prediction of final menstrual period

In this model, we used eight parameters to describe eight characteristics of an individual woman's menstrual pattern. As an example of how this model might be applied to further understanding of the menopausal transition, we consider the degree to which all eight of the menstrual pattern parameters predict age at FMP by regressing the individual level posterior means against the age at FMP in a two-stage approach. (As noted in Section 5, a two-stage approach is inferior to a fully joint model; however, the primary purpose of this section is to provide an example of how our model output can be used to inform understanding of the menopausal transition further.) To accommodate subjects with censored FMPs, we use a semiparametric accelerated failure time model (Jin *et al.*, 2006) which assumes that failure times are linearly related to the covariates while leaving the error distribution unspecified. All individual level posterior means of the eight characteristic parameters are standardized before analysis to facilitate comparisons of their influences of age at FMP.

Table 3 shows the associations between each characteristic and age at FMP, both bivariately and in a single multivariate model. Bivariately, age at the mean and variance change points have
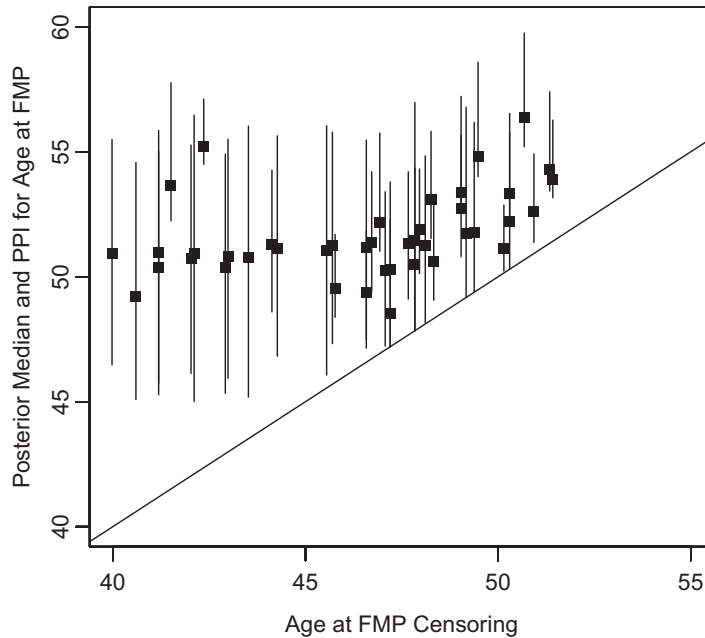
Fig. 4. Posterior medians (■) and 95% posterior predictive intervals for unobserved FMPs, by age at FMP censoring (50 randomly selected women with censored FMP values)

the largest influence on FMP, with a 1-standard-deviation increase in the mean change point associated with a 2.43-year (95% CI 2.24–2.62 years) increase in age at FMP and a 1-standard-deviation increase in the variance change point associated with a 1.88-year (95% CI 1.64–2.11 years) increase in age at FMP. Weaker but significant positive associations are associated with the mean and variance of segment length at age 35 years, whereas modest negative associations are observed with more rapid increases in mean after the mean change point, and less rapid decreases in variance before the variance change point. In the joint model, mean and variance change points' influence on FMPs decreases somewhat, whereas the influence of variance slope before and after change points substantially increases: women with a 1-standard-deviation less rapid decrease in variance before change points have a 1.93- (95% CI from −2.81 to −1.05) year decrease in age at FMP, whereas women with a 1-standard-deviation more rapid increase in variance after change points have a 1.91- (95% CI from −2.54 to −1.28) year decrease in age at FMP. The influence of mean segment lengths at age 35 years is no longer significant, whereas the effect of baseline variance is reversed, with a 1-standard-deviation increase in variance at age 35 years associated with a decreased age at FMP of 0.52 (95% CI from −0.86 to −0.19) years. The model is highly predictive of FMP, explaining 83.6% of the variance in the observed FMP ages.

These results indicate that age at FMP is largely explained by the timing and characteristics of the menopausal transition, confirming that accurate and precise identification of the age at transition would provide important information to women and clinicians which is relevant to contraceptive and surgical decision making and initiation of preventative care. These results also highlight the importance of population differences in cycle length variability in explaining differences in the duration of the menopausal transition and in women's transition experience. To date, most studies of the menopausal transition have described the population-average ex-

Table 3. Influence of each menstruation pattern characteristic on age of FMPs†

Parameter	Estimated effects (95% confidence interval)	
	Bivariate	Multivariate
Mean intercept	*1.12 (0.83,1.41)*	0.08 (−0.09,0.24)
Mean slope before change point	−0.46 (−1.07,0.16)	0.02 (−0.27,0.31)
Mean slope after change point	*−1.09 (−1.41,−0.76)*	*−0.61 (−0.82,−0.40)*
Log-variance intercept	*0.33 (0.00,0.65)*	*−0.52 (−0.86,−0.19)*
Log-variance slope before change point	*−0.61 (−0.97,−0.24)*	*−1.93 (−2.81,−1.05)*
Log-variance slope after change point	0.20 (−0.19,0.58)	*−1.91 (−2.54,−1.28)*
Segment length mean change point	*2.43 (2.24,2.62)*	*1.70 (1.11,2.30)*
Segment length variance change point	*1.88 (1.64,2.11)*	*1.24 (0.55,1.93)*

†Influences are assessed bivariately and in a single multivariate model. The results are from a semiparametric accelerated failure time model which included censored FMPs. 95% confidence intervals are provided. Numbers in italics represent significant associations.

perience with little attention on how this experience might differ between subgroups of women. Factors that are associated with age at FMP, although predictive on average, generally exhibit considerable dispersion, making them less useful in predicting age at FMP for individual women (Santoro *et al.*, 2007; Taffe and Dennerstein, 2002b). This analysis suggests that future research should focus on factors that influence menstrual cycle variance and timing of the transition.

4.5. Sensitivity to departures from assumptions

We consider the effect on our results of deviations from two of the assumptions that were outlined in Section 3: menstrual length distributional assumptions, and the missingness at random assumption for menstrual cycle length data.

A natural alternative to the log-normal distribution for the highly skewed menstrual length data is the gamma distribution:

$$y_{it}|\mu_{it}, \sigma^2_{it} \sim \text{gamma}(\mu_{it}, \sigma^2_{it}). \tag{2}$$

We use the mean–scale parameterization of the gamma distribution, so that

$$f(y_{it}; mu_{it}, \sigma^2_{it}) = \exp\left[\frac{1}{\sigma^2_{it}}\log\left(\frac{y_{it}}{\sigma^2_{it}\mu_{it}}\right) - \frac{y_{it}}{\sigma^2_{it}\mu_{it}} - y_{it} - \log\left\{\Gamma\left(\frac{1}{\sigma^2_{it}}\right)\right\}\right],$$

and use a log-link for both the mean and the scale so that the mean and scale parameters have similar interpretations in both the log-normal and the gamma model

$$\log(\mu_{it}) = \alpha_{0i} + \alpha_{1i}(a_{it} - 35) + \alpha_{2i}(a_{it} - \alpha_{3i})_+,$$
$$\log(\sigma^2_{it}) = \beta_{0i} + \beta_{1i}(a_{it} - 35) + \beta_{2i}(a_{it} - \beta_{3i})_+.$$

Similar hyperpriors were used; the MCMC algorithm that was used for the log-normal model was altered to accommodate the first-stage gamma distribution. Computation was considerably slower than for the log-normal model, since all subject level parameters required a Metropolis-within-Gibbs step, and convergence was considerably slower, with two MCMC chains requiring 100000 iterations each after 40000 draws as burn-in. All of the population and 95% of the individual level parameters had a Gelman–Rubin statistic value of less than 1.2.

The population level parameters governing the mean and scale under the gamma model are given in Table 1. Qualitatively, there is no strong evidence for either increases or decreases in population means and scales before change points; after change points, both increase. The increase in mean after the change point is about twice the rate (31%) under the gamma model than under the log-normal model. The population-average variability change point is approximately 0.6 years earlier in the gamma model (42.21 years) than in the log-normal model (42.84 years), whereas the population-average mean change point is approximately 1.4 years earlier in the gamma model (44.88 years) than in the log-normal model (46.23 years). The differences between the change points under the gamma model and under the log-normal model are clinically significant; however, the model fit for the gamma model is not as good as for the log-normal model. To see this, we compare the histogram of the PPD *p*-values for the χ^2 discrepancy statistic $T(y_i, \mu_i, \sigma_i^2)$ under the normal model with the PPD *p*-values for the χ^2 discrepancy statistic under the gamma model in Fig. 3. (Because the χ^2 discrepancy statistic is no longer exactly $\chi^2_{T^{\mathrm{obs}}}$ under the correct model, we compute the PPD *p*-value empirically as $L^{-1}\Sigma_{l=1}^{L} I[T\{y_i^{\mathrm{obs}}, \mu_i^{(l)}, (\sigma^2)_i^{(l)}\} \leqslant T\{y_i^{(l)}, \mu_i^{(l)}, (\sigma^2)_i^{(l)}\}]$, where $y_i^{(l)}$ is generated under its PPD.) Although the model fit for the gamma model is reasonably adequate, the considerably heavier tails of the gamma model indicate the tendency of some subjects to be either underfitted (means not as well specified or variances underestimated) or, less commonly, overfitted (variances overestimated). A substantive reason for the poorer fit of the gamma model may be that differing
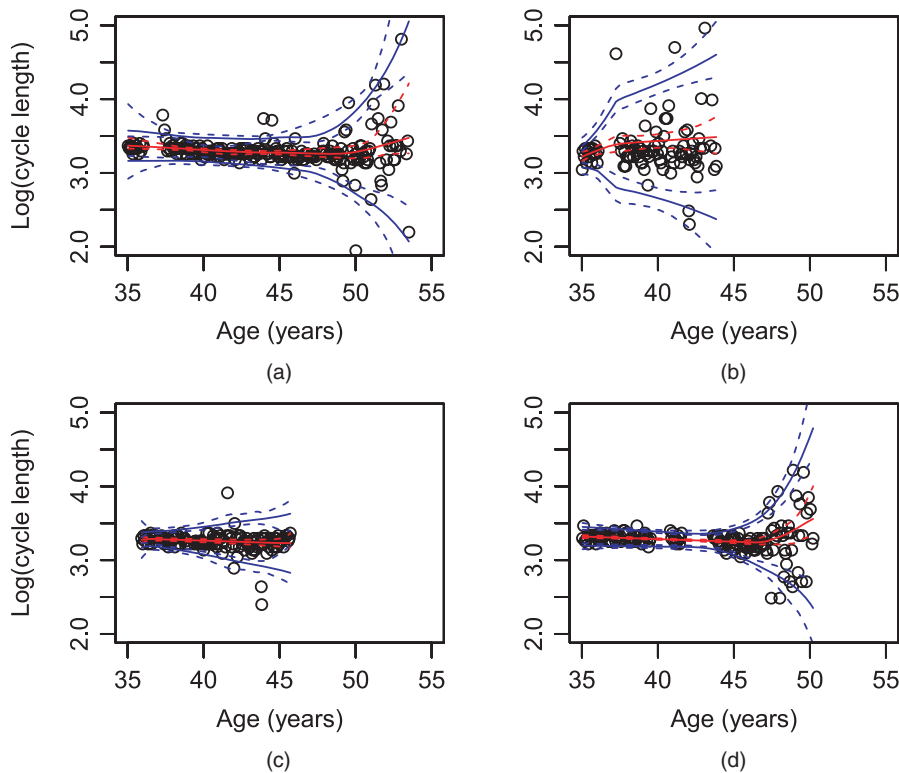

Fig. 5. Change point model applied to the data for four women in Fig. 2 (———, posterior mean of the mean segment length; – – – –, associated 95% CIs; ———, posterior means for the upper and lower 2.5-percentiles for the segment distribution; – – – –, associated 95% CIs; $\bigcirc$, logarithm of observed segment lengths): gamma distributional instead of log-normal distributional assumption for segment length

biological mechanisms are believed to underlie the mean and variance change points; the log-linear model breaks the correlation between mean and variance after the log-transformation, whereas the gamma model assumes that the mean is in part a function of the variance. In addition, the gamma model appears to be more sensitive to early, possibly spurious, long cycles; Fig. 5 gives some evidence of this, particularly if we compare subject (b) with the results in Fig. 1.

Finally, we explore the effect of a missingness not at random mechanism via a sensitivity analysis, focusing on women whose FMPs are censored because of dropout or hormone use. Specifically, we allow for the possibility that women with censored FMPs have a different risk profile from women with observed FMPs via a sensitivity parameter c_i, so that the age at FMP is assumed piecewise exponential with hazard $h_i(t) = c_i \eta_k$ for $A_{k-1} \leqslant t < A_k$ for knots k as defined above. When the FMP is observed $c_i = 1$; when censored, $c_i = c \neq 1$, where c is assumed fixed and known. We consider $c = 0.9$ (a longer average time to FMP for women with censored FMPs) and $c = 1.1$ (a shorter average time to FMP for women with censored FMPs). Table 1 shows the effect of these changes in missingness assumption on the population level parameters under the log-normal modelling assumption. There appears to be very little effect on the posterior means and predictive intervals of the population level parameters resulting from moderate departures from the assumption of missingness at random for women whose FMPs are censored. Similarly, there is little effect on the correlation structure of the regression coefficients (the data are not shown).

5. Discussion

In this paper we have provided a hierarchical change point model for describing the patterns of means and variances of women's menstrual segment lengths as they approach menopause. Our model detects individual change points of the mean and variance of segment lengths for each individual woman. The model is applied to the TREMIN data. Instead of setting splines at a certain fixed point for all women and using traditional random-effect models to study menstrual patterns (Harlow *et al.*, 2000), our model allows the change points to be unknown parameters that vary for different subjects. The hierarchical setting provides a flexible way of capturing both the mean and the variability of each individual's segment length trajectory, as well as imputation of missing data.

Our work develops a data-driven definition of early and late transition defined by subject level variance and mean change points respectively. We observed a 3.2-year difference in age between mean and variance change points at the population level, which is somewhat shorter than that of Lisabeth *et al.* (2004a), who reported a 3.9-year difference between cycle lengths with standard deviations of 6 days and the first cycle of 60 days or more. In addition, our results were consistent with those of Wallace *et al.* (1979), Den Tonkelaar *et al.* (1998) and Lisabeth *et al.* (2004a), who found that longer mean segment lengths were associated with later FMPs. Our results were also consistent with those of Weinstein *et al.* (2003), who found that lower variability was associated with early FMPs. We further found relationships between rates of change in length and variability before and after change points themselves, in particular that greater baseline variability was associated with more rapid declines in variability before variance change points and greater increases thereafter; and later mean change points were associated with greater increases in mean length and more mean variability after mean change points. These data contribute to efforts to define a staging system for reproductive aging and to describe better the heterogeneity in women's experience by menopausal transition

indicators about which standard Bayesian statistical inference (posterior means and CIs) can be made, in contrast with summary statistics of menstrual length that have previously been proposed.

In addition to providing a framework to understand the timing and duration of the menopausal transition better, we show that we can use the results of our model to predict the age at FMP accurately. Of particular interest is the fact that heterogenity in cycle length variability is an even stronger predictor of FMP than the more visually apparent age of change points in mean and variance. Women with greater variability at baseline, with less rapid decreases in variability before the variance change point and with more rapid increases in variability after the variance change point had earlier ages at FMP. These findings emphasize the importance of variance in menstrual cycle length as a key marker of ovarian function and suggest that consideration of menstrual variance may lead to improved precision in clinical predictions of time to FMP. Information that the FMP is more likely 2 *versus* 5 years off has important ramifications for clinical decision making. Furthermore, these results suggest that research on ovarian aging is likely to be enhanced by studies investigating factors that influence cycle variance at the time of the transition. Of particular interest is how the transition experience of women with polycystic ovarian syndrome, who exhibit infrequent cycles and high variability, might best be characterized and how body size is related to cycle variance at the time of transition.

Several extensions and additional applications of the model are possible. To have easily interpreted early and late menopausal transition ages, we have assumed a continuous but non-differentiable mean and variance function. In reality, a smooth transition may be more biologically plausible, suggesting use of differentiable transition models such as the hyperbolic tangent functions of Bacon and Watts (1971), the bent cable regression models of Chiu *et al.* (2006) or the transition polynomial models of Van den Hout *et al.* (2011)—although, as our change points are latent, the actual posteriors of the predictive means and variances at both the individual and the population level trace a smooth curve over age, as evidenced in Figs 1 and 2. Second, the FMP prediction of Section 4.4 could be replaced with a joint model that incorporates the latent subject level mean and variance parameters directly into the modelling of the FMP. We did not pursue this here because of our intention to provide an example of how our model output can be used to inform understanding of the menopausal transition further, as well as the practical limitations of obtaining model fit in a reasonable period of time for such a model; the latter limitation may be overcome with future increases in processing speed or improved MCMC algorithmic construction. A third key extension of this model will add the second TREMIN cohort data to assess changes in women's menstrual patterns in different generations by adding secular cohort (1935–1970 *versus* 1960–1995) as a population level covariate to the model. Fourthly, although developing typologies of the menopausal transition have been considered in the literature (Gorrindo *et al.*, 2007), they have approached the problem from an *ad hoc* direction, based on visual inspection of the cycle length trace plots. Use of a two-stage clustering or joint latent class model may provide a more principled method for developing such typologies and lead to better understanding of the stages of menopausal transition: Huang *et al.* (2012) used the output from the eight-parameter mean–variance change point model together with a k-means clustering algorithm to identify six subgroups of women whose transition experience can be distinguished by age at onset, variability of the menstrual cycle and duration of the early transition. Finally, extension of the model to estimate menopausal transition in the presence of left censoring is also of interest as many recent and on-going studies enrolled prevalent cohorts including women who had already begun the menopausal transition.

Acknowledgements

Appendix A

Gibbs sampling is used to draw from the posterior distribution $p(\mathbf{\Phi}, \Lambda, \Omega | \mathbf{z})$, where $\Phi_i = (\alpha_i^\mu, \beta_i^\mu, \gamma_i^\mu, \theta_i^\mu, \alpha_i^\sigma, \beta_i^\sigma, \gamma_i^\sigma, \theta_i^\sigma)'$. An outline of the algorithm is as follows.

Step 1: initialize $\mathbf{\Phi}$, Λ and Ω. Perform an initial imputation of missing data.
Step 2: for $i = 1, \dots, n$ and z_i consisting of both observed and imputed data,

(a)
$$(\alpha_i^\mu, \beta_i^\mu, \gamma_i^\mu | \text{rest}) \sim N(U, V),$$
$$U = (A_i^{\mu'} W_i^{-1} A_i^\mu + \Omega_\mu^{-1})^{-1} (A_i^{\mu'} W_i^{-1} z_i + \Omega_\mu^{-1} x_i' \Lambda_\mu),$$
$$V = (A_i^{\mu'} W_i^{-1} A_i^\mu + \Omega_\mu^{-1})^{-1}$$

where
$$W_i = \text{diag}(\sigma_{it}^2),$$

$$A_i^\mu = \begin{pmatrix} 1 & (a_{i1} - 35) & (a_{i1} - \theta_i^\mu)_+ \\ \vdots & \vdots & \vdots \\ 1 & (a_{iT_i} - 35) & (a_{iT_i} - \theta_i^\mu)_+ \end{pmatrix},$$

and Λ_μ and Ω_μ are the corresponding part of the prior multivariate normal mean Λ and covariance matrix Ω conditional on other parameters.

(b)
$$p(\alpha_i^\sigma | \text{rest}) \propto \exp\left(-\frac{1}{2} \left[\sum_{t=1}^{T_i} \frac{z_{it} - \{\alpha_i^\mu + \beta_i^\mu (a_{it} - 35) + \gamma_i^\mu (a_{it} - \theta_i^\mu)_+\}^2}{\exp\{\alpha_i^\sigma + \beta_i^\sigma (a_{it} - 35) + \gamma_i^\sigma (a_{it} - \theta_i^\sigma)_+\}} + \frac{(\alpha_i^\sigma - \mu_{\alpha^\sigma})^2}{\Omega_{\alpha^\sigma}} \right] \right)$$
$$\times \exp\left(-\frac{\alpha_i^\sigma T_i}{2} \right)$$

where $\mu_{\alpha^\sigma} = x_i' \Lambda_{\alpha^\sigma}$ and Ω_{α^σ} are the corresponding part of the prior multivariate normal mean and variance conditional on other parameters. The inverse cumulative distribution function (CDF) method is used to obtain the conditional draws.

(c)
$$p(\beta_i^\sigma | \text{rest}) \propto \exp\left(-\frac{1}{2} \left[\sum_{t=1}^{T_i} \frac{z_{it} - \{\alpha_i^\mu + \beta_i^\mu (a_{it} - 35) + \gamma_i^\mu (a_{it} - \theta_i^\mu)_+\}^2}{\exp\{\alpha_i^\sigma + \beta_i^\sigma (a_{it} - 35) + \gamma_i^\sigma (a_{it} - \theta_i^\sigma)_+\}} + \frac{(\beta_i^\sigma - \mu_{\beta^\sigma})^2}{\Omega_{\beta^\sigma}} \right] \right)$$
$$\times \exp\left\{ -\frac{1}{2} \beta_i^\sigma \sum_{t=1}^{T_i} (a_{it} - 35) \right\}$$

where $\mu_{\beta^\sigma} = x_i' \Lambda_{\beta^\sigma}$ and Ω_{β^σ} are the corresponding part of the prior multivariate normal mean and variance conditional on other parameters. The inverse CDF method is used to obtain the conditional draws.

(d)

$$p(\gamma_i^\sigma|\text{rest}) \propto \exp\left(-\frac{1}{2}\left[\sum_{t=1}^{T_i} \frac{z_{it} - \{\alpha_i^\mu + \beta_i^\mu(a_{it}-35) + \gamma_i^\mu(a_{it}-\theta_i^\mu)_+\}^2}{\exp\{\alpha_i^\sigma + \beta_i^\sigma(a_{it}-35) + \gamma_i^\sigma(a_{it}-\theta_i^\sigma)_+\}} + \frac{(\gamma_i^\sigma - \mu_{\gamma^\sigma})^2}{\Omega_{\gamma^\sigma}}\right]\right)$$
$$\times \exp\left\{-\frac{1}{2}\gamma_i^\sigma \sum_{t=1}^{T_i}(a_{it}-\theta_i^\sigma)_+\right\}$$

where $\mu_{\gamma^\sigma} = x_i'\Lambda_{\gamma^\sigma}$ and Ω_{γ^σ} are the corresponding part of the prior multivariate normal mean and variance conditional on other parameters. The inverse CDF method is used to obtain the conditional draws.

(e)

$$p(\theta_i^\mu|\text{rest}) \propto \exp\left(-\frac{1}{2}\left[\sum_{t=1}^{T_i} \frac{z_{it} - \{\alpha_i^\mu + \beta_i^\mu(a_{it}-35) + \gamma_i^\mu(a_{it}-\theta_i^\mu)_+\}^2}{\exp\{\alpha_i^\sigma + \beta_i^\sigma(a_{it}-35) + \gamma_i^\sigma(a_{it}-\theta_i^\sigma)_+\}} + \frac{(\theta_i^\mu - \mu_{\theta^\mu})^2}{\Omega_{\theta^\mu}}\right]\right)$$

where $\mu_{\theta^\mu} = x_i'\Lambda_{\theta^\mu}$ and Ω_{θ^μ} are the corresponding part of the prior multivariate normal mean and variance conditional on other parameters. The inverse CDF method is used to obtain the conditional draws.

(f)

$$p(\theta_i^\sigma|\text{rest}) \propto \exp\left(-\frac{1}{2}\left[\sum_{t=1}^{T_i} \frac{z_{it} - \{\alpha_i^\mu + \beta_i^\mu(a_{it}-35) + \gamma_i^\mu(a_{it}-\theta_i^\mu)_+\}^2}{\exp\{\alpha_i^\sigma + \beta_i^\sigma(a_{it}-35) + \gamma_i^\sigma(a_{it}-\theta_i^\sigma)_+\}} + \frac{(\theta_i^\sigma - \mu_{\theta^\sigma})^2}{\Omega_{\theta^\sigma}}\right]\right)$$
$$\times \exp\left\{-\frac{1}{2}\gamma_i^\sigma \sum_{t=1}^{T_i}(a_{it}-\theta_i^\sigma)_+\right\}$$

where $\mu_{\theta^\sigma} = x_i'\Lambda_{\theta^\sigma}$ and Ω_{θ^σ} are the corresponding part of the prior multivariate normal mean and variance conditional on other parameters. The inverse CDF method is used to obtain the conditional draws.

Step 3:

$$\Lambda|\text{rest} \sim N[\{X'(\Omega \otimes I_N)^{-1}X\}^{-1}X'(\Omega \otimes I_N)^{-1}\Phi, \{X'(\Omega \otimes I_N)^{-1}X\}^{-1}]$$

where X is the covariate matrix of all subjects, which consists of stacked rows of x_i', and Φ consists of the stacked rows of Φ_i'.

Step 4:

$$\Omega|\text{rest} \sim \text{Inv-Wishart}[\Omega|\{\sum_{i=1}^{N}(\Phi_i - x_i'\Lambda)(\Phi_i - x_i'\Lambda)' + I\}].$$

Step 5: use the updated parameters to create a new imputation data set as discussed in Section 3.2. Then go to step 2. Repeat until convergence is obtained.

Appendix B: Piecewise exponential distribution

Assume that Q_i, the age at FMP, follows a piecewise exponential distribution. The baseline hazard is constant within each interval, so

$$\lambda_0(t) = \eta_k, \qquad t \in [A_{k-1}, A_k],$$

$$f(Q_i = t : t \in [A_{k-1}, A_k]) = \eta_k \exp(-\eta_k t).$$

Here, $A_0, \ldots, A_k$ are a set of age knots, which are set at age 40, 42, 43, 44, 45, 46, 46.5, 47 47.5, 48, 48.5, 49, 49.5, 50, 51.5, 52, 52.5, 53, 53.5, 54, 55, 56 and 57 years; we define $A_{-1} = 0$ and $A_{K+1} = \infty$ and assume that $\eta_0 = 0$ (no risk of FMP before age 40 years).

We postulate a very weakly informative prior for $\eta_k : \eta_k \sim \text{gamma}(0.001, 0.001)$. The posterior distribution for η_k is

$$p(\eta_k|\tilde{q}) \propto p(\tilde{q}|\eta_k)\, p(\eta_k) \propto \text{gamma}(m_k + 0.001, r_k + 0.001)$$

where $m_k = \Sigma_i I(A_k - 1 \leqslant Q_i \leqslant A_k)$ is the number of women with FMPs that occur between time A_{k-1} and A_k and $r_k = \Sigma_i I(Q_i \geqslant A_{k-1})$ is the number of women without an FMP at time A_{k-1}.

The hazard function for each interval is

$$\lambda(t) = \eta_k I(A_{k-1} \leqslant t \leqslant A_k).$$

The cumulative hazard and survival functions are then given by

$$\Lambda(t) = \int_0^t \lambda(t)\, \mathrm{d}t = \sum_{j=1}^{k-1} \eta_j (A_j - A_{j-1}) + \eta_k (t - A_{k-1}), \qquad t \in [A_{k-1}, A_k],$$

$$S(t) = \exp\{-\Lambda(t)\}.$$

The probability that the event occurs in the interval $[t_1, t_2]$ given that the event has not occurred by t_1 is

$$P(Q_i \in [t_1, t_2] \mid Q_i > t_1) = \frac{S(t_1) - S(t_2)}{S(t_1)} = 1 - \frac{S(t_2)}{S(t_1)}$$

$$= \begin{cases} 1 - \exp\{-\eta_k(t_2 - t_1)\} & \text{if } A_{k-1} \leqslant t_1 < t_2 \leqslant A_k, \\ 1 - \exp[-\{\eta_k(t_2 - A_{k-1}) - \eta_{k-1}(t_1 - A_{k-1})\}] & \text{if } A_{k-2} \leqslant t_1 < A_{k-1} \leqslant t_2 \leqslant A_k, \\ 1 - \exp[-\{\eta_{k+1}(t_2 - A_k) + \eta_k(A_k - A_{k-1}) - \eta_{k-1}(t_1 - A_{k-1})\}] & \text{if } A_{k-2} \leqslant t_1 < A_{k-1} < A_k \leqslant t_2. \end{cases}$$

References

Avis, N., Assmann, S., Kravitz, H., Ganz, P. and Ory, M. (2004) Quality of life in diverse groups of midlife women: assessing the influence of menopause, health status and psychosocial and demographic factors. *J. Qual. Lif. Res.*, **13**, 933–946.

Bacon, D. W. and Watts, D. G. (1971) Estimating the transition between two intersecting straight lines. *Biometrika*, **58**, 525–534.

Belsey, E. and Farley, T. (1987) The analysis of menstrual bleeding patterns: a review. *App. Stochast. Mod. Data Anal.*, **3**, 125–150.

Bortot, P., Masarotto, G. and Scarpa, B. (2010) Sequential predictions of menstrual cycle lengths. *Biostatistics*, **11**, 741–755.

Brambilla, D., McKinlay, S. and Johannes, C. (1994) Defining the perimenopause for applications in epidemiologic investigations. *Am. J. Epidem.*, **140**, 1091–1095.

Chiu, G., Lockhart, R. and Routledge, R. (2006) Bent-Cable regression theory and applications. *J. Am. Statist. Ass.*, **101**, 542–553.

Colombo, B., Mion, A., Passarin, K. and Scarpa, B. (2006) Cervical mucus symptom and daily fecundability: first results from a new data base. *Statist. Meth. Med. Res.*, **15**, 161–180.

Crainiceanu, C., Ruppert, D., Carroll, R., Joshi, A. and Goodner, B. (2007) Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *J. Computnl Graph. Statist.*, **16**, 265–288.

Davidian, M. and Carroll, R. (1987) Variance function estimation. *J. Am. Statist. Ass.*, **82**, 1079–1091.

Dennerstein, L., Smith, A., Morse, C., Burger, H., Green, A., Hopper, J. and Ryan, M. (1993) Menopausal symptomatology in Australian women. *Med. J. Aust.*, **159**, 232–236.

Den Tonkelaar, I., te Velde, E. and Looman, C. (1998) Menstrual cycle length preceding menopause in relation to age at menopause. *Maturitas*, **29**, 115–123.

Diggle, P. J. (1990) *Time Series: a Biostatistical Introduction*. Oxford: Oxford University Press.

Gelman, A. (2003) A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *Int. Statist. Rev.*, **71**, 369–382.

Gelman, A. (2007) Discussion on 'Bayesian checking of the second levels of hierarchical models', by M. J. Bayarri and M. E. Castellanos. *Statist. Sci.*, **22**, 349–352.

Gelman, A., Meng, X. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statist. Sin.*, **6**, 733–807.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.*, **7**, 457–472.

Gorrindo, T., Lu, Y., Pincus, S., Riley, A., Simon, J. A., Singer, B. H. and Weinstein, M. (2007) Lifelong menstrual histories are typically erratic and trending a taxonomy. *Menopause*, **14**, 74–88.

Guo, Y., Manatunga, A., Chen, S. and Marcus, M. (2006) Modeling menstrual cycle length using a mixture distribution. *Biostatistics*, **7**, 100–114.

Hall, C., Ying, J., Kuo, L. and Lipton, R. (2003) Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computnl Statist. Data Anal.*, **42**, 91–109.

Harlow, S., Cain, K., Crawford, S., Dennerstein, L., Little, R., Mitchell, E., Nan, B., Randolph, Jr, J., Taffe, J. and Yosef, M. (2006) Evaluation of four proposed bleeding criteria for the onset of late menopausal transition. *J. Clin. Endocrin. Metablsm*, **91**, 3432–3438.

Harlow, S., Crawford, S., Dennerstein, L., Burger, H., Mitchell, E. and Sowers, M. (2007) Recommendations from a multi-study evaluation of proposed criteria for staging reproductive aging. *Climacteric*, **10**, 112–119.

Harlow, S., Lin, X. and Ho, M. (2000) Analysis of menstrual diary data across the reproductive life span: applicability of the bipartite model approach and the importance of within-woman variance. *J. Clin. Epidem.*, **53**, 722–733.

Harlow, S., Mitchell, E., Crawford, S., Nan, B., Little, R. and Taffe, J. (2008) The ReSTAGE collaboration: defining optimal bleeding criteria for onset of early menopausal transition. *Fertil. Steril.*, **89**, 129–140.

Harlow, S. D. and Zeger, S. L. (1991) Application of longitudinal methods to the analysis of menstrual diary data. *J. Clin. Epidem.*, **44**, 1015–1025.

Huang, X., Harlow, S. D. and Elliott, M. R. (2012) Patterns of menstrual bleeding and their relation to the onset of menopause, markers of stages, and health outcomes. *Am. J. Epidem.*, **175**, 74–83.

Jin, Z., Lin, D. and Ying, Z. (2006) On least-squares regression with censored data. *Biometrika*, **93**, 147–161.

Lisabeth, L., Harlow, S., Gillespie, B., Lin, X. and Sowers, M. (2004a) Staging reproductive aging: a comparison of proposed bleeding criteria for the menopausal transition. *Menopause*, **11**, 186–197.

Lisabeth, L., Harlow, S. and Qaqish, B. (2004b) Marginal models for mean and variance: a new approach to modelling menstrual variation across the menopausal transition. *J. Clin. Epidem.*, **57**, 484–496.

Little, R. and Rubin, D. (2002) *Statistical Analysis with Missing Data*. New York: Wiley.

Mansfield, P., Carey, M., Anderson, A., Barsom, S. and Koch, P. (2004) Staging the menopausal transition: data from the TREMIN research program on women's health. *Wom. Hlth Iss.*, **24**, 220–226.

Marcus, M., McChesney, R., Golden, A. and Landrigan, P. (2000) Video display terminals and miscarriages. *J. Am. Med. Wom. Ass.*, **55**, 84–88.

Matsumoto, S., Nogami, Y. and Ohkuri, S. (1962) Statistical studies on menstruation: a criticism on the definition of normal menstruation. *Gun. J. Med. Sci.*, **11**, 294–318.

Miolo, L., Colombo, B. and Marshall, J. (1993) A database for biometric research on changes in basal body temperature in the menstrual cycle. *Statistica*, **53**, 563–572.

Mitchell, E., Woods, N. and Mariella, A. (2000) Three stages of the menopausal transition from the Seattle Midlife Women's Health Study: toward a more precise definition. *Menopause*, **7**, 334–349.

Nassaralla, C., Stanford, J., Daly, K., Schneider, M., Schliep, K. and Fehring, R. (2011) Characteristics of the menstrual cycle after discontinuation of oral contraceptives. *J. Wom. Hlth*, **20**, 1–9.

Santoro, N., Brockwell, S., Johnston, J., Crawford, S., Gold, E., Harlow, S., Matthews, K. and Sutton-Tyrrell, K. (2007) Helping midlife women predict the onset of the final menses—SWAN, the Study of Women's Health Across the Nation. *Menopause*, **14**, 1–9.

Soules, M., Sherman, S., Parrott, E., Rebar, R., Santoro, N., Utian, W. and Woods, N. (2001) Stages of Reproductive Aging Workshop (STRAW). *J. Wom. Hlth Gend. Med.*, **10**, 843–848.

Sowers, M., Jannausch, M., McConnell, D., Little, R., Greendale, G., Finkelstein, J., Neer, R., Johnston, J. and Ettinger, B. (2006) Hormone predictors of bone mineral density changes during the menopausal transition. *J. Clin. Endcrin. Metablsm*, **91**, 1261–1267.

Taffe, J. and Dennerstein, L. (2002a) Menstrual patterns leading to the final menstrual period. *Menopause*, **9**, 32–40.

Taffe, J. and Dennerstein, L. (2002b) Time to the final menstrual period. *Fertil. Steril.*, **78**, 397–403.

Thum, Y. and Bhattacharya, S. (2001) Detecting a change in school performance: a Bayesian analysis for a multilevel joint point problem. *J. Educ. Behav. Statist.*, **26**, 443–468.

Treloar, A. (1981) Menstrual cyclicity and the pre-menopause. *Maturitas*, **3**, 249–264.

Treloar, A. and Behn, B. (1971) Effects of oral contraceptives on the human menstrual cycle. In *Proc. 7th Wrld Congr. Fertility and Sterility, Tokyo and Kioto*, pp. 886–890. Amsterdam: Excerpta Medica.

Treloar, A., Boynton, R., Behn, B. and Brown, B. (1967) Variation of the human menstrual cycle through reproductive life. *Int. J. Fertil.*, **12**, 77–113.

Van den Hout, A., Muniz-Terrera, G. and Matthews, F. E. (2011) Smooth random change point models. *Statist. Med.*, **30**, 599–610.

Vollman, R. (1977) *The Menstrual Cycle*. Philadelphia: Saunders.

Wallace, R., Sherman, B., Bean, J., Treloar, A. and Schlabaugh, L. (1979) Probability of menopause with increasing duration of amenorrhea in middle-aged women. *Am. J. Obstet. Gyn.*, **135**, 1021–1024.

Wegienka, G. and Baird, D. (2003) Potential bias due to excluding oral contraceptive users when estimating menstrual cycle characteristics. *Am. J. Epidem.*, **158**, 947–950.

Weinstein, M., Gorrindo, T., Riley, A., Mormino, J., Niedfeldt, J., Singer, B. and Rodriguez, G. (2003) Timing of menopause and patterns of menstrual bleeding. *Am. J. Epidem.*, **158**, 782–791.

Wooldridge, J. (2002) *Econometric Analysis of Cross-section and Panel Data*. Cambridge: MIT Press.

World Health Organization (1996) Research on Menopause in the 1990's. *Technical Report*. World Health Organization, Geneva.