University of Michigan School of Public Health
The University of Michigan Department of Biostatistics Working Paper Series

Year 2010 *Paper* 87

Modeling Menstrual Cycle Length and Variability at the Approach of Menopause Using Bayesian Changepoint Models

Xiaobi Huang* Michael R. Elliott† Sioban D. Harlow‡

*University of Michigan, xiaobi@umich.edu
†University of Michigan, mrelliot@umich.edu
‡University of Michigan, harlow@umich.edu

http://biostats.bepress.com/umichbiostat/paper87

Modeling Menstrual Cycle Length and Variability at the Approach of Menopause Using Bayesian Changepoint Models

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

Abstract

As women approach menopause, the patterns of their menstruation cycle lengths change. To study these changes, we need to jointly model both the mean and variability of the cycle length. The 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. Data are from TREMIN, an ongoing 70-year old longitudinal study that has obtained menstrual calendar data of women throughout their reproductive life course. An additional complexity arises from the fact that these calendars have substantial missingness due to hormone use, surgery, failure to report, and loss of contact. We integrate multiple imputation and time-to event modeling in our Bayesian estimation procedure to deal with different forms of the missingness. Posterior predictive model checks are applied to evaluate the model fit. Our method successfully modeled patterns of women's menstrual cycle trajectories throughout their late reproductive life and identified the change points for mean and variability of segment length, which provides insight into the menopausal process. More generally, our model points the way toward increasing use of joint mean-variance models to predict health outcomes and better understand disease processes.

1 Introduction

Menstrual cycles are the most easily observed markers of ovarian function throughout reproductive life. Changes in menstrual bleeding patterns are important indicators of ovarian aging, endocrine disruption and endocrine risk factors for chronic disease (Harlow 1995). The menopausal transition is increasingly recognized to be a critical period in women's lives as physiologic changes and health practices adopted during this period frequently define women's long term chronic disease risk profile (Wildman et al. 2008, Sowers et al. 2006, Avis et al. 2004). Given this recent interest in the interface between reproductive and somatic aging, several proposals for staging reproductive aging have emerged. The Stages of Reproductive Aging Workshop (STRAW) recommendations (Soules et al. 2001), its modifications (Harlow et al. 2007) and several other proposals (Mitchell et al. 2000,Taffe and Dennerstein 2002, Mansfield et al. 2004) define criteria primarily by menstrual bleeding characteristics to determine onset of the transition, as well as the stages within the transition period.

Information on the patterns of menstrual bleeding across the reproductive lifespan derives mainly from four seminal menstrual calendar studies, including three studies from Caucasian populations (Chiazze et al. 1968, Treloar et al. 1967, Vollman 1977) and one study from a Japanese population (Matsumoto et al. 1962, Matsumoto et al. 1979). Treloar (1981) was the first to estimate age at entry into the menopausal transition by visual inspection of menstrual cycle lengths for the 12 year period prior to 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

estimated 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 amenorrhea or menstrual irregularity. Subsequently, investigators from several longitudinal studies (Melbourne Women's Midlife Health Project [MWMHP] (Dennerstein et al. 1993), Seattle Midlife Women's Health Study [SWMHS] (Mitchell et al. 2000), TREMIN (Treloar et al. 1967)) proposed various bleeding criteria to define the transition period (Taffe and Dennestein 2002, Mitchell et al. 2000, Mansfield et al. 2004, Lisabeth et al. 2004a).

STRAW defined stages principally by changes in menstrual bleeding characteristics and, to a lesser extent, by changes in serum follicle-stimulating hormone (FSH) levels (Soules et al. 2001). STRAW divided reproductive life prior to menopause into the reproductive years (3 stages) and the transition years (2 stages, early and late transition). Entry into the early transition is characterized by increased variability in menstrual cycle length while entry into the late transition is characterized by the occurrence of skipped cycles or amenorrhea. The STRAW recommendations by (Soules et al. 2001), although based on emerging results of the large cohort studies of midlife women, were not data-driven. The multi-study ReSTAGE Collaboration subsequently evaluated bleeding criteria that served as the basis of the STRAW recommendations and documented the extent to which the various proposed criteria identified a similar moment in women's reproductive life (Harlow et al. 2006, Harlow et al. 2007, Harlow et al. 2008). Essentially, 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 papers attempted to model these changepoints longitudinally.

Harlow et al. (2000) longitudinally modeled 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. They fitted a bipartite cubic spline model that modeled the risk of both very short and very long segments using changepoints fixed at ages 34 and 40. Lisabeth et al. (2004b) used generalized estimating equations to model changes in mean cycle length and variance independent of the mean referenced to age at FMP and demonstrated that variance in menstrual cycle lengths increase on average 2 to 6 years before increases in the mean, depending on age at FMP.

Prior descriptive analyses also suggest that there is some heterogeneity in women's menstrual trajectories. Menstrual characteristics in young adult women are associated with fertility (Small et al. 2006) and the timing of menopause (Den Tonkelaar et al. 1998, Wallace et al. 1979, Lisabeth et al. 2004b). 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 two years before menopause than women with earlier menopause. Lisabeth and colleagues (Lisabeth et al. 2004b) 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) had a longer mean cycle length in the nine years prior to menopause than women with an earlier menopause. Weinstein and colleagues (Weinstein et al. 2003) found that low serial irregularity, 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 variability change when women approach menopause. We assume that there are underlying unknown mean and variance changepoints for each individual woman and build a Bayesian change point model to estimate

distributions of these changepoints. Furthermore, we impute cycles that are missing due to hormone use, gaps in the menstrual calendar and gynecological surgery, allowing more subjects and information to be included. Most prior reports have censored women when they began using hormonal contraceptives or hormone therapy (HT) (Weinstein et al. 2003, Guo et al. 2006, Harlow et al. 2006, Harlow et al. 2008).

Statistically, the objective is to model both the mean and variance of a set of curves. Several approaches have been proposed for correlated functional data of this type, including the bipartite spline model proposed by (Harlow et al. 2000) which modelled mean and between-subject variance by a linear random effect model and used a two-stage log-linear regression to study within-subject variance vs. age. Crainiceanu et al. (2007) proposed Bayesian penalized splines to model both mean and variance by using a set of fixed knots for the splines with structural covariance matrix and random effects to depict the heterogeneity of variance. Lisabeth et al. (2004b) modeled means and variances over time separately using independent generalized estimating equations. Gunn and Dunson (2005) modeled hormone patterns in the menstrual cycles using a Bayesian hierarchical model and mapped the posterior draws to a constrained space which guarantees that each curve increases monotonically to an unknown changepoint and decreases afterwards. To model student test achievement, Thum and Bhattacharya (2001) proposed a hierarchical Bayesian regression model which included 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 changepoints 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 changepoints. Here we consider a Bayesian hierarchical model that estimates individual-level mean and variance profiles with unknown changepoints. These changepoints represent measures of menopausal transition, and, together with intercepts and pre- and post-changepoint 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 article is organized as follows. In section 2 we describe the TREMIN study data. In section 3 we describe a Bayesian model to study the trajectories of women's menstrual cycle length that estimates unknown changepoints for both means and variances and allows these changepoints to be functions of subject-level covariates. In addition, we impute different forms of missingness in the data set and incorporate the imputation in the Markov Chain Monte Carlo sampling used to estimate the algorithm. In section 4 we give the results from fitted the model to menstrual data, along with Bayesian posterior predictive model checks. In section 5 we discuss how our results compare to and extend previous menstrual cycle staging research, along with possible extensions of our model.

2 The TREMIN Dataset

Our models are designed for the TREMIN data, one of the only two data sets available providing individual women's menstrual calendar data across their reproductive life span. The study, 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 WHO (Belsey and Farley 1987) were used to summarize the calendar data. A bleeding segment, 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. Bleeding segment length is the dependent variable in our study. Age at menopause is determined by the date of final menstrual period (FMP), which is attributed retrospectively after 12 months of amenorrhea on the calendar cards (WHO 1996).

We used data from 617 women in the 1935-1939 cohort who were a) age 25 or less at enrollment, b) used hormones for less than four years continuously, c) had at least one observed segment before age 40, and d) were not censored before age 40 (Data tape TRUST998.FINAL, March 1993). We consider segment lengths beginning at age 35. 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, the bleeding segment is coded as a treated interval, during which ovarian function is masked. Thus, the segment data are considered to be missing when women use hormones. A one-segment washout period after hormone use 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 four years. Studies have not found that hormonal use influences menstrual segment length after stopping use and allowing for a washout period (Taylor et al. 1977, Treloar and Behn 1971).

The 617 women included in our analysis each 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 (FMPs) were observed for 313 subjects (50.7%). Only 105 (17.0%) have complete data.

Figure 1 displays log segment lengths for four typical women in the TREMIN data set. Subject A has complete data. She has a pregnancy gap which is not included in the analysis, no gynecological surgery or periods of hormone use, and has an observed final menstruation period (FMP). Subject B was coded as using hormones from age 36.07 to age 37.24, and her information for this period is treated as missing. Her FMP is observed, however. Subject C has intermittent missingness at age 36.95. She had a hysterectomy at age 45.78, thus her menstrual history was truncated at this point and no FMP was observed. Subject D has intermittent missing at age 39.59 and from age 41.56 to 43.64. She began hormone therapy after age 50.21, with no untreated bleeds recorded afterwards; thus no FMP was observed.

3 Modeling Menstrual Cycle Data

We construct a Bayesian change point model for the mean and variance of the segment length.

3.1 Change Point Model for Mean and Variance

Let y_{it} denote the t^{th} menstrual segment length of subject i. Let a_{it} denote the age at the beginning of the t^{th} menstrual segment of subject i, where $i = 1, \ldots, N$, $t = 1, \ldots, T_i$, $N = 617$.

We consider a log-normal model with a linear change point for both the mean and variance for each subject:

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

$$\mu_{it} = \alpha^{\mu}_i + \beta^{\mu}_i(a_{it} - 35) + \gamma^{\mu}_i(a_{it} - \theta^{\mu}_i)_+$$

$$log(\sigma^2_{it}) = \alpha^{\sigma}_i + \beta^{\sigma}_i(a_{it} - 35) + \gamma^{\sigma}_i(a_{it} - \theta^{\sigma}_i)_+$$

The function $(x)_+ = x$ if $x \geq 0$, $(x)_+ = 0$ if $x < 0$; θ^{μ}_i and θ^{σ}_i are the unknown change points of mean and variance for subject i. The change points create a linear spline for each mean and variance model. We denote these eight subject-level parameters for each woman as $\Phi_i = (\alpha^{\mu}_i, \beta^{\mu}_i, \gamma^{\mu}_i, \theta^{\mu}_i, \alpha^{\sigma}_i, \beta^{\sigma}_i, \gamma^{\sigma}_i, \theta^{\sigma}_i)'$.

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

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

where x_i are covariates 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.

We complete the model specification by postulating an Inverse-Wishart hyperprior for Λ and Ω:

$$p(\Lambda, \Omega) = Inv - Wishart(\Omega; 1, I)$$

which is completely flat for Λ and weakly informative for Ω.

3.2 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 conditional on the observed data z^{obs}. The posterior based on the complete data z is given by

$$p(\mathbf{\Phi}, \Lambda, \Omega|\boldsymbol{z}) \propto \prod_{i=1}^{N}[\prod_{t=1}^{T_i} p(z_{it}|\Phi_i)p(\Phi_i|\Lambda,\Omega)]p(\Lambda,\Omega) \propto$$

$$[\prod_{i=1}^{N}[\prod_{t=1}^{T_i}\frac{1}{\sigma_{it}}\exp(-\frac{(z_{it}-\mu_{it})^2}{2\sigma_{it}^2})]|\Omega|^{-\frac{1}{2}}\exp(-\frac{1}{2}(\Phi_i - x_i'\Lambda)'\Omega^{-1}(\Phi_i - x_i'\Lambda))]|\Omega|^{-\frac{k+2}{2}}\exp(-\frac{1}{2}tr(\Omega^{-1})) =$$

$$\left[\prod_{i=1}^{N}\prod_{t=1}^{T_i}\sigma_{it}^{-1}\right] \mid \Omega \mid^{-\frac{N+k+2}{2}} \exp\left\{\sum_{i=1}^{N}\left[\sum_{t=1}^{T_i}\frac{(z_{it}-\mu_{it})^2}{\sigma_{it}^2} + (\Phi_i - x_i')'\Omega^{-}1(\Phi_i - x_i')\right] + tr(\Omega^{-1})\right\}$$

where $\mu_{it} = \alpha_i^{\mu} + \beta_i^{\mu}(a_{it} - 35) + \gamma_i^{\mu}(a_{it} - \theta_i^{\mu})_+$, $\sigma_{it}^2 = \exp(\alpha_i^{\sigma} + \beta_i^{\sigma}(a_{it} - 35) + \gamma_i^{\sigma}(a_{it} - \theta_i^{\sigma})_+)$, $k = dim(\Omega) = 8$. We sample the parameters via a MCMC algorithm that uses Metropolis-within-Gibbs sampling. Details of the procedure is in Appendix A.

Missing data are imputed under a missing at random (MAR) assumption (Little and Rubin 2002) using a standard selection model. Imputation is embedded within the MCMC algorithm. Details are provided in the next section.

3.3 Imputation of Missing Data

The majority (512 of the 617 women) have some form of missing data. For 313 women, their segment lengths are censored due to dropout while still menstruating, surgical termination of menstruation due to hysterectomy or bilateral oophorectomy, or hormone use that began before FMP and continued past FMP. 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 missing completely at random. In order to deal with the different types of missingness, we impute the missing data under a missing at random (MAR) assumption. 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 of the FMP in order to terminate the imputation process.

When missingness is intermittent, we ensure that the imputed missing segment lengths sum to the length of the gap 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 $(y_{i,k+1}, ..., y_{i,k+S})' = \tilde{y}_i$ in the gap are independent, subject to the constraint that $\sum_{s=1}^{S} y_{i,k+s} = T$. We obtain a draw $\log(y^{rep}_{i,k+1}) \sim N(\mu_{i,k+1}, \sigma^2_{i,k+1})$ where $\mu_{i,k+1} = \alpha^{\mu}_i + \beta^{\mu}_i(a_{i,k+1} - 35)_+ + \gamma^{\mu}_i(a_{i,k+1} - \theta^{\mu}_i)_+$ and $\sigma^2_{i,k+1} = exp(\alpha^{\sigma}_i + \beta^{\sigma}_i(a_{i,k+1} - 35)_+ + \gamma^{\sigma}_i(a_{i,k+1} - \theta^{\sigma}_i)_+)$ and $a_{i,k+1} = a_{ik} + y_{ik}$ is the age of the start of segment $y^{rep}_{i,k+1}$. A draw of $y^{rep}_{i,k+2}$ is then obtained as for $y^{rep}_{i,k+1}$, where now $a_{i,k+2} = a_{i,k+1} + y^{rep}_{i,k+1}$. This process is repeated until we obtain $y^{rep}_{i,k+S}$ such that $\sum_{s=1}^{S} y^{rep}_{i,k+s} > L_i$. We then replace $y^{rep}_{i,k+S}$ with $\tilde{y}^{rep}_{i,k+S} = L_i - \sum_{s=1}^{S-1} y^{rep}_{i,k+s}$. Let $(y^{(t)}_{i,k+1}, ..., y^{(t)}_{i,k+S-1}, \tilde{y}^{(t)}_{i,k+S}) = \tilde{y}^{(t)}_i$ be the t^{th} vector of imputations, $t = 1, ..., 50$. Finally, we draw one of the 50 sets with probability $p^t = \frac{f(\tilde{y}^{(t)}_i|\Phi_i)}{\sum_t f(\tilde{y}^{(t)}_i|\Phi_i)}$, where $f(\tilde{y}^{(t)}_i|\Phi_i) = \prod_{s=1}^{S-1} \phi\left(\frac{\log(y_{i,k+s}) - \mu_{i,k+s}}{\sigma_{i,k+s}}\right) \times \phi\left(\frac{\log(\tilde{y}_{i,k+S}) - \mu_{i,k+S}}{\sigma_{i,k+S}}\right)$, where $\phi(\cdot)$ is the pdf of the standard normal distribution. On rare occasions where $y^{rep}_{i,k+s} < 4$, the imputed values were truncated to be 4; similarly $y^{rep}_{i,k+s} > 365$ was truncated to 365.

When subjects' segment lengths are censored, we need to impute an FMP since it is unobserved. We model the age at FMP Q_i as a piecewise exponential distribution with hazard $h_i(t) = \eta_k$ for $A_{k-1} \leq t < A_k$ for knots $k = 1, ..., K$. Knots are set at age 40, 42, 43, 44,

45, 46, 46.5, 47, 47.5, 48, 48.5, 49, 49.5, 50, 50.5, 51, 51.5, 52, 52.5, 53, 53.5, 54, 55, 56, 57, and 60. We first obtain a draw from $p(\eta_k \mid Q) \sim GAMMA(\sum_i I(A_{k-1} \leq Q_i \leq A_k), \sum_i I(Q_i \geq A_{k-1}))$ for $k = 1, \ldots, K$, where Q includes both the observed FMP and those imputed at the previous iteration of Gibbs sampler. As in the intermittent missing setting, we then obtain a draw $\log(y^{rep}_{i,T^{rep}_i+1}) \sim N(\mu_{i,T^{rep}_i}, \sigma^2_{i,T^{rep}_i})$ where $\mu_{i,T^{rep}_i} = \alpha^\mu_i + \beta^\mu_i(a_{i,T^{rep}_i} - 35)_+ + \gamma^\mu_i(a_{i,T^{rep}_i} - \theta^\mu_i)_+$ and $\sigma^2_{i,T^{rep}_i} = exp(\alpha^\sigma_i + \beta^\sigma_i(a_{i,T^{rep}_i} - 35)_+ + \gamma^\sigma_i(a_{i,T^{rep}_i} - \theta^\sigma_i)_+)$, $a_{i,T^{rep}_i} = a_{iT^{rep}_i} + y_{iT^{rep}_i}$ is the age of the start of segment y^{rep}_{i,T^{rep}_i}, and T^{rep}_i is the number of observed segments plus the number of imputed segements in any intermittent missing gaps. We then obtain a draw $W_{i,1}$ from a Bernoulli distribution with probability

$$P(a_{i,T^{rep}_i} \leq Q_i \leq a_{i,T^{rep}_i+1} \mid Q_i > \max(a_{i,T^{rep}_i}, \theta^\mu_i, \theta^\sigma_i))$$

$$= \begin{cases} 0 & \text{if } a_{i,T^{rep}_i} \leq \max(\theta^\mu_i, \theta^\sigma_i) \\ 1 - e^{-\eta_k(a_{i,T^{rep}_i+1} - a_{i,T^{rep}_i})} & \text{if } a_{i,T^{rep}_i} > \max(\theta^\mu_i, \theta^\sigma_i), \\ & a_{i,T^{rep}_i}, a_{i,T^{rep}_i+1} \in [A_{k-1}, A_k] \\ 1 - e^{-\left[\eta_k(a_{i,T^{rep}_i+1} - A_{k-1}) - \eta_{k-1}(a_{i,T^{rep}_i} - A_{k-1})\right]} & \text{if } a_{i,T^{rep}_i} > \max(\theta^\mu_i, \theta^\sigma_i), a_{i,T^{rep}_i} \in [A_{k-2}, A_{k-1}], \\ & a_{i,T^{rep}_i+1} \in [A_{k-1}, A_k] \\ 1 - e^{-\left[\eta_{k+1}(a_{i,T^{rep}_i+1} - A_k) + \eta_k(A_k - A_{k-1}) - \eta_{k-1}(a_{i,T^{rep}_i} - A_{k-1})\right]} & \text{if } a_{i,T^{rep}_i} > \max(\theta^\mu_i, \theta^\sigma_i), a_{i,T^{rep}_i} \in [A_{k-2}, A_{k-1}], \\ & a_{i,T^{rep}_i+1} \in [A_k, A_{k+1}] \end{cases}$$

Note that the FMP must occur after both the last observed segment and the latent mean and variance changepoints; also, since none of our knots are less than six months apart, a segment can cover a maximum of 3 intervals. If $W_{i,1} = 1$, y^{rep}_{i,T^{rep}_i+1} is the length of the final FMP. If $W_{i,1} = 0$, we draw $\log(y^{rep}_{i,T^{rep}_i+2}) \sim N(\mu_{i,T^{rep}_i+1}, \sigma^2_{i,T^{rep}_i+1})$ and repeat the process s times until one of the following occurs: $W_{i,T^{rep}_i+s} = 1$, $y^{rep}_{i,T^{rep}_i+1} > 365$ or $a^{rep}_{i,T^{rep}_i+s} \geq 60$. For the vast majority of subjects, the FMP variable trigged the end of the imputation. For more details, see Appendix B.

4 Results

We use the methodology described in section 3 to analyze TREMIN data using MATLAB software. We ran two MCMC chains for 10,000 iterations each after discarding the first 10,000 draws as "burn-in". We assessed convergence using the Gelman and Rubin statistic (Gelman et al. 2004), with a thinning interval of 5 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 visually assess model fit at the individual level, Figure 2 plots the observed segment lengths and predicted means and variances for the same four sampled women described in Figure 1. The model appears to capture the trajectories well, with approximately 5% of cycle lengths excluded from the 95% predictive intervals. The uncertainty in the position of the variance changepoint is highlighted in (b) and (c).

Figure 3 plots the posterior means and 90% credible intervals of the mean and variance changepoints for 50 randomly selected women. As noted by Treloar et al. (1981) and Lisabeth et al. (2004b), variability generally begin to increase before mean length. Subjects with earlier changepoints averaged 4-5 years between mean and variance changepoints, whereas subjects with later changepoints averaged only 1-2 years between mean and variance changepoints, consistent with the findings of Harlow et al. (2008). Uncertainty in the variance changepoints is generally greater than in the mean changepoints.

Figure 4 plots the posterior medians of the mean and variance changepoints against the final menstural periods (FMPs) for the 304 women with observed FMPs. FMPs occured on

average 3.6 years after the mean changepoint, with a standard deviation of 1.4 years. FMPs occured on average 6.5 years after the variance changepoint, with a standard deviation of 2.5 years. The mean time to FMP after the mean changepoint was fairly constant with respect to age at mean changepoint; mean time to FMP after the variance changepoint was considerably shorter in women with later variance changepoints than in younger women.

4.2 Population Level Parameters

Table 1 summarizes the posterior means and associated 95% credible intervals for the population level segment length mean and variance regression parameters. The population mean age at the changepoint for segment length means is 46.20 years (95% CI 45.87-46.54 years), older than the population mean age at changepoints for segment length variability, which is 42.21 years (95% CI 41.84-42.58 years); thus variability in segment length is predicted to begin increasing 3.24 years earlier (95% CI 2.97-3.51 years) in the population than the mean segment length itself. Mean segment length declined about 1% per year before the changepoint and increased about 15% per year afterwards. Variability of log-segment length was stable before the changepoint and increased by 79% per year after the changepoint.

Table 2 presents the posterior mean and associated 95% posterior predictive interval for the correlation matrix corresponding to the covariance matrix Ω. The 95% credible intervals of correlations that exclude zeros are denoted in bold.

- Later changepoints for variance are highly associated with later changepoints for mean.
- Later changepoints for both mean and variance are also correlated with longer and more variable segment lengths, and more rapid increases in mean and variance after the changepoint; consequently mean and variance slopes after changepoints are positively

correlated.

- Greater mean length is associated with greater declines in variability before the variance changepoint and greater increases in variability after.
- Larger segment variability is associated with longer mean segment length.
- Larger segment variability is highly associated with more rapid declines in variability before but larger increases in variability after the variance changepoint: thus change in variability before and after the variance changepoint is negatively correlated.

We conducted a principal components analysis of Ω to determine if the relationships among the eight parameters governing perimenopause segment lengths could be summarized in a smaller number of dimensions. Table 3 shows that four components explained 82% of the variance of the individual level parameters governing menstrual segment length. The first component loads heavily on the inverse relationship between the slope of the variances before and after the variance changepoint, and on late mean and variance changepoints. The second component also loads on the inverse relationship between the slope of the variances before and after the variance changepoint, but picks up a relationship between early changepoints in means and variances and smaller increases in means after the mean changepoint. The third and fourth components load on the relationship between the mean intercepts and slopes: the third component relates longer mean segment lengths at age 35 with more rapid declines in mean length before the mean changepoint and less rapid increases thereafter, while the fourth component relates shorter mean segment lengths at age 35 with more rapid declines in mean length before the mean changepoint and more rapid increases thereafter.

We also fit a two-covariate model, including parity and age at menarche. As 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 checks (Gelman et al. 1996) to assess model fit. We calculated the χ^2 discrepancy statistic for observed segment lengths of each individual woman given by $\sum_t \frac{(y_{it}-\mu_{it}^{rep})^2}{(\sigma_{it}^{rep})^2}$, which will have a $\chi^2_{T_i^{obs}}$ distribution if the model is correct, where T_i^{obs} is the total number of observed segments for the i^{th} woman. We assessed corresponding predictive p-values for these χ^2 test statistics based on 250 replications. Figure 5 shows the predictive p-values for all subjects. No subjects had a posterior predictive p-values greater than 0.95 and only one subject has a posterior predictive p-value smaller than 0.05. Review of subjects with low posterior predictive p-values show 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 toward larger values, yielding small χ^2 discrepancy statistics.

To consider the appropriateness of the final menstrual period modeling, we plot the observed and predicted FMPs together with the censoring ages for 100 randomly selected women in Figure 6. The method for estimating FMP when not observed appears to have worked well, with the distribution for the predicted FMPs corresponding closely to the observed FMPs when the censoring age is relatively early and little information is usually available to predict FMP.

5 Discussion

In this article we have provided a Bayesian changepoint model for describing the patterns of means and variances of women's menstrual segment lengths as they approach menopause. Our model detects individual changepoints of mean and variance of segment lengths for each individual woman. The model is applied to the TREMIN data. Multiple imputations integrated with an MCMC chain are carried out to impute the different kinds of missingness in the data set. 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 changepoints to be unknown parameters that vary for different subjects. This setting provides a flexible way of capturing both the mean and variability of each individual's segment length trajectory.

Our work develops a data-driven definition of early and late transition defined by subject-level variance and mean changepoints respectively. We observed a 3.2 year difference in age between mean and variance changepoints at the population level, somewhat shorter than that of Lisabeth et al. (2004b), 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. (2004b), 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 changepoints themselves, in particular that greater baseline variability was associated with more rapid declines in variability before variance changepoints and greater increases thereafter; and later mean changepoints were

associated with greater increases in mean length and more mean variability after mean changepoints. These data contribute to efforts to define a staging system for reproductive aging as they further our understanding of the timing and duration of the menopausal transition and describe the nature of heterogeneity in women's experience.

Our next step is to add the second TREMIN cohort data to assess changes of women's menstrual pattern in different generations by adding secular cohort (1935-1970 vs. 1960-1995) as a population-level covariate to the model. Also, while model checking showed that the model provides an adequate fit to the data, the model might still be improved. Distributions of individual level variance parameters (not shown) are somewhat skewed or heavy-tailed, suggesting a mixture distribution might be more suitable than one normal distribution for all subjects. Thus, a latent class model with subjects belonging to one of several underlying categories might fit the data even better. Estimation in the presence of left censoring is also of interest as many recent and ongoing studies enrolled prevalent cohorts including women who had already begun the menopausal transition.

Carroll (2003), in a paper entitled "Variances are not Always Nuisance Parameters," called for increased focus on developing methods for "variance structures" in order to better understand how "systematic dependence of variability on known factors" could yield both better prediction and improved inference. We agree with Carroll that incorporating information from subject-level variability in longitudinal data settings is underutilized in clinical and epidemiological research settings, at least in part because of the lack of methods for such analyses. In our application, it would be of interest to identify sub-groups of women who experience distinct patterns of variability during the menopausal transition and evaluate whether these subgroups also differ in their risk for developing chronic disease. We believe the analysis provided here begins to fill in some of the gaps in this area.

6 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, *Journal Quality of Life Research*, 13, 933–946.

Belsey, E. and Farley, T. (1987), The analysis of menstrual bleeding patterns: a review, *Applied Stochastic Models and Data Analysis*, 3, 125–150.

Brambilla, D., McKinlay, S., and Johannes, C. (1994), Defining the perimenopause for applications in epidemiologic investigations, *American Journal of Epidemiology*, 140, 1091–1095.

Carroll, R. (2003), Variance are not always nuisance parameters, *Biometrics*, 59, 211–220.

Chiazze, L., Brayer, F.T., Macisco, J.J., Parker, M.P., and Duffy, B.J. (1968), The length and variability of the human menstrual cycle, *Journal of American Medical Association.* 203, 377–380.

Crainiceanu, C., Ruppert, D., Carroll, R., Jpshi, A., and Goodner, B. (2007), Spatially adaptive Bayesian penalized splines with heteroscedastic errors, *Journal of Computational and Graphical Statistics*, 16, 265–288.

Davidian, M. and Carroll, R. (1987), Variance Function Estimation, *Journal of American Statistical Association*, 82, 1079–1091.

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.

Dennerstein, L., Smith, A., Morse, C., Burger, H., Green, A., Hopper, J., and M., R. (1993),

Menopausal symptomatology in Australian women, *Medical Journal of Australia*, 159, 232–236.

Gelman,A., Carlin,J., Stern,H., and Rubin,D. (2004). *Bayesian Data Analysis, 2nd Edition*, Chapman&Hall/CRC: Boca Raton, FL.

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

Gunn, L. and Dunson, D. (2005), A transformation approach for incorporating monotone or unimodal constraints, *Biostatistics*, 6, 434–449.

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, *Computational Statistics & Data Analysis*, 42, 91–109.

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

Harlow S.D., Crawford S., Dennerstein L., Burger H.G., Mitchell E.S., Sowers M.F. (2007), Recommendations from a multi-study evaluation o f proposed criteria for Staging Reproductive Aging, *Climacteric*, 10, 112–119.

Harlow, S.D., 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, *Journal of Clinical Epidemiology*, 53, 722–733.

Harlow, S.D., 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 Ephross, S. (1995), Epidemiology of menstruation and its relevance to women's health, *Epidemiology Review*, 17, 265–286.

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

Lisabeth, L., Harlow, S., and Qaqish, B. (2004b), A new statistical approach demonstrated that menstrual patterns during the menopausal transition did not vary by age at menopause, *Journal of Clinical Epidemiology*, 57, 484–496.

Little, R. and Rubin, D. (2002), *Statistical Analysis with Missing Data*, John 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, *Women's Health Issues*, 24, 220–226.

Matsumoto, S., Nogami, Y., and Ohkuri, S. (1962), Statistical studies on menstruation: A criticism on the definition of normal menstruation, *Gunma Journal of Medical Sciences.*, 11, 294–318.

Matsumoto, S., Tamada, T., and Konuma, S. (1979), Endocrinological analysis of environ-

mental menstrual disorders, *International Journal of Fertility*, 24, 233–239.

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

Small, C., Manatunga, A., Klein, M., Feigelson, H., Dominguez, C., McChesney, R., and Marcus, M. (2006), Menstrual cycle characteristics: associations with fertility and spontaneous abortion, *Epidemiology*, 17, 52–60.

Soules, M., Sherman, S., Parrott, E., Rebar, R., Santoro, N., Utian, W., and N., W. (2001), Stages of Reproductive Aging Workshop (STRAW), *Fertility and Sterility*, 76, 874–878.

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, *Journal of Clinical Endocrinology & Metabolism*, 91, 1261–1267.

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

Taylor, R. J., Berger, G., and Treloar, A. (1977), Changes in menstrual cycle length and regularity after use of oral contraceptives, *International Journal of Gynaecology and Obstetrics*, 15, 55–59.

Thum, Y. and Bhattacharya, S. (2001), Detecting a change in school performance: a Bayesian analysis for a multilevel joint point problem, *Journal of Educational and Behavioral Statistics*, 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. Proceedings of the VII World Congress, Tokyo and Kioto, Japan, Fertility and Sterility, pp. 886–890.

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

Vollman, R. (1977), *The menstrual cycle.*, Philadelphia, PA: W.B. Saunders and Company.

Wallace, R., Sherman, B., Bean, J., Treloar, A., and Schlabaugh, L. (1979), Probability of Menopause with Increasing Duration of Amenorrhea in Middle-Aged Women, *American Journal of Obstetrics and Gynecology*, 135, 1021–1024.

Wegienka, G. and Baird, D. (2003), Potential bias due to excluding oral contraceptive users when estimating menstrual cycle characteristics, *American Journal of Epidemiology*, 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, *American Journal of Epidemiology*, 158, 782–291.

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

Wildman, R., Colvin, A., Powell, L., Matthews, K., Everson-Rose, S., Hollenberg, S., Johnston, J., and Sutton-Tyrrell, K. (2008), Associations of endogenous sex hormones with the vasculature in menopausal women: the Study of Women's Health Across the Nation (SWAN),

Menopause, 15, 414–421.

APPENDIX

A Gibbs Sampling Algorithm

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

1. Initialize $\mathbf{\Phi}, \Lambda, \Omega$. Perform an initial imputation of missing data.

2. For $i = 1, ..., n$ and z_i consisting of both observed and imputed data:

2a.

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

where $W_i = Diag(\sigma_{it}^2)$, $A_i^\mu = \begin{pmatrix} 1 & (a_{i1} - 35) & (a_{i1} - \theta_i^\mu)_+ \\ \vdots & \vdots & \vdots \\ 1 & (a_{iT_i} - 35) & (a_{iT_i} - \theta_i^\mu)_+ \end{pmatrix}$, and Λ_μ and Ω_μ are the corresponding part of prior multivariate normal mean Λ and covariance matrix Ω conditional on other parameters.

2b.

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

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

2c.

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

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

2d.

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

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

2e.

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

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

2f.

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

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

3.

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

where X is the covariate matrix of all subjects, which consists of stacked rows of x_i', and $\mathbf{\Phi}$ consists of the stacked rows of Φ_i'.

4.

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

5. Use the updated parameters to create a new imputation data set. Then go to step 2.

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 that

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

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

Here, $A_0, ...A_K$ 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; we define $A_{-1} = 0$ and $A_{K+1} = \infty$ and assume $\eta_0 = 0$ (no risk of FMP before age 40).

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

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

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

Since we have no covariates, the hazard function for each interval is

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

The cumulative hazard and survival functions are then given by

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

$$S(t) = exp(-\Lambda(t))$$

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

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

$$\begin{cases} 1 - e^{-\eta_k(t_2 - t_1)} \text{ if } A_{k-1} \leq t_1 < t_2 \leq A_k \\ 1 - e^{-[\eta_k(t_2 - A_{k-1}) - \eta_{k-1}(t_1 - A_{k-1})]} \text{ if } A_{k-2} \leq t_1 < A_{k-1} \leq t_2 \leq A_k \\ 1 - e^{-[\eta_{k+1}(t_2 - A_k) + \eta_k(A_k - A_{k-1}) - \eta_{k-1}(t_1 - A_{k-1})]} \text{ if } A_{k-2} \leq t_1 < A_{k-1} < A_k \leq t_2 \end{cases}$$

Parameter	Λ (95%CI)
Mean intercept	3.313(3.306,3.321)
Mean slope before changepoint	-0.007(-0.010,-0.003)
Mean slope after changepoint	0.139(0.124,0.155)
Segment length mean changepoint	45.95(45.66,46.24)
Log-variance intercept	-4.814(-4.927,-4.704)
Log-variance slope before changepoint	0.016(-0.015,0.047)
Log-variance slope after changepoint	0.583(0.528,0.636)
Segment length variance changepoint	42.71(42.38,43.03)

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

	Mean intercept	Mean slope before changepoint	Mean slope after changepoint	Mean changepoint	Log-variance intercept	Log-variance slope before changepoint	Log-variance slope after changepoint	Variance changepoint
Mean intercept	1.00	**-0.11**	-0.02	**0.28**	**0.16**	**-0.12**	**0.15**	**0.27**
	–	**(-0.19,-0.03)**	(-0.14,0.09)	**(0.19,0.36)**	**(0.07,0.25)**	**(-0.22,-0.03)**	(0.06,0.24)	**(0.19,0.35)**
Mean slope		1.00	-0.02	-0.01	-0.04	0.02	-0.01	-0.01
before changepoint		–	(-0.10,0.07)	(-0.10,0.07)	(-0.12,0.04)	(-0.06,0.10)	(-0.10,0.07)	(-0.10,0.07)
Mean slope			1.00	**0.28**	0.07	-0.08	**0.31**	**0.26**
after changepoint			–	**(0.16,0.40)**	(-0.05,0.25)	(-0.20,0.05)	**(0.20,0.42)**	**(0.13,0.38)**
Mean changepoint				1.00	**0.16**	**-0.25**	**0.46**	**0.81**
				–	**(0.07,0.25)**	**(-0.37,-0.15)**	**(0.37,0.55)**	**(0.77,0.85)**
Log-variance					1.00	**-0.69**	**0.45**	0.04
intercept					–	**(-0.74,-0.65)**	**(0.38,0.53)**	(-.04,0.13)
Log-variance slope						1.00	**-0.75**	0.06
before changepoint						–	**(-0.79,-0.70)**	(-0.05,0.16)
Log-variance							1.00	**0.37**
slope after changepoint							–	**(0.29,0.45)**
Variance								1.00
changepoint								

	Loadings			
	PC1	PC2	PC3	PC4
Mean intercept	0.22	-0.14	0.63	-0.32
Mean slope before changepoint	-0.03	0.00	-0.61	-0.75
Mean slope after changepoint	0.24	-0.19	-0.44	0.55
Mean changepoint	0.45	-0.40	-0.01	-0.10
Log-variance intercept	0.36	0.45	0.09	-0.06
Log-variance slope before changepoint	-0.42	-0.50	0.02	0.03
Log-variance slope after changepoint	0.51	0.16	-0.15	0.06
Log-variance changepoint	0.36	-0.55	0.00	-0.10
Cumulative percent variance explained	35.3	56.1	69.9	81.9

Table 3: Principal components analysis of menstrual segment length parameters.

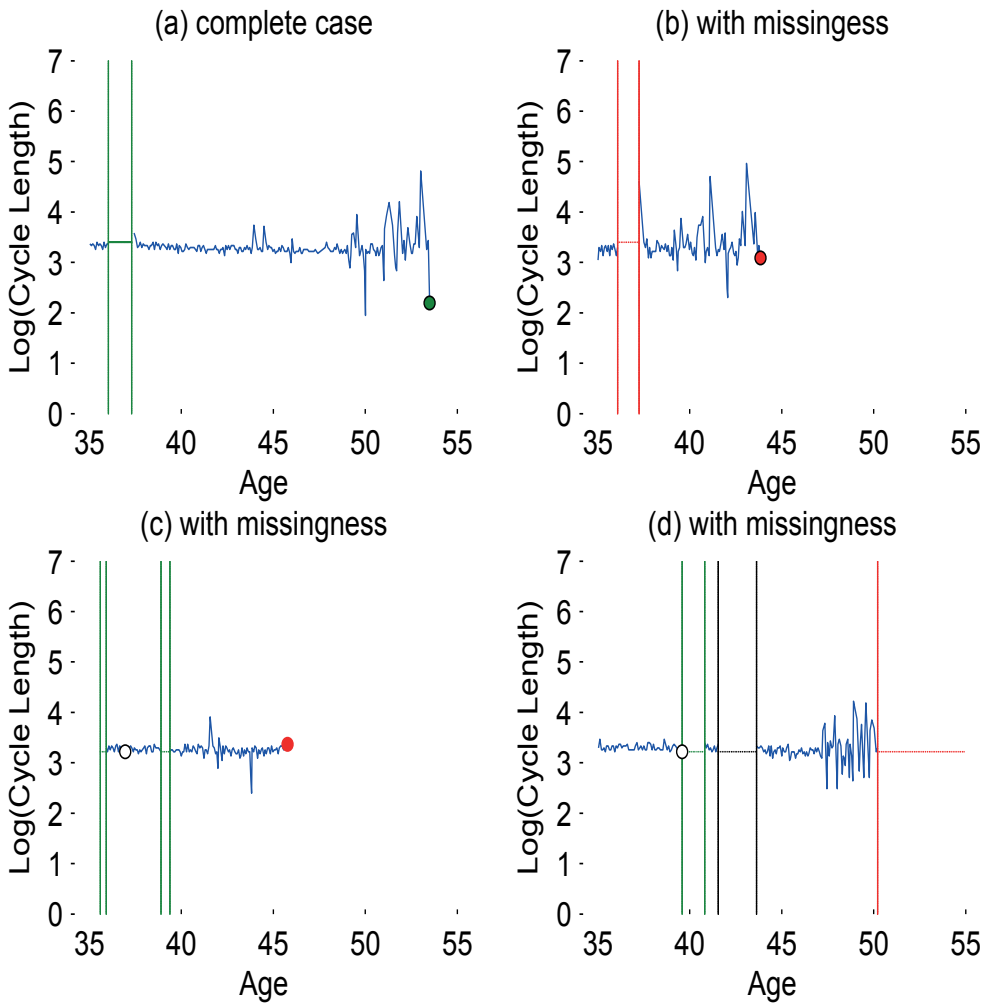

Figure 1: Four sampled women's log-segment-length trajectory after age 35: subject (a) has no missing data, the green gap is due to pregnancy and no imputation is needed. The black dot at the end means that FMP was observed for this subject. Subject (b) used hormones for a period of time, the red gap is due to hormone use. FMP is observed for this subject. Subject (c) has two pregnancy gaps (green gaps) and intermittent missingness at around age 36 (black circle). The red dot at the end represent that the subject's menstruation was truncated by surgery. Subject (d) has a pregnancy gap (green gap), an intermittent missingness (black circle) and a loss of contact gap (black gap). Her menstruation was censored due to hormone use (red line) and missing afterwards.

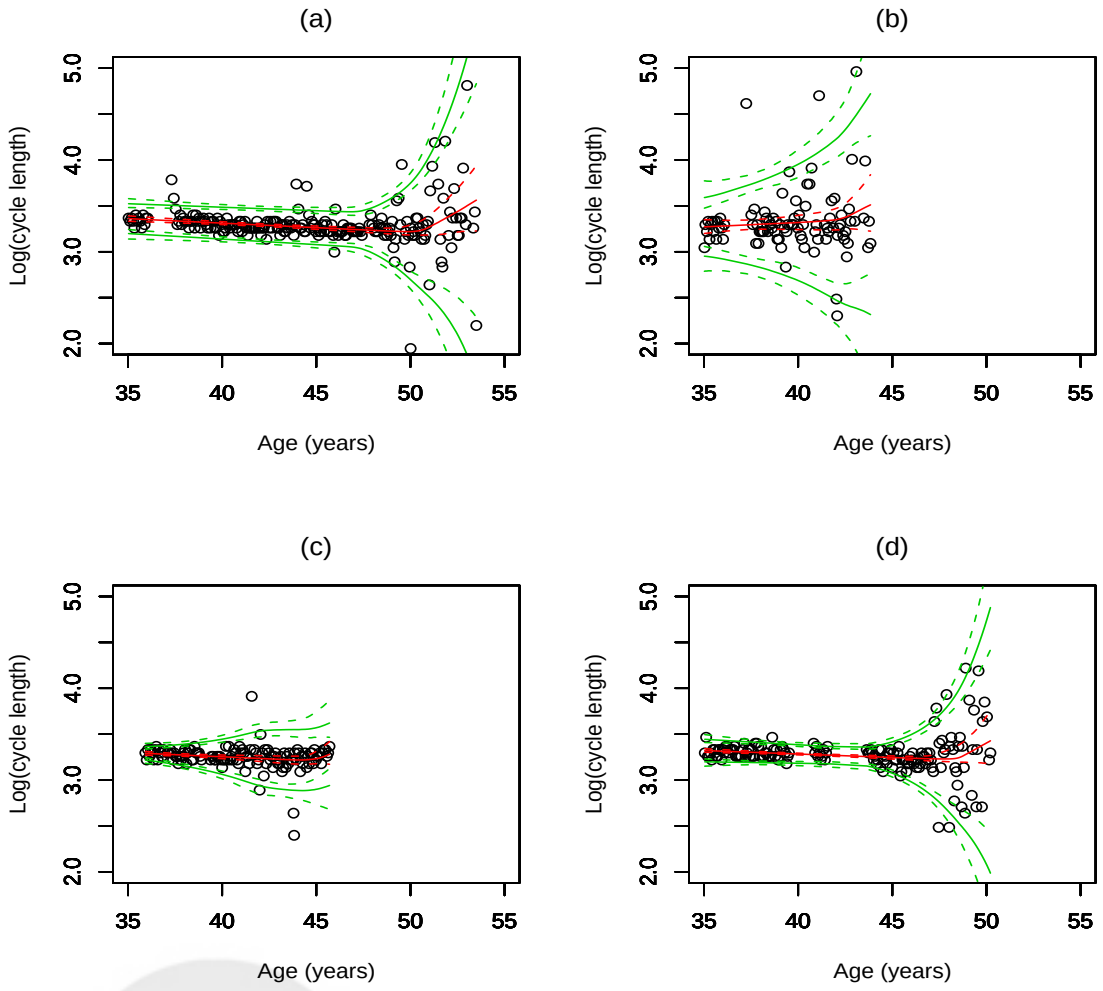

Figure 2: Data from four women: Red lines represent posterior mean of the mean segment length and associated 95% credible intervals; green lines represent posterior mean for the upper and lower 2.5 percentiles for the segment distribution and their associated 95% credible intervals. Black dots represent log of observed segment lengths.

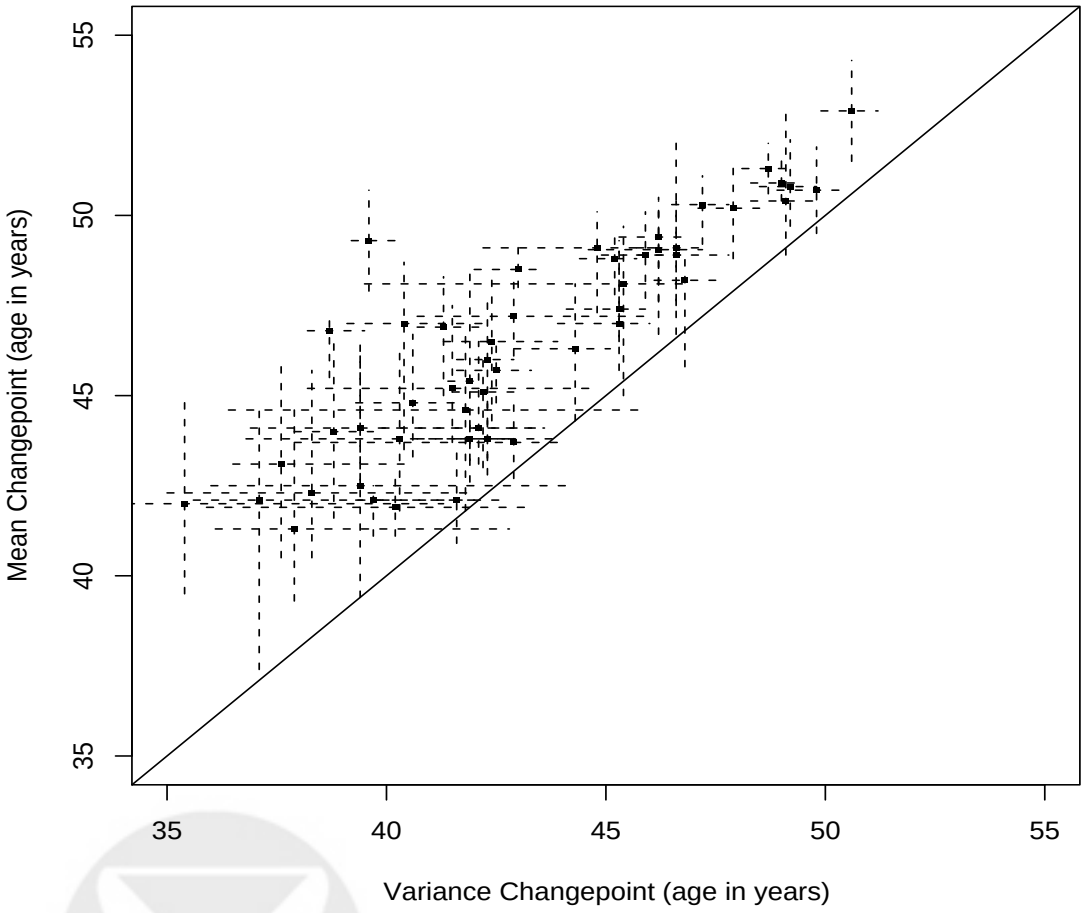

Figure 3: Posterior means and 90% posterior predictive intervals for mean changepoints and variance changepoints (100 randomly selected women).

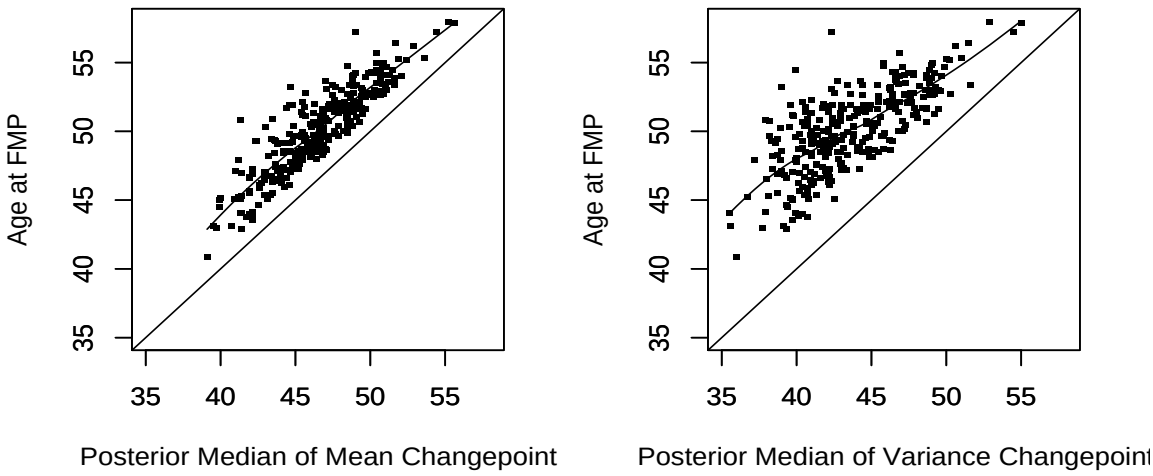

Figure 4: Posterior means of mean and variance changepoints versus final menstrual period for 304 women with observed final menstrual periods.

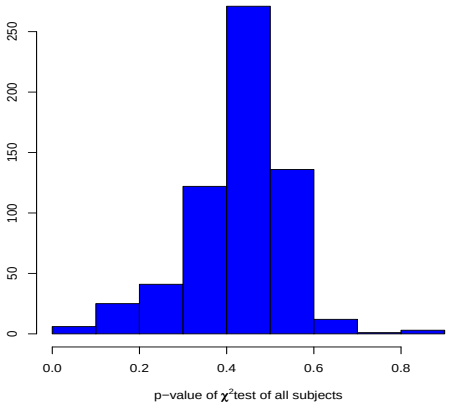

Figure 5: Histogram of p-values of subject level posterior predictive χ^2 tests.

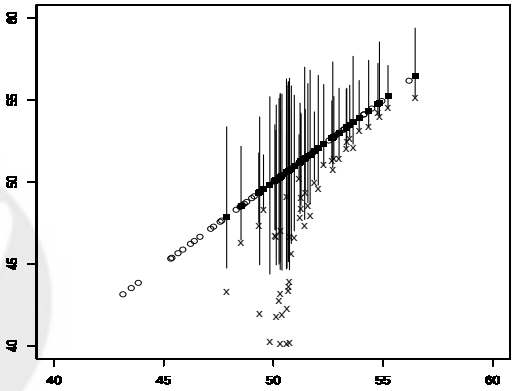

Figure 6: Observed FMP (circle) and posterior medians (squares) and 95% predictive interval for unobserved FMPs. X indicates age at censoring. (100 randomly selected women.)