Appendix 1

Analyses were implemented in Matlab7.0 (The MathWorks, Inc.), SAS version 9.1, SAS macro language and SAS/IML (SAS Institute, Cary, NC.).

A: Semi/non-parametric stochastic mixed model

Relationships between estradiol change and time to final menstrual period (FMP) could not be appropriately modeled by using quadratic or cubic terms; therefore, a semi-parametric stochastic mixed modeling was used. In general, the semi-parametric stochastic mixed model can be formulated by:

$$Y_{ij} = \boldsymbol{X}_{ij}^T \boldsymbol{\beta} + f(t_{ij}) + \boldsymbol{Z}_{ij}^T \boldsymbol{b}_i + U_i(t_{ij}) + \varepsilon_{ij} \qquad \text{(A.1)}$$

where:

$\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients associated with covariates $\boldsymbol{X}_{\text{ij}}$

$f(t)$ is a twice-differentiable smooth function of time;

$\boldsymbol{b}_{\text{i}}$ are independent $q \times 1$ vectors of random effects associated with covariates $\boldsymbol{Z}_{\text{ij;}}$

$U_i(t_{ij})$ are independent random processes used to model serial correlation;

ε_{ij} are independent measurement errors.

The fundamental assumptions for this model are: $\boldsymbol{b}_{\text{i}}$, $U_i(t_{ij})$, and ε_{ij} are mutually independent. $\boldsymbol{b}_{\text{i}}$ ~ normal (**0**, $\boldsymbol{D}$(**φ**)), $\boldsymbol{D}$ is a positive definite matrix depending on a parameter vector **φ;** $U_i(t_{ij})$ is a mean zero Gaussian process with covariance function or a non-homogeneous Ornstein-Uhlenbeck (NOU) process, cov($U_i(t)$,$U_i(s)$) = γ(**ζ**, α; t, s) depending on a parameter vector **ζ** and a scalar α, which is used to characterize the variance and correlation of the process $U_i(t)$; ε_{ij} ~ iid $N(0, \sigma^2)$.

In this study, the potential covariates considered included body mass index, smoking behavior, parity, and age at menarche.

More specifically, in order to capture the characteristics of $_{\log}$hormone mean and variance varying over time, the model of $_{\log}$hormone was formulated as:

$$Y_{ij} = f(t_{ij}) + b_{0i} + b_{1i}t_{ij} + U_i(t_{ij}) + \varepsilon_{ij} \tag{A.2}$$

where $U_i(t)$ is a non-homogeneous Ornstein-Uhlenbeck (NOU) process satisfying:

$$\text{var}(U_i(t)) = \xi(t) \text{ with } \log(\xi(t)) = A_0 + A_1 t + A_2 t^2 \tag{A.3}$$

$$corr(U_i(t), U_i(s)) = \rho^{|t-s|} \tag{A.4}$$

And assuming each woman's serial correlation wss the same. The smoothing function $f(t)$ represents the $_{\log}$hormone mean profile for the population of women over time.

B: Change characteristics of mean profile $_{log}$hormone trajectory

Hormone trajectory changes over time followi a non-linear pattern. The instantaneous changes of these trajectories (i.e., mean $_{\log}$hormone profile) can be characterized by rate of change, acceleration / deceleration, and curvature, where are first-, second-order derivatives of the mean curve, and the hinge/bend of the mean curve integrating the rate of change and acceleration, respectively. The cubic spline approach was used to estimate the rate of change as well as acceleration or deceleration.

Assume the time *t* was equally spaced with step $h = t_{k+1} - t_k$ ($k = 1,2,...,n-1$), where n was the total number of distinguished time points. Let $f(t)$ be the $_{\log}$hormone mean profile (trajectory) and $S_3(t)$ be its cubic spline approximation with the nice property:

$$\left| f^{(\alpha)}(t) - S_3^{(\alpha)}(t) \right| = O(h^{4-\alpha}), \alpha = 0,1,2,3 \tag{A.5}$$

The rate of change can be approximated by solving "m" equations:

$$m_{k-1} + 4m_k + m_{k+1} = 3\frac{f(t_{k+1}) - f(t_{k-1})}{h}, k = 2,3,...,n-1 \tag{A.6}$$

where $m_k = f'(t_k)$, $k = 2,3,...,n-1$. The m_1 and m_n can be approximated by 5-points method using first and last 5-points, respectively,

$$m_1 = \frac{1}{12h}\left[-25f(t_1) + 48f(t_2) - 36f(t_3) + 16f(t_4) - 3f(t_5)\right] \tag{A.7}$$

$$m_n = \frac{1}{12h}\left[3f(t_{n-4}) - 16f(t_{n-3}) + 36f(t_{n-2}) - 48f(t_{n-1}) + 25f(t_n)\right] \tag{A.8}$$

The acceleration / deceleration can be approximated by solving " M " equations:

$$M_{k-1} + 4M_k + M_{k+1} = 6\frac{f(t_{k+1}) - 2f(t_k) + f(t_{k-1})}{h^2}, k = 2,3,\ldots,n-1 \tag{A.9}$$

where $M_k = f''(t_k)$, $k = 2,3,\ldots,n-1$. The M_1 and M_n satisfied the boundary conditions

$2M_1 + M_2 = \frac{6}{h}\left(\frac{f(t_2) - f(t_1)}{h} - m_1\right)$ and $M_{n-1} + 2M_n = \frac{6}{h}\left(m_n - \frac{f(t_n) - f(t_{n-1})}{h}\right)$.

The 95% confidence bands of these characteristics were obtained using bootstrapping approach with 100 bootstrap samples by replicating the above processes.

***C:* Piecewise linear mixed model related time (e.g., time to final menstrual period (FMP))**

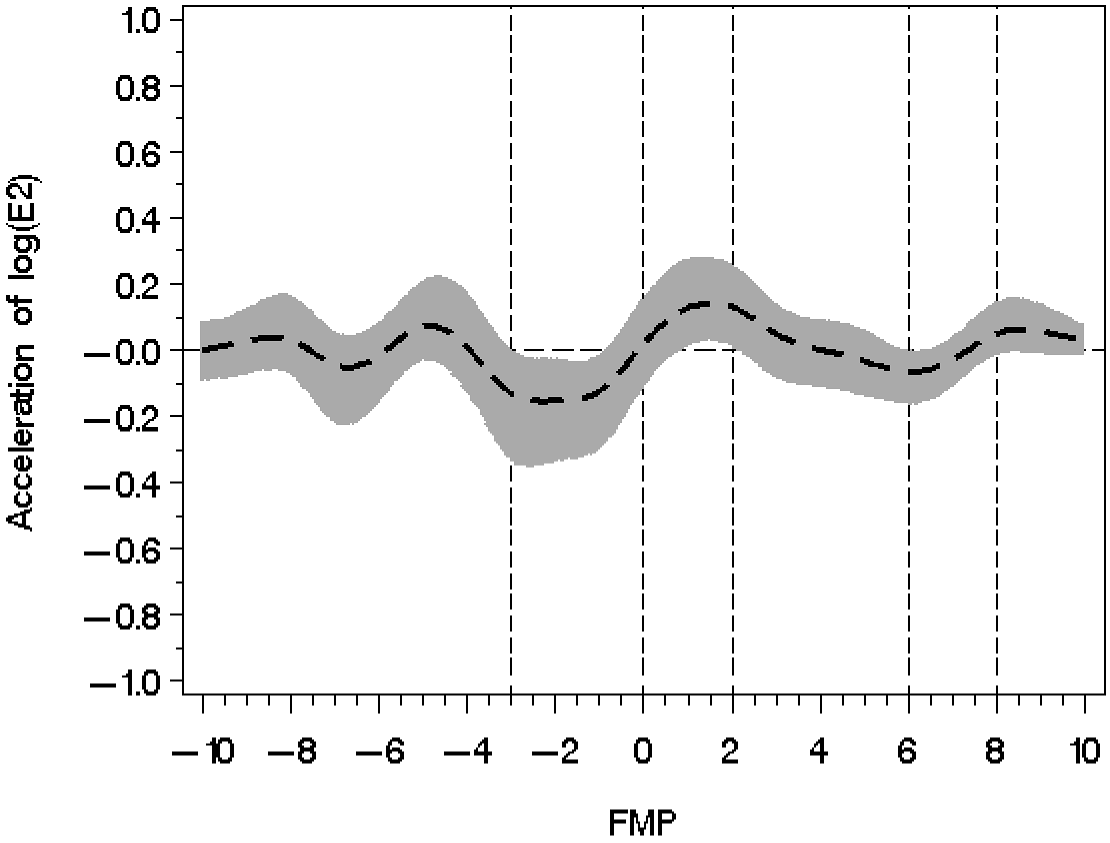
The nodes (or turning points) of the population hormone profile were identified based on the changing characteristics obtained by the above processes and they were further used to segment the hormone trajectory into stages. Piecewise linear mixed model was developed to capture these segmented characteristics (i.e., rate of change at each segment). Statistical comparisons of these slopes from two consecutive intervals around a turning point were tested to ascertain if one slope was different than the adjacent slope. The piece wise linear mixed model was formulated as follows.

Assume the independent variable of interest $t \in T \subset R^1$ (e.g., the time to FMP in this study). Let $\Omega = \left\{t^*_{(k)}, 1 \le k \le K \middle| t^*_{(1)} < t^*_{(2)} < \ldots < t^*_{(K)}\right\}$ be one known division of T, where K is the total number of turning points used to split T into $(K+1)$ non-overlapped intervals. The mean structure of piecewise linear mixed effect model was given by:

$$E[y_{ij}] = \beta_0 + \beta_1 t_{ij} + \sum_{k=1}^{K}\beta_k^{(1)}\left(t_{ij} - t^*_{(k)}\right)_+ + \sum_{l=1}^{L}\beta_l^{(2)}x_{ijl} + \sum_{l=1}^{L}\sum_{k=1}^{K}\beta_{l,k}\left(t_{ij} - t^*_{(k)}\right)_+ x_{ijl} \tag{A.10}$$

where $\left(t_{ij} - t_{(k)}^{*}\right)_{+} = \begin{cases} t_{ij} - t_{(k)}^{*} & , t_{ij} > t_{(k)}^{*} \\ 0 & , t_{ij} \le t_{(k)}^{*} \end{cases}$; $x_{ijl}\,(l = 1,2,...,L)$ are the covariates of interest. If there are no other covariates (e.g., pure "time" effect is of interest) then $\beta_{l}^{(2)}$ and $\beta_{l,k}$ can be dropped from the model. The random effects will be appropriately specified, e.g., random intercept and random slopes. The variance-covariance structure and model assumptions follow general linear mixed models.

Appendix Figure 1. Significant decelerations and accelerations occurred in ${}_{\text{log}}$estradiol (E2) levels at approximately 3 years prior to the FMP and 2, 6, and 8 years following the FMP as identified where broken vertical lines indicate where the 95% confidence intervals (in gray shaded bands) exclude the null values of no acceleration/deceleration.


Acceleration of log(E2)
1.0
0.8
0.6
0.4
0.2
−0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−10
−8
−6
−4
−2
0
2
4
6
8
10
FMP