VARIANCE AS A PREDICTOR OF HEALTH OUTCOMES: SUBJECT-LEVEL TRAJECTORIES AND VARIABILITY OF SEX HORMONES TO PREDICT BODY FAT CHANGES IN PERI- AND POST-MENOPAUSAL WOMEN

A PREPRINT

Irena Chen
Department of Biostatistics
University of Michigan
irena@umich.edu

Zhenke Wu
Department of Biostatistics
University of Michigan

Siobàn D. Harlow
Department of Epidemiology
University of Michigan

Carrie A. Karvonen-Gutierrez
Department of Epidemiology
University of Michigan

Michelle M. Hood
Department of Epidemiology
University of Michigan

Michael R. Elliott
Department of Biostatistics
University of Michigan

July 15, 2022

ABSTRACT

Longitudinal biomarker data and cross-sectional outcomes are routinely collected in modern epidemiology studies, often with the goal of informing tailored early intervention decisions. For example, hormones such as estradiol (E2) and follicle-stimulating hormone (FSH) may predict changes in womens' health during the midlife. Most existing methods focus on constructing predictors from mean marker trajectories. However, subject-level biomarker variability may also provide critical information about disease risks and health outcomes. In this paper, we develop a joint model that estimates subject-level means and variances of longitudinal biomarkers to predict a cross-sectional health outcome. Simulations demonstrate excellent recovery of true model parameters. The proposed method provides less biased and more efficient estimates, relative to alternative approaches that either ignore subject-level differences in variances or perform two-stage estimation where estimated marker variances are treated as observed. Analyses of women's health data reveal larger variability of E2 or larger variability of FSH were associated with higher levels of fat mass change and higher levels of lean mass change across the menopausal transition.

Keywords Estradiol · Follicle-stimulating hormone · Hamiltonian Monte Carlo · Joint models · Menopause · Subject-level variability · Study of Women's Health Across the Nation (SWAN) · Variance component priors

1 Introduction

The menopausal transition is a critical lifestage that can shape women's midlife and long-term health. Reproductive aging and the menopausal transition are characterized by decreasing levels of estradiol (E2) and increasing levels of follicle-stimulation hormone (FSH) [Randolph et al., 2004]. These two hormones, in addition to having important roles in regulating reproductive functionality, are also important for regulating other health outcomes [Karvonen-Gutierrez and Harlow, 2017]. E2 is known to regulate fat mass and a decline in E2 is associated with increased fat mass

[Colleluori et al., 2018]. Further, across the menopausal transition, women with obesity had lower E2 levels during the premenopause, but higher levels of E2 during post-menopause as compared to women without obesity [Randolph et al., 2011]. FSH has also been linked to health outcomes in women; for example, higher levels of FSH may be associated with increased risk of osteoporosis [Zaidi et al., 2018]. Additionally, FSH appears to be an important predictor of increased adiposity, reduced energy expenditure [Sponton and Kajimura, 2017, Kohrt and Wierman, 2017] and lower lean mass during the postmenopause [Gourlay et al., 2012]. Data from animal models suggest that blockage of the FSH receptor in mice was sufficient to reduce body fat and increase lean mass. [Liu et al., 2017].

A large and well-developed body of literature over the past two decades uses longitudinal biomarker or questionnaire data to predict health outcomes: an early example is given by Henderson et al. [2000], who tied psychiatric disorder measures over time to prediction dropout in schizophrenia trials via a subject level random effect in the disorder trajectory that is also present as a frailty in the time-to-event model. Proust-Lima et al. [2014] link latent classes of prostate specific antigen to survival models. More recent work by Wang et al. [2017] considers multiple longitudinal predictors - in their case, measures of daily functioning - to predict onset of Parkinson's disorder. However, methods that assess the ability of measures of residual variability are largely lacking in the literature, despite calls for increased focus on developing methods for variance structures. Such methods may elucidate whether systemic dependence of variability on known factors could yield both better prediction and improved inference [Carroll, 2003]. This oversight appears to be a substantial gap in the statistical methods literature given that increased variance can be an early predictor of instability in biological systems; for example, heart rate variability may be a marker of autonomic dysregulation given its predictive nature with poor health outcomes [Young and Benton, 2018].

The existing literature often uses a simple two-stage approach with a squared-error estimate of variance obtained from observed data to predict a single outcome, ignoring the inherent uncertainty in the constructed variance estimates. Joint modeling efforts are rarer but critical given that statistical uncertainty of the constructed predictors, e.g., mean and residual variance estimates, can lead to extremely biased estimates of association and uncontrolled sampling variances [e.g., Ogburn et al., 2021, Wang et al., 2020]. Finally, prediction approaches based on multiple trajectories - which allow for consideration of residual covariances as well as residual variances as predictors – appears to be completely absent from the biostatistics toolkit.

This paper is organized as follows. Section 2 provides a brief overview of previous work on variance as a predictor of health outcomes, highlighting the statistical consequences of ignoring the statistical uncertainty in multiple predictors for the outcome. Section 3 describes our joint model framework linking the mean and variance trajectory model for the longitudinal predictors with the model for the outcome. Section 4 conducts extensive simulation studies to show that our proposed approach produces less biased estimates of the outcome regression coefficients with proper control of the sampling variances relative to a variety of increasingly complex two-stage competitors that fail to account for the statistical uncertainty in the subject-level means and in the variances. Section 5 applies our method to assess the associations between the mean trajectories, variance, and covariance of E2 and FSH and changes in fat and lean body mass during the menopausal transition using data from the Study of Women's Health Across the Nation. Section 6 discusses the implications of our work along with limitations and directions for future research.

2 Previous Work on Variances as a Predictor of Outcomes

Joint modeling of longitudinal trajectories and cross-sectional outcomes is a rich area of statistical research [Chi and Ibrahim, 2006, Ibrahim et al., 2010, Lawrence Gould et al., 2015]. For example, including longitudinal measurements in joint longitudinal - survival models can provide better estimates of the survival outcome by accounting for the individual trajectories over time, as shown by Long and Mills [2018], who used a joint model to predict motor diagnosis using longitudinal characterizations of Huntington's disease progression taken at annual visits; also see Papageorgiou et al. [2019] for a comprehensive review. Until recently, most of the focus in the area of modeling individual trajectories has been on modeling the mean trends. The variability associated with individual biomarkers has traditionally been treated as a nuisance parameter.

However, interest in both modeling the residual variability of individual trajectories and using these estimates as predictive variables has been growing. Elliott et al. [2012] studied the relationship between individual variability in short-term memory tests and long term onset of senility. They found that increased variability in the memory tests were associated with increased risk of senility. Furthermore, this variability was a stronger predictor of senility than the mean trajectories. With regards to women's health, some previous research has considered variance as a predictor of health outcomes. Harlow et al. [2000] found that women who had increased menstrual cycle variability at a younger age were more likely to experience abnormal uterine bleeding. The variability of menstrual cycle length was also found to be an important predictor of menopausal onset [Huang et al., 2014]. Furthermore, Sammel et al. [2001] found that E2 variability during the menopausal transition was highly predictive of experiencing hot flashes. Jiang et al. [2015]

modeled the individual mean and variances of FSH hormone levels over time to predict risk of hot flash and found that the variance of FSH was strongly associated with the risk.

Our work makes two primary contributions to the existing statistical literature: 1) we model individual-level variances and co-variances of multiple longitudinal trajectories by proposing flexible hierarchical priors for residual variance-covariance matrices that may flexibly vary by subject and 2) we incorporate these subject-level means and variance-covariance matrices for longitudinal data and a cross-sectional outcome in a joint modeling framework, rather than a two-stage approach that we show to be statistically inferior. Also, to the best of our knowledge, this is the first rigorous assessment of the role of individual variability of E2 and FSH hormones in predicting body mass composition changes.

2.1 Measurement Error in Multivariate Linear Models

An advantage of joint models relative to two-stage models is that the uncertainty associated with the parameters estimated in the first stage is carried over to the second stage. Consider a simple linear relationship of the form:

$$Y = X\beta + \epsilon, \tag{1}$$

where X is a $n \times K$ matrix of K predictors and ϵ is an $n \times 1$ vector of independent normal error terms with mean 0 and variance σ^2. Suppose that the true relationship between Y and X is described by (1), but instead, we observe $\widetilde{X}$, where $\widetilde{X} = X + U$, where U is the matrix of normally-distributed independent measurement errors with mean 0 and variance-covariance Σ_U.

If $U \perp\!\!\!\perp X$, then in the $K = 1$ scenario, we know that the estimate of β will be attenuated towards the null [Carroll et al., 2006, p.42-43]. For $K > 1$, with multiple predictors measured with error, the estimates of the β are still biased, but the direction of the bias now depends on the correlation between the measurement errors [Carroll et al., 2006, p.53-55]. Consider the following equation for $K = 2$ predictors:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon, \tag{2}$$

and suppose we measure X_1, X_2 with some error:

$$\widetilde{X}_1 = X_1 + U_1, \widetilde{X}_1 = X_2 + U_2, \tag{3}$$

Griliches and Intriligator [1987, p.1477–1479] derive the bias of estimating β_1 as:

$$\text{plim}(\widehat{b}_1 - \beta_1) = -\frac{\beta_1 \lambda_1}{(1-\rho^2)} + \frac{\beta_2 \lambda_2 \rho}{(1-\rho^2)}, \tag{4}$$

where $\widehat{b}_1$ is the coefficient obtained from regressing Y on $\widetilde{X}_1$ in the multiple regression model, $\lambda_1 = \text{Var}(U_1)/\sigma_1^2$ is the relative amount of measurement error in X_1 where σ_1^2 refers to the variance of X_1, and ρ is the (true) correlation between X_1, X_2; $plim$ refers to convergence in probability. A similar equation can be derived for the bias of estimating β_2 in the presence of such measurement errors. We can see, then, that the bias is increased by a factor of $\frac{1}{(1-\rho^2)}$.

The overall effect of the additional variable $\widetilde{X}_2$ is a bias towards the null [Griliches and Intriligator, 1987, p.1479]. For $K > 2$ variables, the expressions for the bias of each predictor become more complicated to derive. We show via simulations in Section 4.1.1 that the bias in the mean parameters is clear in the two-stage model linear regression alternatives to the joint model. Furthermore, we see that that bias in the estimates of the variance-covariance parameters persists in the two-stage model alternatives. This would be an issue if we believe that individual variability (covariability) is actually predictive of an outcome of interest.

3 Proposed Model

3.1 Notation

Let $\mathcal{D} = (Y_i, \boldsymbol{X}_{ij}, t_{ij}, \boldsymbol{W}_i)$ be the observed data for subject $i = 1, \ldots, N$, where $Y_i \in \mathbb{R}^1$ is a continuous outcome, $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijQ})^\mathsf{T}$ is a vector of Q time-varying marker values at observation time points $t_{ij}, j = 1, \ldots, n_i$, that may differ by subjects, and $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{id})^\mathsf{T}$ is a vector of d time-invariant covariates, e.g., race/ethnicity, activity class.

The proposed model has two connected components. We first specify the regression model for irregularly and longitudinally observed multiple markers $(t_{ij}, \boldsymbol{X}_{ij}), j = 1, \ldots, n_{ij}$; the second component links the outcome Y_i to time invariant covariates $\boldsymbol{W}_i$ and unobserved individual-specific vectors of regression coefficients and residual variance-covariance matrices in the model for longitudinal markers, enabling inference about how the mean trajectories and residual variations are associated with the outcome Y_i.

3.2 Likelihood

3.2.1 Component 1: Longitudinal Markers

We specify the model for the longitudinal marker data as follows:

$$\boldsymbol{X}_{ij} \mid \mathbf{B}_i, \mathbf{S}_i \sim \mathcal{N}_Q\left(\boldsymbol{\mu}(t_{ij}; \mathbf{B}_i), \mathbf{S}_i\right), j = 1, \ldots, n_i, \text{ independently for } i = 1, \ldots, N, \tag{5}$$

$$\boldsymbol{b}_{iq} \overset{\text{indep.}}{\sim} \mathcal{N}_P(\boldsymbol{\beta}_q, \Sigma_q), q = 1, \ldots, Q \tag{6}$$

where $\mathcal{N}_Q(\boldsymbol{\mu}, \mathbf{S})$ is a generic notation that represents a Q-dimensional multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{S}$; $\boldsymbol{\mu}(t; \mathbf{B}_i)$ is a Q-dimensional function of time given by $\mathbf{B}_i = [\boldsymbol{b}_{i1}, ..., \boldsymbol{b}_{iQ}]^\mathsf{T}$ and $\boldsymbol{b}_{iq} = (b_{iq1}, \ldots, b_{iqP})^\mathsf{T}$ is a vector of P regression coefficients for the q-th marker. Here P is the number of basis functions of time; to simplify presentation in this paper, we assume the same number of coefficients for each marker, e.g., intercept and slope. In addition, $\boldsymbol{\beta}_q = (\beta_{q1}, \ldots, \beta_{qP})^\mathsf{T}$ is a vector of population mean regression coefficients that are specific to the q-th marker. In addition, Equation (6) has assumed individual-specific regression coefficients for any two markers are conditionally independent.

3.2.2 Component 2: Outcome Regression (OR) Model

The outcome variable Y_i is assumed to be related to individual-specific mean and variance-covariance parameters $\mathbf{B}_i$ and $\mathbf{S}_i$ in the longitudinal marker model (5) as follows:

$$Y_i \mid \mathbf{B}_i, \mathbf{S}_i, \boldsymbol{W}_i \sim \mathcal{N}\left(\eta_i, \sigma^2\right), i = 1, \ldots, N, \tag{7}$$

where $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{id})^\mathsf{T}$ is a vector of d time-invariant covariates; $\eta_i = \eta(\mathbf{B}_i, \mathbf{S}_i, \boldsymbol{W}_i; \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W)$ and $\eta(\cdot; \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W)$ is a generic mean outcome regression function parameterized by $\boldsymbol{\alpha}$ (for $\mathbf{B}_i$), $\boldsymbol{\gamma}$ (for $\mathbf{S}_i$), and $\boldsymbol{\gamma}^W$ (for $\boldsymbol{W}_i$); $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\gamma}^W$ are of dimension PQ, $Q(Q+1)/2$, d, respectively. In this paper, we will illustrate the statistical performance of such a formulation by focusing on simple specifications of $\eta(\cdot)$, e.g., linear models. The framework readily generalizes to general outcome models; in Section 5.3, we illustrate a model with a Gaussian scale-mixture outcome regression model for lean mass rate-of-change outcomes.

3.3 Priors

In this section, we specify the prior and hyperprior distributions for the unknown parameters in the two likelihood components.

Prior for $\mathbf{S}_i$

We rewrite $\mathbf{S}_i = \mathbf{D}_i \mathbf{R}_i \mathbf{D}_i$, where $\mathbf{D}_i = \text{diag}(d_{i1}, ..., d_{iQ})$ is a diagonal matrix of residual variances and $\mathbf{R}_i$ is the associated subject-level correlation matrix. We assume

$$\log(d_{iq}) \sim \mathcal{N}(\nu_q, \psi_q^2), \tag{8}$$

independently for marker $q = 1, \ldots, Q$, and subject $i = 1, \ldots, N$. Because $\mathbf{R}_i$ is a correlation matrix, we only need to specify the prior distribution for the off-diagonal elements. We consider the special case of $Q = 2$, where r_{12i} is unconstrained, and separately the general case of $Q > 2$, where the components of $\mathbf{R}_i$ must meet the positive definite criterion.

Special Case: $Q = 2$.

For a 2×2 correlation matrix, we place the following prior on the off-diagonal value r_{12}:

$$(r_{i12} + 1)/2 \sim \mathsf{Beta}(a'_{12}, b'_{12}), \text{ independently for } i = 1, \ldots, N. \tag{9}$$

Finally, we specify hyperpriors for ν_q, ψ_q and (a'_{12}, b'_{12}) by

$$\nu_q \sim \mathcal{N}(m, \xi_q^2), \psi_q \sim \textsf{half-Cauchy}(0, \tau), \text{ independently for } q \leq Q, \tag{10}$$

$$a'_{12} \sim \mathsf{Exp}(\kappa), \; b'_{12} \sim \mathsf{Exp}(\kappa'). \tag{11}$$

In our application, we set $m = 0$ and $\xi_q = 10$. In addition, we use a half-Cauchy hyperprior on ψ_q instead of the inverse-Gamma distribution as this prior is recommended for datasets where the variance ψ_q may be small [Gelman, 2006]. In this setting, inferences using the inverse-Gamma distribution are extremely sensitive to the choice of hyperparameter values [Gelman, 2006, p. 524], which makes the inverse-Gamma prior "not at all uninformative". The half-Cauchy distribution avoids this potential issue due to its heavier tail, which still allows for higher variance values.

Remark 1. *The defining feature of our framework is the individual-specific variance-covariance matrices, $\{\boldsymbol{S}_i, i = 1, \ldots, N\}$, over which we must specify a hierarchical prior distribution. Such priors must not be restrictive in capturing between-subject differences and similarities in the variance-covariance matrices. Focusing on the prior specification for individual-specific correlation matrices $\{\mathbf{R}_i, i = 1, \ldots, N\}$, standard priors designed for a single unknown population correlation matrix, e.g., Lewandowski-Kurowicka-Joe (LKJ) prior [Lewandowski et al., 2009], have severe drawbacks. In particular, the LKJ distribution is governed by a single positive scale parameter, ζ, that tunes the strengths of the correlations. The off-diagonal elements of a $K \times K$ correlation matrix are marginally distributed as: $(r_{ilk} + 1)/2 \sim \mathsf{Beta}(\mathsf{a}, \mathsf{b})$, where $a = b = \zeta - 1 + K/2$. This implies that the correlations will a priori be concentrated around 0. However, in our motivating application, $\{r_{ilk}, i = 1, \ldots, N\}$ represent the individual-specific residual correlations between the l- and the k-th hormone, which 1) by domain knowledge are a priori unlikely to have a strong prior of being near zero, and 2) may vary between subjects in a way far from the implied distribution of $2 \cdot \mathsf{Beta}(a, b) - 1$. In Equation (11), we have removed this identity restriction and specified hyperpriors on a and b, which provides greater flexibility in allowing the data to estimate the true a and b. The same argument can be applied to the Inverse-Wishart distribution for variance-covariance matrices, which is also governed by a single scale parameter.*

In addition, we argue that the Exponential distribution in Equation (11) is a sensible hyperprior on the Beta parameters as follows: Let $x_1, \ldots, x_n$ be data from a $\mathsf{Beta}(a, b)$ distribution. Assume that $a \sim \mathsf{Exp}(\lambda_a), b \sim \mathsf{Exp}(\lambda_b)$. We can also assume a and b are independent *a priori*. Then the posterior distribution, $p(a, b|x) \propto L(a, b)p(a)p(b) \propto \prod_{i=1}^n \exp(-\ln B(a, b) + (a-1)\ln(x_i) + (b-1)\ln(1-x_i) \times \exp(\ln(\lambda_a) - \lambda_a a + \ln(\lambda_b) - \lambda_b b) \propto \exp(-n \ln B(a, b) + a\sum_{i=1}^n(\ln x_i - \lambda_a) + b\sum_{i=1}^n(\ln(1 - x_i) - \lambda_b)$, which suggests that the posterior distribution of a, b would be updated from a flat prior by subtracting (λ_a, λ_b) from the sufficient statistics $\sum_{i=1}^n \ln x_i$ and $\sum_{i=1}^n \ln(1 - x_i)$. The reason for this is that the Maximum Likelihood Estimators (MLEs) for a, b respectively can be approximated by: $\widehat{a}_{MLE} = \dfrac{1}{2} + \dfrac{\widehat{G}_x}{1 - \widehat{G}_x - \widehat{G}_{(1-x)}}, \widehat{b}_{MLE} = \dfrac{1}{2} + \dfrac{\widehat{G}_{(1-x)}}{1 - \widehat{G}_x - \widehat{G}_{(1-x)}}$ where $\widehat{G}_x = \exp(n^{-1}\sum_{i=1}^n \ln(x_i)), \widehat{G}_{1-x} = \exp(n^{-1}\sum_{i=1}^n \ln(1 - x_i))$, where $(a, b) > 1$. The posterior modes using the Exponential priors become $\dfrac{1}{2} + \dfrac{\widehat{G}_x \exp(\lambda_a)}{1 - \widehat{G}_x \exp(\lambda_a) - \widehat{G}_{(1-x)} \exp(\lambda_a)}, \dfrac{1}{2} + \dfrac{\widehat{G}_{(1-x)} \exp(\lambda_b)}{1 - \widehat{G}_x \exp(\lambda_b) - \widehat{G}_{(1-x)} \exp(\lambda_b)}$. When $\lambda_a, \lambda_b \to \infty$, the posterior modes of a, b shrink towards $\dfrac{1}{2}$, which is the Jeffrey's prior. When $\lambda_a, \lambda_b \to 0$, we recover the likelihood. Overall, these results suggest that the choice of the Exponential distribution is a flexible hyperprior on a, b and thus is a reasonable choice.

General Case: $Q \geq 3$.

For the general case of a $Q \times Q$ correlation matrix where $Q \geq 3$, the off-diagonal values of the individual correlation matrices are now more complicated to estimate since the space of valid correlation matrices is a proper subset of the space of all possible $Q \times Q$ matrices. We address this constraint by following the approach of Ghosh et al. [2021] where the off-diagonal values are parameterized in terms of hyperspherical coordinates. The angles are allowed to vary freely over $[0, \pi]$ before being back-transformed into valid correlation values. To illustrate, we specify the prior for $\mathbf{R}_i$ for when $Q = 3$ as follows (similarly for $Q > 3$):

$$r_{12} = \cos(\theta_{12}), r_{13} = \cos(\theta_{13}), r_{23} = \sin(\theta_{12}) \cdot \sin(\theta_{13}) \cdot \cos(\theta_{23}) + \cos(\theta_{12}) \cdot \cos(\theta_{13}),$$

where $\theta_{12} = \arccos(c_{12}), \theta_{13} = \arccos(c_{13}), \theta_{23} = \arccos(c_{23})$, and

$$(c_{i12} + 1)/2 \quad \sim \mathsf{Beta}(a'_{12}, b'_{12}), (c_{i13} + 1)/2 \sim \mathsf{Beta}(a'_{13}, b'_{13}), (c_{i23} + 1)/2 \sim \mathsf{Beta}(a'_{23}, b'_{23}).$$

As in the case where $Q = 2$, we specify hyper-priors for $a'_{kl}, b'_{kl}, k < l$, e.g., the Exponential prior.

Priors for population longitudinal marker regression coefficients:

$$\boldsymbol{\beta}_q \sim \mathcal{N}_P(0, \xi_q^2 I_{P \times P}), \text{ independently for } q = 1, \ldots, Q \tag{12}$$

$$\Sigma_q = \mathbf{K}_q \mathbf{L}_q \mathbf{K}_q, \ \mathbf{K}_q = \mathrm{diag}\{k_{q1}, \ldots, k_{qP}\}, q = 1, \ldots, Q, \tag{13}$$

$$k_{qp} \sim \text{half-Cauchy}(0, \tau_0), p = 1, \ldots, P, \text{ and } \mathbf{L}_q \sim \mathsf{LKJ}(\zeta), \text{ independently for } q = 1, \ldots, Q \tag{14}$$

where $\mathbf{K}_q = \mathrm{diag}\{k_{q1}, \ldots, k_{qP}\}$ is a diagonal matrix and $\mathbf{L}_q$ is a correlation matrix. The τ_0, ζ parameters are set in practice as 2.5 and 1 respectively. It is fine to use the half-Cauchy and the LKJ priors in Equation (14) since, for each marker q, they are standard hyperpriors for a *single* population variance matrix $\mathbf{K}_q$ and a *single* population-level

correlation matrix $\mathbf{L}_q$, which is different from Equation (11) that specifies the prior for multiple and individual-specific variance-covariance matrices. In addition, Gelman [2006] shows that the Inverse-Gamma prior distribution is overly informative for a hierarchical variance and this issue persists for an Inverse-Wishart on a hierarchical covariance matrix. We therefore choose to specify half-Cauchy for variances $\{k_{qp}, p = 1, \ldots, P\}$, and the LKJ distribution for $\mathbf{L}_q$.

Prior for parameters in the outcome regression model For the outcome model, we place diffuse independent Gaussian priors for each element of the outcome regression parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^W$). Finally, to complete the prior specification, we place a diffuse prior on the outcome residual standard deviation parameter $\sigma \sim \textsf{half-Cauchy}(0, \tau_1)$. In our simulation studies and data analysis, we set the priors on the regression parameters as $\mathcal{N}(0, 10^2)$ (a very diffuse prior in order to allow the data to estimate the parameters) and $\tau_1 = 2.5$ (the default suggested by Carpenter et al. [2017]).

Let $\Theta = (\boldsymbol{\beta}_q, \Sigma_q, \xi, \nu_q, \psi_q, a'_{kl}, b'_{kl}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W, \sigma)$ denote the unknowns of interest in the proposed model. Let $\pi(\Theta)$ denote the prior distribution where we have assumed that all parameters in Θ have independent components:

$\pi(\Theta) = \Pi_{q=1}^{Q}[\pi(\boldsymbol{\beta}_q)\pi(\Sigma_q)\pi(\xi_q)\pi(\nu_q, \psi_q)]\Pi_{k<l}[\pi(a'_{kl}, b'_{kl})]\pi(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W))\pi(\sigma)$.

Joint Distribution The joint distribution of the data and unknowns is

$$
\begin{aligned}
P(\Theta, \mathcal{D}) \propto \prod_{i=1}^{n}\prod_{q=1}^{Q} \Big\{ & \frac{1}{\sqrt{(2\pi)^q|\Sigma|}}\exp(-\frac{1}{2}(\boldsymbol{b}_{iq} - \boldsymbol{\beta}_q)^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{b}_{iq} - \boldsymbol{\beta}_q)) \\
& \times \frac{1}{\sqrt{2\pi\xi_q^2}}\exp\left[\frac{(\log(d_{iq}) - \nu_q)^2}{2\xi_q^2}\right]\left(\frac{[(r_{ikl}+1)/2]^{a'_{kl}-1}\{1 - ([(r_{ikl}+1)/2]\}^{b'_{kl}-1}}{\textsf{Beta}(a'_{kl}, b'_{kl})}\right) \\
& \times \prod_{j=1}^{n_i}\frac{1}{\sqrt{(2\pi)^{n_i}|\mathbf{S}_i|}}\exp\left(-\frac{1}{2}\{\boldsymbol{X}_{ij} - \boldsymbol{\mu}(t_{ij}; \mathbf{B}_i)\}^{\mathsf{T}}\mathbf{S}_i^{-1}\{\boldsymbol{X}_{ij} - \boldsymbol{\mu}(t_{ij}; \mathbf{B}_i)\}\right)\Big\} \\
& \times \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{(Y_i - \eta(\mathbf{B}_i, \mathbf{S}_i, \boldsymbol{W}_i; \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W))^2}{2\sigma^2}\right] \times \pi(\Theta). \qquad (15)
\end{aligned}
$$

Figure S1 in the Supplementary Materials [Chen et al., 2022] uses a directed acyclic graph to visualize and summarize the hierarchical relationships between the different components of our modeling framework.

3.4 Posterior Inference

In a Bayesian framework, the inference is conducted based on the posterior distribution $P(\Theta \mid \mathcal{D})$. However, it is not feasible to derive the closed-form posterior distribution owing to the lack of prior-likelihood conjugacy in our proposed model. We therefore used Hamiltonian Monte Carlo to draw sequential samples and approximate the posterior distribution. We implement the model using Stan and the `rstan` package [Stan Development Team, 2020] as the interface for running the model and obtaining the posterior estimates. Code to run the joint model and generate the data used in our simulation studies are provided in attached supplementary files. For our simulations studies in Section 4.1, 4.2 and Section 3 in the Supplementary Material [Chen et al., 2022], we run two chains per independent replicate data set, with $2,000$ iterations and $1,000$ burn-in. For the data application in Section 5, we ran 4 chains each and 4,000 iterations with 2,000 burn-in. Visual inspection of the traceplots for all model parameters indicated non-divergent chains. All chains were combined for calculating posterior summaries.

We also examined Stan's R-hat convergence diagnostic [Vehtari et al., 2021] and the Effective Sample Sizes (ESS) to determine if the chains had mixed well. The R-hat value for all model parameters was less than 1.05. In the fat mass rate of change model, the ESS for all of the model parameters was at least 100 times the number of chains used, except for the model parameter corresponding to individual-level E2 variance and the parameter corresponding to fat mass proportion at the first visit. This was also the case for the lean mass rate of change model (E2 variance and lean mass proportion at the first visit). However, the ESS values for these parameter were still above the recommended sample size number, $5\widehat{m}$, where $\widehat{m}$ is the number of chains [Gelman et al., 2013, p.384]. The R-hat values of these parameters was also effectively 1.00 in both models. Based on these diagnostics, we concluded that our models had converged. The posterior predictive checks we conducted (see S@ in the Supplementary Material, Chen et al. [2022]) suggest that our model generates reasonable estimates for the observed outcomes and trajectories.

4 Simulation Study

We conducted simulation studies to 1) evaluate our model's operating characteristics and 2) compare against common alternatives that could also be used in modeling individual means and variances as predictors of outcomes. We evaluated our model performance across independent simulation replicates using three criteria: for each method and each parameter θ, we assess the 1) bias (defined as $1/R\sum_{r=1}^{R}(\widehat{\theta}^{(r)}-\theta_0)$ where $\widehat{\theta}^{(r)}$ is the posterior mean of θ obtained from the r-th replication), 2) the coverage rate of the nominal 95% credible intervals (CrI; defined as $1/R\sum_{r=1}^{R}\mathbb{1}\{\theta_0\in I_r\}$ where I_r is the 95% CrI for parameter θ obtained by computing the 2.5% and 97.5% percentiles of the draws from the posterior distribution for the r-th replication, and 3) average length of the 95% CrIs obtained across simulation replicates, defined as $1/R\sum_{r=1}^{R}T_r$, where T_r is the length of I_r, i.e., the range of the estimated 2.5% and 97.5% posterior quantiles for θ in replicate r.

In this section, we present the results from simulation studies with $Q=2$ and $Q=3$ biomarkers. Additionally, in Section S3 in the Supplementary Material [Chen et al., 2022], we present a simulation study examining our model performance when we approximate the true nonlinear relationship between the marker means and variances and the outcome as linear.

4.1 Simulation 1: Two Biomarkers.

In this simulation, we assume the mean trajectories can be expressed linearly with individual intercepts and slopes. We generated $n_i=6$ to 15 time points for $N=1,000$ individuals, which mimics the data used in Section 5. We then simulated two trajectories for each individual using the following parameters.

Component 1: Longitudinal Markers

$$X_{itq}=b_{iq1}+b_{iq2}t+\epsilon_{iq}, q=1,2; \boldsymbol{B}_{i1}\sim\mathcal{N}_2\left(\boldsymbol{\beta}_1,\Sigma_1\right), \boldsymbol{B}_{i2}\sim\mathcal{N}_2\left(\boldsymbol{\beta}_2,\Sigma_2\right)$$
$$\boldsymbol{\beta}_1=(0,2)^{\mathsf{T}}, \boldsymbol{\beta}_2=(2,1)^{\mathsf{T}}, \Sigma_1=\begin{pmatrix}1 & -0.05\\ -0.05 & 1\end{pmatrix}, \Sigma_2=\begin{pmatrix}1 & -0.1\\ -0.1 & 0.5\end{pmatrix},$$
$$\log(d_{i1})\sim\mathcal{N}(0,0.75)/2, \log(d_{i2})\sim\mathcal{N}(0.5,0.5)/2, (r_{i12}+1)/2\sim\mathsf{Beta}(1,5).$$

Component 2: Outcome Regression Model

To generate the outcome for each individual, we assume $Y_i\sim\mathcal{N}(\eta(\mathbf{B}_i,\mathbf{S}_i),\sigma^2)$ and set

$$\eta(\mathbf{B}_i,\mathbf{S}_i)=\alpha_{11}b_{i11}+\alpha_{12}b_{i12}+\alpha_{21}b_{i21}+\alpha_{22}b_{i22}+\gamma_{11}s_{i11}+\gamma_{21}s_{i21}+\gamma_{22}s_{i22},$$

where the true values of $\boldsymbol{\alpha},\boldsymbol{\gamma}$ are $\boldsymbol{\alpha}=(-3,-3,-3,3), \boldsymbol{\gamma}=(2,-1,2)$; we did not include other time-invariant covariates $\boldsymbol{W}_i$. These particular truth values were chosen so that the distribution of the outcome Y_i would be similar to the distribution of the SWAN body mass outcomes (our data analysis application). Lastly, we set $\sigma^2=1$. We present the results for $R=200$ replicates in Table 1 for the outcome submodel parameters $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$. See Table S5 in the Supplementary Materials [Chen et al., 2022] for the results for the other model parameters.

4.1.1 Alternative Methods

We briefly introduce three common alternatives in our comparative simulation study: two-stage simple linear model (TSLM), two-stage linear mixed model (TSLMM), and two-stage individual-variance (TSIV) model. We refer to our joint model as the "Joint Model with Individual Variances", or JMIV.

Two-Stage Simple Linear Regression (TSLM) One of the most simple alternative models we could use is the linear regression model in two stages. We used the `lm()` function in R and first fit the following model:

$$X_{itq}=\beta_{iq0}+\beta_{iq1}t+\epsilon_{iq}, q=1,2.$$

Here, we will obtain $\widehat{\beta}_{iq0},\widehat{\beta}_{iq1}$ via ordinary least squares estimates for the mean parameters b_{iq0},b_{iq1}. To estimate s_{i11},s_{i22}, we collected the residuals from each regression, e.g., $r_{ij1}=(x_{it1}-(\widehat{\beta}_{i10}+\widehat{\beta}_{i11}t_{ij})), j=1,\ldots,n_i$, and computed the sample variance of these residuals, which we term "estimated residual variance" for each individual, i.e., $\widehat{s}_{i11}$; similarly, we obtain r_{ij2} and $\widehat{s}_{i22}$. The residual covariance, $\widehat{s}_{i12}$, was estimated by sample covariance of (r_{ij1},r_{ij2}). We then used linear regression to model the outcome based on the estimated coefficients and residual variances and covariances from the first-stage model: $\mathbb{E}(Y_i\mid\text{others})=\sum_{q=1,2}\left\{\alpha_{q1}\widehat{\beta}_{iq0}+\alpha_{q2}\widehat{\beta}_{iq1}+\gamma_{qq}\widehat{s}_{qq}\right\}+\sum_{q'<q}\gamma_{q'q}\widehat{s}_{q'q}$. For each replicate, we saved the point estimates and 95% confidence intervals to compute the bias, coverage, and average interval length.

Two-Stage Linear Mixed Model and Linear Regression (TSLMM) This alternative is a slightly more sophisticated approach than TSLM. In the first stage, we fit a Bayesian bivariate response linear mixed model with the **brms** package [Bürkner, 2017]

$$X_{iq}(t) = \beta_{q0} + b_{iq0} + \beta_{q1}t + b_{iq1}t + \epsilon_{itq}, q = 1, 2.$$

We chose to use a Bayesian framework for this model since fitting linear mixed models with multivariate outcomes is more complicated to implement in a frequentist setting. Standard Bayesian software such as the **brms** package allows for easier implementation of multivariate outcome linear mixed models. We place independent $\mathcal{N}(0, 10^2)$ priors on the intercept and slope parameters. We use the preset prior distribution for the random-effects correlation matrix, an LKJ prior with scale parameter 1, as suggested by Bürkner [2017]. For all other prior specifications, we used the default prior settings in the **brms** package.

We approximated the "B_i" coefficients for each individual trajectory with the "overall" coefficient estimates: $\widehat{B}_{iq0}$ = $\widehat{\beta}_{q0} + \widehat{b}_{iq0}$ and $\widehat{B}_{iq1}$ = $\widehat{\beta}_{q1} + \widehat{b}_{iq1}$, where $\widehat{\beta}_{qp}$ and $\widehat{b}_{iqp}, p = 0, 1$ are the estimated posterior means of the fixed and random effects respectively. As in the previous model, we estimated $\mathbf{S}_i$ by computing the model residuals (e.g. $X_{itq} - (\widehat{B}_{i0q} + \widehat{B}_{i1q}t)$) and then computed the variance across all residuals. We also computed the residual covariance to estimate of $\widehat{s}_{12}$. We then fit the same second-stage model as in the TSLM setup to get the estimated posterior means and corresponding 95% confidence intervals for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

Two-Stage Individual Variances (TSIV) Model Here, we fit the longitudinal outcome model using Equations (5) and (6) only (together with their prior specifications in Equation (9), (11) to (12)) and use the estimates of the posterior means, $\widehat{\mathbf{B}}_i$ and $\widehat{\mathbf{S}}_i$ in the model 7 (together with prior specifications 3.3). Note that we do not consider this to be a practical alternative to our first two models, since if one goes to the effort of using a non-standard multilevel model for subject-specific variance-covariance matrices, one might as well go the extra step of bringing them together within a joint model. However, we do this to investigate the effect of not propagating the statistical uncertainty across the two components of the model.

4.1.2 Simulation I: Results

Table 1 presents the results of Simulation I. For the two-stage linear regression model, we can see that the point estimates of the outcome model parameters are attenuated towards the null. This result makes sense given what we know about bias resulting from measurement error (Section 2.1). Furthermore, the actual coverage rate is quite poor, especially for the regression coefficients of the variances and covariances ($\boldsymbol{\gamma}$).

For the TSLMM approach, the coverage and bias of the α parameters have significantly improved compared to the TSLM approach, likely due to the linear mixed model appropriately capturing the dependence between individuals' data points (i.e., appropriately capturing the measurement error in the mean parameters). However, TSLMM still has difficulty in recovering the coefficients of the variances and covariances, as can be seen by the poor coverage and high bias of these parameters. This makes sense since this framework assumes that individual random effects variability can be drawn from a population level variance-covariance matrix (not capturing measurement error in the variance parameters). This result suggests that if the individual variances and covariances do have an influential role in estimating the outcome, neither TSLMM nor TSLM will be able to recover the true values of these parameters. Interestingly, the TSLMM results also show an attenuation towards the null for the $\boldsymbol{\gamma}$ parameters, but not for the $\boldsymbol{\alpha}$ parameters (although the bias is negligible). This indicates that the TSLMM alternative is able to better estimate the individual intercepts and slopes, but not the residual variability.

Compared to the TSLM and the TSLMM approaches, the TSIV approach has noticeably better coverage and lower bias of the $\boldsymbol{\gamma}$ parameters. However, compared to our proposed JMIV, TSIV is still uniformly 'worse' across the three metrics. The bias of the three $\boldsymbol{\gamma}$ parameters is higher when compared to the bias produced by JMIV. Also, none of the $\boldsymbol{\gamma}$ parameters have higher than 90% coverage across the 200 replicates and the average length of the 95% credible intervals is higher than the 95% credible intervals from the JMIV approach. Across all of the simulation replicates, JMIV achieved greater than 90% coverage of the true parameters. JMIV also achieved low bias across the simulation replicates. We do note that the average 95% CrI interval lengths are larger for the $\mathbf{R}_i$ parameters than for the $\mathbf{D}_i$ parameters (see Supplementary Material S5 [Chen et al., 2022]). This is likely due to the higher uncertainty in estimating these correlation parameters, which has been captured appropriately. This higher uncertainty is also likely the same mechanism behind the larger average 95% CrI interval length of the γ_{12} parameter (corresponding to the covariance of the two trajectories). Overall, these results demonstrate that our model is able to successfully recover the data generating parameters while maintaining good coverage and low bias.

Truth	Model	Bias	Coverage (%)	Average Interval Length
α_{11} = -3	**JMIV**	0.01	98.0	0.29
	TSLM	0.33	2.5	0.30
	TSLMM	-0.01	93.5	0.38
	TSIV	-0.02	97.5	0.35
α_{12} = -3	**JMIV**	0.01	95.0	0.27
	TSLM	0.13	67.5	0.34
	TSLMM	0.00	94.0	0.34
	TSIV	-0.01	93.0	0.29
α_{21} = -3	**JMIV**	0.00	97.0	0.25
	TSLM	0.46	0.0	0.29
	TSLMM	-0.01	92.5	0.40
	TSIV	-0.01	89.0	0.32
α_{22} = 3	**JMIV**	-0.02	93.5	0.37
	TSLM	-0.17	71.5	0.36
	TSLMM	-0.01	97.0	0.49
	TSIV	0.01	95.0	0.43
γ_{11} = 2	**JMIV**	0.01	94.0	0.52
	TSLM	-0.38	17.0	0.36
	TSLMM	-0.38	15.0	0.35
	TSIV	0.03	76.0	0.41
γ_{12} = -1	**JMIV**	0.00	95.5	0.86
	TSLM	0.41	35.0	0.65
	TSLMM	-0.41	31.0	0.62
	TSIV	0.04	88.5	0.86
γ_{22} = 2	**JMIV**	0.00	98.0	0.43
	TSLM	0.62	0.0	0.33
	TSLMM	0.62	0.0	0.32
	TSIV	-0.01	88.5	0.40

Table 1: Simulation I: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates. We compare our Bayesian joint model (JMIV) to the 1) simple two-stage linear regression (TSLM) 2) the two-stage linear mixed model-linear regression (TSLMM) and 3) the two-stage individual variances (TSIV) model. See Section 4.1.1 for details about the alternative methods.

4.2 Simulation 2: Three Biomarkers

Here we again simulate $n_i = 6$ to 15 timepoints each for $N = 1,000$ individuals. The simulated longitudinal data is generated by $X_{itq} = b_{iq1} + b_{iq2}t + \epsilon_{iq}, q = 1, 2$, where $\boldsymbol{B}_{i1} \sim \mathcal{N}_2(\boldsymbol{\beta}_1, \Sigma_1), \boldsymbol{B}_{i2} \sim \mathcal{N}_2(\boldsymbol{\beta}_2, \Sigma_2), \boldsymbol{B}_{i3} \sim \mathcal{N}_2(\boldsymbol{\beta}_3, \Sigma_3)$, and $\boldsymbol{\beta}_1 = (0, 2)^\mathsf{T}$, $\boldsymbol{\beta}_2 = (2, 1)^\mathsf{T}$, $\boldsymbol{\beta}_3 = (1, 1)^\mathsf{T}$,

$$\Sigma_1 = \begin{pmatrix} 1 & -0.05 \\ -0.05 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & -0.1 \\ -0.1 & 0.5 \end{pmatrix}, \Sigma_3 = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 1 \end{pmatrix},$$
$$\log(d_{i1}) \sim \mathcal{N}(0, 0.75)/2, \log(d_{i2}) \sim \mathcal{N}(0.5, 0.5)/2, \log(d_{i3}) \sim \mathcal{N}(0, 1)/2.$$

We first generate the following values $(c_{12} + 1)/2 \sim \mathsf{Beta}(1, 5)$, $(c_{13} + 1)/2 \sim \mathsf{Beta}(1, 5)$, $(c_{23} + 1)/2 \sim \mathsf{Beta}(2, 2)$ and use the approach for $Q = 3$ markers described in Section 3.3 to generate the individual correlation matrices, $\mathbf{R}_i$.

For the outcome submodel, we set the true values of the regression coefficients as: $\boldsymbol{\alpha} = (-3, -3, 3, -3, 3, 3), \boldsymbol{\gamma} = (2, -1, 2, -2, 2, 1)$, where these values were again chosen so that the distribution of the outcome y_i would be similar to the distribution of the SWAN body mass outcomes. Lastly, we set $\sigma^2 = 1$ (the variance parameter for the outcome).

We present the results of this simulation study in Tables 2 and 3. We note that the proposed JMIV achieves above 90% coverage for both the mean ($\boldsymbol{\alpha}$) and variance-covariance ($\boldsymbol{\gamma}$) parameters, which the other models fail to do. With respect to the $\boldsymbol{\alpha}$ parameters, TSLMM and TSIV both perform well in terms of both coverage and bias. However, substantial differences in performance are present in the $\boldsymbol{\gamma}$ parameters. We see that that our model, JMIV, consistently has lower bias except in the case of γ_{23} where TSIV achieves lower bias and γ_{33} where TSLMM achieves lower bias. However, in both cases, JMIV outperforms the other models in terms of higher coverage (substantially higher coverage in the case of γ_{33}), indicating that JMIV is still a better model choice.

Truth	Model	Bias	Coverage (%)	Average Interval Length
α_{11} = -3	**JMIV**	0.01	96.5	0.49
	TSLM	0.61	2.0	0.56
	TSLMM	0.00	93.5	0.69
	TSIV	-0.01	96.0	0.64
α_{12} = -3	**JMIV**	-0.01	93.0	0.45
	TSLM	0.18	77.5	0.62
	TSLMM	-0.01	94.0	0.62
	TSIV	0.00	94.5	0.55
α_{13} = 3	**JMIV**	-0.01	96.5	0.50
	TSLM	-0.98	0.0	0.55
	TSLMM	0.03	93.0	0.74
	TSIV	-0.01	92.0	0.59
α_{21} = -3	**JMIV**	-0.01	95.0	0.43
	TSLM	0.28	47.0	0.53
	TSLMM	0.003	95.0	0.73
	TSIV	-0.01	96.0	0.81
α_{22} = 3	**JMIV**	-0.01	92.5	0.63
	TSLM	-0.17	71.5	0.36
	TSLMM	0.02	93.5	0.90
	TSIV	0.01	93.5	0.65
α_{23} = 3	**JMIV**	-0.01	97.5	0.47
	TSLM	-0.44	20.5	0.63
	TSLMM	-0.02	91.5	0.65
	TSIV	0.01	96.0	0.59

Table 2: Simulation II: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates for the $\boldsymbol{\alpha}$ parameters. We compare our Bayesian joint model (JMIV) to the 1) simple two stage linear regression (TSLM) 2) the two stage linear mixed model-linear regression (TSLMM) and 3) the Bayesian two stage model (TSIV).

Truth Parameter	Model	Bias	Coverage (%)	Average Interval Length
γ_{11} = 2	**JMIV**	-0.01	93.5	1.08
	TSLM	-0.43	47.5	0.82
	TSLMM	-0.44	43.0	0.80
	TSIV	-0.02	83.0	0.97
γ_{12} = -1	**JMIV**	-0.04	92.0	1.66
	TSLM	1.22	8.5	1.32
	TSLMM	1.22	8.0	1.28
	TSIV	-0.08	91.5	1.77
γ_{22} = 2	**JMIV**	0.03	96.5	0.79
	TSLM	-1.23	0.0	0.61
	TSLMM	-1.24	0.0	0.59
	TSIV	-0.03	87.5	0.76
γ_{13} = -2	**JMIV**	0.03	94.5	1.91
	TSLM	1.63	2.5	1.33
	TSLMM	1.62	2.5	1.29
	TSIV	-0.06	82.0	1.78
γ_{23} = 2	**JMIV**	0.04	94.5	1.50
	TSLM	-0.90	17.5	1.07
	TSLMM	-0.92	14.5	1.03
	TSIV	-0.01	90.5	1.52
γ_{33} = 1	**JMIV**	0.01	94.0	0.54
	TSLM	0.01	68.5	0.38
	TSLMM	-0.002	69.0	0.37
	TSIV	-0.02	64.4	0.42

Table 3: Simulation II: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates for the $\boldsymbol{\gamma}$ parameters. We compare our Bayesian joint model (JMIV) to the 1) simple two stage linear regression (TSLM) 2) the two stage linear mixed model-linear regression (TSLMM) and 3) the Bayesian two stage model (TSIV).

5 Hormone Trajectories and Changes in Body Mass Across the Menopausal Transition

5.1 Study of Women's Health Across the Nation

The Study of Women's Health Across the Nation (SWAN) is a longitudinal cohort study that followed women over the menopausal transition [Sowers et al., 2000]. Seven clinical sites across the United States participated in this study. To be eligible for the cohort, women had to be between 42-52 years old, had to have had at least one menstrual period and not used reproductive hormones (e.g. hormonal contraceptives or other exogenous hormones) in the past three months, had to reside in the geographic area of the clinical site, and had to self-identify as White, Black, Chinese, Japanese or Hispanic. Serum E2 and FSH biomarker measurements were collected at baseline and during 13 approximately annual follow-up visits, along with other health measurements. Body composition was measured via dual-energy X-ray absorptiometry (DXA) at five of the seven clinical site visits. Women also completed questionnaires regarding lifestyle and socio-demographic characteristics. The comprehensive and longitudinal aspect of this dataset makes it an ideal application for understanding how individual hormone trajectories affect the rate of change of fat mass and lean mass across the menopausal transition.

We initially started with women who were enrolled at one of the five sites with body composition measures and who were not on hormone therapy during the clinical visits where E2 and FSH measurements were taken. After restricting our analysis to women who had an observed FMP, we had 887 individuals in the dataset. Although there are five racial/ethnic groups in the SWAN study, the site with Hispanic women did not have body composition data; hence, Hispanic women are not included in this analysis. After computing the body mass composition windows (see below), an additional 47 women were deleted from the analysis. The final analytical dataset was completed on 841 individuals, with a total number of 9,902 hormone measurements.

For the individual trajectory model, we use the log values of FSH (mIU/mL) and E2 (pg/mL) measured at each visit. The log scale minimizes the skewness of the data and allows us to better determine the overall population trend. We removed this population trend by fitting a lowess curve to each (log) hormone. This was done by using time to FMP at each visit as the numeric predictor for the corresponding hormone measurement and using weighted least squares to obtain the predicted fit at each timepoint. We then subtracted the individual measurements from the lowess estimates. By removing the common population trends in the data, our model can better approximate the individual trajectories and individual level variances using a simpler (lower dimensional) subject-level trajectory model. Figure 1 summarizes the subject-level longitudinal data model fit results for two randomly-selected women. These individual plots in Figure 1 indicate that our model generates reasonable estimates of the individual hormone values.

For the outcomes of interest, we selected fat mass rate of change and lean mass rate of change over a selected window. We define this window to be from the visit closest to 5 years before the FMP to the visit closest to 5 years after the FMP, with the requirement that the closest visit be at least 3 or more years before/after the FMP. By doing this, we aimed to capture the most accurate trend in body composition change that was not fully dependent on a measurement right before menopause. The fat mass (or lean mass) rate of change is the difference between the 'first visit' (pre FMP) and the 'last visit' (post FMP) divided by the amount of time (in years) within each individual window. We normalized the body mass measurements by using the proportion of fat mass (or lean mass) to body weight (grams) rather than using the unadjusted body mass measurements (both measured in grams). Figure 2 displays the histograms of these outcomes, after performing the normalization and rate adjustments.

In the outcome models, we used the correlation between E2 and FSH rather than the covariance as a variable of interest, since the correlation measure has a more straightforward interpretation and is normalized to the E2, FSH variances. We also explored interaction terms in the outcome regression, in order to better understand the relationships between the two hormones, and how these may affect the body mass outcomes. We interacted the individual E2 and FSH intercepts, individual E2 and FSH slopes, and individual E2 and FSH variances parameters separately for both the fat mass and lean mass outcomes (for a total of six different models). For both the fat mass and lean mass models, significant interactions between the E2 and FSH variances were observed.

Finally, by including additional covariates $\boldsymbol{W}_i$ in the outcome regression model, we adjusted for: fat mass (lean mass) body weight proportion at the 'first' visit, race/ethnicity (White, Black, Chinese and Japanese) and sports activity category. We included race/ethnicity in the models given previous research using SWAN data that found differences in body mass composition changes among ethnic groups during the menopausal transition [Greendale et al., 2019, 2021]. The physical activity category is a measure of the individual level physical activity trajectories for each subject in the SWAN study, grouped into categories reflecting: (1) lowest, (2) increasing, (3) decreasing, (4) middle, and (5) highest physical activity during follow up. For a more detailed description, please refer to Pettee Gabriel et al. [2017]. Table 4 displays the descriptive statistics for the individuals in our analysis, including demographic and physical activity information. The following sections contain the results from applying our joint model to the SWAN dataset.

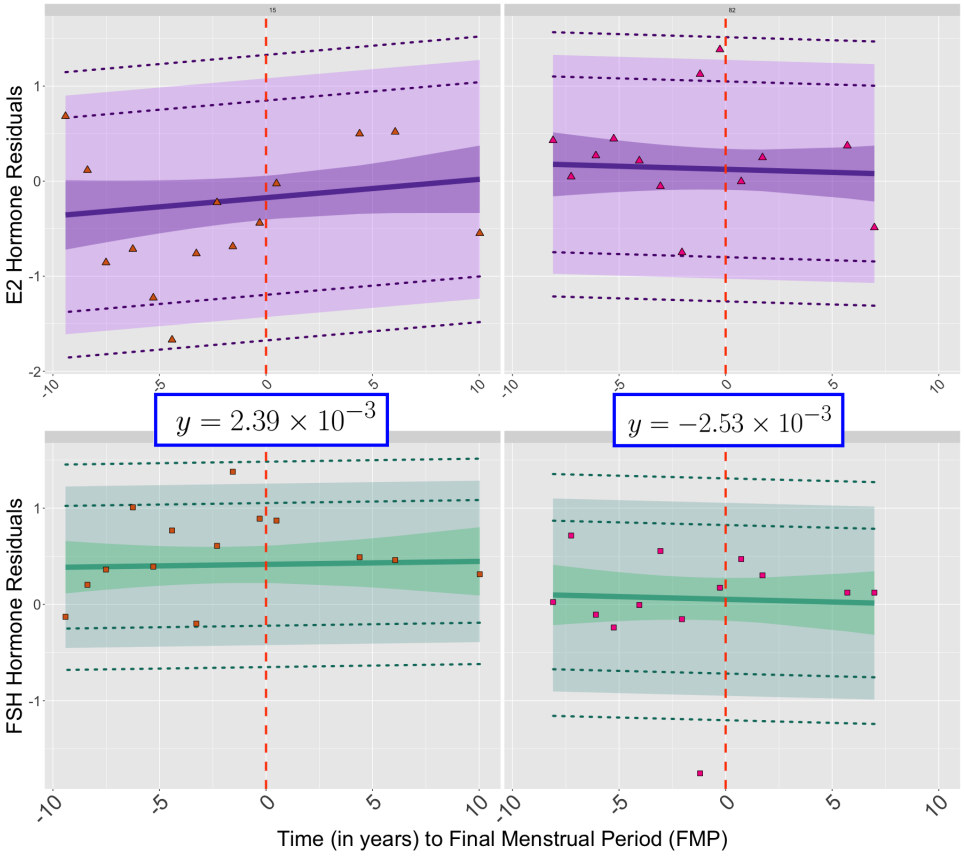

Figure 1: Plots of estimated hormone trajectories for two individuals from the fat mass rate of change model. The solid lines are the estimated individual mean trajectories, based on the posterior means of $\boldsymbol{B_i}$, i.e. $\widehat{b}_{i0} + \widehat{b}_{i1}t$. The darker inner intervals around the solid lines are $+/-1.64 \times var(\widehat{b}_{i0} + \widehat{b}_{i1}t)$ and the lighter band is $+/-1.64 \times \widehat{\sigma}_{iq}$, where $\widehat{\sigma}_{iq}$ is the square root of the estimated posterior mean of the individual level variance of hormone q. The dotted lines represent $+/-1.64 \times \sigma_{iq5}$ and $+/-1.64 \times \sigma_{iq95}$ where $\sigma_{iq5}, \sigma_{iq95}$ are the values of the 5th and 95th percentiles of the posterior samples of the individual variances for each hormone q. The triangles and squares are the observed E2 and FSH residuals, respectively. The observed individual fat mass rates of change are shown in bordered boxes.

Variable	**Statistic**	**Value**	**n**
Longitudinal Predictors	Mean/SD		
E2 Residuals		-0.04 (0.81)	9,902
FSH Residuals		0.02 (0.61)	9,902
Body Mass Outcome	Mean/SD		
Fat Mass Rate of Change		0.001 (0.004)	841
Lean Mass Rate of Change		-0.001 (0.004)	841
Baseline Body Mass	Mean/SD		
Fat Mass Prop. at Visit 1		0.36 (0.07)	841
Lean Mass Prop. at Visit 1		0.55 (0.06)	841
Race/Ethnicity	Percent		
White (Reference)		47.2%	397
Black		24.9%	209
Japanese		15.3%	129
Chinese		12.6%	106
Physical Activity	Percent		
Lowest Activity (Reference)		23.6%	199
Increasing Activity		12.7%	107
Decreasing Activity		22.7%	191
Middle Activity		25.6%	215
Highest Activity		15.3%	129

Table 4: Descriptive statistics of the SWAN dataset based on 840 individuals.

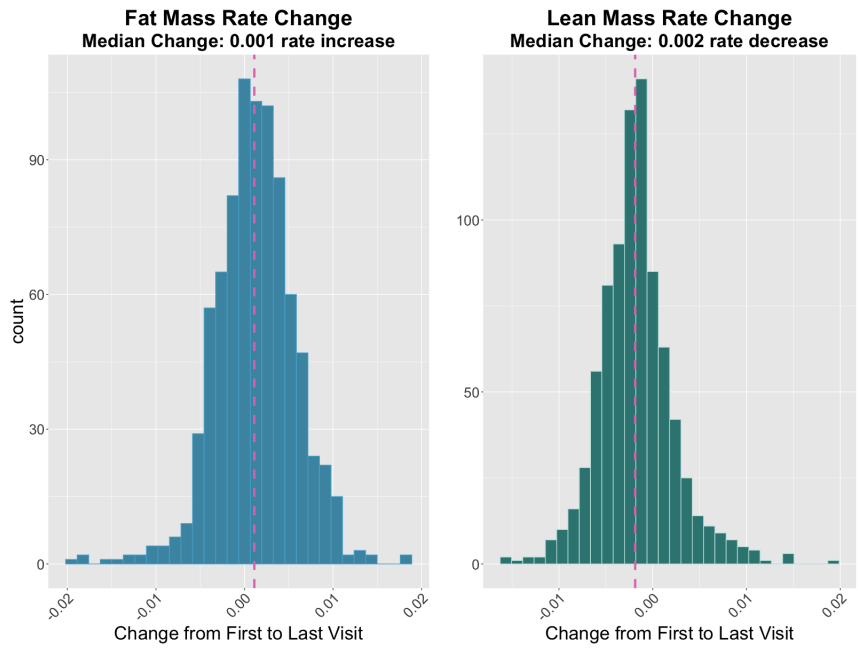

Figure 2: Histograms of the observed rate of change in fat mass composition (left) and the observed rate of change in lean mass composition (right).

Coefficient	Variable	Post. Mean	2.5% CrI	97.5% CrI
α_{11}	**E2 Intercept**	**5.02**	**2.56**	**7.77**
α_{21}	**FSH Intercept**	**0.94**	**-0.20**	**2.09**
α_{12}	**E2 Slope**	**33.70**	**4.28**	**68.54**
α_{22}	**FSH Slope**	**-23.82**	**-42.83**	**-4.82**
γ_{11}	**E2 Var.**	**7.03**	**2.63**	**10.92**
γ_{12}	E2, FSH Cor.	2.44	-1.51	6.11
γ_{22}	**FSH Var.**	**16.38**	**5.43**	**26.42**
γ_3	**E2 Var. x FSH Var.**	**-33.78**	**-51.62**	**-14.47**
γ_1^W	**Fat Mass Prop. (First Visit)**	**-2.05**	**-2.38**	**-1.74**
γ_2^W	**Black**	**-0.73**	**-1.42**	**0.00**
γ_3^W	Chinese	0.11	-0.80	0.98
γ_4^W	**Japanese**	**-3.23**	**-4.07**	**-2.41**
γ_5^W	Increasing Activity (Cat. 2)	0.30	-0.64	1.26
γ_6^W	Decreasing Activity (Cat. 3)	0.66	-0.12	1.45
γ_7^W	Middle Activity (Cat. 4)	0.08	-0.69	0.84
γ_8^W	Highest Activity (Cat. 5)	-0.76	-1.66	0.18

Table 5: Estimated posterior means and 95% credible intervals for the fat mass rate of change model with variance interactions. All estimated posterior means and 95% CrI values have been multiplied by 10^3.

5.2 Fat Mass Rate of Change

Table 5 displays the results of the fat mass model with variance interactions. We found that the E2, FSH intercepts and the E2, FSH slopes were all signfiicantly associated with fat mass rate of change. For a one-unit increase in starting E2 (from the population average), this was associated with an average 5.02×10^{-3} increase in fat mass per year.

Since we standardized the hormones to be relative to the population average, the effects of the E2 and FSH variances can be interpreted as the effect of one hormone's variability when the other hormone variability is at the population average. So, when FSH variability was the same as the population variance, then a one unit increase in E2 variability was associated with a 7.03×10^{-3} increase in fat mass per year. Similarly, when E2 variability was the same as the population variance, then a one unit increase in FSH variability was associated with a 16.38×10^{-3} inrease in fat mass per year.

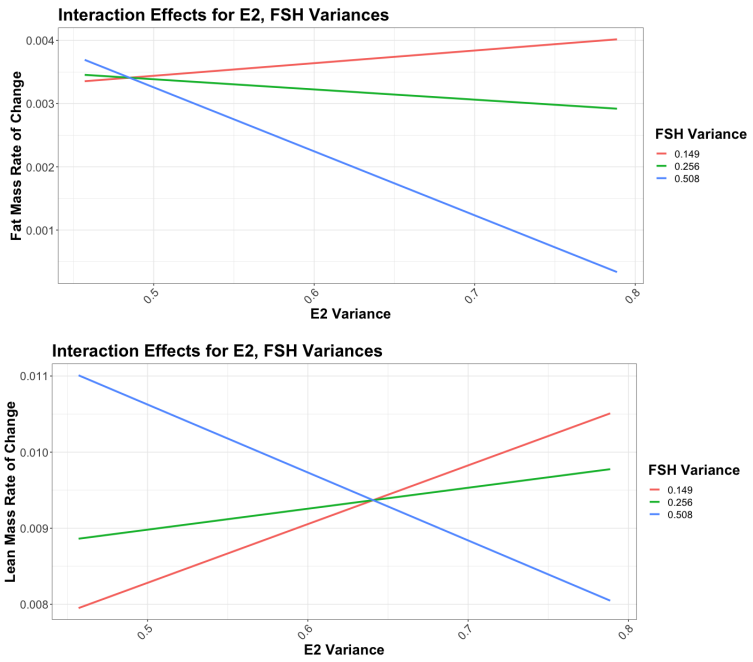

Figure 3: Visualizations of the effects of the E2, FSH variance terms for the fat mass model (top) and lean mass model(bottoms).

The top plot in Figure 3 shows a visualization of how the interaction between E2, FSH variances is related to fat mass change. We can see that when E2 variability was high and FSH variability was low, there was an average higher fat mass rate of change. However, when both E2 and FSH variability were high, their individual effects on fat mass change were moderated by the interaction term, resulting in lower fat mass rates of change. This finding suggested that women who had highly variable E2 and FSH also had, on average, slower fat mass increases than women who had only one highly variable hormone. E2 and FSH have a tightly coupled relationship since the two hormones balance each other; thus, if one hormone is highly variable when the other one is not, this could indicate higher deterioration or malfunction in the reproductive system's functioning beyond the expected decline during the menopausal transition. If this is the case, then it would make sense that a greater decline in reproductive functioning may be associated with increased fat mass gain, since E2 is known to regulate adipose tissue.

In this fully adjusted model, fat mass as a proportion of overall body mass at the first visit was also negatively associated with the outcome (an average -2.05×10^{-3} fat mass decline per year). Black women and Japanese women, compared to white women, had slower fat mass gain (-0.73×10^{-3} per year and -3.23×10^{-3} per year, respectively).

5.3 Lean Mass Rate of Change

Initially the Gaussian outcome assumption did not appear to be a good fit for the observed outcome. In particular, the residuals suggested overdispersed variances with a common mean, so we allowed the outcome to be modeled as a mixture of two Gaussian distributions with equal means but different variances:

$$
\begin{aligned}
&Y_i | z_i, \eta(\mathbf{B}_i, \mathbf{S}_i, \boldsymbol{W}_i), \sigma_1^2, \sigma_2^2 \sim \mathcal{N}(\eta_i, \sigma_{z_i}^2), \\
&z_i | \pi \sim \mathsf{Bernoulli}(\pi), \\
&\pi \sim \mathsf{Beta}(1/2, 1/2), \sigma_1 \sim \mathsf{half-Cauchy}(0, 2.5), \sigma_2 \sim \mathsf{half-Cauchy}(0, 5)
\end{aligned}
$$

where z_i is an unobserved indicator variable indicating membership in the first mixture component; $z_i = 1$ for the first component. Because the mean is equal across the mixture components, the interpretation of the regression parameters will be the same as for the fat mass models despite the additional variance parameter.

Table 6 displays the estimated coefficients for the lean mass model with variance interactions. A one-unit higher E2 intercept (above the population mean) was associated with an average -2.54×10^3 decline in lean mass per year. A higher individual FSH intercept was also negatively associated with lean mass; a one-unit higher starting FSH was associated with an average -1.49×10^3 decline in lean mass per year.

Coefficient	Variable	Post. Mean	2.5% CrI	97.5% CrI
α_{11}	**E2 Intercept**	**-2.54**	**-4.74**	**-0.40**
α_{21}	**FSH Intercept**	**-1.49**	**-2.47**	**-0.59**
α_{12}	E2 Slope	-18.46	-46.02	6.65
α_{22}	FSH Slope	16.27	-1.02	33.40
γ_{11}	**E2 Var.**	**14.62**	**9.56**	**20.27**
γ_{12}	E2, FSH Cor.	-3.53	-7.67	0.14
γ_{22}	**FSH Var.**	**29.70**	**18.17**	**41.88**
γ_3	**E2 Var. x FSH Var.**	**-46.35**	**-67.53**	**-26.88**
γ_1^W	**Lean Mass Prop. (First Visit)**	**-24.92**	**-29.69**	**-19.89**
γ_2^W	**Black**	**0.86**	**0.26**	**1.48**
γ_3^W	**Chinese**	**1.59**	**0.81**	**2.36**
γ_4^W	Japanese	0.51	-0.19	1.22
γ_5^W	Increasing Activity (Cat. 2)	-0.04	-0.83	0.77
γ_6^W	Decreasing Activity (Cat. 3)	-0.09	-0.75	0.62
γ_7^W	Middle Activity (Cat. 4)	0.16	-0.51	0.85
γ_8^W	**Highest Activity (Cat. 5)**	**1.02**	**0.23**	**1.81**

Table 6: Estimated posterior means and 95% credible intervals for the lean mass rate of change model with variance interactions. All estimated posterior means and 95% CrI values have been multiplied by 10^3.

We found that when the individual E2 variance was at the population mean level, a one unit increase in FSH variance was associated with a 29.70×10^3 increase in lean mass per year. The interaction term of E2, FSH variances was also negative, which indicates that the positive relationship between higher E2 variance and lean mass rate of change was weaker when FSH variability is also high (and vice versa for the relationship between FSH variance and lean mass change). This also corresponds with the E2, FSH correlation term, which is negatively associated with the outcome (although in this model, the 95% CrI contains 0 so the correlation is not significantly associated with the outcome).

The bottom plot of Figure 3 displays how the variance interaction affects lean mass change for different values of E2 variance and FSH variance. When E2 variability was high and FSH variability was low, there was an average higher lean mass rate of change. This was also true when FSH variability was high and E2 variability was low. However, when both E2 and FSH variabilities were high, their effects on lean mass change was moderated by the interaction term, resulting in average lower lean mass changes. Women who had highly variable E2 and FSH also had, on average, slower lean mass increases than women who had one, but not both, highly variable hormone. This result mirrors the result from the fat mass rate of change model, where we also saw that the variance interaction term moderated the effects from the E2 and FSH variance terms.

Unsurprisingly, lean mass proportion at first visit was negatively associated with lean mass change (-24.92×10^3 decline per year). Black and Chinese women also had faster lean mass gains compared to white women on average (0.86×10^3 increase per year and 1.59×10^3 increase per year, respectively). Finally, women in the highest activity group also had higher lean mass increases than women in the lowest activity group (1.02×10^3 increase per year).

6 Discussion

We have presented a joint modeling approach for estimating individual-level mean and variance-covariance matrices based on longitudinal marker trajectories, which are then linked to a cross-sectional outcome. Simulations show that our model outperforms alternative approaches to this research problem. Analysis of hormone trajectories data reveal that the variability of E2 does have a statistically significant association with the rate of change of body mass composition across the menopausal transition.

Our work is important for both methodological development of joint models and for women's health research. Our model estimates both mean longitudinal trends and the residual variability of these individual trajectories, and propagates the estimation uncertainty into the second submodel. This joint modelling is important for obtaining accurate estimates (in terms of low bias, higher coverage and lower variability) of how the individual-level parameters are linked to the outcome. Simulation results demonstrate that our model outperforms common two-stage approaches. Substantively, understanding how the individual-level variability of two hormones (E2 and FSH) is associated with changes in body mass composition had not been well explored. Our analysis of the SWAN data revealed that higher variability in E2 and FSH are linked to both faster rates of fat mass and lean mass change. Our analysis of the SWAN data revealed that higher variability in E2 and FSH are linked to both faster rates of fat mass and lean mass change. Additionally, women with highly variable E2 and FSH tended to have slower fat mass increases and slower lean mass increases than

women who had only one highly variable hormone. These findings can serve as a basis for further investigation of how hormone variability affects such health outcomes.

As mentioned above, joint estimation of longitudinal variables and scalar outcomes can be useful for understanding of scientific questions in many areas. With longitudinal biomarker data becoming more readily available (e.g. wearable devices), we need statistical methods for analyzing these types of data. Our proposed method addresses the gap in methods by 1) providing a framework for jointly modeling longitudinal and cross-sectional data and 2) explicitly modeling individual-level variability in the longitudinal trajectories, which can improve understanding of the relationship between longitudinal predictors and health outcomes.

6.1 Remark

In our simulation studies and scientific application, we made the simplifying assumption to exclude covariates $\boldsymbol{W}_i$ in modeling the longitudinal markers (Equation 5). In mathematical terms, this means that the likelihood functions of $\boldsymbol{X}_{it}, \boldsymbol{B}_i, \boldsymbol{S}_i, \boldsymbol{W}_i$ are:

$$f(\boldsymbol{X}_{it} \mid \boldsymbol{B}_i, \boldsymbol{S}_i, \boldsymbol{W}_i) = f(\boldsymbol{X}_{it} \mid \boldsymbol{B}_i, \boldsymbol{S}_i)$$
$$f(\boldsymbol{B}_i, \boldsymbol{S}_i \mid \boldsymbol{W}_i) = f(\boldsymbol{B}_i, \boldsymbol{S}_i)$$

For the scientific application, our main focus was to evaluate the overall marginal effects of the biomarker hormone means and variances on the body mass outcomes of interest. In particular, the effects of the variance and correlation parameters were of key interest, since the associations between individual E2 and FSH variabilities (and co-variability) and body mass changes had not been previously explored. The estimated coefficients for the $\boldsymbol{B}_i, \boldsymbol{S}_i$ described in Section 5 should be interpreted as marginal effects, rather than conditional on the other adjusted covariates. For this particular scientific application, we believed that this assumption resulted in a more straightforward interpretation of the mean, variance, and correlation parameters. The decision to include or exclude $\boldsymbol{W}_i$ in the longitudinal submodel should be made with the specific research application in mind and whether or not the simplifying assumption makes sense for the particular context.

6.2 Future Work

One extension of this work could be to model the individual variances as being functions of time, i.e. $\mathbf{S}_{it}$. E2 and FSH are known to be highly variable as women approach their final menstrual cycle, so estimating $\mathbf{S}_{it}$ may better capture such changes in the biomarker variances. To obtain these estimates, we would likely need a larger dataset (with both more individuals and timepoints) than is currently available with the SWAN study. Another methodological extension could be to extend this model to account for missingness in both the trajectory data and the outcome data. For this analysis, we removed the missing values in the hormone data and only analyzed individuals with observed body mass outcomes. Although less than 5% of the values in our dataset were missing, analyzing complete case data only could still result in biased inference. In the SWAN dataset, individuals can be subject to intermittent missingness as well as dropout; these types of missing data patterns could be addressed in future work. Finally, an alternative approach could extend the work of [Jiang et al., 2015] in using latent class models to capture the impact of the means and variances of multiple longitudinal trajectories on health outcomes. Finally, we note that increasing the number of longitudinal trajectories may result in a form of $\eta(\cdot)$ in the OR model that is complicated to estimate, since the number of variance-covariance parameters increases quadratically with the number of trajectories. Some type of dimension reduction procedure may be useful in these settings, although retaining interpretability may be challenging.

References

John F. Randolph, Jr., MaryFran Sowers, Irina V. Bondarenko, Siobán D. Harlow, Judith L. Luborsky, and Roderick J. Little. Change in estradiol and follicle-stimulating hormone across the early menopausal transition: Effects of ethnicity and age. *The Journal of Clinical Endocrinology & Metabolism*, 89:1555–1561, 2004. ISSN 0021-972X. doi:10.1210/jc.2003-031183. URL `https://doi.org/10.1210/jc.2003-031183`.

Carrie Karvonen-Gutierrez and Sioban D. Harlow. Menopause and midlife health changes. In Jeffrey B. Halter, Joseph G. Ouslander, Stephanie Studenski, Kevin P. High, Sanjay Asthana, Mark A. Supiano, and Christine Ritchie, editors, *Hazzard's Geriatric Medicine and Gerontology*. McGraw-Hill Education, 7 edition, 2017. URL `accessmedicine.mhmedical.com/content.aspx?aid=1136589663`.

Georgia Colleluori, Rui Chen, Nicola Napoli, Lina E. Aguirre, Clifford Qualls, Dennis T. Villareal, and Reina Armamento-Villareal. Fat mass follows a u-shaped distribution based on estradiol levels in postmenopausal women. *Frontiers in Endocrinology*, 9:315, 2018. ISSN 1664-2392. doi:10.3389/fendo.2018.00315. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6036116/.

John F. Randolph, Huiyong Zheng, MaryFran R. Sowers, Carolyn Crandall, Sybil Crawford, Ellen B. Gold, and Marike Vuga. Change in follicle-stimulating hormone and estradiol across the menopausal transition: Effect of age at the final menstrual period. *The Journal of Clinical Endocrinology and Metabolism*, 96:746–754, 2011. ISSN 0021-972X. doi:10.1210/jc.2010-1746. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3047231/.

Mone Zaidi, Daria Lizneva, Se-Min Kim, Li Sun, Jameel Iqbal, Maria I New, Clifford J Rosen, and Tony Yuen. FSH, bone mass, body fat, and biological aging. *Endocrinology*, 159:3503–3514, 2018. ISSN 0013-7227. doi:10.1210/en.2018-00601. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6134257/.

Carlos Henrique Sponton and Shingo Kajimura. Burning fat and building bone by FSH blockade. *Cell Metabolism*, 26: 285–287, 2017. ISSN 1932-7420. doi:10.1016/j.cmet.2017.07.018.

Wendy M. Kohrt and Margaret E. Wierman. Preventing fat gain by blocking follicle-stimulating hormone. *The New England Journal of Medicine*, 377:293–295, 2017. ISSN 1533-4406. doi:10.1056/NEJMcibr1704542.

Margaret L. Gourlay, Bonny L. Specker, Chenxi Li, Catherine A. Hammett-Stabler, Jordan B. Renner, and Janet E. Rubin. Follicle-stimulating hormone is independently associated with lean mass but not BMD in younger postmenopausal women. *Bone*, 50:311–316, 2012. ISSN 8756-3282. doi:10.1016/j.bone.2011.11.001. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3246561/.

Peng Liu, Yaoting Ji, Tony Yuen, Elizabeth Rendina-Ruedy, Victoria E. DeMambro, Samarth Dhawan, Wahid Abu-Amer, Sudeh Izadmehr, Bin Zhou, Andrew C. Shin, Rauf Latif, Priyanthan Thangeswaran, Animesh Gupta, Jianhua Li, Valeria Shnayder, Samuel T. Robinson, Yue Eric Yu, Xingjian Zhang, Feiran Yang, Ping Lu, Yu Zhou, Ling-Ling Zhu, Douglas J. Oberlin, Terry F. Davies, Michaela R. Reagan, Aaron Brown, T. Rajendra Kumar, Solomon Epstein, Jameel Iqbal, Narayan G. Avadhani, Maria I. New, Henrik Molina, Jan B. van Klinken, Edward X. Guo, Christoph Buettner, Shozeb Haider, Zhuan Bian, Li Sun, Clifford J. Rosen, and Mone Zaidi. Blocking FSH induces thermogenic adipose tissue and reduces body fat. *Nature*, 546:107–112, 2017. ISSN 1476-4687. doi:10.1038/nature22342.

R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1:465–480, 2000. ISSN 1465-4644. doi:10.1093/biostatistics/1.4.465.

Cécile Proust-Lima, Mbéry Séne, Jeremy M. G. Taylor, and Hélène Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research*, 23:74–90, 2014. ISSN 1477-0334. doi:10.1177/0962280212445839.

Jue Wang, Sheng Luo, and Liang Li. Dynamic prediction for multiple repeated measures and event time data: An application to parkinson's disease. *The Annals of Applied Statistics*, 11:1787–1809, 2017. ISSN 1932-6157, 1941-7330. doi:10.1214/17-AOAS1059. URL https://projecteuclid.org/journals/annals-of-applied-statistics/volume-11/issue-3/Dynamic-prediction-for-multiple-repeated-measures-and-event-time-data/10.1214/17-AOAS1059.full.

Raymond Carroll. Variances are not always nuisance parameters. *Biometrics*, 59:211–220, 2003. doi:10.1111/1541-0420.t01-1-00027.

Hayley A. Young and David Benton. Heart-rate variability: a biomarker to study the influence of nutrition on physiological and psychological health? *Behavioural Pharmacology*, 29(2):140–151, 2018. ISSN 0955-8810. doi:10.1097/FBP.0000000000000383. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5882295/.

Elizabeth L Ogburn, Kara E Rudolph, Rachel Morello-Frosch, Amber Khan, and Joan A Casey. A warning about using predicted values from regression models for epidemiologic inquiry. *American Journal of Epidemiology*, 190: 1142–1147, 2021. doi:10.1093/aje/kwaa282.

Siruo Wang, Tyler H McCormick, and Jeffrey T Leek. Methods for correcting inference based on outcomes predicted by machine learning. *Proceedings of the National Academy of Sciences*, 117:30266–30275, 2020.

Yueh-Yun Chi and Joseph G. Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62:432–445, 2006. ISSN 0006-341X. doi:10.1111/j.1541-0420.2005.00448.x.

Joseph G. Ibrahim, Haitao Chu, and Liddy M. Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28:2796–2801, 2010. ISSN 0732-183X. doi:10.1200/JCO.2009.25.0654. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4503792/.

A. Lawrence Gould, Mark Ernest Boye, Michael J. Crowther, Joseph G. Ibrahim, George Quartey, Sandrine Micallef, and Frederic Y. Bois. Joint modeling of survival and longitudinal non-survival data: current methods and issues; Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine*, 34:2181–2195, 2015. ISSN 1097-0258. doi:10.1002/sim.6141.

Jeffrey D. Long and James A. Mills. Joint modeling of multivariate longitudinal data and survival data in several observational studies of huntington's disease. *BMC Medical Research Methodology*, 18:138, 2018. ISSN 1471-2288. doi:10.1186/s12874-018-0592-9. URL `https://doi.org/10.1186/s12874-018-0592-9`.

Grigorios Papageorgiou, Katya Mauff, Anirudh Tomer, and Dimitris Rizopoulos. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Review of Statistics and Its Application*, 6:223–240, 2019. doi:10.1146/annurev-statistics-030718-105048. URL `https://doi.org/10.1146/annurev-statistics-030718-105048`.

Michael R. Elliott, Mary D. Sammel, and Jessica Faul. Associations between variability of risk factors and health outcomes in longitudinal studies. *Statistics in Medicine*, 31:2745–2756, 2012. ISSN 0277-6715. doi:10.1002/sim.5370. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3470883/`.

S. D. Harlow, X. Lin, and M. J. Ho. 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, 2000. ISSN 0895-4356. doi:10.1016/s0895-4356(99)00202-4.

Xiaobi Huang, Michael R. Elliott, and Siobán D. Harlow. Modeling menstrual cycle length and variability at the approach of menopause using hierarchical change point models. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 63:445–466, 2014. ISSN 0035-9254. doi:10.1111/rssc.12044.

Mary Sammel, Y Wang, S Ratcliie, Ellen Freeman, and K Propert. Models for within-subject heterogeneity as predictors for disease. In *Proceedings of the Annual Meeting of the American Statistical Association*, 2001.

Bei Jiang, Michael R. Elliott, Mary D. Sammel, and Naisyin Wang. Joint modeling of cross-sectional health outcomes and longitudinal predictors via mixtures of means and variances. *Biometrics*, 71:487–497, 2015. ISSN 1541-0420. doi:https://doi.org/10.1111/biom.12284. URL `http://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12284`.

Raymond J. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. CRC Press LLC, 2006. ISBN 978-1-4200-1013-8. URL `http://ebookcentral.proquest.com/lib/umichigan/detail.action?docID=274076`.

Zvi Griliches and Michael D. Intriligator. *Handbook of Econometrics*, chapter 25, pages 1465–1514. North Holland, 1987.

Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1:515–534, 2006. ISSN 1936-0975, 1931-6690. doi:10.1214/06-BA117A. URL `http://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-3/Prior-distributions-for-variance-parameters-in-hierarchical-models-comment-on/10.1214/06-BA117A.full`.

Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100:1989–2001, 2009. doi:https://doi.org/10.1016/j.jmva.2009.04.008.

Riddhi Pratim Ghosh, Bani Mallick, and Mohsen Pourahmadi. Bayesian estimation of correlation matrices of longitudinal data. *Bayesian Analysis*, 16:1039–1058, 2021. ISSN 1936-0975, 1931-6690. doi:10.1214/20-BA1237. URL `https://projecteuclid.org/journals/bayesian-analysis/volume-16/issue-3/Bayesian-Estimation-of-Correlation-Matrices-of-Longitudinal-Data/10.1214/20-BA1237.full`.

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76:1–32, 2017. doi:10.18637/jss.v076.i01. URL `https://www.jstatsoft.org/index.php/jss/article/view/v076i01`.

Irena Chen, Zhenke Wu, Siobán D. Harlow, Carrie A. Karvonen-Gutierrez, Michelle M. Hood, and Michael R. Elliott. Supplement to variance as a predictor of health outcomes: using subject-level trajectories and variability of sex hormones to predict body fat changes in peri- and post-menopausal women. 2022.

Stan Development Team. RStan: the R interface to Stan, 2020. URL `http://mc-stan.org/`. R package version 2.21.2.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved r^ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16:667–718, 2021. ISSN 1936-0975, 1931-6690. doi:10.1214/20-BA1221. URL https://projecteuclid.org/journals/bayesian-analysis/volume-16/issue-2/Rank-Normalization-Folding-and-Localization--An-Improved-R%cb%86-for/10.1214/20-BA1221.full. Publisher: International Society for Bayesian Analysis.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. CRC Press, 2013. ISBN 978-1-4398-9820-8.

Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80:1–28, 2017. doi:10.18637/jss.v080.i01.

MARYFRAN Sowers, SYBIL L. Crawford, BARBARA Sternfeld, DAVID Morganstein, ELLEN B. Gold, GAIL A. Greendale, DENIS Evans, ROBERT Neer, KAREN Matthews, SHERRY Sherman, ANNIE Lo, GERSON Weiss, and JENNIFER Kelsey. Swan: A multicenter, multiethnic, community-based cohort study of women and the menopausal transition. In ROGERIO A. Lobo, JENNIFER Kelsey, and ROBERT Marcus, editors, *Menopause: Biology and Pathology*, pages 175–188. Academic Press, 2000. ISBN 978-0-12-453790-3. doi:10.1016/B978-012453790-3/50012-3. URL https://www.sciencedirect.com/science/article/pii/B9780124537903500123.

Gail A. Greendale, Barbara Sternfeld, MeiHua Huang, Weijuan Han, Carrie Karvonen-Gutierrez, Kristine Ruppert, Jane A. Cauley, Joel S. Finkelstein, Sheng-Fang Jiang, and Arun S. Karlamangla. Changes in body composition and weight during the menopause transition. *JCI Insight*, 4:e124865, 2019. ISSN 2379-3708. doi:10.1172/jci.insight.124865. URL https://insight.jci.org/articles/view/124865.

Gail A. Greendale, Weijuan Han, Joel S. Finkelstein, Sherri-Ann M. Burnett-Bowie, MeiHua Huang, Deborah Martin, and Arun S. Karlamangla. Changes in regional fat distribution and anthropometric measures across the menopause transition. *The Journal of Clinical Endocrinology and Metabolism*, 106:2520–2534, 2021. ISSN 1945-7197. doi:10.1210/clinem/dgab389.

Kelley Pettee Gabriel, Barbara Sternfeld, Alicia Colvin, Andrea Stewart, Elsa S. Strotmeyer, Jane A. Cauley, Sheila Dugan, and Carrie Karvonen-Gutierrez. Physical activity trajectories during midlife and subsequent risk of physical functioning decline in late mid-life: The Study of Women's Health Across the Nation (SWAN). *Preventive Medicine*, 105:287–294, 2017. ISSN 1096-0260. doi:10.1016/j.ypmed.2017.10.005.