Joint Modeling Methods for Individual-level Variances as Predictors of Health Outcomes

by

Irena B. Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2023

Doctoral Committee:

Professor Michael R. Elliott, Co Chair
Professor Zhenke Wu, Co Chair
Professor Veerabhadran Baladandayuthapani
Professor Siobàn D. Harlow
Professor Abigail Z. Jacobs

Irena B. Chen
irena@umich.edu
ORCID iD: 0000-0002-9366-8506

DEDICATION

I would like to dedicate this dissertation to my family, Hao, Ruth, Travis, Kandace and Max, and to my partner, Logan Walls.

ACKNOWLEDGMENTS

I have received so much support and kindness from many individuals during my graduate studies. I would like to acknowledge some of them here.

First and foremost, I would like to thank my advisors, Michael Elliott and Zhenke Wu. I do not think I would have made it to this point without your unwavering patience and guidance. I have learned so much from working with you and I am grateful for your knowledge and expertise. I would also like to thank my committee members: Sioban Harlow, Veera Baladandayuthapani, and Abbie Jacobs. The feedback and support I received from all of you has helped me grow immensely in my understanding of both the statistical and scientific contexts for my research directions. I would also like to acknowledge my Michigan SWAN and MBHS collaborators, without whom I would not have been able to carry out the research questions described in this dissertation.

The Biostatistics department at the University of Michigan has been an amazing home for the past five years. I have met many passionate fellow students who have inspired and supported me here, and I will treasure your friendship and kindness. In particular, I would like to thank Mengbing Li, Elizabeth Chase, Fatema Shafie Khorassani, Margaret Banker, Kim Hochstedler, Madeline Abbott, and Lam Tran. I also would like to extend a very warm thank you to the wonderful staff members who have answered my numerous emails, helped me get into locked conference rooms, and guided me through technical difficulties regarding the computing clusters. My thanks go out to Nicole Fenech, Sabrina Olson, Mandi Larson, Dave Kubacki, Kyle Terwillegar, and Dan Barker. I also want to give thank the members of the MiSiL lab for providing helpful feedback for my lab presentations, and for trusting me to navigate our bike ride to Ypsilanti where we achieved a 100% survival rate with only 1 flat tire.

Of course, I need to acknowledge the good people of Ann Arbor who have enriched my life here in so many ways. A special thanks in particular to Andy, Alex Bell, Alex Stefanov, Elyse, Lina, Brad, Janelle, Roxie, Steve and Trevor. Thank you for all of the pub trivia, game nights, pop up menu tasting, and rock climbing sessions. Thank you especially for never having once asked me such terrifying questions like “why aren’t you done with your dissertation yet?” I will treasure all of these memories.

Finally, a thank you to my family, Hao, Ruth, Travis and Kandace, and to Logan, my favorite person. Without your for unending love and support, I would not have been able to make it this far.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

LIST OF ACRONYMS

BMC Bone Mineral Content

BMD Bone Mineral Density

E2 Estradiol

FMP Final Menstrual Period

FSH Follicle-Stimulating Hormone

JMIV Joint Model of Individual Means and Variances

JELO Joint Estimation of Longitudinal Outcomes

MBHS Michigan Bone Health Study

MT Menopausal Transition

SWAN Study of Women's Health Across the Nation

ABSTRACT

Precision medicine has the potential to improve early health diagnostics and support individualized treatment plans. The study and identification of repeatedly measured biomarkers for diseases and health risks is essential to advance this field. Existing joint models developed for modeling longitudinal biomarkers have usually focused on estimating the means of the trajectories. However, the variabilities and covariabilities of these trajectories may be informative for health outcomes. This dissertation develops a family of Bayesian hierarchical models that model the individual-level variances and covariances of the trajectories and correlate them to outcomes of interest. The methods presented in this dissertation are designed to handle varying levels of data complexity such as multiple marker trajectories, repeatedly measured and cross-sectional outcomes, and individual time-varying (co-)variances. This body of work supports advances in personalized healthcare by modeling the complex interplay between biomarker means and variances, and corresponding health outcomes. In Chapter 2, I develop a joint model that links estimates of the individual means, variances and covariances of multiple biomarker trajectories to a cross-sectional outcome of interest. This framework can accommodate multiple individual markers by specifying individual variance-covariance matrices in the longitudinal submodel. I propose hierarchical priors on the individual variance-covariance matrices, which allow the model to flexibly capture between-subject differences and similarities in the residual variances and covariances. Simulations demonstrate that this joint model outperforms alternative two-stage approaches. In an application to women's health, I find that higher individual variability of estradiol (E2) is associated with increased fat mass gain across the menopausal transition. This finding indicates that E2 variability may be protective against large increases in waist circumference in midlife women and raises new questions regarding the role ofE2 variability in predicting fat distribution changes during menopause. In Chapter 3, I examine the setting of simultaneously estimating multiple longitudinal trajectories, in order to understand associations between variables over time. I explore a linear parameterization of time-varying individual variances in the predictor model so that the individual variances, as well as the means, are used to predict the outcome at the same point in time. I demonstrate via simulation studies that this model is able to recover the true data generating parameters while maintaining low bias and proper coverage. I apply this method to women's hormone markers and bone density measurements during the midlife and find that higher follicle-stimulating hormone

(FSH) variability is associated with slower declines in bone density. Our findings suggest that FSH variability, but not E2 variability, is a more predictive measurement of bone health in midlife women. Chapter 4 introduces a joint model of individual-level mean and covariances trajectories of multiple markers for estimating a repeatedly measured health outcome. In the predictor sub-model, the individual variance-covariance matrices comprised a shared residual covariance matrix and individual-specific regression coefficients that characterize the evolution of the variances and covariances over time. This method is applied to estimate the associations between FSH and testosterone variabilities and bone mineral content declines in women undergoing menopause. We find for the first time, high variability of testosterone is associated with faster declines in bone mineral content (BMC) for post-menopausal women. Conversely, higher covariability between FSH and testosterone post-FMP was also associated with slower declines in BMC. A simulation study validates that the model can recover the parameters of interest with low bias and high coverage.

CHAPTER 1

Introduction

Laird and Ware (1982) discussed a family of models for analyzing longitudinal data that included specifying individual level random effect parameters. Since then, there have been extensive advances in statistical methods for analyzing longitudinal data. Over the past two decades, there has been a broad body of literature that uses longitudinal data from biomarkers or questionnaires 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. 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.

A joint model is particularly useful for linking information from longitudinal markers to health outcomes. A standard joint model usually has the following specification that involves a trajectory model and an outcome model:

$$X_{ij} \sim \mathcal{N}(\mu(\boldsymbol{b}_i, t_{ij})\ S), \boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{\beta}, \Psi) \tag{1.1}$$

$$Y_i \sim \mathcal{N}(\eta(\boldsymbol{b}_i), \sigma^2) \tag{1.2}$$

where X_{ij} is the value of the biomarker for individual i at time t_{ij}, Y_i is the outcome of interest for each individual, and the two are linked by the mean and variance parameters ($\boldsymbol{b}_i, S$ respectively) governing the individual trajectories. Typically, a prior is also set on the $\boldsymbol{b}_i$ parameters. S is usually also either specified in advance or estimated from the data with an appropriate prior (e.g. inverse-Gamma or half-Cauchy). The advantage of joint models is that they fully account for the uncertainty in the longitudinal parameter estimates, by estimating both sub-models simultaneously (Sayers et al., 2016).

Recently, there has been some exploration of modeling individual-level variabilities rather than defining a single common population variance for all individuals (Elliott et al., 2012; Huang et al.,

2014). A joint model that allows for intra-individual variability could look something like:

$$X_{ij} \sim \mathcal{N}(\mu\,(\,\boldsymbol{b}_i, t_{ij})\;S_i), \tag{1.3}$$

$$Y_i \sim \mathcal{N}(\eta(\boldsymbol{b}_i, S_i), \sigma^2), \tag{1.4}$$

where S_i, like $\boldsymbol{b}_i$, can then be estimated with a hierarchical prior. Additionally, in this model, the residual variances, S_i, can be used to predict the outcome Y_i along with the predictor means $\boldsymbol{b}_i$. Joint models that use individual-level variances as predictors have been applied to health outcomes such as cardiovascular disease risk (Parker et al., 2021) and kidney transplant risk complications (Campbell et al., 2021).

The models in this dissertation add substantial contributions to the current literature on variance as a predictor for health outcomes. In this dissertation, I develop and implement new statistical methods for modeling these individual level variances with the goal of using these variances to predict health outcomes. I present methods for utilizing individual variances as predictors within a joint model framework, modeling the variances and covariances of multiple marker trajectories, and modeling time-varying variances, as shown in Table 1.1.

Dissertation Project	Multiple biomarkers	Longitudinal outcome	Time-varying variance
Chapter 2	✓		
Chapter 3		✓	✓
Chapter 4	✓		✓

Table 1.1: **Overview of joint modeling methods for individual variances developed in each dissertation chapter.**

In Chapter 2, I propose a model for estimating the individual means, variances, and covariances of multiple biomarkers. These estimates are then used in the outcome submodel to link the trajectories to a cross-sectional outcome of interest. I describe a flexible method for modeling the individual variance-covariance matrices S_i by using the decompositon approach from Barnard et al. (2000) and placing weakly informative priors on the components. I demonstrate via simulations that this joint model is able to recover the true data generating values and also outperforms comparative two-stage approaches to the same estimation setting. I also apply this model to women's health data to evaluate the associations between longitudinal Estradiol (E2) and Follicle-Stimulating Hormone (FSH) measurements and changes in fat mass and waist circumference across the menopausal transition.

In Chapter 3, I turn to the setting of repeatedly measured health outcomes. I extend the joint modeling framework from Chapter 2 to capture the associations between a longitudinal predictor and a longitudinal outcome of interest. The model simultaneously estimates the longitudinal pre-

dictor and outcome, where the two trajectories are linked via shared random effects. I also explore a linear parameterization of individual time-varying variance and incorporate this into the overall joint model. Simulation studies indicate that the model can recover the data-generating parameters with low bias and high coverage in the setting with constant variances and heteroskedastic variances. Simulation studies demonstrate 1) the advantage of the joint model over comparative two-stage approaches and 2) the relative performance of the time-varying variance specification over the time-invariant model under different variance scenarios. Finally, I apply this model to analyze the associations between certain hormone markers (E2 and FSH) and Bone Mineral Density (BMD) declines in middle-aged women.

In Chapter 4, I introduce a joint model for modeling individual time-varying variabilities and covariabilities of multiple markers. This chapter extends the application of covariance regression to a joint modeling setting. The individual variance-covariance matrix is parameterized by a residual time-invariant covariance matrix that is shared across individuals and a set of individual-level regression coefficients that describe the evolution of the variances over time. We apply this model to women's hormone and bone data to evaluate the associations between certain hormone marker varabilities and correlations and declines in bone mineral content. This model is also able to estimate the evolution of FSH and testosterone variances and covariances over time, providing novel insights into the individual-level variance trajectories of these hormones.

Finally, in Chapter 5, I summarize the major contributions developed in this dissertation, and suggest future areas of research. The following sections provide an overview of the existing literature and areas of research on models of individual variability.

1.1 Individual Variances as Predictors

The potential importance of individual variance as a predictor or measurement of health outcomes has been recognized in many fields. Intra-individual variability of reaction times is known to be significantly associated with outcomes such as cognitive performance (Jensen, 1992; MacDonald et al., 2006) and individual variability in memory speed tasks is considered to be an index for central nervous system function (Ram et al., 2005; Dykiert et al., 2012). High short term variability in blood pressure measurements is associated with increased mortality risk for patients with untreated hypertension (Hsu et al., 2016). 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 an important predictor of menopausal onset (Huang et al., 2014).

Sammel et al. (2001) designed a two-stage model that linked individual means and variances of longitudinal profiles to a corresponding health outcome and applied this to peri-menopausal

womens' health data. In the first stage, they estimated the individual mean profiles and residual variances of a reproductive hormone (estradiol), and then used these estimates in the second stage to predict prevalence of hot flashes. They found that high variability in these hormone levels was associated with experiencing more severe hot flashes.

Elliott et al. (2012) extended this idea of using individual variances from longitudinal data to a joint model setting. The individual mean trends, residual variances, and outcome of interest were estimated simultaneously:

$$X_{ij} \mid \beta_i, \sigma_i^2 \sim \mathcal{N}(\mu(\beta_i, t_{ij}), \sigma_i^2)$$
$$Y_i \mid \beta_i, \sigma_i^2 \sim BER(\pi_i), log(\frac{\pi_i}{1-\pi_i}) = \eta(\beta_i, \sigma_i^2)$$

where prior distributions were assigned to β_i, σ_i^2 and the posterior distributions were estimated using MCMC methods. Jiang et al. (2015) also used a Bayesian joint model framework to link estimates of individual variances to an outcome of interest. In this model, the means and variances of longitudinal profiles are connected to the outcome either as shared random effects or by latent classes. We note that a common feature lacking in these models is the handling of multiple markers in the longitudinal submodel. Additionally, these methods mostly focus on cross-sectional or summary outcomes of interest, rather than a longitudinal outcome, and also assume a constant individual variance σ_i across all timepoints.

This dissertation adds to the current research on individual variability by addressing these gaps in the literature. Chapter 2 focuses on two main objectives: 1) modeling individual-level variances and co-variances of multiple longitudinal trajectories by proposing flexible hierarchical priors for residual variance-covariance matrices and 2) incorporating these subject-level means and variance-covariance matrices for longitudinal data and a cross-sectional outcome in a joint modeling framework. Chapter 3 develops a joint model framework that incorporates individual time-varying variability as a predictor for longitudinal outcome. Chapter 4 presents a method for estimating time-varying covariance matrices, so that these heteroskedastic variances and covariances can also be used as predictors in the outcome model. I design several simulation studies for these models to evaluate their behavior under various scenarios and compare their performance against alternative two-stage methods. I also apply these models to women's health applications in order to understand how the variability and co-variability of certain biomarkers affects health outcomes such as body mass changes and bone loss trajectories during the women's midlife.

1.2 Individual variances of multiple markers

A standard formulation for joint models with multiple longitudinal predictors is as follows:

$$\begin{aligned}\boldsymbol{X}_{ij} \mid \mathbf{B}_i, \Psi &\sim \mathcal{N}_Q\left(\mu(\mathbf{B}_i, t_{ij}), \Psi\right) \\ Y_i \mid \mathbf{B}_i, \boldsymbol{\beta}, \sigma^2 &\sim \mathcal{N}(\eta(\mathbf{B}_i, \boldsymbol{\beta}), \sigma^2),\end{aligned}$$

where Q is the number of predictors, $\mathbf{B}_i = [\boldsymbol{b}_{i1}, ..., \boldsymbol{b}_{iQ}]^\mathsf{T}$ and $\boldsymbol{b}_{iq} = (b_{iq1}, \ldots, b_{iqP})^\mathsf{T}$ are vectors of P regression coefficients for the q-th marker and Ψ is the population-level variance covariance matrix for the multiple markers. Estimation of Ψ can either be done via banding methods (Wu and Pourahmadi, 2009; Bickel and Gel, 2011), or penalized regression for precision matrix components (Meinshausen and Bühlmann, 2006). Alternatively, in a Bayesian framework, a standard prior such as the Inverse-Wishart can be specified for Ψ.

Long et al. (2016) presented a Bayesian framework for modeling the correlation between multiple longitudinal markers. If Z_{ip} are the measurements for biomarker p for individual i and subject to some measurement error, then the observed values z_{ipk} can be written as:

$$\begin{aligned}z_{ipk} &= Z_{ip}(t_{ik}) + \widetilde{\epsilon}_{ipk}, i = 1, \ldots, n, p = 1, \ldots, P, k = 1, \ldots, m_i \\ \widetilde{\epsilon}_{ipk} &= b_i + \epsilon_{ipk}\end{aligned}$$

where b_i is the subject level random effect that captures the additional correlation between the measurements of different biomarkers from subject i. They placed a Normal $\mathcal{N}(0, \sigma_b^2)$ prior on this correlation parameter and applied this model to a study on biomarkers of colorectal adenoma data and risk of colorectal cancer. Li et al. (2021) modeled the correlation between multiple longitudinal biomarkers as a function of an individual specific stochastic trend over time, $\mu_i(t)$, which "characterizes the overall underlying trend shared by the biomarkers", and used this as a regressor in the time-to-event outcome submodel. The marker correlations $\mu_i(t)$ are related to the j-th biomarker with an factor loading coefficient β_j, common across all individuals. In this specification, it is the latent process $\mu_i(t)$ and not the individual residual correlations between biomarkers themselves, that is used in the outcome submodel.

In these models of multiple marker variability, the individual residual variances and covariances are not treated as predictors of interest for the corresponding outcome, which is one of the main contributions in Chapter 2. Additionally, these approaches have focused on modeling the correlation between markers as a population-level parameter, rather than on modeling individual-specific residual variance-covariance matrices. This is also addressed in Chapter 2, where I formulate a method for estimating individual variance-covariance matrices for multiple biomarkers. In this

model, the observed marker values X_{ij} are defined as:

$$\begin{aligned}\boldsymbol{X}_{ij} \mid \boldsymbol{B}_i, \mathbf{S}_i &\sim \mathcal{N}_Q(\mu\left(\boldsymbol{B}_i, t_{ij}\right), \mathbf{S}_i) \\ Y_i \mid \boldsymbol{B}_i, \boldsymbol{S}_i, \sigma^2 &\sim \mathcal{N}_Q(\eta(\boldsymbol{B}_i, \mathbf{S}_i), \sigma^2),\end{aligned}$$

where $\mathbf{S}_i$ is estimated for each individual. In the application, I evaluate two hormone markers of interest and use the estimated residual variances and covariances to predict changes in body fat mass composition. I also apply this method to 3 markers via a simulation study and find that it estimates the variance-covariance matrices with reasonable runtime.

1.3 Joint models with longitudinal outcomes

There is a voluminous body of literature regarding random effects in longitudinal models (Diez-Roux, 2000; Greene, 2005; Diggle et al., 2013). These models are typically specified as:

$$\begin{aligned}Y_{ij} &= \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}\boldsymbol{b}_i + \epsilon_{ij}, \\ \epsilon_{ij} &\sim \mathcal{N}(0, \sigma^2)\end{aligned}$$

Where Y_{ij} is the longitudinal outcome of interest, $\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}$ are the covariates for the population (fixed) effects $\boldsymbol{\beta}$ and individual (random) effects $\boldsymbol{b}_i$. To extend this to a joint modeling framework, we would model the longitudinal predictor X_{ij} as a function of time and individual-specific regression coefficients, e.g. $X_{ij} \mid \beta_i, t_{ij}, \omega, \sim \mathcal{N}(\eta(\beta_i, t_{ij}), \omega)$ where β_i are the individual-level mean parameters. This allows us to estimate individual level mean and variances from the predictor and relate these to the health outcome of interest.

In Chapter 3, I present a joint model for a longitudinal predictor and longitudinal outcome that estimates the individual-level variability in the predictor. The individual variances are used as predictors in the outcome model so that the effect of individual variance on the outcome of interest is explicitly captured. The proposed model borrows from the framework of the mixed effects model:

$$\begin{aligned}X_{ij} \mid \boldsymbol{b}_i, s_i \ &\sim \ \mathcal{N}\left(\boldsymbol{\mu}(t_{ij}; \boldsymbol{b}_i), s_i^2\right), j = 1, \ldots, n_i, i = 1, \ldots, N, \\ Y_{ij} \mid \boldsymbol{b}_i, \boldsymbol{a}_i, t_{ij}, s_i &\sim \mathcal{N}\left(\eta(\boldsymbol{b}_i, \boldsymbol{a}_i, t_{ij}, s_i), \sigma^2\right), j = 1, \ldots, n_i, i = 1, \ldots, N,\end{aligned}$$

This specification directly links the predictor X_{ij} to the outcome Y_{ij} via the means $\boldsymbol{b}_i$ and the variances s_i. This model links the two trajectories via the random effects from the predictor variable

and modeling the random effects as regression coefficients. Additionally, this model treats the individual variances s_i's as being predictive of the outcome and includes these estimates in the outcome model.

1.4 Time-varying variances and covariances

The methods reviewed in the previous sections 1.1,1.2 and 1.3 generally do not consider the setting of time-varying variability, where the individual residual variances (and covariances) could change over time. In many applications, the assumption of a homoskedastic variance-covariance matrix is often unreasonable. For financial time series analyses, a well known model for time-varying variability in the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Proposed by Bollerslev (1986), the GARCH model parameterizes the variance of the current time period as a function of previous time periods' variances. The stochastic volatility model, a common alternative to GARCH models, parameterizes the variance of a time series with a latent stochastic process (Chan and Grant, 2015).

Within the setting of multivariate regression, there are many applications where the covariance matrix, as well as the mean, may depend on external covariates and/or change over time. Li et al. (2014) descrbie how the variability of questionnaire responses from hospital nurses may be dependent on covariates such as hospital location and years of work experience. Fox and Dunson (2015) studied the evolution of weekly flu activity trends across the United States, where the means and variances of the estimated disease trajectories may be changing over time and space. The outcome of interest is described as follows:

$$\boldsymbol{X}_i \mid \boldsymbol{Z}_i \sim \mathcal{N}(\mu\left(\boldsymbol{Z}_i\right), \Sigma\left(\boldsymbol{Z}_i\right))$$

,

so that the mean $\mu\left(\boldsymbol{Z}_i\right)$ and covariance matrix $\Sigma\left(\boldsymbol{Z}_i\right)$ of $\boldsymbol{X}_i$ are both functions of covariates $\boldsymbol{Z}_i$, which can include time. This model can be extended to the longitudinal data setting to estimate individual or group-specific means and covariance matrices (Niu and Hoff, 2019; Li et al., 2014).

There are several approaches for estimating these covariances matrices in the multivariate regression setting. Chiu et al. (1996) modeled the logarithm of the covariance matrix, which allows the elements of the log-transformed covariance matrix to vary freely, but also still ensures that the corresponding covariance matrix is postive-definite. Other methods utilize the covariance matrix's Cholesky factor (Pourahmadi, 1999; Pan, 2003), or a regression function (Hoff and Niu, 2012; Li et al., 2014; Fox and Dunson, 2015). Although these model the covariance matrices as a function of covariates, they do not utilize these variance estimates for inference or prediction

of another outcome of interest, in either a two-stage or joint model framework. In Chapter 4, I extend the covariance regression method proposed by Hoff and Niu (2012) to the joint modeling setting. We return to the setting of multiple biomarkers and develop a joint model that allows for the individual-level marker variances and correlations to vary over time. Our model uses the covariance regression model in the predictor submodel:

$$\begin{aligned}\boldsymbol{X}_{ij} \mid \mathbf{B}_i, \mathbf{S}_i &\sim \mathcal{N}_Q\left(\boldsymbol{\mu}_{ij}(\mathbf{B}_i, t_{ij}), \mathbf{S}_{ij}(\Theta_i, t_{ij})\right) \\ Y_{ij} \mid \boldsymbol{\mu}_{ij}, \boldsymbol{S}_{ij}, \boldsymbol{\beta}, \sigma^2 &\sim \mathcal{N}(\eta(\boldsymbol{\mu}_{ij}, \mathbf{S}_{ij}, \boldsymbol{\beta}), \sigma^2),\end{aligned}$$

where $\boldsymbol{\mu}_{ij}, \boldsymbol{S}_{ij}$ are now both functions of time. Using this model, we are able to estimate how these variances and covariances may change over time. As in the previous chapters, these estimated variances (and covariances) are used in the outcome submodel to predict individual health trajectories.

1.5 Applications to women's heath

The applications presented in this dissertation have important implications for women's health research. As mentioned before, variability of biological mechanisms can indicate disregulations or the breakdown of certain systems within the body. E2 and FSH are two hormones are known to be important regulators of the female reproductive system, as well as other tissues (e.g. fat or bone) (Ryan, 1982; Colleluori et al., 2018; Zaidi et al., 2018). As women approach their Final Menstrual Period (FMP), mean E2 tends to decline while mean FSH tends to rise in response. Additionally, both of these hormones are known to have extreme fluctuations during the Menopausal Transition (MT). In particular, high variability of E2 may be predictive of depressive events during the MT (Gordon et al., 2016) and high variability of FSH is associated with increased risk of hot flash (Jiang et al., 2015). Chapter 2 focuses on evaluating the associations between individual level E2 and FSH variances and covariances and body mass rate of change across the MT.

Chapter 3's model is motivated by observations of women's bone health as they transition through menopause. Early detection and intervention for declining BMD is critical for reducing the risk of poor outcomes (e.g. fractures) post-menopause. E2, and more recently FSH, have both been studied as important biomarker predictors for BMD in midlife women (Ebeling et al., 1996; Chin, 2018). By characterizing the longitudinal relationship between E2 (FSH) and physical function during the midlife, this research can support the design of effective interventions to preclude substantial bone loss as women transition through the midlife into their later years. Chapter 2 and Chapter 3 used data from the Study of Women's Health Across the Nation (SWAN) study, a

multi-site, multi-ethnic cohort spanning almost three decades of data collection.

Chapter 4 focuses on modeling time-varying variances and covariances of certain hormone markers as women transition through menopause and relate these variance estimates to declines in Bone Mineral Content (BMC). Here, we examine women's testosterone levels as a potential marker of BMC decline, something that few existing studies have examined previously. We jointly model the time-varying means, variances, and covariance of FSH and testosterone and their associations with BMC trajectories in midlife women. We use data from the Michigan Bone Health Study (MBHS), a 33-year cohort study that investigates bone strength declines during the MT. By estimating the time-varying variances and covariances of FSH and testosterone as women enter menopause, we can evaluate how these hormone variabilities at different stages of menopause may impact BMC decline.

CHAPTER 2

Chapter 2: 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

2.1 Introduction

The menopausal transition is a critical lifestage that can shape women's midlife and long-term health. The US census bureau estimates that by 2050, approximately 47 million women will be aged 45 to 64 years (U.S. Census Bureau, 2017), and women are projected to spend more than one-third of their life post-menopause (Mohammadalizadeh Charandabi et al., 2015). Therefore, understanding how the midlife can affect health outcomes is vital for supporting a healthy aging population.

Reproductive aging and the menopausal transition are characterized by well-established patterns of falling levels of estradiol (E2) and rising levels of follicle-stimulation hormone (FSH) (Randolph et al., 2004). In addition to regulating reproductive functionality, E2 and FSH have also been found to be highly associated with risk of adverse health outcomes (Karvonen-Gutierrez and Harlow, 2017; Zaidi et al., 2018). Since E2 regulates adipose tissue, women tend to gain fat mass post-menopause (Colleluori et al., 2018). Fat mass distribution also changes, with body fat shifting to the intraabdominal region during the menopausal transition (Carr, 2003). Excess abdominal fat is one of the symptoms of metabolic syndrome, which can place individuals at higher risk of health conditions such as heart disease, diabetes, and stroke. Waist circumference is a commonly used measure of abdominal fat and previous research suggests that waist circumference may be an important indicator of health risks (Ross et al., 2020; Darsini et al., 2020).

Higher increases in FSH levels are also associated with higher fat mass increases in women undergoing menopause (Sowers et al., 2007). Additionally, FSH appears to be an important pre-

dictor 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). This motivates further investigation into how E2 and FSH can jointly predict body mass composition in women. Identifying these associations is important since excess body weight can increase the risk of adverse health outcomes and mortality in midlife women (Stevens et al., 2002)

Growing evidence also suggests that the variability of these hormones may be critical for predicting adverse health outcomes. Gordon et al. (2016) found that higher E2 variability in women over a period of 14 months was predictive of greater depressive symptoms at month 14. Lower levels of FSH variability in perimenopausal and postmenopausal women was strongly associated with reduced risk of hot flash, while changes in individual mean FSH trajectories were not similarly predictive of hot flash risk (Jiang et al., 2015). By understanding how the variability of these biomarkers relate to changes in body mass, we can improve health diagnostics for women and support individualized treatment plans. Despite this emerging work, the majority of current research has still focused on using mean hormone measurements or group based trajectories to predict health outcomes. The existing literature does not account for how individual variabilities or co-variability of E2 and FSH may be related to changes in body mass and waist circumference across the menopausal transition. Thus, there is a dearth of statistical models that properly extract and use these individual hormone trends and variabilities when predicting an outcome.

2.1.1 SWAN Dataset

Our motivating dataset comes from the Study of Women's Health Across the Nation (SWAN), a multi-site US-based longitudinal cohort study that followed women over the menopausal transition (Sowers et al., 2000a). The SWAN dataset has made it possible to establish longitudinal associations between hormone trajectories and health outcomes (Park et al., 2017; Sowers et al., 2007), rather than relying on baseline hormone measurements to predict health risks. The comprehensive and longitudinal aspect of this dataset makes it ideal for understanding how individual hormone trends can predict changes in body mass composition.

To be eligible for the SWAN 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 prior to enrollment in the study, 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 of the 15 approximately annual follow-up visits, along with other health measurements. Figure 2.1 shows E2 and FSH measurements collected from SWAN participants, along with a loess curve to estimate the overall population trend. Body composition was measured via dual-

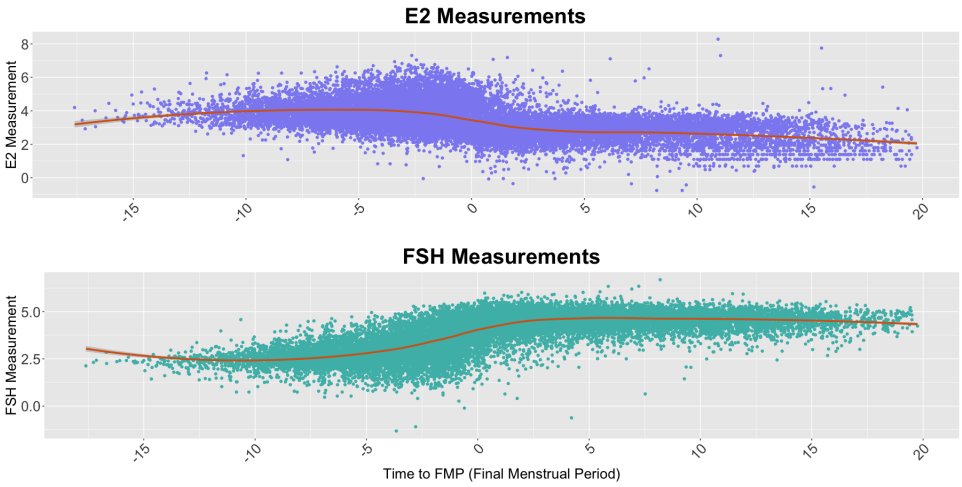


Figure 2.1: **Plots of the observed E2 and FSH measurements. In our analysis, we log-transformed these measurements and then detrended them by subtracting the individual observations from a population loess fit.**

energy X-ray absorptiometry (DXA) at five of the seven clinical site visits. Women also completed questionnaires regarding lifestyle and socio-demographic characteristics.

For the fat mass dataset, we initially started with women who were enrolled at one of the five sites with body composition measures. If a woman was on hormone replacement therapy during a clinical visit, we removed that observation from the dataset. Additionally, we removed women who did not have an observed FMP. Although the SWAN study enrolled five racial/ethnic groups, the site with Hispanic women did not have body composition data; hence, Hispanic women are not included in the fat mass analysis. After computing the fat mass composition window (see below), an additional 47 women were excluded from the analysis, due to either not having both pre and post FMP observations or not having observations that fell within the desired time range before and after FMP (i.e. observations outside of the 3-7 year range before and after FMP). The 3-7 year range was chosen in order to ensure that the changes in fat mass distribution was captured sufficiently before the start (and end) of the menopausal transition, and the number of women who had measurements beyond the 7 year cutpoint before and after FMP was scarce. The final analytical fat mass dataset was completed on 841 individuals, with a total number of 9,902 hormone measurements.

For waist circumference analysis, since all seven sites collected waist circumference measurements, we were able to have a larger sample size for this analysis. As in the fat mass analysis, we removed women who did not have an observed FMP and we also removed observations where the woman was on hormone replacement therapy. Our final analytical dataset for waist circumference was completed on 1,029 individuals and 12,059 hormone measurements.

For the individual trajectory model, we use the log values of FSH (mIU/mL) and E2 (pg/mL) measured at each visit. We removed the E2 and FSH population trends by fitting a lowess curve

to each (log) hormone. The lowess curve was fit 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 2.3 summarizes the subject-level longitudinal data model fit results for two randomly-selected women.

For the outcomes of interest, we selected fat mass rate of change and waist circumference rate of change over a selected time 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 (waist circumference) rate of change is the difference between the 'last visit' (post FMP) and the 'first visit' (pre FMP) divided by the amount of time (in years) within each individual window. Figure 2.2 shows the individual observations of these rates. We normalized the fat mass measurements by using the proportion (i.e. ratio) of fat mass to body weight (grams) rather than using the unadjusted fat mass measurements (also in grams), thus creating a measure of percent fat mass for each woman. Raw fat mass values are highly correlated with body weight, and previous work has demonstrated that there is no menopausal effect of body weight change beyond normal aging. The proportion of fat mass, however, has a strong curvilinear relationship across the menopausal transition. For this reason, we have used fat mass proportion in our analyses to reflect the impact of the MT on this measure. (Greendale et al., 2019). Figure 2.2 displays the histograms of these outcomes, after performing the normalization (for fat mass) and rate adjustments (for both models).

2.1.2 Methods for longitudinal markers and health outcomes

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 predict dropout in schizophrenia trials; they used 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 utility of residual variability in predicting health outcomes

Variable	Statistic	Value	n
Longitudinal Predictors	Mean/SD		
E2 Residuals		-0.04 (0.81)	9,902
FSH Residuals		0.02 (0.61)	9,902
Health Outcome	Mean/SD		
Fat 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
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 2.1: **Descriptive statistics of the fat mass dataset based on 841 individuals.**

Variable	Statistic	Value	n
Longitudinal Predictors	Mean/SD		
E2 Residuals		-0.003 (0.80)	12,059
FSH Residuals		-0.008 (0.62)	12,059
Health Outcome	Mean/SD		
Waist Circumference Rate of Change		0.41 (0.82)	1,029
Baseline Waist Circumference	Mean/SD		
Waist Circumference at Visit 1		86.03 (15.90)	1,029
Race/Ethnicity	Percent		
White (Reference)		47.0%	484
Black		26.2 %	270
Japanese		15.3%	128
Chinese		11.5%	119
Hispanic		2.7%	28
Physical Activity	Percent		
Lowest Activity (Reference)		23.6%	267
Increasing Activity		12.7%	131
Decreasing Activity		22.6%	233
Middle Activity		24.0%	247
Highest Activity		14.6%	151

Table 2.2: **Descriptive statistics of waist circumference dataset based on 1,029 individuals.**

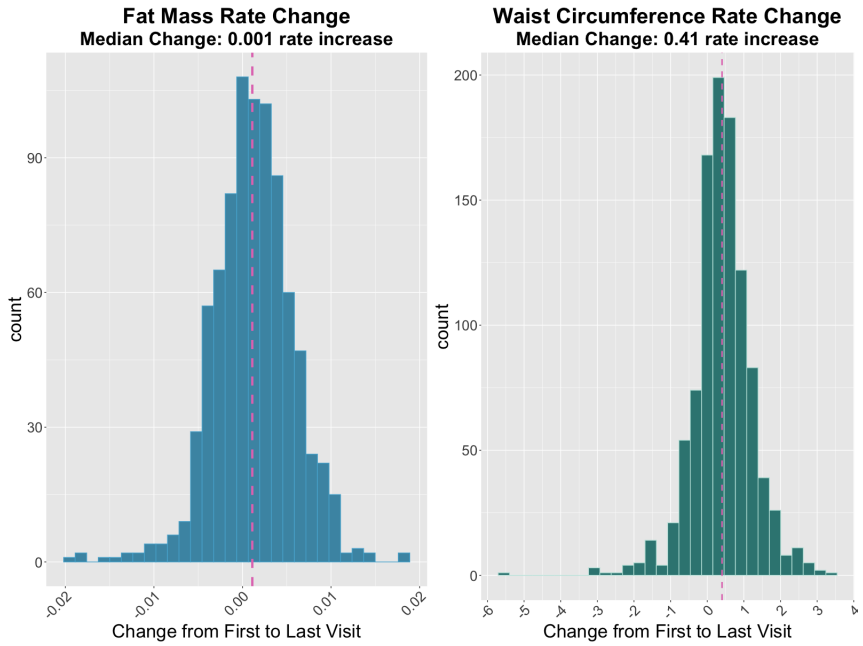


Figure 2.2: **Histograms of the observed rate of change in fat mass composition (left) and the observed rate of change in waist circumference (right). The observed values for the fat mass outcome are in fat mass proportional to body weight (both in grams) per year and waist circumference (cm) per year.**

are largely lacking in the literature, despite calls for increased focus on developing methods for variance structures. Such methods may elucidate whether variability as predictors 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). Our proposed method addresses this gap by explicitly parameterizing individual variances (and co-variances) and uses these estimates as predictors in the outcome model.

Early literature used simple two-stage approaches with a squared-error estimate of variance obtained from the observed data to predict a single outcome (Sammel et al., 2001), ignoring the inherent uncertainty in the constructed variance estimates. More recent methods have focused on a joint model for the predictors and outcomes (Elliott et al., 2012; Jiang et al., 2015, e.g.,). The joint modeling approach is critical given that statistical uncertainty of the constructed predictors, e.g., mean and residual variance estimates, if unaccounted for, can lead to extremely biased estimates of the effects of these individual-level mean trajectories and the residual variances on the outcomes (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. In a Bayesian frame-

work, our joint model properly accounts for the uncertainty in estimating the residual variances and covariances of multiple biomarkers, thereby propagating the uncertainty into outcome prediction.

This paper is organized as follows. Section 2.3 describes our joint model framework linking the mean trajectory and residual variance model for the longitudinal predictors with the model for the outcome. Section 2.4 conducts extensive simulation studies to show that our proposed approach produces less biased estimates of the outcome regression coefficients with valid statistical uncertainty assumptions 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 2.5 applies our method to assess the associations between the mean trajectories, variance, and covariance of E2 and FSH and the changes in fat body mass and fat distribution during the menopausal transition using women's health data from the SWAN study. Section 2.6 discusses the implications of our work along with limitations and directions for future research.

2.2 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. In mathematical notation, these joint models typically take the form:

$$\begin{aligned} \boldsymbol{X}_{ij} \mid \boldsymbol{B}_i, \Psi &\sim \mathcal{N}(\mu(\boldsymbol{B}_i, t_{ij}), \Psi), \\ Y_i \mid \boldsymbol{B}_i, \beta, \sigma^2 &\sim \mathcal{N}(\eta(\beta, \boldsymbol{B}_i), \sigma^2), \end{aligned}$$

where X are the observed markers/predictors, Y is the observed outcome of interest, the main predictors of interest are the mean parameters $\boldsymbol{B}_i$ and the variance-covariance matrix Ψ is assumed to be a population level parameter, rather than allowing each individual to have their own variance-covariance matrix (e.g. S_i).

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 relation-

ship 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). Sammel et al. (2001) used a two-stage model to that linked individual means and variances of longitudinal profiles to a corresponding health outcome. They found that E2 variability during the menopausal transition was highly predictive of experiencing hot flashes. Jiang et al. (2015) proposed a joint model of individual means and variances of FSH hormone trajectories and risk of hot flash as the health outcome. Low FSH variability was predictive of substantial reduction in risk of hot flash. We note that a common feature of these models is the lack of multiple predictors in the longitudinal sub-model.

Our work's most important contribution over the previous work is to consider the individual-level means and variances of multiple predictor trajectories, rather than a single biomarker trajectory, in a joint modeling framework. This allows for investigation into how these trajectories may both independently and interactively associate with outcomes, and in turn requires substantial methodological development, particularly to decompose an individual-level variance-covariance matrix for use in a joint modeling setting. This work also adds to the still small set of literature showing the use of variability as a predictor of health outcomes, and more generally emphasizes the need for a joint modeling framework over the less efficient two-stage approach, as we show in Section 2.4. Also, to the best of our knowledge this is the first rigorous assessment of the role of individual trends and variability of E2 and FSH hormones in jointly predicting body mass changes.

2.3 Proposed Model

2.3.1 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.

2.3.2 Likelihood

2.3.2.1 Component 1: Longitudinal Markers

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

$$\boldsymbol{X}_{ij} \mid \mathbf{B}_i, \mathbf{S}_i \sim \mathcal{N}_Q\left(\boldsymbol{\mu}(t_{ij}; \mathbf{B}_i), \mathbf{S}_i\right), j = 1, \ldots, n_i, \text{ independently for } i = 1, \ldots, N, \quad (2.1)$$

$$\boldsymbol{b}_{iq} \overset{\text{indep.}}{\sim} \mathcal{N}_P(\boldsymbol{\beta}_q, \Sigma_q), q = 1, \ldots, Q, \quad (2.2)$$

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 our simulation and data application, we specify these basis functions in advance to be linear functions of time. 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 (2.2) has assumed individual-specific regression coefficients for any two markers are conditionally independent. Note that in this model, we have given each individual a variance covariance matrix $\mathbf{S}_i$.

2.3.2.2 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 (2.1) as follows:

$$Y_i \mid \mathbf{B}_i, \mathbf{S}_i, \boldsymbol{W}_i \sim \mathcal{N}\left(\eta_i, \sigma^2\right), i = 1, \ldots, N, \quad (2.3)$$

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 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 2.5.3, we illustrate a model with a Gaussian scale-mixture outcome regression model for waist circumference rate-of-change outcome.

2.3.3 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

$$\log(d_{iq}) \sim \mathcal{N}(\nu_q, \psi_q^2), \tag{2.4}$$

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}:

$$(r_{i12} + 1)/2 \sim \mathsf{Beta}(a'_{12}, b'_{12}), \text{ independently for } i = 1, \ldots, N. \tag{2.5}$$

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

$$\nu_q \sim \mathcal{N}(m, \xi_q^2), \psi_q \sim \mathsf{half\text{-}Cauchy}(0, \tau), \text{ independently for } q \leq Q, \tag{2.6}$$

$$a'_{12} \sim \mathsf{Exp}(\kappa), \; b'_{12} \sim \mathsf{Exp}(\kappa'). \tag{2.7}$$

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 signal of the variance ψ_q may be weak (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 estimates of the variance, but constrains the posterior distribution "to an extent allowed by the data".

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., 2009a), 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 (2.7), 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 using Inverse-Wishart distribution as a prior for variance-covariance matrices, which is also governed by a single scale parameter and suffers from the same drawbacks as the LKJ distribution.*

There is little existing literature on hyperprior recommendations for the parameters of the Beta distribution. Robert and Casella (2010) note that "there exists a family of conjugate priors on a, b", however, they also note that these prior distributions are often intractable, due to the "difficulty of dealing with the Gamma function". Instead, we opt for a simpler approach by allowing the a, b parameters to be independently drawn from an Exponential prior. We argue that the Exponential distribution in Equation (2.7) is a reasonable choice for a hyperprior 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: $\widehat{a}_{MLE} = \frac{1}{2} + \frac{\widehat{G}_x}{1 - \widehat{G}_x - \widehat{G}_{(1-x)}}, \widehat{b}_{MLE} = \frac{1}{2} + \frac{\widehat{G}_{(1-x)}}{1 - \widehat{G}_x - \widehat{G}_{(1-x)}}$ 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 $\frac{1}{2} + \frac{\widehat{G}_x \exp(\lambda_a)}{1 - \widehat{G}_x \exp(\lambda_a) - \widehat{G}_{(1-x)} \exp(\lambda_a)}, \frac{1}{2} + \frac{\widehat{G}_{(1-x)} \exp(\lambda_b)}{1 - \widehat{G}_x \exp(\lambda_b) - \widehat{G}_{(1-x)} \exp(\lambda_b)}$. When $\lambda_a, \lambda_b \to \infty$, the posterior modes of a, b shrink towards $\frac{1}{2}$, 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

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

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

2.3.3.1 Priors for population longitudinal marker regression coefficients:

$$\boldsymbol{\beta}_q \sim \mathcal{N}_P(0, \xi_q^2 I_{P\times P}), \text{ independently for } q = 1, \ldots, Q, \tag{2.8}$$

$$\Sigma_q = \mathbf{K}_q \mathbf{L}_q \mathbf{K}_q, \ \mathbf{K}_q = \mathrm{diag}\{k_{q1}, \ldots, k_{qP}\}, q = 1, \ldots, Q, \tag{2.9}$$

$$k_{qp} \sim \mathsf{half\text{-}Cauchy}(0, \tau_0), p = 1, \ldots, P, \text{ and } \mathbf{L}_q \sim \mathsf{LKJ}(\zeta), \tag{2.10}$$

independently for $q = 1, \ldots, Q$ 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 (2.10) 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 Equations (2.4 - 2.5) that specifies the prior for multiple and individual-specific variance-covariance matrices.

2.3.3.2 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 $\sigma \sim \mathsf{half\text{-}Cauchy}(0, \tau_1)$. In our

simulation studies and data analysis, we set the priors on the regression parameters as $\mathcal{N}(0, 10^2)$ (a weakly informative 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 $Z = (\boldsymbol{B}_i, \boldsymbol{S}_i)$ and let $\Theta = (\boldsymbol{\beta}_q, \Sigma_q, \xi, \nu_q, \psi_q, a'_{kl}, b'_{kl}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W, \sigma)$, where these sets denote the unknown parameters 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 unknown parameters is then

$$
\begin{aligned}
P(\Theta, Z, \mathcal{D}) \propto \prod_{i=1}^{n}\prod_{q=1}^{Q}\Big\{ & \frac{1}{\sqrt{(2\pi)^q|\Sigma|}}\exp(-\frac{1}{2}(\boldsymbol{b}_{iq} - \boldsymbol{\beta}_q)^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{b}_{iq} - \boldsymbol{\beta}_q)) \\
& \times \frac{1}{\sqrt{2\pi\xi_q^2}}\exp\left[\frac{(\log(d_{iq}) - \nu_q)^2}{2\xi_q^2}\right]\left(\frac{[(r_{ikl}+1)/2]^{a'_{kl}-1}\{1 - ([(r_{ikl}+1)/2]\}^{b'_{kl}-1}}{\mathsf{Beta}(a'_{kl}, b'_{kl})}\right) \\
& \times \prod_{j=1}^{n_i}\frac{1}{\sqrt{(2\pi)|\mathbf{S}_i|}}\exp\left(-\frac{1}{2}\{\boldsymbol{X}_{ij} - \boldsymbol{\mu}(t_{ij}; \mathbf{B}_i)\}^{\mathsf{T}}\mathbf{S}_i^{-1}\{\boldsymbol{X}_{ij} - \boldsymbol{\mu}(t_{ij}; \mathbf{B}_i)\}\right)\Big\} \\
& \times \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{(Y_i - \eta(\mathbf{B}_i, \mathbf{S}_i, \boldsymbol{W}_i; \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W))^2}{2\sigma^2}\right] \times \pi(\Theta). \qquad (2.11)
\end{aligned}
$$

Figure S1 in the Supplementary Materials uses a directed acyclic graph to visualize and summarize the hierarchical relationships between the different components of our modeling framework.

2.3.4 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 2.4.1, 2.4.2 and in B.5, we run two chains per independent replicate data set, with $2,000$ iterations and $1,000$ burn-in. For the data application in Section 2.5, we ran 4 chains each for 2,000 iterations with 1,000 burn-in. Visual inspection of the traceplots for all model parameters indicated non-divergent chains (see Figures B.4 and B.5 in Appendix B.2.1). 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. The R-hat values of these two 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 Appendix B.2) suggest that our model generates reasonable estimates for the observed outcomes and trajectories.

2.4 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 $\frac{1}{R}\sum_{r=1}^{R}(\widehat{\theta}^{(r)}-\theta_0)$ 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 $\frac{1}{R}\sum_{r=1}^{R}\mathbb{1}\{\theta_0 \in I_r\}$ 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 $\frac{1}{R}\sum_{r=1}^{R}T_r$, 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.

2.4.1 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 2.5. 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 2.3 for the outcome submodel parameters $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$. See Table B.3 in Appendix B for the results for the other model parameters.

2.4.1.1 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: $\mathbb{E}(Y_i \mid \mathsf{others}) = \sum_{q=1,2} \left\{ \alpha_{q1} \widehat{\beta}_{iq0} + \alpha_{q2} \widehat{\beta}_{iq1} + \gamma_{qq} \widehat{s}_{qq} \right\} + \sum_{q' < q} \gamma_{q'q} \widehat{s}_{q'q}$. 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 (2.1) and (2.2) only (together with their prior specifications in Equation (2.5), (2.7) to (2.8)) and use the estimates of the posterior means, $\widehat{\mathbf{B}}_i$ and $\widehat{\mathbf{S}}_i$ in the model 2.3 (together with prior specifications in Section 2.3.3.2). 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.

2.4.1.2 Simulation I: Results

Table 2.3 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 (see Appendix A. 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 cap-

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	−001	93.5	0.38
	TSIV	−002	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	−001	93.0	0.29
α_{21} = -3	**JMIV**	0.00	97.0	0.25
	TSLM	0.46	0.0	0.29
	TSLMM	−001	92.5	0.40
	TSIV	−001	89.0	0.32
α_{22} = 3	**JMIV**	−002	93.5	0.37
	TSLM	−017	71.5	0.36
	TSLMM	−001	97.0	0.49
	TSIV	0.01	95.0	0.43
γ_{11} = 2	**JMIV**	0.01	94.0	0.52
	TSLM	−038	17.0	0.36
	TSLMM	−038	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	−041	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	−001	88.5	0.40

Table 2.3: **Simulation I: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates. We compare 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 2.4.1.1 for details about the alternative methods.**

turing 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 also 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 Table B.3 in Appendix B.4). 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.

2.4.2 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 $\boldsymbol{B}_{i1} \sim \mathcal{N}_2\left(\boldsymbol{\beta}_1, \Sigma_1\right), \boldsymbol{B}_{i2} \sim \mathcal{N}_2\left(\boldsymbol{\beta}_2, \Sigma_2\right), \boldsymbol{B}_{i3} \sim \mathcal{N}_2\left(\boldsymbol{\beta}_3, \Sigma_3\right)$, 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 presented in Equations (2.3.3) and (2.8) 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 2.4 and 2.5. 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

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 2.4: **Simulation II: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates for the α parameters. We compare our 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).**

are present in the γ 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.

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

2.5.1 Study of Women's Health Across the Nation

We examine the joint association of E2 and FSH on rate of change in fat mass and waist circumference as women transition from the premenopause to the postmenopause using data from the Study of Women's Health Across the Nation (SWAN). Previous research with SWAN data has

Truth	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 2.5: **Simulation II: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates for the γ parameters. We compare our 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).**

demonstrated that fat mass increases and lean mass decreases in a non-linear fashion across the menopausal transition (Greendale et al., 2019). The rate of change in body fat mass and lean mass accelerates approximately two years prior to the FMP and persists until approximately 2 years after the FMP (Greendale 2013). Body weight and BMI, however, have a consistent positive linear relationship throughout the menopause transition, suggesting no unique menopausal effect on body weight or BMI despite the changes to a more adverse body composition profile (i.e., more fat mass and less lean mass). Given that increases in weight and fat mass in midlife contribute to women's risk of chronic disease, improving understanding of the physiologic mechanisms that underlie these increases is important. Yet, the association of mean body size parameters with both mean E2 and mean FSH is complex, especially late in the menopausal transition and into the postmenopause, because fat is a significant source of estrogen and a known negative feedback regulator of FSH in the hypothalamus and pituitary. The role of fat in moderating a woman's endocrine profile may help explain why, in women who are obese, E2 is lower prior to menopause and higher postmenopause while mean FSH is much lower, compared to women who are not obese (Ran-

dolph et al., 2011). Since not all fat is metabolically equal and functionally varies by anatomic distribution, a model that can evaluate the contribution of individual variability in hormones would advance understanding of the complex relationship between body composition and reproductive hormones.

In the predictor submodels, we specified a linear mean trend consisting of an individual intercept and slope. We explored a higher order quadratic form for the mean trend, but found that the quadratic terms (e.g. E2 Intercept^2) did not significantly predict either outcome of interest, likely due to the sample size of our datasets, as well as the limited individual-level information that remained after detrending the hormone population trends. For this reason, we focused on a linear specification for the hormone markers mean trajectories. In the outcome submodels, 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.

In the outcome regression model, we adjusted for the following covariates: fat mass body weight proportion (or waist circumference) 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). Tables 2.1 and 2.2 display the descriptive statistics for the individuals in our two analyses, including demographic and physical activity information.

2.5.2 Fat Mass Rate of Change

Table 2.6 displays the results of the fat mass model. For ease of interpretation, the coefficients relating to the individual means (E2, FSH intercepts and E2, FSH slopes) and the individual variances (E2, FSH variances and E2, FSH correlation) have been multiplied by their respective sample standard deviation estimates. Table S3 in the Supplementary Material displays these SD estimates.

We found that the E2 intercept and the E2 and FSH slopes were all significantly associated with fat mass rate of change. A one standard deviation higher E2 intercept (compared to the population mean) was was associated with an average 0.11% increase in fat mass proportion per year. Since E2 tends to decline over the menopausal transition, we can interpret the E2 coefficient as follows: one standard deviation lower E2 slope than the population average was associated with a mean decrease of 0.09% in fat mass proportion per year. Conversely, higher increases in FSH

Variable	Post. Mean	2.5% CrI	97.5% CrI
E2 Intercept	**0.11**	**0.05**	**0.17**
FSH Intercept	0.02	-0.01	0.06
E2 Slope	**0.09**	**0.02**	**0.17**
FSH Slope	**-0.06**	**-0.11**	**-0.02**
E2 Var.	-0.03	-0.08	0.02
E2, FSH Cor.	0.01	-0.05	0.07
FSH Var.	-0.03	-0.08	0.01
Fat Mass Prop. (First Visit)	**-2.97**	**-3.44**	**-3.51**
Black	**-0.08**	**-0.16**	**-0.01**
Chinese	0.01	-0.08	0.11
Japanese	**-0.32**	**-0.41**	**-0.24**
Increasing Activity (Cat. 2)	0.02	-0.07	0.11
Decreasing Activity (Cat. 3)	0.07	-0.01	0.15
Middle Activity (Cat. 4)	0.01	-0.06	0.09
Highest Activity (Cat. 5)	-0.09	-0.17	0.00

Table 2.6: **Estimated posterior means and 95% credible intervals for the fat mass rate of change model. The variables related to E2 and FSH (intercepts, slopes, and variances/correlation) have been standardized by the sample standard deviation of their posterior estimates. The presented values have been multiplied by 10^2.**

were negatively associated with fat mass rate of change, with a one standard deviation increase in FSH (compared to the population average) being associated with -0.06% increase in fat mass proportion per year.

We adjusted for other covariates in the outcome model. Fat mass proportion at first visit was negatively associated with fat mass rate of change (-2.97% average decline per year). Black and Japanese women also had slower fat mass gains compared to white women on average (-0.08% increase per year and -0.32% per year, respectively).

2.5.3 Waist Circumference 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

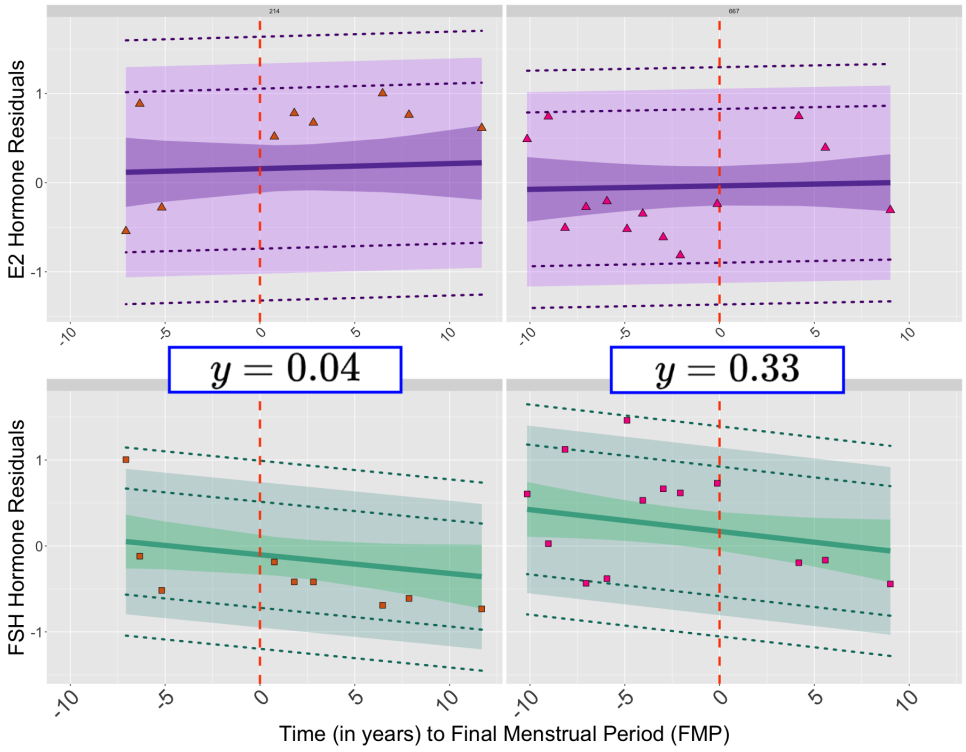


Figure 2.3: **Plots of estimated hormone trajectories for two individuals from the waist circumference rate of change model. The solid lines are the estimated individual mean trajectories, based on the posterior means of B_i, i.e. $\widehat{b}_{i0} + \widehat{b}_{i1}t$. The darker inner intervals around 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 waist circumference rates of change are shown in bordered boxes.**

Variable	Post. Mean	2.5% CrI	97.5% CrI
E2 Intercept	**0.19**	**0.11**	**0.29**
FSH Intercept	**-0.07**	**-0.13**	**-0.19**
E2 Slope	**0.26**	**0.16**	**0.37**
FSH Slope	**-0.17**	**-0.11**	**-4.35**
E2 Var.	**-0.11**	**-0.13**	**-0.03**
E2, FSH Cor.	-0.03	-0.12	0.06
FSH Var.	-0.05	-0.82	0.11
Waist Circum. (First visit)	**-0.02**	**-0.02**	**-0.01**
Black	-0.08	-0.20	0.03
Chinese	**-0.30**	**-0.45**	**-0.16**
Japanese	**-0.24**	**-0.38**	**-0.10**
Hispanic	-0.06	-0.33	0.21
Increasing Activity (Cat. 2)	0.00	-0.14	0.15
Decreasing Activity (Cat. 3)	-0.04	-0.16	0.08
Middle Activity (Cat. 4)	-0.09	-0.21	0.04
Highest Activity (Cat. 5)	-0.13	-0.27	0.02

Table 2.7: **Estimated posterior means and 95% credible intervals for the waist circumference rate of change model. The variables related to E2 and FSH (intercepts, slopes, and variances/correlation) have been standardized by the sample standard deviation of their posterior estimates.**

additional variance parameter.

Table 2.7 displays the estimated coefficients for the waist circumference model. As in the fat mass model, the coefficients related to the individual means and variances have been adjusted by their respective sample standard deviations (also found in Table S3 in the Supplementary Material). A one-unit higher E2 intercept (above the population mean) was associated with an average 0.19 faster increase in waist circumference (cm/year). A one standard deviation lower E2 slope was also associated with slower declines in waist circumference per year (-0.26 cm/year). A higher individual FSH intercept was negatively associated with waist circumference rate of change; a one-unit higher starting FSH was associated with an average -0.07 cm/year decrease in waist circumference.

E2 variability was negatively associated with the outcome, meaning that women with a one standard deviation higher E2 variability had, on average, -0.11 cm/year decrease in waist circumference. Neither FSH variability nor E2, FSH correlation were significantly associated with changes in waist circumference.

Unsurprisingly, waist circumference at first visit was negatively associated with waist circumference change, meaning that women with a higher starting waist circumference tended to have slower increases in waist circumference (-0.02 cm/year increase). Chinese and Japanese women also had slower increases compared to white women on average (-0.30 cm/year increase and -0.24 cm/year increase, respectively).

2.6 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. Our analysis of hormone trajectories data revealed E2 variability had a statistically significant association with waist circumference change, but not overall 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 modeling is important for obtaining accurate estimates (in terms of low bias, higher coverage and shorter interval lengths) of how the individual-level parameters are linked to the outcome. Simulation results demonstrate that our model outperforms common two-stage approaches.

Substantively, our analyses are in line with the established literature on the associations between average hormone levels and fat mass and distribution changes during menopause. As noted above, the association of mean body mass with both mean E2 and mean FSH is complex, especially as women transition into the postmenopause, because adipose tissue is a significant source of estrogen and a known negative feedback regulator of FSH in the hypothalamus and pituitary. The known E2 results are echoed by our analyses, which showed a 1) positive relationship between increasing E2 and fat mass gains and 2) a positive relationship between increasing E2 and waist circumference gains and 3) a negative relationship between increasing FSH and both outcomes of interest. However, evidence suggests that increased FSH itself may directly influence adiposity by reductions in energy expenditure after menopause (Sponton and Kajimura, 2017; Kohrt and Wierman, 2017; Liu et al., 2017). Thus, our findings of a negative relationship between increasing FSH and both outcomes of interest do not support this recent work. This may be due to the complex and complicated relationship of concurrent E2 changes during the menopausal transition.

However, the associations between individual-level variability and co-variability of two hormones (E2 and FSH) and changes in fat mass and waist circumference, a surrogate for fat distribution, had not been well explored. Previous analyses have either only evaluated mean associations of E2 and FSH on health outcomes, or separately analyzed the hormone's variability. Early models of hormone mean trajectories were far too crude to assess individual variability or associations with fat distribution. Our analyses revealed that individual E2 variability was highly predictive of waist circumference changes, but not overall fat mass changes. Since not all fat is metabolically equal and functionally varies by anatomic distribution, this finding could indicate that changes in fat distribution, in particular waist adiposity, are more driven by E2 hormonal variability during

menopause, while other factors could be driving overall fat mass increases. Future analyses would be required to more fully investigate that hypothesis. Additionally, this joint analysis of E2 and FSH can serve as a basis for further investigation of how hormone variability and co-variability may affect other health outcomes. As mentioned above, joint estimation of longitudinal variables and scalar outcomes can be useful for investigating scientific questions in many areas. With longitudinal biomarker data becoming more readily available (e.g. from 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.

2.6.1 Remark

In our simulation studies and SWAN data analysis, we made the simplifying assumption to exclude covariates $\boldsymbol{W}_i$ in modeling the longitudinal markers (Equation 2.1). In mathematical terms, this means that the likelihood functions of $\boldsymbol{X}_{it}, \boldsymbol{B}_i, \boldsymbol{S}_i, \boldsymbol{W}_i$ are:

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

where f is a generic notation for the probability density function. 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 2.5 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.

2.6.2 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 slightly 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.

It may be of interest in future applications to simultaneously model multiple cross-section individual outcomes, as our outcome submodel specification only considered univariate outcomes. Exploring Bayesian semi-parametric approaches to modeling the subject specific parameters, e.g. with a Dirichlet process prior on the unknown parameters, would be another methodological extension. In addition to the increased model flexibility, this could also allow for clustering of individuals with similar mean trajectories and/or residual variances and covariances.

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.

CHAPTER 3

Chapter 3: A Joint Modeling Approach to Study the Association between Subject-level Longitudinal Marker Variabilities and Repeated Putcomes

3.1 Introduction

Bone mass starts to decline in adults during the midlife (Hunter and Sambrook, 2000). Bone mineral density (BMD) is an essential component of bone mass and a key indicator for risk of osteoporosis. People with significantly lower BMD beyond the expected age-related decline are at higher risk for bone fracture. The prevalance of osteoporosis in women over 50 tends to rapidly increase at around 60 years of age, and by age 70, this prevalence has tripled (Alswat, 2017). Bone mineral loss in women is known to accelerate during the menopausal transition (Riggs and Melton, 1992; Ji and Yu, 2015). High rates of BMD declines are associated with adverse health outcomes, such as increased risk of bone fractures (Marshall et al., 1996). Population-level patterns of BMD during the midlife and older age are well-established; however, it is equally vital to model and predict individual bone trajectories of in order to support advances in precision medicine and tailored individualized treatments.

Women undergoing menopause are particularly vulnerable to more rapid declines in BMD, likely due in part to hormones such as estradiol (E2) and follicle-stimulating hormone (FSH) undergoing large changes in overall levels (Sirola et al., 2003; Recker et al., 2000; Finkelstein et al., 2008). Previous studies have examined associations between E2, one of the most potent naturally-occuring estrogens, and BMD during the menopausal transition (Ebeling et al., 1996; Sowers et al., 2006). These studies have established a general positive association between mean E2 levels and BMD. However, most of these studies have either been cross-sectional or over a relatively short time period. Importantly, these studies have focused on mean associations between hormones and bone outcomes, mostly at the population level. The association between individual E2 trajectories, and in particular, individual level E2 variability and BMD changes in peri- and post-menopausal

women over a longer time period has not yet been studied.

Researchers have also studied the relationship between levels of FSH, another key reproductive hormone, and their corresponding associations with bone health outcomes. Park et al. (2021) showed that bone loss during menopause was significantly associated with higher FSH levels in a cross-sectional analysis of midlife women. Chin (2018) examined the longitudinal relationship between FSH and BMD in peri-menopausal women and revealed that "rate of bone loss was inversely associated with FSH level in all subjects, regardless of BMD value". However, Gourlay et al. (2011) did not find a statistically significant relationship between high baseline FSH and decreases in BMD in postmenopausal younger women. These findings suggest that FSH levels may have the strongest association with bone loss during the midlife before the menopausal transition.

These previous research findings estimated the associations between mean levels of E2 and FSH on BMD trajectories. However, these hormones can fluctuate substantially within individual women (Bjørnerem et al., 2006), and these within-individual variabilities have been shown to be significantly associated with health outcomes such as abnormal uterine bleeding and ovarian response to in vitro fertilisation treatments (Harlow et al., 2000; Uhler et al., 2005). Since bone health is an important aspect of overall wellbeing as women enter the midlife, it is important to understand how these intra-individual variabilities, as well as overall mean hormone levels, affect women's bone outcomes as they age.

3.1.1 Study of Women's Health Across the Nation Dataset

The Study of Women's Health Across the Nation (SWAN) is an ongoing multi-site longitudinal cohort study. E2 (pg/mL) and FSH (mIU/m) measurements were collected at baseline and during annual follow-up visits. Five of the seven SWAN sites also collected annual femoral neck BMD (g/cm^2) measurements. Additional socio-demographic covariates, such as age and BMI, were collected at baseline and at each additional visit. For detailed descriptions of the measurement timings and cohort characteristics, we refer the reader to Sowers et al. (2000b).

Since E2 and FSH follow a well-established population trend during the menopausal transition (Randolph et al., 2004), we fit a loess curve to each hormone using the `loess` function in R and subtracted each women's measurements from this loess fit. We performed all of the following analyses in Section 3.4 using these hormone residuals. Figure 3.1 displays the E2 residuals and the FSH residuals. In our analysis, we also lagged these residuals by 1 visit, since the associations between increases or decreases in individual hormones and BMD outcomes may manifest later in time, rather that at the concurrent visit. The corresponding interpretation of the coefficients takes this into account. We also excluded women who had received hormone replacement therapy (HRT) at any point during their clinical visits, since these medications can suppress BMD decline even

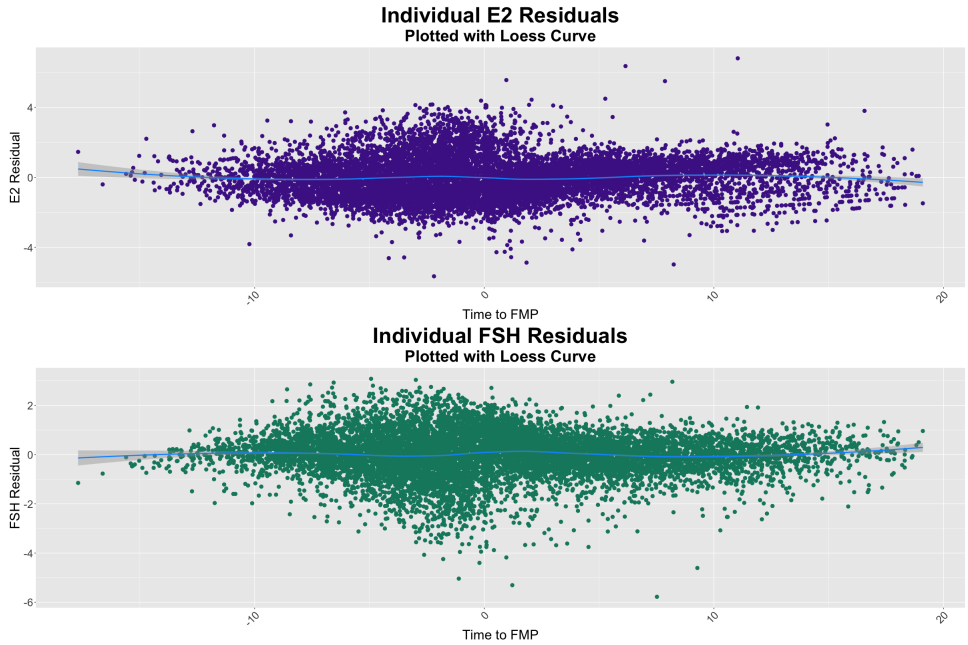


Figure 3.1: **Plot of E2 observations (top figure) and FSH observations (bottom figure) for all individuals in our dataset, plotted over time to FMP. A loess curve has been added to each plot to show the average population trend. Both E2 and FSH observations (before detrending) had been $\log_2$ transformed.**

after stopping HRT (Gambacciani and Levancini, 2014).

All of these longitudinal measurements are measured relative to time to FMP, which are recorded in year. Our aim is to estimate the association between hormone variability and bone measurement outcomes in a longitudinal setting. Additionally, we want to understand if these relationships may change over time, which we can evaluate via interaction terms in the outcome submodel. Thus, our objective is to develop a modeling framework that can estimate the individual level mean and variance of a longitudinal marker, and also relate these estimates to the longitudinal outcome of interest using a regression model while properly propagating the uncertainty in studying these associations.

We also used a base 2 log transformation on the outcome of interest (BMD), both to improve the normality assumption and for interpretability reasons, as a unit change in $\log_2$ BMD can be interpreted as a doubling or halving of BMD. We used the same approach of detrending the BMD measurements as we did in the hormone trajectories by fitting a loess curve to all measurements and then subtracting the individual measurements from the fitted loess values. Figure 3.2 displays the residual BMD values after our detrending, in order to simplify subject level trends as lower-level polynomial functions; in practice a linear approximation appeared sufficient after population-level detrending.

We adjusted for baseline BMI and baseline age in the outcome model. We standardized baseline

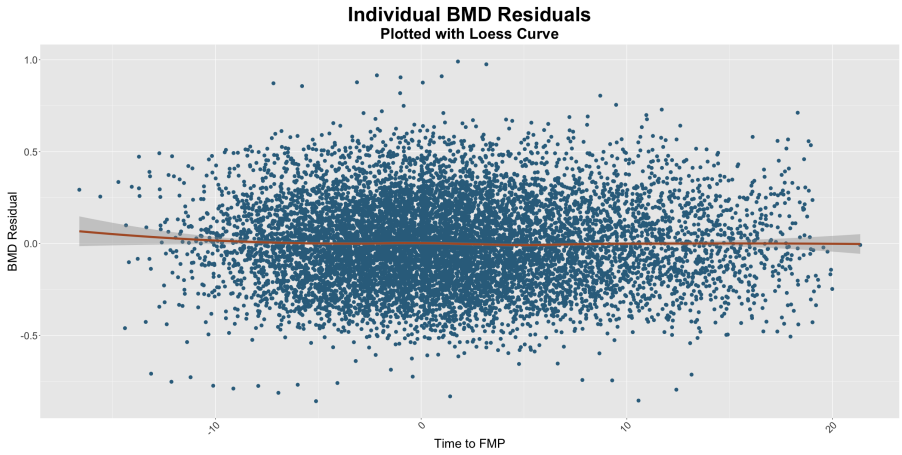


Figure 3.2: **Plot of BMD residuals for all individuals in our dataset, plotted over time to FMP. A loess curve has been added to show the average population trend of the residuals. Prior to detrending, the BMD observations had been $\log_2$ transformed.**

Variable	Statistic	Value	n
Longitudinal Predictor	Mean/SD		
E2 Residuals		-0.01 (1.15)	9,858
FSH Residuals		0.02 (0.90)	9,858
Health Outcome	Mean/SD		
BMD Residuals		0.00 (0.23)	9,858
Adjusted Covariates	Mean/SD		
Starting BMI		27.28 (6.83)	9,858
Starting Age		46.29 (2.60)	9,858

Table 3.1: **Descriptive statistics of the SWAN dataset with BMD measurements, based on 986 individuals.**

BMI measurements to be relative to the population mean (baseline) BMI. We also centered the baseline age measurements relative to the population mean (baseline) age. Our final BMD outcome dataset comprised 986 women with 9,858 measurements. See Table 3.1 for summary statistics of all variables used in our data analyses.

3.1.2 Statistical Models for Longitudinal Outcomes

Models with longitudinal outcomes allow researchers to understand relationships between variables across time, and potentially how these associations change across time. These methods tend to fall into two broad classes: generalized estimating equations (Liang and Zeger, 1986) methods and mixed-effects models or structural equation modeling (SEM)-type growth curve models. While the GEE approach has several advantages, such as robustness in the case of misspecified marginal correlations, it is not designed for studying individual-level random effects. In our par-

ticular setting, we are primarily interested in relating the individual-level means and variances to the health outcome, and so our modeling approach needs to explicitly estimate the individual random effects. Mixed-effects models have been studied extensively (Laird and Ware, 1982; Greene, 2005; Diggle et al., 2013) and are well-suited for estimating multilevel or hierarchical data. These models can also easily handle interactions with time, which can be specified as an additional covariate. A drawback of the mixed effects model is that the predictor variable is treated as fixed data, rather than being estimated as part of the model. Latent growth curve (LGC) models on the other hand provide a framework for estimating two or more trajectories simultaneously. In the latent growth curve specification, the random effects are modeled as latent variables, which can be shown to be mathematically equivalent to the random effects specification in mixed effects models (Zhang et al., 2022). Generally, LGC models are not well-suited for datasets with moderate to large number of observations per individuals, since the LGC framework requires each visit or time point to be a unique predictor. Mixed-effects models, on the other hand, allow the time variable to be univariate and can more easily handle many observations per individual (McNeish and Matta, 2018).

One noticeable gap in both frameworks is that the individual residual variances in the predictors are usually treated as nuisance parameters, rather than as potentially important entities for predicting the outcome. There has been a growing collection of research (Elliott et al., 2012; Jiang et al., 2015; Gao et al., 2022) on joint models that evaluate the associations between within subject variability and outcomes of interest; however, these models have focused on cross-sectional outcomes. Methods for estimating individual variances from longitudinal markers to predict longitudinal outcomes are currently lacking in the literature. Our proposed model contributes to this research by providing a general method for estimating individual-level variability from a longitudinal marker and using this variability to predict a longitudinal time-varying outcome. Our specification can easily extend to incorporating a time-varying variance parameterization, as demonstrated by simulation studies.

The rest of this paper is organized as follows. We describe our Bayesian joint model framework in Section 3.2. Section 3.3 introduces simulation studies that demonstrate that our model produces less biased estimates of the outcome regression coefficients than alternative two-stage models that do not account for the statistical uncertainty in the individual means and variances. We apply our model in Section 3.4 to the SWAN dataset, where we focus on the study of the associations between E2 and FSH mean trajectories and variances and BMD outcomes for women in the midlife. To the best of our knowledge, this is the first scientific assessment of using individual level variances of E2 and FSH to predict declines in longitudinal bone measures. Finally, in Section 3.5, we discuss the implications of our findings along with future directions. Appendix B contains further details about the model validation for the data analysis (e.g. posterior predictive checks).

3.2 Statistical Model

In this section, we describe our proposed model, which specifies our longitudinal predictor X_{ij}, which is measured at each timepoint $t_{ij}, j = 1, \ldots, n_i$, for each subject $i = 1, \ldots, N$ where n_i is the total number of observations for individual i. Our model links the individual-specific vectors of regression coefficients and residual variance estimates of X_{ij} to the longitudinal outcome Y_{ij}.

3.2.1 Likelihood

3.2.1.1 Longitudinal Marker Model

The model for the longitudinal predictor is given by:

$$X_{ij} \mid \boldsymbol{b}_i, s_i \sim \mathcal{N}\left(\mu(t_{ij}; \boldsymbol{b}_i), s_i^2\right), j = 1, \ldots, n_i, \text{ independently for } i = 1, \ldots, N, \tag{3.1}$$

$$\boldsymbol{b}_i \overset{\text{indep.}}{\sim} \mathcal{N}_P(\boldsymbol{\alpha}, \Sigma), \text{ independently for } i = 1, \ldots, N, \tag{3.2}$$

where $\mathcal{N}(\mu, s)$ represents a Gaussian distribution with mean μ and variance parameter s and $\mathcal{N}_P(\boldsymbol{\alpha}, \Sigma)$ is the P-dimensional generalization the Gaussian distribution with mean vector $\boldsymbol{\alpha}$ and variance-covariance matrix Σ. $\mu(t; \boldsymbol{b}_i)$ is function of time and a vector of P regression coefficients, $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iP})^\mathsf{T}$. $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_P)^\mathsf{T}$ is a vector of population mean regression coefficients.

Prior for s_i^2 We assume that the individual residual variances, s_i^2's are drawn from a log-Normal distribution, i.e.:

$$\log(s_i^2) \sim \mathcal{N}(\nu, \psi^2), \tag{3.3}$$

$$\nu \sim \mathcal{N}(m, \xi^2), \psi \sim \text{half-Cauchy}(0, \tau), \tag{3.4}$$

where the values of hyperprior paramters m, ξ, τ are set a priori. In our application, we set $m = 0$, $\xi = 10$, $\tau = 2.5$, so that the priors are weakly informative. Following the recommendations made in Gelman (2006), we use the half-Cauchy hyperprior on ψ, the square root of the variance parameter. Although the inverse-Gamma distribution is the conjugate prior for the variance parameter of a Gaussian distribution, inferences using the inverse-Gamma distribution can be extremely sensitive to the choice of hyperparameter values (Gelman, 2006, p. 524). The half-Cauchy distribution avoids this potential issue due to its heavier tail, which still allows for higher variance values.

Priors for $\boldsymbol{\alpha}$: We assume that $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_P)^{\mathsf{T}}$, are drawn from a P dimensional Normal distribution:

$$\boldsymbol{\alpha} \sim \mathcal{N}_P(0, \xi_0^2 I_{P\times P}), \tag{3.5}$$

$$\Sigma = KLK,\ K = \text{diag}\{k_1, \ldots, k_P\}, \tag{3.6}$$

$$k_p \sim \textsf{half-Cauchy}(0, \tau_0), p = 1, \ldots, P,\ L \sim \textsf{LKJ}(\zeta), \tag{3.7}$$

where $K = \text{diag}\{k_1, \ldots, k_P\}$ is a diagonal matrix and L is a correlation matrix, with a Lewandowski-Kurowicka-Joe (LKJ) diffuse prior (Lewandowski et al., 2009b). As in the case of the hyperpriors for the individual variance parameters, the values of ξ_0, τ_0, ζ parameters are set a priori, here as $\xi = 1, \tau_0 = 2.5, \zeta = 1$.

3.2.1.2 Longitudinal Outcome Model

The outcome variable, Y_{ij}, is related to individual-specific mean and variance parameters $\boldsymbol{b}_i$ and s_i (Equation 3.1) via the following specification:

$$Y_{ij} \mid \boldsymbol{b}_i, \boldsymbol{a}_i, t_{ij}, s_i \sim \mathcal{N}\left(\eta_{ij}, \sigma^2\right), i = 1, \ldots, N, \tag{3.8}$$

where $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{id})^{\mathsf{T}}$ ia vector of d covariates that can either be baseline covariates (e.g., race/ethnicity, age at the first visit) or also time-varying (e.g. body weight measured at each time-point), and $\eta_{ij} = \eta_{ij}(\boldsymbol{b}_i, s_i, t_{ij}, \boldsymbol{W}_i; \boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{\beta}^W)$ is a regression predictor parameterized by $\boldsymbol{\beta}$ (for $\boldsymbol{b}_i$ and s_i), and β^W (for $\boldsymbol{W}_i$). We let $\boldsymbol{a}_i$ be a vector of random effects to capture the correlation between the longitudinal outcome measurements. The dimensions of $\boldsymbol{\beta}$ in general will depend on the number of individual mean regression coefficients, $\boldsymbol{b}_i$ and the number of desired interaction variables. For example, if $P = 2$ in the longitudinal marker model and we have a random intercept in the outcome model, then η_{ij} can be written as:

$$\eta_{ij}(\boldsymbol{b}_i, \boldsymbol{a}_i, t_{ij}, s_i) = \beta_0 + \beta_1 b_{i0} + \beta_2 b_{i1} + \beta_3 s_i^2 + \beta_4 t_{ij} \beta_5 b_{i0} t_{ij} + \beta_6 b_{i1} t_{ij} + \beta_7 s_i^2 t_{ij} + a_i. \tag{3.9}$$

Alternative Specification In Equation (3.9), we used the mean coefficients $\boldsymbol{b}_i$ directly in the mean function for the outcome variable. An alternative specification of the mean that a) eases interpretation and b) avoids contamination from the time-invariant parameters $\boldsymbol{b}_i$, can be written as:

$$Y_{ij} \mid \boldsymbol{b}_i, \boldsymbol{a}_i, t_{ij}, s_i, \boldsymbol{W}_i \sim \mathcal{N}\left(\eta_{ij}(\mu_{ij}, t_{ij}, s_i, \boldsymbol{W}_i), \sigma^2\right), i = 1, \ldots, N, \tag{3.10}$$

where μ_{ij} is the mean from the marker model. This specification allows us to evaluate the direct relationship between the estimate μ_{ij} and the outcome Y_{ij}, rather than linking X_{ij} and Y_{ij} via the mean coefficients $\boldsymbol{b}_i$. This ease of interpretation may be preferred in scientific applications. The individual residual marker variances, s_i^2, could also be specified similarly if we allow these to vary over time (see Simulation study 3.3.2).

In our application, we focus on simple specifications of $\eta_{ij}(\cdot)$, e.g., linear models with two-way interactions in order to maintain interpretability of the coefficients. However, extensions with more complex interactions or non-linear mean structures are also possible, with the caveat that the interpretability of the coefficients may be challenging (or near impossible) with higher-order terms.

Priors for $\boldsymbol{\beta}$, $\boldsymbol{a}_i$, σ For the outcome model, we use independent $\mathcal{N}(0, 10^2)$ priors for each element of the outcome regression parameters ($\boldsymbol{\beta}$, $\boldsymbol{\beta}^W$), and a diffuse prior on the outcome residual standard deviation parameter $\sigma \sim \text{half-Cauchy}(0, 2.5)$, as recommended by Carpenter et al. (2017). For $\boldsymbol{a}_i$, the random effects, we place a multivariate Gaussian prior with mean zero and precision τ_a, i.e., $\boldsymbol{a}_i \sim MVN(0, \tau_a)$. In the case of a random intercept a_i, τ_a can be drawn from a half-Cauchy distribution or, in the case of a vector-valued $\boldsymbol{a}_i$, τ_a is a covariance matrix, whose values can be drawn from the prior described in Equation 3.7.

Joint Distribution Let $D = (Y_{ij}, X_{ij}, t_{ij}, \boldsymbol{W}_i)$ denote the observed data, Z = ($\boldsymbol{b}_i$, s_i, $\boldsymbol{a}_i$) denote the subject-level latent variables, and $\Theta = (\boldsymbol{\alpha}, \Sigma, \xi, \nu_q, \psi, \boldsymbol{\beta}, \boldsymbol{\beta}^W, \tau_a, \sigma)$ denote the model parameters. We also let $\pi(\Theta)$ denote the prior distribution of the parameters in Θ:

$$\pi(\Theta) = \pi(\boldsymbol{\alpha})\pi(\Sigma)\pi(\xi)\pi(\nu, \psi)\pi(\boldsymbol{\beta}, \boldsymbol{\beta}^W)\pi(\sigma).$$

We can then write the joint distribution of D, Z, and Θ as

$$
\begin{aligned}
P(\Theta, D, Z) \propto \prod_{i=1}^{N} \Big\{ & \frac{1}{\sqrt{(2\pi)|\Sigma|}} \exp(-\frac{1}{2}(\boldsymbol{b}_i - \boldsymbol{\alpha})^{\mathsf{T}} \Sigma^{-1} (\boldsymbol{b}_i - \boldsymbol{\alpha})) \\
& \times \frac{1}{\sqrt{2\pi\xi^2}} \exp\left[\frac{(\log(s_i^2) - \nu)^2}{2\xi^2} \right] \qquad (3.11) \\
& \times \frac{1}{\sqrt{2\pi|\tau_a|}} \exp(-\frac{1}{2} \boldsymbol{a}_i^{\mathsf{T}} \tau_a^{-1} \boldsymbol{a}_i) \Big\} \\
& \times \prod_{i=1}^{N} \prod_{j=1}^{n_i} \Big\{ \frac{1}{\sqrt{2\pi s_i^2}} \exp\left(-\frac{1}{2} \left\{ \frac{X_{ij} - \mu_{ij}(t_{ij}; \boldsymbol{b}_i)}{s_i} \right\}^2 \right) \\
& \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(y_{ij} - \eta_{ij}(\boldsymbol{b}_i, s_i, \boldsymbol{W}_i; \boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{\beta}^W)^2}{\sigma^2} \right) \Big\} \times \pi(\Theta). \qquad (3.12)
\end{aligned}
$$

3.2.2 Posterior Inference

We implemented our joint model using Stan and the `rstan` package (Stan Development Team, 2020) and obtained posterior estimates via the Hamiltonian Monte Carlo sampler. For our two simulations studies in Section 3.3, we ran two chains per independent replicate data set, with $2,000$ iterations with $1,000$ burn-in. For the E2 predictor model in Section 3.4, we ran 2 chains each with 4,000 iterations and 2,000 burn in. For the FSH model, we found that running 2 chains with 2,000 iterations with 1,000 burn in was sufficient for the model to achieve convergence. We conducted visual inspection of the traceplots for all model parameters which indicated non-divergent chains. All of the chains in each model were combined for computing posterior summaries.

We also used Stan's R-hat convergence diagnostic (Vehtari et al., 2021) to evaluate model convergence. The outcome parameters in both of our models had R-hat values < 1.02. Additionally, model checks of the posterior predictive distribution in Appendix C.1 indicated that our models generated reasonable predictions for our datasets.

3.3 Simulation Studies

The goal of our simulation studies was 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 longitudinal outcomes. For each proposed method and each parameter θ, we assessed 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

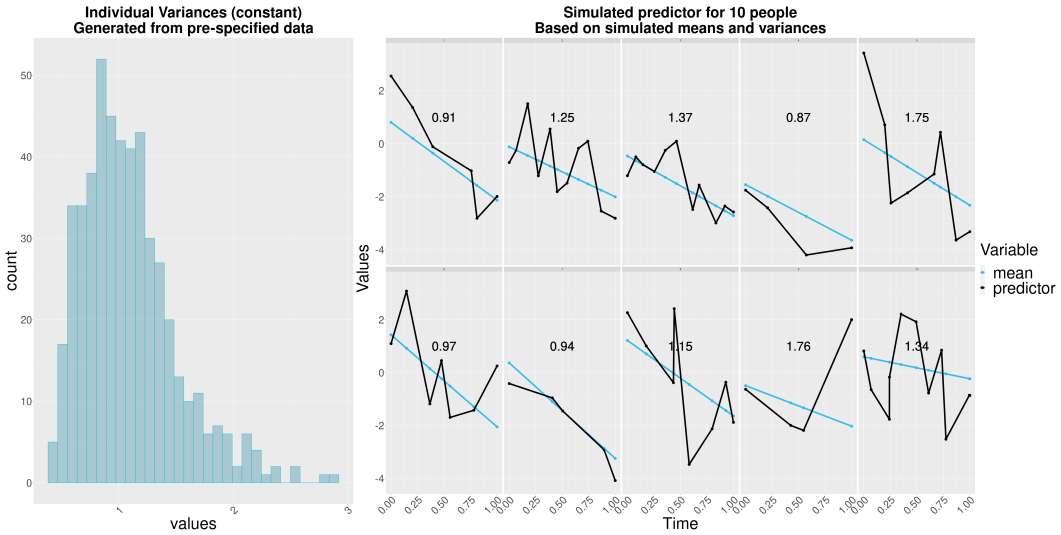


Figure 3.3: **Histogram of individual variances (left plot) and individual mean and marker trajectories (line plots) for 10 individuals, along with the generated individual variance labelled in each plot.**

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.

3.3.1 Simulation Study 1: Comparisons to two-stage approaches

For this simulation study, we generated data for N = 300 individuals. We simulate for each individual between 2 to 15 timepoints, which mimics the our application dataset from the SWAN study. Based on these individual timepoints, we simulate the marker values for each individual using the following data generating parameters:

$$
\begin{aligned}
&X_{ij} \sim \mathcal{N}(\boldsymbol{\mu}_{ij}, s_i^2), \boldsymbol{\mu}_{ij} = b_{i1} + b_{i2}t_{ij}, \\
&\boldsymbol{b}_i \sim \mathcal{N}_2\left(\boldsymbol{\alpha}, \Sigma\right), \boldsymbol{\alpha} = (0, 2)^{\mathsf{T}}, \\
&\Sigma = \begin{pmatrix} 1 & -0.05 \\ -0.05 & 1 \end{pmatrix}, \log(s_i^2) \sim \mathcal{N}(0, 0.375).
\end{aligned}
$$

Figure 3.3 displays the simulated marker means, marker trajectories, and constant variances that are generated by these specified parameters.

For the longitudinal outcome, we assumed the following model: $Y_{ij} \sim \mathcal{N}(\eta_{ij}(\boldsymbol{b}_i, s_i, a_i, t_{ij}), \sigma^2)$

and set

$$\eta_{ij}(\boldsymbol{b}_i, s_i) = \beta_0 + \beta_1\mu_{ij} + \beta_2 s_i^2 + (\beta_3 + \beta_4\mu_{ij} + \beta_5 s_i^2)t_{ij} + a_i,$$

where the true values of $\boldsymbol{\beta}$ are shown in Table 3.2. We specified the random intercept a_i as follows:

$$a_i \sim \mathcal{N}(0, 0.5).$$

Lastly, we set $\sigma^2 = 0.1$. In this simulation, we did not adjust for other covariates $\boldsymbol{W}_i$ in either submodel. We present the results for 200 replicates in Table 3.2 for the outcome submodel parameters $\boldsymbol{\beta}$.

3.3.1.1 Alternative Methods

We compared our approach to two alternative two-stage methods: a two-stage linear mixed model (TSLMM), and a two-stage Bayesian model with longitudinal outcome (TSLO). The TSLO approach is essentially the two-stage version of our joint model, using the posterior mean estimates from the means and variances of the longitudinal marker to predict the outcome in the second stage. We refer to our joint model as the Joint Estimation of Longitudinal Outcomes (JELO). In the absence of a joint model, these approaches would be reasonable methods for scientific researchers who wish to analyze the associations between a longitudinal predictor and longitudinal outcome. However, as shown in previous literature, two-stage methods often do not correctly preserve the uncertainty associated with estimating the individual random effects from the predictor marker variable (Hickey et al., 2016).

Two-Stage Linear Mixed Models (TSLMM) In the first stage, we fit a linear mixed model with the **nlme** package (Pinheiro et al., 2022)

$$X_{ij} = \beta_0 + b_{i0} + \beta_1 t_{ij} + b_{i1}t_{ij} + \epsilon_{ij}.$$

We used an unstructured variance-covariance structure for the random effects, which is the default specification for this package.

We obtained the predicted values of x_{ij} using the `predict()` function. We estimated s_i by computing the model residuals (e.g. $X_{ij} - (\widehat{B}_{i0} + \widehat{B}_{i1}t_{ij})$), where these "$B_i$" coefficients are defined as $\widehat{B}_{i0} = \widehat{\beta}_0 + \widehat{b}_{i1}$ and $\widehat{B}_{i1} = \widehat{\beta}_1 + \widehat{b}_{iq}$, where $\widehat{\beta}_{pi}$ and $\widehat{b}_{pi}, p = 0, 1$, and then computed the variance across all residuals.

In the second stage model, again using the **nlme** package in R, we fit another linear mixed

Table 3.2: **Simulation I: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates.**

Truth	Model	Bias	Coverage (%)	Average Interval Length
β_0 = 2	JELO	0.00	95.5	0.28
	TSLMM	-1.63	53.5	0.44
	TSLO	0.05	0.0	0.43
β_1 = -0.1	JELO	0.00	95.5	0.06
	TSLMM	-0.09	2.0	0.09
	TSLO	0.00	44.0	0.07
β_2 = -1	JELO	0.00	95.0	0.22
	TSLMM	0.43	57.0	0.20
	TSLO	-0.02	43.5	0.21
β_3 = -0.75	JELO	0.00	94.5	0.26
	TSLMM	0.58	0.0	0.06
	TSLO	0.00	72.5	0.38
β_4 = -0.5	JELO	0.00	96.5	0.04
	TSLMM	-0.04	0.0	0.08
	TSLO	-0.003	43.5	0.03
β_5 = 0.2	JELO	0.00	94.0	0.26
	TSLMM	0.01	0.0	0.07
	TSLO	0.00	52.5	0.15

model with the following specification:

$$Y_{ij} = \beta_0 + \beta_1 \widehat{x}_{ij} + \beta_2 t_{ij} + \beta_3 \widehat{S}_i + \beta_4 \widehat{x}_{ij} t_{ij} + \beta_5 \widehat{S}_i t_{ij} + \epsilon_{ij}.$$

Two-Stage Individual Variances (TSLO) Model We used Equations (3.1) and (3.2), and the prior specifications in Equation (3.4) to (3.7) to fit the longitudinal predictor model. We then collected the posterior mean estimates of $\widehat{X}_{ij}$ and $\widehat{s}_i$, and used these in the outcome model (Equation 3.8), along with the prior specifications in 3.2.1.2).

3.3.1.2 Simulation I: Results

Table 3.2 presents the results of Simulation I. We see that for β_3 and β_7, the coefficients of the variance parameters in the outcome submodel, the biases from the TSLMM approach and the TSLO approach are higher than the bias from our proposed model. Additionally, the coverage of the true parameters is extremely low, with neither alternative being able to achieve $> 50\%$ coverage. This indicates that if the variability of the longitudinal predictor is indeed important for estimating the outcome, neither two-stage alternative would be able to consistently estimate this association.

In particular, we see that our model outperforms the two competitors with the regards to estimating the variance coefficients (β_3, β_5). The TSLMM has extremely low coverage, which makes

sense because this model framework does not account for individual variability. The two stage approach, TSLO, performs somewhat better, but fails to achieve $> 50\%$ coverage for either parameter. Our joint modeling framework explicitly models the individual level variances and thus appropriately carries over the uncertainty from the variances into the second submodel, which improves estimation of the parameters in the outcome regression.

3.3.2 Simulation Study 2: Comparison of constant variance and time varying variance

There were two main objectives of this study. The first was to understand how well our model could recover the data generating parameters with a time-varying individual variance component. The second objective was to compare this approach to the approach with the time-invariant individual variance. This comparison gave us more insight into the situations where not specifying the time-varying component could result in large biases or high undercoverage of the true parameters.

We evaluated two scenarios with time-varying individual variances. For each simulation replicate, we generated data for N = 500 individuals and gave each individual between 4 to 12 time-points.

In the first scenario, we simulated the marker values for each individual using the following parameters:

$$
\begin{aligned}
&X_{ij} \sim \mathcal{N}(\boldsymbol{\mu}_{ij}, \sigma^2_{ij}), \boldsymbol{\mu}_{ij} = b_{i1} + b_{i2} t_{ij}, \\
&\boldsymbol{b}_i \sim \mathcal{N}_2\left(\boldsymbol{\alpha}, \Sigma\right), \boldsymbol{\alpha} = (0, -2)^\mathsf{T}, \Sigma = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 0.5 \end{pmatrix}, \\
&log(\sigma^2_{ij}) = s_{0i} + s_{1i} t_{ij}, \boldsymbol{s}_i \sim \mathcal{N}_2(\boldsymbol{\alpha}_s, \Sigma_s), \\
&\boldsymbol{\alpha}_s = (-1, 0.5)^\mathsf{T}, \Sigma_s = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 0.5 \end{pmatrix},
\end{aligned}
$$

so that the individual intercepts and slopes for the variance trends are larger in magnitude. Figure 3.4 displays histograms of individual intercepts and slopes of the variances for one simulation replicate and also 10 individual marker trajectories, based on these simulated variances and means. We will refer to this scenario as the "high-variability" (HV) case.

In the second scenario, we keep the same $\boldsymbol{\alpha}, \Sigma$ values, but change $\boldsymbol{\alpha}_s, \Sigma_s$ to be:

$$
\boldsymbol{\alpha}_s = (0, 0)^\mathsf{T}, \Sigma_s = \begin{pmatrix} 0.5 & -0.01 \\ -0.01 & 0.05 \end{pmatrix},
$$

so that the intercepts and slopes for the individual variances are smaller in magnitude. Figure 3.5

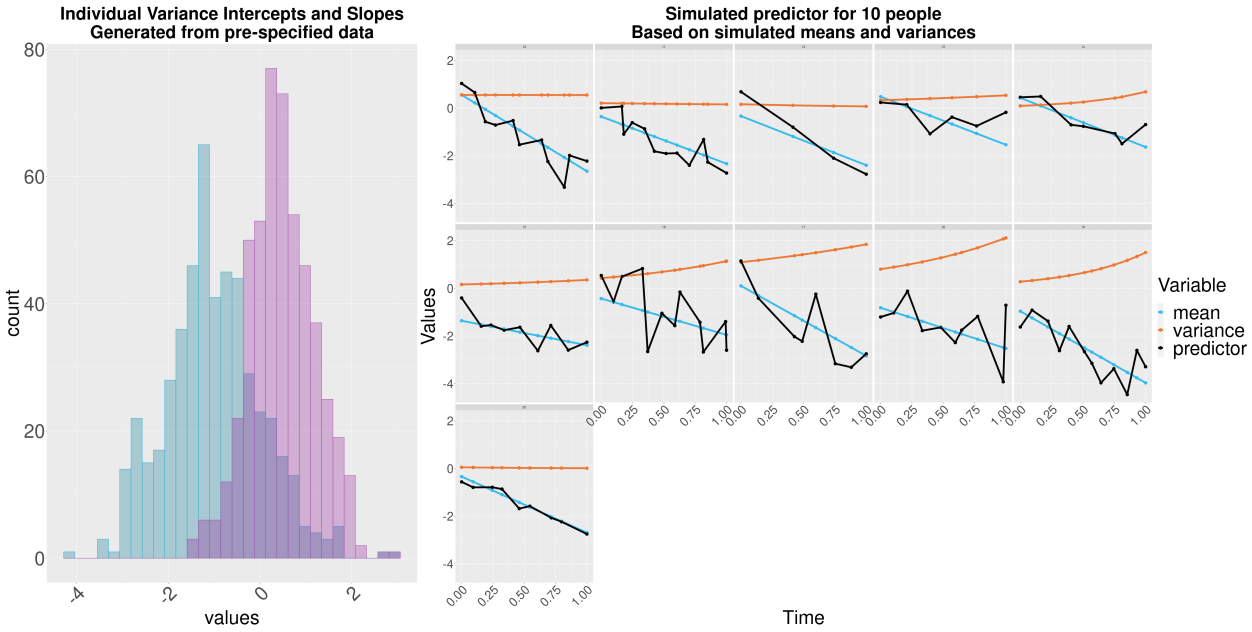


Figure 3.4: **Individual variance intercepts (left, flatter histogram) and slopes (right, peaked histogram) and individual mean, variance and marker trajectories (line plots). When the individual variability trajectory is increasing, we see that the marker trajectory has higher variability between measurements.**

displays histograms of the generated individual intercepts and slopes from one simulation replicate. We will refer to this scenario as the "low-variability" (LV) case.

To generate the longitudinal outcome in each scenario, we used the following mean specification:

$$\eta_{ij}(\boldsymbol{b}_i, \boldsymbol{s}_i) = \beta_0 + \beta_1\mu_{ij} + \beta_2\sigma^2_{ij} + (\beta_3 + \beta_4\mu_{ij} + \beta_5\sigma^2_{ij})t_{ij} + a_i,$$

where the true values of $\boldsymbol{\beta}$ are shown in Table 3.3.

3.3.2.1 Model Comparisons

Since standard methods for longitudinal outcomes (e.g. linear mixed models) do not have functionality to model a time-varying residual error variance, we compared our model JELO with a time-varying variance performance against our model with a constant variance parameterization (JELO CV). For the JELO CV simulation replicates, we used the same simulated datasets generated by the time-varying variance setup described above, but we fit the model for individual variances described in Equations (3) and (4).

From Table 3.3, it is clear that when there are large individual time-varying variances, incorrectly assuming constant individual variances lead to undercoverage of all of the outcome param-

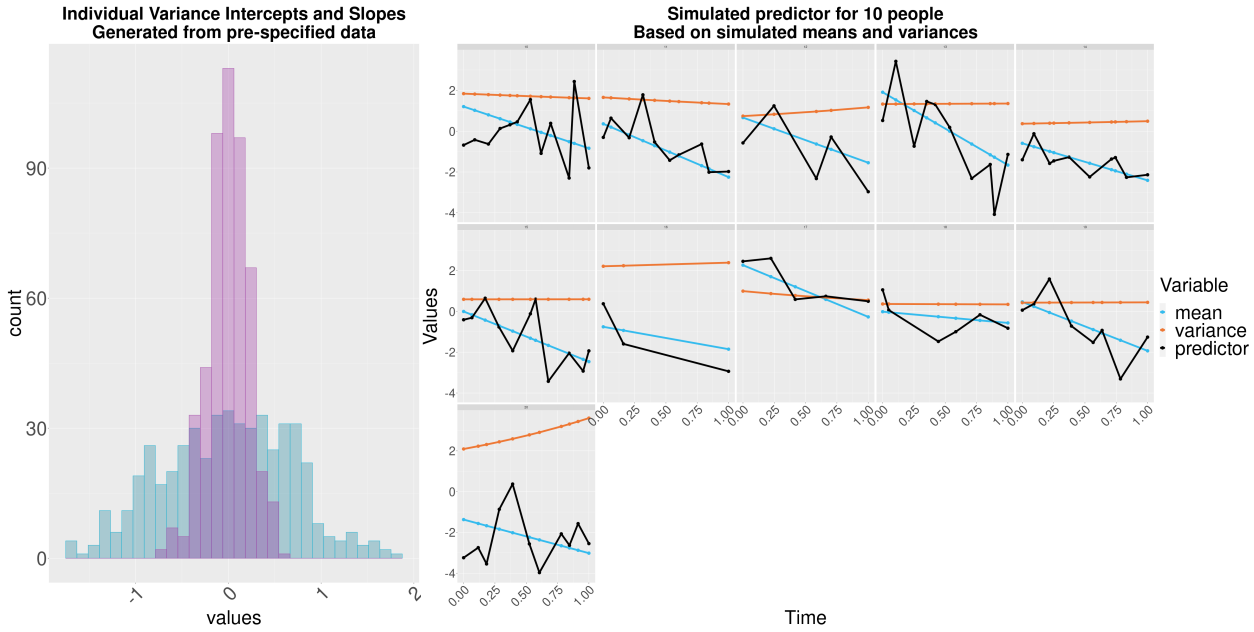


Figure 3.5: **Histograms (left plot) of individual intercepts (flatter histogram) and slopes (peaked histogram) and individual mean, variance and predictor trajectories (line plots).**

eters. Coverage of the coefficient for the individual variance β_2 and the time-variance interaction coefficient β_5 are essentially non-existent. However, if the individual variances are smaller, then the assumption of constant variances does not appear to result in undercoverage or substantially biased estimates of the true parameters. We can be reasonably confident that the constant variance assumption may suffice in scenarios where the the individual variances are time-varying, but not substantially different across time. However, if the time-varying variances are substantially changing across time within an individual, then this assumption would likely result in higher bias and substantial undercoverage of the true parameters.

3.4 Application

We now apply our proposed model to analyze the hormone and bone trajectory data described in Section 3.1.1. Our outcome model formulation was specified as follows:

$$\mathbb{E}(\log_2(BMD_{ij})) = \beta_0 + \beta_1\mu_{ij} + \beta_2 s_i^2 + (\beta_3 + \beta_4\mu_{ij} + \beta_5 s_i^2)t_{ij} + \beta_6 BMI_i^* + \beta_7 Age_i^* + a_{0i},$$

where μ_{ij} is the mean E2 (FSH) residual at time t_{ij} from the longitudinal submodel, s_i^2 is the individual-level variance, t_{ij} is the time to FMP for each individual woman, BMI_i^*, Age_i^* are the standardized BMI and baseline age values for each individual, and a_{0i} is a random intercept for

Truth	Scenario	Model	Bias	Coverage (%)	Average Interval Length
β_0 = 2	HV	JELO	0.00	96.5	0.12
	HV	JELO (CV)	0.05	68.5	0.12
	LV	JELO	0.00	95.0	0.24
	LV	JELO (CV)	0.00	93.0	0.24
β_1 = -1.5	HV	JELO	0.00	91.0	0.08
	HV	JELO (CV)	0.00	93.5	0.08
	LV	JELO	0.00	92.5	0.10
	LV	JELO (CV)	0.00	91.0	0.10
β_2 = 0.25	HV	JELO	0.00	94.5	0.15
	HV	JELO (CV)	-0.20	0.5	0.11
	LV	JELO	0.00	95.0	0.19
	LV	JELO (CV)	0.00	95.4	0.19
β_3 = 1	HV	JELO	0.00	98.5	0.18
	HV	JELO (CV)	-0.05	82.5	0.18
	LV	JELO	-0.01	97.5	0.06
	LV	JELO (CV)	0.00	97.0	0.06
β_4 =0.75	HV	JELO	0.00	95.5	0.05
	HV	JELO (CV)	0.01	82.5	0.05
	LV	JELO	0.00	95.0	0.27
	LV	JELO (CV)	-0.01	96.4	0.26
β_5 = -0.10	HV	JELO	0.00	96.0	0.12
	HV	JELO (CV)	0.29	0.0	0.10
	LV	JELO	0.00	99.0	0.06
	LV	JELO (CV)	0.01	96.4	0.13

Table 3.3: **Simulation II: bias, coverage, and 95% credible interval length across 200 simulation replicates for each data scenario.**

Table 3.4: **Estimated posterior means and 95% credible intervals for the E2-BMD model with time interactions. All estimated posterior means and 95% CrI values have been multiplied by 10^2.**

Variable	Post. Mean	95% CrI
Predicted E2	**29.46**	**(25.25,35.00)**
E2 Var.	-2.38	(-6.76, 2.05)
Time to FMP	**-0.79**	**(-1.12, -0.49)**
Time to FMP x Predicted E2	**0.31**	**(0.12, 0.58)**
Time to FMP x E2 Var.	0.21	(-0.01,0.46)
BMI	**12.21**	**(11.00, 13.41)**
Age	0.11	(-0.36, 0.57)

each woman. We also explored models with a linear random slope and quadratic random slope, but found that the estimates of the linear random slopes and the quadratic random slopes were essentially 0. Furthermore, the E2-BMD model with a quadratic random slope failed to converge. We thus concluded that a random intercept was sufficient to capture the within-individual BMD measurement correlations. The estimated variances of the random intercept in both models can be found in Tables C.1 and C.2 in Appendix C.

We ran two separate models, one with E2 measurements as the main biomarker measurement of interest and one with FSH measurements as the main predictor of interest. For the longitudinal predictor, we used the E2 (FSH) measurement obtained at the previous visit to predict BMD at the following visit. This is to better capture how differences in E2 at an earlier time may be associated with BMD declines later, rather than analyzing E2 and BMD values at the same timepoint.

The models in the following sections used a time-invariant individual variance for the predictor hormone. We also attempted to fit a time-varying individual variance component (see Simulation 3.3.2) on the SWAN dataset, but encountered severe model convergence issues. We suspected that the number of timepoints in the dataset was not large enough to capture a linear variance trend.

3.4.1 E2 Predictor Model

In this model, we included interaction between the (lagged) estimated E2 residual and time to FMP when the BMD measurement was collected, and the interaction between E2 variability and time to FMP. Table 3.4 displays the estimated posterior means and 95% credible intervals for the outcome coefficients.

The predicted E2 residual at the visit before FMP was significantly associated with BMD at FMP (i.e., when t_{ij} = 0). The interpretation of the coefficient is that for women with a 1 unit (mg/l) higher predicted E2 residual at the visit before FMP, there was an average corresponding $(2^{0.3474} - 1) \times 100\% = 22.8\%(19.1\%, 27.5\%)$ higher BMD at FMP. This effect was slightly moderated by

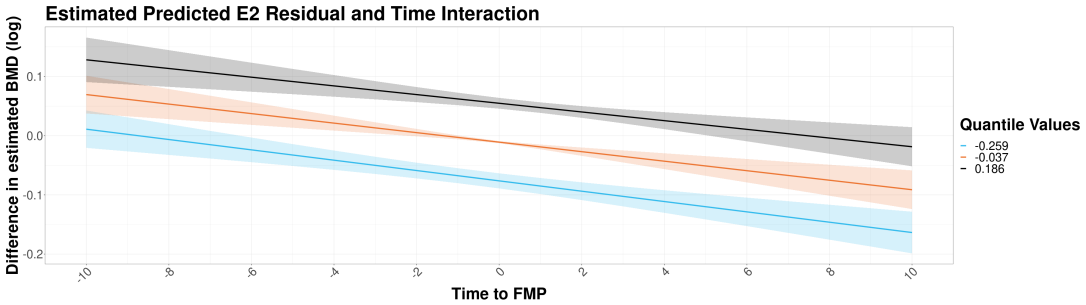


Figure 3.6: **Plots of predicted E2 residuals and time interaction for the BMD outcome model. The solid lines represent the 25th, 50th, and 75th quantile values of the E2 variable, along with the prediction band for this value.**

time, since the coefficient for the interaction of predicted E2 and time is negative. However, the credible interval for the estimated interaction effect of time and E2 contains 0, meaning that the estimated effect of E2 on BMD does not significantly change over time. This is also evident from Figure 3.6, where the estimated slopes of the BMD trajectories do not change over the MT.

When predicted E2 is at the population average, an additional one year was associated with a $(1 - 2^{-0.0079}) \times 100\% = -0.54\%(-0.77\%, -0.34\%)$ change in BMD. If we hold predicted E2 constant, then each additional year is associated with a $(1 - 2^{-0.0079} \times 2^{0.0031}) \times 100\% = -0.33\%(-0.70\%, 0.06\%)$ decrease in predicted BMD.

Higher E2 variability at FMP was negatively associated with BMD; a one unit increase in E2 variance was associated with a $(1-2^{-0.0238}) \times 100\% = -1.6\%(-4.74\%, 1.43\%)$ change in BMD. However, since the 95% CrIs contained 0, we cannot say that this relationship was statistically significant. The interaction term for E2 variance and time to FMP was also not significant.

Finally, baseline BMI was positively associated with BMD, indicating that women with higher BMI tended, on average, to have higher 8.8% BMD $(8.0\%, 9.74\%)$, holding all else constant. Baseline age, however, was not significantly associated with BMD.

3.4.2 FSH Predictor Model

Table 3.5 displays the estimated posterior means and 95% credible intervals for the FSH model outcome coefficients. Predicted FSH residual at the visit before FMP was significantly associated with BMD at FMP (when $t_{ij} = 0$). For women with a 1 unit (pg/mL) higher predicted FSH residual at the visit before FMP, there was an average corresponding $(1-2^{-0.2700}) \times 100\% = -17.1\%(-18.7\%, -15.4\%)$ lower BMD at FMP. When predicted FSH is at the population average, an additional one year was associated with a $(1-2^{-0.0028}) \times 100\% = -0.19\%(-0.33\%, -0.06\%)$ change in BMD. The FSH and time interaction was also significant and indicated that the association between mean FSH and BMD becomes amplified over time. If we hold predicted FSH

Variable	Post. Mean	95% CrI
Predicted FSH	**-27.00**	**(-29.83, -24.25)**
FSH Var.	0.80	(-4.29, 5.95)
Time to FMP	**-0.28**	**(-0.48, -0.09)**
Time to FMP x Predicted FSH	**-0.97**	**(-1.20, -0.74)**
Time to FMP x FSH Var.	**0.38**	**(0.12, 0.64)**
BMI	**9.86**	**(8.58, 11.24)**
Age	-0.22	(-0.70, 0.29)

Table 3.5: **Estimated posterior means and 95% credible intervals for the FSH-BMD model with time interactions. All estimated posterior means and 95% CrI values have been multiplied by 10^2.**

constant, then each additional year was associated with a $(1 - (2^{-0.0028} \times 2^{-0.0097}) \times 100\% = -0.86\%(-1.15\%, -0.57\%)$ change in predicted BMD residual. This can also be seen in Figure 3.7, where the estimated BMD trajectories tend to diverge sharply after FMP.

FSH variability at FMP was not significantly associated with BMD. The interaction term, however, between variability and time to FMP was significant. Holding FSH variability constant, a one year increase in time to FMP is associated with a $(1-2^{-0.0028} \times 2^{0.0038}) = 0.07\%(-0.25\%, 0.38\%)$ change in BMD. Since this interaction term has a positive sign, this indicates that the association of FSH variability with BMD is moderated over time. The bottom plot in Figure 3.7 shows this moderating effect where the estimated difference in BMD trajectories converge around 2 years before FMP. After FMP, higher FSH variability is associated with slower decreases in BMD.

Finally, as in the E2 predictor model, BMI was significantly associated with BMD. Women with higher BMI values tended to have higher BMD values, holding all else constant. Baseline age was not significantly associated with BMD.

3.5 Discussion

Maintaining bone health in women is particularly important during the midlife, when BMD naturally tends to decline (Finkelstein et al., 2008). Although much research has been conducted on the associations between mean hormone levels and BMD outcomes, there was a gap in understanding the associations between hormone variabilities and BMD loss in women undergoing menopause. To expand our knowledge base of the relationships between hormone variabilities and BMD trajectories, we developed a joint model that estimates individual means and variances of a longitudinal predictor and longitudinal outcome.

Our analyses found that higher mean E2 was associated with higher average predicted BMD, which supports the hypothesis that higher levels of E2 are protective against BMD loss, and support

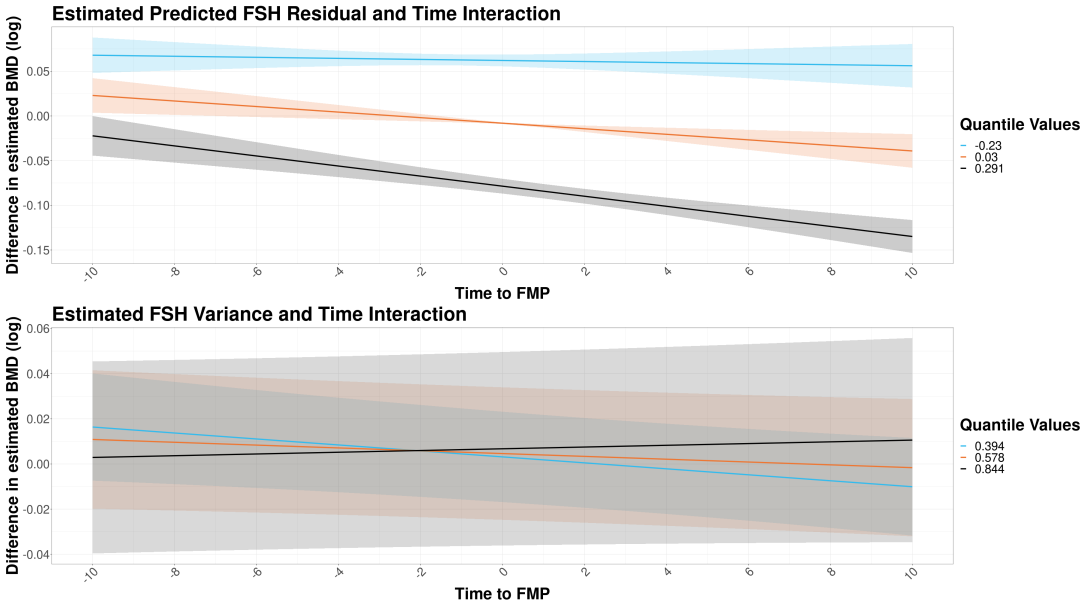


Figure 3.7: **Plots of predicted FSH and time interaction (top figure) and FSH variance and time interaction (bottom figure) for the BMD outcome model. The solid lines represent the 25th, 50th, and 75th quantile values of the FSH variable, along with the prediction band for this value. We can see the moderating effect of the variance-time interaction term as the BMD residual trajectories converge at about 2 years before FMP.**

bone health during the midlife. Conversely, higher FSH was associated with lower average BMD, which suggests that higher levels of FSH may negatively affect bone health in women during the midlife. These findings are consistent with the literature on the relationship between E2 and FSH levels and BMD outcomes in midlife women (Zaidi et al., 2018; Park et al., 2021).

Furthermore, we found that FSH variability was significantly associated with higher predicted BMD loss across the menopausal transition, with a one-unit higher FSH variability at FMP being associated with a 0.55% higher BMD. This relationship persisted past the MT, where higher individual FSH variability was associated with slower declines in BMD. We hypothesize that higher individual-level FSH variability could indicate a deviation beyond the expected reproductive system cessations during the menopauseal transition. In this case, the rapid period of bone loss during the MT may be also lower for these women if they are not experiencing the expected FSH trajectory. We did not find a similar relationship between BMD and E2 variability. This suggests that FSH variability is more strongly predictive of BMD trajectories in menopausal women than E2 variability. These findings should motivate further research into the role of FSH variability on BMD declines to more fully understand the relationship between individual variances of hormones and women's BMD trajectories.

Our analysis had a few limitations. As noted in the introduction, BMD has a nonlinear rapid decrease as women approach the menopausal transition, and then stabilizes around 2-5 years after

FMP. Originally, we attempted to fit an outcome model with a linear spline on time to FMP, with knots at 2 years before and after FMP, the model had trouble converging. In particular, the coefficients that represented the mean and variance interacted with the spline on time to FMP failed to converge. To address this, we decided to remove the nonlinear BMD trend by using the individual BMD residuals as the main outcome of interest; this left subject-level trends that could be modeled linearly. Additionally, as mentioned in Section 3.4, when we attempted to fit a linear time-varying variance on the SWAN dataset, the model had convergence problems, most likely due to the insufficient number of timepoints necessary to capture a time-varying variance trend.

3.5.0.1 Future Work

Although BMD has historically been the standard metric of choice used to evaluate bone health, some experts have cautioned against solely relying on BMD to measure bone health. Prentice et al. (1994) argued that since the formula for BMD assumes a constant proportional relationship between bone area and bone mineral content (BMC), this can lead to spurious correlations between BMD and health outcomes when the relationship between BMC and bone area is not directly proportional. "If BMD is used when the relationship between BMC and BA [bone area] is not one of simple direct proportion, part of its variation within a population will be due to differences in bone size between individuals." After having analyzed the association between hormone variability and BMD, the next question of interest would be to understand if these associations are also present with women's BMC trajectories and bone area trajectories.

The SWAN study has collected women's femoral neck BMC (g) measurements and femoral neck area (cm^2) measurements visits, but due to changes in the dual-energy x-ray absorptiometry (DEXA) collection machines over the course of the study, these measurements were not appropriately calibrated for longitudinal analyses. When the calibrated measurements become available, we plan to apply our model to BMC trajectories and bone area trajectories to better understand how E2 and FSH variabilities can predict BMC or bone area. Another possible extension of the model would be to simultaneously model both BMC and bone area trajectories as a multivariate outcome, rather than separately analyzing each variable.

An interesting future area of research could be to model complex formulation of the individual variances. In particular, decomposing individual time-varying variances into short-term and long-term trends may be of scientific interest for epidemiologists and physicians. To explore these higher-order trends and obtain sufficiently stable estimates of the variance trends, we will likely need a larger dataset than the currently available one from SWAN.

Finally, our joint model is specified for one longitudinal marker, but this could be extended to multiple longitudinal markers by specifying $\mathbf{B}_i = [\boldsymbol{b}_{i1}, ..., \boldsymbol{b}_{iQ}]^\mathsf{T}$, where Q is the number of markers, and a $Q \times Q$ variance-covariance matrix $\mathbf{S}_i$ for each individual. This extension would

have several considerations. The computation cost of estimating $\mathbf{S}_i$ would grow non-linearly as Q increases. Additionally, if we wanted to model both a mean regression and covariance regression (e.g. time-varying covariance matrices) in the predictor model, then this would further increase the computational burden of estimating the model.

CHAPTER 4

Chapter 4: A Joint Model with Covariance Regression of Longitudinal Hormone Biomarkers to Estimate Bone Health Outcomes Across the Menopausal Transition

4.1 Introduction

4.1.1 Motivation: bone health outcomes for midlife women

As longevity increases and the aging population continues to grow, programs and interventions that promote healthy aging are vital. Bone health is particularly important for older adults as fractures due to osteoporosis are a significant cause of mortality and lowered life expectancy (Sullivan et al., 2017; Bliuc et al., 2009). In women, bone mineral density declines sharply during the menopausal transition (MT), with women losing an average of 10% of bone in the 5-6 years surrounding the final menstrual period (FMP) (Ji and Yu, 2015). However, some women lose bone much faster, thus identifying women most at risk of clinically significant bone loss is key for establishing early interventions and treatment.

The menopausal transition is hallmarked by important endocrinologic changes including increases in follicle-stimulating hormone (FSH) and declines in estradiol (E2). Changes in E2 are associated with bone loss (Sowers et al., 2006; Crandall et al., 2013), and more recent work suggests that FSH may also be an important biomarker for bone loss. In a cross-sectional study of adult Chinese women, Wu et al. (2013) found that early declines in BMD were significantly correlated with FSH levels, but not with E2. Similarly, Shieh et al. (2019) found that FSH was a stronger predictor of "imminent bone loss" than E2. In animal models, "sharp rises" in FSH levels were associated with more rapid bone loss (Zaidi et al., 2018).

While testosterone levels do not change as dramatically during the menopause transition as

do E2 and FSH, it is an important biomarker associated with health outcomes such as depressive symptoms and sexual dysfunction, in midlife women (Maharjan et al., 2021; Janssen et al., 2010). R). In terms of bone health, testosterone has an important role in maintaining bone in men, but less is known about the importance of testosterone levels during the menopause and bone outcomes. In a recent paper, Zhang et al. (2022) found that higher mean testosterone predicted higher BMD in women aged 40-60. This relationship was also present in older post-menopausal women (Rariy et al., 2011), thus suggesting that female testosterone levels during the menopausal transition could be predictive of bone health.

To date, most studies evaluating sex steroid hormones (i.e., E2, FSH, testosterone) and bone outcomes have used approaches that associate population mean levels of the hormones with population mean level or change in BMD. However, other aspects of hormone trajectories including individual-level variability, may be informative biologically.

Bone mineral density (BMD) is the standard clinical metric of measuring bone health for many fields. Recently, Jepsen et al. (2023) found that BMD alone did not uniquely predict bone strength and suggested that other clinical measures in addition to BMD may be needed in order to accurately assess bone health. Because BMD is a ratio of bone mineral content (BMC) and bone area, it is of great interest to understand independent predictors of increases or decreases in BMC and bone area. To date, the bone loss literature on has focused mainly on BMD as an outcome, rather than on the components that make up this metric. Our previous work in Chapter 3 examined the relationships between sex steroid hormones and BMD and found that higher variability of FSH independently predicted slower declines in BMD across the menopausal transition. Scientifically, we extend the research questions from Chapter 3 by examining relationships between FSH and testosterone mean levels, variabilities, and their covariance and BMC across the MT.

4.1.2 Joint Models for Multiple Biomarkers

Multivariate analysis of biomarkers is a growing area of interest in disease prediction and diagnostics, since one singular biomarker is usually not completely predictive for a disease outcome. There is a large established work on joint models of longitudinal correlated biomarkers (Brown et al., 2005; Long et al., 2016; Li et al., 2021). These methods generally assume a constant covariance matrix for the residual variances and covariances. Recently, Zhao et al. (2021) presented a method that allows for time-varying correlations between longitudinal markers. The majority of these methods use only the estimates from the individual mean marker trajectories to predict the outcome, ignoring the variances and covariances from the trajectories.

In the setting with one biomarker of interest, research suggests that these residual marker variabilities may also predict individual health outcomes (Elliott et al., 2012; Jiang et al., 2015; Parker

et al., 2021). A well-established example of variability as a predictor is heart-rate variability (HRV), which has been shown to be significantly predictive of cardiovascular outcomes (Fang et al., 2020; Goldenberg et al., 2019). Higher variability of estradiol and higher variability of follicle-stimulating hormone were found to be predictive of increased depression risk in women transitioning through menopause (Freeman et al., 2006). Martins (2022) found that higher estimated CD4 count variability was associated with a higher hazard of death in HIV positive patients. These established associations suggests that biomarker variability may be a powerful tool for predicting and diagnosing disease progression. To date, however, the literature is lacking in methods that utilize the variances and covariances of multiple markers to predict health outcomes. Our first two chapters also focused on modeling 1) individual variances and covariances of multiple markers and 2) modeling individual (possibly time-varying) variances to predict time-varying outcomes. This motivates our development of a joint model to extend the existing work to the multivariate joint modeling setting to infer repeatedly measured outcomes. We develop a model that estimates the variances and covariances of *multiple* biomarkers to predict health outcomes. Our model allows individuals' variances and correlations, in addition to their means, to change over time by specifying mean and covariance regressions. This chapter extends the setting from Chapter 3 to the multivariate biomarker setting.

4.1.2.1 Covariance Regression Models

One major challenge of modeling a covariance matrix is ensuring that it remains positive definite when modeled as a function of time or other overrates. In Chapter 3, we proposed a model for the time-varying variance where the variance at each timepoint is a function of time-invariant coefficients and the corresponding timepoint. Extending this to the covariance matrix setting, Chiu et al. (1996) modelled the logarithm of the variance-covariance matrix as linear functions of covariates. Another approach, proposed by (Pourahmadi, 1999), is to model the elements of the matrix's Cholesky decomposition as functions of covariates. These two approaches are straightforward to implement, but the number of parameters to be estimated in both approaches can grow quite large, since each of the unique elements of the logarithm (or decomposition) is a function of coefficients. This would be especially computationally intensive in the case of estimating individual-specific covariance matrices. For this reason, we choose to explore a different parameterization for estimating our covariance matrices.

An alternative to estimating non-constant covariance matrices is with covariance regression models, as introduced by Hoff and Niu (2012). They specified a method for covariance regression as part of the standard multivariate regression model. They specify a covariance matrix as: $S(x) = \Sigma_0 + \Theta\boldsymbol{x}(\Theta\boldsymbol{x})^\intercal$, where Σ_0 is a positive-definite matrix and Θ is a vector of coefficients for predictors $\boldsymbol{x}$. Fox and Dunson (2015) used a Bayesian nonparameteric approach to this covariance

regression model that could further reduce the number of unknown covariance regression parameters to be estimated, by modeling a high-dimensional response as a function of lower-order latent factors. Li et al. (2014) extended the covariance regression model from Hoff and Niu (2012) to allow for the covariance matrices to also be a function of individual random effects, in addition to covariates.

Although these methods model the mean trajectories and the covariance trajectories of the response variable, the main focus of these models is to obtain estimates of the variance and covariance regression parameters within a linear mixed model setting, and not to relate these estimates to other outcomes of interest, i.e. within a joint model. In our motivating application, we are not only interested in modeling multiple biomarkers, but also using the subject-level random effects and residual variance-covariance estimates to predict a health outcome.

In this paper, we extend this work on covariance regression and joint models for correlated markers by introducing a joint model in Section 4.2 that uses both mean and variance-covariance regressions to model the multiple markers. We estimate the longitudinal markers by modeling time-varying individual mean trajectories and time varying variance-covariance trajectories for sex steroid hormones across the menopausal transition. These estimates are then used in to predict a longitudinal outcome (BMC). Given the high dimensionality and hierarchical nature of the model, we use a Bayesian approach that allows uncertainty at the multiple model levels to be properly accounted for when estimating variances of the parameters of interest. In Section 4.3, we apply our method to data from the Michigan Bone Health Study (MBHS). Section 4.4 provides a simulation study to assess the repeated sampling properties of our proposed model. We conclude with a summary of our findings and thoughts for next steps in Section 4.5.

4.2 Proposed Model

In this section, we outline the proposed joint model with mean and variance-covariance regressions. In the first submodel, we estimate individual-specific regression coefficients for a time-varying mean and covariance trajectories, and then link these individual random effects to an repeated outcome of interest in the second submodel.

4.2.1 Biomarker Submodel:

Let $\boldsymbol{X}_{ij}$ represent a vector Q predictor markers, measured at time $j = 1, \ldots, n_i$, where n_i is the total number of observations for individual i where $i = 1, \ldots N$. We model $\boldsymbol{X}_{ij}$ as:

$$\boldsymbol{X}_{ij} \mid \boldsymbol{B}_i, \boldsymbol{\Theta}_i, t_{ij} \sim \mathcal{N}_Q(\boldsymbol{\mu}(\boldsymbol{B}_i, f(t_{ij})), \boldsymbol{S}(\boldsymbol{\Theta}_i, g(t_{ij}))), j = 1, \ldots, n_i, \text{ independently for } i = 1, \ldots, N, \tag{4.1}$$

where $\boldsymbol{B}_i, \boldsymbol{\Theta}_i$ are matrices containing the individual-specific coefficients for the mean and covariance regression functions, respectively, and $f(t_{ij}), g(t_{ij})$ are basis expansions on t_{ij}, which may differ for the mean and variance trajectories.

4.2.1.1 Mean regression

The regression model for the mean trajectories is expressed as:

$$\boldsymbol{\mu}(\boldsymbol{B}_i, f(t_{ij})) = f(t_{ij})\boldsymbol{B}_i \tag{4.2}$$

where $\boldsymbol{\mu}(f(t_{ij}); \boldsymbol{B}_i)$ is a Q-dimensional function of time given by $\boldsymbol{B}_i = [\boldsymbol{b}_{i1}, ..., \boldsymbol{b}_{iQ}]^\mathsf{T}$ and $\boldsymbol{b}_{iq} = (b_{iq1}, \ldots, b_{iqP})^\mathsf{T}$ is a vector of P regression coefficients for the q-th marker. $f(t_{ij})$ is the design matrix resulting from a P dimensional basis-expansion function of time, where P is determined in advance by the degree of the basis expansion on time as well as the number of knots. For simplicity, we let P be the same for each marker in the modeling notation, although this can be relaxed in practice to allow each marker to be modeled with different basis expansion functions.

4.2.1.2 Covariance regression

Specifying a covariance regression model allows us to independently model the variability not already captured by the model for $\boldsymbol{\mu}_{ij}$. We have a $Q \times Q$ variance covariance matrix, $\boldsymbol{S}_{ij}$, whose regression model is specified with the following equation:

$$S(\boldsymbol{\Theta}_i, t_{ij})) = \boldsymbol{\Theta}_i g(t_{ij}) g(t_{ij})^\mathsf{T} \boldsymbol{\Theta}_i^\mathsf{T} + \Sigma_0, \tag{4.3}$$

where $\boldsymbol{\Theta}_i = [\boldsymbol{\theta}_{i1}, ..., \boldsymbol{\theta}_{iQ}]^\mathsf{T}$ is a $Q \times P_S$ matrix of coefficients and $g(t_{ij})$ is the design matrix from a P_S dimensional-basis expansion on time. P_S can be, but is not necessarily, equal to P, the number of individual mean regression coefficients. In practice, we choose $P_S < P$ to provide a more parsimonious covariance regression, in order to minimize the computational burden of estimating large $\boldsymbol{\Theta}_i$. We discuss the choices of P and P_S in our application (Section 4.3) and simulation study (Section 4.4).

This regression equation ensures that $\boldsymbol{S}_{ij}$ is positive semi-definite for each timepoints t_{ij}. Σ_0 is set to be a valid covariance matrix and the set of valid $Q \times Q$ variance-covariance matrices is

convex and closed under addition. Σ_0 can be thought of as the 'baseline' covariance matrix that is common across all individuals. To relax this assumption, we can include an intercept term in $g(t_{ij})$, so that at baseline, each individual's variance-covariance matrix is comprised of the population variance-covariance matrix and their own starting deviations. Σ_0 in this case, can be thought of as a 'residual' covariance matrix that is shared across all individuals when time is 0.

Hoff and Niu (2012) proved that $\boldsymbol{\Theta}_i$ is identifiable up to a sign change. In a Bayesian framework, this can be troublesome for the sampler, since the chains could move between Θ_i and $-\Theta_i$ without achieving convergence. We follow the suggestion from Li et al. (2014) and pass $\boldsymbol{\Theta}_i g(t_{ij})$ through an exponential link function. This function is applied element-wise to the $q \times 1$ vector, $\Theta_{iq} g(t_{ij})$. We note that this assumes that the random effects from the covariance regression have a multiplicative effect on the variance trajectories, similar to the assumption made by Parker et al. (2021) in the case of a single time-varying variance trajectory.

4.2.2 Outcome Submodel

In this model, our outcome of interest, Y_{ij}, is a one-dimensional vector of responses. We write the model for Y_{ij} as:

$$Y_{ij} \mid \mu_{ij}, \boldsymbol{S}_{ij}, \sim \mathcal{N}\left(\eta_{ij}(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W, \boldsymbol{\mu}_{ij}, \boldsymbol{S}_{ij}, \boldsymbol{W}_{ij}, t_{ij}, \boldsymbol{a}_i), \gamma^2\right), i = 1, \ldots, N, j = 1, \ldots, n_i \quad (4.4)$$

where $\eta_{ij}(\cdot) = \mu_{ij}$ is the mean regression function, $\boldsymbol{S}_{ij}$ are the estimated means, variances, and covariances from the predictor submodel, $\boldsymbol{W}_{ij}$ are observed covariates that can be time-invariant or time-varying, $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W$ are the population coefficients associated with the marker means, variances/covariances, and adjusted covariates respectively, and $\boldsymbol{a}_i$ is a vector of random effects. The specification of $\eta(\cdot)$ is flexible; for ease of interpretation in our data analysis, we explored linear basis expansion functions of η_{ij} for our outcome in order to maintain interpretability. An example of η_{ij} as a linear function of $\boldsymbol{\mu}_{ij}, \boldsymbol{S}_{ij}$ and with time-interactions is presented in the simulation study (Section 4.4).

Prior Specification

Prior for Σ_0 Following the approach of (Barnard et al., 2000), we let $\Sigma_0 = \mathbf{DRD}$, where $\mathbf{D} = \text{diag}(d_1, ..., d_Q)$ is a diagonal matrix of standard deviations and $\mathbf{R}$ is a correlation matrix. We assume that the standard deviations can be drawn from a half-Cauchy distribution and $\mathbf{R}$ can be drawn from a Lewandowski-Kurowicka-Joe (LKJ) prior (Lewandowski et al., 2009b):

$$d_q \sim \text{half-Cauchy}(0, \tau), \text{independently for} q = 1, \ldots, Q, \tag{4.5}$$

$$R \sim \text{LKJ}(\phi). \tag{4.6}$$

The parameters τ, ϕ can be specified in practice; we followed the suggested weakly informative values of $\tau = 2.5, \phi = 1$ in our simulation studies and application.

Prior for Θ_i For each marker q, we draw each set of individual coefficients $\boldsymbol{\theta}_{iq}$ from a multivariate Gaussian with mean zero:

$$\boldsymbol{\theta}_{iq} \overset{\text{indep.}}{\sim} \mathcal{N}_{P_S}(\mathbf{0}, \Psi_q), q = 1, \ldots, Q, \tag{4.7}$$

where Ψ_q is estimated with the same decomposition and priors specified in Equations 4.5-4.6, i.e., half-Cauchy and LKJ priors.

Prior for $\boldsymbol{b}_{iq}$ We can draw each $\boldsymbol{b}_{iq}$, the subject-specific mean regression coefficients for marker q from a multivariate Gaussian distribution:

$$\boldsymbol{b}_{iq} \overset{\text{indep.}}{\sim} \mathcal{N}_P(\boldsymbol{\beta}_q, \Omega_q), q = 1, \ldots, Q, \tag{4.8}$$

The priors on $\boldsymbol{\beta}_q, \Omega_q$ are specified as follows:

$$\boldsymbol{\beta}_q \sim \mathcal{N}_P(0, \xi_q^2 I_{P\times P}), \text{ independently for } q = 1, \ldots, Q \tag{4.9}$$

$$\Omega_q = \mathbf{K}_q \mathbf{L}_q \mathbf{K}_q, \ \mathbf{K}_q = \text{diag}\{k_{q1}, \ldots, k_{qP}\}, q = 1, \ldots, Q, \tag{4.10}$$

$$k_{qp} \sim \text{half-Cauchy}(0, \tau_0), p = 1, \ldots, P, \text{ and } \mathbf{L}_q \sim \text{LKJ}(\phi_0), \tag{4.11}$$

independently for $q = 1, \ldots, Q$ where $\mathbf{K}_q = \text{diag}\{k_{q1}, \ldots, k_{qP}\}$ is a diagonal matrix and $\mathbf{L}_q$ is a correlation matrix. τ_0 and phi_0 are specified in advance; in our analysis, we use the values of 2.5 and 1 (respectively) so that these priors are weakly informative.

Priors for $\boldsymbol{\beta}$, $\mathbf{a}_i$, σ For the outcome model, we use independent $\mathcal{N}(0, 10^2)$ priors for each element of the outcome regression parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^W$), and a diffuse prior on the outcome residual standard deviation parameter $\sigma \sim \text{half-Cauchy}(0, 2.5)$, as recommended by Carpenter et al. (2017). For $\boldsymbol{a}_i$, the random effects, we place a multivariate Gaussian prior with mean zero and precision τ_a, i.e., $\boldsymbol{a}_i \sim \mathcal{MVN}(0, \tau_a)$. In the case of a random intercept a_i, τ_a can be drawn from a half-Cauchy distribution or, in the case of a vector-valued $\boldsymbol{a}_i$, τ_a is a covariance matrix, whose values can be

drawn from the prior described in Equations 4.5-4.6 (i.e., half-Cauchy and LKJ prior).

Joint Distribution Let $D = (Y_{ij}, X_{ij}, t_{ij}, \mathbf{W}_i)$ denote the observed data, Z = ($\mathbf{B}_i$, $\mathbf{\Theta}_i$, $\boldsymbol{a}_i$) denote the subject-level latent variables, and $\boldsymbol{\zeta} = (\boldsymbol{\alpha}, \boldsymbol{\Sigma}_0, \xi, \boldsymbol{\Omega}, \boldsymbol{\Psi}, \boldsymbol{a}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W, \tau_a, \sigma)$ denote the model parameters. We also let $\pi(\boldsymbol{\zeta})$ denote the prior distribution of the parameters in $\boldsymbol{\zeta}$:

$$\pi(\boldsymbol{\zeta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\Omega})\pi(\boldsymbol{\Psi})\pi(\Sigma_0)\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\gamma}^W)\pi(\sigma).$$

If we let C be the normalizing constant for the LKJ prior Lewandowski et al. (2009b), we can then write the joint distribution of D, Z, and $\boldsymbol{\zeta}$ as

$$\begin{aligned}
P(\boldsymbol{\zeta}, D, Z) \propto \prod_{i=1}^{N}\prod_{q=1}^{Q} \Bigg\{ & \frac{1}{\sqrt{(2\pi)|\Psi_q|}} \exp\left(-\frac{1}{2}(\mathbf{b}_i - \boldsymbol{\beta})^{\mathsf{T}} \Psi_q^{-1} (\mathbf{b}_i - \boldsymbol{\beta})\right) \\
& \times \left(\frac{1}{\sqrt{2\pi|\Psi_q|}} \exp\left(-\frac{1}{2}\mathbf{\Theta}_i |\Psi_q|^{-1} \mathbf{\Theta}_i\right) + \frac{2}{\pi\tau} \frac{1}{1 + \frac{(d_q^2)}{\tau^2}} \times C[\det(R)]^{\phi-1} \right) \Bigg\} \\
& \times \frac{1}{\sqrt{2\pi|\tau_a|}} \exp(-\frac{1}{2}\boldsymbol{a}_i^{\mathsf{T}} \tau_a^{-1} \boldsymbol{a}_i) \\
& \times \prod_{i=1}^{N}\prod_{j=1}^{n_i} \Bigg\{ \frac{1}{\sqrt{(2\pi)|\mathbf{S}_{ij}|}} \exp\left(-\frac{1}{2}\{\boldsymbol{X}_{ij} - \boldsymbol{\mu}(t_{ij}; \mathbf{B}_i)\}^{\mathsf{T}} \mathbf{S}_{ij}^{-1} \{\boldsymbol{X}_{ij} - \boldsymbol{\mu}(t_{ij}; \mathbf{B}_i)\}\right) \\
& \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_{ij} - \eta_{ij}(\boldsymbol{\mu}_{ij}, \boldsymbol{S}_{ij}, \mathbf{W}_{ij}, \boldsymbol{a}_i; \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\gamma}^W))^2}{\sigma^2}\right) \Bigg\} \\
& \times \pi(\boldsymbol{\zeta}). \qquad (4.12)
\end{aligned}$$

4.2.3 Model Estimation

We developed a Hamiltonian Monte Carlo program for our proposed model, using the STAN probabilistic programming language and `rstan` R package. For our simulations, we ran 2 chains each for 2,000 iterations and 1,000 burnin. For the analysis, we found that 2 chains with 4,000 iterations and 2,000 burnin was sufficient for achieving convergence. We evaluated model convergence from visual examinations of traceplots and from Stan's R-hat convergence diagnostic (Vehtari et al., 2021). The outcome parameters our models reported R-hat values < 1.05. Additional model checking is discussed in Section 4.3.4.

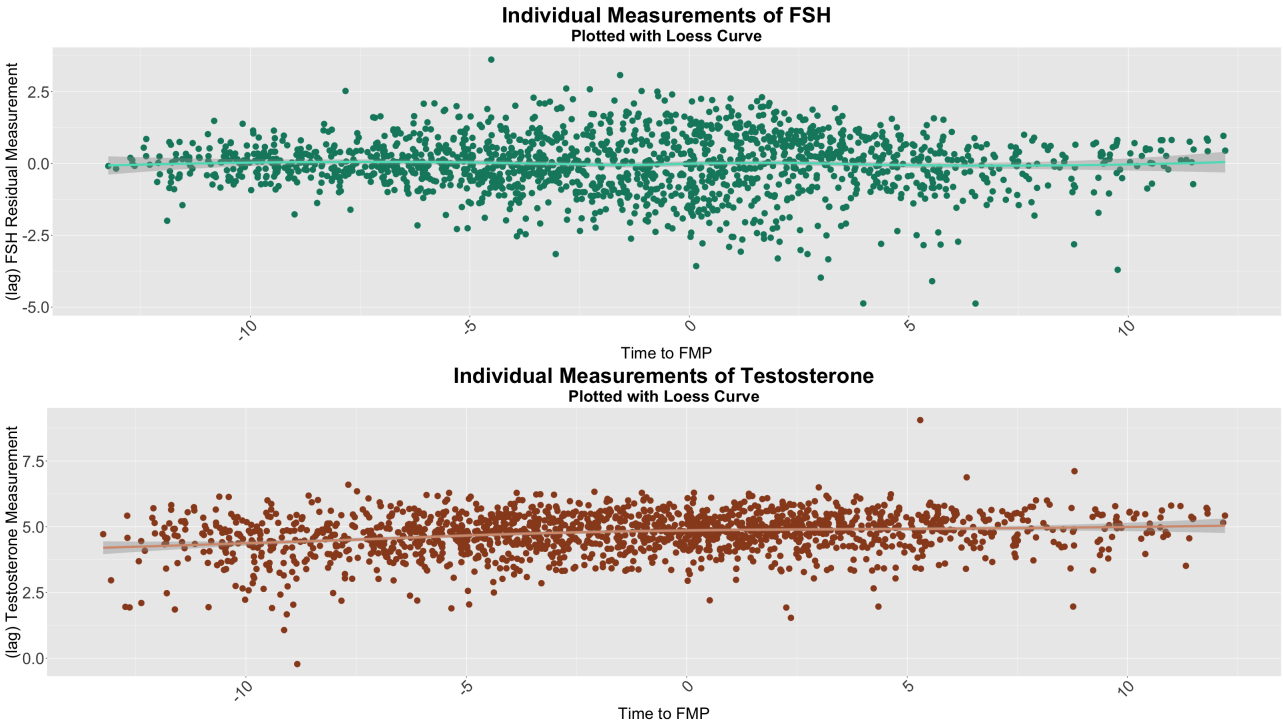


Figure 4.1: **Individual FSH residuals and testosterone measurements over time, plotted with a loess curve to visualize the overall trend over time.**

4.3 Application: Michigan Bone Health Study

4.3.1 Dataset

The Michigan Bone Health and Metabolism Study (MBHS) is a population-based longitudinal study of women's sex steroid hormones and their relation to the initiation and development of musculoskeletal conditions, metabolic diseases, and functional limitations as women transition through the menopause Sowers et al. (1998). A total of 664 white women were initially enrolled in the study in 1992. After excluding women who did not have a non-surgical and observable FMP (i.e., not using hormones that obscured FMP, n=326) or who did not have pre- and/or post-FMP DEXA scans (n=127), our final analytic dataset included 211 women representing 1,662 unique observations.

BMC was measured using dual-energy X-ray absorptiometry (DEXA) machines at baseline and each of the subsequent annual follow-up visits (up to 14 in total). At each visit, a fasted blood sample was obtained during days 2-7 of the follicular phase of the menstrual cycle and assayed for FSH and testosterone. Information about the biospecimen collection protocol and hormone assays has been previously published (Sowers et al., 1998). At each annual exam, women completed questionnaires related to their medical history and other socio-demographic variables (e.g. physical

Table 4.1: **Descriptive statistics of the BMC dataset, based on 210 individuals in the MBHS cohort.**

Variable	Statistic	Value	n
Longitudinal Predictor	Mean/SD		
FSH Residuals		-0.01 (0.99)	1,662
Testosterone (log_2 transformed)		4.63 (0.82)	1,662
Health Outcome	Mean/SD		
BMC Residuals		0.00 (0.24)	1,662
Adjusted Covariates	Mean/SD		
Baseline BMI		27.28 (6.83)	1,662
Baseline Age		46.29 (2.60)	1,662

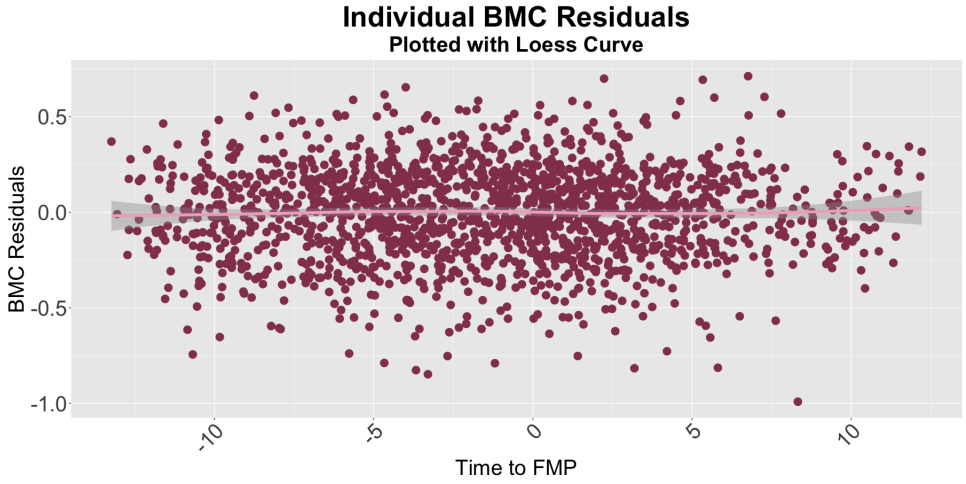


Figure 4.2: **Individual BMC residuals over time, plotted with a loess curve to visualize the overall trend over time.**

activity levels and smoking status). Table 4.1 contains the dataset's descriptive statistics.

Previous research using data from the Study of Women's Health Across the Nation has indicated that higher FSH is associated with lower BMD values (Shieh et al., 2019). Since BMD is computed as a ratio of BMC divided by bone area, we would expect mean FSH to be negatively associated with BMC. Testosterone has also been positively associated with higher BMD in adult women (Zhang et al., 2022).

In Chapter 3, we also found that time-invariant FSH variability was positively associated with BMD. We would expect to find similar results with our covariance regression joint model (i.e., higher variability of FSH predicts higher BMC). Since our model in Chapter 3 specified a single predictive marker, we were not able to investigate associations between hormone covariances and bone outcomes.

4.3.2 Implementing the Proposed Model

We now apply our proposed model to analyze the hormone and bone trajectory data described in the previous section. Based on previous research on the approximate timeframes where mean FSH has noticeable changes (Randolph et al., 2011), we placed knots at (-7, -2, 2, 7) years to FMP in the individual marker mean regression. For the variances and covariances, we placed knots at (-2, 2) years before and after the FMP, allowing for the possibility that the individual variances and covariances could change across the menopausal transition. We conducted some sensitivity analyses on these knot placements in Appendix D.1.3.

Our outcome model formulation was specified as follows:

$$\mathbb{E}(\log_2(BMC_{ij})) = \beta_0 + \beta_1\mu_{ij1} + \beta_2\mu_{ij2} + \beta_3 S_{ij11} + \beta_4 S_{ij12} + \beta_5 S_{ij22}$$
$$+(\beta_6 + \beta_7\mu_{ij1} + \beta_8\mu_{ij2} + \beta_9 S_{ij11} + \beta_{10} S_{ij12} + \beta_{11} S_{ij22})t_{ij} + \beta_{12} BMI_i^* + \beta_{13} Age_i^* + a_{0i},$$

where our $Q = 2$ markers of interest were FSH residuals and testosterone measurements. Figure 4.1 shows the observed FSH and testosterone measurements used in the predictor submodel. We adjusted for baseline BMI (BMI^*) and baseline age (Age^*), and included a random intercept for each woman to account for the residual correlations between bone measurements. For the longitudinal predictors, we used the hormone measurement obtained at the previous visit to predict BMC at the following visit. This is to better capture how differences in hormone measurements at an earlier time may be associated with BMC declines later, rather than analyzing hormone and bone values at the same timepoint.

4.3.3 Results

4.3.3.1 Individual variance trajectories

A defining feature of this covariance regression joint model is that we are able to examine the individual-level variance trajectories. As an example, Figure 4.3 displays the estimated posterior mean values of individual FSH variances over time. The dashed lines indicate the pre-specified knots used in the model. During the [-2, 2] years to FMP window, we see that the average estimated FSH variance increases, compared to the average trend before -2 years to FMP.

Additionally, we can see that the overall estimated mean FSH variability tends to decline after FMP. This may suggest that mean FSH levels tend to stabilize after the menopausal transition. However, the individual estimates of the variance also become more variable after FMP. This could be due to the smaller number of post-FMP observations in the dataset, but could also indicate that post-FMP, there may be larger individual level variability of FSH observations. We obtained

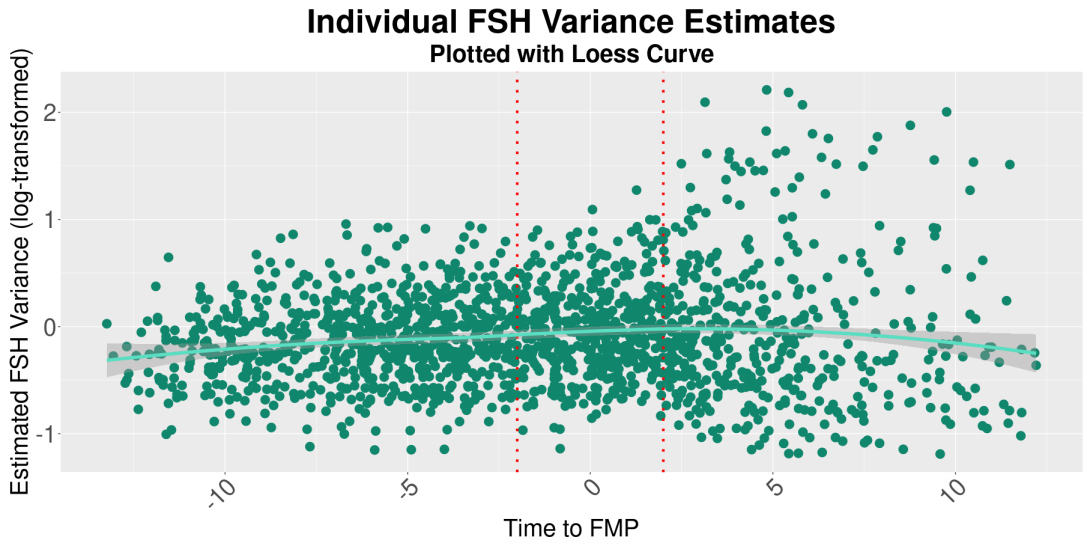


Figure 4.3: **Plots of the posterior means of estimated individual FSH variances. A loess curve has been added to the plot to show the overall trend of FSH variability over time.**

posterior estimates of the individual mean coefficients and plotted these histograms for each time period defined by the prespecified knots.

Figure 4.4 displays the estimated posterior mean values of individual testosterone variance trajectories. Compared to the FSH variance trajectory estimates, there is more estimated variation in the estimated testosterone trajectories before and during FMP. Average estimated testosterone variability tends to remain constant before and during the [-2, 2] FMP window. After FMP, we see that the average trend for testosterone variability tends to slightly increase. Compared to the FSH variability trend, however, the average change in slope is much smaller post-FMP. Unlike FSH, testosterone levels in women do not experience a rapid change during the MT. It follows that estimated testosterone variability also does not display sudden changes around FMP.

The estimated correlation trajectories (Figure 4.5), in contrast, tend to be more stable. The estimated population correlation from the MBHS data was 0.012 and the majority of the individual estimates appear to be centered around 0, although there is still noticeable variability in theses estimates. After FMP, there appears to be a slight upward trend in correlation, indicating that average correlation between FSH and testosterone may increase post-menopause.

4.3.3.2 Associations between hormone variabilities and BMC

Table 4.2 displays the estimated coefficients produced by the model. Mean FSH at the visit before FMP was significantly associated with BMC at FMP. At FMP, a one unit higher predicted FSH at the previous visit was associated with a $1-2^{-0.036}\times 100\% = -2.5\%(-3.7\%, -1.3\%)$ lower BMC. This relationship is not estimated to change over time, since the interaction between predicted FSH

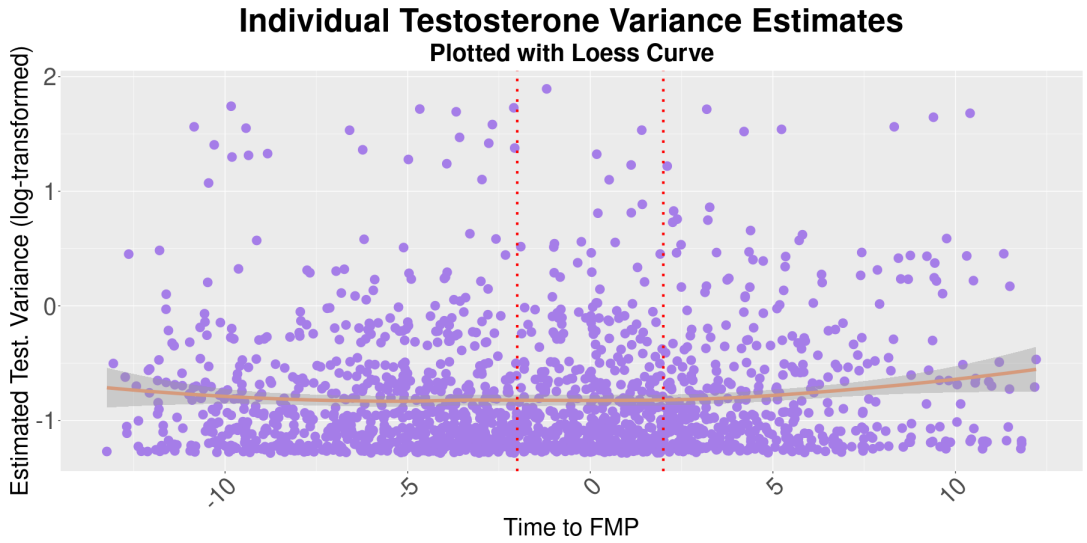


Figure 4.4: **Plots of the posterior means of estimated individual testosterone variances. A loess curve has been added to the plot to show the overall trend of over time.**

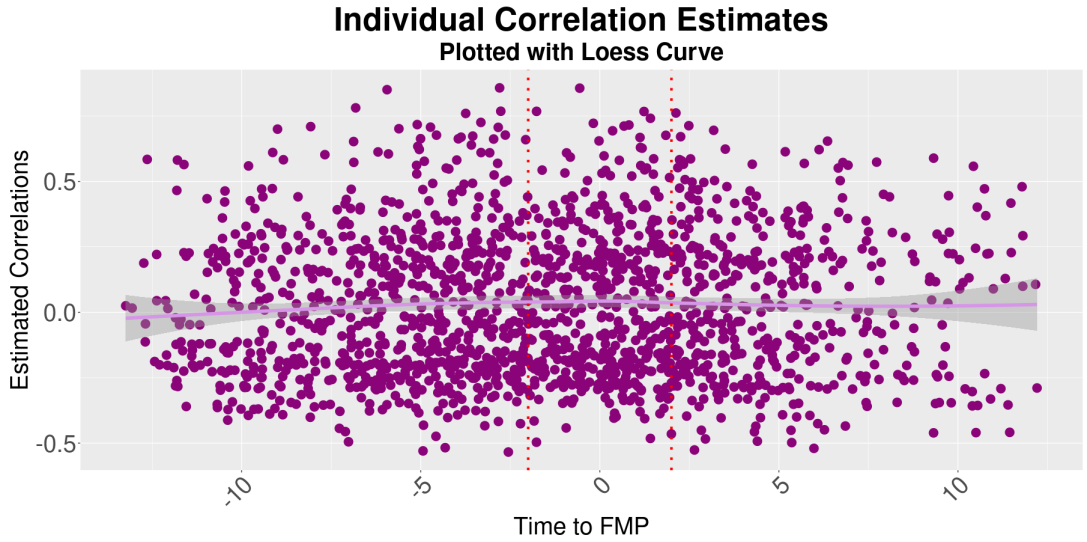


Figure 4.5: **Plots of the posterior means of estimated individual FSH, testosterone correlations. A loess curve has been added to the plot to show the overall trend of over time.**

Variable	Post. Mean	95% CrI
Predicted FSH	**-3.68**	**(-5.62, -1.96)**
Predicted Test.	1.97	(-0.19, 4.22)
FSH Var.	-4.50	(-11.44, 0.94)
FSH, Test. Cov.	-3.64	(-28.05, 15.92)
Test. Var.	**-11.38**	**(-24.79, -2.29)**
Time to FMP	0.96	(-0.67, 2.73)
Time to FMP x Predicted FSH	0.23	(-0.85, 0.03)
Time to FMP x Predicted Test.	0.12	(-0.15, 0.39)
Time to FMP x FSH Var.	-0.53	(-1.13, 0.04)
Time to FMP x FSH, Test. Cov.	**5.63**	**(4.13, 7.53)**
Time to FMP x Test. Var	**-2.96**	**(-4.77, -1.29)**
BMI	**11.44**	**(8.42, 13.87)**
Age	-0.61	(-1.33, 0.22)

Table 4.2: **Estimated posterior means and 95% credible intervals for the BMC model with FSH and Testosterone markers. All estimated posterior means and 95% CrI values have been multiplied by 10^2.**

and time is not significant. Mean testosterone, on the other hand, was not significantly associated with BMC.

Testosterone variability was estimated to be significantly predictive of BMC. At FMP, testosterone variability at the previous visit had a negative relationship with BMC. This effect becomes amplified over time. Holding testosterone variability constant, each additional year is associated with a $1 - 2^{0.0096} \times 2^{-0.0296} \times 100\% = -1.37\%(-3.7\%, 1.0\%)$ change in BMC. This effect over different stages of the midlife can be seen in 4.6. The largest difference in estimated BMC is 5 years post FMP, where higher testosterone variability is associated with lower BMC residuals (i.e. lower than average BMC than the population average).

FSH and testosterone covariance was also significantly associated with BMC. At FMP, the predicted relationship between covariance and BMC is negative (although not significant). After FMP, however, this relationship changes to become positive over time. Holding covariance constant, each additional year is associated with a $2^{0.0096} \times 2^{0.0563} - 1 \times 100\% = 4.3\%(2.4\%, 7.3\%)$ change in BMC residual. Figure 4.7 shows the estimated associations between covariance and BMC residuals for different timepoints. Before FMP, a positive covariance between FSH and testosterone is associated with lower values of BMC residual (lower than average BMC). After FMP, however, a positive covariance is associated with higher than average BMC.

Finally, higher BMI was positively associated with BMC, holding all else equal. This is consistent with the literature on the associations between BMI and bone. Age, however, was not significantly associated with BMC.

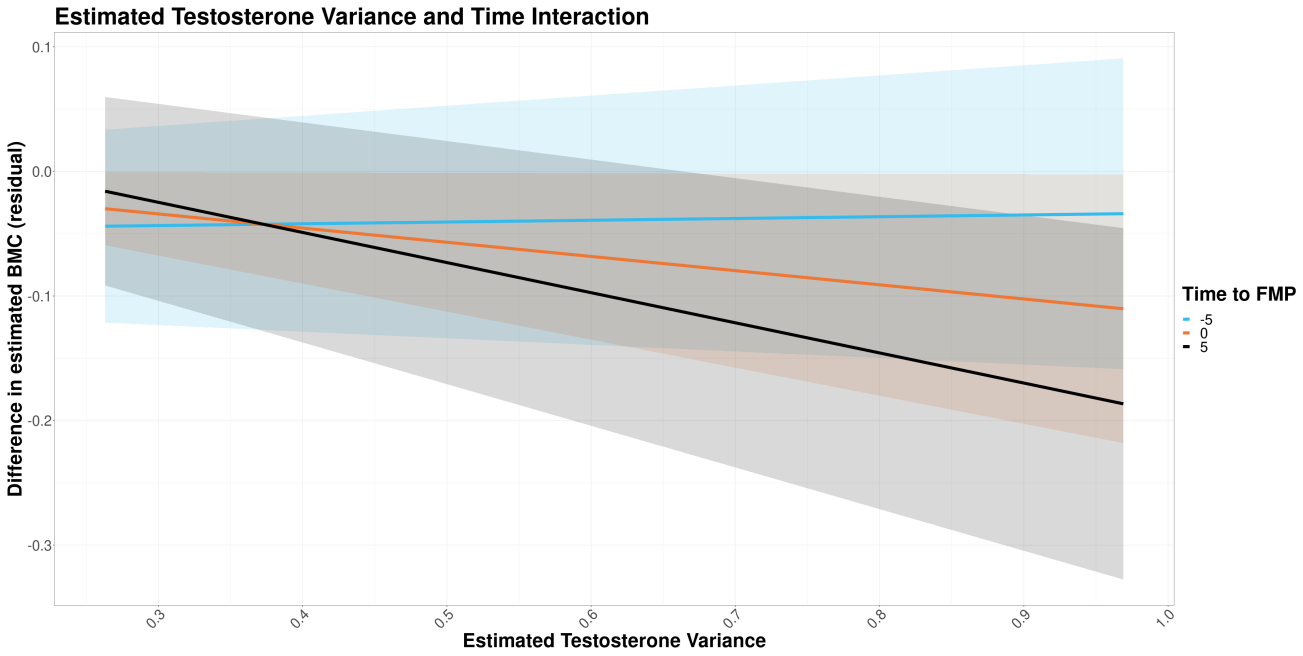


Figure 4.6: **Plots of predicted testosterone variance and time interaction for the BMC outcome model. Different values of estimated testosterone variance quantiles are shown on the x-axis, with corresponding estimated BMC differences on the y-axis.The solid lines represent different years to FMP (-5, 0, 5).**

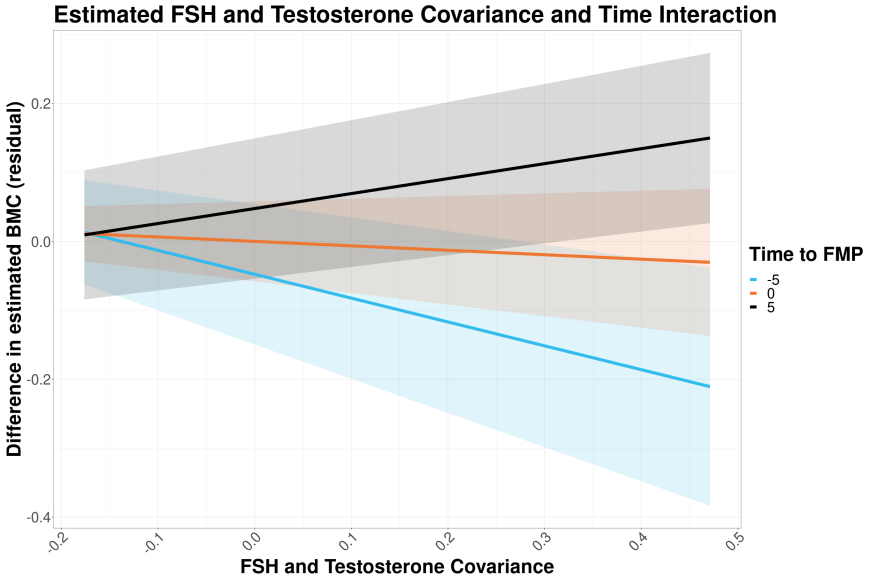


Figure 4.7: **Plots of predicted covariances and time interaction. Different values of estimated covariance quantiles are shown on the x-axis, with corresponding estimated BMC differences on the y-axis.The solid lines represent different years to FMP (-5, 0, 5).**

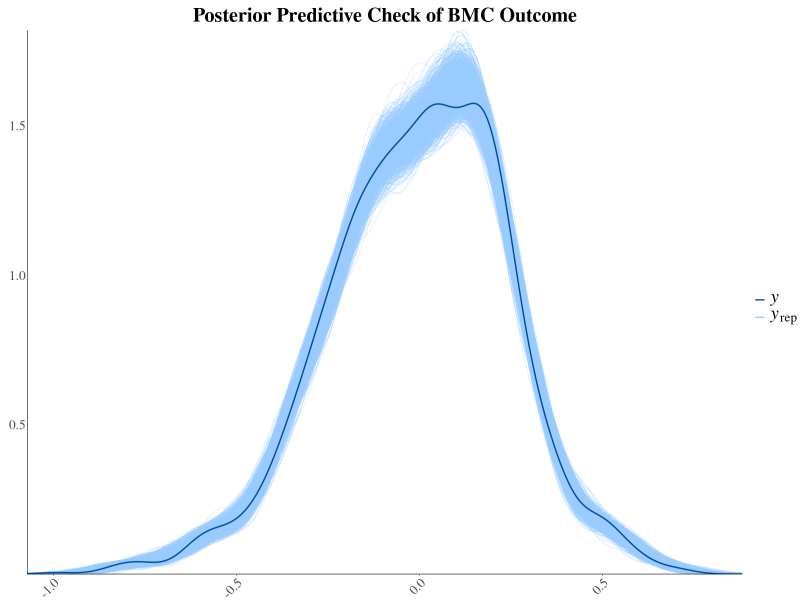


Figure 4.8: **Posterior predictive check of the BMC residuals, with the solid dark line indicating the observed data values and the thinner, lighter bands representing kernel density estimates of the posterior predictive distributions.**

4.3.4 Model Checking

We conducted several model checking procedures to understand if the modeling assumptions are validated in the MBHS data. For the outcome of interest, we performed a posterior predictive check where we visually plotted the observed distribution of the BMC residuals against model-generated BMC residuals. Figure 4.8 displays this graph. We can see that the model-generated kernel density estimates of the outcome overlap onto the the observed distribution of BMC residuals. Figure D.1 in Appendix D.1 display the graphs for the FSH and testosterone predictors.

Another approach for model checking is to compute the posterior predictive distribution p-values. We use the following statistic: $T(x_{ij}; b_{ip}, t_{ij}) = \sum_t (x_{ij} - \mu(b_{ipq}, t_{ij}))^2$ where $\mu(b_{ipq}, t_{ij}))$ is the estimated individual i's mean trajectory for marker q. By doing this, we can compare $T_i(x_{ij}^{obs}; b_{ipq}, t_{ij})$ (which is a function of the observed data and the estimated parameters) with $T_i(x_{ij}^{rep}; b_{ipq}, t_{ij})$, a function of the model generated data using the model estimated parameters). To compare these two $T(x_{ij})$ statistics, we compute the 'posterior predictive p-value', which is $P(T_i(x_{ij}^{obs}; b_{ipq}, t_{ij}, \sigma_{iq}^2)_q < T_i(x_{ijq}^{sim}; b_{ipq}, t_{ij}, \sigma_{iq}^2)_q | (x_{ijq}^{obs}))$. For the $q-th$ marker, we keep (x_{ij}^{obs}) fixed at the observed values and compute 1,000 values of $T_i(x_{ij}^{sim}; b_{ipq}, t_{ij})$ from the posterior of $\boldsymbol{b}_{ipq}$. We then compared these values with 1,000 draws from $T_i(x_{ij}^{sim}; \boldsymbol{b}_{ipq}$ and computed the corresponding p-values.

Figure 4.9 shows the histograms of computed p-values for FSH and testeosterone. The mean FSH p-value and mean testosterone p-value were both 0.6. For both markers, the majority of p-

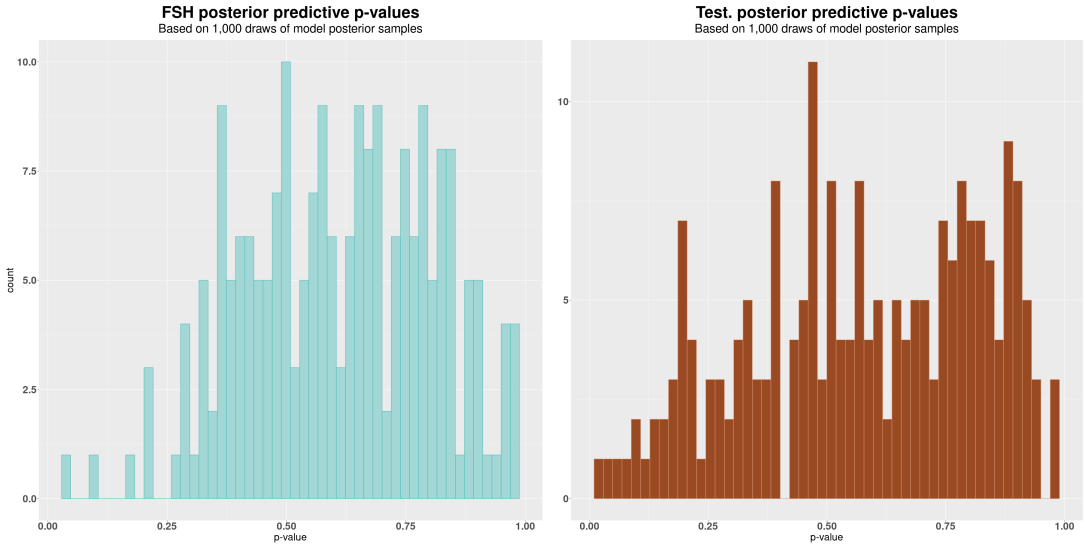


Figure 4.9: **Posterior predictive check of the FSH residuals and testosterone measurements.**

values are between 0.4 and 0.75. Ideally, the mean p-value would be at 0.5, which suggests that our model may be over-estimating the observed data.

To further investigate our model, we randomly select individuals based on the quantiles of the p-values for each marker. For each individual, we computed the 50th and 95th quantiles of their estimated trajectories ($\boldsymbol{\mu}_{ij}$) from 500 randomly selected draws of the posterior samples. We then plotted these against their observed measurements in Figure 4.10.

The model estimated individual trajectories appear to capture the individual level mean trajectories for both markers. The majority of observations are also covered by the 95th quantile estimates of the trajectories. Further model checks can be found in Appendix D.1.

4.4 Simulation Study

To confirm the repeated sampling properties of our model, we also performed a simulation study, where the goal was to evaluate the performance of our joint model in recovering the true data-generating parameters. We evaluated our model in terms of three metrics: 1) bias 2) the coverage rate of the nominal 95% credible intervals and 3) average length of the 95% CrIs obtained across simulation replicates. We focus on evaluation of the proposed model because there is not a clear simple competitor for our approach.

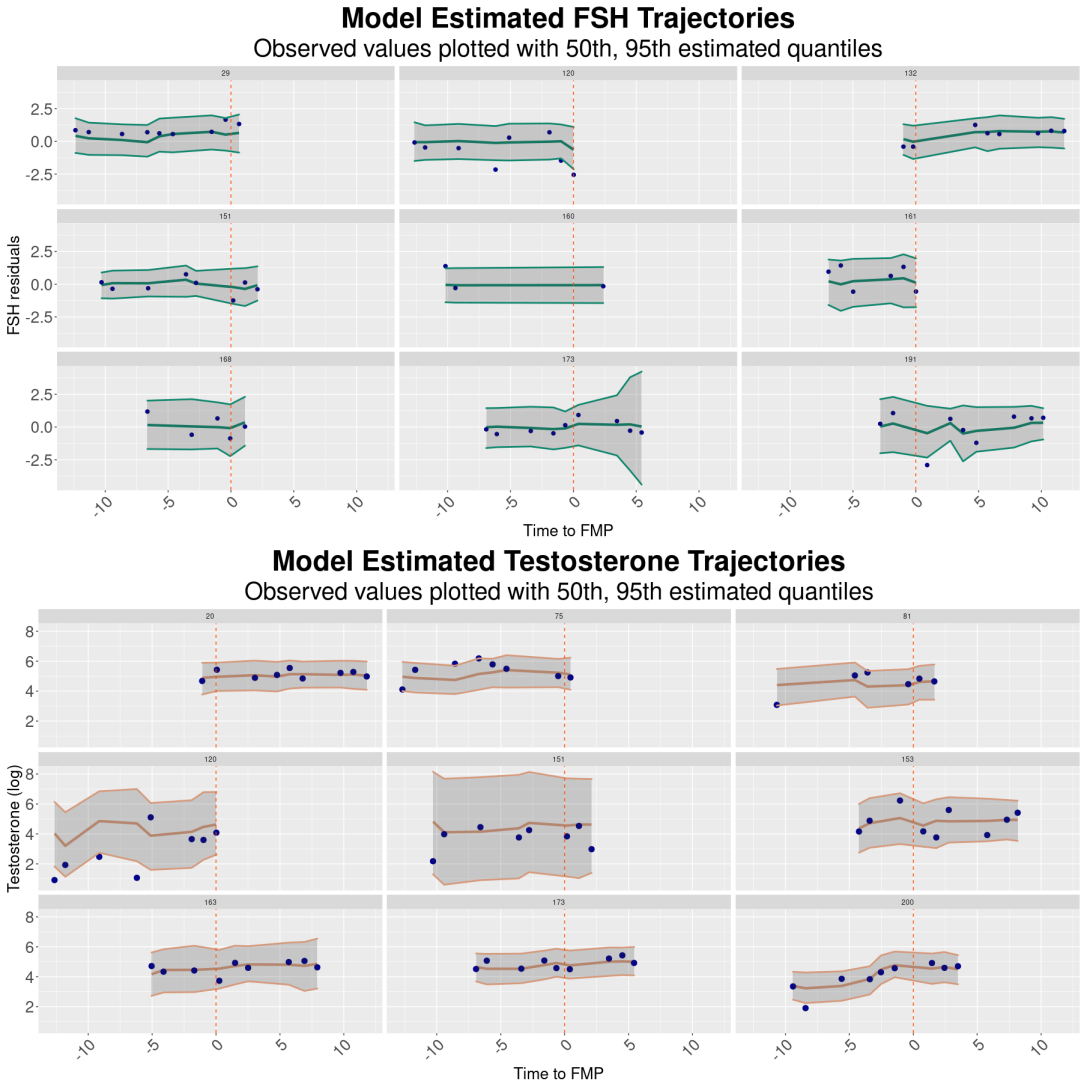


Figure 4.10: **Individual level model-estimated and observed marker trajectories of FSH residuals (top figure) and testosterone (bottom figure). The observed values are shown as points and and the 50th quantile estimated trajectories are shown as lines. The plotted ribbons represent the 95th quantile estimates.**

4.4.1 Data Generation

For this simulation study, we generated data for N = 200 individuals and $Q = 2$ markers. Mimicking our application dataset, we simulate timepoints based on years pre- and post-FMP, with each individual having between 1-24 timepoints. For the marker means, we apply a linear B-spline expansion to each individual's timepoints, placing knots at timepoints (-7, -2, 2, 7). We also add a global intercept term to the marker means.

$$\boldsymbol{X}_{ij} \mid \boldsymbol{\mu}_{ij}, S_{ij} \sim N_Q(\boldsymbol{\mu}_{ij}, S_{ij}), \mu_{ijq} = b_{0q} + \boldsymbol{B}_i f(t_{ij}), \boldsymbol{b}_{iq} \sim \mathcal{N}_{\mathcal{P}}(\beta_q, \Omega_q),$$
$$\beta_1 = (-0.57, 0.18, -0.06, 0.36, 0.34)^{\mathsf{T}}, \beta_2 = (0.26, -0.32, 0.00, -0.40, -0.21)^{\mathsf{T}},$$
$$\Omega_1 = \begin{pmatrix} 0.21 & 0.02 & 0.07 & 0.04 & 0.02 \\ 0.02 & 0.36 & 0.13 & -0.04 & -0.04 \\ 0.07 & 0.14 & 0.41 & 0.05 & 0.29 \\ 0.04 & 0.11 & 0.10 & 0.02 & -0.04 \\ 0.023 & -0.04 & 0.29 & 0.10 & 0.56 \end{pmatrix}, \Omega_2 = \begin{pmatrix} 0.09 & 0.11 & 0.02 & 0.01 & 0.01 \\ 0.11 & 0.44 & 0.15 & 0.11 & 0.08 \\ 0.02 & 0.15 & 0.37 & 0.28 & 0.25 \\ 0.01 & 0.11 & 0.28 & 0.35 & 0.29 \\ 0.01 & 0.08 & 0.25 & 0.29 & 0.33 \end{pmatrix}.$$

We obtained these values of β_q, Ω_q by fitting this model on a subset of individuals in our dataset and collected the posterior mean estimates of β_q, Ω_q. These were then used to generate $\boldsymbol{X}_{ij}$.

For the covariance regression, we also apply a linear B-spline expansion with knots at (-2, 2). We generate the individual variance-covariance matrices for each individual using the following data generating parameters:

$$S_{ij} \mid \boldsymbol{\Theta}_i, \Sigma_0 = \rho(\boldsymbol{\Theta}_i g(t_{ij}))\rho(\boldsymbol{\Theta}_i g(t_{ij}))^{\mathsf{T}} + \Sigma_0,$$
$$\Theta_{iq} \sim \mathcal{N}(\boldsymbol{0}, \Psi_q),$$
$$\Psi_1 = \begin{pmatrix} 0.09 & -0.02 & -0.001 \\ -0.02 & 0.075 & -0.001 \\ -0.001 & -0.001 & 0.001 \end{pmatrix}, \Psi_2 = \begin{pmatrix} 0.21 & -0.02 & -0.01 \\ -0.02 & 0.17 & -0.01 \\ -0.01 & -0.01 & 0.002 \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} 0.81 & -0.3 \\ -0.3 & 0.21 \end{pmatrix}.$$

For the outcome submodel, we specify the outcome, Y_{ij} as:

$$\begin{aligned} &Y_{ij} \sim \mathcal{N}(\eta_{ij}, 0.01^2), \eta_{ij} = \eta(\boldsymbol{\mu}_{ij}, S_{ij}) \\ &\eta(\boldsymbol{\mu}_{ij}, S_{ij}) = \alpha_0 + \alpha_1 \mu_{ij1} + \alpha_2 \mu_{ij2} + \gamma_1 s_{ij11} + \gamma_2 s_{ij21} + \gamma_3 s_{ij22} \\ &+ (\alpha_3 + \alpha_4 \mu_{ij1} + \alpha_5 \mu_{ij2} + \gamma_4 s_{ij11} + \gamma_5 s_{ij21} + \gamma_6 s_{ij22}) t_{ij} + a_{0i}, \\ &a_{0i} \sim \mathcal{N}(0, 0.1^2), \end{aligned}$$

where the pre-set values of $\boldsymbol{\alpha}, \boldsymbol{\gamma}$, can be found in the first column of Tables 4.3 and 4.4.

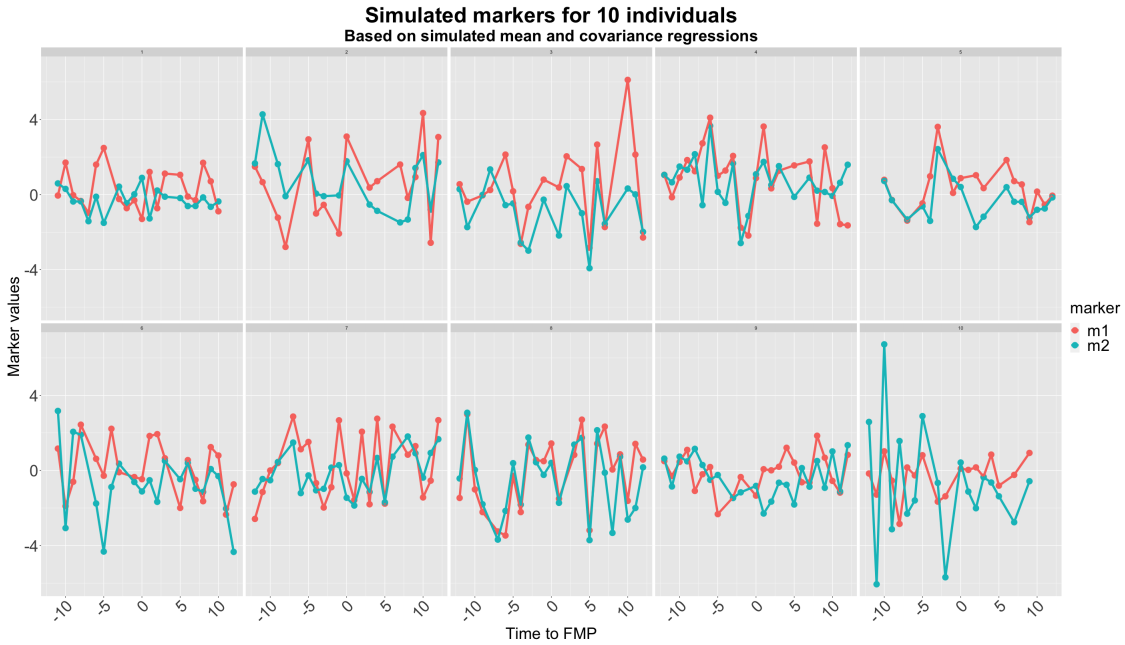


Figure 4.11: **Simulated marker trajectories for 10 individuals, based on the pre-specified data generating parameters and simulated timepoints.**

Table 4.3: **Simulation I: bias, coverage, and mean 95% credible interval length for the mean parameters across 55 simulation replicates.**

Truth	Bias	Coverage (%)	Ave. Interval Length
α_1 = -0.5 (Mean Biomarker 1)	0.01	90.74	0.16
α_2 = 0.3 (Mean Biomarker 2)	-0.01	90.74	0.21
α_4 = 0.1 (Mean 1 $\times$ time)	0.00	98.15	0.16
α_5 = -0.03 (Mean 2 $\times$ time)	-0.01	98.15	0.20

4.4.2 Results

Tables 4.3 and 4.4 display the estimated average bias, coverage rates, and average interval lengths for 55 simulation replicates. The joint model is able to recover the parameters with high coverage and low bias. This assures us that the model is indeed able to estimate the data-generating parameters with consistency and accuracy, when the modeling assumptions are valid. Tables D.3, D.4, and D.5 in Appendix D.3 contain the results for the other model parameters.

Table 4.4: **Simulation I: bias, coverage, and mean 95% credible interval length for the variance parameters across 55 simulation replicates.**

Truth	Bias	Coverage (%)	Ave. Interval Length
γ_1 = 0.1 (Var Biomarker 1)	0.00	96.30	0.24
γ_2 = 0.3 (Cov Biomarker 1,2)	-0.01	98.15	0.42
γ_3 = 0.2 (Var Biomarker 2)	0.01	98.15	0.20
γ_4 = -0.05 (Var 1 $\times$ time)	-0.02	96.30	0.23
γ_5 = 0.1 (Cov 1,2 $\times$ time)	-0.01	100.0	0.35
γ_6 = -0.08 (Var 2 $\times$ time)	-0.02	98.15	0.17

4.5 Discussion

Women are at increased risk of bone fractures after the menopausal transition (Ahlborg et al., 2003; Sowers et al., 2010; Sullivan et al., 2017). In order to support diagnostics for such health risks, it is important to identify biomarkers that can predict bone health. Previous research has shown that higher overall FSH predicts lower bone density. However, the literature has not yet examined associations between time-varying variability of FSH on bone outcomes such as BMC. Furthermore, the role of testosterone in maintaining women's bone health, particularly during the menopausal transition, is unknown.

We have proposed a method that jointly models individual mean, variance, and covariance trajectories over time using mean and covariance regression strategies. This model allows us to estimate the changes over time in biomarker variances and covariances, and relates these estimates to a health outcome of interest. To the best of our knowledge, this is also the first time that *time-varying* variance and correlation estimates of FSH, and testosterone have been used to predict BMC outcomes in midlife women.

As expected, we found that higher average FSH at FMP was found to be significantly associated with faster BMC declines. It is hypothesized that FSH may intensify bone resorption by stimulatinge osteoclasts, the cells that break down and repair bone structure. (Chin, 2018; Zaidi et al., 2018), which could explain this estimated negative relationship with BMC. This relationship did not appear to change over the menopausal transition. Mean testosterone, on the other hand, was not significantly associated with BMC. The direction of the relationship (positive), was, however, consistent with the literature on testosterone and bone health in women (Rariy et al., 2011).

We also estimated relationships between hormone variances/covariances and bone outcomes. Time varying FSH, testosterone covariance was significantly associated with BMC. After FMP, higher covariance (correlation) between testosterone and FSH was associated with slower declines

in BMC. Scientifically, a higher correlation between FSH and testostone could mean that these two hormones are behaving more synchronously with each other. This may in turn be a potential indicator that the individual hormone trajectories have settled post-menopause. We also found that higher variability of testosterone, but not FSH, was associated with faster declines in BMC. Additionally, testosterone variance over time was also significantly associated with BMC. Higher testosterone variance post-FMP tended to predict lower BMC values than high testosterone variances pre-FMP. To the best of our knowledge, this is the first time that testosterone variability in midlife women has been linked to BMC declines. Highly variable testosterone continuing post-FMP could point to disruptions in the reproductive system beyond the expected fluctuations during the MT, which could result in additional bone loss. This finding suggests that for post-menopausal women, testosterone variability may be a useful marker for bone loss risk, although we recommend additional research to validate these findings.

Remark In our proposed model, the mean and covariance structures are inextricably related. Allowing for a highly flexible mean structure for the individual trajectories increases the difficulty of estimating patterns in the variances and covariances. Similarly, specifying a highly flexible structure for the covariance regression impacts the estimation of the individual mean trajectories, which are assumed to be not overly complex. Thus, a limitation to the model is that the mean structure and covariance structure cannot not both be exceedingly flexible. Generally, the previous literature on covariance regression (Hoff and Niu, 2012; Li et al., 2014) also specify a more flexible mean structure (e.g. larger number of knots or larger number of covariates) and a more parsimonious covariance structure in their applications. We recommend that the decision regarding mean and covariance specifications should be discussed with scientific collaborators who can provide insight into the suitability of different characterizations for the specific application context.

4.5.1 Future Work

Our work has some limitations. After deriving our analytic dataset, only 211 women remained. This meant that our sample size was relatively small. Since the MBHS only enrolled white women, our results may not generalize to other ethnic groups of women. Future analysis could use datasets such as the Study of Women's Health Across the Nation, which is a multi-ethnic, multi-site study that collects hormone and bone measurements.

Methodologically, our model assumes that the individual covariance regression coefficients are independent for each marker. For our application, this means that the individual FSH variance trend at a given point in time is modeled independently of the individual testosterone variance trend at the same point in time. For highly correlated markers, this assumption of independence may be too

strong. We do note that in the Bayesian framework, correlations between the individual coefficients will be induced during the posterior sampling, even if *a priori* we have assumed independence. Nontheless, a future extension to this could be to relax the independence assumption and allow the individual coefficients to be correlated between markers. In the case of highly correlated markers, this assumption of independent regression coefficients may be too strong. A future extension to this proposed method could be to allow for dependency between the individual variance coefficients. This would likely increase the computational time necessary for estimating the full likelihood.

In Chapter 3, we evaluated the univariate associations between E2 and FSH and BMD. In this chapter, we have examined the association between FSH and testosterone. A logical next step could be to model E2, FSH, and testosterone together in the predictor model. This would mean estimating a 3×3 variance covariance matrix at each timepoint and the computational power needed to estimate the time-varying covariance matrices would also increases nonlinearly. Specifying latent classes of individuals (Jiang et al., 2015) may ease computational burden, since individuals within the same latent class could share the same mean and variance regression coefficients (i.e. a manifest model), rather than needing to estimate each individual's coefficients for the mean and covariance regressions. This could also provide insight into different subgroups of, characterized by different evolutions of their means and variance trajectories, e.g. women with unchanging variabilities after FMP or women who have increasing variabilities after FMP.

In our application, the dimension of the individual covariance matrices is small because we focused on two biomarkers. However, if the number of available biomarkers becomes larger, maintaining interpretability of the covariances in the outcome model may become extremely difficult, if not impossible. In this case, an interesting extension to the joint model presented here could be to explore tensor regression or other methods for high dimensional matrix regression in the outcome model. Incorporating shrinkage priors in the tensor regression, such as those proposed by (Guhaniyogi et al., 2017), might reduce the overall burden of estimating the outcome model in this setting.

CHAPTER 5

Discussion and Future Work

5.1 Discussion

Emerging technologies have made it easier for biomedical and epidemiological to collect biomarker data at the individual level. Advances in precision medicine rely on the study of such biomarkers that can predict disease prognoses and risks (Coffman and Richmond-Bryant, 2015; Talukdar et al., 2023). Researchers are recognizing that intra-individual variability of repeatedly measured biomarkers is a potentially powerful predictor for health risks (Goldenberg et al., 2019; Barrett et al., 2019; Parker et al., 2021; Martins, 2022). In order to explore these associations between individual variances and health outcomes, we need to develop models that can capture these individual variances and utilize them as predictors. In the multivariate biomarker setting, statistical methods that capture the individual covariance structures between the predictor markers would be particularly useful, since these correlations could also be predictive of disease outcomes.

The joint models presented in this dissertation contribute to this by explicitly modeling the marker heterogeneity at the individual level, and then using these estimates to predict of health outcomes. We have demonstrated in simulation studies that these models can recover the parameters of interest with low bias and high coverage, and that these models outperform two-stage competitor approaches. Although this dissertation focused on applications in women's health, these models can serve as general tools for researchers who want to better understand the complex associations between longitudinal biomarker predictors and corresponding health outcomes.

Multivariate analysis of biomedical markers is important for understanding their complex interactions, and how these interactions can predict health outcomes. We are particularly motivated by questions regarding how the variabilities and covariabilities of certain reproductive hormone markers can predict health outcomes for women during the midlife. We used data from the SWAN and MBHS studies to examine the associations between E2, FSH, and testosterone variances (and covariances) and body mass and bone health outcomes in peri- and post-menopausal women. Although the applications in this dissertation focused on women's midlife health outcomes, the joint

models presented here are designed to be broadly applicable to open research questions where individual-level variabilities could be used as predictors.

In Chapter 2, we proposed a method for estimating the individual variances and covariances of multiple markers in order to predict a cross-sectional health outcome. Each individual is given their own variance-covariance matrix, rather than assuming a common population-level covariance matrix across all individuals. Estimation of the individual covariance matrices is efficiently carried out by using a decomposition method that separately estimates the standard deviations and residual correlations. As expected, our joint model outperformed two-stage competitors by achieving lower bias, higher coverage, while maintaining appropriate estimation uncertainty of the pre-specified data generating parameters. Surprisingly, we found that larger individual E2 variances on average, were associated with slower increases in waist circumference, suggesting that E2 may drive waist adiposity changes more than overall fat mass changes. We believe that this finding can serve as a starting point for further investigations into the role of E2 variability in predicting fat mass composition changes as women grow older.

In Chapter 3, we are motivated by the setting of repeatedly measured health outcomes. Our main application of interest is to relate individual trajectories and variances of a hormone biomarker to predict declines in BMD as women transition through menopause. The joint model further allows for the individual residual variances to change over time, by modeling these variances with a regression function. Simulation studies compared our joint model to other two-stage methods, and also highlighted different scenarios to evaluate the relative performances of of the time-varying variance parameterization over a time-invariant one. We found that the time-invariant variance model had similar coverage to the time-varying model when the individual variance trends were smaller in magnitude, but that it had noticeably poorer coverage when there were larger individual variance trends. In our application, we found that higher time-invariant FSH variability was strongly associated with slower declines in BMD. This suggested that women with higher variability in their FSH measurements were more likely to have higher BMD values, holding all else equal. These associations were strongest after FMP; pre-FMP, the predicted BMD trajectories at different quantile estimates of FSH variability were predicted to converge. This suggested that higher FSH variability may be more strongly associated with higher rates of bone loss after the FMP. These results support the hypothesis that FSH, rather than E2, is more closely associated with bone loss in midlife women, and, for the first time, quantify a relationship between FSH variability and bone outcomes in women during the MT.

Chapter 4 extended Chapter 3's univariate marker setting to the multivariate marker setting by studying the evolution over time of variances and covariances from multiple markers. Our motivating scientific question was to understand if the variabilities of women's hormones also varied over time across the MT, and also if these changing variabilities could predict declines in BMC. We

drew upon established work in the area of covariance regression by extending this idea to the joint modeling setting. The proposed method jointly modeled time-varying variances and covariances of multiple markers, by estimating individual-level regression coefficients for these trajectories. Our analysis provided new insights into the associations between FSH and testosterone variances and covariances and women's BMC trajectories. We found that higher testosterone variability after FMP was associated with faster declines in BMC, a result that had not yet been established in the current bone literature regarding midlife women. Furthermore, we found that when FSH and testosterone were more highly correlated after FMP, there were slower declines in women's BMC. A simulation study provided a sanity check for our model by establishing that our model was able to consistently recover the model parameters with low bias.

5.2 Future Work

There are several areas of future work that can be done. One natural extension of the presented models would be to explicitly model the missing data mechanism (e.g. Missing at Random). Longitudinal data frequently suffers from missing data problems that are caused by a variety of factors, such as patient refusal for measurements or medical equipment malfunction. The SWAN study, and to a lesser extent the MBHS study, are not exempt from missing data issues. However, since the relative number of missing data in these two datasets were small ($< 5\%$ in both datasets), our analysis was performed on complete cases. Since our models use a Bayesian approach to estimating the parameters of interest, we could utilize methods like multiple imputation to account for missingness in the biomarker predictors. Previous work on joint modeling with missing longitudinal data (Chen et al., 2014; Takeda et al., 2022) can serve as guidance for integrating missing data estimation methods within our model frameworks.

A fundamental and active research area in multivariate analysis is the efficient estimation of covariance matrices, particularly as the number of variables increases. In Chapter 2, we found that Barnard et al. (2000)'s decomposition method was sufficiently fast for two and three biomarkers; however, we would expect the computational cost of estimating the covariance matrices to increase non-linearly with each additional marker. Additionally, the time-varying covariance regression method in Chapter 4 further increases the potential runtime needed to fit the joint model. Bringing in methods for efficient covariance matrix estimation into the joint modeling setting would be particularly useful for handling higher-dimensional biomarker datasets.

Latent class models could alleviate the computational costs of estimating individual level covariance matrices. Extending our models to the latent class setting would also allow us to study trajectories of means and variances across latent subgroups of individuals. For example, we could assign women to different mean and variance trajectory classes. Within each class, women could

share the same mean and variance coefficients. This would mean that we only need to estimate coefficients for each class rather than coefficients for each individual. For larger datasets, this would greatly reduce the computational burden and also provide insight into different subgroups of patients, based on their means and variances.

With regards to women's health, our findings have significant implications for advancing women's health both on the individual level and on a broader scientific level. Our findings on hormone variabilities and co-variabilities as markers for health outcomes should motivate new research questions about the underlying scientific mechanisms that may drive these patterns. Since the predictive power of women's intra-individual hormone variabilities have not been widely modeled yet, we do not yet have a comprehensive understanding of what may be driving these associations. Future research that delves deeper into the biological mechanisms that may drive these variances, and their associations with fat mass and bone loss changes during the MT, would give us a better understanding of these complex dynamics. This would be a meaningful opportunity to further establish the literature on variances as predictors of women's health outcomes. These investigations would also enhance our ability to deliver effective individualized treatments and interventions for midlife women.

APPENDIX A

Appendix 01

A.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:

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

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 (A.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:

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

and suppose we measure X_1, X_2 with some error:

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

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

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

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, 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 Chapters 2.4 and 3.3.1 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 (and/or covariability) is actually predictive of an outcome of interest.

APPENDIX B

Appendix 02

B.1 Visualization of the Joint Model

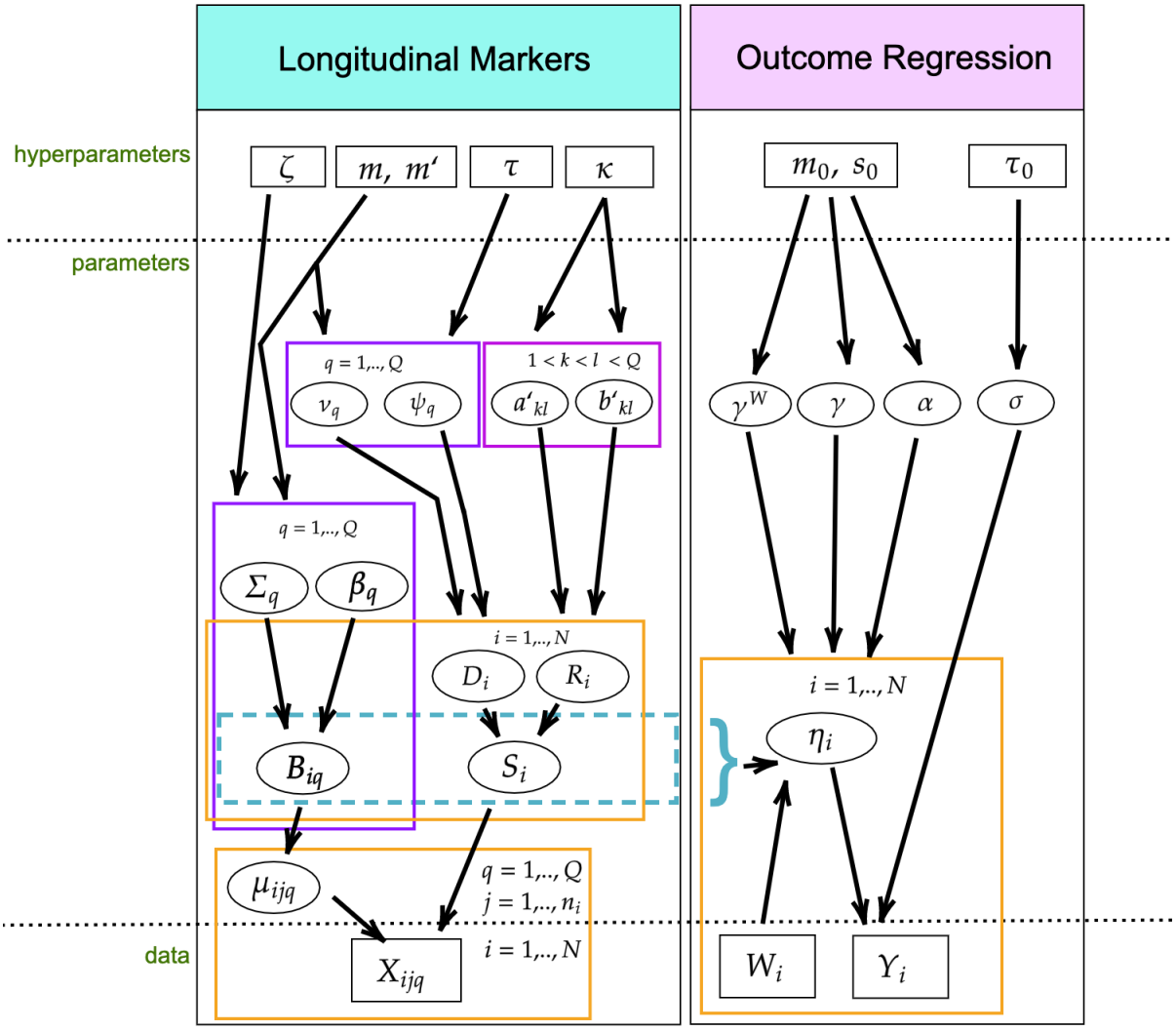


Figure B.1: **A visualization of the relationships between the model parameters and data. This directed graph shows the hierarchical form of our model framework. The quantities in squares are either data or hyperparameters; the unknown quantities are displayed in circles. The arrows connecting variables indicate that the parent parameterizes the distribution of the child node. The rectangular "plates" that enclose variables indicate that a similar graphical structure is repeated over the index. The index in a plate indicates nodes, hyperparameter levels and subjects.**

B.2 Posterior Predictive Model Checking

To assess our model's validity on the SWAN data, we conduct posterior predictive checks for both the trajectories submodel and the outcome submodel.

For the outcome submodel, we generated simulated data from the posterior predictive distribu-

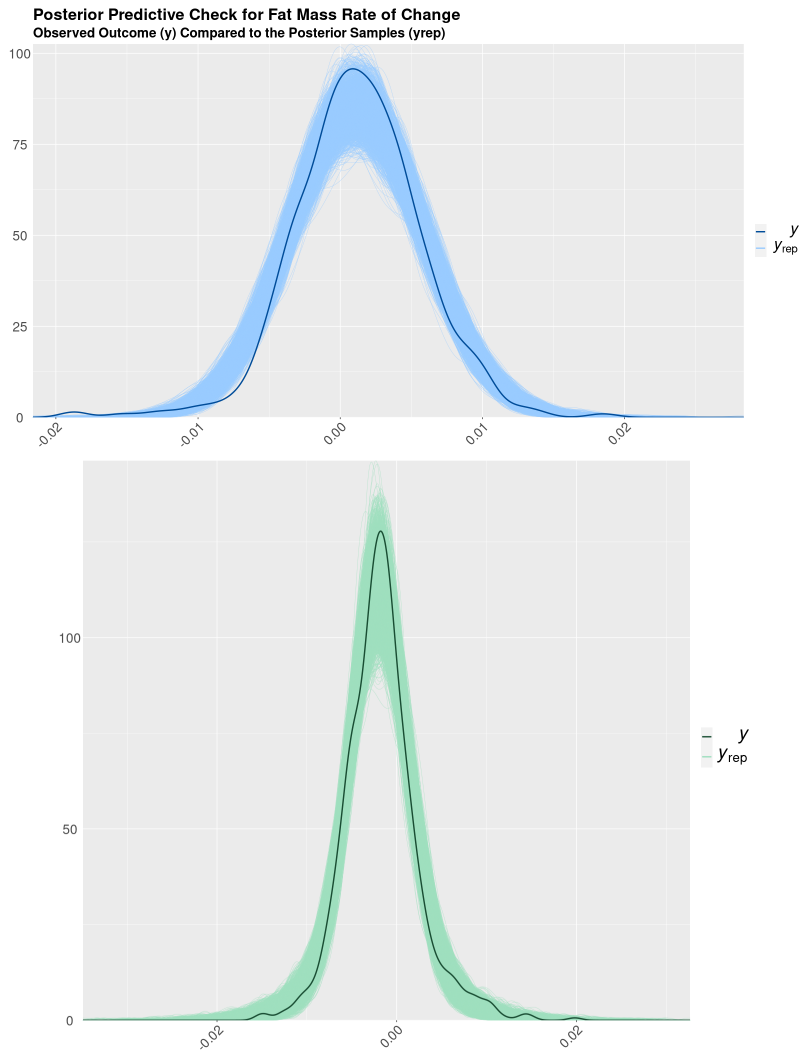


Figure B.2: **Visualizations of the posterior predictive checks performed for the fat mass rate of change (top) and lean mass rate of change (bottom). The observed outcomes (y) are represented by the solid lines and the model-generated outcomes (yrep) are represented by the thin semi-opaque lines. We see that the model-generated outcomes cover the observed outcomes for both models, indicating that our model is generating reasonable estimates of the outcomes.**

tion. The posterior predictive distribution for the predicted outcome, $\widetilde{Y}$ can be written as:

$$p(\widetilde{Y}|Y) = \int p(\widetilde{Y}|\theta, X)p(\theta, X|Y)d\theta dX$$

where θ are the unknown model parameters and X are the predictor variables used in the outcome regression. For each draw of the model parameters from the posterior distribution, $p(\theta|Y, X)$, we can draw a vector $\widetilde{Y}$ from the posterior predictive distribution by conditioning on the draw of the model parameters and then simulating from the data model (Gabry et al., 2019).

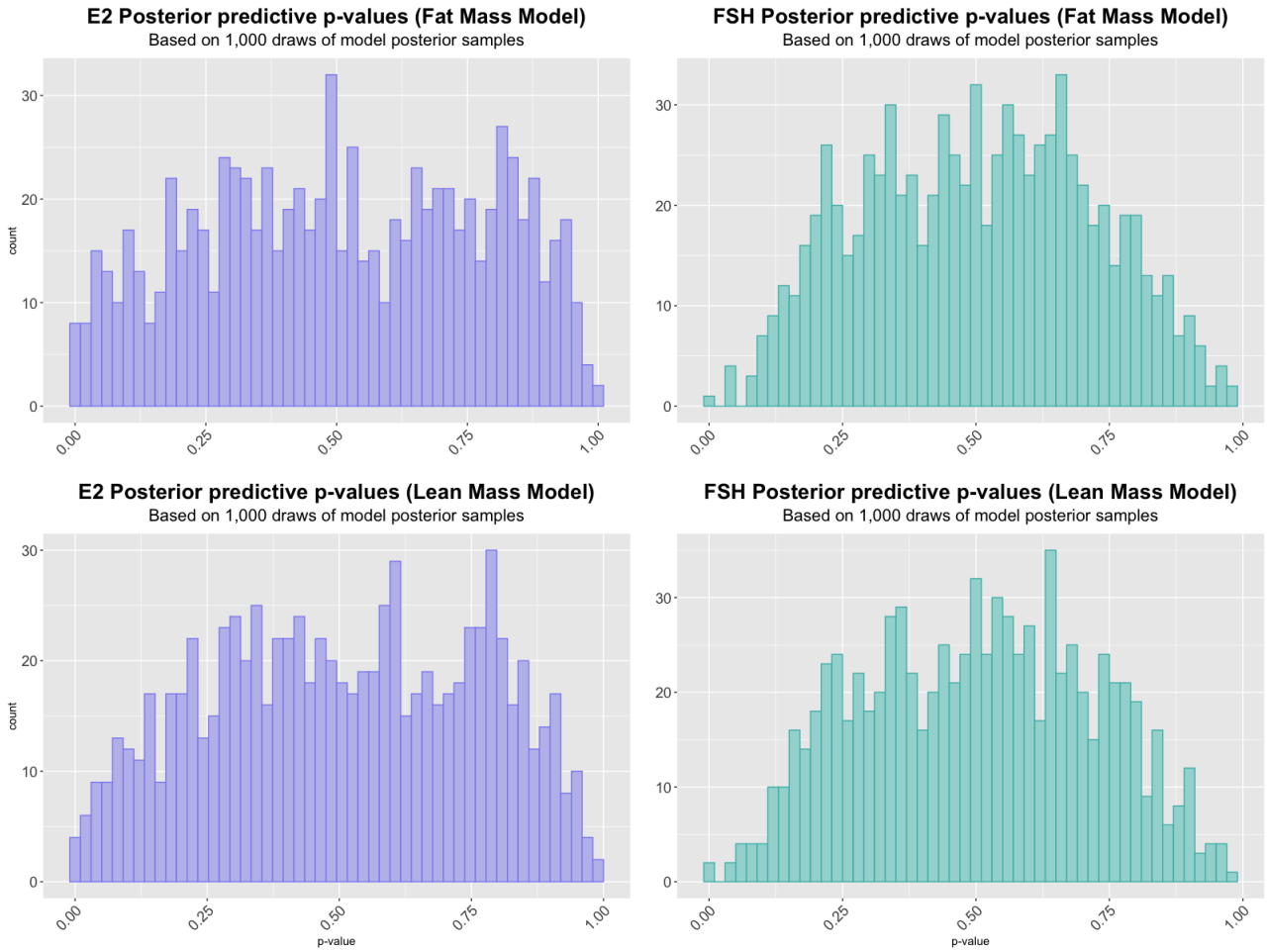


Figure B.3: **Posterior predictive check of E2, FSH trajectories across all individuals for both the fat mass and lean mass models. The median p-value for each 1,000 draws of posterior samples was 0.5.**

We then plotted 1,000 draws of this model-generated data against the observed outcome, which is shown in Figure B.2. For both the fat mass rate of change and the lean mass rate of change models, we see that the simulated replicated data from the model overlap the observed data, indicating that our model is producing reasonable predictions.

For the trajectories submodel, we define the following statistic: $T(x_{itq}; b_{iqp}, t)_q = \sum_t (x_{itq} - \mu(b_{iqp}, t))^2/(\sigma^2_{iq}))$ where $\mu(b_{iqp}, t))$ is the estimated individual i's mean trajectory for hormone q and σ^2_{iq} is the estimated variance of individual i's trajectory for hormone q. By doing this, we can compare $T_i(x^{obs}_{itq}; b_{iqp}, t)_q$ (which is a function of the observed data and the estimated parameters) with $T_i(x^{sim}_{itq}; b_{ipq}, t)_q$ (which is a function of the model generated data using the model estimated parameters). If there are large discrepancies between $T_i(x^{obs}_{itq}; b_{iqp}, t, \sigma^2_{iq})_q$ and $T_i(x^{sim}_{itq}; b_{iqp}, t, \sigma^2_{iq})_q$, this could indicate poor model fit (Gelman et al., 2013).

One way to compare these two $T(x_{itq})$ statistics is to compute the 'posterior predictive p-value',

which is $P(T_i(x_{itq}^{obs}; b_{iqp}, t, \sigma_{iq}^2)_q < T_i(x_{itq}^{sim}; b_{iqp}, t, \sigma_{iq}^2)_q | (x_{itq}^{obs}))$. For E2 and FSH, we keep (x_{it}^{obs}) fixed at the observed values and compute 1,000 values of $T_i(x_{itq}^{sim}; b_{iqp}, t)$ from the posterior of $b_{iqp}, t, \sigma_{iq}^2$. We then compare these values with 1,000 draws from $T_i(x_{itq}^{sim}; b_{iqp}, t, \sigma_{iq}^2)_q$. Figure B.3 displays the histograms of the resulting p-values for each individual's hormone trajectory for the two models. Across all of the hormones, most of the the computed p-values were between 0.25 and 0.75. Further analysis of the p-values across the quantiles of the distribution shows that the generated data from the model reasonably captures the individual trends. This provides justification that both the trajectories submodel and the outcome submodel are good fits for the data.

B.2.1 Traceplots

We present traceplots of the predictor variables for both the fat mass (Figure B.4) and waist circumference models (Figure B.5). Traceplots of the other model parameters are available upon request.

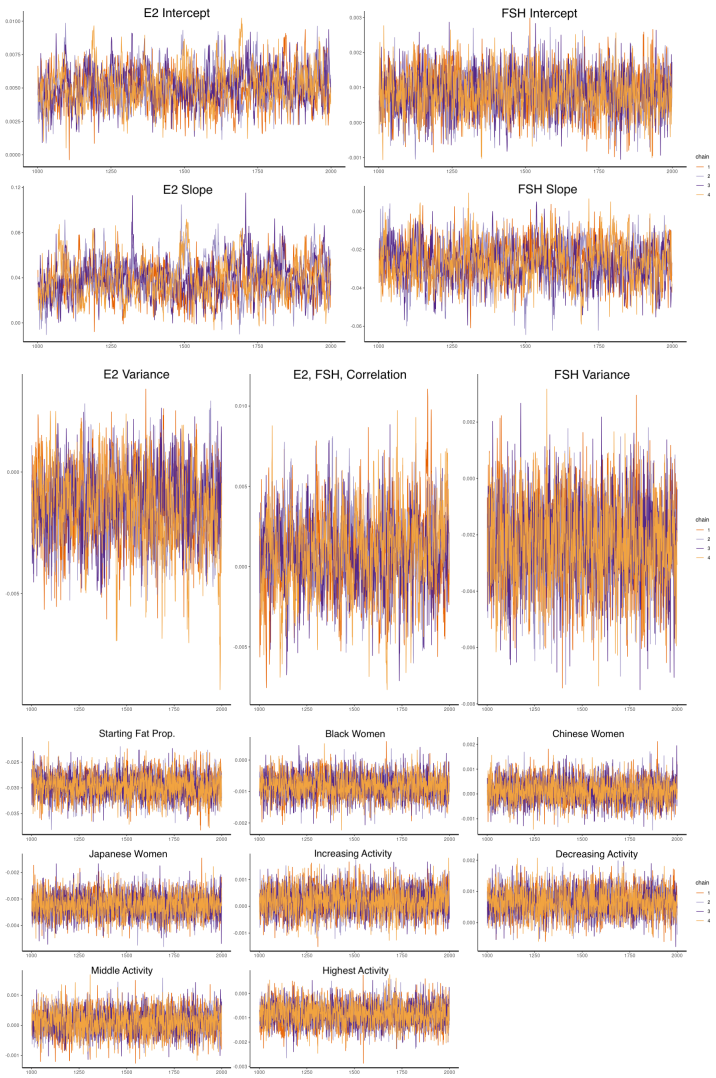


Figure B.4: **Traceplots for the mean parameters (top figure), variance parameters (middle figure) and adjusted covariates (bottom figure) in the fat mass rate of change model.**

B.3 Data Analysis: Posterior Means and 95% CrIs for Additional Model Parameters

In this section, we present the estimated posterior means and 95% credible intervals for the other parameters from the data application from Section 5 in the main text.

Parameter	Post. Mean	2.5% CrI	97.5% CrI
β_{11}	-35.82	-56.39	-14.80
β_{12}	3.46	0.34	6.63
β_{21}	26.29	4.36	48.12
β_{22}	-6.44	-8.82	-3.99
$\Sigma_1[1,1]$	47.31	38.47	56.86
$\Sigma_1[1,2]$	-1.49	-2.38	0.61
$\Sigma_1[2,2]$	-0.51	0.34	0.70
$\Sigma_2[1,1]$	80.44	70.60	91.12
$\Sigma_2[1,2]$	0.96	-0.68	0.84
$\Sigma_2[2,2]$	0.56	0.45	0.68
ν_1	-286	-304	268
ν_2	-286	-724	-677
ξ_1	111	81.74	140
ξ_2	245	223	266
α_1*	8.55	6.47	11.53
β_1*	32.14	24.05	43.60
σ	3.71	3.35	4.03

Table B.1: **Evaluation of the posterior means and 95% CrI estimates for the other parameters in the fat mass rate of change model. All values except for α_1, β_1 (indicated with asterisk) have been multiplied by 10^3. α_1, β_1 have been presented in their original values.**

Parameter	Post. Mean	2.5% CrI	97.5% CrI
β_{11}	-35.84	-57.46	-14.93
β_{12}	3.80	0.71	6.89
β_{21}	26.07	4.23	48.50
β_{22}	-6.53	-4.23	-4.01
$\Sigma_1[1,1]$	47.22	38.36	56.93
$\Sigma_1[1,2]$	-1.51	-2.45	-0.62
$\Sigma_1[2,2]$	-0.51	-0.36	0.70
$\Sigma_2[1,1]$	80.46	70.93	90.73
$\Sigma_2[1,2]$	-0.10	-0.67	-0.87
$\Sigma_2[2,2]$	0.56	0.45	0.68
ν_1	-0.29	-0.31	-0.27
ν_2	-0.70	-0.73	-0.68
ξ_1	0.14	0.11	0.16
ξ_2	0.25	0.22	0.27
α_1*	8.70	6.57	11.70
β_1*	32.66	24.41	44.61
σ_1	2.53	1.95	3.02
σ_2	5.87	4.70	7.63
Π_1*	0.73	0.52	0.90

Table B.2: **Evaluation of of the posterior means and 95% CrI estimates for the other model parameters in the lean mass rate of change model. All values except for α_1, β_1, Π_1 (indicated with asterisk) have been multiplied by 10^3. α_1, β_1, Π_1 have been presented in their original values.**

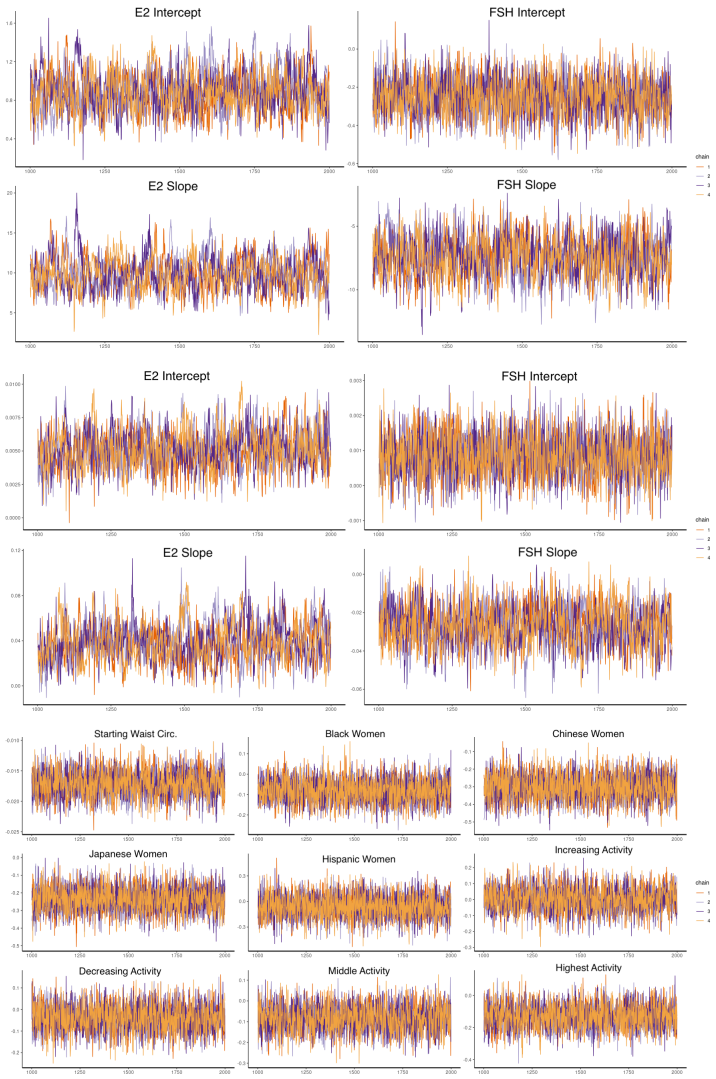


Figure B.5: **Traceplots for the mean parameters (top figure), variance parameters (middle figure) and adjusted covariates (bottom figure) in the waist circumference rate of change model.**

B.4 Other Simulation Replicates for Two Biomarker Simulation Study and Three Biomarker Simulation Study

In this section, we present the bias, coverage and average interval length for the other JMIV model parameters from running 200 simulation replicates. For each simulation replicate, we ran two chains with 2,000 steps and 1,000 burn in. The data generation parameters are detailed in Section 4.1 and 4.2 of the main text.

Parameter	Truth	Average Post. Mean	Bias	Coverage %	Average Interval Length
β_{11}	0	0.00	0.00	95.5	0.14
β_{12}	2	2.00	0.00	95.0	0.13
β_{21}	2	2.00	0.01	95.5	0.15
β_{22}	1	1.00	0.00	95.5	0.09
$\Sigma_1[1,1]$	1	1.00	0.00	95.0	0.22
$\Sigma_1[1,2]$	-0.05	-0.05	0.00	96.0	0.14
$\Sigma_1[2,2]$	1.01	1.00	0.01	94.0	0.19
$\Sigma_2[1,1]$	1	1.00	0.00	96.0	0.24
$\Sigma_2[1,2]$	-0.1	-0.10	0.00	97.5	0.10
$\Sigma_2[2,2]$	0.5	0.50	0.00	95.0	0.10
ν_1	0	0.00	0.00	95.5	0.06
ν_2	0.25	0.25	0.00	97.0	0.04
ξ_1	0.38	0.37	0.00	95.0	0.05
ξ_2	0.25	0.25	0.00	97.0	0.04
a_1	1	1.01	0.01	96.0	0.01
b_1	5	5.06	0.06	96.0	1.32

Table B.3: **Two Trajectory Simulation Setting: Evaluation of bias, coverage, and 95% credible interval length across 200 simulation replicates for the B_i and S_i parameters for the JMIV model. Our model achieves $>$ 90% coverage across all parameters and maintains low bias.**

B.5 Simulation 3: Linear Approximation of Nonlinearity

In this simulation study, we study how well our model performs when the true relationship between some of the longitudinal means and variances terms and the cross-sectional outcome is nonlinear, but we approximate this relationship with a linear form.

Step 1: Estimate Linear Approximation Coefficients

We use the same data generation parameters for the longitudinal markers as in Section 4.1 of the main text. For the outcome model, we generate the mean $\eta(\boldsymbol{B}_i, \boldsymbol{S}_i)$ as:

$$\eta(\mathbf{B}_i, \mathbf{S}_i) = 2b_{i11} + b_{i12} - b_{i21} + 0.5b_{i22} + 2s_{i11} - s_{i21} + 2s_{i22} + 0.5b_{i21}^2 + 0.75s_{i11}^2$$

so that the individual slope of the first biomarker (b_{i12}) and the variance of the first biomarker (s_{i11}) are quadratically related to the outcome.

To estimate the "linear approximation" coefficients, we simulate data for 1 million individuals and generate the outcome data as:

$$Y_i \sim \mathcal{N}(\eta(\mathbf{B}_i, \mathbf{S}_i), 0.01)$$

Parameter	Truth	Average Post. Mean	Bias	Coverage %	Average Interval Length
β_{11}	0	0.00	0.00	95.0	0.14
β_{12}	2	1.99	0.00	93.5	0.35
β_{21}	2	1.99	0.00	96.0	0.14
β_{22}	1	0.99	0.00	95.5	0.09
β_{31}	1	1.99	0.00	93.0	0.14
β_{32}	1	1.00	0.00	95.5	0.13
$\Sigma_1[1,1]$	1	1.00	0.00	94.0	0.21
$\Sigma_1[1,2]$	-0.05	-0.05	0.00	97.0	0.14
$\Sigma_1[2,2]$	1.00	0.99	0.01	96.0	0.18
$\Sigma_2[1,1]$	1	1.00	0.00	96.5	0.22
$\Sigma_2[1,2]$	-0.1	-0.10	0.00	93.5	0.10
$\Sigma_2[2,2]$	0.5	0.50	0.00	97.5	0.09
$\Sigma_3[1,1]$	1	1.00	0.00	96.0	0.21
$\Sigma_3[1,2]$	-0.25	-0.25	0.00	94.5	0.14
$\Sigma_3[2,2]$	1	1.00	0.00	92.0	0.18
ν_1	0.00	0.00	0.00	94.0	0.06
ν_2	0.25	0.25	0.00	95.5	0.04
ν_3	0.25	0.25	0.00	94.5	0.07
ξ_1	0.375	0.38	0.00	97.0	0.06
ξ_2	0.25	0.25	0.00	94.5	0.04
ξ_3	0.25	0.25	0.00	94.5	0.07
a_{12}	1	1.01	0.01	96.0	0.19
a_{13}	1	1.01	0.01	94.5	0.19
a_{23}	2	2.04	0.04	93.5	0.50
b_{12}	5	5.09	0.09	96.5	1.29
b_{13}	5	5.07	0.07	92.5	1.29
b_{23}	2	2.05	0.05	94.0	0.50

Table B.4: **Three Trajectory Simulation Setting: Evaluation of bias, coverage, and 95% credible interval length across 200 simulation replicates for the B_i and S_i parameters for the JMIV model. Our model achieves $> 90\%$ coverage across all parameters and maintains low bias.**

We then fit a linear approximation using the `lm()` function in R and the following model:

$$Y_i = \alpha_{11} * b_{i11} + \alpha_{12} * b_{i12} + \alpha_{21} * b_{i21} + \alpha_{22} * b_{i22} + \gamma_{11} * s_{i11} + \gamma_{21} * s_{i21} + \gamma_{22} * s_{i22}$$

and collect the estimated coefficients. These coefficients are the "target" coefficients that we want to approximate. These targets are shown in Table B.5 as the "truth" values.

Step 2: Joint Model Simulation Replicates

We follow the same data generation in Step 1 for the longitudinal markers. For the outcome model, we generate the data as:

$$Y_i \sim \mathcal{N}(\eta(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{B}_i, \mathbf{S}_i), 0.5)$$

After generating this data, we apply our joint model, but modeled with the same linear mean function, $\eta(\mathbf{B}_i, \mathbf{S}_i)$, as in Section 4 of main text and collect the estimated coefficients. We do this for 200 independent replicates and evaluate model performance using the same criteria (bias, coverage, and average 95% CrI length) as in the previous simulation studies. Table B.5 displays the results for the outcome mean and variance parameters. We find that the model maintains low bias and high coverage of the truth (¿ 90% coverage). This indicates that our model can recover the estimated parameters from a linear approximation of the model, when the true form of the outcome mean may be nonlinear.

Truth	Bias	Coverage (%)	Average Interval Length
α_{11} = 1.99	0.02	93.50	0.41
α_{12} = 1.44	0.03	93.00	0.34
α_{21} = -1.06	0.00	94.58	0.38
α_{22} = 0.56	-0.03	94.50	0.51
γ_{11} = 1.97	-0.01	96.00	0.69
γ_{12} = -0.99	0.03	94.50	1.17
γ_{22} = -2.06	-0.03	92.50	0.58

Table B.5: **Simulation III: bias, coverage, and 95% credible interval (or confidence interval) length across 200 simulation replicates. With the linear approximation, our model maintains low bias and high coverage of the true (linear approximating) parameters.**

APPENDIX C

Appendix 03

C.1 Posterior Predictive Model Checking

To assess our model's validity on the SWAN data, we conduct posterior predictive checks, using the same procedures described in Appendix B.2. For both models (Figure C.1.), we see that the simulated replicated data from the model overlap the observed data, indicating that our model is producing reasonable predictions.

For the longitudinal biomarker predictor submodel, we use $T(x)$ statistic described in Appendix B.2.

Figure C.2 displays the histograms of the resulting p-values for each individual's E2 trajectory. The majority of the computed p-values were between 0.25 and 0.75. Further analysis of the p-values across the quantiles of the distribution shows that the generated data from the model reasonably captures the individual trends. This provides justification that both the predictor submodel and the outcome submodel are good fits for the data.

C.2 Data Analysis: Posterior Means and 95% CrIs for Additional Model Parameters

In this section, we present the estimated posterior means and 95% credible intervals for the other parameters from the BMD outcome models in Section 4 of the main text. All R-hat values for these parameters were < 1.01, indicating the models had converged.

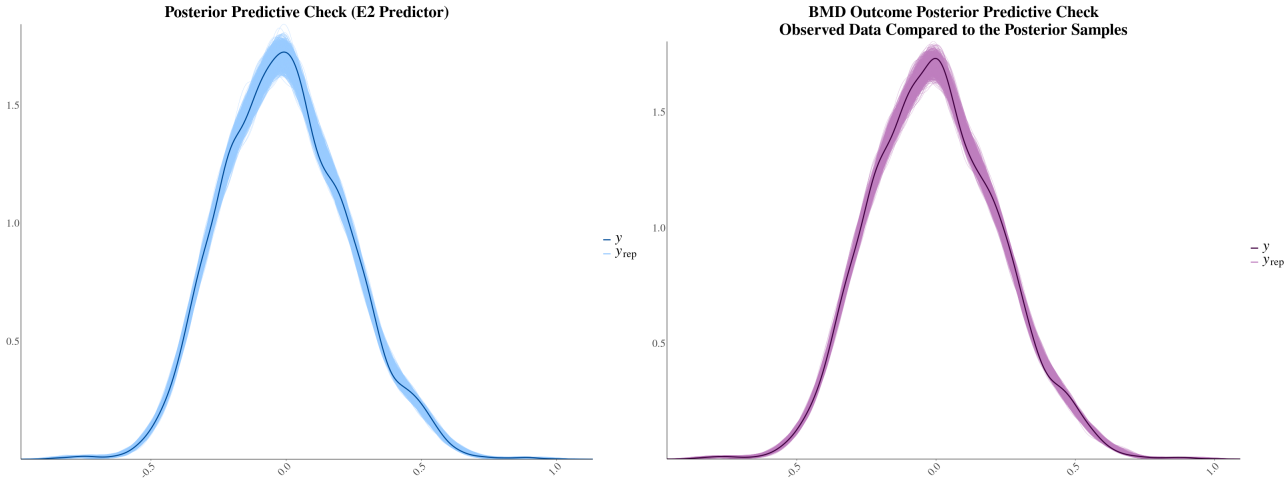
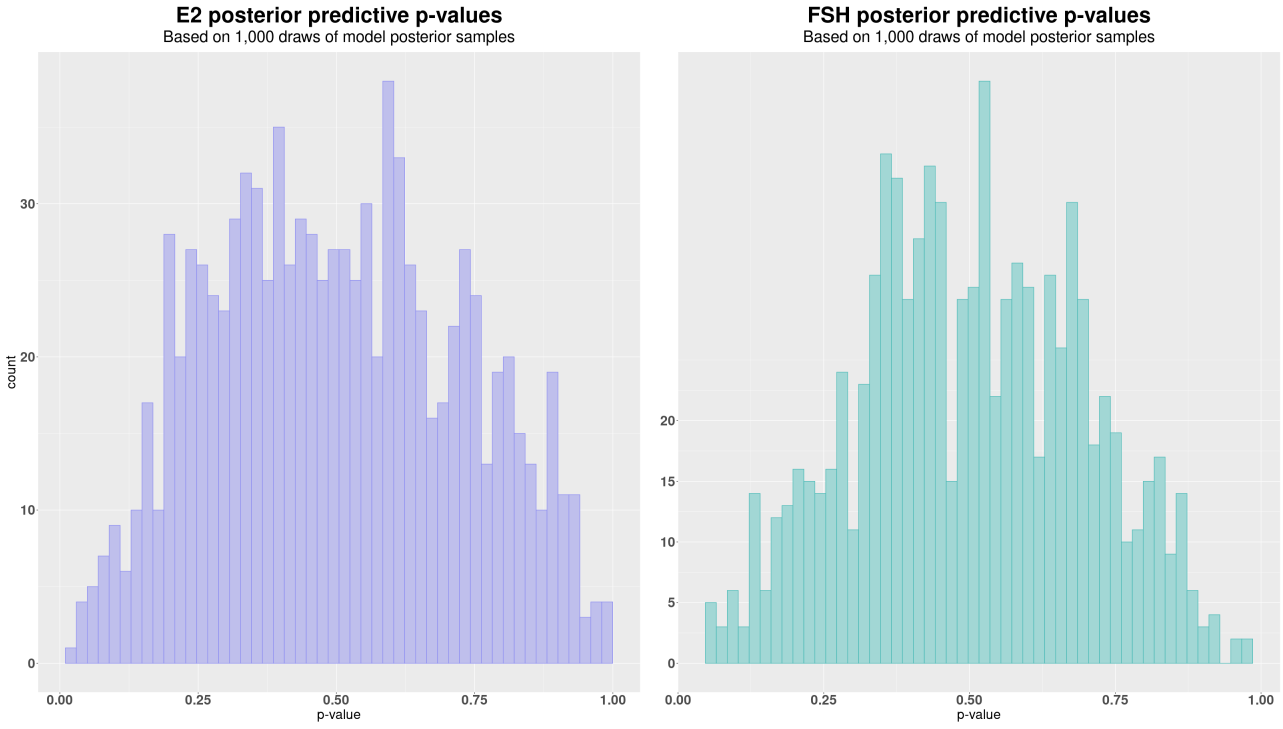


Figure C.1: **Visualization of the posterior predictive checks performed for the E2 predictor model (left) and the FSH predictor model (right). The observed outcomes (y) are plotted with the solid lines and the model-generated outcomes (yrep) are plotted with the thin semi-opaque lines. These plots show that the model-generated outcomes cover the observed BMD residuals, indicating that our model is generating reasonable estimates of the outcomes.**

Figure C.2: **Visualization of the computed p-values performed for the BMD outcome model. The distributions of the p-values indicate that our model was able to fit the hormone residual data reasonably well.**



Parameter	Post. Mean	2.5% CrI	97.5% CrI
α_1	-0.03	-0.06	0.00
α_2	0.02	0.01	0.02
$\Sigma[1,1]$	0.09	0.07	0.11
$\Sigma[1,2]$	0.00	0.00	0.00
$\Sigma[2,2]$	0.00	0.00	0.00
ν	0.16	0.11	0.20
ξ	0.38	0.34	0.43
σ	0.04	0.04	0.04
τ	0.19	0.18	0.19
β_0	0.04	-0.02	0.09

Table C.1: **Evaluation of the posterior means and 95% CrI estimates for the other parameters in the E2-BMD outcome model. All posterior means and 95% CrIs are presented in their original values.**

Parameter	Post. Mean	2.5% CrI	97.5% CrI
α_1	0.02	-0.01	0.06
α_2	0.00	0.00	0.00
$\Sigma[1,1]$	0.16	0.13	0.19
$\Sigma[1,2]$	0.00	0.00	0.00
$\Sigma[2,2]$	0.00	0.00	0.00
ν	-0.55	-0.61	-0.50
ξ	0.56	0.51	0.61
σ	0.04	0.04	0.04
τ	0.19	0.18	0.20
β_0	0.11	-3.54	3.77

Table C.2: **Evaluation of the posterior means and 95% CrI estimates for the other parameters in the FSH-BMD outcome model. All posterior means and 95% CrIs are presented in their original values.**

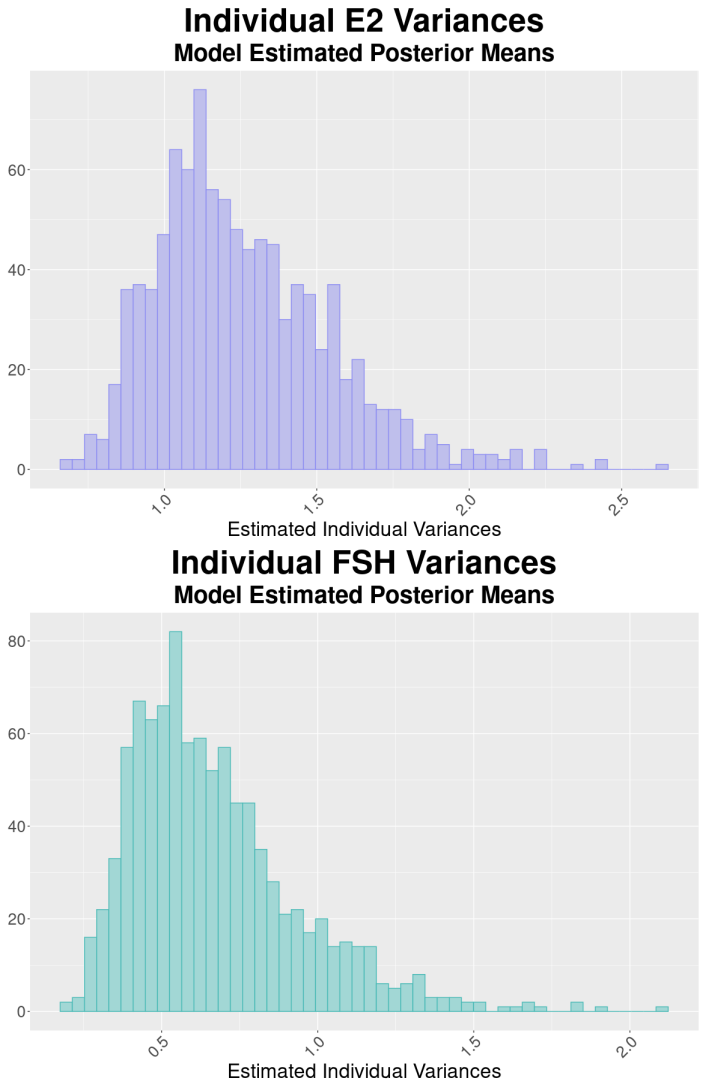


Figure C.3: **Histograms of estimated E2 individual variances and estimated FSH individual variances. The estimated mean E2 variance was 1.21 and the estimated mean FSH variance was 0.67.**

Parameter	Truth	Average Post. Mean	Bias	Coverage %	Average Interval Length
α_1	0	0.00	0.00	95.5	0.14
α_2	2	2.00	0.00	95.0	0.13
$= \Sigma[1,1]$	1	1.00	0.00	95.0	0.22
$\Sigma[1,2]$	-0.05	-0.05	0.00	96.0	0.14
$\Sigma[2,2]$	1.01	1.00	0.01	94.0	0.19
ν	0	0.00	0.00	95.5	0.06
ξ	0.38	0.37	0.00	95.0	0.05
τ	1	1.01	0.01	96.0	0.01
σ	5	5.06	0.06	96.0	1.32

Table C.3: **Simulation Study 1: Evaluation of the JELO model bias, coverage, and 95% credible interval length across 200 simulation replicates.**

Parameter	Truth	Bias	Coverage %	Average Interval Length
α_1	0	0.00	91.0	0.19
α_2	-2	0.00	95.0	0.17
α_{s1}	0	0.00	95.5	0.14
α_{s2}	2	0.00	95.0	0.13
$\Sigma[1,1]$	1	0.00	96.0	0.26
$\Sigma[1,2]$	-0.25	0.00	93.0	0.14
$\Sigma[2,2]$	0.5	0.00	93.5	0.14
$\Sigma_S[1,1]$	1	0.00	93.0	0.47
$\Sigma_S[1,2]$	0.1	-0.01	93.0	0.37
$\Sigma_S[2,2]$	0.5	0.00	95.0	0.52
τ	0.25	0.00	96.0	0.07
σ	0.1	0.00	92.0	0.01

Table C.4: **Simulation Study 2 (High Variance Setting): Evaluation of the JELO (with time-varying variance) model bias, coverage, and 95% credible interval length across 200 simulation replicates.**

C.3 Simulation Studies: Additional Parameters

This section contains the bias, coverage, and average interval length statistics for the other JELO model parameters. For each simulation replicate, we ran two chains with 2,000 steps and 1,000 burn in. The data generation parameters are detailed in Section of the main text. Table C.3 contains the estimates for the constant variance simulation study and Tables C.4 and C.5 contains the estimates for our time-varying variance simulation study.

Parameter	Truth	Bias	Coverage %	Average Interval Length
α_1	0	0.00	91.0	0.20
α_2	-2	0.00	93.0	0.20
α_{s1}	0	0.00	95.5	0.14
α_{s2}	0	0.00	95.0	0.13
$\Sigma[1,1]$	1	0.00	95.5	0.28
$\Sigma[1,2]$	-0.25	0.00	92.0	0.15
$\Sigma[2,2]$	0.5	0.00	93.5	0.15
$\Sigma_S[1,1]$	0.5	0.01	96.0	0.28
$\Sigma_S[1,2]$	-0.01	0.02	93.0	0.22
$\Sigma_S[2,2]$	0.05	-0.03	79.5	0.20
τ	0.25	0.00	93.0	0.10
σ	0.1	0.00	92.5	0.01

Table C.5: **Simulation Study 2 (Low Variance Setting): Evaluation of the JELO (with time-varying variance) model bias, coverage, and 95% credible interval length across 200 simulation replicates.**

APPENDIX D

Appendix 04

D.1 Additional Model Checking

D.1.1 Posterior Predictive Plots

Figure D.1 display the posterior predictive checks for the FSH residuals and testosterone measurements. The solid, dark lines indicate the observed hormone measurements, while the thinner lines are kernel density estimates of the posterior distributions. We can see that while most of the data is covered by the model-generated estimates, there is still some skewness for both markers that is not fully captured. However, further model checks conducted in 4.3.4 show that our model does a reasonable job in capturing the hormone trajectories.

D.1.2 Bayesian "Residuals"

As an alternative to Figure D.1, we can also compute Bayesian "residuals" by using the following equation:

$$r_{ij}^{(}s) = x_{ij} - E(x_{ij}|\theta^{(}s))$$

where θ^s is the estimated value of the model parameters at the s-th iteration. These residuals should generally be centered at 0, with large deviations representing outliers. Systemic deviations from this can indicate poor model fit. We do this for the predictor submodel to check the model fits of FSH and testosterone. Figure D.2 shows histograms of residuals across all marker observations estimated at 20 different iterations of the model. In general, these residuals should be centered around 0, with large deviations on either side indicating outliers. In general, the distributions looks reasonable, with some of the testosterone residual histograms having a left skew. This indicates that the predictor submodel may be over-estimating small values of testosterone.

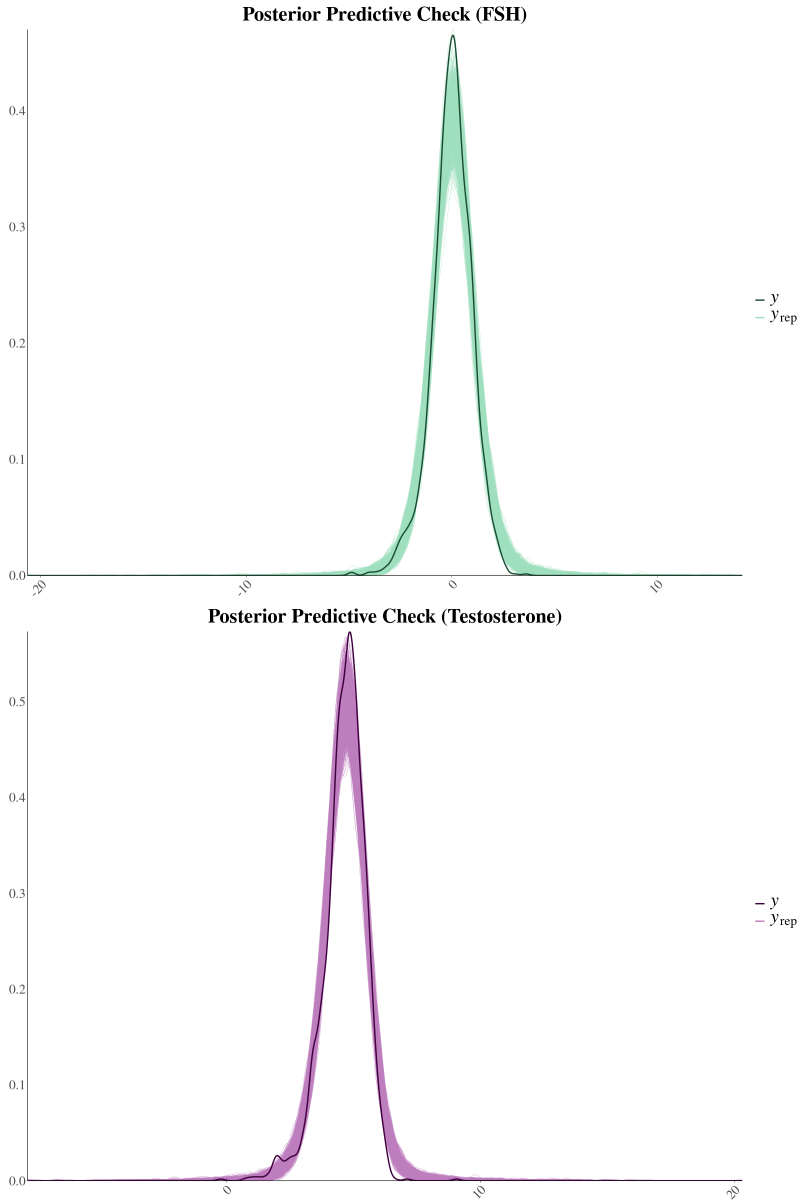


Figure D.1: **Posterior predictive check of the FSH residuals and the testosterone measurements. The darker lines indicate the observed data and the thinner bands are kernel density estimates of the posterior distributions.**

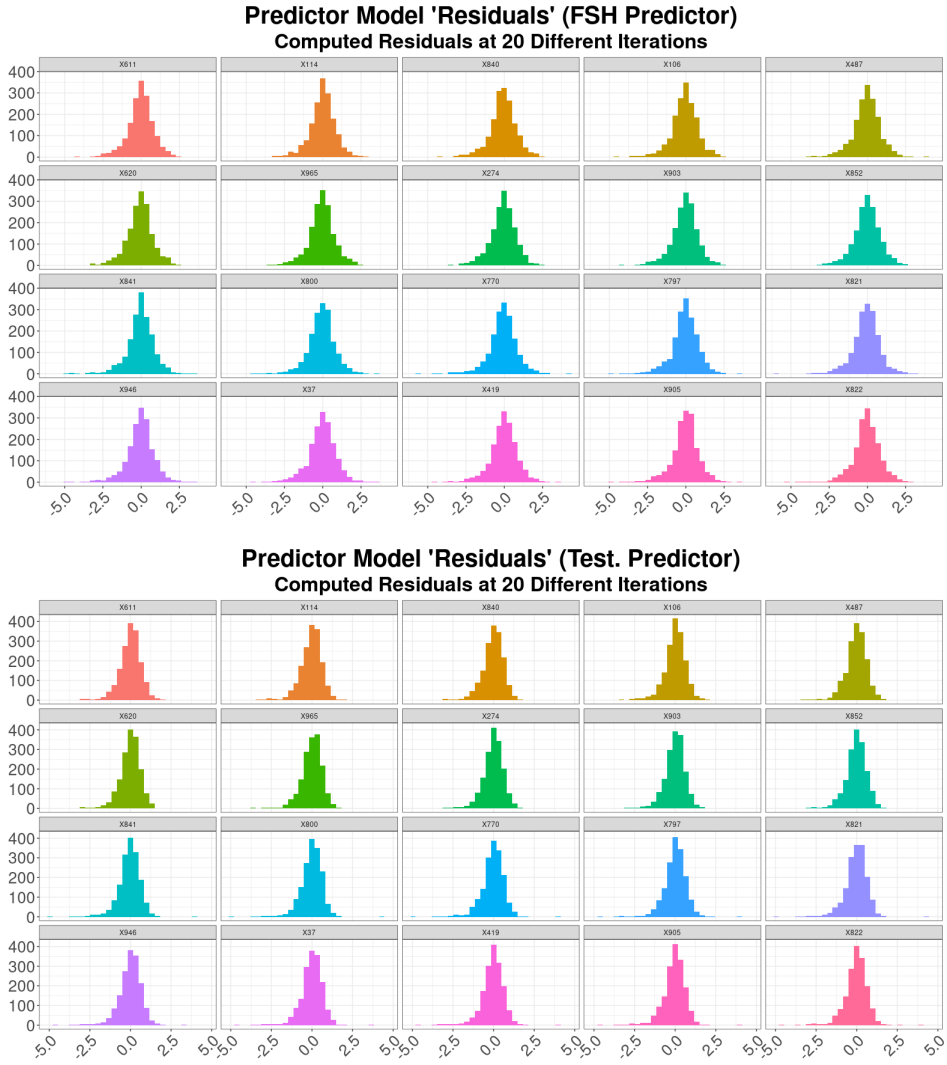


Figure D.2: **Histogram of model 'residuals' for the FSH predictor, computed at 20 different iterations.**

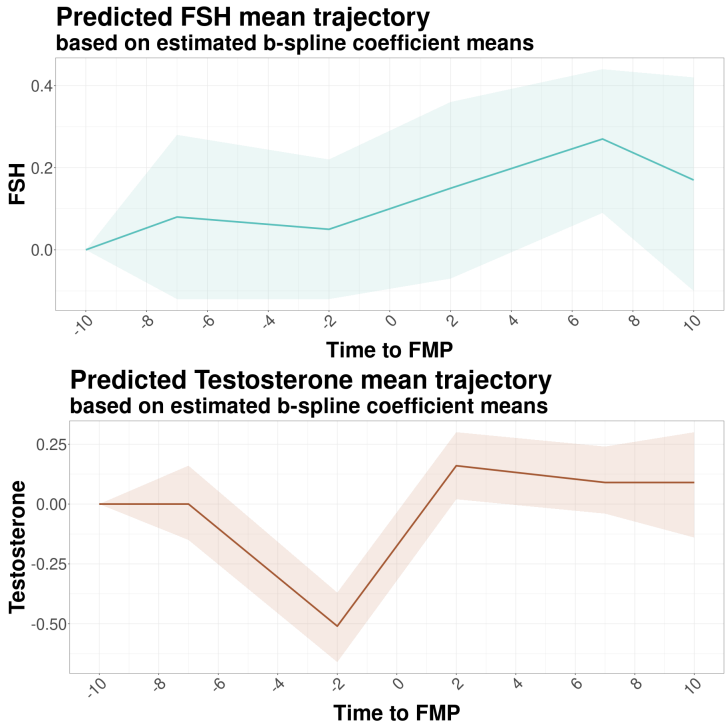


Figure D.3: **Estimated mean FSH (top figure) and estimated mean testosterone (bottom figure), based on the estimated β coefficients.**

D.1.3 Model-Estimated Mean Coefficients

As discussed in 4.3, the choice of knots used for the individual mean hormone trajectories was made based on previous scientific evidence for FSH mean levels Randolph et al. (2011). Testosterone, on the other hand, is generally linear over time and does not have the "S-shape" that we expect to see from FSH. To validate this choice on our data, Figure D.3 plots the estimated posterior means (solid lines) and 95% CrI estimates (ribbons) of the β coefficients (population mean parameters for the individual b-spline coefficients) for FSH and testosterone, based on different values for FMP. (See Table REF for the estimated posterior means and 95% CrIs for the β parameters.) The solid lines are the model-estimated posterior means of β for each hormone marker, and the ribbons represent the 95% credible interval values for each hormone.

For FSH, the model estimated mean trajectory does follow the expected pattern of FSH, with FSH rising around -2 years to FMP. FSH is predicted to continue to rise after the MT as well, before tapering at 7 years post FMP. We do see that testosterone appears to decline between -7 and -2 years to FMP, before rising again in the [-2, 2] FMP window. After 2 years past FMP, mean testosterone stablizes and the change in slopes between [2,7] and $>$ 7 years past FMP appears to be very minimal.

Parameter	Post. Mean	95% CrI
$\lambda_{1,1}$	0.21	(0.11, 0.38)
$\lambda_{1,2}$	0.21	(0.11, 0.38)
$\lambda_{1,3}$	0.21	(0.11, 0.38)
$\lambda_{2,1}$	0.21	(0.11, 0.38)
$\lambda_{2,2}$	0.21	(0.11, 0.38)
$\lambda_{2,3}$	0.21	(0.11, 0.38)
$\Psi_1[1,1]$	0.21	(0.11, 0.38)
$\Psi_1[1,2]$	0.01	(-0.03, 0.06)
$\Psi_1[1,3]$	-0.07	(-0.22, 0.02)
$\Psi_1[1,4]$	0.01	(-0.16, 0.19)
$\Psi_1[2,2]$	0.02	(0.00, 0.11)
$\Psi_1[2,3]$	-0.01	(-0.07, 0.02)
$\Psi_1[2,4]$	0.00	(-0.08, 0.07)
$\Psi_1[3,3]$	0.22	(0.04, 0.56)
$\Psi_1[3,4]$	0.13	(-0.03, 0.39)
$\Psi_1[4,4]$	0.69	(0.20, 1.53)
$\Psi_2[1,1]$	5.31	(2.25, 10.12)
$\Psi_2[1,2]$	-0.04	(-0.45, 0.32)
$\Psi_2[1,3]$	0.18	(-0.50, 1.10)
$\Psi_2[1,4]$	0.13	(-0.74, 1.25)
$\Psi_2[2,2]$	0.04	(0.00, 0.16)
$\Psi_2[2,3]$	0.00	(-0.07, 0.06)
$\Psi_2[2,4]$	0.00	(-0.07, 0.09)
$\Psi_2[3,3]$	0.19	(0.03, 0.62)
$\Psi_2[3,4]$	0.10	(-0.02, 0.33)
$\Psi_2[4,4]$	0.27	(0.03, 0.79)
b_{01}	-0.05	(-0.18, 0.07)
b_{02}	4.74	(4.65, 4.84)
τ_a	0.18	(0.16, 0.21)
σ	0.05	(0.05, 0.06)

Table D.1: **Evaluation of the posterior means and 95% CrI estimates for the covariance regression parameters and the outcome regression parameters in the FSH and Testosterone BMC outcome model.**

D.2 Data Analysis: Posterior Means and 95% CrIs for Additional Model Parameters

In this section, we present the estimated posterior means and 95% credible intervals for the other parameters from Chapter 4's data application (Section 4.3.3.2) in the main text. Table D.1 contains the estimates for the other outcome submodel parameters. Table D.2 contains the estimates for the biomarker submodel parameters.

Parameter	Post. Mean	95% CrI
β_{11}	0.08	(-0.12, 0.15)
β_{12}	0.05	(-0.12, 0.22)
β_{13}	0.15	(-0.07, 0.36)
β_{14}	0.27	(0.09, 0.44)
β_{15}	0.17	(-0.1, 0.42)
β_{21}	0.00	(-0.15, 0.16)
β_{22}	-0.51	(-0.66, -0.37)
β_{23}	0.16	(0.02, 0.30)
β_{24}	0.09	(-0.04,0.24)
β_{25}	0.09	(-0.14, 0.3)
$\Omega_1[1,1]$	0.43	(0.21, 0.69)
$\Omega_1[1,2]$	0.04	(-0.02, 0.13)
$\Omega_1[1,3]$	0.09	(-0.07, 0.26)
$\Omega_1[1,4]$	0.03	(0.07, 0.16)
$\Omega_1[1,5]$	0.00	(-0.18, 0.19)
$\Omega_1[2,2]$	0.03	(0.00, 0.01)
$\Omega_1[2,3]$	0.01	(-0.05, 0.09)
$\Omega_1[2,4]$	0.01	(-0.04, 0.07)
$\Omega_1[2,5]$	0.00	(-0.05, 0.07)
$\Omega_1[3,3]$	0.75	(0.47, 1.10)
$\Omega_1[3,4]$	0.33	(0.18, 0.52)
$\Omega_1[3,5]$	0.27	(0.06, 0.52)
$\Omega_1[4,4]$	0.34	(0.17, 0.58)
$\Omega_1[4,5]$	0.19	(0.04, 0.37)
$\Omega_1[5,5]$	0.29	(0.08, 0.59)
$\Omega_2[1,1]$	0.31	(0.19, 0.47)
$\Omega_2[1,2]$	0.21	(0.12, 0.32)
$\Omega_2[1,3]$	0.13	(0.06, 0.21)
$\Omega_2[1,4]$	0.19	(0.12, 0.29)
$\Omega_2[1,5]$	0.08	(0.00, 0.20)
$\Omega_2[2,2]$	0.43	(0.29, 0.60)
$\Omega_2[2,3]$	0.10	(0.02, 0.17)
$\Omega_2[2,4]$	0.17	(0.09, 0.27)
$\Omega_2[2,5]$	0.05	(-0.05, 0.18)
$\Omega_2[3,3]$	0.12	(0.05, 0.22)
$\Omega_2[3,4]$	0.11	(0.05, 0.18)
$\Omega_2[3,5]$	0.07	(0.00, 0.16)
$\Omega_2[4,4]$	0.21	(0.11, 0.38)
$\Omega_2[4,5]$	0.09	(0.00, 0.02)
$\Omega_2[5,5]$	0.13	(0.01, 0.35)

Table D.2: **Evaluation of the posterior means and 95% CrI estimates for the mean marker parameters in the MBHS applications. Estimates have been rounded to 2 decimal places.**

Parameter	Truth	Bias	Coverage %	Average Interval Length
b_0	0.07	-0.01	87.0	0.24
b_1	-0.02	-0.01	90.7	0.24
τ_a	0.1	-0.01	87.3	0.08
σ	0.05	0.00	98.1	0.01

Table D.3: **Simulation Study I: Evaluation of bias, coverage, and 95% credible interval length across 55 simulation replicates for the b_0, τ_a, σ parameters.**

D.3 Simulation Study- Results

In this section, we present the coverage, bias, and mean 95% CrI length for the other parameters in our Simulation Study (Section 4).

Parameter	Truth	Bias	Coverage %	Average Interval Length
β_{11}	-0.57	0.01	96.3	0.31
β_{12}	0.17	-0.01	96.3	0.32
β_{13}	-0.06	-0.01	96.3	0.33
β_{14}	0.36	0.02	94.4	0.30
β_{15}	0.34	0.01	90.7	0.38
β_{21}	0.26	0.02	90.7	0.34
β_{22}	-0.33	0.01	96.3	0.35
β_{23}	0.00	-0.01	92.6	0.36
β_{24}	-0.40	-0.02	92.6	0.35
β_{25}	-0.21	0.00	92.6	0.35
$\Omega_1[1,1]$	0.22	0.02	94.4	0.15
$\Omega_1[1,2]$	0.02	0.01	92.6	0.14
$\Omega_1[1,3]$	0.07	-0.03	100.0	0.16
$\Omega_1[1,4]$	0.04	-0.01	96.3	0.09
$\Omega_1[1,5]$	0.02	0.00	98.2	0.14
$\Omega_1[2,2]$	0.36	0.28	98.2	0.77
$\Omega_1[2,3]$	0.14	-0.19	100.0	0.55
$\Omega_1[2,4]$	-0.04	-0.06	96.3	0.21
$\Omega_1[2,5]$	-0.04	-0.01	96.3	0.17
$\Omega_1[3,3]$	0.41	0.22	94.4	0.69
$\Omega_1[3,4]$	0.05	0.06	94.4	0.24
$\Omega_1[3,5]$	0.29	-0.07	92.6	0.33
$\Omega_1[4,4]$	0.10	0.03	92.6	0.16
$\Omega_1[4,5]$	0.10	-0.04	90.74	0.18
$\Omega_1[5,5]$	0.57	0.06	92.6	0.45
$\Omega_2[1,1]$	0.10	0.00	92.6	0.14
$\Omega_2[1,2]$	0.11	-0.02	88.2	0.14
$\Omega_2[1,3]$	0.02	-0.01	96.3	0.12
$\Omega_2[1,4]$	0.01	0.00	98.1	0.12
$\Omega_2[1,5]$	0.01	0.01	100.0	0.15
$\Omega_2[2,2]$	0.45	0.00	90.74	0.31
$\Omega_2[2,3]$	0.16	-0.02	100.0	0.26
$\Omega_2[2,4]$	0.11	-0.01	96.3	0.26
$\Omega_2[2,5]$	0.08	0.04	94.4	0.18
$\Omega_2[3,3]$	0.36	0.01	92.6	0.26
$\Omega_2[3,4]$	0.27	-0.01	94.4	0.18
$\Omega_2[3,5]$	0.25	-0.01	94.4	0.21
$\Omega_2[4,4]$	0.35	-0.01	98.1	0.24
$\Omega_2[4,5]$	0.29	0.03	96.3	0.31
$\Omega_2[5,5]$	0.33	0.5	94.4	1.26
b_{01}	-0.02	0.00	87.04	0.24
b_{02}	-0.02	0.00	90.74	0.24

Table D.4: **Simulation Study I: Evaluation of bias, coverage, and 95% credible interval length across 55 simulation replicates for the B_i parameters.**

Parameter	Truth	Bias	Coverage %	Average Interval Length
$\Sigma_0[1,1]$	0.82	0.00	90.7	0.19
$\Sigma_0[1,2]$	-0.30	0.01	94.4	0.13
$\Sigma_0[2,2]$	0.21	0.01	92.6	0.14
$\Psi_1[1,1]$	0.01	0.01	94.4	0.10
$\Psi_1[1,2]$	-0.02	0.01	94.4	0.17
$\Psi_1[1,3]$	-0.02	0.00	100.0	0.04
$\Psi_1[2,2]$	0.07	0.31	90.7	0.80
$\Psi_1[2,3]$	-0.001	-0.01	100.0	0.06
$\Psi_1[3,3]$	0.001	0.00	100.0	0.02
$\Psi_2[1,1]$	0.21	0.00	94.4	0.12
$\Psi_2[1,2]$	-0.02	0.00	96.2	0.08
$\Psi_2[1,3]$	-0.02	0.02	94.4	0.07
$\Psi_2[2,2]$	0.17	0.02	94.4	0.12
$\Psi_2[2,3]$	-0.009	-0.01	94.4	0.05
$\Psi_2[3,3]$	0.001	0.31	98.2	0.67

Table D.5: **Simulation Study I: Evaluation of bias, coverage, and 95% credible interval length across 55 simulation replicates for the Θ_i parameters.**

BIBLIOGRAPHY

Ahlborg, H. G., O. Johnell, C. H. Turner, G. Rannevik, and M. K. Karlsson (2003). Bone loss and bone size after menopause. *New England Journal of Medicine 349*(4), 327–334.

Alswat, K. A. (2017). Gender disparities in osteoporosis. *Journal of Clinical Medicine Research 9*(5), 382–387.

Barnard, J., R. McCulloch, and X.-L. Meng (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica 10*, 1281–1311.

Barrett, J. K., R. Huille, R. Parker, Y. Yano, and M. Griswold (2019). Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC study. *Statistics in Medicine 38*(10), 1855–1868.

Bickel, P. J. and Y. R. Gel (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 73*(5), 711–728.

Bjørnerem, Å., B. Straume, P. Øian, and G. K. R. Berntsen (2006). Seasonal variation of estradiol, follicle stimulating hormone, and dehydroepiandrosterone sulfate in women and men. *The Journal of Clinical Endocrinology & Metabolism 91*(10), 3798–3802.

Bliuc, D., N. D. Nguyen, V. E. Milch, T. V. Nguyen, J. A. Eisman, and J. R. Center (2009). Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women. *JAMA 301*(5), 513–521.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Brown, E. R., J. G. Ibrahim, and V. DeGruttola (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics 61*(1), 64–73.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software 80*, 1–28.

Campbell, K. R., R. Martins, S. Davis, and E. Juarez-Colunga (2021). Dynamic prediction based on variability of a longitudinal biomarker. *BMC Medical Research Methodology 21*(1), 104.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of Statistical Software 76*, 1–32.

Carr, M. C. (2003). The emergence of the metabolic syndrome with menopause. *The Journal of Clinical Endocrinology & Metabolism 88*(6), 2404–2411.

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

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. CRC Press LLC.

Chan, J. C. C. and A. Grant (2015). Modeling energy price dynamics: GARCH versus stochastic volatility. In *Social Science Research Network*.

Chen, Q., R. C. May, J. G. Ibrahim, H. Chu, and S. R. Cole (2014). Joint modeling of longitudinal and survival data with missing and left-censored time-varying covariates. *Statistics in medicine 33*(26), 4560–4576.

Chi, Y.-Y. and J. G. Ibrahim (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics 62*, 432–445.

Chin, K.-Y. (2018). The relationship between follicle-stimulating hormone and bone health: Alternative explanation for bone loss beyond oestrogen? *International Journal of Medical Sciences 15*(12), 1373–1383.

Chiu, T. Y. M., T. Leonard, and K.-W. Tsui (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association 91*(433), 198–210.

Coffman, E. and J. Richmond-Bryant (2015). Multiple biomarker models for improved risk estimation of specific cardiovascular diseases related to metabolic syndrome: a cross-sectional study. *Population Health Metrics 13*(1), 7.

Colleluori, G., R. Chen, N. Napoli, L. E. Aguirre, C. Qualls, D. T. Villareal, and R. Armamento-Villareal (2018). Fat mass follows a u-shaped distribution based on estradiol levels in postmenopausal women. *Frontiers in Endocrinology 9*, 315.

Crandall, C. J., C.-H. Tseng, A. S. Karlamangla, J. S. Finkelstein, J. F. Randolph, R. C. Thurston, M.-H. Huang, H. Zheng, and G. A. Greendale (2013). Serum sex steroid levels and longitudinal changes in bone density in relation to the final menstrual period. *The Journal of Clinical Endocrinology and Metabolism 98*(4), E654–663.

Darsini, D., H. Hamidah, H. B. Notobroto, and E. A. Cahyono (2020). Health risks associated with high waist circumference: A systematic review. *Journal of Public Health Research 9*(2), 1811.

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health 21*(1), 171–192.

Diggle, P., P. Heagerty, K.-Y. Liang, and S. Zeger (2013). *Analysis of Longitudinal Data* (Second Edition ed.). Oxford Statistical Science Series. Oxford University Press.

Dykiert, D., G. Der, J. M. Starr, and I. J. Deary (2012). Age differences in intra-individual variability in simple and choice reaction time: systematic review and meta-analysis. *PloS One 7*(10), e45759.

Ebeling, P. R., L. M. Atley, J. R. Guthrie, H. G. Burger, L. Dennerstein, J. L. Hopper, and J. D. Wark (1996). Bone turnover markers and bone density across the menopausal transition. *The Journal of Clinical Endocrinology and Metabolism 81*(9), 3366–3371.

Elliott, M. R., M. D. Sammel, and J. Faul (2012). Associations between variability of risk factors and health outcomes in longitudinal studies. *Statistics in Medicine 31*, 2745–2756.

Fang, S.-C., Y.-L. Wu, and P.-S. Tsai (2020). Heart rate variability and risk of all-cause death and cardiovascular events in patients with cardiovascular disease: A meta-analysis of cohort studies. *Biological Research for Nursing 22*(1), 45–56.

Finkelstein, J. S., S. E. Brockwell, V. Mehta, G. A. Greendale, M. R. Sowers, B. Ettinger, J. C. Lo, J. M. Johnston, J. A. Cauley, M. E. Danielson, and R. M. Neer (2008). Bone mineral density changes during the menopause transition in a multiethnic cohort of women. *The Journal of Clinical Endocrinology and Metabolism 93*(3), 861–868.

Fox, E. B. and D. B. Dunson (2015). Bayesian nonparametric covariance regression. *Journal of Machine Learning Research 16*(77), 2501–2542.

Freeman, E. W., M. D. Sammel, H. Lin, and D. B. Nelson (2006). Associations of hormones and menopausal status with depressed mood in women with no history of depression. *Archives of General Psychiatry 63*(4), 375–382.

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman (2019). Visualization in bayesian workflow. *Journal of Royal Statistical Society: Series A 182*, 389–402.

Gambacciani, M. and M. Levancini (2014). Hormone replacement therapy and the prevention of postmenopausal osteoporosis. *Przeglad Menopauzalny (Menopause Review) 13*(4), 213–220.

Gao, F., J. Luo, J. Liu, F. Wan, G. Wang, M. Gordon, and C. Xiong (2022). Comparing statistical methods in assessing the prognostic effect of biomarker variability on time-to-event clinical outcomes. *BMC medical research methodology 22*(1), 201.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis 1*, 515–534.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis*. CRC Press.

Ghosh, R. P., B. Mallick, and M. Pourahmadi (2021). Bayesian estimation of correlation matrices of longitudinal data. *Bayesian Analysis 16*, 1039–1058.

Goldenberg, I., R. Goldkorn, N. Shlomo, M. Einhorn, J. Levitan, R. Kuperstein, R. Klempfner, and B. Johnson (2019). Heart rate variability for risk assessment of myocardial ischemia in patients without known coronary artery disease: The HRV-DETECT (heart rate variability for

the detection of myocardial ischemia) study†. *Journal of the American Heart Association 8*(24), e014540. Publisher: American Heart Association.

Gordon, J. L., D. R. Rubinow, T. A. Eisenlohr-Moul, J. Leserman, and S. S. Girdler (2016). Estradiol variability, stressful life events and the emergence of depressive symptomatology during the menopause transition. *Menopause (New York, N.Y.) 23*(3), 257–266.

Gourlay, M. L., J. S. Preisser, C. A. Hammett-Stabler, J. B. Renner, and J. Rubin (2011). Follicle-stimulating hormone and bioavailable estradiol are less important than weight and race in determining bone density in younger postmenopausal women. *Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 22*(10), 2699–2708.

Gourlay, M. L., B. L. Specker, C. Li, C. A. Hammett-Stabler, J. B. Renner, and J. E. Rubin (2012). Follicle-stimulating hormone is independently associated with lean mass but not BMD in younger postmenopausal women. *Bone 50*, 311–316.

Greendale, G. A., W. Han, J. S. Finkelstein, S.-A. M. Burnett-Bowie, M. Huang, D. Martin, and A. S. Karlamangla (2021). Changes in regional fat distribution and anthropometric measures across the menopause transition. *The Journal of Clinical Endocrinology and Metabolism 106*, 2520–2534.

Greendale, G. A., B. Sternfeld, M. Huang, W. Han, C. Karvonen-Gutierrez, K. Ruppert, J. A. Cauley, J. S. Finkelstein, S.-F. Jiang, and A. S. Karlamangla (2019). Changes in body composition and weight during the menopause transition. *JCI Insight 4*, e124865.

Greene, W. (2005). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis 23*(1), 7–32.

Griliches, Z. and M. D. Intriligator (1987). *Handbook of Econometrics*, Chapter 25, pp. 1465–1514. North Holland.

Guhaniyogi, R., S. Qamar, and D. B. Dunson (2017). Bayesian tensor regression. *Journal of Machine Learning Research 18*(79), 1–31.

Harlow, S. D., X. Lin, and M. J. Ho (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.

Henderson, R., P. Diggle, and A. Dobson (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics 1*, 465–480.

Hickey, G. L., P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology 16*(1), 117.

Hoff, P. D. and X. Niu (2012). A covariance regression model. *Statistica Sinica 22*(2), 729–753. Publisher: Institute of Statistical Science, Academia Sinica.

Hsu, P.-F., H.-M. Cheng, C.-H. Wu, S.-H. Sung, S.-Y. Chuang, E. G. Lakatta, F. C. Yin, P. Chou, and C.-H. Chen (2016). High short-term blood pressure variability predicts long-term cardiovascular mortality in untreated hypertensives but not in normotensives. *American Journal of Hypertension 29*(7), 806–813.

Huang, X., M. R. Elliott, and S. D. Harlow (2014). 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.

Hunter, D. J. and P. N. Sambrook (2000). Bone loss. epidemiology of bone loss. *Arthritis Research 2*(6), 441–445.

Ibrahim, J. G., H. Chu, and L. M. Chen (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology 28*, 2796–2801.

Janssen, I., L. H. Powell, R. Kazlauskaite, and S. A. Dugan (2010). Testosterone and visceral fat in midlife women: The study of women's health across the nation (SWAN) fat patterning study. *Obesity (Silver Spring, Md.) 18*(3), 604–610.

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences 13*(8), 869–881.

Jepsen, K. J., E. M. Bigelow, M. A. Casden, R. W. Goulet, K. Kennedy, S. Hertz, C. Kadur, B. T. Nolan, K. Richards-McCullough, S. Merillat, C. A. Karvonen-Gutierrez, G. Clines, and T. L. Bredbenner (2023). Associations among hip structure, bone mineral density, and strength vary with external bone size in white women. *JBMR Plus 7*(3), e10715.

Ji, M.-X. and Q. Yu (2015). Primary osteoporosis in postmenopausal women. *Chronic Diseases and Translational Medicine 1*(1), 9–13.

Jiang, B., M. R. Elliott, M. D. Sammel, and N. Wang (2015). Joint modeling of cross-sectional health outcomes and longitudinal predictors via mixtures of means and variances. *Biometrics 71*, 487–497.

Karvonen-Gutierrez, C. and S. D. Harlow (2017). Menopause and midlife health changes. In J. B. Halter, J. G. Ouslander, S. Studenski, K. P. High, S. Asthana, M. A. Supiano, and C. Ritchie (Eds.), *Hazzard's Geriatric Medicine and Gerontology* (7 ed.). McGraw-Hill Education.

Kohrt, W. M. and M. E. Wierman (2017). Preventing fat gain by blocking follicle-stimulating hormone. *The New England Journal of Medicine 377*, 293–295.

Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics 38*(4), 963–974.

Lawrence Gould, A., M. E. Boye, M. J. Crowther, J. G. Ibrahim, G. Quartey, S. Micallef, and F. Y. Bois (2015). 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.

Lewandowski, D., D. Kurowicka, and H. Joe (2009a). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis 100*, 1989–2001.

Lewandowski, D., D. Kurowicka, and H. Joe (2009b). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis 100*(9), 1989–2001.

Li, B., L. Bruyneel, and E. Lesaffre (2014). A multivariate multilevel gaussian model with a mixed effects structure in the mean and covariance part. *Statistics in Medicine 33*(11), 1877–1899.

Li, N., Y. Liu, S. Li, R. M. Elashoff, and G. Li (2021). A flexible joint model for multiple longitudinal biomarkers and a time-to-event outcome: With applications to dynamic prediction using highly correlated biomarkers. *Biometrical Journal 63*(8), 1575–1586.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Liu, P., Y. Ji, T. Yuen, E. Rendina-Ruedy, V. E. DeMambro, S. Dhawan, W. Abu-Amer, S. Izadmehr, B. Zhou, A. C. Shin, R. Latif, P. Thangeswaran, A. Gupta, J. Li, V. Shnayder, S. T. Robinson, Y. E. Yu, X. Zhang, F. Yang, P. Lu, Y. Zhou, L.-L. Zhu, D. J. Oberlin, T. F. Davies, M. R. Reagan, A. Brown, T. R. Kumar, S. Epstein, J. Iqbal, N. G. Avadhani, M. I. New, H. Molina, J. B. van Klinken, E. X. Guo, C. Buettner, S. Haider, Z. Bian, L. Sun, C. J. Rosen, and M. Zaidi (2017). Blocking FSH induces thermogenic adipose tissue and reduces body fat. *Nature 546*, 107–112.

Long, J. D. and J. A. Mills (2018). Joint modeling of multivariate longitudinal data and survival data in several observational studies of huntington's disease. *BMC Medical Research Methodology 18*, 138.

Long, Q., X. Zhang, Y. Zhao, B. A. Johnson, and R. M. Bostick (2016). Modeling clinical outcome using multiple correlated functional biomarkers: A bayesian approach. *Statistical methods in medical research 25*(2), 520–537.

MacDonald, S. W. S., L. Nyberg, and L. Bäckman (2006). Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences 29*(8), 474–480.

Maharjan, D. T., A. A. S. Syed, G. N. Lin, and W. Ying (2021). Testosterone in female depression: A meta-analysis and mendelian randomization study. *Biomolecules 11*(3), 409.

Marshall, D., O. Johnell, and H. Wedel (1996). Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ (Clinical research ed.) 312*(7041), 1254–1259.

Martins, R. (2022). A flexible link for joint modelling longitudinal and survival data accounting for individual longitudinal heterogeneity. *Statistical Methods & Applications 31*(1), 41–61.

McNeish, D. and T. Matta (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods 50*(4), 1398–1414.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics 34*(3), 1436–1462.

Mohammadalizadeh Charandabi, S., N. Rezaei, S. Hakimi, A. Montazeri, S. Taheri, H. Taghinejad, and K. Sayehmiri (2015). Quality of life of postmenopausal women and their spouses: A community-based study. *Iranian Red Crescent Medical Journal 17*(3), e21599.

Niu, X. and P. D. Hoff (2019). Joint mean and covariance modeling of multiple health outcome measures. *The annals of applied statistics 13*(1), 321–339.

Ogburn, E. L., K. E. Rudolph, R. Morello-Frosch, A. Khan, and J. A. Casey (2021). A warning about using predicted values from regression models for epidemiologic inquiry. *American Journal of Epidemiology 190*, 1142–1147.

Pan, J. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika 90*(1), 239–244.

Papageorgiou, G., K. Mauff, A. Tomer, and D. Rizopoulos (2019). An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Review of Statistics and Its Application 6*, 223–240.

Park, S. K., S. D. Harlow, H. Zheng, C. Karvonen-Gutierrez, R. C. Thurston, K. Ruppert, I. Janssen, and J. F. Randolph (2017). Association between changes in oestradiol and follicle-stimulating hormone levels during the menopausal transition and risk of diabetes. *Diabetic medicine : a journal of the British Diabetic Association 34*(4), 531–538.

Park, Y.-M., C. M. Jankowski, C. M. Swanson, K. L. Hildreth, W. M. Kohrt, and K. L. Moreau (2021). Bone mineral density in different menopause stages is associated with follicle stimulating hormone levels in healthy women. *International Journal of Environmental Research and Public Health 18*(3), 1200.

Parker, R. M. A., G. Leckie, H. Goldstein, L. D. Howe, J. Heron, A. D. Hughes, D. M. Phillippo, and K. Tilling (2021). Joint modeling of individual trajectories, within-individual variability, and a later outcome: Systolic blood pressure through childhood and left ventricular mass in early adulthood. *American Journal of Epidemiology 190*(4), 652–662.

Pettee Gabriel, K., B. Sternfeld, A. Colvin, A. Stewart, E. S. Strotmeyer, J. A. Cauley, S. Dugan, and C. Karvonen-Gutierrez (2017). 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.

Pinheiro, J., D. Bates, and R Core Team (2022). *nlme: Linear and Nonlinear Mixed Effects Models*. CRAN: The Comprehensive R Archive Network. R package version 3.1-157.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika 86*(3), 677–690. Publisher: [Oxford University Press, Biometrika Trust].

Prentice, A., T. J. Parsons, and T. J. Cole (1994). Uncritical use of bone mineral density in absorptiometry may lead to size-related artifacts in the identification of bone mineral determinants. *The American Journal of Clinical Nutrition 60*(6), 837–842.

Proust-Lima, C., M. Séne, J. M. G. Taylor, and H. Jacqmin-Gadda (2014). Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research 23*, 74–90.

Ram, N., P. Rabbitt, B. Stollery, and J. R. Nesselroade (2005). Cognitive performance inconsistency: intraindividual change and variability. *Psychology and Aging 20*(4), 623–633.

Randolph, J. F., H. Zheng, M. R. Sowers, C. Crandall, S. Crawford, E. B. Gold, and M. Vuga (2011). 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.

Randolph, Jr., J. F., M. Sowers, I. V. Bondarenko, S. D. Harlow, J. L. Luborsky, and R. J. Little (2004). 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.

Rariy, C. M., S. J. Ratcliffe, R. Weinstein, S. Bhasin, M. R. Blackman, J. A. Cauley, J. Robbins, J. M. Zmuda, T. B. Harris, and A. R. Cappola (2011). Higher serum free testosterone concentration in older women is associated with greater bone mineral density, lean body mass, and total fat mass: The cardiovascular health study. *The Journal of Clinical Endocrinology & Metabolism 96*(4), 989–996.

Recker, R., J. Lappe, K. Davies, and R. Heaney (2000). Characterization of perimenopausal bone loss: a prospective study. *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research 15*(10), 1965–1973.

Riggs, B. L. and L. J. Melton (1992). The prevention and treatment of osteoporosis. *New England Journal of Medicine 327*(9), 620–627.

Robert, C. P. and G. Casella (2010). Monte carlo integration. In C. Robert and G. Casella (Eds.), *Introducing Monte Carlo Methods with R*, Use R, pp. 61–88. Springer.

Ross, R., I. J. Neeland, S. Yamashita, I. Shai, J. Seidell, P. Magni, R. D. Santos, B. Arsenault, A. Cuevas, F. B. Hu, B. A. Griffin, A. Zambon, P. Barter, J.-C. Fruchart, R. H. Eckel, Y. Matsuzawa, and J.-P. Després (2020). Waist circumference as a vital sign in clinical practice: a consensus statement from the IAS and ICCR working group on visceral obesity. *Nature Reviews Endocrinology 16*(3), 177–189. Number: 3 Publisher: Nature Publishing Group.

Ryan, K. J. (1982). Biochemistry of aromatase: Significance to female reproductive physiology1. *Cancer Research 42*(8), 3342s–3344s.

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

Sayers, A., J. Heron, A. Smith, C. Macdonald-Wallis, M. Gilthorpe, F. Steele, and K. Tilling (2016). Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP. *Statistical Methods in Medical Research 26*(1), 437–452.

Shieh, A., G. A. Greendale, J. A. Cauley, C. Karvonen-Gutierrez, C. J. Crandall, and A. S. Karlamangla (2019). Estradiol and follicle-stimulating hormone as predictors of onset of menopause transition-related bone loss in pre- and perimenopausal women. *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research 34*(12), 2246–2253.

Sirola, J., H. Kröger, R. Honkanen, J. S. Jurvelin, L. Sandini, M. T. Tuppurainen, S. Saarikoski, and OSTPRE Study Group (2003). Factors affecting bone loss around menopause in women without HRT: a prospective study. *Maturitas 45*(3), 159–167.

Sowers, M., S. L. Crawford, B. Sternfeld, D. Morganstein, E. B. Gold, G. A. Greendale, D. Evans, R. Neer, K. Matthews, S. Sherman, A. Lo, G. Weiss, and J. Kelsey (2000a). Swan: A multicenter, multiethnic, community-based cohort study of women and the menopausal transition. In R. A. Lobo, J. Kelsey, and R. Marcus (Eds.), *Menopause: Biology and Pathology*, pp. 175–188. Academic Press.

Sowers, M., S. L. Crawford, B. Sternfeld, D. Morganstein, E. B. Gold, G. A. Greendale, D. A. Evans, R. Neer, K. Matthews, S. Sherman, A. Lo, G. Weiss, and J. Kelsey (2000b). Design, survey sampling and recruitment methods of SWAN: A multi-center, multi-ethnic, community-based cohort study of women and the menopausal transition. In *Menopause: biology and pathobiology*, Volume 175. Academic Press.

Sowers, M., M. Crutchfield, R. Bandekar, J. F. Randolph, B. Shapiro, M. A. Schork, and M. Jannausch (1998). Bone mineral density and its change in pre- and perimenopausal white women: The michigan bone health study. *Journal of Bone and Mineral Research 13*(7), 1134–1140.

Sowers, M., H. Zheng, K. Tomey, C. Karvonen-Gutierrez, M. Jannausch, X. Li, M. Yosef, and J. Symons (2007). 6-year changes in body composition in women at mid-life: ovarian and chronological aging. *The Journal of clinical endocrinology and metabolism 92*(3), 895–901.

Sowers, M. R., M. Jannausch, D. McConnell, R. Little, G. A. Greendale, J. S. Finkelstein, R. M. Neer, J. Johnston, and B. Ettinger (2006). Hormone predictors of bone mineral density changes during the menopausal transition. *The Journal of Clinical Endocrinology & Metabolism 91*(4), 1261–1267.

Sowers, M. R., H. Zheng, M. L. Jannausch, D. McConnell, B. Nan, S. Harlow, and J. F. Randolph, Jr. (2010). Amount of bone loss in relation to time around the final menstrual period and follicle-stimulating hormone staging of the transmenopause. *The Journal of Clinical Endocrinology & Metabolism 95*(5), 2155–2162.

Sponton, C. H. and S. Kajimura (2017). Burning fat and building bone by FSH blockade. *Cell Metabolism 26*, 285–287.

Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2.

Stevens, J., J. Cai, K. R. Evenson, and R. Thomas (2002). Fitness and fatness as predictors of mortality from all causes and from cardiovascular disease in men and women in the lipid research clinics study. *American Journal of Epidemiology 156*(9), 832–841.

Sullivan, S. D., A. Lehman, N. K. Nathan, C. A. Thomson, and B. V. Howard (2017). Age of menopause and fracture risk in post-menopausal women randomized to calcium + vitamin d, hormone therapy, or the combination: Results from the women's health initiative clinical trials. *Menopause (New York, N.Y.) 24*(4), 371–378.

Takeda, Y., T. Misumi, and K. Yamamoto (2022). Joint models for incomplete longitudinal data and time-to-event data. *Mathematics 10*(19), 3656. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

Talukdar, T., C. E. Zwilling, and A. K. Barbey (2023). Integrating nutrient biomarkers, cognitive function, and structural MRI data to build multivariate phenotypes of healthy aging. *The Journal of Nutrition 153*(5), 1338–1346.

Uhler, M. L., R. P. Rao, A. N. Beltsos, H. Grotjan, and A. S. Lifchez (2005). High intercycle variability of day 3 FSH levels is useful for predicting ovarian responses but not pregnancy outcomes in IVF. *Fertility and Sterility 84*, S265. Publisher: Elsevier.

U.S. Census Bureau (2017). 2017 national population projections tables: Main series. Section: Government.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner (2021). Rank-normalization, folding, and localization: An improved r^ for assessing convergence of MCMC (with discussion). *Bayesian Analysis 16*, 667–718. Publisher: International Society for Bayesian Analysis.

Wang, J., S. Luo, and L. Li (2017). Dynamic prediction for multiple repeated measures and event time data: An application to parkinson's disease. *The Annals of Applied Statistics 11*, 1787–1809.

Wang, S., T. H. McCormick, and J. T. Leek (2020). Methods for correcting inference based on outcomes predicted by machine learning. *Proceedings of the National Academy of Sciences 117*, 30266–30275.

Wu, W. B. and M. Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica 19*(4), 1755–1768.

Wu, X.-Y., S.-J. Yu, H. Zhang, H. Xie, X.-H. Luo, Y.-Q. Peng, L.-Q. Yuan, R.-C. Dai, Z.-F. Sheng, S.-P. Liu, X.-P. Wu, and E.-Y. Liao (2013). Early bone mineral density decrease is associated with FSH and LH, not estrogen. *Clinica Chimica Acta; International Journal of Clinical Chemistry 415*, 69–73.

Young, H. A. and D. Benton (2018). Heart-rate variability: a biomarker to study the influence of nutrition on physiological and psychological health? *Behavioural Pharmacology 29*(2), 140–151.

Zaidi, M., D. Lizneva, S.-M. Kim, L. Sun, J. Iqbal, M. I. New, C. J. Rosen, and T. Yuen (2018). FSH, bone mass, body fat, and biological aging. *Endocrinology 159*(10), 3503–3514.

Zhang, H., K. Ma, R.-M. Li, J.-N. Li, S.-F. Gao, and L.-N. Ma (2022). Association between testosterone levels and bone mineral density in females aged 40–60 years from NHANES 2011–2016. *Scientific Reports 12*(1), 16426. Number: 1 Publisher: Nature Publishing Group.

Zhang, Z., C. T. Rohloff, and N. Kohli (2022). Model fit indices for random effects models: Translating model fit ideas from latent growth curve models. *Structural Equation Modeling: A Multidisciplinary Journal 0*(0), 1–9.

Zhao, L., T. Chen, V. Novitsky, and R. Wang (2021). Joint penalized spline modeling of multivariate longitudinal data, with application to HIV-1 RNA load levels and CD4 cell counts. *Biometrics 77*(3), 1061–1074.