Title

WiSER: Robust and Scalable Estimation and Inference of Within-Subject Variances from Intensive Longitudinal Data

Journal

Biometrics, 78(4)

ISSN

0006-341X

Authors

German, Christopher A

Sinsheimer, Janet S

Zhou, Jin

et al.

Publication Date

2022-12-01

DOI

10.1111/biom.13506

Peer reviewed

Published in final edited form as:

Biometrics. 2022 December ; 78(4): 1313–1327. doi:10.1111/biom.13506.

WiSER: Robust and Scalable Estimation and Inference of Within-Subject Variances from Intensive Longitudinal Data

Christopher A. German[1], **Janet S. Sinsheimer**[1,2,3], **Jin Zhou**[4,5], **Hua Zhou**[1,*]

[1]Department of Biostatistics, University of California, Los Angeles, CA 90095, U.S.A

[2]Department of Computational Medicine, University of California, Los Angeles, CA 90095, U.S.A

[3]Department of Human Genetics, University of California, Los Angeles, CA 90095, U.S.A

[4]Department of Medicine, University of California, Los Angeles, CA 90095, U.S.A

[5]Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ 85721, U.S.A.

Summary:

The availability of vast amounts of longitudinal data from electronic health records (EHR) and personal wearable devices opens the door to numerous new research questions. In many studies, individual variability of a longitudinal outcome is as important as the mean. Blood pressure fluctuations, glycemic variations, and mood swings are prime examples where it is critical to identify factors that affect the within-individual variability. We propose a scalable method, within-subject variance estimator by robust regression (WiSER), for the estimation and inference of the effects of both time-varying and time-invariant predictors on within-subject variance. It is robust against the misspecification of the conditional distribution of responses or the distribution of random effects. It shows similar performance as the correctly specified likelihood methods but is $10^3 \sim 10^5$ times faster. The estimation algorithm scales linearly in the total number of observations, making it applicable to massive longitudinal data sets. The effectiveness of WiSER is evaluated in extensive simulation studies. Its broad applicability is illustrated using the accelerometry data from the Women's Health Study and a clinical trial for longitudinal diabetes care.

Keywords

blood pressure variability; electronic health record (EHR); glycemic variation; intra-individual variability; mHealth; method of moments

* huazhou@ucla.edu .

Supporting Information

Web Appendices A-I, referenced in this article, and the codes and scripts for data pre-processing and reproducing all results in this paper are available with this paper at the Biometrics website on Wiley Online Library.

Data Availability Statement

The Women's Health Study (WHS) accelerometry data is available from NIH dbGaP (2021) (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001964.v1.p1) by request. The Action to Control Cardiovascular Disease (ACCORD) data is available from NIH BioLINCC (2021) (https://biolincc.nhlbi.nih.gov/studies/accord/) by request. The codes and scripts for data pre-processing and reproducing all results in this paper are available at GitHub (2021) (https://github.com/chris-german/WiSER-Reproducibility)

1. Introduction

Electronic health records (EHR) and personal wearable devices generate massive longitudinal measurements. In many studies, the within-subject (intra-individual) variability of certain responses is of primary scientific interest, not their mean levels. Here are a few examples.

Blood pressure variability is associated with the increased risk of stroke (Rothwell et al., 2010) and received intensive attention. Rothwell et al. (2010) analyze data from a large randomized clinical trial of over 18,000 individuals comparing two classes of blood pressure lowering medications. They find that calcium-channel blockers reduce blood pressure variability while β-blockers increase systolic blood pressure variability, explaining part of the difference in the reduction of stroke risk of people on the two regimens.

Glycemic variation may play an important role in the development of diabetes complications (DeVries, 2013; Ceriello et al., 2019). Zhou et al. (2018) analyze data from the Veterans Affairs Diabetes Trial where fasting glucose is measured repeatedly in over 1,700 veterans. High blood glucose variability is associated with increased cardiovascular disease in patients with type 2 diabetes (T2D) even after accounting for mean levels.

The popularity of smart phones and handheld devices make ecological momentary assessment (EMA) methods powerful tools in modern behavioral, social, and psychological studies. Compared to retrospective self-reports collected at research or clinic visits, which are subject to recall bias, EMA repeatedly samples subjects' current behaviors and experiences in real time (Heron et al., 2017; Russell and Gajos, 2020). EMA generates enormous amounts of longitudinal data and sparks new methodology development (Ruwaard et al., 2018). *Mood swings*, defined as mood fluctuations measured on the Visual Analogue Scale, are intensively studied EMA outcomes (Ruwaard et al., 2018, Chapter 5). They are linked to stress, substance abuse, depressive symptoms, and mood disorders.

These applications need effective methods to identify the covariates (risk factors, genetic variants, environmental factors) that affect the intra-individual variances. Analyzing such longitudinal data is challenging. First, the model needs to properly account for the correlation of longitudinal measurements. Second, the model needs to discern sources of variation at the mean level, between subjects (BS), and within subjects (WS). Third, the real data often violate the statistical model's distribution assumptions. Lastly, the scale of EHR and personal wearable device data makes computation challenging. These sources not only generate data for a massive number of individuals, e.g., UK Biobank (Sudlow et al., 2015) has EHR data on 2×10^5 individuals and the Million Veteran Project (MVP) (Gaziano et al., 2016) has EHR data on 7×10^5 individuals, but also for a massive number of longitudinal measurements, e.g., Apple Watches sample heart rate every 5 minutes in standby mode and continuously as 5-second averages during workouts (Tison et al., 2018). The large size in the longitudinal dimension is particularly damaging to computing. Methods such as linear mixed models (LMMs) and generalized estimation equation (GEE) scale as the cube of the longitudinal dimension because of inversion of the covariance matrices.

1.1 Previous work and our contributions

Current applications employ heuristic strategies to calculate subject-level longitudinal variation such as standard deviation (SD), average real variability (ARV), or coefficient of variation (CV), and then model them as the responses with covariates (Smit et al., 2018; Ivarsdottir et al., 2017). This framework implicitly assumes that an individual's variability is constant over time, and cannot be affected by time-varying covariates. Additionally, this approach does not recognize that these standard deviations can be based on very different numbers of observations, as is often the case in health applications. Figure 1 depicts a hypothetical but commonly observed scenario where the WS variability is affected by both time-varying (e.g. medication use) and time-invariant features (e,g. gender). Regressing the subject-level variability summaries on predictors leads to serious bias (Barrett et al., 2019). In a simulation experiment in Web Appendix F, we demonstrate that this heuristic approach can lead to serious inflation of type I error and power loss.

LMMs are powerful tools for modeling variation in the longitudinal setting (Fitzmaurice et al., 2011; Verbeke and Molenberghs, 2009). Motivated by a smartphone-based EMA study of adolescent smoking behavior, Hedeker et al. (2008) introduce a *mixed-effects location scale model* for longitudinal data which allows both WS and BS variability to be modeled through covariates. They model the mood assessment y_{ij} of student i at occasion $j \in \{1, 2, \ldots, n_i\}$ as

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i + \epsilon_{ij},$$

where $\boldsymbol{x}_{ij}$ is the $p \times 1$ vector of regressors typically including the intercept and $\boldsymbol{\beta}$ is the corresponding regression coefficients. The random intercepts v_i are independently distributed as normal with mean zero and variance σ_v^2. The errors ϵ_{ij} are independently distributed as normal with mean zero and variance σ_ϵ^2, independent of v_i. Here σ_v^2 represents the BS variance and σ_ϵ^2 represents the WS variance. To allow covariates to influence BS and WS variances, a log-linear model is employed: $\sigma_{v_i}^2 = \exp\left(\boldsymbol{u}_i^T \boldsymbol{\alpha}\right)$, $\sigma_{\epsilon_{ij}}^2 = \exp\left(\boldsymbol{w}_{ij}^T \boldsymbol{\tau}\right)$. The variances are subscripted by i and j to indicate that their values change depending on the values of the covariates $\boldsymbol{u}_i$ and $\boldsymbol{w}_{ij}$ (and their parameters). The WS variance can further vary across individuals beyond the contribution of the covariates by, $\sigma_{\epsilon_{ij}}^2 = \exp\left(\boldsymbol{w}_{ij}^T \boldsymbol{\tau} + \omega_i\right)$, where the random intercepts ω_i have mean 0 and variance σ_ω^2. If ω_i is specified as normal, then the WS variances follow a log-normal distribution at the individual level. The mixed-effects location scale model has been estimated using Bayesian approaches by several authors, allowing for more flexibility in assumed distributions (Rast et al., 2012; Goldstein et al., 2017; Barrett et al., 2019).

The mixed-effects location scale model has many advantages over the heuristic methods. It allows for simultaneous modeling of the mean and variability of the longitudinal measurement, increases power, and reduces bias. It leverages information across individuals to get more precise estimates (Barrett et al., 2019).

Lin et al. (1997) use a model similar to the mixed-effects location scale model except that the WS variance has an inverse Gamma distribution whose mean is related to WS predictors via the log-linear link. By using quasi-likelihoods and method of moments, they avoid numerical integration. However the WS predictors are linked to the subject-level mean of WS variance, which excludes modeling time-varying covariate effects on WS variability.

Dzubur et al. (2020) further expand the mixed-effects location scale model to a *mixed-effects multiple location scale model* that allows for multiple random effects in the mean component. This model motivates us and is discussed in Section 2.1. However, fitting such a model is extremely challenging because it requires numerical integration in each iteration. Another concern is that real data can violate the restrictive distribution assumptions for both the response and random effects and compromise the estimation and inference.

We propose an estimation method, within-subject variance estimator by robust regression (WiSER), which is robust to misspecification of the response (conditional on random effects) and the random effects distributions. WiSER is a method of moments (MoM) adaptation of the likelihood approach by Dzubur et al. (2020). It is similar to Lin et al. (1997) but allows time-varying predictors for WS variability. The estimation algorithm avoids numerical integration and large matrix inversion and scales linearly in the total number of longitudinal measurements. WiSER's close connection to the quadratic estimating equation (QEE) is shown in Web Appendix B.

Table 1 contrasts WiSER estimates and run times with those of maximum likelihood estimation (MLE) as implemented in the MixWILD software (Hedeker and Nordgren, 2013; Dzubur et al., 2020) on two simulated data sets with 1000 individuals and 10 observations per individual. MixWILD run times range from 40 minutes to 10+ hours according to the different assumptions being made. In contrast, WiSER takes less than one second to obtain point estimates and confidence intervals, which are almost identical to MLE. The WiSER method is introduced in the next section and more extensive simulation studies are presented in Section 5 to evaluate its estimation and inference accuracy in various scenarios.

2. Model

Table 2 summarizes the notation used in this article.

2.1 Method of moment estimator

We first motivate our method by developing a method of moment (MoM) estimator for the mixed-effects multiple location scale models (Dzubur et al., 2020)

$$\begin{aligned} y_{ij} &= \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\gamma}_i + \epsilon_{ij}, \quad \epsilon_{ij}\sim N\left(0, \sigma^2_{\epsilon_{ij}}\right), \\ \sigma^2_{\epsilon_{ij}} &= \exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\tau} + \boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\gamma}_i + \omega_i\right), \quad \omega_i\sim N\left(0, \sigma^2_{\omega}\right), \end{aligned} \tag{1}$$

where $\sigma^2_{\epsilon_{ij}}$ represents the WS variance and $\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T$ comes from the Cholesky factor of the covariance matrix of the random effects joint distribution

$$\begin{pmatrix}\gamma_i\\ w_i\end{pmatrix}\sim N\left(\mathbf{0}_{q+1},\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}\right).$$

We denote the Cholesky decomposition of the random effects covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}w}$ as

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}=\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} & \boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}\\ \boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}^T & \sigma_\omega^2\end{pmatrix}=\begin{pmatrix}\boldsymbol{L}_{\boldsymbol{\gamma}} & \mathbf{0}\\ \boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T & \ell_\omega\end{pmatrix}\begin{pmatrix}\boldsymbol{L}_{\boldsymbol{\gamma}}^T & \boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}\\ \mathbf{0}^T & \ell_\omega\end{pmatrix},$$

where $\boldsymbol{L}_{\boldsymbol{\gamma}}$ is a $q\times q$ lower triangular matrix with positive diagonal entries and $\ell_\omega>0$. The elements of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}$ are expressed in terms of the Cholesky factors as

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}=\boldsymbol{L}_{\boldsymbol{\gamma}}\boldsymbol{L}_{\boldsymbol{\gamma}}^T,\quad \boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}=\boldsymbol{L}_{\boldsymbol{\gamma}}\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega},\quad \sigma_\omega^2=\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}+\ell_\omega^2.$$

The model (1) allows covariates to affect both WS variability and the mean. $\boldsymbol{w}_{ij}$ reflects covariates modeling WS variability; it is not necessarily a subset of $\boldsymbol{x}_{ij}$. $\boldsymbol{\gamma}_i$ in the model for $\sigma_{\epsilon ij}^2$ allows random location effects, which represent BS variability, to be correlated with the WS variability. To derive a MoM estimator, we note the conditional distribution of the response given random effects is

$$\boldsymbol{Y}_i\mid\boldsymbol{\gamma}_i,\omega_i\sim N\left(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{\gamma}_i,\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}\right),\quad \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}=\operatorname{diag}\left(\sigma_{\epsilon i1}^2,\sigma_{\epsilon i2}^2,\ldots,\sigma_{\epsilon in_i}^2\right). \tag{2}$$

Then the iterated expectation formula yields the marginal mean and covariance

$$\mathbb{E}(\boldsymbol{Y}_i)=\mathbb{E}[\mathbb{E}(\boldsymbol{Y}_i\mid\boldsymbol{\gamma}_i,\omega_i)]=\boldsymbol{X}_i\boldsymbol{\beta}$$

$$\operatorname{Var}(\boldsymbol{Y}_i)=\mathbb{E}[\operatorname{Var}(\boldsymbol{Y}_i\mid\boldsymbol{\gamma}_i,\omega_i)]+\operatorname{Var}[\mathbb{E}(\boldsymbol{Y}_i\mid\boldsymbol{\gamma}_i,\omega_i)]=\operatorname{diag}\left(\mathbb{E}\sigma_{\epsilon i1}^2,\mathbb{E}\sigma_{\epsilon i2}^2,\ldots,\mathbb{E}\sigma_{\epsilon in_i}^2\right)+\boldsymbol{Z}_i\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{Z}_i^T.$$

The expectation

$$\mathbb{E}\sigma_{\epsilon ij}^2=\mathbb{E}\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\tau}+\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\gamma}_i+\omega_i\right)=\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\tau}\right)\mathbb{E}\exp\left(\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\gamma}_i+\omega_i\right)$$

evaluates to the moment generating function of a normal random variable with mean 0 and variance $\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}+\sigma_\omega^2+2\,\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}$. Thus

$$\mathbb{E}\sigma_{\epsilon ij}^2=\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\tau}+0.5\left(\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}+\sigma_\omega^2+2\,\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}\right)\right)=e\left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}\right)\cdot\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\tau}\right),$$

where the constant

$$e(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}) = \exp\Big(0.5\Big(\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\ \boldsymbol{\ell}_{\boldsymbol{\gamma}\omega} + \ell_\omega^2 + 2\,\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\,\boldsymbol{L}_{\boldsymbol{\gamma}}\,\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega} + \boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\,\boldsymbol{L}_{\boldsymbol{\gamma}}\boldsymbol{L}_{\boldsymbol{\gamma}}^T\,\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}\Big)\Big)$$

encapsulates the contribution to the population WS variance due to random effects. This leads to the expression for the variance of $\boldsymbol{Y}_i$

$$\boldsymbol{V}_i(\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}) = e(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}) \begin{pmatrix} \exp\left(\boldsymbol{w}_{i1}^T\boldsymbol{\tau}\right) & & \\ & \ddots & \\ & & \exp\left(\boldsymbol{w}_{in_i}^T\boldsymbol{\tau}\right) \end{pmatrix} + \boldsymbol{Z}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{Z}_i^T.$$

To obtain an MoM estimator for the model parameters, we minimize the squared error between the subject empirical covariance matrices and their theoretical ones

$$\frac{1}{2}\sum_{i=1}^{m}\left\|\left(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}\right)\left(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}\right)^T - \boldsymbol{V}_i(\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega})\right\|_{\mathrm{F}}^2, \tag{3}$$

where $\widehat{\boldsymbol{\beta}} = \left(\sum_i \boldsymbol{X}_i^T\boldsymbol{X}_i\right)^{-1}\left(\sum_i \boldsymbol{X}_i^T\boldsymbol{y}_i\right)$ is the ordinary least squares estimate of $\boldsymbol{\beta}$. Here $\|\cdot\|_{\mathrm{F}}$ indicates the Frobenius norm of a matrix.

2.2 Robust estimation by WiSER

The MoM estimator enjoys a "double robustness" property. Unlike the usual sense where an estimator is robust to a violation of either one of two assumptions, the MoM estimator is robust to violation of both assumptions. It is robust to the misspecification of both the distribution of random effects ($\boldsymbol{\gamma}_i$, ω_i) and the conditional distribution of $\boldsymbol{Y}_i$ given ($\boldsymbol{\gamma}_i$, ω_i). The derivation only requires the conditional moments

$$\mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i, \omega_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i,\ \ \mathrm{Var}(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i, \omega_i) = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_i}.$$

Furthermore, the joint normality of random effects ($\boldsymbol{\gamma}_i$, ω_i) is not critical. The only requirements are the existence of the covariance matrix $\mathrm{Var}(\boldsymbol{\gamma}_i, \omega_i) = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}$ and the expectation $e(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega}) = \mathbb{E}\exp\left(\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}^T\,\boldsymbol{\gamma}_i + \omega_i\right)$. Because our scientific interests lie in the non-intercept coefficients in $\boldsymbol{\tau}$, the constant term $e(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\omega})$ is absorbed into the intercept in $\boldsymbol{\tau}$. The nuisance parameters ℓ_ω and $\boldsymbol{\ell}_{\boldsymbol{\gamma}\omega}$, thus σ_ω^2 and $\boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}$, are not identifiable in (3); however this lends us robustness against the misspecification of random effects distribution. If the primary interest is to estimate σ_ω^2 and $\boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}$, then one invokes higher moments, since they characterize the BS variance of WS variances, or uses the full likelihood approach.

We seek an estimation method that inherits the robustness and computational simplicity of the MoM, while improving its statistical efficiency. This leads to the WiSER estimator

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2}\sum_i (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T\left(\boldsymbol{V}_i^{(0)}\right)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$$

$$\widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}} \frac{1}{2} \sum_i \mathrm{tr}\left(\left(\boldsymbol{V}_i^{(0)} \right)^{-1} \boldsymbol{R}_i \left(\boldsymbol{V}_i^{(0)} \right)^{-1} \boldsymbol{R}_i \right), \tag{4}$$

where

$$\boldsymbol{R}_i = \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} \right) \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} \right)^T - \boldsymbol{V}_i \left(\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \right),$$

$$\boldsymbol{V}_i \left(\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \right) = \begin{pmatrix} \exp\left(\boldsymbol{w}_{i1}^T \boldsymbol{\tau} \right) & & \\ & \ddots & \\ & & \exp\left(\boldsymbol{w}_{in_i}^T \boldsymbol{\tau} \right) \end{pmatrix} + \boldsymbol{Z}_i \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \boldsymbol{Z}_i^T, \tag{5}$$

and $\boldsymbol{V}_i^{(0)} = \boldsymbol{V}_i \left(\boldsymbol{\tau}^{(0)}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{(0)} \right)$ is an initial estimator of $\mathrm{Var}(\boldsymbol{Y}_i)$. We emphasize that the WS covariate matrices $\boldsymbol{W}_i$ must include an intercept, which encapsulates the population level baseline WS variance plus BS variance of WS variances. Taking $\boldsymbol{V}_i^{(0)} = \boldsymbol{I}_{n_i}$, WiSER reduces to the MoM. In practice we find that setting initial $\boldsymbol{V}_i^{(0)}$ to a least squares estimator of $\boldsymbol{\tau}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ leads to good performance (see Section 4). Iterating the WiSER procedure (4) improves estimation accuracy. That is, before each round of WiSER, we update $\boldsymbol{V}_i^{(0)}$ with the current WiSER estimates of $\boldsymbol{\tau}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ and repeat. In this paper, unless specified otherwise, we report the results of setting $\boldsymbol{V}_i^{(0)}$ to an initial least squares estimate and then running two rounds of WiSER.

Remark 1: WiSER estimator (4) is a special case of the quadratic estimation equation for estimating variance parameters (Prentice, 1988; Zhao and Prentice, 1990; Ye and Pan, 2006; Leng et al., 2010). Specifically, in Web Appendix B, we show that WiSER is equivalent to a specific quadratic generalized estimation equation with a working covariance structure assuming marginal normality of $\boldsymbol{Y}_i$. This particular working covariance strikes a balance between statistical efficiency and computational scalability.

3. Statistical properties

3.1 Consistency and asymptotic normality

Theorem 1 establishes the consistency and asymptotic normality of the WiSER estimator under regularity conditions. A sketch of the proof, following the M-estimation framework (van der Vaart, 1998), is given in Web Appendix D. We use notation $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\tau}, \mathrm{vech}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$ to collect all model parameters. vech $\boldsymbol{A}$ stacks the entries of the lower triangular part of a square matrix $\boldsymbol{A}$ into a long vector in the column-major order. Corresponding to the WiSER empirical loss functions in (4), we define the population criterion function

$$f_1(\boldsymbol{\theta}) = \frac{1}{2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \left(\boldsymbol{V}^{(0)} \right)^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}),$$

$$f_2(\boldsymbol{\theta}) = \frac{1}{2}\left\|\left(\boldsymbol{V}^{(0)}\right)^{-1}\boldsymbol{R}\left(\boldsymbol{V}^{(0)}\right)^{-1}\boldsymbol{R}\right\|_{\mathrm{F}}^2$$

with gradient $\nabla f(\boldsymbol{\theta}) = \left[\nabla_{\boldsymbol{\beta}} f_1(\boldsymbol{\theta})^T, \nabla_{\boldsymbol{\tau},\,\mathrm{vech}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}} f_2(\boldsymbol{\theta})^T\right]^T$. Explicit expressions for the gradient are detailed in Web Appendix C.2. We make the following assumptions:

(A1) (Model) Observation tuples $(\boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i)$, $i = 1, \ldots, m$, are independently and identically distributed (iid) from $F = F(\boldsymbol{\theta}_0)$ and satisfy the conditional moment conditions

$$\mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i) = \boldsymbol{X}_i\boldsymbol{\beta}_0,$$

$$\mathrm{Var}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i) = \boldsymbol{V}_i\left(\boldsymbol{\tau}_0, \boldsymbol{\Sigma}_{\boldsymbol{\gamma},0}\right),$$

where $\boldsymbol{V}_i(\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$ takes the form (5). We denote the dimension of $\boldsymbol{Y}_i$ (number of observations for the i-th individual) by N_i, which is random under F.

(A2) (Compactness) $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\tau}, \mathrm{vech}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$ lies within a compact set $\boldsymbol{\Theta}$ and $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \mathrm{vech}\boldsymbol{\Sigma}_{\boldsymbol{\gamma},0})$ is in the interior of $\boldsymbol{\Theta}$.

(A3) (Identifiability) $\| \mathbb{E}\nabla f(\boldsymbol{\theta}) \|_2 > 0$ under F for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ in $\boldsymbol{\Theta}$.

(A4) (Moment condition) These moments are finite under, F: $\mathbb{E}\|\boldsymbol{Y}_i\|_2^8$, $\mathbb{E}\lambda_{\max}^2\left(\boldsymbol{W}_i^T\boldsymbol{W}_i\right)$, $\mathbb{E}N_i^2$, and $\mathbb{E}\lambda_{\max}^4\left(\boldsymbol{Z}_i^T\boldsymbol{Z}_i\right)$. Here $\lambda_{\max}(\boldsymbol{M})$ is the maximal eigenvalue of a symmetric matrix $\boldsymbol{M}$.

(A5) (Nonsingularity) The matrices

$$\boldsymbol{A}_1(\theta_0) = \mathbb{E}_F\boldsymbol{X}_i^T\left(\boldsymbol{V}_i^{(0)}\right)^{-1}\boldsymbol{X}_i,$$

$$\boldsymbol{A}_2(\boldsymbol{\theta}_0) = \mathbb{E}_F\begin{pmatrix}\boldsymbol{W}_i^T\mathrm{diag}\left(e^{\boldsymbol{W}_i\boldsymbol{\tau}_0}\right)\boldsymbol{Q}_{N_i}^T\\ \boldsymbol{C}_q^T\left(\boldsymbol{Z}_i^T\otimes\boldsymbol{Z}_i^T\right)\end{pmatrix}\left(\boldsymbol{V}_i^{(0)}\otimes\boldsymbol{V}_i^{(0)}\right)^{-1}\begin{pmatrix}\boldsymbol{W}_i^T\mathrm{diag}\left(e^{\boldsymbol{W}_i\boldsymbol{\tau}_0}\right)\boldsymbol{Q}_{N_i}^T\\ \boldsymbol{C}_q^T\left(\boldsymbol{Z}_i^T\otimes\boldsymbol{Z}_i^T\right)\end{pmatrix}^T$$

are positive definite. $\boldsymbol{C}_q$ is the $q^2 \times q(q+1)/2$ copying matrix such that $\boldsymbol{C}_q \cdot \mathrm{vech}\boldsymbol{M} = \mathrm{vec}\boldsymbol{M}$ for arbitrary $q \times q$ lower triangular matrix $\boldsymbol{M}$ and $\boldsymbol{Q}_n$ is the $n^2 \times n$ diagonal selection matrix such that $\mathrm{diag}(\boldsymbol{M}) = \boldsymbol{Q}_n^T\mathrm{vec}\,\boldsymbol{M}$ for any $n \times n$ square matrix $\boldsymbol{M}$.

(A6) (Boundedness) Entries of $\boldsymbol{W}_i$ and $\left(\boldsymbol{V}_i^{(0)}\right)^{-1}$ are uniformly bounded with probability one.

Theorem 1: Under (A1)–(A6), the WiSER estimator $\widehat{\theta}_m = \left(\widehat{\beta}_m, \widehat{\tau}_m, \mathrm{vech}\widehat{\Sigma}_{\gamma_m}\right)$ defined by (4) is strongly consistent as $m \to \infty$ and $\sqrt{m}\left(\widehat{\theta}_m - \theta_0\right)$ is asymptotically normal with mean zero and covariance

$$S(\theta_0) = \begin{pmatrix} A_1^{-1}(\theta_0) & O \\ O & A_2^{-1}(\theta_0) \end{pmatrix} \cdot \left[\mathbb{E}_F \nabla f(\theta_0) \nabla f(\theta_0)^T\right] \cdot \begin{pmatrix} A_1^{-1}(\theta_0) & O \\ O & A_2^{-1}(\theta_0) \end{pmatrix}.$$

A few remarks are in order.

Remark 2: WiSER's only structural assumption is the conditional moment condition (A1), which guarantees unbiasedness of the estimation equation $\mathbb{E}_F[\nabla f(\theta_0)] = \mathbf{0}$. The mixed-effects multiple location scale model (1) satisfies (A1) whenever the moment generating function of the random effects (γ_i, ω_i) exists (Section 2.1). This relaxes normality assumptions on the conditional distribution of Y_i and the distribution of random effects (γ_i, ω_i).

Remark 3: Under (A1), the WiSER estimate $\widehat{\beta}$ is semiparametric efficient (Tsiatis, 2006); it has the smallest asymptotic variance among all semiparametric estimators of β.

Remark 4: If we assume that N_i and the entries of Z_i are bounded by a finite constant with probability one, together with the boundedness condition (A6), then the moment condition (A4) reduces to just $\mathbb{E}\|Y_i\|_2^4 < \infty$.

3.2 Sandwich estimator

We use the plug-in estimator

$$\widehat{A}_{1,m} = \frac{1}{m}\sum_i X_i^T \left(V_i^{(0)}\right)^{-1} X_i$$

$$\widehat{A}_{2,m} = \frac{1}{m}\sum_i \begin{pmatrix} W_i^T \mathrm{diag}\left(e^{W_i \widehat{\tau}_m}\right) Q_{n_i}^T \\ C_q^T\left(Z_i^T \otimes Z_i^T\right) \end{pmatrix} \left(V_i^T \otimes V_i^{(0)}\right)^{-1} \begin{pmatrix} W_i^T \mathrm{diag}\left(e^{W_i \widehat{\tau}_m}\right) Q_{n_i}^T \\ C_q^T\left(Z_i^T \otimes Z_i^T\right) \end{pmatrix}^T$$

for $A_1(\theta_0)$ and $A_2(\theta_0)$ respectively and the empirical estimator

$$\widehat{B}_m = \frac{1}{m}\sum_i \nabla f\left(\widehat{\theta}_m; y_i, X_i, Z_i, W_i\right) \nabla f\left(\widehat{\theta}_m; y_i, X_i, Z_i, W_i\right)^T$$

for $B(\theta_0)$. Then the sandwich estimator for the asymptotic covariance of $\sqrt{m}\left(\widehat{\theta}_m - \theta_0\right)$ is

$$\widehat{S}_m = \begin{pmatrix} \widehat{A}_{1,m}^{-1} & O \\ O & \widehat{A}_{2,m}^{-1} \end{pmatrix} \widehat{B}_m \begin{pmatrix} \widehat{A}_{1,m}^{-1} & O \\ O & \widehat{A}_{2,m}^{-1} \end{pmatrix}.$$

The consistency of $\widehat{S}_m$ for estimating $S(\theta_0)$ is guaranteed by showing that the second and third derivatives of f_i, $i = 1,2$, are bounded above by an integrable function (Boos and Stefanski, 2013, Theorem 7.3) under the moment condition (A4). Details are omitted.

3.3 Hypothesis testing

We partition the parameter θ as $\theta_1 \in \mathbb{R}^r$ and $\theta_2 \in \mathbb{R}^{p+\ell+q(q+1)/2-r}$. In our applications, θ_1 is always a sub-vector of (β, τ). Inference on the variance component Σ_γ is difficult due to the boundary conditions and is subject to a parametric bootstrap. The Wald test statistic for testing $H_0 : \theta_1 = \theta_{10}$ is $T_W = \left(\widehat{\theta}_{m,1} - \theta_{10}\right)^T \left(\widehat{S}_{m,11}\right)^{-1} \left(\widehat{\theta}_{m,1} - \theta_{10}\right)$, where $\widehat{S}_{m,11}$ is the sub-block of the sandwich estimator $\widehat{S}_m$ corresponding to θ_1. T_W is asymptotically distributed as χ_r^2 under H_0. A score test (Boos, 1992) can be derived but is not pursued here.

4. Computational strategy

The optimization task in WiSER (4) is a nonlinear least squares problem and subject to standard algorithms such as the Gauss-Newton and Levenberg-Marquardt methods. Our implementation, an open source Julia package WiSER.jl (2021), offers a choice of many open source nonlinear programming solvers, such as Ipopt (Wächter and Biegler, 2006) and NLopt (Johnson, 2020), and commercial ones, such as KNITRO (Byrd et al., 2006). With careful implementation, each iteration of the optimization algorithms scales linearly in the total number of observations $\sum_i n_i$; therefore WiSER can be applied to very large longitudinal data sets. In Web Appendix C, we provide a detailed account of how to efficiently evaluate the objective function, gradient, and expected Hessian matrix. The key is to utilize the Woodbury structure (Hager, 1989) in V_i and $\left(V_i^{(0)}\right)^{-1}$ to avoid the storage and computation of potentially large $n_i \times n_i$ matrices. Each iteration costs $O\left(\left(\sum_i n_i\right) \ell\, q^2 + q^4\right)$ flops. Convergence is achieved from a few to a few dozen iterations in most scenarios, depending on the algorithm, solver, sample size, generative model, and signal-to-noise ratio. If users have the time and resources, exploring different solvers and starting values may be worthwhile in some applications. We recommend using the solution with the best objective value in that case. Figure 2 demonstrates the linear scalability of WiSER on simulated data sets. Run times scale linearly with the number of independent individuals and with the number of measures per individual (left panel); and the average time per observation stabilizes quickly within one million observations (right panel).

To get an initial estimate of $V_i^{(0)}$, we start from the regular least squares estimate of β

$$\boldsymbol{\beta}^{(0)} = \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1}\left(\sum_i \boldsymbol{X}_i^T \boldsymbol{y}_i\right),$$

compute the corresponding residuals $\boldsymbol{r}_i^{(0)} = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}^{(0)}$, and then set $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{(0)}$ to be the minimizer of the least squares criterion $\sum_i \left\|\text{offdiag}\left(\boldsymbol{r}_i^{(0)}\boldsymbol{r}_i^{(0)T} - \boldsymbol{Z}_i\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{Z}_i^T\right)\right\|_{\text{F}}^2$. Here offdiag($\boldsymbol{M}$) sets the diagonal entries of a matrix $\boldsymbol{M}$ to zero. We initialize $\boldsymbol{\tau}^{(0)}$ by regressing $\log\left(\boldsymbol{r}_i^2\right) = \left(\log r_{i1}^2, \ldots, \log r_{in_i}^2\right)^T$ on $\boldsymbol{W}_i$; that is

$$\boldsymbol{\tau}^{(0)} = \left(\sum_i \boldsymbol{W}_i^T \boldsymbol{W}_i\right)^{-1}\left[\sum_i \boldsymbol{W}_i^T \log\left(\boldsymbol{r}_i^2\right)\right].$$

5. Simulations

We evaluate WiSER's estimation accuracy and confidence interval coverage in two scenarios. The first (Section 5.1) is the LMM normal-normal model. The second (Section 5.2) investigates the robustness of WiSER by using non-normal distributions for the conditional distribution of $\boldsymbol{Y}_i$ and distributions of the random effects ($\boldsymbol{\gamma}_i$, ω_i). In both scenarios, non-intercept entries of covariate matrices $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$, and $\boldsymbol{W}_i$ are generated from independent standard normal and the true regression coefficients are $\boldsymbol{\beta}_{\text{true}}$ = (0.1, 6.5, −3.5, 1.0, 5)^T and $\boldsymbol{\tau}_{\text{true}}$ = (0.0, 0.5, −0.2, 0.5, 0.0)^T. In Section 5.3, WiSER estimates are compared to the computational intensive MLE on two representative simulation replicates. In both scenarios, we vary subjects $m \in$ {1000, 2000, …, 6000} and observations per subject n_i $\in$ {10, 25, 50, 100, 1000}. Each simulation scenario was run on 1000 replicates. These scenarios reflect the sample sizes in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial in Section 6.2.

5.1 (Normal, Normal, Log-Normal) model

We set the conditional distribution of $\boldsymbol{Y}_i$ to be a multivariate normal with mean $\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_{i\gamma i}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}$ (2) and generate the random effects ($\boldsymbol{\gamma}_i$, ω_i) from the multivariate normal distribution with mean zero and covariance

$$\sum_{\boldsymbol{\gamma}\omega} = \begin{pmatrix} 1.5 & 0.5 & 0.3 & 0.2 \\ 0.5 & 1.0 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.5 & 0.05 \\ 0.2 & 0.1 & 0.05 & 1.0 \end{pmatrix}.$$

ω is a single random variable so the covariance matrix corresponds to 3 random location effects and 1 random scale effect, where $\sigma_\omega^2 = 1.0$.

5.2 (Multivariate T, Multivariate Gamma, Inverse Gamma) model

We set the conditional distribution of $\boldsymbol{Y}_i$ to be a multivariate T with degree of freedom $\nu = 6$, mean $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_{i\gamma i}$, and covariance $\boldsymbol{\Sigma}_{\epsilon_i}$, the random effects $\boldsymbol{\gamma}_i$ to be a multivariate Gamma shifted to have mean 0, and the WS random effect ω_i to be the natural logarithm of an inverse-gamma random deviate. In Bayesian statistics, inverse-gamma is commonly used as a conjugate prior for the variance of a normal model. Parameters of the Gamma and inverse Gamma deviates are chosen such that the covariance of $(\boldsymbol{\gamma}_i, \omega_i)$ is

$$\sum_{\gamma\omega} = \begin{pmatrix} 1.5 & 0.5 & 0.3 & 0.0 \\ 0.5 & 1.0 & 0.2 & 0.0 \\ 0.3 & 0.2 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}.$$

$\boldsymbol{\gamma}_i$ is independent of ω_i here but WiSER does not require this independence.

The parameter estimate mean squared error (MSE) for the two simulation scenarios at each combination of sample size (m) and observations per subject (n_i) are shown in Figure 3. The MSEs for the majority of parameter estimates are below 10^{-2}. There are a few outliers when estimating $\boldsymbol{\tau}$ in the (Multivariate T, Multivariate Gamma, Inverse Gamma) simulation, reflecting difficulty with heavy-tail distributions. Across all scenarios and parameters, the maximum percentage of outliers is 4% and median is 1%. These occur when the observations per individual are low ($n_i = 10$), which can be remedied by choosing a different starting point or using a different nonlinear optimization solver. Coverage at $\alpha = 0.05$ for each scenario is reported in Tables S.1 and S.2 (all close to the nominal value of 95%). We also report results of these simulations under smaller sample sizes ($m = 250$ and $m = 500$) in Web Appendix G, when asymptotic properties are less likely to hold.

5.3 Comparison with MLE

MLE for the mixed-effects multiple location scale model (1) is implemented in a comprehensive GUI software MixWILD (Dzubur et al., 2020), which wraps an efficient FORTRAN MLE engine (Hedeker and Nordgren (2013)). Unfortunately, despite its efficiency, MixWILD run times and its GUI design prevent a full scale comparison. Instead we choose the representative simulation replicate with the median MSE for estimating $\boldsymbol{\tau}$ by WiSER from the smallest sample size scenario ($m = 1000$, $n_i = 10$) and tally the results by WiSER and MixWILD along with the true parameter values in Table 1.

In Table 1, Models 1–3 represent different assumptions MixWILD makes in the mixed-effects multiple location scale model (1). Model 1 assumes $\boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega} = \mathbf{0}$ and $\sigma^2_\omega = 0$; Model 2 assumes $\boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega} = \mathbf{0}$; and Model 3 is the most general model which estimates all parameters $\left(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}, \sigma^2_\omega\right)$. Note WiSER can only estimate $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ because $\boldsymbol{\sigma}_{\boldsymbol{\gamma}\omega}$ and σ^2_ω are not identifiable. We observe that (1) WiSER estimates and standard errors for both $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are almost identical to MLEs, differing only in the third decimal place, (2) the run times of WiSER are $10^3 \sim 10^5$ times faster than MLE, and (3) the standard errors of WiSER estimates are overall larger (in the third decimal place) than those from MLE, reflecting a slight loss of efficiency in WiSER due to relaxing the distributional assumption. The efficiency

loss may have more impact for smaller sample sizes where the differences in computation time will be less pronounced. In these cases, likelihood-based methods are preferred. In this scenario with four random effects, the likelihood method requires numerical integration over Q^4 points (where Q is the number of Gaussian quadrature knots in one dimension of the integration). The computational time difference, while still notable, will be less so in a model with fewer random effects.

6. Real data analyses

6.1 An application to mobile health: Women's Health Study (WHS) accelerometry data

Habitual, lengthy sedentary behavior is a risk factor for a wide variety of long-term poor health outcomes that are distinct from negative health consequences due to a lack of regular exercise (moderate-to-vigorous physical activity) (Owen et al., 2010). Understanding factors associated with persistent sedentary behavior will lead to better-targeted interventions to encourage breaks in sedentary behavior. WHS is a randomized two-by-two factorial trial that took place between 1994 and 2002 to investigate the effects of vitamin E and aspirin in preventing cardiovascular disease and cancer among healthy women in the United States (Ridker et al., 2005). An ancillary study began in 2011, investigating links with physical activity (Lee et al., 2018). Women were sent accelerometers and asked to wear them for 7 days during waking hours. We apply WiSER to these data, looking at factors related to changes in the mean and within-subject variability of step count. To avoid strong daily periodicity and problems synchronizing the data between subjects, we restrict to the two most active hours for each individual each day. Vector magnitude, a measure of physical activity intensity, is reported in 1-minute epochs; these measurements are accumulated over each hour in order to identify a person's two most active hours in each day (Santos-Lozano et al., 2013). We use the number of steps taken over 5 minute epochs during these two hours as our outcomes.

The initial response variable, total steps every 5 minutes, has many zeros and a heavy tail. Although WiSER is robust to distributional assumptions, we compare its estimated mean effects $\widehat{\boldsymbol{\beta}}$ to the standard LMM, which assumes normality. In order to achieve a more normal distribution (for comparison with the LMM), we add 0.5 to each step count and take the $\log_{10}$ transformation to use as the response variable. The data set contains 2,534,015 observations on 15,390 individuals. Summary statistics are reported in Web Appendix H. Table 3 lists the mean effect estimates $\boldsymbol{\beta}$ by LMM and WiSER on the left and the estimates of the WS variability fixed effects $\boldsymbol{\tau}$ by WiSER on the right. The estimated mean effects $\widehat{\boldsymbol{\beta}}$ by LMM and WiSER are almost identical. Both LMM and WiSER also include a random intercept and a random slope for the day the device was worn (1 to 7). Their estimates are also similar and listed in Web Appendix H. The variable hour refers to the hour of the day (hour = 13 means during the 1 PM hour). WiSER reveals factors that are associated with activity level WS variability. For example, compared to Sunday, the participants have higher activity levels on Monday to Saturday but the variability is reduced. This may reflect the pattern that the two most active hours coincide with rush hours on weekdays while they are more sporadic on weekends (Althoff et al., 2017). Body mass index (BMI), hour of day, age, and smoking status are found to be associated with the WS activity variability. The negative

association of age and current smoking for the mean and variability in steps suggests older and smoking individuals are more sedentary, and thus a potential target population for interventions.

It takes our software package WiSER.jl 33 seconds to complete four WiSER estimation rounds on the WHS accelerometry data. LMM results come from the software package MixedModels.jl (Bates et al., 2020). We are not successful obtaining the MLE from MixWILD in a reasonable amount of time so no results from MixWILD are provided.

6.2 Which diabetes drug classes best control glycemic variation?

Although average glycemic levels, e.g., glycated hemoglobin (HbA1c), was considered the gold standard for assessing overall glycemic control (ADA, 2020), glycemic variability may be an even more meaningful measure in diabetes care (DeVries, 2013). Many pathophysiologic mechanisms could explain how glucose fluctuations cause vascular injury (Brownlee and Hirsch, 2006; Ceriello et al., 2008). Despite its clinical significance, there is no consensus on the optimal method for characterizing glycemic variability, partially due to the lack of statistical methodologies. Applying WiSER to the ACCORD trial, we evaluate and compare the effects of four widely adopted glucose lowering medication classes on both the mean glucose levels and the intra-individual glycemic variability. Our results show that metformin, meglitinides, and thiazolidinediones are more favorable treatments than insulin or sulphonylureas for controlling glycemic variability.

ACCORD was a double-blinded, two-by-two factorial, randomized, parallel treatment trial in which 10,251 participants were assigned to receive either an intensive treatment targeting HbA1c of < 6.0% (42.1 mmol/mol) or a standard treatment targeting HbA1c of 7.0–7.9% (53–62.8 mmol/mol). Participants had T2D, HbA1c concentrations of 7.5% (58.5 mmol/mol) or more, and were 40–79 years old with a history of cardiovascular disease or 55–79 years old with evidence of significant atherosclerosis, albuminuria, left ventricular hypertrophy, or at least two risk factors for cardiovascular disease (dyslipidemia, hypertension, smoking, or obesity). During the study, glucose concentrations were measured every 4 months in the initial year, then annually up to a maximum of 84 months. The design and principal results of ACCORD trial were reported previously (Ismail-Beigi et al., 2010; ACCORD et al., 2008).

Our analysis uses all in-study glucose measures of the full ACCORD study, 67,063 observations on 10,195 individuals. Data preparation details are provided in Web Appendix I and summary statistics are reported in Supplementary Table S.3. In order to control glucose at specific levels within each of the treatment arms in ACCORD, glycemic management is well-documented, including the type and dose of medications taken at each visit throughout the study period. Table 4 reports WiSER estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$. In addition to the covariates in the table, we include a random intercept and a random slope effect for treatment month in the model. Their estimates are listed in Web Appendix I. We follow Siraj et al. (2015) and use insulin units per body weight in kg (adjusted insulin) instead of raw total insulin units. We find the month of treatment, BMI, age, race, cardiovascular disease history at baseline, and adjusted insulin (combined dosage from Basal, Bolus, and premixed), and certain oral medication classes to be significantly associated with the mean and intra-

individual variability of fasting plasma glucose. Interestingly, adjusted insulin is associated with a lower mean and sulphonylureas have little effect on the mean, but they increase the intra-individual variability of fasting plasma glucose. Meglitinides are associated with significantly lower glucose variability. Sulphonylureas and meglitinides are both second line oral-therapies for T2D patients and have similar clinical effects, but meglitinides lead to fewer hypoglycemic events than sulphonylureas (Grant and Graven, 2016). Although our findings require validation in other clinical studies, they demonstrate the capability of WiSER to characterize glucose variability using complex longitudinal data and modifiable factors identified can be used to develop future interventions.

7. Discussion

We demonstrated WiSER as an efficient tool for analyzing WS variance with massive intensive longitudinal data. While relaxing the strict distributional assumptions in mixed models, WiSER estimates show comparable efficiency as the correctly specified MLE but take orders of magnitude less time. However, when the interest lies in estimating the random scale variance σ^2_ω, the random scale location covariances $\ell_{\gamma\omega}$, or individual level estimates of random effects $\boldsymbol{\gamma}_i$ or ω_i, the likelihood approach should be used. Estimate of ω_i can be useful for identifying unusual subjects, e.g., those that have extremely high or low within-subject variance. Another obvious application is to use WiSER estimates as warm starts for likelihood methods. This strategy could reduce the number of iterations in the expensive likelihood-based inference procedures. WiSER can be extended in many directions, which we outline here.

We focused on quantitative outcomes as dictated by our motivating examples. In principle, WiSER accommodates qualitative responses since only the conditional moment condition is assumed. Alternatively, as with GEEs, we can apply a link function to the mean systematic component $\boldsymbol{X}_i\boldsymbol{\beta}$ and model the intra-individual covariance as $\boldsymbol{V}_i = \mathrm{diag}\left(e^{0.5\boldsymbol{W}_i\boldsymbol{\tau}}\right) \cdot \boldsymbol{R}_i \cdot \mathrm{diag}\left(e^{0.5\boldsymbol{W}_i\boldsymbol{\tau}}\right)$, where $\boldsymbol{R}_i$ is an $n_i \times n_i$ working correlation matrix. However we lose the obvious interpretation of WS and BS variability and the computational scalability in the intensive longitudinal measurement setting is a concern. The log-linear link for the WS variance systematic component $\boldsymbol{W}\boldsymbol{\tau}$ can be relaxed to any monotone, positive link function.

Consistency and asymptotic normality of WiSER for fixed numbers of parameters are established assuming that the observation tuples $(\boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i)$ are iid, recognizing the great variability in the number of observations per individual. The large n_i (Xie and Yang, 2003) and diverging p (Wang, 2011) asymptotics are particularly relevant in the high-dimensional GEE models and needs to be investigated in the WiSER setting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was partially funded by the Burroughs Wellcome Fund Inter-school Training Program in Chronic Diseases (CAG) and grants from the National Institute of General Medical Sciences (GM053275, JSS and HZ; GM141798, JSS and HZ), the National Human Genome Research Institute (HG009120, JSS; HG006139, HZ), the National Science Foundation (DMS-1264153, JSS; DMS-2054253, HZ), the National Institute of Diabetes and Digestive and Kidney Disease (K01DK106116, JJZ), and the National Heart, Lung, and Blood Institute (R21HL150374, JJZ).

References

ACCORD SG, Gerstein H, Miller M, Byington R, Goff DJ, Bigger J, Buse J, Cushman W, Genuth S, Ismail-Beigi F, Grimm RJ Jr, Probstfield JL, Simons-Morton DG, and Friedewald WT (2008). Effects of intensive glucose lowering in type 2 diabetes. New England Journal of Medicine 358, 2545–2559. [PubMed: 18539917]

ADA (2020). 2. classification and diagnosis of diabetes: Standards of medical care in diabetes-2020. Diabetes Care 43, S14. [PubMed: 31862745]

Althoff T, Sosič R, Hicks JL, King AC, Delp SL, and Leskovec J (2017). Largescale physical activity data reveal worldwide activity inequality. Nature 547, 336–339. [PubMed: 28693034]

Barrett JK, Huille R, Parker R, Yano Y, and Griswold M (2019). Estimating the association between blood pressure variability and cardiovascular disease: an application using the ARIC study. Statistics in Medicine 38, 1855–1868. [PubMed: 30575102]

Bates D, Alday P, Kleinschmidt D, Calderón JBS, Noack A, Kelman T, Bouchet-Valat M, Gagnon YL, Babayan S, Mogensen PK, Piibeleht M, Hatherly M, Saba E, and Baldassari A (2020). Juliastats/mixedmodels.jl: v2.3.0.

Boos DD (1992). On generalized score tests. The American Statistician 46, 327–333.

Boos DD and Stefanski LA (2013). Essential Statistical Inference. Springer Texts in Statistics. Springer, New York. Theory and methods.

Brownlee M and Hirsch IB (2006). Glycemic variability: a hemoglobin a1c–independent risk factor for diabetic complications. JAMA 295, 1707–1708. [PubMed: 16609094]

Byrd RH, Nocedal J, and Waltz RA (2006). Knitro: An integrated package for nonlinear optimization. In Large-Scale Nonlinear Optimization, pages 35–59. Springer.

Ceriello A, Esposito K, Piconi L, Ihnat MA, Thorpe JE, Testa R, Boemi M, and Giugliano D (2008). Oscillating glucose is more deleterious to endothelial function and oxidative stress than mean glucose in normal and type 2 diabetic patients. Diabetes 57, 1349–1354. [PubMed: 18299315]

Ceriello A, Monnier L, and Owens D (2019). Glycaemic variability in diabetes: clinical and therapeutic implications. The Lancet Diabetes & Endocrinology 7, 221–230. [PubMed: 30115599]

DeVries JH (2013). Glucose variability: where it is important and how to measure it. Diabetes 62, 1405–1408. [PubMed: 23613566]

Dzubur E, Ponnada A, Nordgren R, Yang C-H, Intille S, Dunton G, and Hedeker D (2020). MixWILD: A program for examining the effects of variance and slope of time-varying variables in intensive longitudinal data. Behavior Research Methods 52, 1403–1427. [PubMed: 31898295]

Fitzmaurice GM, Laird NM, and Ware JH (2011). Applied Longitudinal Analysis. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition.

Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, and O'Leary TJ (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. Journal of Clinical Epidemiology 70, 214–223. [PubMed: 26441289]

GitHub (2021). https://github.com/chris-german/WiSER-Reproducibility.

Goldstein H, Leckie G, Charlton C, Tilling K, and Browne WJ (2017). Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. Statistical Methods in Medical Research 27, 3478–3491. [PubMed: 28459180]

Grant JS and Graven LJ (2016). Progressing from metformin to sulfonylureas or meglitinides. Workplace Health & Safety 64, 433–439. [PubMed: 27621259]

Hager WW (1989). Updating the inverse of a matrix. SIAM Rev. 31, 221–239.

Hedeker D, Mermelstein RJ, and Demirtas H (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics 64, 627–634. [PubMed: 17970819]

Hedeker D and Nordgren R (2013). MIXREGLS: A program for mixed-effects location scale analysis. Journal of Statistical Software 52, 1–38. [PubMed: 23761062]

Heron KE, Everhart RS, McHale SM, and Smyth JM (2017). Using mobiletechnology-based ecological momentary assessment (EMA) methods with youth: a systematic review and recommendations. Journal of Pediatric Psychology 42, 1087–1107. [PubMed: 28475765]

Ismail-Beigi F, Craven T, Banerji MA, Basile J, Calles J, Cohen RM, Cuddihy R, Cushman WC, Genuth S, Grimm RH Jr, et al. (2010). Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomised trial. The Lancet 376, 419–430.

Ivarsdottir EV, Steinthorsdottir V, Daneshpour MS, Thorleifsson G, Sulem P, Holm H, Sigurdsson S, Hreidarsson AB, Sigurdsson G, Bjarnason R, et al. (2017). Effect of sequence variants on variance in glucose levels predicts type 2 diabetes risk and accounts for heritability. Nature Genetics 49, 1398. [PubMed: 28783164]

Johnson SG (2020). The NLopt nonlinear-optimization package.

Lee I-M, Shiroma EJ, Evenson KR, Kamada M, LaCroix AZ, and Buring JE (2018). Accelerometer-measured physical activity and sedentary behavior in relation to all-cause mortality. Circulation 137, 203–205. [PubMed: 29109088]

Leng C, Zhang W, and Pan J (2010). Semiparametric mean-covariance regression analysis for longitudinal data. J. Amer. Statist. Assoc 105, 181–193.

Lin X, Raz J, and Harlow SD (1997). Linear mixed models with heterogeneous within-cluster variances. Biometrics 53, 910–923. [PubMed: 9290222]

NIH BioLINCC (2021). https://biolincc.nhlbi.nih.gov/studies/accord/.

NIH dbGaP (2021). https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001964.v1.p1.

Owen N, Sparling PB, Healy GN, Dunstan DW, and Matthews CE (2010). Sedentary behavior: emerging evidence for a new health risk. Mayo Clinic Proceedings 85, 1138–1141. [PubMed: 21123641]

Prentice RL (1988). Correlated binary regression with covariates specific to each binary observation. Biometrics 44, 1033–1048. [PubMed: 3233244]

Rast P, Hofer SM, and Sparks C (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. Multivariate Behavioral Research 47, 177–200. [PubMed: 26734847]

Ridker PM, Cook NR, Lee I-M, Gordon D, Gaziano JM, Manson JE, Hennekens CH, and Buring JE (2005). A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. New England Journal of Medicine 352, 1293–1304. [PubMed: 15753114]

Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlöf B, Poulter NR, and Sever PS (2010). Effects of β blockers and calcium-channel blockers on within-individual variability in blood pressure and risk of stroke. The Lancet Neurology 9, 469–480. [PubMed: 20227347]

Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlöf B, Sever PS, and Poulter NR (2010). Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. The Lancet 375, 895–905.

Russell MA and Gajos JM (2020). Annual research review: ecological momentary assessment studies in child psychology and psychiatry. Journal of Child Psychology and Psychiatry 61, 376–394. [PubMed: 31997358]

Ruwaard J, Kooistra L, and Thong M (2018). Ecological Momentary Assessment in Mental Health Research: A Practical Introduction, with Examples in R. Amsterdam: APH Mental Health, 1st edition.

Santos-Lozano A, Santin-Medeiros F, Cardon G, Torres-Luque G, Bailon R, Bergmeir C, Ruiz JR, Lucía Mulas A, and Garatachea N (2013). Actigraph GT3X: Validation and determination of physical activity intensity cut points. Int J Sports Med 34, 975–982. [PubMed: 23700330]

Siraj ES, Rubin DJ, Riddle MC, Miller ME, Hsu F-C, Ismail-Beigi F, Chen S-H, Ambrosius WT, Thomas A, Bestermann W, Buse JB, Genuth S, Joyce C, Kovacs CS, O'Connor PJ, Sigal RJ, and Solomon S (2015). Insulin dose and cardiovascular mortality in the ACCORD trial. Diabetes Care 38, 2000. [PubMed: 26464212]

Smit RAJ, Jukema JW, Postmus I, Ford I, Slagboom PE, Heijmans BT, Le Cessie S, and Trompet S (2018). Visit-to-visit lipid variability: Clinical significance, effects of lipid-lowering treatment, and (pharmaco) genetics. Journal of Clinical Lipidology 12, 266–276.e3. [PubMed: 29422286]

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Medicine 12, e1001779. [PubMed: 25826379]

Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, Vittinghoff E, Lee ES, Fan SM, et al. (2018). Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiology 3, 409–416. [PubMed: 29562087]

Tsiatis AA (2006). Semiparametric Theory and Missing Data. Springer Series in Statistics. Springer, New York.

van der Vaart AW (1998). Asymptotic Statistics, volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Verbeke G and Molenberghs G (2009). Linear Mixed Models for Longitudinal Data. Springer Series in Statistics. Springer, New York. Reprint of the 2000 original.

Wächter A and Biegler LT (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106, 25–57.

Wang L (2011). GEE analysis of clustered binary data with diverging number of covariates. Ann. Statist 39, 389–417.

WiSER.jl (2021). https://github.com/OpenMendel/WiSER.jl.

Xie M and Yang Y (2003). Asymptotics for generalized estimating equations with large cluster sizes. Ann. Statist 31, 310–347.

Ye H and Pan J (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. Biometrika 93, 927–941.

Zhao LP and Prentice RL (1990). Correlated binary regression using a quadratic exponential model. Biometrika 77, 642–648.

Zhou JJ, Schwenke DC, Bahn G, and Reaven P (2018). Glycemic variation and cardiovascular risk in the veterans affairs diabetes trial. Diabetes Care 41, 2187. [PubMed: 30082325]

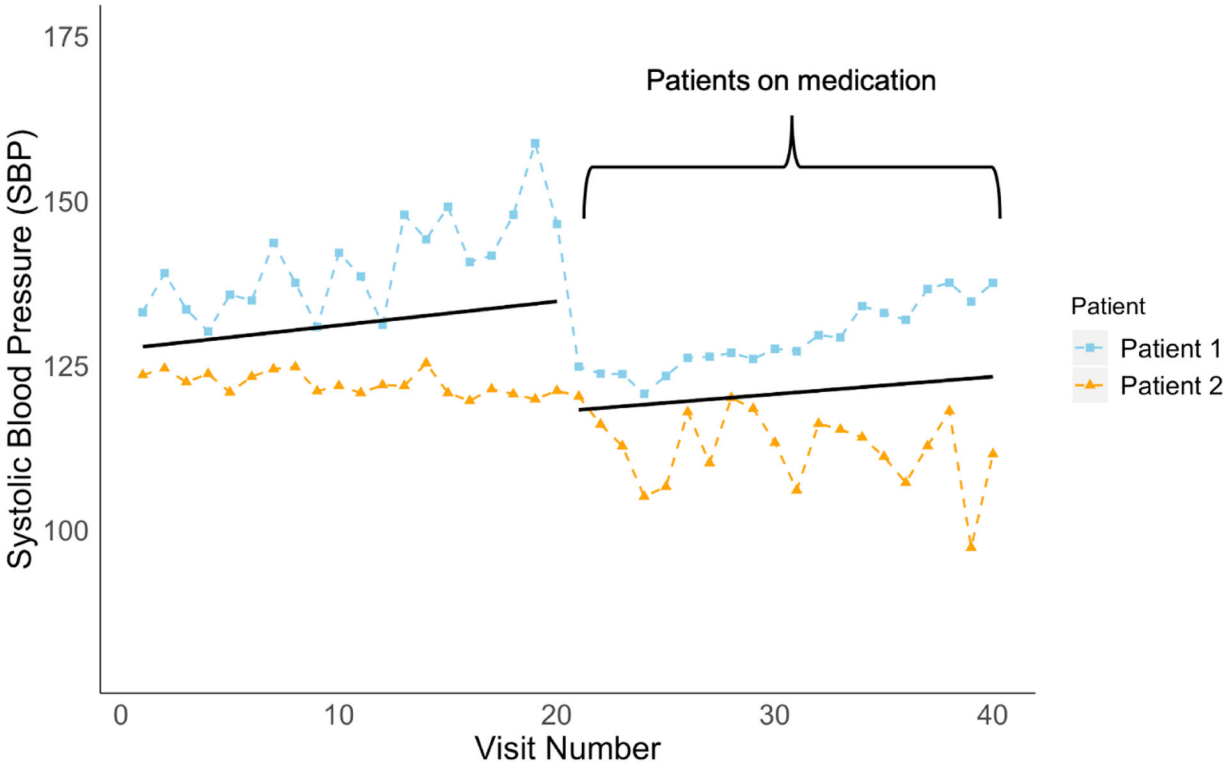

Figure 1:
Within-subject variability changes with time-varying covariates such as medication use. Patient 1 has higher blood pressure (BP) variability than Patient 2 before starting medication due to gender. After starting BP lowering medications, Patient 1 (on a calcium channel-blocker) has decreased BP variability and Patient 2 (on a β-blocker) has increased BP variability. WiSER models both time-varying and time-invariant influences on within-subject BP variability.

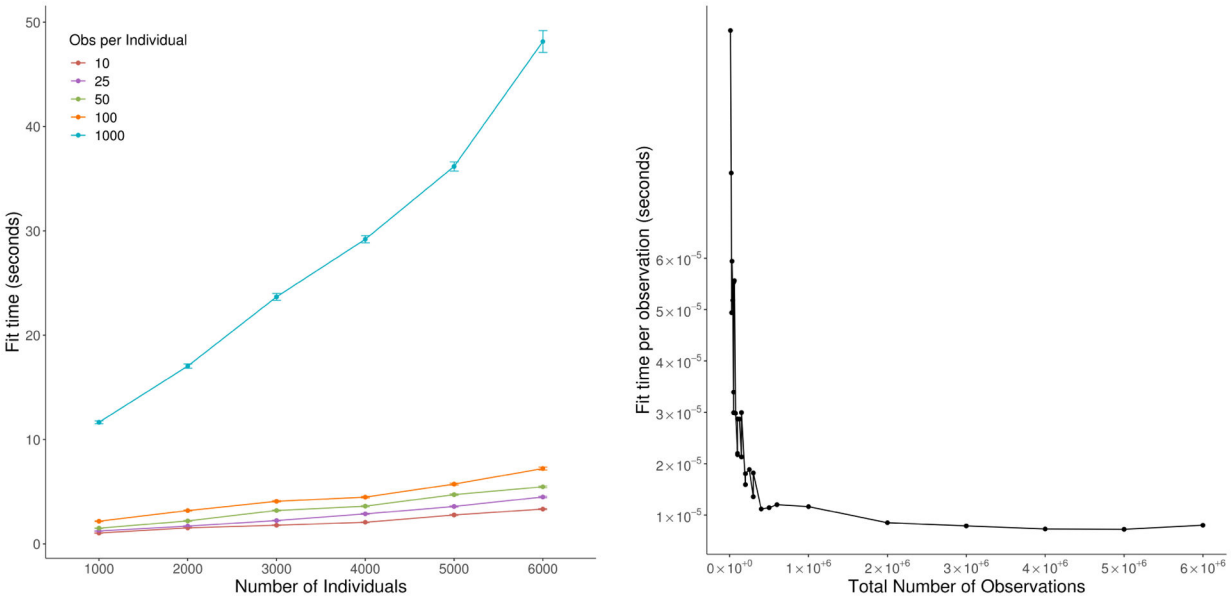

Figure 2:

Computational complexity of WiSER scales linearly in the total number of observations. The left panel plots the total run times versus the number of individuals; each line represents a fixed number of observations per individual. The right panel demonstrates that the average time per observation stabilizes to a constant at large sample sizes.

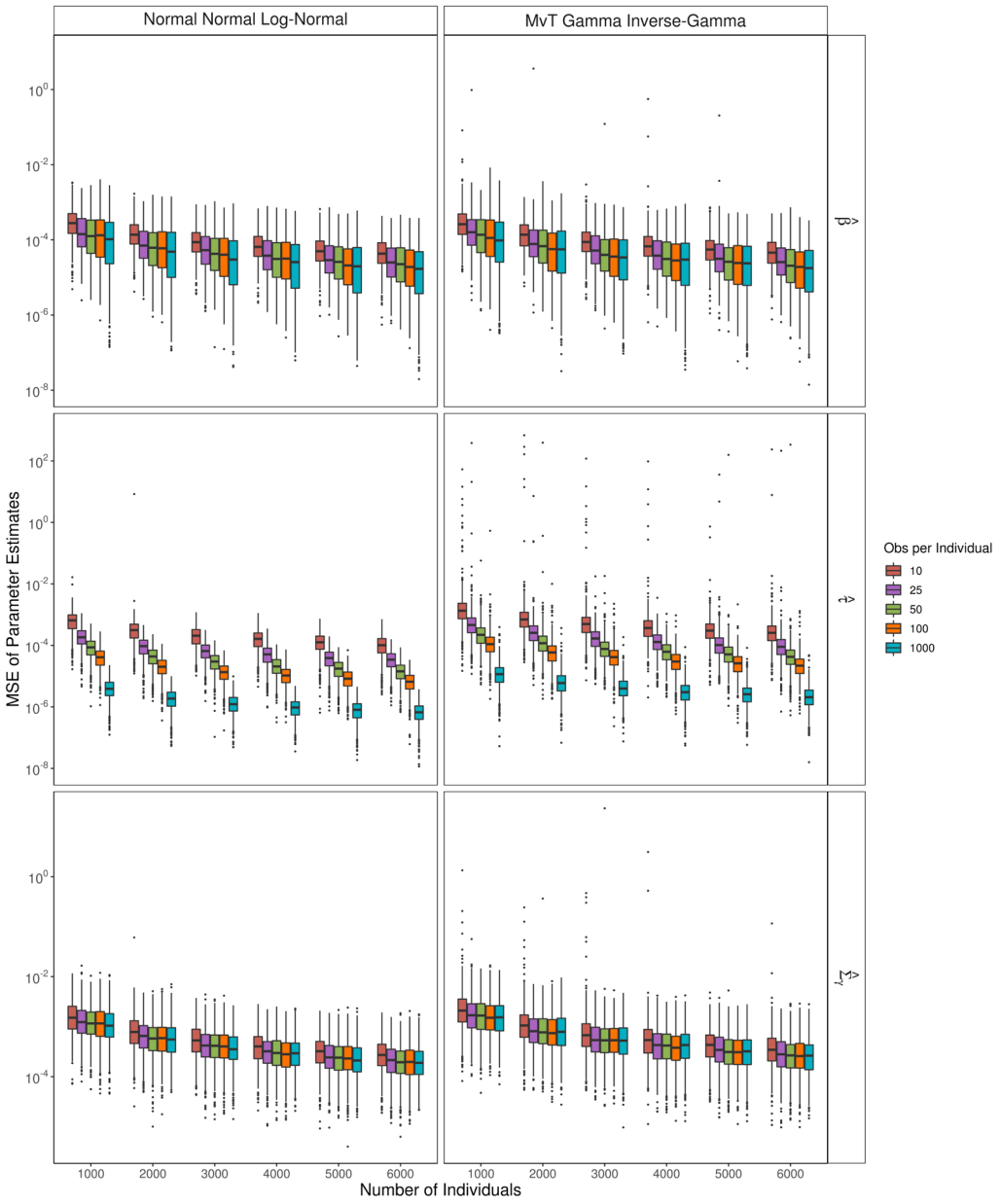

Figure 3:
Mean squared error (MSE) of WiSER parameter estimates of $\boldsymbol{\beta}$ (top row), $\boldsymbol{\tau}$ (middle row), and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ (bottom row) under the (Normal, Normal, Log-Normal) (left column) and (Multivariate T, Multivariate Gamma, Inverse Gamma) (right column) models. Each scenario reports results from 1000 replicates.

Table 1:

WiSER achieves nearly the same accuracy as the maximum likelihood estimate (as implemented in MixWILD) but is $10^3 \sim 10^5$ faster on two simulated data sets with 1000 individuals and 10 observations per individual. Displayed are point estimates with standard errors in the parentheses. Simulation details are described in Sections 5.1–5.3.

			Maximum likelihood estimate (MixWILD)		
Coefficient	**Truth**	**WiSER**	**Model 1**	**Model 2**	**Model 3**
β_1	0.1	0.110 (0.037)	0.110 (0.037)	0.089 (0.034)	0.109 (0.035)
β_2	6.5	6.509 (0.013)	6.510 (0.013)	6.512 (0.010)	6.513 (0.010)
β_3	−3.5	−3.489 (0.013)	−3.490 (0.013)	−3.503 (0.011)	−3.502 (0.010)
β_4	1.0	0.984 (0.013)	0.984 (0.013)	0.986 (0.010)	0.985 (0.010)
β_5	5.0	4.979 (0.012)	4.979 (0.013)	4.981 (0.010)	4.980 (0.010)
τ_1	0.0	0.358 (0.037)	0.358 (0.017)	0.051 (0.031)	0.061 (0.029)
τ_2	0.5	0.545 (0.029)	0.545 (0.018)	0.514 (0.021)	0.519 (0.021)
τ_3	−0.2	−0.189 (0.027)	−0.190 (0.018)	−0.191 (0.020)	−0.188 (0.020)
τ_4	0.5	0.490 (0.024)	0.492 (0.019)	0.485 (0.021)	0.487 (0.020)
τ_5	0.0	−0.012 (0.027)	−0.012 (0.018)	0.009 (0.020)	0.010 (0.020)
Runtime (s)		0.37	2350	30030	34129

(a) (Normal, Normal, Log-Normal) model.

			Maximum likelihood estimate (MixWILD)		
Coefficient	**Truth**	**WiSER**	**Model 1**	**Model 2**	**Model 3**
β_1	0.1	0.149 (0.038)	0.150 (0.038)	0.156 (0.032)	0.151 (0.032)
β_2	6.5	6.514 (0.011)	6.514 (0.012)	6.515 (0.010)	6.515 (0.010)
β_3	−3.5	−3.503 (0.012)	−3.503 (0.012)	−3.511 (0.010)	−3.512 (0.010)
β_4	1.0	1.031 (0.013)	1.032 (0.012)	1.032 (0.010)	1.032 (0.010)
β_5	5.0	5.007 (0.012)	5.008 (0.012)	5.004 (0.010)	5.004 (0.010)
τ_1	0.0	0.237 (0.040)	0.238 (0.017)	−0.080 (0.031)	−0.081 (0.031)
τ_2	0.5	0.540 (0.030)	0.536 (0.019)	0.532 (0.021)	0.531 (0.021)
τ_3	−0.2	−0.213 (0.032)	−0.213 (0.019)	−0.229 (0.021)	−0.228 (0.021)
τ_4	0.5	0.471 (0.028)	0.464 (0.019)	0.495 (0.022)	0.494 (0.022)
τ_5	0.0	0.051 (0.032)	0.050 (0.018)	0.014 (0.020)	0.015 (0.020)
Runtime (s)		0.49	2490	56788	29977

(b) (Multivariate T, Gamma, Inverse Gamma) model.

Table 2:

Symbols used to describe the WiSER model.

m	≜	number of subjects
n_i	≜	number of observations for subject i
q	≜	number of random effects
p	≜	number of fixed effects
ℓ	≜	number of variables affecting within-subject (WS) variance
$\boldsymbol{\beta}$	≜	$p\times 1$ coefficient vector of fixed effects
$\boldsymbol{\gamma}_i$	≜	$q \times 1$ coefficient vector of random effects of subject i (random-location effects) with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$
$\boldsymbol{\tau}$	≜	$\ell\times 1$ coefficient vector of WS effects
ω_i	≜	random intercept in WS variance of subject i (random-scale parameter) with mean 0 and variance σ^2_{ω}
$\boldsymbol{y}_i$	≜	$n_i \times 1$ vector of observed responses for subject i
$\boldsymbol{X}_i$	≜	$n_i \times p$ matrix of fixed effects covariates for subject i
$\boldsymbol{Z}_i$	≜	$n_i \times q$ matrix of random effects covariates for subject i
$\boldsymbol{W}_i$	≜	$n_i \times \ell$ matrix of covariates affecting WS variance, $\sigma^2_{\epsilon_{ij}}$, for subject i
$\boldsymbol{e}_i$	≜	$n_i \times 1$ vector of error term reflecting WS variance

Table 3:

WiSER identifies factors associated with mean and variation of women's activity levels from the Women's Health Study (WHS) accelerometry data with 2.5 million observations on 15,390 women.

	LMM β		WiSER β	
Covariate	**Estimate**	**p value**	**Estimate**	**p value**
(Intercept)	2.4789	<1e-99	2.5266	<1e-99
BMI	−0.0169	<1e-99	−0.0169	<1e-99
Weekday: Mon	0.0842	<1e-99	0.0844	<1e-99
Weekday: Tues	0.0676	<1e-99	0.0678	<1e-81
Weekday: Wed	0.0650	<1e-99	0.0653	<1e-72
Weekday: Thurs	0.0570	<1e-99	0.0574	<1e-55
Weekday: Fri	0.0722	<1e-99	0.0724	<1e-89
Weekday: Sat	0.0735	<1e-99	0.0730	<1e-99
Hour	−0.0029	<1e-99	−0.0037	<1e-45
Race: African American	−0.0160	0.3216	−0.0166	0.3039
Race: Asian	−0.0849	<1e-5	−0.0827	<1e-5
Race: Hispanic	0.0481	0.0214	0.0480	0.0135
Race: Native American	−0.0109	0.8150	−0.0089	0.8493
Race: Other	0.0069	0.9058	0.0085	0.8893
Stairs	0.0134	<1e-22	0.0134	<1e-23
Age	−0.0174	<1e-99	−0.0174	<1e-99
Smoker: Past	0.0034	0.4000	0.0028	0.4832
Smoker: Current	−0.1492	<1e-43	−0.1497	<1e-36
Season: Spring	−0.0047	0.4064	−0.0038	0.5025
Season: Summer	0.0050	0.3456	0.0048	0.3508
Season: Winter	−0.0367	<1e-9	−0.0360	<1e-9
Total Minutes Worn (Daily)	0.0007	<1e-99	0.0007	<1e-99

(a) Mean fixed effects, $\boldsymbol{\beta}$, estimated by LMM and WiSER.

	WiSER τ	
Covariate	**Estimate**	**p value**
(Intercept)	−0.1551	0.0011
BMI	0.0010	0.0912
Weekday: Mon	−0.0647	<1e-25
Weekday: Tues	−0.0574	<1e-20
Weekday: Wed	−0.0620	<1e-22
Weekday: Thurs	−0.0633	<1e-23
Weekday: Fri	−0.0818	<1e-40
Weekday: Sat	−0.0813	<1e-40
Hour	−0.0099	<1e-54
Race: African American	−0.1078	<1e-5
Race: Asian	0.0413	0.1509
Race: Hispanic	−0.0251	0.4420
Race: Native American	0.0597	0.3675
Race: Other	0.0276	0.8011
Age	−0.0014	0.0058
Smoker: Past	0.0063	0.3006
Smoker: Current	−0.0449	0.0036
Total Minutes Worn (Daily)	−0.0003	<1e-52

(b) Within-subject variance fixed effects, $\boldsymbol{\tau}$, estimated by WiSER.

Table 4:

WiSER estimates the effects of various factors on the mean glucose level and glycemic variation from the ACCORD data with 67,063 observations on 10,195 individuals.

	WiSER β		WiSER τ	
Covariate	**Estimate**	**p-value**	**Estimate**	**p-value**
Intercept	219.0090	<1e-99	8.4800	<1e-99
Visit Number	−0.2144	<1e-94	−0.0078	<1e-27
BMI	−0.0368	0.5230	−0.0138	<1e-7
Female	−1.3908	0.0392	0.0229	0.3799
Baseline Age	−0.7471	<1e-48	−0.0121	<1e-7
Race: Black	−8.5492	<1e-22	0.2493	<1e-12
Race: Hispanic	−2.2693	0.0801	0.2066	<1e-4
Race: Other	−1.2686	0.2578	0.0836	0.0335
Baseline CVD History	0.9638	0.1594	0.0595	0.0236
Total Injected Insulin (units/kg body weight)	−14.8855	<1e-61	0.8075	<1e-99
Sulphonylureas	−0.5211	0.3407	0.3036	<1e-29
Metformin	−5.5822	<1e-15	−0.1356	<1e-4
Meglitinides	−13.4449	<1e-99	−0.3021	<1e-16
Thiazolidinediones	−20.2340	<1e-99	−0.0194	0.4222