Linear mixed models with endogenous covariates: modeling sequential treatment effects with application to a mobile health study

Tianchen Qian, Predrag Klasnja and Susan A. Murphy

Department of Statistics, Harvard University, Cambridge, MA 02138
e-mail: qiantianchen@fas.harvard.edu; samurphy@fas.harvard.edu.

School of Information, University of Michigan, Ann Arbor, MI 48109
e-mail: klasnja@umich.edu.

Abstract: Mobile health is a rapidly developing field in which behavioral treatments are delivered to individuals via wearables or smartphones to facilitate health-related behavior change. Micro-randomized trials (MRT) are an experimental design for developing mobile health interventions. In an MRT the treatments are randomized numerous times for each individual over course of the trial. Along with assessing treatment effects, behavioral scientists aim to understand between-person heterogeneity in the treatment effect. A natural approach is the familiar linear mixed model. However, directly applying linear mixed models is problematic because potential moderators of the treatment effect are frequently endogenous—that is, may depend on prior treatment. We discuss model interpretation and biases that arise in the absence of additional assumptions when endogenous covariates are included in a linear mixed model. In particular, when there are endogenous covariates, the coefficients no longer have the customary marginal interpretation. However, these coefficients still have a conditional-on-the-random-effect interpretation. We provide an additional assumption that, if true, allows scientists to use standard software to fit linear mixed model with endogenous covariates, and person-specific predictions of effects can be provided. As an illustration, we assess the effect of activity suggestion in the HeartSteps MRT and analyze the between-person treatment effect heterogeneity.

Keywords and phrases: linear mixed model, endogenous covariates, micro-randomized trial, causal inference.

1. Introduction

Mobile health (mHealth) refers to the use of mobile phones and other wireless devices to improve health outcomes, often by providing individuals with support for health-related behavior change. One major category of time-varying treatments delivered through mobile devices, which is the focus of this paper, are "push interventions"; in this setting, the mobile device determines when a treatment will be provided, rather than the individual seeking the intervention of her own accord (e.g., by opening the app). Push interventions are usually provided via some kind of a notification, such as an audible ping, vibration, or the lock screen of a phone lightening up. For example, to encourage physical activity in sedentary individuals, the HeartSteps intervention sends users push notifications that contain contextually-tailored activity suggestions (Klasnja et al., 2018).

Micro-randomized trials (MRTs) provide an experimental design for developing mHealth interventions. These trials provide longitudinal data to assess whether there is an effect of a time-varying

treatment, how this effect changes over time, and whether aspects of the current context impact the effect (Liao et al., 2016; Dempsey et al., 2015). In an MRT, each individual is randomized repeatedly to different versions of a treatment (or no treatment) with a known probability over the course of the trial (often hundreds or even thousands of times). Between randomizations, the trial collects covariate data on the individual's current/recent context via sensors and self-report, and after each randomization it assesses a proximal outcome. The large number of randomization points likely covers a wide range of contexts, and methods that exploit this for assessing effect moderation of a time-varying treatment have been developed (Boruvka et al., 2018).

Random effects models (Laird and Ware, 1982; Raudenbush and Bryk, 2002), sometimes also known as mixed effect models, hierarchical models, or multilevel models, have been used with great success in the analysis of longitudinal studies. Behavioral scientists, and researchers from many other scientific fields, have long used random effects model in research involving longitudinal data (Agresti et al., 2000; Berger and Tan, 2004; Cheung, 2008; Luger, Suls and Vander Weg, 2014). A particularly appealing feature of random effects models is the ability to predict person-specific random effects, which enables quantitative characterization of between-person heterogeneity due to unobserved factors (Schwartz and Stone, 2007; Bolger and Laurenceau, 2013). Understanding such heterogeneity can bring forth new scientific hypotheses for further studies. In addition, the random effects provide a model for the within-person dependence in the time-varying outcome, which improves efficiency in parameter estimation. Because data from an MRT is longitudinal, it is natural to consider a random effects model when making inference about treatment effects using MRT data.

However, random effects models were designed for settings where the covariates are considered fixed, and inferential challenges arise when one tries to apply the standard random effects model if there are endogenous time-varying covariates. A time-varying covariate is endogenous if this covariate is not independent of previous treatment or outcomes; we give a more precise definition in Section 1.2. As written above, MRTs are conducted to make inference about the effect of a time-varying treatment, how this effect changes over time, and whether certain aspects of the current context impact the effect. Covariates, often endogenous, describe the individual's context, and it is often of scientific interest to assess if the time-varying treatment is moderated by certain endogenous covariates. Furthermore, to reduce variance in assessing treatment effects, it is very useful to control for an endogenous covariate in the analysis (Boruvka et al., 2018). For example, consider HeartSteps, an MRT of an intervention that aims to increase physical activity among sedentary adults (Klasnja et al., 2018). In this study the treatments are contextually-tailored activity suggestions. The steps taken by the individual during the 30 minutes prior to randomization is likely highly correlated with the primary proximal outcome, the step count in the subsequent 30 minutes. Thus it is useful to control for this covariate in the analysis as well as to assess whether this covariate moderates the effect of the activity suggestion on the subsequent 30-minute step count. However, because the activity suggestions are randomized roughly every 2 hours, it is likely that the 30-minute step count prior to randomization is related to past step counts (i.e., past outcomes) as well as past treatment, which makes it an endogenous covariate. As we discuss below, including endogenous covariates in random effects models can result in biased estimates.

A related but different concept to an endogenous covariate is a time-varying confounder. Recall that a time-varying confounder, sometimes also called a time-dependent confounder, is a covariate that is affected by previous treatment (hence is endogenous) and affects future treatment assignment (Daniel et al., 2013; Hernán and Robins, 2019). To our surprise, even without time-varying confounding (e.g., when the randomization probability is constant in an MRT), the inclusion of

endogenous covariates in random effects models can cause bias in assessment of the treatment effects.

Pepe and Anderson (1994) pointed out that when using generalized estimating equations (GEE) with endogenous covariates, one should use working independence correlation structure to avoid biased estimates. Diggle et al. (2002), in their classic monograph on longitudinal data analysis, noted that:

> "Although Pepe and Anderson (1994) focused on the use of GEE, the issue that they raise is important for all longitudinal data analysis methods including likelihood-based methods such as linear and generalized linear mixed models."

In this paper, we focus on linear mixed models (LMM), a simple form of random effects models where the outcome is continuous and the link function is identity. We review how problems arise when endogenous covariates are included in LMM. Coefficients, and specifically treatment effects, in a standard LMM with fixed covariates have both marginal and conditional-on-the-random-effect interpretations. But the marginal interpretation is no longer valid with endogenous covariates.

Fortunately, despite losing the marginal interpretation, the conditional interpretation of the parameters is consistent with scientific interest in the prediction of person-specific effects in MRTs. Here we propose to interpret treatment effects as conditional on the random effect in LMM with possibly endogenous covariates. We provide an additional assumption under which valid estimates of the effect (conditional on the random effect) of the time-varying treatment, estimates of the variance components, and person-specific predictions of these treatment effects can be obtained through standard LMM software, even if some covariates are endogenous. Simulation studies are conducted to support the main result.

Lastly, we discuss whether and when the aforementioned assumption makes sense in HeartSteps, and analyze the data using the proposed method.

The paper is organized as follows. We provide an overview of the HeartSteps MRT in Section 1.1. We introduce notation and definition in Section 1.2. In Section 2 we give a detailed account of the issue regarding endogenous covariates in a standard LMM, and review related literature in causal inference (Section 2.3) and econometrics (Section 2.4). Next we provide an assumption under which treatment effects can be estimated based on LMM with endogenous covariates in Section 3. In Section 4 we present results from a simulation study. We apply the proposed model to analyzing the HeartSteps data in Section 5. Section 6 concludes with discussion.

1.1. Motivating Example: HeartSteps

Our motivating example is from HeartSteps, a 6-week MRT of an mHealth intervention to encourage regular walking among sedentary adults (Klasnja et al., 2018). The intervention package in HeartSteps includes multiple components; in this paper we focus on one push intervention component as the treatment, which is the activity suggestions. Each individual is in the study for 42 days, and is randomized 5 times a day, each time with probability 0.6 to receive an activity suggestion. The 5 randomization times are pre-specified and individual-specific, corresponding to each individual's morning commute, lunchtime, mid-afternoon, evening commute, and after-dinner. The content of the suggestion was tailored to the current time of day, weekend vs weekday, weather, and the individuals current location. The activity suggestions were designed to help individuals get activity throughout the day. Due to the tailoring of the suggestions to the individuals current context, the research team expected to see the greatest impact of the activity suggestions on near

time, proximal activity, so the proximal outcome is defined as the individual's step count during the 30 minutes following each randomization. In addition to the step counts, at each randomization the individual's context is also recorded, including current location, weather and 30-minute step count prior to randomization. Note that the 30-minute step count prior to the time of randomization is likely impacted by prior treatment and thus is an endogenous covariate. In addition to the measured information, there are other unobserved variables that may impact the treatment effect, such as each individual's commitment to becoming more active, conscientiousness, degree of social support and so on. Therefore, it is of interest to provide person-specific predictions of treatment effect. We will apply methods developed in this paper to the HeartSteps data in Section 5.

1.2. Notation and definition

We will consider two settings in the paper. In the first setting we consider a longitudinal study without treatment, and in the second one with a sequentially randomized treatment. The first setting will be used to explain bias incurred by the inclusion of endogenous covariates in random effects models, as this issue also occurs without treatment and is easier to explain there. The second setting involves time-varying treatment that is sequentially randomized; thus it's relevant to data from MRTs. We will see that randomized treatment assignment in MRT does not necessarily alleviate the biases resulting from the inclusion of endogenous time-varying covariates in LMMs. We will consider assumptions that allow valid estimation under this second setting. The setting under consideration will be clear from the context.

For the first setting without treatment, we denote data for individual i by $X_{i1}, Y_{i2}, X_{i2}, Y_{i3}, \ldots, X_{iT_i}, Y_{iT_i+1}$, where T_i denotes the total number of observations for individual i. X_{it} is a vector of covariates prior to the t-th time point and Y_{it+1} is the outcome subsequent to the t-th time point. Note that the time index for the outcome Y is augmented by 1 to make it consistent with the second setting. We use overbar to denote history; for example, $\bar{X}_{it} = (X_{i1}, X_{i2}, \ldots, X_{it})$. The individual's history information up to the t-th time is denoted by $H_{it} = (X_{i1}, Y_{i2}, \ldots, X_{it-1}, Y_{it}, X_{it}) = (\bar{Y}_{it}, \bar{X}_{it})$.

For the second setting with treatment, the data for individual i is $X_{i1}, A_{i1}, Y_{i2}, X_{i2}, A_{i2}, Y_{i3}, \ldots, X_{iT_i}, A_{iT_i}, Y_{iT_i+1}$, where X_{it} is the covariate vector prior to the t-th time, A_{it} is the randomized treatment at the t-th time, and Y_{it+1} is the proximal outcome subsequent to the t-th time. To maintain expositional clarity, throughout we assume there are only two types of treatment and $A_{it} \in \{0,1\}$. The history is defined as $H_{it} = (X_{i1}, A_{i1}, Y_{i2}, \ldots, X_{it-1}, A_{it-1}, Y_{it}, X_{it}) = (\bar{Y}_{it}, \bar{X}_{it}, \bar{A}_{it-1})$. We define $X_{i0} = \emptyset$, $A_{i0} = \emptyset$, and $Y_{i1} = \emptyset$.

In both settings, we use b_i to denote the random effect of individual i.

We use $\perp$ to denote statistical independence; for example, $A \perp B \mid C$ means that A is independent of B conditional on C. In the first setting, a covariate process X_{it} is called *exogenous* (with respect to the outcome process Y_{it}) if $X_{it} \perp \bar{Y}_{it} \mid \bar{X}_{it-1}$; otherwise, X_{it} is *endogenous*. In the second setting, X_{it} is called *exogenous* if $X_{it} \perp (\bar{Y}_{it}, \bar{A}_{it-1}) \mid \bar{X}_{it-1}$; otherwise, X_{it} is *endogenous*. In a longitudinal study, examples of exogenous covariates include baseline variables (age, gender, etc.), functions of time, and time-varying variables that are not impacted by prior treatment or prior outcome, such as weather.

2. Issue of linear mixed models with endogenous covariates

In this section, we start by considering the situation where no treatment is involved, as endogenous covariates give rise to issues even without considering causal inference. We give a brief review

of standard LMM in Section 2.1, and explain the issue of endogenous covariates in Section 2.2. In Section 2.3, we briefly review causal inference literature on a related topic, time-varying confounding, which is a more restrictive definition than endogeneity. In Section 2.4, we discuss connections to the econometric literature. We comment on why the methods reviewed in Sections 2.3 and 2.4 do not directly solve the issue of LMM with endogenous covariates in MRTs.

2.1. Brief overview of standard LMM with exogenous covariates

A standard linear mixed model (LMM) (Laird and Ware, 1982) assumes a relationship between the covariate X_{it} and the outcome Y_{it+1} such as the following:

$$Y_{it+1} = X_{it}^T\beta + Z_{it}^T b_i + \epsilon_{it+1}. \tag{1}$$

Here, $b_i \sim N(0,G)$ denotes the vector of person-specific random effects, $Z_{it} \subset X_{it}$ and $\epsilon_{it+1} \sim N(0,\sigma_\epsilon^2)$ is a random noise. It is typically assumed that ϵ_{it+1}'s are independent of each other and of b_i, and we will adopt this assumption throughout this paper. This model specifies the conditional distribution of Y_{it+1} given X_{it} and b_i; in particular, this is a Gaussian distribution with mean:

$$E(Y_{it+1} \mid X_{it}, b_i) = X_{it}^T\beta + Z_{it}^T b_i. \tag{2}$$

Furthermore, use of the standard LMM assumes, though not always explicitly, that all covariates are fixed, or at least exogenous and independent of b_i. Thus, the marginal mean of Y_{it} is

$$E(Y_{it+1} \mid X_{it}) = X_{it}^T\beta, \tag{3}$$

because $E(b_i \mid X_{it}) = 0$. Thus, when the covariates are exogenous and independent of b_i, β has both a conditional interpretation and a marginal interpretation[1]. This dual interpretation provides the opportunity to estimate β with alternative approaches such as with generalized estimating equations (GEE) (Zeger and Liang, 1986), depending on the desired robustness of the estimator of β to deviations from the LMM assumptions.

Assuming the covariates are indeed exogenous and independent of b_i, the maximum likelihood score equation for β is:

$$\frac{1}{n}\sum_{i=1}^{n} X_i V_i^{-1}(Y_i - X_i^T\beta) = 0, \tag{4}$$

where $X_i = (X_{i1},\ldots,X_{iT_i})$, $Z_i = (Z_{i1},\ldots,Z_{iT_i})$ and $Y_i = (Y_{i2},\ldots,Y_{iT_i+1})^T$, $V_i = Z_i^T G Z_i + R_i$ is a $T_i \times T_i$ covariance matrix, and R_i is a $T_i \times T_i$ diagonal matrix with all diagonal entries equal to σ_ϵ^2.

2.2. Issue with endogenous covariates: marginal interpretation is no longer valid

Any LMM solves the same estimating equation as a GEE with a corresponding non-independence working correlation structure (e.g., an LMM with a random intercept solves the same estimating equation as a GEE with compound symmetric working correlation structure). In fact, (4) is the

[1]In this paper, we use the term "conditional (model/interpretation)" to denote a model that is conditional on the random effect, and we use "marginal (model/interpretation)" to denote a model where the random effect is marginalized over. This is consistent with the terminology in Zeger and Liang (1992) and Heagerty and Zeger (2000).

estimating equation for GEE with marginal mean model (3) and working correlation matrix V_i. In the GEE literature, estimation bias due to the inclusion of endogenous covariates has been discussed repeatedly. We first review this briefly.

Pepe and Anderson (1994) first pointed out that when using GEE to estimate parameters in $E(Y_{it+1} \mid X_{it})$, a sufficient condition for estimation consistency is either

$$E(Y_{it+1} \mid X_{it}) = E(Y_{it+1} \mid X_{i1},\ldots,X_{iT}) \tag{5}$$

or the use of a working independence correlation structure. When (5) is violated and a correlation structure other than working independence is used, they provided simulation results to show that bias could occur. Diggle et al. (2002, Chapter 12) reiterated this point, and referred to (5) as "full covariate conditional mean (FCCM)" assumption. Schildcrout and Heagerty (2005) analyzed the bias-efficiency trade-off associated with working correlation choices of GEE for longitudinal binary data, when FCCM is violated due to exogenous covariates being time-varying, through simulation studies. This potential bias from the violation of FCCM have also been warned about by Pan, Louis and Connett (2000) in the context of linear regression via analytic calculations. Tchetgen et al. (2012) showed, in the context of marginal structural models (Robins, 1998), that when GEE is combined with inverse probability weighting for handling dropout, parameter estimation is generally biased in the presence of endogenous covariates unless either a condition similar to (5) holds or a working independence correlation structure is used.

When there are endogenous covariates, the FCCM assumption (5) is unlikely to hold because Y_{it+1} may impact future X_{is} for $s \geq t+1$. In this case, Pepe and Anderson (1994) suggested the use of working independence GEE to guarantee consistent estimation of parameters in $E(Y_{it+1} \mid X_{it})$. Because of the close tie between the estimating equations of LMM and GEE, Pepe and Anderson's point about GEE implies that estimators fitted using the standard LMM could be inconsistent when there are endogenous covariates. Indeed, if one intends to estimate parameters in the marginal mean $E(Y_{it+1} \mid X_{it})$, then using LMM *as an estimation procedure* can result in inconsistent estimators because of the biased estimating equations. However, in our opinion, this is not the fundamental issue of LMM under endogeneity, but rather a technical consequence.

More fundamentally, when there are endogenous covariates, LMM (1) *as a model* can imply a marginal mean relationship different from (3). X_{it} being endogenous means it may depend on previous outcomes, which in turn implies dependence on the random effect b_i. Thus, $E(b_i \mid X_{it})$ is usually nonzero and the conditional model (2) may no longer imply the marginal model (3). The marginal model implied by (2) becomes, instead,

$$E(Y_{it+1} \mid X_{it}) = X_{it}^T\beta + Z_{it}^T E(b_i \mid X_{it}). \tag{6}$$

As a concrete example, consider the case where each individual is observed for 2 time points ($T_i = 2$), and the covariate at the second time point is the lag-1 outcome: $X_{i2} = Y_{i2}$. Suppose the variables are generated from the following LMM with a random intercept: $b_i \sim N(0,\sigma_u^2)$, $X_{i1} \sim N(0,\sigma_{X_1}^2)$ independently of b_i, $Y_{i2} \mid X_{i1},b_i \sim N(\beta_0+\beta_1 X_{i1}+b_i,\sigma_\epsilon^2)$, $X_{i2} = Y_{i2}$, and $Y_{i3} \mid X_{i1},Y_{i2},X_{i2},b_i \sim N(\beta_0+\beta_1 X_{i2}+b_i,\sigma_\epsilon^2)$. This implies a parsimonious conditional relationship: $E(Y_{it+1} \mid X_{it},b_i) = \beta_0+\beta_1 X_{it}+b_i$, but the induced marginal relationship is rather complex:

$$\begin{aligned} E(Y_{i2} \mid X_{i1}) &= \beta_0+\beta_1 X_{i1}, \\ E(Y_{i3} \mid X_{i2}) &= (1-\rho\zeta-\rho)\beta_0+\{(1-\rho\zeta)\beta_1+\rho\}X_{i2}, \end{aligned}$$

with $\rho = \sigma_u^2/(\sigma_u^2+\sigma_\epsilon^2)$ and $\zeta = \beta_1\sigma_{X_1}^2/(\beta_1\sigma_{X_1}^2+\sigma_u^2+\sigma_\epsilon^2)$.

Therefore, when building LMM with endogenous covariates, one needs to be aware that the modeling assumption is on the conditional relationship $E(Y_{it+1}|X_{it},b_i)$, not the marginal relationship $E(Y_{it+1}|X_{it})$. Although it is attractive to treat β in (1) with not only a conditional interpretation but also a marginal interpretation, which is true with exogenous covariates, the latter interpretation can be invalid with endogenous covariates. In addition to this model interpretation issue, endogenous covariates also give rise to additional concerns in model fitting, which will be discussed in Section 3.

As a side note, for generalized linear mixed models, it is well known that even when all covariates are exogenous, the conditional parameter and the marginal parameter are different due to the nonlinear link function, and there has been work in the literature on connecting the two interpretations (Zeger, Liang and Albert, 1988; Heagerty, 1999; Wang and Louis, 2004). For LMMs, the discrepancy in the two interpretations only occurs when there are endogenous covariates.

2.3. Connection to time-varying confounding in causal inference literature

In the setting with treatment, a related issue, often called "time-varying confounding" or "time-dependent confounding", has been well studied in the causal inference literature. A time-varying covariate is a time-varying confounder if it is affected by previous treatment (hence is endogenous) and it affects future treatment assignment (Daniel et al., 2013; Hernán and Robins, 2019). Time-varying confounders are usually intermediate variables (that lie in the causal pathway between the treatment and the outcome), and this gives rise to inferential challenges for conventional regression-based methods due to the following dilemma: confounders should be adjusted for in the analysis, but intermediate variables should not (Diggle et al., 2002).

Causal inference methods have been developed to estimate treatment effects in the presence of time-varying confounding. These methods include g-computation (Robins, 1986), structural nested models (Robins, 1994, 1997), inverse probability weighting in marginal structural models (Robins, 1998, 2000), history-restricted marginal structural models (Neugebauer et al., 2007), sequential conditional mean models (Vansteelandt, 2007; Keogh et al., 2017), and weighted and centered least-squares for MRTs (Boruvka et al., 2018). These methods cover a variety of estimands that characterize the effect of a time-varying treatment from various aspects, but all the treatment effects are marginal in the sense that no random effect is considered.

Estimators of conditional-on-the-random-effect versions of the above estimands will be potentially biased as discussed in Section 2.2. Furthermore, the issue with bias persists even when A_{it} is not confounded by observed or unobserved variables (e.g., when the randomization probability is constant). Take, for example, the sequential conditional mean models in Vansteelandt (2007), which considers the marginal expected mean $E(Y_{it+1} \mid \bar{A}_{it},\bar{X}_{it})$. When random effect is incorporated, the model becomes the conditional expected mean $E(Y_{it+1} \mid \bar{A}_{it},\bar{X}_{it},b_i)$. When X_{it} is endogenous, even if X_{it} does not confound A_{it}, the same argument in Section 2.2 applies, and the parameter in the conditional model $E(Y_{it+1} \mid \bar{A}_{it},\bar{X}_{it},b_i)$ generally does not have the marginal interpretation. This means the methods for estimating marginal treatment effect cannot be used to estimate parameters in the conditional model, let alone used to predict the random effects in the conditional model.

2.4. Connection to level-2 endogeneity in econometric literature

Violation of the assumption that the random effect being independent of the covariates, $b_i \perp X_{it}$, is sometimes called "level-2 endogeneity" in the econometric literature (Wooldridge, 2002;

Grilli and Rampichini, 2011). It is well known that level-2 endogeneity can lead to biased parameter estimates (Ebbes, Böckenholt and Wedel, 2004); in particular, Kim and Frees (2007) gave a display similar to (6), and warned about the bias that could occur when one uses an estimator intended for the marginal parameter (such as the ordinary least-squares) to estimate the conditional parameter—this is the counterpart of our discussion in Section 2.2, that using LMM to estimate the marginal parameter will incur bias with endogenous covariates.

Various estimators have been proposed in the econometric literature for the conditional parameter under level-2 endogeneity, many of which are based on explicitly modeling the conditional distribution of the random effects given the endogenous covariates (Mundlak, 1978), centering the time-varying covariate and the time-varying outcome by their average over time (Hausman and Taylor, 1981; Arellano and Bover, 1995; Neuhaus and McCulloch, 2006; Kim and Frees, 2006; Hanchane and Mostafa, 2012), constructing internal instrumental variables (Amemiya and MaCurdy, 1986; Arellano and Bond, 1991; Semykina and Wooldridge, 2010), or using semiparametric efficiency theory by not specifying the distribution of the random effects (Liu and Xiang, 2014; Garcia and Ma, 2016).

In those works, it is usually assumed that the error term ϵ_{it} is independent of the history of the time-varying covariate, $\bar{X}_{iT_i}$; thus these methods are not directly applicable to the MRT setting where future covariates can depend on previous outcomes (hence previous error terms). In addition, many of these methods focus on estimating the conditional parameter while treating the random effect as a nuisance parameter. We argue that in MRTs, prediction of the random effects are of equal importance to estimation of the conditional parameter; otherwise, one could have used the causal inference methods mentioned in Section 2.3 to estimate the marginal treatment effect. It is an open question whether the ideas behind the above methods can be adapted for LMM-based inference in MRTs.

3. A conditional independence assumption

In an MRT, the observed history up to time t is defined as $H_{it} = (X_{i1}, A_{i1}, Y_{i2}, \ldots, X_{it-1}, A_{it-1}, Y_{it}, X_{it})$. We consider the following LMM:

$$Y_{it+1} = f_0(H_{it})^T \beta_0 + A_{it} f_1(H_{it})^T \beta_1 + g_0(H_{it})^T b_{0i} + A_{it} g_1(H_{it})^T b_{1i} + \epsilon_{it+1} \tag{7}$$

for $t = 1, \ldots, T$, where $f_0(H_{it}), f_1(H_{it}), g_0(H_{it}), g_1(H_{it})$ are known functions of H_{it}. For example, if we believe that the outcome depends linearly on time, current covariate and previous outcome, that the treatment also interacts with these three variables, and that the outcome has no residual association with other information in H_{it}, we may set each of $f_0(H_{it}), f_1(H_{it}), g_0(H_{it}), g_1(H_{it})$ to be $(1, t, X_{it}, Y_{it})$. Recall that for simplicity we consider only binary treatment. In this section, we provide an additional assumption that, if true, ensures valid treatment inference and person-specific predictions via standard software even when there are endogenous covariates.

We make the standard LMM assumptions. The random effects (b_{0i}^T, b_{1i}^T) are assumed to marginally follow a multivariate Gaussian distribution with mean 0 and variance-covariance matrix G. A_{it} is assumed to be randomized with randomization probability depending only on H_{it}, not b_{i0} or b_{i1}; this is ensured by the MRT design. The random noise ϵ_{it+1} is assumed to be independent of $(H_{it}, A_{it}, b_{0i}, b_{1i})$ and follows $N(0, \sigma_\epsilon^2)$. $f_0(H_{it})$ and $f_1(H_{it})$ can include possibly endogenous covariates X_{it} and lagged outcomes such as Y_{it}.

Equation (7) along with the above assumptions completely specifies the conditional distribution of the outcome Y_{it+1} conditional on $b_{0i}, b_{1i}, H_{it}, A_{it}$. It implies the following treatment effect that is

conditional on the random effects

$$E(Y_{it+1} \mid b_{0i}, b_{1i}, H_{it}, A_{it}=1) - E(Y_{it+1} \mid b_{0i}, b_{1i}, H_{it}, A_{it}=0) = f_1(H_{it})^T \beta_1 + g_1(H_{it})^T b_{1i}. \quad (8)$$

Furthermore due to endogeneity, it is likely that

$$E(Y_{it+1} \mid H_{it}, A_{it}=1) - E(Y_{it+1} \mid H_{it}, A_{it}=0) \neq f_1(H_{it})^T \beta_1. \quad (9)$$

In other words, the treatment effect (8) implied by model (7) is interpreted as *conditional-on-the-random-effect*; $\beta = (\beta_0^T, \beta_1^T)^T$ does not have a *marginal* interpretation. A similar point for when there is no treatment has been extensively discussed in Section 2.

The above model provides the distribution of Y_{it+1} conditional on $(b_{0i}, b_{1i}, H_{it}, A_{it})$ as opposed to conditional on $(b_{0i}, b_{1i}, X_{it}, A_{it})$. Thus β_1 in (8) has a causal interpretation even when the randomization probability for A_{it} depends on H_{it} in an MRT. Likelihood-based inference and model fitting through standard LMM software can be conducted as described below. Note that since $f_0(H_{it})$ and $f_1(H_{it})$ can include lagged outcomes, the dependence between outcomes is explicitly modeled in (7). The purpose of introducing random effects here is mainly to model the between-person heterogeneity.

To estimate the conditional-on-the-random-effect β, we make an additional conditional independence assumption. The *conditional independence assumption* is

$$X_{it} \perp (b_{0i}, b_{1i}) \mid H_{it-1}, A_{it-1}, Y_{it}. \quad (10)$$

This does allow X_{it} to be endogenous, but the endogenous covariate X_{it} can only depend on the random effects through the variables observed prior to X_{it}: H_{it-1}, A_{it-1}, and Y_{it}. If the only endogenous covariates are functions of prior treatments and prior outcomes, then assumption (10) automatically holds. In general, assumption (10) needs to be verified from the domain science perspective. We discuss this assumption in the context of HeartSteps in Section 5.

Assumption (10) allows us to decompose the likelihood. This likelihood decomposition will provide a justification for the use of estimators from standard LMM software. Denote by X_i, A_i and Y_i the vectors of observations for individual i, and X, A and Y the collection of observations for all individuals. Denote by $b_i = (b_{0i}, b_{1i})$. Suppose G, the covariance matrix of the random effects, is parametrized by θ. The joint likelihood of the observed data, $\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon \mid X, A, Y)$, can be written as

$$\begin{aligned}
\prod_i p(X_i, A_i, Y_i \mid \alpha, \beta, \theta, \sigma_\epsilon) &= \prod_i \int p(X_i, A_i, Y_i \mid b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i) \\
&= \prod_i \Big\{ \int \prod_t p(X_{it} \mid H_{it-1}, A_{it-1}, Y_{it}, b_i) p(A_{it} \mid H_{it}, b_i) \\
&\qquad \times p(Y_{it+1} \mid H_{it}, A_{it}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i) \Big\}. \quad (11)
\end{aligned}$$

By the conditional independence assumption (10) and given that A_{it} is randomized conditional on H_{it}, the joint likelihood in (11) becomes

$$\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon \mid X, A, Y) = \Big\{ \prod_i \prod_t p(X_{it} \mid H_{it-1}, A_{it-1}, Y_{it}) p(A_{it} \mid H_{it}) \Big\} \mathcal{L}_1(\alpha, \beta, \theta, \sigma_\epsilon \mid X, A, Y), \quad (12)$$

where

$$\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon \mid X,A,Y) = \prod_i \Big\{ \int \prod_t p(Y_{it+1} \mid H_{it},A_{it},b_i;\alpha,\beta,\theta,\sigma_\epsilon) dF(b_i) \Big\}. \tag{13}$$

Because the first factor on the right hand side of (12) does not involve $(\alpha,\beta,\theta,\sigma_\epsilon)$, any inference for $(\alpha,\beta,\theta,\sigma_\epsilon)$ that is based on the joint likelihood $\mathcal{L}(\alpha,\beta,\theta,\sigma_\epsilon \mid X,A,Y)$ can be equivalently based on the partial likelihood $\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon \mid X,A,Y)$. Observe that $\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon \mid X,A,Y)$ is actually the likelihood function for a standard LMM where X_{it} and A_{it} are treated as fixed covariates. Thus, the maximum likelihood estimators that are obtained through standard LMM software are valid maximum likelihood estimators for the joint likelihood $\mathcal{L}(\alpha,\beta,\theta,\sigma_\epsilon \mid X,A,Y)$ under the conditional independence assumption, and (4) with X redefined to include the treatment indicator is a likelihood score equation for β in the conditional-on-the-random-effect model. Note that even though the form of (4) appears to indicate estimation of a regression coefficient in a marginal model, this is a false impression in the case of endogenous covariates. Furthermore, recall that restricted maximum likelihood (REML) estimation can be viewed as maximum *a posteriori* in a Bayesian hierarchical model (Laird and Ware, 1982). This latter interpretation continues to hold for the REML estimators obtained through standard LMM software when there are endogenous covariates. In addition, it can be shown that the empirical Bayes predictor of the random effects $\hat{b}_i$ obtained through standard LMM software is valid empirical Bayes predictor for model (7) with endogenous covariates. We include proofs of these claims in the Appendix.

The conditional independence assumption (10) is similar to an assumption used by Sitlani et al. (2012). Sitlani et al. (2012) aimed to use an LMM to assess causal effects in the context of noncompliance in surgical trials. They assumed conditional independence between the treatment assignment and the random effect given the observed history. This assumption allowed them to decompose the likelihood as is done above and thus use standard LMM estimators.

It is worth noting, as pointed out by a reviewer, that if the analyst poses a model as (7) but without the $A_{it}g_1(H_{it})^T b_{1i}$ term (i.e., the random effect in the model does not interact with A_{it}), then (9) becomes an equality. In other words, in this case β_1 recovers its marginal interpretation

$$E(Y_{it+1} \mid H_{it}, A_{it}=1) - E(Y_{it+1} \mid H_{it}, A_{it}=0) = f_1(H_{it})^T \beta_1,$$

and furthermore it can be interpreted marginally over $H_{it} \setminus f_1(H_{it})$:

$$E\big\{ E(Y_{it+1} \mid H_{it}, A_{it}=1) - E(Y_{it+1} \mid H_{it}, A_{it}=0) \mid f_1(H_{it}) \big\} = f_1(H_{it})^T \beta_1. \tag{14}$$

Note that β_0 still has only the conditional-on-the-random-effect interpretation. In absence of b_{1i}, the conditional independence assumption (10) becomes

$$X_{it} \perp b_{0i} \mid H_{it-1}, A_{it-1}, Y_{it};$$

this assumption justifies the use of over-the-counter LMM software's via the likelihood factorization (12).

4. Simulation

In the simulation, we considered three generative models (GMs), in all of which the covariate is endogenous. In the first two GMs, the endogenous covariate X_{it} equals the previous outcome Y_{it}

plus some random noise, so the conditional independence assumption (10) is valid. In GM 3, the endogenous covariate depends directly on b_i, so the assumption (10) is violated. Details of the generative models are described in the following.

In GM 1, we considered a simple case with only a random intercept and a random slope for A_{it}, so that $Z_{it}^{(0)} = Z_{it}^{(2)} = 1$ in model (7). The outcome is generated as $Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + \beta_1 X_{it} + b_{i2}) + \epsilon_{it+1}$. The random effects $b_{i0} \sim N(0, \sigma_{b0}^2)$ and $b_{i2} \sim N(0, \sigma_{b2}^2)$ are independent of each other. We generated the covariate to be $X_{i1} \sim N(0,1)$, $X_{it} = Y_{it} + N(0,1)$ for $t \geq 2$. The randomization probability p_t is constant $1/2$. The exogenous noise $\epsilon_{it+1} \sim N(0, \sigma_\epsilon^2)$.

In GM 2, we considered the case where $Z_{it}^{(0)} = Z_{it}^{(2)} = (1, X_{it})$, and the randomization probability is time-varying. The outcome is generated as $Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + b_{i1} X_{it} + A_{it}(\beta_0 + \beta_1 X_{it} + b_{i2} + b_{i3} X_{it}) + \epsilon_{it+1}$. The random effects $b_{ij} \sim N(0, \sigma_{bj}^2)$, $0 \leq j \leq 3$, are independent of each other. We generated the covariate to be $X_{i1} \sim N(0,1)$, $X_{it} = Y_{it} + N(0,1)$ for $t \geq 2$. The randomization probability depends on X_{it}: $p_t = 0.7 \cdot \mathbb{1}(X_{it} > -1.27) + 0.3 \cdot \mathbb{1}(X_{it} \leq -1.27)$. Here $\mathbb{1}(\cdot)$ represents the indicator function, and the cutoff -1.27 was chosen so that p_t equals 0.7 or 0.3 each for about half of the time. The exogenous noise $\epsilon_{it+1} \sim N(0, \sigma_\epsilon^2)$.

GM 3 is the same as GM 1, except that the covariate X_{it} depends directly on b_i: $X_{i1} \sim N(b_{i0}, 1)$, $X_{it} = Y_{it} + N(b_{i0}, 1)$ for $t \geq 2$.

We chose the parameter values as follows: $\alpha_0 = -2$, $\alpha_1 = -0.3$, $\beta_0 = 1$, $\beta_1 = 0.3$, $\sigma_{b0}^2 = 4$, $\sigma_{b1}^2 = 1/4$, $\sigma_{b2}^2 = 1$, $\sigma_{b3}^2 = 1/4$, $\sigma_\epsilon^2 = 1$.

For each of the three GMs, we simulated for sample size $n = 30, 100, 200$ and the number of observations per individual $T_i = T = 10, 30$. Each setting was replicated 1,000 times. The estimation was done using the R package lmer (Bates et al., 2015) for standard LMM, and 95% confidence interval was computed based on the t distribution with degrees of freedom obtained by Satterthwaite approximation (Satterthwaite, 1941), which is implemented in the R package lmerTest (Kuznetsova, Brockhoff and Christensen, 2017). Bias, standard deviation (sd) and coverage probability (cp) of 95% nominal confidence interval for the estimated β_0 and β_1 are presented in Table 1. As expected, the estimators are consistent for GM 1 and GM 2, and they are inconsistent for GM 3 because of the violation of the conditional independence assumption (10). For GM 1 and GM 2, the confidence interval coverage probability can be slightly lower than the nominal level for some of the parameters for small n or small T, but it gets back to the nominal level as the sample size or total number of time points gets larger. Additional simulation results for more choices of n and T, the performance of estimated α_0, α_1, and variance components $\sigma_{bj}^2, 0 \leq j \leq 3$ and σ_ϵ^2 are in the Appendix, and the conclusion is similar to the results for the β's as shown here.

5. Illustrative data analysis of HeartSteps

5.1. Data and model assumptions

As described in Section 1.1, HeartSteps (Klasnja et al., 2018) is a 6-week micro-randomized trial of an mHealth intervention to encourage activity among sedentary adults. The following analysis focuses on the time-varying treatment consisting of contextually-tailored activity suggestions.

Prior to the randomization at each time point, software on the smartphone determined whether an individual is *available* for treatment at the time. If the activity recognition on the phone determined that an individual was operating a vehicle, the individual was considered unavailable for safety reasons. If an individual had just finished an activity bout in the prior 90 seconds, they were considered unavailable for treatment in order to minimize user burden and aggravation. Lastly,

			β_0			β_1		
GM	T	n	bias	sd	cp	bias	sd	cp
1	10	30	-0.001	0.249	0.943	0.002	0.091	0.897
		100	-0.003	0.135	0.941	-0.001	0.049	0.898
		200	-0.001	0.096	0.926	-0.001	0.034	0.899
1	30	30	-0.002	0.206	0.946	0.001	0.053	0.913
		100	-0.005	0.112	0.949	-0.001	0.028	0.935
		200	0.000	0.081	0.944	-0.001	0.022	0.902
2	10	30	-0.010	0.269	0.939	-0.004	0.105	0.903
		100	0.009	0.145	0.933	-0.001	0.056	0.915
		200	-0.008	0.105	0.931	-0.002	0.038	0.934
2	30	30	-0.006	0.216	0.943	-0.001	0.070	0.939
		100	0.006	0.115	0.947	-0.001	0.039	0.948
		200	-0.004	0.084	0.935	-0.000	0.027	0.940
3	10	30	-0.048	0.245	0.949	-0.043	0.075	0.725
		100	-0.060	0.134	0.927	-0.047	0.041	0.548
		200	-0.052	0.095	0.907	-0.046	0.029	0.355
3	30	30	-0.023	0.207	0.946	-0.017	0.041	0.847
		100	-0.028	0.112	0.942	-0.019	0.022	0.762
		200	-0.024	0.079	0.941	-0.019	0.015	0.628

TABLE 1

Bias, standard deviation (sd) and coverage probability (cp) of 95% nominal confidence interval for estimated β_0 and β_1 in the simulation study. n denotes sample size; T denotes total number of observations for each individual; GM denotes generative model. The result is based on 1,000 replicates for each setting.

because the software on the server and smartphone required an internet connection to send a suggestion, if the smartphone did not have wireless connectivity the individual was deemed unavailable. At each of the five points each day for each individual, availability was assessed, the context was recorded, and if the individual was available then HeartSteps randomized to deliver an activity suggestion to the individual with probability 3/5. The sample for this analysis consisted of 7,540 time points from 37 individuals. The individuals were available for 6,061 (80.4%) time points, unavailable due to no internet connection for 602 (8.0%) time points, unavailable due to being detected as in transit for 841 (11.1%) time points, and unavailable due to being detected to have just finished an activity bout in the prior 90 seconds for 36 (0.5%) time points.

Let $A_{it} = 1$ if an activity suggestion is delivered at time t for individual i and equal to 0 otherwise. The proximal outcome Y_{it+1} is the (log-transformed) 30-minute step count following time point t. We used three covariates in the model:

- $X_{it,1}$: day in the study for the time point t, coded as 0,1,...,41.
- $X_{it,2}$: whether the individual was at home or work at time point t; $X_{it,2} = 1$ if at home or work, 0 if at some other location.
- $X_{it,3}$: (log-transformed) 30-minute step count preceding time point t.

We specify model (7) in the HeartSteps context as follows: $f_0(H_{it}) = (X_{it,1}, X_{it,2}, X_{it,3})$; $f_1(H_{it}) = (X_{it,1},\ X_{it,2})$; the model contains a random intercept, $g_0(H_{it}) = 1$, and a random slope for A_{it}, $g_1(H_{it}) = 1$. We denote the availability status of individual i at time t by I_{it} ($I_{it} = 1$ if available; 0 otherwise). In the model, we multiply A_{it} with I_{it} to operationalize the notion that the treatment may only be delivered when the individual is available. Because the relationship between Y_{it+1} and the $f_0(H_{it})$ can depend on the availability status, we included an interaction between I_{it} and

$f_0(H_{it})$. Thus, the LMM is given by

$$\begin{aligned} Y_{it+1} = {} & \alpha_0 + \alpha_1 X_{it,1} + \alpha_2 X_{it,2} + \alpha_3 X_{it,3} + I_{it}(\tilde{\alpha}_0 + \tilde{\alpha}_1 X_{it,1} + \tilde{\alpha}_2 X_{it,2} + \tilde{\alpha}_3 X_{it,3}) + b_{0i} \\ & + A_{it} I_{it}(\beta_0 + \beta_1 X_{it,1} + \beta_2 X_{it,2} + b_{1i}) + \epsilon_{it+1} \end{aligned} \tag{15}$$

where $\epsilon_{it+1} \sim N(0, \sigma_\epsilon^2)$, and the random effects $(b_{0i}, b_{1i}) \sim N(0, G)$ with G being a 2×2 variance-covariance matrix. b_{0i} accounts for the between-individual variation in the 30-minute step count under no treatment, and b_{1i} accounts for the between-individual variation in the treatment effect on the 30-minute step count.

In model (15), $X_{it,2}$, $X_{it,3}$ and I_{it} are possibly endogenous. Location, $X_{it,2}$, is most likely exogenous but might be endogenous because the number of steps an individual took following a prior time point, combined with the location s/he was at then, might be predictive of whether s/he would be at home/work or other places at the subsequent time point. Prior time t 30-minute step count, $X_{it,3}$, might be correlated with 30-minute step count after time $t-1$, Y_{it}, because an individual might walk less if s/he had already walked earlier in the day. For the availability status I_{it}, unavailability due to being in transit is likely exogenous but may be endogenous for a reason similar to that of location, $X_{it,2}$. Unavailability due to having just finished an activity bout may be endogenous for a reason similar to that of prior time t 30-minute step count, $X_{it,3}$. We argue that the conditional independence assumption (10) is plausible for all three variables. For location, $X_{it,2}$, because the enrollment criterion required each individual to either have a full-time daytime job or be a student, the time-varying location of such individuals with regular schedule is unlikely to depend on some unmeasured baseline factors (i.e., the random effects) that impact step count. For prior time t 30-minute step count, $X_{it,3}$, the impact of random effects should be largely explainable through earlier outcomes and covariates, as those are also step counts but just for other time windows. For I_{it}, most of the unavailability (1443/1479) instances are due to being in transit or loss of internet connection; the conditional independence is likely to approximately hold for I_{it} for a similar reason to that of $X_{it,2}$.

5.2. Results

We fitted model (15) using the R package lmer (Bates et al., 2015) for standard LMM, because standard LMM yields valid estimators under the conditional independence assumption (10).

The first three columns in Table 2 show the estimated fixed effects with 95% confidence interval and the estimated variance components. The estimated variance for b_{1i} is extremely small and the estimated correlation between b_{0i} and b_{1i} is 1.000, suggesting that we might not have enough data to fit two separate random effects so the fitting collapsed onto a linear combination of the two. We conducted the likelihood ratio test for nonzero variance of b_{1i}, and the p-value was 0.72. Note that likelihood ratio tests for nonzero variance components can be conservative because the null value $(\mathrm{Var}(b_{1i}) = 0)$ is on the boundary of the parameter space (Self and Liang, 1987; Stram and Lee, 1994; Crainiceanu and Ruppert, 2004), and we are just using this test and the critical value as a guideline. The result suggests that the potential heterogeneity in the treatment effect may not be large enough to be detected from the data. Model fit of (15) with b_{1i} removed is presented in the last two columns in Table 2.

The estimated treatment effects, which are conditional on the observed history and the unobserved random effects, are similar from both model fits in the point estimates as well as the confidence intervals. The data indicates that, for an individual, the treatment has a positive effect

at the beginning of the study ($\hat{\beta}_0 > 0$), and the effect decreased over time ($\hat{\beta}_1 < 0$). This is likely due to the individual's habituation to the activity suggestions, which is consistent with the exit interviews reported by Klasnja et al. (2018) in which individuals reported that "the suggestions became boring after 2–4 weeks". On the other hand, the data indicates no moderating influence of location (whether an individual was at home/work or some other place) on the treatment effect for an individual.

	Model with b_{1i}		Model without b_{1i}	
coefficient	estimate	95% CI	estimate	95% CI
α_0	1.990	(1.643, 2.338)	1.997	(1.646, 2.348)
α_1	-0.009	(-0.021, 0.002)	-0.009	(-0.021, 0.002)
α_2	0.851	(0.238, 1.465)	0.840	(0.226, 1.453)
α_3	0.539	(0.495, 0.583)	0.537	(0.493, 0.582)
$\tilde{\alpha}_0$	-0.177	(-0.586, 0.232)	-0.182	(-0.591, 0.228)
$\tilde{\alpha}_1$	0.008	(-0.006, 0.023)	0.008	(-0.007, 0.023)
$\tilde{\alpha}_2$	-0.871	(-1.522, -0.221)	-0.863	(-1.514, -0.212)
$\tilde{\alpha}_3$	-0.156	(-0.206, -0.107)	-0.154	(-0.204, -0.104)
β_0	0.415	(0.105, 0.724)	0.410	(0.100, 0.719)
β_1	-0.017	(-0.028, -0.005)	-0.017	(-0.028, -0.005)
β_2	0.122	(-0.156, 0.400)	0.130	(-0.148, 0.408)
$\text{Var}(b_{0i})$	0.160		0.182	
$\text{Var}(b_{1i})$	0.003		-	
$\text{Corr}(b_{0i}, b_{1i})$	1.000		-	
$\text{Var}(\epsilon_{it+1})$	7.138		7.139	

TABLE 2

Estimated coefficients and 95% confidence interval for model (15) of HeartSteps data. Estimators are obtained using R package lmer*, and the 95% confidence interval are based on t distribution with Satterthwaite approximation implemented in R package* lmerTest*.*

As a point of contrast, we also analyzed the data using the weighted and centered least-squares (WCLS) estimator in Boruvka et al. (2018) for a related but different model. We used WCLS to estimate $\psi = (\psi_0, \psi_1, \psi_2)$ in the following model:

$$E\big\{E(Y_{it+1} \mid H_{it}, A_{it} = 1) - E(Y_{it+1} \mid H_{it}, A_{it} = 0) \mid X_{it,1}, X_{it,2}, I_{it} = 1\big\} = \psi_0 + \psi_1 X_{it,1} + \psi_2 X_{it,2}. \tag{16}$$

Boruvka et al. (2018) called (16) the causal excursion effect; ψ is marginal over both the random effects and $H_{it} \setminus \{X_{it,1}, X_{it,2}\}$, which is different from β in (15). We used $\gamma_0 + \gamma_1 X_{it,1} + \gamma_2 X_{it,2} + \gamma_3 X_{it,3}$ as the working model for $E(Y_{it+1} \mid H_{it}, A_{it} = 0, I_{it} = 0)$ in WCLS; this working model does not need to be correctly specified to guarantee the consistent of the estimator for ψ. The estimated ψ and the 95% confidence interval are listed in Table 3. Although β and ψ are different estimands with different interpretation, their estimated value and confidence interval are qualitatively similar. These results are consistent with the comments made in the last paragraph regarding the direction of how different variables moderate the treatment effect.

coefficient	estimate	95% CI
ψ_0	0.454	(0.156, 0.753)
ψ_1	-0.018	(-0.029, -0.006)
ψ_2	0.096	(-0.219, 0.410)

TABLE 3

Estimated coefficients and 95% confidence interval for model (16) using WCLS estimator in Boruvka et al. (2018).

6. Discussion

Linear mixed models (LMM) were originally developed for settings with fixed covariates, and it has been natural for researchers to think about the induced marginal model when building and interpreting the fixed effects in LMM. In this paper, we review related literature on the potential bias that would arise when including endogenous covariates into LMM. We argued that the fundamental issue in LMM with endogenous covariates is that the fixed effects, including the treatment effect, will only have a conditional-on-the-random-effect interpretation, and the marginal interpretation is no longer valid. In terms of estimation for LMM with endogenous covariates, we introduced a conditional independence assumption, and showed that under this assumption standard LMM software can still be used to obtain valid estimator of the fixed effects and the variance components, as well as valid prediction of the random effects. We used an LMM to model the effect of sequentially assigned treatment in HeartSteps MRT in which the covariates are likely endogenous, and we discussed the plausibility of the conditional independence assumption for these covariates.

The potential bias resulting from endogenous covariates in the without-treatment longitudinal setting has been known for decades since Pepe and Anderson (1994). However, it was quite surprising to us that in the MRT setting, this issue occurs even with randomized treatment with constant randomization probability (no confounding). The method in this paper utilizes the randomization to the extent that the treatment indicator A_{it} automatically satisfies a conditional independence assumption similar to (10). Furthermore, (7) is a mechanistic model for the outcome, which implies that how well the estimated β approximates the true treatment effect is contingent on how well the mechanistic model approximates the true data generating distribution. When the marginal treatment effect is of interest, there are many tools in causal inference that consistently estimate the effect with a possibly misspecified nuisance model (Robins, 1994, 2000; Hernán, Brumback and Robins, 2001; Brumback et al., 2003; Goetgeluk and Vansteelandt, 2008; Boruvka et al., 2018). It is an open question whether the randomization can be further leveraged in LMM to increase robustness to misspecified nuisance models.

The inclusion of endogenous covariates to an LMM implies that the fixed effects should only be interpreted as conditional on an individual. Thus, a future research question is to develop estimation methods for the parameters in the marginal mean model that are coherent with fixed effect parameters in an LMM where there are endogenous covariates. Related work in generalized linear mixed models but with exogenous covariates includes Heagerty (1999), Heagerty and Zeger (2000), and Larsen et al. (2000).

In a standard LMM with exogenous covariates, the empirical best linear unbiased predictor (eBLUP) equals the empirical Bayes estimator where a noninformative prior is imposed on the fixed effect and the variance components are estimated through REML (Lindley and Smith, 1972; Dempfle, 1977). In Section 3 we showed through partial likelihood argument that the empirical Bayes estimator of random effects from standard LMM is still a valid empirical Bayes estimator in the case of endogenous covariates. However, it is unknown whether it is still eBLUP absent further assumptions.

Along the same lines, in a standard LMM the restricted maximum likelihood (REML) estimator of the variance components can be viewed as the maximum *a posteriori* estimator in a Bayesian hierarchical model (Laird and Ware, 1982), and in Section 3 we showed that this latter interpretation is valid for the REML estimators obtained through standard LMM software when there are endogenous covariates. Another interpretation of the REML estimator in a standard LMM is the maximizer for the likelihood of linear combinations of the outcome that is orthogonal to the fixed

effects. It is unknown whether this interpretation continues to hold for the endogenous covariate case.

In the literature, there has been work on handling endogenous covariates in longitudinal data via jointly modeling of the covariate process and the outcome process, which could be alternative approaches to the method proposed in this paper for situations where the conditional independence assumption is questionable. Note that each of these alternative approaches require certain assumptions on the covariate process, and these assumptions themselves need to be verified in the context of each application. For example, Miglioretti and Heagerty (2004) modeled the covariate process, and assumed that $X_{it} \perp b_i \mid X_{i1}, X_{i2}, \ldots, X_{it-1}$. Roy et al. (2006) proposed to model the distribution of covariates given the history to infer the dependence of a Poisson process outcome on the endogenous covariates. Sitlani et al. (2012) proposed to use joint modeling for analyzing the effect of a surgical trial (where the time-varying treatment is a jump process) under noncompliance. Shardell and Ferrucci (2018) proposed to use a joint model approach, by assuming either that the distribution of X_{it} can be correctly modeled, or that the endogenous covariate is the lagged outcome.

Appendix A: Estimation and prediction through standard LMM software

In this Appendix, we provide a proof for the claims in Section 3 that maximum likelihood estimators, maximum *a posterior* estimators, and the empirical Bayes prediction of the random effects can be obtained through standard LMM software.

A.1. Estimation of fixed effects and variance components

This subsection focuses on estimation of the fixed effects α and β and the variance components θ and σ_ϵ^2 in model (7).

That the maximum likelihood estimator for the fixed effects and the variance component can be obtained through standard LMM software is immediate from the likelihood factorization (12).

The restricted maximum likelihood (REML) estimator of the variance components θ and σ_ϵ in a standard LMM can be obtained through Bayesian maximum *a posteriori* (MAP) estimation with a non-informative prior on the fixed effects α, β (Laird and Ware, 1982; Searle, Casella and McCulloch, 1992). For our case, the marginal likelihood for θ, σ_ϵ, where α and β are integrated over with respect to non-informative priors $p(\alpha)$ and $p(\beta)$, is

$$L(\theta,\sigma_\epsilon \mid X_i, A_i, Y_i, 1 \le i \le n) = \int p(\alpha)p(\beta)\prod_i p(X_i, A_i, Y_i \mid \alpha,\beta,\theta,\sigma_\epsilon)d\alpha d\beta,$$

which by (12) equals

$$\begin{aligned}
&\prod_i \Big\{ \prod_t p(X_{it} \mid H_{it-1}, A_{it-1}, Y_{it}) p(A_{it} \mid H_{it}) \Big\} \\
&\quad \times \int p(\alpha)p(\beta)\prod_i \Big\{ \int \prod_t p(Y_{it+1} \mid H_{it}, A_{it}, b_i; \alpha,\beta,\theta,\sigma_\epsilon) dF(b_i) \Big\} d\alpha d\beta \\
&\propto \int p(\alpha)p(\beta)\prod_i \Big\{ \int \prod_t p(Y_{it+1} \mid H_{it}, A_{it}, b_i; \alpha,\beta,\theta,\sigma_\epsilon) dF(b_i) \Big\} d\alpha d\beta. \qquad (17)
\end{aligned}$$

Expression (17) is the marginal likelihood for θ,σ_ϵ in a standard LMM; hence, the MAP estimator of the variance components can be obtained through standard LMM fitting procedure with the REML option.

A.2. Prediction of random effects

Prediction of random effects in a standard LMM is through best linear unbiased predictors (BLUPs, Henderson (1975)), which can be alternatively derived as empirical Bayes estimates using REML estimator of the variance components and fixed effects (Lindley and Smith, 1972; Dempfle, 1977).

Denote by $b=(b_1,\ldots,b_n)$, $X=(X_1,\ldots,X_n)$, $A=(A_1,\ldots,A_n)$, and $Y=(Y_1,\ldots,Y_n)$. In our proposed model, the posterior distribution of b is

$$p(b\mid X,A,Y;\theta,\sigma_\epsilon)=\frac{p(b,X,A,Y\mid\theta,\sigma_\epsilon)}{p(X,A,Y\mid\theta,\sigma_\epsilon)}. \tag{18}$$

We omit the notational dependence on θ,σ_ϵ hereafter. Let $p(\alpha)$ and $p(\beta)$ denote the prior distribution of α and β. The numerator of the right hand side of (18) equals

$$\begin{aligned}\int p(b,X,A,Y,\alpha,\beta)d\alpha d\beta &= \int p(\alpha)p(\beta)\prod_i p(b_i)\prod_t p(X_{it}\mid H_{it-1},A_{it-1},Y_{it},b_i,\alpha,\beta)\\ &\quad\times p(A_{it}\mid H_{it},b_i,\alpha,\beta)p(Y_{it+1}\mid H_{it},A_{it},b_i;\alpha,\beta)d\alpha d\beta\\ &=\Big\{\prod_i\prod_t p(X_{it}\mid H_{it-1},A_{it-1},Y_{it})p(A_{it}\mid H_{it})\Big\}\\ &\quad\times\int p(\alpha)p(\beta)\prod_i p(b_i)\prod_t p(Y_{it+1}\mid H_{it},A_{it},b_i;\alpha,\beta)d\alpha d\beta,\end{aligned} \tag{19}$$

where the last equality follows from the conditional independence assumption and the randomization of A_{it}. The denominator of the right hand side of (18) is $\int\int p(b,X,A,Y,\alpha,\beta)d\alpha d\beta db$. Thus, the posterior distribution (18) equals

$$\frac{\int p(\alpha)p(\beta)\prod_i p(b_i)\prod_t p(Y_{it+1}\mid H_{it},A_{it},b_i;\alpha,\beta)d\alpha d\beta}{\int p(\alpha)p(\beta)\prod_i p(b_i)\prod_t p(Y_{it+1}\mid H_{it},A_{it},b_i;\alpha,\beta)d\alpha d\beta db}, \tag{20}$$

which is the posterior distribution of b in a standard LMM when X and A are treated as fixed or exogenous.

Therefore, the Bayesian MAP estimator of b can be obtained through standard LMM fitting procedure. Along the same line, the empirical Bayes estimator of b with plug-in variance component estimates can also be obtained through standard LMM.

Appendix B: Additional simulation results

In the additional simulation results, we included simulations for sample size $n=30,50,100,200$ and the number of observations per individual $T_i=T=10,20,30$. Each setting was replicated 1,000 times. Bias, standard deviation (sd) and coverage probability (cp) of 95% nominal confidence interval for the estimated fixed effects (β's and α's) are presented in Table 4. Table 5 presents the bias and

standard deviation for the estimated variance components $\sigma^2_{bj}, 0 \leq j \leq 3$ and σ^2_ϵ. For GM 1 and GM 3, the model doesn't include b_{i1} and b_{i3}, so the variance components only include σ^2_{b0},σ^2_{b2} and σ^2_ϵ. Conclusion to Section 4 can be made: for GM 1 and GM 2, the variance components are consistently estimated, whereas for GM 3 the estimators are inconsistent. Again, this is due to violation of the conditional independence assumption (10) in GM 3.

Acknowledgements

Research presented in this paper was supported by the National Heart, Lung and Blood Institute under award number R01HL125440; the National Institute on Alcohol Abuse and Alcoholism under award number R01AA023187; the National Institute on Drug Abuse under award number P50DA039838; and the National Institute of Biomedical Imaging and Bioengineering under award number U54EB020404. The authors would like to thank Peng Liao, Walter Dempsey, two reviewers, and the associate editor for helpful suggestions.

References

Agresti, A., Booth, J. G., Hobert, J. P. and Caffo, B. (2000). Random-Effects Modeling of Categorical Response Data. *Sociological Methodology* **30** 27–80.

Amemiya, T. and MaCurdy, T. E. (1986). Instrumental-variable estimation of an error-components model. *Econometrica: Journal of the Econometric Society* 869–880.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies* **58** 277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics* **68** 29–51.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67** 1–48.

Berger, M. P. and Tan, F. E. (2004). Robust designs for linear mixed effects models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **53** 569–581.

Bolger, N. and Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.

Boruvka, A., Almirall, D., Witkiewitz, K. and Murphy, S. A. (2018). Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association* **113** 1112–1121.

Brumback, B., Greenland, S., Redman, M., Kiviat, N. and Diehr, P. (2003). The intensity-score approach to adjusting for confounding. *Biometrics* **59** 274–285.

Cheung, M. W.-L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological methods* **13** 182.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** 165–185.

Daniel, R. M., Cousens, S., De Stavola, B., Kenward, M. G. and Sterne, J. (2013). Methods for dealing with time-dependent confounding. *Statistics in medicine* **32** 1584–1618.

Dempfle, L. (1977). Comparison of several sire evaluation methods in dairy cattle breeding. *Livestock Production Science* **4** 129–139.

			β_0			β_1			α_0			α_1		
GM	T	n	bias	sd	cp	bias	sd	cp	bias	sd	cp	bias	sd	cp
		30	-0.001	0.249	0.943	0.002	0.091	0.897	-0.021	0.377	0.951	-0.002	0.065	0.915
		50	-0.002	0.187	0.953	-0.001	0.068	0.897	-0.019	0.295	0.947	-0.001	0.048	0.930
1	10	100	-0.003	0.135	0.941	-0.001	0.049	0.898	-0.011	0.210	0.949	-0.001	0.033	0.920
		200	-0.001	0.096	0.926	-0.001	0.034	0.899	-0.009	0.150	0.941	0.000	0.025	0.909
		30	-0.001	0.217	0.943	0.001	0.063	0.919	-0.020	0.372	0.950	-0.002	0.046	0.928
		50	0.001	0.168	0.947	-0.000	0.048	0.916	-0.018	0.288	0.945	-0.002	0.034	0.935
1	20	100	-0.002	0.117	0.950	-0.000	0.035	0.906	-0.010	0.207	0.946	-0.000	0.025	0.930
		200	-0.001	0.085	0.943	-0.001	0.026	0.892	-0.008	0.147	0.944	0.000	0.018	0.921
		30	-0.002	0.206	0.946	0.001	0.053	0.913	-0.020	0.367	0.952	-0.001	0.038	0.924
		50	-0.000	0.160	0.949	0.001	0.040	0.930	-0.017	0.288	0.945	-0.001	0.028	0.940
1	30	100	-0.005	0.112	0.949	-0.001	0.028	0.935	-0.009	0.205	0.944	0.000	0.020	0.938
		200	0.000	0.081	0.944	-0.001	0.022	0.902	-0.009	0.146	0.946	0.000	0.015	0.923
		30	-0.010	0.269	0.939	-0.004	0.105	0.903	-0.015	0.391	0.950	-0.003	0.079	0.933
		50	-0.011	0.209	0.932	-0.000	0.078	0.909	-0.010	0.302	0.941	0.001	0.062	0.931
2	10	100	0.009	0.145	0.933	-0.001	0.056	0.915	-0.012	0.222	0.934	-0.002	0.045	0.929
		200	-0.008	0.105	0.931	-0.002	0.038	0.934	-0.007	0.150	0.960	0.001	0.031	0.935
		30	-0.005	0.229	0.943	-0.001	0.079	0.930	-0.014	0.377	0.951	-0.002	0.067	0.940
		50	-0.008	0.180	0.944	0.001	0.061	0.929	-0.014	0.292	0.951	-0.001	0.053	0.931
2	20	100	0.007	0.123	0.942	0.001	0.044	0.931	-0.012	0.213	0.945	-0.003	0.038	0.940
		200	-0.007	0.090	0.933	-0.001	0.030	0.939	-0.006	0.147	0.957	0.001	0.026	0.945
		30	-0.006	0.216	0.943	-0.001	0.070	0.939	-0.014	0.374	0.951	-0.002	0.062	0.946
		50	-0.008	0.168	0.957	0.001	0.055	0.945	-0.016	0.289	0.951	-0.002	0.049	0.942
2	30	100	0.006	0.115	0.947	-0.001	0.039	0.948	-0.010	0.210	0.943	-0.002	0.035	0.934
		200	-0.004	0.084	0.935	-0.000	0.027	0.940	-0.008	0.145	0.950	0.000	0.025	0.942
		30	-0.048	0.245	0.949	-0.043	0.075	0.725	0.048	0.341	0.951	0.057	0.060	0.629
		50	-0.049	0.189	0.940	-0.045	0.055	0.674	0.053	0.265	0.949	0.059	0.044	0.519
3	10	100	-0.060	0.134	0.927	-0.047	0.041	0.548	0.063	0.190	0.931	0.061	0.031	0.283
		200	-0.052	0.095	0.907	-0.046	0.029	0.355	0.064	0.135	0.924	0.061	0.022	0.079
		30	-0.029	0.216	0.945	-0.024	0.051	0.798	0.016	0.351	0.955	0.028	0.038	0.766
		50	-0.035	0.168	0.950	-0.027	0.039	0.762	0.022	0.273	0.949	0.030	0.028	0.714
3	20	100	-0.038	0.119	0.931	-0.027	0.028	0.666	0.029	0.194	0.948	0.030	0.021	0.548
		200	-0.034	0.083	0.935	-0.027	0.019	0.514	0.031	0.137	0.953	0.031	0.014	0.272
		30	-0.023	0.207	0.946	-0.017	0.041	0.847	0.005	0.354	0.954	0.018	0.031	0.832
		50	-0.026	0.159	0.946	-0.018	0.031	0.822	0.010	0.275	0.948	0.019	0.022	0.794
3	30	100	-0.028	0.112	0.942	-0.019	0.022	0.762	0.016	0.197	0.950	0.020	0.016	0.658
		200	-0.024	0.079	0.941	-0.019	0.015	0.628	0.018	0.139	0.950	0.021	0.011	0.438

TABLE 4

Bias, standard deviation (sd) and coverage probability (cp) of 95% nominal confidence interval for the fixed effect parameters in the simulation study. n denotes sample size; T denotes total number of observations for each individual; GM denotes generative model. The result is based on 1,000 replicates for each setting.

			σ^2_{b2}		σ^2_{b3}		σ^2_{b0}		σ^2_{b1}		σ^2_{ϵ}	
n	T	GM	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
		30	0.024	0.400	-	-	-0.008	1.137	-	-	-0.003	0.049
		50	0.013	0.300	-	-	-0.020	0.868	-	-	-0.002	0.035
1	10	100	0.017	0.210	-	-	-0.031	0.614	-	-	-0.001	0.024
		200	0.004	0.151	-	-	-0.021	0.431	-	-	-0.000	0.017
		30	0.012	0.319	-	-	-0.025	1.067	-	-	-0.003	0.032
		50	0.010	0.246	-	-	-0.026	0.822	-	-	-0.001	0.023
1	20	100	0.008	0.174	-	-	-0.041	0.579	-	-	-0.001	0.016
		200	0.004	0.126	-	-	-0.021	0.403	-	-	-0.000	0.011
		30	0.003	0.293	-	-	-0.036	1.036	-	-	-0.002	0.025
		50	0.001	0.232	-	-	-0.037	0.809	-	-	-0.001	0.018
1	30	100	0.008	0.163	-	-	-0.040	0.569	-	-	-0.001	0.013
		200	0.000	0.116	-	-	-0.023	0.399	-	-	-0.000	0.009
		30	0.047	0.498	-0.001	0.058	-0.003	1.238	-0.003	0.040	-0.003	0.048
		50	0.048	0.392	-0.005	0.046	-0.057	0.935	-0.004	0.033	-0.001	0.038
2	10	100	0.000	0.260	-0.003	0.033	-0.019	0.646	-0.001	0.022	0.000	0.027
		200	0.005	0.184	-0.003	0.021	-0.043	0.451	-0.001	0.015	-0.001	0.019
		30	0.009	0.367	-0.003	0.043	-0.029	1.094	-0.003	0.032	-0.000	0.031
		50	0.022	0.302	-0.003	0.033	-0.045	0.854	-0.002	0.025	0.000	0.025
2	20	100	0.002	0.200	-0.002	0.021	-0.016	0.597	-0.000	0.017	0.001	0.017
		200	-0.001	0.142	-0.001	0.015	-0.029	0.418	-0.001	0.012	-0.001	0.012
		30	0.001	0.334	-0.002	0.036	-0.045	1.065	-0.003	0.029	0.000	0.025
		50	0.012	0.268	-0.003	0.027	-0.049	0.826	-0.002	0.022	0.000	0.019
2	30	100	0.002	0.183	-0.001	0.019	-0.028	0.584	0.000	0.016	0.000	0.013
		200	-0.003	0.127	-0.001	0.013	-0.029	0.409	-0.001	0.011	-0.000	0.009
		30	0.126	0.434	-	-	-0.710	1.159	-	-	0.004	0.046
		50	0.105	0.329	-	-	-0.771	0.860	-	-	0.005	0.034
3	10	100	0.094	0.228	-	-	-0.810	0.604	-	-	0.005	0.025
		200	0.080	0.159	-	-	-0.796	0.429	-	-	0.006	0.018
		30	0.059	0.329	-	-	-0.380	1.056	-	-	0.000	0.029
		50	0.053	0.262	-	-	-0.428	0.800	-	-	0.001	0.023
3	20	100	0.040	0.174	-	-	-0.429	0.575	-	-	0.001	0.017
		200	0.038	0.125	-	-	-0.430	0.406	-	-	0.002	0.011
		30	0.040	0.304	-	-	-0.268	1.029	-	-	-0.000	0.024
		50	0.030	0.237	-	-	-0.296	0.782	-	-	-0.001	0.018
3	30	100	0.027	0.162	-	-	-0.306	0.569	-	-	0.000	0.013
		200	0.023	0.115	-	-	-0.299	0.395	-	-	0.001	0.009

TABLE 5

Bias and standard deviation (sd) for the estimated variance components $\sigma^2_{bj}, 0 \leq j \leq 3$ and σ^2_{ϵ} in the simulation study. n denotes sample size; T denotes total number of observations for each individual; GM denotes generative model. For GM 1 and GM 3, the model doesn't include b_{i1} and b_{i3}, so the corresponding entries in the table are left blank. The result is based on 1,000 replicates for each setting.

Dempsey, W., Liao, P., Klasnja, P., Nahum-Shani, I. and Murphy, S. A. (2015). Randomised trials for the Fitbit generation. *Significance* **12** 20–23.

Diggle, P., Heagerty, P., Liang, K.-Y., Zeger, S. et al. (2002). *Analysis of longitudinal data*. Oxford university press.

Ebbes, P., Böckenholt, U. and Wedel, M. (2004). Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica* **58** 161–178.

Garcia, T. P. and Ma, Y. (2016). Optimal Estimator for Logistic Model with Distribution-free Random Intercept. *Scandinavian Journal of Statistics* **43** 156–171.

Goetgeluk, S. and Vansteelandt, S. (2008). Conditional generalized estimating equations for the analysis of clustered and longitudinal data. *Biometrics* **64** 772–780.

Grilli, L. and Rampichini, C. (2011). The role of sample cluster means in multilevel models. *Methodology*.

Hanchane, S. and Mostafa, T. (2012). Solving endogeneity problems in multilevel estimation: an example using education production functions. *Journal of Applied Statistics* **39** 1101–1114.

Hausman, J. A. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric Society* 1377–1398.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* **55** 688–698.

Heagerty, P. J. and Zeger, S. L. (2000). Marginalized multilevel models and likelihood inference. *Statistical Science* **15** 1–26.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 423–447.

Hernán, M. A., Brumback, B. and Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* **96** 440–448.

Hernán, M. A. and Robins, J. M. (2019). *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming.

Keogh, R. H., Daniel, R. M., VanderWeele, T. J. and Vansteelandt, S. (2017). Analysis of longitudinal studies with repeated outcome measures: adjusting for time-dependent confounding using conventional methods. *American journal of epidemiology* **187** 1085–1092.

Kim, J.-S. and Frees, E. W. (2006). Omitted variables in multilevel models. *Psychometrika* **71** 659.

Kim, J.-S. and Frees, E. W. (2007). Multilevel modeling with correlated effects. *Psychometrika* **72** 505–533.

Klasnja, P., Smith, S., Seewald, N. J., Lee, A., Hall, K., Luers, B., Hekler, E. B. and Murphy, S. A. (2018). Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. *Annals of Behavioral Medicine*.

Kuznetsova, A., Brockhoff, P. B. and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82** 1–26.

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

Larsen, K., Petersen, J. H., Budtz-Jørgensen, E. and Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics* **56** 909–914.

Liao, P., Klasnja, P., Tewari, A. and Murphy, S. A. (2016). Sample size calculations for micro-randomized trials in mHealth. *Statistics in medicine* **35** 1944–1971.

Lindley, D. V. and Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the*

Royal Statistical Society. Series B (Methodological) 1–41.

Liu, L. and Xiang, L. (2014). Semiparametric estimation in generalized linear mixed models with auxiliary covariates: A pairwise likelihood approach. *Biometrics* **70** 910–919.

Luger, T. M., Suls, J. and Vander Weg, M. W. (2014). How robust is the association between smoking and depression in adults? A meta-analysis using linear mixed-effects models. *Addictive behaviors* **39** 1418–1429.

Miglioretti, D. L. and Heagerty, P. J. (2004). Marginal modeling of multilevel binary data with time-varying covariates. *Biostatistics* **5** 381–398.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society* 69–85.

Neugebauer, R., van der Laan, M. J., Joffe, M. M. and Tager, I. B. (2007). Causal inference in longitudinal studies with history-restricted marginal structural models. *Electronic journal of statistics* **1** 119.

Neuhaus, J. M. and McCulloch, C. E. (2006). Separating between-and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 859–872.

Pan, W., Louis, T. A. and Connett, J. E. (2000). A note on marginal linear regression with correlated response data. *The American Statistician* **54** 191–195.

Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation and Computation* **23** 939–951.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* **1**. Sage.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. *Mathematical modelling* **7** 1393–1512.

Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods* **23** 2379–2412.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality* 69–117. Springer.

Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in medicine* **17** 269–302.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials* 95–133. Springer.

Roy, J., Alderson, D., Hogan, J. W. and Tashima, K. T. (2006). Conditional inference methods for incomplete Poisson data with endogenous time-varying covariates: Emergency department use among HIV-infected women. *Journal of the American Statistical Association* **101** 424–434.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika* **6** 309–316.

Schildcrout, J. S. and Heagerty, P. J. (2005). Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics* **6** 633–652.

Schwartz, J. E. and Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In *The science of real-time data capture: Self-reports in health research* (A. A. Stone, A. Atienza, L. Nebeling and S. Shiffman, eds.) 76–113. Oxford University Press New York, NY.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance components*. John Wiley & Sons.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82** 605–610.

Semykina, A. and Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* **157** 375–380.

Shardell, M. and Ferrucci, L. (2018). Joint mixed-effects models for causal inference with longitudinal data. *Statistics in medicine* **37** 829–846.

Sitlani, C. M., Heagerty, P. J., Blood, E. A. and Tosteson, T. D. (2012). Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Statistics in medicine* **31** 1738–1760.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* 1171–1177.

Tchetgen, E. J. T., Glymour, M. M., Weuve, J. and Robins, J. (2012). Specifying the correlation structure in inverse-probability-weighting estimation for repeated measures. *Epidemiology* **23** 644–646.

Vansteelandt, S. (2007). On confounding, prediction and efficiency in the analysis of longitudinal and cross-sectional clustered data. *Scandinavian journal of statistics* **34** 478–498.

Wang, Z. and Louis, T. A. (2004). Marginalized binary mixed-effects models with covariate-dependent random effects and likelihood inference. *Biometrics* **60** 884–891.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT press.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 121–130.

Zeger, S. L., Liang, K.-Y. and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1049–1060.

Zeger, S. L. and Liang, K.-Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in medicine* **11** 1825–1839.