Published in final edited form as:

Stat Med. 2018 February 28; 37(5): 829–846. doi:10.1002/sim.7567.

Joint mixed-effects models for causal inference with longitudinal data

Michelle Shardell^{*} and Luigi Ferrucci

National Institute on Aging, 3001 S. Hanover Street, Baltimore, Maryland 21225

Abstract

Causal inference with observational longitudinal data and time-varying exposures is complicated due to the potential for time-dependent confounding and unmeasured confounding. Most causal inference methods that handle time-dependent confounding rely on either the assumption of no unmeasured confounders or the availability of an unconfounded variable that is associated with the exposure (e.g., an instrumental variable). Furthermore, when data are incomplete, validity of many methods often depends on the assumption of missing at random. We propose an approach that combines a parametric joint mixed-effects model for the study outcome and the exposure with g-computation to identify and estimate causal effects in the presence of time-dependent confounding and unmeasured confounding. G-computation can estimate participant-specific or population-average causal effects using parameters of the joint model. The joint model is a type of shared parameter model where the outcome and exposure-selection models share common random effect(s). We also extend the joint model to handle missing data and truncation by death when missingness is possibly not at random. We evaluate the performance of the proposed method using simulation studies and compare the method to both linear mixed-effects models and fixed-effects models combined with g-computation as well as to targeted maximum likelihood estimation. We apply the method to an epidemiologic study of vitamin D and depressive symptoms in older adults and include code using SAS PROC NLMIXED software to enhance the accessibility of the method to applied researchers.

Keywords

Causal inference; longitudinal data; missing not at random; time-dependent confounding; unmeasured confounding

1. Introduction

Estimating causal effects from observational longitudinal studies with time-varying exposures is rife with challenges, including time-dependent confounding and unmeasured confounding. Appropriately addressing these challenges is essential to unbiased estimation of causal effects.

^{*}Correspondence to: Harbor Hospital Rm NM529, National Institute on Aging, 3001 S. Hanover Street, Baltimore, Maryland 21225, USA. michelle.shardell@nih.gov, Phone: 410-350-7370.

The available methodology for performing causal inference with longitudinal data has largely focused on semiparametric modeling of population-average causal effects. Most methods require either the assumption of no unmeasured confounding (e.g., marginal structural models, targeted maximum likelihood estimation) or availability of an unconfounded variable that is associated with the exposure (e.g., instrumental variable methods or g-estimation of structural nested models).^{1–4}

There has been relatively little work on causal inference with longitudinal data using mixed-effects models⁵ in which causal effects are defined using potential outcomes. Sitlani *et al*⁶ considered linear structural mixed models for randomized trials with non-compliance. Although the authors mentioned a joint model for outcome and exposure to handle unmeasured confounding and an outcome model that includes serial correlation as a source of time-dependent confounding, the authors focused evaluation and discussion on outcome models without serial correlation conditioned on the random effects and exposure models that were independent of random effects. Bind *et al*⁷ considered causal mediation analysis with separate mixed-effects models for the mediator and the outcome assuming no unmeasured confounding and no time-dependent confounding. The models were linked through (possibly) correlated random effects. Kennedy *et al*⁸ used mixed-effects models as the first of a two-stage approach to appropriately overcome time-dependent confounding in time-to-event analysis. Small *et al*⁹ and Holmes *et al*¹⁰ proposed mixed-effects models to handle non-adherence in randomized trials. Zhang *et al*¹¹ addressed heterogeneity of treatment effects in randomized trials by sensitivity analysis of a mixed-effects model in which potential outcomes are presumed conditionally independent given observed and unobserved factors.

In this paper, we focus on joint mixed-effects models for longitudinal outcomes and time-varying exposures in settings with time-dependent confounding. Joint models were originally introduced as a method for handling nonignorably missing data^{12, 13} when estimating the outcome data-generating model. We adapt the method to handle unmeasured confounding and to identify causal effects by performing g-computation¹⁴ using the joint-model parameters. We additionally consider these issues in conjunction with selective attrition in the form of data missing not at random in the sense of Rubin;¹⁵ that is, nonignorably missing data, and truncation by death. A simulation study compares and contrasts the performance of g-computation using parameters from the joint mixed-effects model with that of g-computation using parameters from conventional linear mixed models and fixed-effects models. We also compare the joint mixed-effects model to targeted maximum likelihood estimation, a semiparametric method that depends on the assumption of no unmeasured confounding and is doubly robust in that correct estimation of only one of the outcome-generating model or exposure-selection model is required for unbiased estimation.²

We consider epidemiological studies of older adults to motivate the methods. When estimating causal effects of potential risk factors for aging-related adverse outcomes, important sources of outcome heterogeneity and unmeasured confounding include lifecourse factors; that is, exposures that occurred during gestation, childhood, adolescence, young adulthood, and adult life that may affect health in old age.¹⁶ Thus, mixed-effects models

may be particularly attractive when studying the older adult population. To be concrete, we focus on a study examining the causal effects of serum vitamin D on depressive symptoms over time in older adults. It is plausible that unmeasured antecedent factors may affect severity of depressive symptoms in later life, and that some of these same factors may influence later lifestyle choices and physical health with implications on serum vitamin D concentrations.

2. Data, Definitions, Models, and Estimation

2.1. Data and Definitions

Consider a longitudinal study with n participants where participant i has J_i scheduled follow-up visits. Let Y_{ij} denote the study outcome for participant i at visit j ($j = 0, \dots, J_i$) such that Y_{i0} is the baseline value of the outcome. Similarly, A_{ij} is the study exposure for participant i at visit j . Let $\mathbf{X}_i$ be a vector of observed baseline exposure-outcome confounders.

The analytical objective is to estimate causal effects of exposure history through visit $j-1$, denoted $\mathbf{a}_{i,j-1} = \{A_{i0}, A_{i1}, \dots, A_{i,j-2}, A_{i,j-1}\}$, on Y_{ij} . Let $Y_{ij}(\mathbf{a}_{i,j-1})$ denote the potential outcome for participant i at visit j that would have been observed if exposure history $\mathbf{a}_{i,j-1}$ were set to $\mathbf{a}_{i,j-1} = \{a_0, \dots, a_{j-1}\}$. The causal effect of $\mathbf{a}_{i,j-1}$ on Y_{ij} is a contrast of the average of $Y_{ij}(\mathbf{a}_{i,j-1})$ versus $Y_{ij}(\mathbf{a}_{i,j-1}')$, where $\mathbf{a}_{i,j-1}'$ is an alternative exposure history. We make the consistency assumption to link potential outcomes with observed outcomes. That is, we assume that $Y_{ij} = Y_{ij}(\mathbf{a}_{i,j-1})$ if $\mathbf{a}_{i,j-1} = \mathbf{a}_{i,j-1}$. We make the stable unit treatment value assumption,¹⁷ which means that the exposure history of a participant does not influence the outcomes of other participants and that there are not multiple versions of the exposure. This assumption allows us to denote potential outcomes with $Y_{ij}(\mathbf{a}_{i,j-1})$ rather than $Y_{ij}(\mathbf{a}_{i,j-1})$, where $\mathbf{a}_{i,j-1} = \{a_{1,j-1}, \dots, a_{n,j-1}\}$.

2.2. Models

We posit a linear mixed-effects model for the data-generating mechanism of outcome Y . In general, let

$$Y_{ij} = \beta_0 + \mathbf{X}_i^t \boldsymbol{\beta}_1 + \bar{\mathbf{Y}}_{i,j-1}^t \boldsymbol{\beta}_2 + \bar{\mathbf{A}}_{i,j-1}^t \boldsymbol{\beta}_3 + (\mathbf{X}_i \otimes \bar{\mathbf{A}}_{i,j-1})^t \boldsymbol{\beta}_4 + \beta_5 j + \mathbf{Z}_{ij}^t \mathbf{b}_i + \varepsilon_{ij} \text{ for } j=0, \dots, J_i,$$

(1)

where $\mathbf{b}_i$ is a vector of random effects, $\mathbf{Z}_{ij}$ is a design matrix for the random effects, ε_{ij} is a measurement error, and $\bar{\mathbf{A}}_{i,j-1}^t \boldsymbol{\beta}_3 = \beta_{30} A_{i0} + \dots + \beta_{3,j-1} A_{i,j-1}$ with other terms similarly defined. Note that serial correlation (and hence time-dependent confounding) is considered by allowing Y_{ij} to depend on past values of Y , $\bar{\mathbf{Y}}_{i,j-1}$. Thus, coefficients of $\mathbf{A}$ may not have a causal interpretation.

Throughout, we assume that

$$\varepsilon_i \text{ iid } N(0, \sigma^2) \quad (2)$$

and

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}).$$

To be more specific and parsimonious, we will focus on binary exposures using the following outcome model:

$$Y_{ij} = \beta_0 + \mathbf{X}_i^t \boldsymbol{\beta}_1 + \beta_2 Y_{i,j-1} + \bar{A}_{i,j-1}^t \boldsymbol{\beta}_3 + (\mathbf{X}_{i1} \bar{A}_{i,j-1})^t \boldsymbol{\beta}_4 + \beta_5 j + b_{i0} + b_{i1} \left(\mathbf{X}_{i1} + \mathbf{X}_{i1} \sum_{k=0}^{j-1} A_{ik} \right) + \varepsilon_{ij}, \quad (3)$$

where $\bar{Y}_{i,k} = \bar{Y}_{i,k}$ for $k < 0$.

Thus, we consider a random intercept (b_{i0}) as well as a random slope (b_{i1}) for a baseline covariate $\mathbf{X}_{i1}$ that also depends on cumulative exposure history. These random effects can be interpreted as unmeasured sources of heterogeneity that influence the value of the outcome and the effect of observed factors on the outcome. For example, in an epidemiologic study of vitamin D and depressive symptoms in older adults, $\mathbf{X}_{i1}$ may be age at study baseline; thus b_{i1} accounts for the possibility that heterogeneity of age and age-specific vitamin D effects on depressive symptoms may be due, for example, to unmeasured lifecourse factors or subclinical biological factors. As noted by Sitlani *et al.*,⁶ a special case of Equation (3) that does not include heterogeneity of exposure effects (i.e., $b_{i1} = 0$ for all i) is an example of a rank-preserving model,¹⁸ because the participants' outcome ranks are preserved across all possible exposure histories. In contrast, Equation (3) in general is an example of a non-rank-preserving model and may be more realistic in some applications.

The same unmeasured factors that influence the study outcome may also influence selection of exposure over time. The selection process for binary exposures may follow a model such as

$$\text{logit}(p_{i,j-1}^A) = \alpha_0 + \mathbf{X}_i^t \boldsymbol{\alpha}_1 + \alpha_2 Y_{i,j-1} + \bar{A}_{i,j-2}^t \boldsymbol{\alpha}_3 + (\mathbf{X}_i Y_{i,j-1})^t \boldsymbol{\alpha}_4 + \alpha_5 j + \alpha_6 b_{i0} + \alpha_7 b_{i1} \text{ for } j=1, \dots, J_i, \quad (4)$$

where $p_{i,j-1}^A$ denotes $E[A_{i,j-1} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}, \mathbf{b}_i]$.

If the outcome and exposure-selection models are correctly specified by Equations (3) and (4), respectively, then unbiased parameter estimation can proceed by maximizing the joint likelihood for the outcome and exposure selection,

$$\begin{aligned} & \prod_i \int f(\bar{Y}_{iJ_i}, \bar{A}_{iJ_i-1} | \mathbf{X}_i, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}) f(\mathbf{b}_i; \mathbf{D}) d\mathbf{b}_i \\ &= \prod_i \int \prod_j f(Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) \times f(A_{i,j-1} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}, \mathbf{b}_i; \boldsymbol{\alpha}) f(\mathbf{b}_i; \mathbf{D}) d\mathbf{b}_i \end{aligned} \quad (5)$$

where $f(\cdot)$ denotes a probability density (or mass) function. As per Equation (2), $f(Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2)$ is the normal density. As per Equation (4), The joint mixed-effects model Equation (5) is an example of a shared parameter model,^{12, 13, 19} because the outcome model and exposure-selection model have common (i.e, shared) random effects. In Equation (5), the outcome model and exposure-selection model are simultaneously fit to derive unbiased estimates of $\boldsymbol{\beta}$. If, however, $\mathbf{b}_i$ does not influence exposure selection, then Equation (5) can be factored as

$$\begin{aligned} & \prod_i \int f(\bar{Y}_{iJ_i}, \bar{A}_{iJ_i-1} | \mathbf{X}_i, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}) f(\mathbf{b}_i; \mathbf{D}) d\mathbf{b}_i \\ &= \prod_i \int f(A_{i,j-1} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}; \boldsymbol{\alpha}) \\ & \quad \times \int \prod_j f(Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) f(\mathbf{b}_i; \mathbf{D}) d\mathbf{b}_i. \end{aligned} \quad (6)$$

In model (6), the exposure-selection model factors from the outcome model. Thus, parameters $\boldsymbol{\beta}$ can be estimated by maximizing the outcome likelihood using standard methods for linear mixed-effects models.

We have hitherto focused on estimating parameters that describe the data-generating mechanism, but that may not have a causal interpretation. We now turn our attention to estimating causal effects using these parameters. First, we consider a directed acyclic graph (Figure 1) to convey the assumptions of unmeasured confounding via arrows from $\mathbf{b}_i$ to both the exposure A and the outcome Y . Figure 1 also conveys time-dependent confounding via the pathways $A_{i,j-1} \rightarrow Y_{ij} \rightarrow A_{ij} \rightarrow Y_{i,j+1}$ and $A_{i,j-1} \rightarrow Y_{i,j+1}$. Maximizing Equation (5) addresses unmeasured confounding when estimating the data-generating process. The parameters $\boldsymbol{\beta}$ do not, in general, have a causal interpretation owing to the time-dependent confounding. However, under certain assumptions, functions of $\boldsymbol{\beta}$ can identify causal effects. Namely, the following assumption, sequential randomization conditional on random effects, must hold to identify causal effects from $\boldsymbol{\beta}$:

$$\bar{Y}_i(\bar{a}) \perp A_{i,j-1} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2} = \bar{a}_{j-2}, \mathbf{b}_i \text{ for all } \bar{a} \in \bar{\mathcal{A}} \text{ and } j=1, \dots, J_i, \quad (7)$$

where $\bar{Y}_i(\cdot) = \{Y_{i1}(a_0), Y_{i2}(\cdot), \dots, Y_{iJ_i}(\cdot)\}$ and $\bar{\mathcal{A}}$ is the support of $\bar{a}$; that is, $\bar{\mathcal{A}}$ is the collection of all possible exposure histories. Assumption (7) holds if there are no unmeasured exposure-outcome confounders at time j after conditioning on baseline covariates, outcome and exposure history, and random effects. For example, causal effects can be identified by maximizing Equation (5) under (correct specification of) the exposure-selection model (4) and outcome model (3). If $\alpha_6 = 0$ and $\alpha_7 = 0$, then causal effects can be identified by fitting the outcome-model using standard methods for linear mixed-effects models; that is, the exposure-selection model need not be estimated owing to the factoring in Equation (6). Alternatively, methods that can produce unbiased estimates by correctly modeling the exposure-selection mechanism can be used in this case including marginal structural models or doubly robust methods such as longitudinal targeted maximum likelihood.^{1, 2}

We use g-computation¹⁴ to identify causal effects as functions of β . G-computation generalizes the idea of covariate standardization to complex longitudinal settings. The idea is to integrate outcome expected values over covariate histories while treating the exposure as fixed (see Daniel et al.²⁰ for an accessible illustration). Thus, we define

$$g_j(\bar{a}_{j-1} | \mathbf{X}_i, \mathbf{b}_i) = \int E[Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1} = \bar{a}_{j-1}, \mathbf{b}_i] \prod_{k=0}^{j-1} f(Y_{i,k} | \mathbf{X}_i, \bar{Y}_{i,k-1}, \bar{A}_{i,k-1} = \bar{a}_{k-1}, \mathbf{b}_i) dY_{i,k}. \quad (8)$$

We refer to (8) as conditional g-computation, because contrasts of the form

$g_j(\bar{a}_{j-1} | \mathbf{X}_i, \mathbf{b}_i) - g_j(\bar{a}'_{j-1} | \mathbf{X}_i, \mathbf{b}_i)$ are causal contrasts conditioned on baseline covariates and random effects. Depending on study objectives, researchers may prefer to additionally integrate over baseline covariates and/or random effects. Computing

$$g_j(\bar{a}_{j-1}) = \int g_j(\bar{a}_{j-1} | \mathbf{X}_i, \mathbf{b}_i) f(\mathbf{X}_i) d\mathbf{X}_i f(\mathbf{b}_i) d\mathbf{b}_i \quad (9)$$

and $g_j(\bar{a}_{j-1}) - g_j(\bar{a}'_{j-1})$ results in population-average causal contrasts. For example, consider the model in Equation (3) for $j = 1, 2$. At $j = 1$,

$$E[Y_{i1} | \mathbf{X}_i, Y_{i0}, A_{i0} = a_0, \mathbf{b}_i] = (\beta_0 + \beta_5) + \mathbf{X}_i^t \beta_1 + \beta_2 Y_{i0} + \beta_{30} a_0 + \beta_{40} X_{i1} a_0 + b_{i0} + b_{i1} (X_{i1} + X_{i1} a_0).$$

Integrating over $f(Y_{j0} | \mathbf{X}_j, \mathbf{b}_j)$ involves replacing Y_{j0} with its expected value conditioned on $\mathbf{X}_j$ and $\mathbf{b}_j$,

$$E[Y_{j0} | \mathbf{X}_j, \mathbf{b}_j] = \beta_0 + \mathbf{X}_j^t \boldsymbol{\beta}_1 + b_{j0} + b_{j1} X_{j1},$$

resulting in

$$g_1(a_0 | \mathbf{X}_i, \mathbf{b}_i) = (\beta_0 + \beta_5) + \mathbf{X}_i^t \boldsymbol{\beta}_1 + \beta_2(\beta_0 + \mathbf{X}_i^t \boldsymbol{\beta}_1 + b_{i0} + b_{i1} X_{i1}) + \beta_{30} a_0 + \beta_{40} X_{i1} a_0 + b_{i0} + b_{i1} (X_{i1} + X_{i1} a_0) = (\beta_0 + \beta_5 + b_{i0} + \beta_2 \beta_0 + \beta_2 b_{i0}) + \mathbf{X}_i^t (\boldsymbol{\beta}_1 + \beta_2 \boldsymbol{\beta}_1) + X_{i1} (b_{i1} + \beta_2 b_{i1}) + a_0 (\beta_{30} + \beta_{40} X_{i1} + b_{i1} X_{i1}).$$

(10)

Therefore, $g_1(a_0 = 1 | \mathbf{X}_j, \mathbf{b}_j) - g_1(a_0 = 0 | \mathbf{X}_j, \mathbf{b}_j) = \beta_{30} + \beta_{40} X_{j1} + b_{j1} X_{j1}$. Furthermore, $g_1(a_0 = 1) - g_1(a_0 = 0) = \beta_{30} + \beta_{40} E[X_{j1}]$, because random effects are assumed to be independent of $\mathbf{X}_j$ and to have mean 0.

To perform g-computation at $j = 2$, note that

$$E[Y_{i2} | \mathbf{X}_i, \bar{Y}_{i1}, \bar{A}_{i1} = \bar{a}_i, \mathbf{b}_i] = (\beta_0 + 2\beta_5) + \mathbf{X}_i^t \boldsymbol{\beta}_1 + \beta_2 \bar{Y}_{i1} + \beta_{30} a_0 + \beta_{31} a_1 + \beta_{40} X_{i1} a_0 + \beta_{41} X_{i1} a_1 + b_{i0} + b_{i1} [X_{i1} + X_{i1} (a_0 + a_1)].$$

Integrating over $f(Y_{j1} | \mathbf{X}_j, Y_{j0}, A_{j0} = a_0, \mathbf{b}_j)$ and $f(Y_{j0} | \mathbf{X}_j, \mathbf{b}_j)$ results in

$$g_2(\bar{a}_1 | \mathbf{X}_i, \mathbf{b}_i) - g_2(\{0, 0\} | \mathbf{X}_i, \mathbf{b}_i) = a_0 (\beta_{30} + \beta_2 \beta_{30} + \beta_{40} X_{i1} + \beta_2 \beta_{40} X_{i1} + b_{i1} X_{i1} + \beta_2 b_{i1} X_{i1}) + a_1 (\beta_{31} + \beta_{41} X_{i1} + b_{i1} X_{i1}).$$

Further integration over $\mathbf{X}_j$ and $\mathbf{b}_j$ results in

$$g_2(\bar{a}_1) - g_2(\{0, 0\}) = a_0 (\beta_{30} + \beta_2 \beta_{30} + \beta_{40} E[X_{i1}] + \beta_2 \beta_{40} E[X_{i1}]) + a_1 (\beta_{31} + \beta_{41} E[X_{i1}]).$$

2.3. Some Extensions

2.3.1. Missing Data and Truncation by Death—Joint mixed-effects models were originally proposed to address nonignorable attrition.^{12, 13} These models have evolved to handle joint longitudinal and time-to-event analysis.²¹ The previously proposed methods have largely focused on point exposure studies (e.g., intent-to-treat analysis of randomized trials) for estimating the data-generating process; but they have not been motivated by the objective of estimating causal effects of time-varying exposures using the potential outcomes framework.

To tackle this problem in which we treat the longitudinal outcome as the focus and death as a nuisance, let $S_{ij} = 1$ indicate that participant i was alive at the time of scheduled visit j , and $S_{ij} = 0$ indicate that participant i was dead at the time of scheduled visit j . Therefore, $S_{ij} = 0$ implies that $S_{ik} = 0$ for $k > j$, and $S_{ij} = 1$ implies that $S_{ik} = 1$ for $k < j$. Similarly, let $R_{ij} = 1$ indicate that participant i responded (i.e., provided data) at scheduled visit j , and $R_{ij} = 0$ indicate that participant i did not respond at scheduled visit j . If missingness is monotone; e.g., where participants are permanently lost to follow-up if they miss a visit, then $R_{ij} = 0$ implies that $R_{ik} = 0$ for $k > j$, and $R_{ij} = 1$ implies that $R_{ik} = 1$ for $k < j$.

Let p_{ij}^S denote $E[S_{ij} | S_{i,j-1} = 1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i] = E[S_{ij} | S_{i,j-1} = 1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i]$. Consider the following model for the death process:

$$\text{logit}(p_{ij}^S) = \gamma_0 + \mathbf{X}_i^t \boldsymbol{\gamma}_1 + \gamma_2 Y_{i,j-1} + \bar{A}_{i,j-1}^t \boldsymbol{\gamma}_3 + \gamma_4 X_{i1} Y_{i,j-1} + \gamma_5 j + \gamma_6 b_{i0} + \gamma_7 b_{i1}, \quad (11)$$

where S_{ij} follows a Bernoulli distribution,

$$f(S_{ij} | S_{i,j-1} = 1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\gamma}) = (p_{ij}^S)^{S_{ij}} (1 - p_{ij}^S)^{(1-S_{ij})}.$$

Model (11) implies that among participants who survived to visit $j-1$, survival to visit j is statistically independent of A_{ij} and Y_{ij} given the observed histories and random effects.

Similarly, let p_{ij}^R denote $E[R_{ij} | S_{ij} = 1, R_{i,j-1} = 1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i] = E[R_{ij} | S_{ij} = 1, R_{i,j-1} = 1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i]$. Consider the following model for a monotone response process among survivors

$$\text{logit}(p_{ij}^R) = \delta_0 + \mathbf{X}_i^t \boldsymbol{\delta}_1 + \delta_2 Y_{i,j-1} + \bar{A}_{i,j-1}^t \boldsymbol{\delta}_3 + \delta_4 X_{i1} Y_{i,j-1} + \delta_5 j + \delta_6 b_{i0} + \delta_7 b_{i1}, \quad (12)$$

where R_{ij} follows a Bernoulli distribution,

$$f(R_{ij} | S_{ij} = 1, R_{i,j-1} = 1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\delta}) = (p_{ij}^R)^{R_{ij}} (1 - p_{ij}^R)^{(1-R_{ij})}.$$

Model (12) implies that, among survivors who have thus far remained in the study, response at visit j is statistically independent of A_{ij} and Y_{ij} given the observed histories and random effects. Model (12) includes a missing not at random mechanism, because R_{ij} may depend on possibly missing Y_{ij} after integrating over $\mathbf{b}_i$.¹⁵ that is, in general missingness is nonignorable. If $\delta_6 = \delta_7 = 0$, then the likelihood factors and the missingness mechanism simplifies to missing at random;¹⁵ that is, missingness is ignorable.

We can estimate the parameters of the data-generating process by extending model (5) to include the death and response processes in the joint likelihood:

$$\begin{aligned}
& \prod_i \int f(\bar{Y}_{iJ_i}, \bar{R}_{iJ_i}, \bar{S}_{iJ_i}, \bar{A}_{iJ_i-1} | \mathbf{X}_i, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}) f(\mathbf{b}_i; \mathbf{D}) d\mathbf{b}_i \\
&= \prod_i \int \prod_j f(Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) \times f(R_{ij} | S_{ij}=1, R_{i,j-1}=1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\delta}) \\
&\times f(S_{ij} | S_{i,j-1}=1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\gamma}) f(A_{i,j-1} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}, \mathbf{b}_i; \boldsymbol{\alpha}) f(\mathbf{b}_i; \mathbf{D}) d\mathbf{b}_i.
\end{aligned}
\tag{13}$$

Models (11) and (12) imply

$$\begin{aligned}
& f(Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) \\
&= f(Y_{ij} | S_{ij}=1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) \\
&= f(Y_{ij} | R_{ij}) \\
&= 1, S_{ij} \\
&= 1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2).
\end{aligned}$$

Maximizing the likelihood (13) estimates $\boldsymbol{\beta}$, however, once again, Equation (3) does not have a causal interpretation. Furthermore, defining causal effects in the presence of truncation by death is particularly challenging, because death renders missing outcomes to be undefined, whereas nonresponse simply renders them unobserved. Tchetgen Tchetgen *et al*²² and Shardell *et al*²³ showed that, in general, g-computation in the presence of truncation by death can identify a causal effect that is a weighted sum of principal strata (PS) direct effects and indirect effects.²⁴ That is, it identifies a causal contrast of exposure history $j-1$ versus $\bar{a}'_{j-1}$ on the outcome at visit j within the latent subgroup who would have survived to visit j whether they followed exposure history $j-1$ or $\bar{a}'_{j-1}$. Unlike in previous work on causal inference with truncation by death,^{22, 23} we consider unmeasured common causes of the outcome, missingness, and death in conjunction with possible unmeasured confounding. G-computation is performed as in Equation (8) by conditioning on survival,

$$\begin{aligned}
& g_j(\bar{a}_{j-1} | \mathbf{X}_i, \mathbf{b}_i) \\
&= \int E[Y_{ij} | R_{ij}=1, S_{ij}=1, \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}=\bar{a}_{j-1}, \mathbf{b}_i] \\
&\times \prod_{k=0}^{j-1} f(Y_{ik} | R_{ik}=1, S_{ik}=1, \mathbf{X}_i, \bar{Y}_{i,k-1}, \bar{A}_{i,k-1}=\bar{a}_{k-1}, \mathbf{b}_i) dY_{ik}.
\end{aligned}
\tag{14}$$

Whether contrasts $g_j(\bar{a}_{j-1} | \mathbf{X}_i, \mathbf{b}_i) - g_j(\bar{a}'_{j-1} | \mathbf{X}_i, \mathbf{b}_i)$ computed from Equation (14) have a causal interpretation depends on whether an unidentifiable assumption is satisfied. Namely,

$$Y_{ij}(\bar{a}_{j-1}, \bar{y}_{j-1}^*) \perp S_{ij}(\bar{a}'_{j-1}, \bar{y}'_{j-1}) | S_{ij}(\bar{a}_{j-1}, \bar{y}_{j-1}^*), \mathbf{X}_i, Y_{i0}, \quad (15)$$

where $\bar{y}_{j-1}^* = \{y_1, \dots, y_{j-1}\}$, i.e., post-exposure outcome history through visit $j-1$, which are the outcomes at all past visits except at $j=0$; and $Y_{ij}(\bar{a}_{j-1}, \bar{y}_{j-1}^*)$ and $S_{ij}(\bar{a}_{j-1}, \bar{y}_{j-1}^*)$ are counterfactual outcome and survival, respectively, at visit j of exposure history and post-exposure outcome history through visit $j-1$. Assumption (15) is a “cross-world” independence assumption that depends on counterfactuals of $\bar{y}_{j-1}$ and variables that mediate the effect of $\bar{y}_{j-1}$ on Y_{ij} and S_{ij} . The assumption implies that the counterfactual outcome set to one joint exposure and outcome history is conditionally independent of survival set to another joint exposure and outcome history.

Using the arguments of Tchetgen Tchetgen *et al*²² and Shardell *et al*,²³ the causal contrast at $j=1$, $g_1(a_0 | \mathbf{X}_i, \mathbf{b}_i) - g_1(a'_0 | \mathbf{X}_i, \mathbf{b}_i)$, equals

$$\int E[Y_{i1}(a_0) - Y_{i1}(a'_0) | S_{i1}(a_0) = S_{i1}(a'_0) = 1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i] f(Y_{i0} | \mathbf{X}_i, \mathbf{b}_i) dY_{i0}. \quad (16)$$

The causal contrast $E[Y_{i1}(a_0) - Y_{i1}(a'_0) | S_{i1}(a_0) = S_{i1}(a'_0) = 1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i]$ is a PS effect of the exposure within the latent subgroup who would have survived if A_0 were set to a_0 or a'_0 , conditioned on $\{Y_{i0}, \mathbf{X}_i, \mathbf{b}_i\}$.²⁴ Thus, the causal contrast in Equation (16) is a weighted average of PS effects.

The causal contrast at $j=2$, $g_2(\bar{a}_1 | \mathbf{X}_i, \mathbf{b}_i) - g_2(\bar{a}'_1 | \mathbf{X}_i, \mathbf{b}_i)$ is more complicated owing to a possible $A_0 \rightarrow Y_1 \rightarrow A_1 \rightarrow Y_2$ pathway. Specifically, it equals^{22, 23}

$$\begin{aligned} & \int E[Y_{i2}(\bar{a}_1, y_1) - Y_{i2}(\bar{a}'_1, y_1) | S_{i2}(\bar{a}_1, y_1) = S_{i2}(\bar{a}'_1, y_1) = 1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i] \\ & \times f_{Y_{i1}(a'_0)}(Y_{i1}=y_1 | S_{i1}(a_0)=1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i) dy_1 \times f(Y_{i0} | \mathbf{X}_i, \mathbf{b}_i) dY_{i0} \\ & + \int E[Y_{i2}(\bar{a}_1, y_1) - Y_{i2}(\bar{a}_1, y'_1) | S_{i2}(\bar{a}_1, y_1) = S_{i2}(\bar{a}_1, y'_1) = 1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i] \\ & \times [f_{Y_{i1}(a_0)}(Y_{i1}=y_1 | S_{i1}(a_0)=1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i) - f_{Y_{i1}(a'_0)}(Y_{i1}=y_1 | S_{i1}(a'_0)=1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i)] dy_1 \times f(Y_{i0} | \mathbf{X}_i, \mathbf{b}_i) dY_{i0}. \\ & = \text{Direct Effect} + \text{Indirect Effect} \end{aligned} \quad (17)$$

The Direct Effect in Equation (17) is a weighted average of the PS-controlled direct effect $E[Y_{i2}(\bar{a}_1, y_1) - Y_{i2}(\bar{a}'_1, y_1) | S_{i2}(\bar{a}_1, y_1) = S_{i2}(\bar{a}'_1, y_1) = 1, Y_{i0}, \mathbf{X}_i, \mathbf{b}_i]$ averaged over the distribution of $\{Y_{i1}(a'_0), Y_{i0}\}$. The Indirect Effect is the difference in average PS effects of

Y_{i1} on Y_{i2} setting a_{i1} to a_{i1}^* versus $\bar{a}_1^*$, comparing the weighted average over the distributions of $Y_{i1}(a_0)$ versus $Y_{i1}(a_0^*)$.

Equation (17) differs from the target of estimation for most methods that jointly model longitudinal and mortality data. Most methods that focus on the longitudinal outcome aim to estimate the data-generating process of the longitudinal outcome if there were no deaths. In contrast, we aim to estimate causal effects, and Equation (17) is the causal effect identified by g-computation in the presence of truncation by death.

Slight variations are needed in studies with non-monotone missingness. In particular, Equation (12) would not require conditioning on $R_{i,j-1} = 1$, but would involve conditioning on the vector $(R_{i0}, R_{i1}, \dots, R_{i,j-1})$ as a way to model the joint distribution of $\mathbf{R}_i = (R_{i0}, R_{i1}, \dots, R_{i,j})$. Similarly, the probability for R_{ij} in Equation (13) would not condition on $R_{i,j-1} = 1$ for the same reason. The use of JMM is advantageous here because it allows integration over missing data with relative ease.^{25, 26}

2.3.2. Additional Time-Dependent Confounding—We have heretofore only considered time-dependent confounding due to serial correlation. However, other endogenous time-varying covariates may also confound the effect of X_i on Y_i , resulting in time-dependent confounding. That is, some factors may affect the exposure and the outcome at one time point, but may be affected by the exposure, directly or indirectly, at an earlier time point.^{1, 20}

Let $\mathbf{L}_{ik}$ be a set of time-dependent confounders at visit k with density $f(\mathbf{L}_{ik} | \mathbf{X}_i, \bar{Y}_{i,k-1}, \bar{L}_{i,k-1}, \dots, \bar{L}_{i,k-1}, \mathbf{b}_i; \phi)$, $k = 0, \dots, J_i - 1$, indexed by parameters ϕ . If $\mathbf{L}_{ik}$ is independent of $\mathbf{b}_i$ for all k , then the likelihood for $\mathbf{L}_i = \{\mathbf{L}_{i0}, \dots, \mathbf{L}_{i,J_i-1}\}$ factors from that of $\{\bar{Y}_i, \bar{L}_i\}$ and can be maximized separately. However, if $\mathbf{L}_i$ shares random effects with $\{\bar{Y}_i, \bar{L}_i\}$, then the likelihood (5) needs to be extended to include ϕ .

$$\begin{aligned} & \prod_i \int f(\bar{Y}_{i,J_i}, \bar{A}_{i,J_i-1}, \bar{L}_{i,J_i-1} | \mathbf{X}_i, \mathbf{b}_i; \beta, \sigma^2, \alpha, \phi) f(\mathbf{b}_i; D) d\mathbf{b}_i \\ &= \prod_i \int \prod_j f(Y_{ij} | \mathbf{X}_i, \bar{L}_{i,j-1}, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \beta, \sigma^2) f(A_{i,j-1} | \mathbf{X}_i, \bar{L}_{i,j-1}, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}, \mathbf{b}_i; \alpha) \\ & \quad \times f(\mathbf{L}_{i,j-1} | \mathbf{X}_i, \bar{L}_{i,j-2}, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}, \mathbf{b}_i; \phi) f(\mathbf{b}_i; D) d\mathbf{b}_i. \end{aligned}$$

(18)

Equation (18) shows that $\bar{L}_{i,j-1}$ may have an effect on Y_{ij} and $A_{i,j-1}$. This phenomenon can be accommodated by including components of $\mathbf{L}_i$ on the right-hand side of Equations (3) and (4). To identify causal effects, g-computation involves integrating over $\bar{L}_i$ (e.g., see Daniel et al²⁰),

$$g_j(\bar{a}_{j-1}|\mathbf{X}_i, \mathbf{b}_i) = \int E \left[Y_{ij} | \mathbf{X}_i, \bar{\mathbf{L}}_{i,j-1}, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1} = \bar{a}_{j-1}, \mathbf{b}_i \right] \\ \times \prod_{k=0}^{j-1} f(\mathbf{L}_{ik} | \mathbf{X}_i, \bar{\mathbf{L}}_{i,k-1}, \bar{Y}_{ik}, \bar{A}_{i,k-1} = \bar{a}_{k-1}, \mathbf{b}_i) d\mathbf{L}_{ik} f(Y_{ik} | \mathbf{X}_i, \bar{\mathbf{L}}_{i,k-1}, \bar{Y}_{i,k-1}, \bar{A}_{i,k-1} = \bar{a}_{k-1}, \mathbf{b}_i) dY_{ik}.$$

(19)

Equation (18) exemplifies the case where variables assessed at some visit k are assumed to be realized in the order Y_k , then L_k , then A_k . Slight variations of Equation (18) are needed for alternative ordering. Unbiased estimation of causal effects requires the model for L to be correctly specified. Exogenous time-varying covariates; that is, covariates that are not influenced (directly or indirectly) by past exposures; need not be modeled.

2.3.3. Process-Specific Random Effects—Thus far, the only random effects that we have considered for the exposure-selection process are those that are shared with the outcome process. Thus, in this model, the assumption of no unmeasured confounders implies the assumption of no exposure-selection heterogeneity. We therefore extend the model to allow separate random effects for the exposure-selection and outcome processes, some of which may be shared between the two processes. For example, let

$$\text{logit}(p_{i,j-1}^A) = \alpha_0 + \mathbf{X}_i^t \boldsymbol{\alpha}_1 + \alpha_2 Y_{i,j-1} + \bar{A}_{i,j-2}^t \boldsymbol{\alpha}_3 + \boldsymbol{\alpha}_4 X_{i1} Y_{i,j-1} + \alpha_5 (j-1) + \alpha_6 b_{i0} + \alpha_7 b_{i1} + \mathbf{V}_i^t \mathbf{g}_i,$$

where $\mathbf{g}_i$ is a vector of random effects for the exposure-selection process and $\mathbf{V}_i$ is the design matrix for the exposure-selection process random effects. Parameter estimation proceeds by maximizing the joint likelihood

$$\prod_i \int f(\bar{Y}_{i,j}, \bar{A}_{i,j-1} | \mathbf{X}_i, \mathbf{b}_i, \mathbf{g}_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}) f(\mathbf{b}_i, \mathbf{g}_i; \mathbf{D}_{\mathbf{b},\mathbf{g}}) d(\mathbf{b}_i, \mathbf{g}_i) \\ = \prod_i \int \prod_j f(Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) f(A_{i,j-1} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}, \mathbf{b}_i, \mathbf{g}_i; \boldsymbol{\alpha}) f(\mathbf{b}_i, \mathbf{g}_i; \mathbf{D}_{\mathbf{b},\mathbf{g}}) d(\mathbf{b}_i, \mathbf{g}_i),$$

(20)

where $\mathbf{D}_{\mathbf{b},\mathbf{g}}$ is the variance-covariance matrix of $\{\mathbf{b}_i, \mathbf{g}_i\}$. A likelihood analogous to Equation (20) has been proposed to handle nonignorable missing data or time-to-event data and has been referred to as a generalized shared-parameter model.^{27, 28} Likelihood (20) has the benefit of accounting for potential heterogeneity in the exposure-selection process, even in the absence of unmeasured confounding. If $\alpha_6 = \alpha_7 = 0$, then Equation (20) factors as

$$\begin{aligned}
& \prod_i \int f(\bar{Y}_{iJ_i}, \bar{A}_{iJ_i-1} | \mathbf{X}_i, \mathbf{b}_i, \mathbf{g}_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}) f(\mathbf{b}_i, \mathbf{g}_i; \mathbf{D}_{\mathbf{b}, \mathbf{g}}) d(\mathbf{b}_i, \mathbf{g}_i) \\
&= \prod_i \int \prod_j f(A_{i,j-1} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-2}, \mathbf{g}_i; \boldsymbol{\alpha}) d(\mathbf{g}_i; \mathbf{D}_{\mathbf{g}}) d\mathbf{g}_i \int \prod_j f(Y_{ij} | \mathbf{X}_i, \bar{Y}_{i,j-1}, \bar{A}_{i,j-1}, \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) f(\mathbf{b}_i; \mathbf{D}_{\mathbf{b}}) d\mathbf{b}_i,
\end{aligned}
\tag{21}$$

where $\mathbf{D}_{\mathbf{g}}$ is the variance-covariance matrix for $\mathbf{g}_i$ and $\mathbf{D}_{\mathbf{b}}$ is the variance-covariance matrix for $\mathbf{b}_i$, since the outcome and exposure-selection processes have no shared random effects (i.e., no unmeasured confounding). Therefore, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be estimated separately. A variation of model (21) has been proposed for mediation analysis in which the mediator and outcome have separate, yet possibly correlated, random effects.⁷

3. Simulation Studies

We performed a series of simulation studies to evaluate the performance of the joint mixed-effects model combined with g-computation to estimate causal effects. We aimed to compare the joint mixed effects model (JMM) with conventional linear mixed-effects models (LMM) combined with g-computation, fixed-effects models (FEM) combined with g-computation, and longitudinal targeted maximum likelihood estimation (LTMLE). Like JMM, LTMLE involves fitting a model for the outcome and a model for the exposure. Unlike JMM, LTMLE is semiparametric, relies on the assumption of no unmeasured confounders, and is doubly robust, which means that it is unbiased if at least one of the models is correctly specified.

We simulated 1,000 datasets of size $n = 1,000$. We fit JMM, LMM, and FEM using SAS version 9.3 with PROC NLMIXED. Numerical integration was performed by Gaussian adaptive quadrature with five quadrature points; optimization was performed via the dual quasi-Newton algorithm using the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm²⁹ to update the Cholesky factor of the Hessian matrix. LTMLE was fit using the ltmle package in R Statistical Software.³⁰

We performed three sets of simulation studies. The first set involved unmeasured confounding only; the second set involved unmeasured confounding and a misspecified random effects distribution (JMM and LMM); and the third set involved unmeasured confounding and nonignorable dropout. JMM was implemented using SAS PROC NLMIXED by fitting likelihoods for Y , A , and (for the third set of simulations) R and including a RANDOM statement for random effects. LMM was implemented in PROC NLMIXED by fitting a likelihood for Y (deleting likelihoods for A and R) and including a RANDOM statement for random effects. Lastly, FEM was implemented in PROC NLMIXED by fitting a likelihood for Y without random effects and deleting the corresponding RANDOM statement (and deleting likelihoods for A and R).

For all simulations, we considered $J_i = 2$ follow-up visits, a single baseline covariate, X_i , and binary time-varying exposure, A_{ij} . X_i was generated as a standard normal random variable. The outcome Y_{ij} was generated to have error variance $\sigma^2 = 1$ and mean as in Equation (3) with $\beta_0 = 1$, $\beta_1 = -1$, $\beta_2 = 1$, $\beta_{30} = 1$, $\beta_{31} = 2$, $\beta_{40} = 0$, $\beta_{41} = 0$, and $\beta_5 = -1$. Thus, at $j = 1$, $g(a_0 = 1) - g(a_0 = 0) = \beta_{30} = 1$, and at $j = 2$, $g(\cdot) - g(\{0, 0\}) = a_0(\beta_{30} + \beta_2\beta_{30}) + a_1\beta_{31} = 2a_0 + 2a_1$.

We performed studies with random intercept only (i.e., $b_{i1} = 0$ for all participants, hence $\text{Var}(b_{i1}) = 0.0$) and studies with both random intercept and slope. In the exposure-selection model, $A_{i,j-1}$ was generated as a Bernoulli random variable to have mean as in Equation (4) with $\alpha_0 = 0$, $\alpha_1 = 0.5$, $\alpha_2 = -0.75$, $\alpha_3 = 1.0$, $\alpha_4 = 0.25$, and $\alpha_5 = 0$. For studies with random intercept only, we set $\alpha_6 = 0$ or 1 and $\alpha_7 = 0$. For studies with a random slope and intercept, we set $\alpha_6 = 0$ and $\alpha_7 = 0$ or 1.

In the third set of simulations, we additionally simulated R_{ij} , conditional on $R_{i,j-1} = 1$ (where $R_{i0} = 1$ for all) as a Bernoulli random variable to have mean as in Equation (12) with $\delta_0 = 1.25$, $\delta_1 = 0.5$, $\delta_2 = 0.5$, $\delta_{30} = 0.5$, $\delta_{31} = 0.5$, $\delta_4 = 0.25$, and $\delta_5 = -0.25$. For studies with random intercept only, we set $\delta_6 = 0$ or 0.5 and $\delta_7 = 0$. For studies with a random slope and intercept, we set $\delta_6 = 0$ and $\delta_7 = 0$ or 0.5. When $R_{ij} = 0$, Y_{ik} and A_{ik} were set to missing for $k \neq j$.

We set $\text{Var}(b_{i0}) = 1$. For studies that additionally had a random slope, we set $\text{Var}(b_{i1}) = 0.3$ or 1 and $\text{Corr}(b_{i0}, b_{i1}) = 0.5$. In the first and third set of simulations, b_{i0} was simulated from a normal distribution, and (b_{i0}, b_{i1}) were simulated from a bivariate normal distribution. In the second set of simulations, b_{i0} was simulated from an exponential distribution and centered to have mean 0; (b_{i0}, b_{i1}) were simulated from a bivariate exponential distribution with correlation 0.5 and centered to have mean 0.³¹

Table 1 displays the simulation results in scenarios with only unmeasured confounding and normally distributed random effects. In the absence of random slope ($\text{Var}(b_{i1})=0$) and unmeasured confounding ($\alpha_6 = 0$); JMM, LMM, and LTMLE produced unbiased estimates; but LMM was the most efficient and LTMLE was the least efficient. In contrast, using FEM and failing to account for heterogeneity resulted in non-negligible bias. These findings are consistent with expectations given that JMM, LMM, and LTMLE account for within-person clustering whereas FEM does not. In the absence of random slope ($\text{Var}(b_{i1})=0$) and the presence of unmeasured confounding ($\alpha_6 = 1$), JMM remained unbiased whereas LMM, FEM, and LTMLE all demonstrated bias exceeding 10% for all causal effects. These findings are also consistent with expectations because JMM was the only method that does not rely on the assumption of no unmeasured confounding.

In the presence of a random slope ($\text{Var}(b_{i1}) \in \{0.3, 1.0\}$), Table 1 shows that JMM produced unbiased estimates regardless of the values of α_7 and $\text{Var}(b_{i1})$, and results from LMM were unbiased in the absence of unmeasured confounding ($\alpha_7 = 0$). However, LMM demonstrated bias exceeding 10% for at least one causal effect in the presence of unmeasured confounding ($\alpha_7 = 1$), where larger biases occurred when $\text{Var}(b_{i1}) = 1.0$ than when $\text{Var}(b_{i1}) = 0.3$, which reflects the strength of unmeasured confounding. FEM exhibited bias exceeding 10% for at

least one causal effect for all scenarios. LTMLE was unbiased when $\text{Var}(b_{11}) = 0.3$ regardless of α_7 ; however, LTMLE exhibited some (2.0% to 10%) bias when $\text{Var}(b_{11}) = 1.0$ regardless of α_7 . JMM was expected to be unbiased irrespective of the presence of unmeasured confounding because it was the only method that does not rely on the assumption of no unmeasured confounding. However, LTMLE was more robust to unmeasured confounding than was LMM, which may reflect the robustness of empirical models of exposure over parametric assumptions of outcome models in this case. Once again, LTMLE was the least efficient method, and LMM tended to be somewhat more efficient than JMM. Non-convergence was rare (0% to 0.3%) across all methods.

Table 2 displays the simulation results in scenarios with unmeasured confounding and exponentially distributed random effects. In this case, JMM and LMM were fit by misspecifying the random effects distribution as normal (random intercept only) or bivariate normal (random intercept and slope). In the absence of random slope ($\text{Var}(b_{11})=0$) and unmeasured confounding ($\alpha_6 = 0$), LTMLE was unbiased and LMM showed little bias whereas FEM had 7.3% to 19.1% bias and JMM estimates all had bias exceeding 10%. In contrast, in the presence of unmeasured confounding ($\alpha_6 = 1$), JMM was unbiased whereas all other methods' estimates had bias exceeding 10%.

In the presence of a random slope ($\text{Var}(b_{11}) \in \{0.3, 1.0\}$), Table 2 shows that JMM produced unbiased estimates in the presence of unmeasured confounding ($\alpha_7 = 1$) and small (1.6% to 7.9%) bias in the absence of unmeasured confounding ($\alpha_7 = 0$). LMM produced estimates with small bias when $\text{Var}(b_{11}) = 1.0$ in the absence of unmeasured confounding ($\alpha_7 = 0$), estimates with small (2.9% to 6.2%) bias when $\text{Var}(b_{11}) = 0.3$ in the absence of unmeasured confounding ($\alpha_7 = 0$), and at least one estimate exceeding 10% bias in the presence of unmeasured confounding ($\alpha_7 = 1$). FEM exhibited bias exceeding 10% for at least one causal effect for all scenarios. LTMLE displayed small to no bias in all scenarios with its largest (5.6% to 9.6%) bias when $\text{Var}(b_{11}) = 1.0$ in the presence of unmeasured confounding ($\alpha_7 = 1$). These findings reflect the robustness of LTMLE and the sensitivity of parametric methods to the outcome distribution in the absence of unmeasured confounding, and that the benefit of accommodating unmeasured confounding may overcome the limitation of misspecifying the random effect distribution. Once again, LTMLE was the least efficient estimation method, and LMM tended to be somewhat more efficient than JMM. Non-convergence was frequent (8.8% to 23.4%) when $\text{Var}(b_{11}) = 1.0$ in JMM or LMM.

Table 3 displays the simulation results in scenarios with unmeasured confounding, nonignorable dropout, and normally distributed random effects. In the absence of random slope ($\text{Var}(b_{11})=0$), and when dropout was ignorable ($\delta_6 = 0$), JMM performed well whether unmeasured confounding was present ($\alpha_6 = 1$) or absent ($\alpha_6 = 0$); LMM performed well when unmeasured confounding was absent ($\alpha_6 = 0$), but not present ($\alpha_6 = 1$); and FEM and LTMLE were biased whether unmeasured confounding was absent ($\alpha_6 = 0$) or present ($\alpha_6 = 1$). This pattern was largely repeated when dropout was nonignorable ($\delta_6 = 0.5$). LTMLE tended to be the least efficient method across all specifications, and LMM was more efficient than JMM.

In the presence of a random slope, Table 3 shows that JMM was unbiased regardless of whether unmeasured confounding was present ($\alpha_7 = 1$) or absent ($\alpha_7 = 0$), or whether dropout was nonignorable ($\delta_7 = 0.5$) or not ($\delta_7 = 0$). LMM was unbiased in the absence of unmeasured confounding ($\alpha_7 = 0$) regardless of the dropout mechanism (δ_7), but LMM produced at least one estimate with bias exceeding 10% in the presence of unmeasured confounding ($\alpha_7 = 1$). LTMLE produced small bias when $\text{Var}(b_{i1}) = 0.3$ regardless of unmeasured confounding (α_7) and the dropout mechanism (δ_7); however, LTMLE produced some estimates that had at least 10% bias when $\text{Var}(b_{i1}) = 1.0$ in the presence of unmeasured confounding ($\alpha_7 = 1$) or nonignorable dropout ($\delta_7 = 0.5$). FEM exhibited bias exceeding 10% for at least one causal effect for all scenarios. LTMLE was the least efficient method across all specifications, and LMM was more efficient than JMM. Non-convergence ranged from 2.5% to 28.7% in JMM and was higher in random intercept models than in models with random both a random intercept and slope.

In summary, no single method outperformed all other methods in every scenario studied; however, all methods outperformed FEM. JMM performed well except when estimating random intercept models with a misspecified random effect distribution in the absence of unmeasured confounding, which reflects the sensitivity to parametric assumptions. In contrast, LTMLE performed particularly well in this scenario, which reflects the fact that it does not rely on parametric assumptions and the assumption of no unmeasured confounders was satisfied. Not surprisingly, LTMLE uniformly outperformed FEM, although both methods omitted the random effects thereby misspecifying the outcome model. The difference was that LTMLE was doubly robust; it estimated the exposure-selection mechanism and, where relevant, the dropout models. Therefore, in cases where the random effect did not affect the exposure mechanism and where dropout was ignorable, LTMLE correctly specified these models. LTMLE also demonstrated small bias when the random slope explained a small portion of the outcome variance ($\text{Var}(b_{i1}) = 0.3$), which reflected a moderate misspecification of the outcome model. The limitation of LTMLE is that it was the least efficient method studied. Interestingly, LMM performed as well as JMM, and with greater efficiency, in the absence of unmeasured confounding even when drop-out was nonignorable. In general, dropout may induce a type of bias referred to as collider-stratification bias.³² This bias occurs when the exposure (A) and random effect (b) both affect the dropout mechanism (R) due to restricting analysis on observations that are not lost to follow-up. That is, arrows in a DAG from A and b collide at R : $A \rightarrow R \leftarrow b$. This restriction induces an association between the exposure and random effect resulting in unmeasured confounding. Hernan et al³² demonstrated that this bias does not occur when A and b have a multiplicative effect on the cumulative probability that $R = 1$. In the simulation herein, A and b had a multiplicative effect on the conditional odds that $R = 1$; thus, except in the presence of very high dropout rates, bias from nonignorable dropout in LMM was expected to be small. This phenomenon was also empirically demonstrated by a recently published set of simulation studies.³³ Lastly, an additional caveat of JMM is its potential for non-convergence.

4. Data Application

We applied the proposed joint mixed-effects model to a study of circulating vitamin D and depressive symptoms among participants enrolled in “Aging in Chianti” (InCHIANTI), a study of older adults residing in the Chianti region of Italy (near Tuscany).³⁴ Specifically, we estimated the effect of serum concentrations of 25-hydroxyvitamin D [25(OH)D], the gold-standard measure of vitamin D body stores, on severity of depressive symptoms in community-dwelling older adults.³⁵

The analysis included 1,203 participants from the InCHIANTI study. Study visits in InCHIANTI were scheduled to be approximately three years apart. In this analysis, we included as exposure 25(OH)D measured at baseline and at the first and second follow-up visits; we included as outcome depressive symptoms assessed at the first, second, and third follow-up visits. Depressive symptoms were measured using the Center for Epidemiologic Studies Depression (CES-D) Scale,³⁶ a measurement scale comprising 20 items resulting in a score ranging from 0 to 60, with higher scores indicating more severe depressive symptoms. We dichotomized 25(OH)D at 20 ng/mL. We regressed CES-D at time j ($j=1, 2$, or 3) on CES-D at time $j-1$, 25(OH)D at time 0 through $j-1$, random slope (models with an additional random intercept did not converge), and covariates. Covariates included participant demographics, lifestyle factors, season of blood collection, body mass index, renal function (creatinine clearance), and comorbid conditions. We jointly logistically regressed 25(OH)D (≥ 20 ng/mL versus <20 ng/mL) at time $j-1$ on 25(OH)D at time 0 through $j-2$, CES-D at time $j-1$, the random slope, and covariates. Lastly, we logistically regressed missingness (yes versus no) at time j on CES-D at time $j-1$, baseline 25(OH)D, covariates, and the random slope. Missingness was not monotone in this study, so CES-D at time $j-1$ was multiplied by $R_{i,j-1}$ in the missing-data model.

We estimated causal mean differences using JMM, LMM, and FEM; all combined with g-computation; and using LTMLE. We computed structural distributed lags by the joint causal effects of 25(OH)D at times $j-1$, $j-2$, and $j-3$ on depressive symptoms at time j . We refer to the effects of 25(OH)D at times $j-1$, $j-2$, and $j-3$ as the lag 1, lag 2, and lag 3 effects, respectively. The estimated distributed lags are interpreted as the causal effect of setting 25(OH)D ≥ 20 ng/mL at all three time points versus setting 25(OH)D <20 ng/mL at all three time points on CES-D.

Table 4 displays participant characteristics by time (study visit). The mean (standard deviation) age was 74.9 (7.9) years, and 56.7% of participants were women. By the last study visit, over half of the participants were missing or dead.

Table 5 shows the estimated causal mean differences of 25(OH)D on depressive symptoms. All methods (JMM, LMM, FEM, and LTMLE) led to the same qualitative conclusion that the joint effect of 25(OH)D ≥ 20 ng/mL at all three time points causes less severe depressive symptoms than 25(OH)D < 20 ng/mL; although JMM and FEM were the only methods for which the 95% confidence interval excluded 0. JMM estimated the largest magnitude structural distributed lag (-1.80 ; 95% confidence interval, $-3.42, -0.18$); LTMLE estimated the smallest magnitude structural distributed lag (-1.01 ; 95% confidence interval, -3.00 ,

0.98) and had the widest confidence intervals. JMM, LMM, and FEM all estimated that lag 1 25(OH)D had the strongest causal effect and lag 3 25(OH)D had the weakest causal effect. Surprisingly, LTMLE estimated that lag 3 25(OH)D 20 ng/mL caused more severe depressive symptoms.

JMM estimated the coefficient of the random slope as 2.51 (95% confidence interval, -0.73, 5.75) for 25(OH)D and -11.88 (95% confidence interval, -27.31, 3.55) for missingness. Furthermore, the estimated variance of the random slope was 0.046 (95% confidence interval, 0.003, 0.613). After accounting for measured factors, including lagged depressive symptoms, unmeasured sources of heterogeneity that impact the effect of 25(OH)D on depressive symptoms may be relatively small (i.e., very little unmeasured confounding); thus the generally concordant results between JMM, LMM, and FEM. the difference between LTMLE and the other three methods may be explained, in part, by the fact that LTMLE tends to have greater variability than the other methods that may reflect sensitivity to large inverse probability of exposure weights, and it is less dependent on correct specification of the outcome model than the other methods. Lastly, unlike the other three methods, LTMLE constrains the estimated potential outcomes to be in the observed range of the observed outcomes.

5. Discussion

Causal effects are a common target of estimation in observational longitudinal studies. However, such studies have the potential for unmeasured confounding and time-dependent confounding. Combining joint mixed-effects models with g-computation is an attractive approach for simultaneously addressing these challenges.

A particular advantage of the joint mixed-effects model is that it performs well in most situations with either the presence or absence of unmeasured confounding. The method's accessibility is enhanced because the model can be estimated using off-the-shelf statistical packages, such as SAS PROC NL MIXED (see Appendix for sample code). Since the joint mixed-effects model relies on neither the assumption of no unmeasured confounding, in contrast with conventional g-computation and marginal structural models, nor the availability of an unconfounded variable, in contrast with instrumental variable methods, the approach can expand the scenarios within which unbiased estimation of causal effects can occur.

A caveat of the joint mixed-effects model is that unbiased estimation depends on correct specification of the outcome data-generating process, the exposure-selection process, and the random-effects distribution.^{37–39} Robustness to misspecification in the random-effects distribution in the context of joint models has been reported.⁴⁰ In our simulation study, JMM performed well with a misspecified random effect distribution except with random intercept only models in the absence of unmeasured confounding. In this case, LTMLE performed the best of all four methods assessed. Furthermore, as with any g-computation, estimation of causal effects can be computationally intensive in the presence of many time-varying confounders. The joint-mixed effects model addresses unmeasured confounding via random effects, which serve as proxies for the unmeasured confounders. However, the

random effects may not capture unmeasured *time-varying* confounders. Despite these caveats, the proposed method may be particularly beneficial when a known strong confounder is not measured and there is sufficient scientific background knowledge to inform the outcome and exposure-selection models. This scenario is plausible when analyzing data from large cohort studies, because typically not all hypotheses are specified at the study design stage and not all information is collected in order to reduce participant burden and to satisfy budget constraints.

A practical way to balance the strengths and limitations of JMM with those of other methods is to perform multiple methods as part of a sensitivity analysis. In particular, JMM was motivated by studies where important strong confounders were not measured, as exemplified by the analysis of vitamin D and depressive symptoms in InCHIANTI, where sunlight exposure was not measured. However, measured confounders may be omitted resulting in misspecification. For example, if confounder X is modeled, but transformed X should have also been modeled, then this omission is a type of unmeasured (more accurately, omitted) confounding. Although the random effects may capture this omission, we recommend researchers perform extensive exploratory analysis, informed by scientific background knowledge, using the measured covariates to identify strong confounders and appropriate functional forms to help build the model. Prioritizing modeling known strong measured confounders may help reduce bias without a major sacrifice in efficiency. Bias from omitted or mis-modeled weak confounders may be partially overcome by the JMM, and such bias may be small relative to the gains in efficiency from omitting some functions of weak confounders. Given the strong parametric assumptions of JMM, we recommend that researchers perform sensitivity analysis using flexible methods that rely on different assumptions. For example, LTMLE is a flexible method that can help researchers relax parametric assumptions and consider more complex functional forms. However, this type of approach relies on the assumption of no unmeasured confounders; therefore, if results differ between the methods, it is not clear if this is due to violating the assumption of no unmeasured confounders in LTMLE or due to misspecifying the random effect distribution or functional form in JMM. Thus, the results of LTMLE can be used to motivate changes to the functional form of JMM or variable transformations to better satisfy JMM assumptions.

The proposed method contributes to a growing body of literature aiming to account for participant heterogeneity while estimating causal effects.^{6, 7} This and other work has mainly focused on heterogeneity in the context of longitudinal studies,^{6, 7} but others have noted that heterogeneity may be due to clustering.⁴¹ Lastly, future extensions of the joint mixed-effects model should consider categorical and count outcomes.

6. Data Accessibility

Computer code used to generate and analyze simulated data is available for readers. Data from the InCHIANTI study is available with permission from the InCHIANTI Publications Committee at www.inchiantistudy.net.

Acknowledgments

Contract/grant sponsor: National Institute on Aging Intramural Research Program

References

1. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
2. van der Laan, MJ., Rose, S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; New York: 2011.
3. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006; 17(4):360–372. [PubMed: 16755261]
4. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods*. 1994; 23:2379–2412.
5. Laird N, Ware J. Random-effect models for longitudinal data. *Biometrics*. 1982; 38(4):963–974. [PubMed: 7168798]
6. Sitlani CM, Heagerty PJ, Blood EA, Tosteson TD. Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Statistics in Medicine*. 2012; 31:1738–1760. [PubMed: 22344923]
7. Bind MC, Vanderweele TJ, Coull BA, Schwartz JD. Causal mediation analysis for longitudinal data with exogenous exposure. *Biostatistics*. 2016; 17:122–134. [PubMed: 26272993]
8. Kennedy EH, Taylor JMG, Schaubel DE, Williams S. The effect of salvage therapy on survival in a longitudinal study with treatment by indication. *Statistics in Medicine*. 2010; 29:2569–2580. [PubMed: 20809480]
9. Small DS, Ten Have TR, Joffe MM, Cheng J. Random effects logistic models for analysing efficacy of a longitudinal randomized treatment with non-adherence. *Statistics in Medicine*. 2006; 25:1981–2007. [PubMed: 16220487]
10. Holmes TH, Zulman DM, Kushida CA. Adjustment for variable adherence under hierarchical structure: instrumental variable modeling through compound residual inclusion. *Medical Care*. in press.
11. Zhang Z, Wang C, Nie L, Soon G. Assessing the heterogeneity of treatment effects via potential outcomes of individual patients. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2013; 62:687–704.
12. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*. 1988; 44:175–188.
13. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics*. 1995; 51:151–168. [PubMed: 7766771]
14. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512.
15. Rubin DB. Inference and missing data (with discussion). *Biometrika*. 1976; 63:581–592.
16. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. *Journal of Epidemiology and Community Health*. 2003; 57:778–783. [PubMed: 14573579]
17. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association*. 2005; 100:322331.doi: 10.1198/016214504000001880
18. Robins JM. Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*. 1992; 79:321–334.
19. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*. 1995; 90:11121121.
20. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. *Statistics in Medicine*. 2013; 32:1584–1618. [PubMed: 23208861]
21. Rizopoulos, D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. New York, NY: Chapman & Hall; 2012.
22. Tchetgen Tchetgen EJ, Glymour MM, Shpitser I, Weuve J. To weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology*. 2012; 23:132–137.

23. Shardell M, Hicks GE, Ferrucci L. Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. *Biostatistics*. 2015; 16:155–168. [PubMed: 24997309]
24. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
25. Tsonaka R, Verbeke G, Lesaffre E. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*. 2009; 65:81–87. [PubMed: 18373713]
26. Tsonaka R, Rizopoulos D, Verbeke G, Lesaffre E. Nonignorable models for intermittently missing categorical longitudinal responses. *Biometrics*. 2010; 66:834–844. [PubMed: 19995352]
27. Creemers A, Hens N, Aerts M, Molenberghs G, Verbeke G, Kenward MG. Generalized shared-parameter models and missingness at random. *Statistical Modelling*. 2011; 11:279–310.
28. Njagi EN, Molenberghs G, Kenward MG, Verbeke G, Rizopoulos D. A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biometrical Journal*. 2014; 56:1001–1015. [PubMed: 24947904]
29. Luenberger, DG., Ye, Y. *Linear and Nonlinear Programming*. 4. New York, NY: Springer; 2016.
30. Ihaka R, Gentleman R. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*. 1996; 5:299–314.
31. Song P. Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*. 2000; 27:305320.
32. Herna MA, Hernández-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004; 15:615–625. [PubMed: 15308962]
33. Mayeda ER, Tchetgen Tchetgen EJ, Power MC, Weuve J, Jacqmin-Gadda H, Marden JR, Vittinghoff E, Keiding N, Glymour MM. A simulation platform for quantifying survival bias: an application to research on determinants of cognitive decline. *American Journal of Epidemiology*. 2016; 184:378–387. [PubMed: 27578690]
34. Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, Guralnik JM. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *Journal of the American Geriatrics Society*. 2000; 48(12):1618–1625. [PubMed: 11129752]
35. Milaneschi Y, Shardell M, Corsi AM, Vazzana R, Bandinelli S, Guralnik JM, Ferrucci L. Serum 25-hydroxyvitamin D and depressive symptoms in older women and men. *Journal of Clinical Endocrinology and Metabolism*. 2010; 95(7):3225–3233. [PubMed: 20444911]
36. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401.
37. Neuhaus J, Hauck W, Kalbfleisch J. The effects of mixture distribution misspecifications when fitting mixed-effects logistic models. *Biometrika*. 1992; 79:755–762.
38. Heagerty P, Kurland B. Misspecified maximum likelihood estimates and generalized linear mixed models. *Biometrika*. 2001; 88:973–985.
39. Komárek A, Lesaffre E. Generalized linear mixed model with a penalized Gaussian mixture as a random-effects distribution. *Computational Statistics and Data Analysis*. 2008; 52:3441–3458.
40. Rizopoulos D, Verbeke G, Molenberghs G. Shared parameter models under random effects misspecification. *Biometrika*. 2008; 95:63–74.
41. Loeys T, Vansteelandt S, Goetghebeur E. Accounting for correlation and compliance in cluster randomized trials. *Statistics in Medicine*. 2001; 20:3753–3767. [PubMed: 11782031]

Appendix: SAS Code and Example Data

We include SAS code using PROC NLMIXED to fit the joint mixed-effects model and perform g-computation. The variable ‘theoutcome’ is a placeholder. Below are example data from four participants. In InCHIANTI, missingness was not necessarily monotone. For example, the participant with ID = 4 has *Y* missing at time 1 and observed at times 2 and 3.

Thus whether this participant contributes data to the likelihood for Y at time 2 or 3 depends on whether Y at time 2 or 3 is regressed on data collected at time 1. Lagged variables were set to 0 for times prior to data collection (time = 0); $X1$ was standardized.

ID	time	Y	Alag1	Ylag1	RYlag1	Alag2	Alag3	Atime0
X1	X2							
1	1	20	0	31	31	0	0	0
1.3875	1							
1	2	.	0	20	20	0	0	0
1.3875	1							
1	3	.	.	.	0	0	0	0
1.3875	1							
2	1	36	0	34	34	0	0	0
-0.1125	1							
2	2	33	1	36	36	0	0	0
-0.1125	1							
2	3	20	1	33	33	1	0	0
-0.1125	1							
3	1	24	0	7	7	0	0	0
-0.9875	0							
3	2	9	1	24	24	0	0	0
-0.9875	0							
3	3	19	1	9	9	1	0	0
-0.9875	0							
4	1	.	1	3	3	0	0	1
-0.7375	0							
4	2	10	.	.	0	1	0	1
-0.7375	0							
4	3	13	0	10	10	.	1	1
-0.7375	0							

```

proc nlmixed data=jmmdata qpoints=5;
parms beta0=14 beta11=0 beta12=1 beta2=0 beta30=0 beta31=0 beta32=0
      beta40=0 beta41=0 beta5=0 logsigmasq=2 /*f(Y) parms */
      alpha0=0 alpha11=0 alpha12=0 alpha2=0 alpha30=0 alpha31=0
      alpha41=0 alpha5=0 alpha7=0 /* f(A) parms */
      delta0=0 delta11=0 delta12=0 delta2=0 delta3=0 delta4=0
      delta5=0 /* f(R) parms */
      logvarb1=0; /*f(b) parms */
/*normal likelihood for Y */
mu = beta0 + beta11*X1 + beta12*X2 + beta2*Ylag1 + beta30*Alag1 +
      beta31*Alag2 + beta32*Alag3 + beta40*X1*Alag1 + beta41*X1*Alag2 +
      beta42*X1*Alag3 + beta5*time + b1*(1 + Alag1 + Alag2 + Alag3)*X1;
py = pdf('NORMAL',Y,mu,exp(logsigmasq));
llhY = log(py);

```

```

if llhY=. then llhY=0;
/* Bernoulli likelihood for A */
logitA=alpha0 + alpha11*X1 + alpha12*X2 + alpha2*Ylag1 + alpha30*Alag2 +
      alpha31*Alag3 + alpha41*X1*Ylag1 + alpha5*time + alpha7*b1;
pA=exp(logitA)/(1 + exp(logitA));
llhA = Alag1*log(pA) + (1-Alag1)*log(1-pA);
if llhA=. then llhA=0;
/* Bernoulli likelihood for (non-monotone) R */
logitR = delta0 + delta11*X1 + delta12*X2 + delta2*Atime0 + delta3*RYlag1 +
      delta4*time + delta5*b1;
pR = exp(logitR)/(1 + exp(logitR));
llhR= R*log(pR) + (1-R)*log(1-pR);
llik= llhY + llhA +llhR;
model theoutcome~general(llik);
/*normal likelihood for b1 */
random b1 ~normal(0,exp(logvarb1)) subject=id;
/* g-computation (X1 centered at 0) */
estimate 'g-computation: lag 1' beta30;
estimate 'g-computation: lag 2' beta2*beta30 + beta31;
estimate 'g-computation: lag 3' beta2*beta2*beta30 + beta2*beta31 + beta32;
estimate 'g-computation: lag 1-3' beta30 + beta2*beta30 + beta31 +
beta2*beta2*beta30 +
      beta2*beta31 + beta32;
ods output ParameterEstimates=ParameterEstimates
      AdditionalEstimates=AdditionalEstimates;
run;

```

Figure 1.
Directed acyclic graph for the joint mixed-effects model conveying unmeasured confounding and time-varying confounding.

Table 1

Simulation study with unmeasured confounding and normally distributed random effects. Simulations include either random intercept only ($\text{Var}(b_{1i}) = 0.0$) or both random slope and intercept ($\text{Var}(b_{1i}) \in \{0.3, 1.0\}$).

$\text{Var}(b_{1i})$	α_6	α_7	Method	$\hat{\beta}_1(a_0 = 1) - \hat{\beta}_1(a_0 = 0)$			$\hat{\beta}_2(a_0 = 1, a_1) - \hat{\beta}_2(a_0 = 0, a_1)$			$\hat{\beta}_2(a_0, a_1 = 1) - \hat{\beta}_2(a_0, a_1 = 0)$		
				Estimate	SE	ESD	Estimate	SE	ESD	Estimate	SE	ESD
0.0	0	0	JMM	0.986	0.119	0.120	1.973	0.253	0.257	1.996	0.108	0.109
			LMM	0.998	0.068	0.066	1.996	0.140	0.136	2.003	0.082	0.083
			FEM	0.858	0.065	0.060	1.998	0.155	0.143	2.303	0.093	0.087
			LTMLE	0.975	0.123	0.130	1.942	0.258	0.253	1.928	0.182	0.197
0.3	1	0	JMM	1.001	0.095	0.094	2.004	0.195	0.194	2.002	0.088	0.088
			LMM	1.351	0.065	0.062	2.756	0.137	0.131	2.236	0.078	0.077
			FEM	1.258	0.063	0.058	2.943	0.148	0.135	2.533	0.088	0.079
			LTMLE	1.443	0.115	0.121	2.862	0.217	0.215	2.242	0.150	0.167
1.0	1	0	JMM	0.997	0.082	0.081	1.996	0.169	0.168	1.996	0.090	0.091
			LMM	1.000	0.075	0.076	2.003	0.155	0.157	1.998	0.087	0.089
			FEM	0.903	0.072	0.076	2.229	0.181	0.194	2.576	0.103	0.106
			LTMLE	0.984	0.142	0.158	2.010	0.303	0.228	1.999	0.204	0.228
1.0	0	1	JMM	0.997	0.081	0.082	1.994	0.165	0.167	2.001	0.086	0.086
			LMM	1.102	0.074	0.075	2.213	0.153	0.155	2.068	0.084	0.085
			FEM	1.027	0.070	0.074	2.549	0.178	0.188	2.637	0.101	0.105
			LTMLE	1.005	0.133	0.149	2.054	0.286	0.328	1.992	0.186	0.201
1.0	1	0	JMM	0.998	0.083	0.083	1.996	0.171	0.171	1.999	0.091	0.091
			LMM	0.996	0.082	0.083	1.990	0.167	0.169	1.997	0.090	0.097
			FEM	0.885	0.082	0.091	2.328	0.220	0.246	2.843	0.118	0.132
			LTMLE	0.926	0.167	0.179	1.960	0.388	0.397	2.057	0.246	0.271
1.0	0	1	JMM	1.000	0.080	0.083	2.001	0.162	0.168	1.996	0.085	0.082
			LMM	1.137	0.079	0.081	2.281	0.160	0.166	2.071	0.085	0.083
			FEM	1.005	0.080	0.091	2.647	0.212	0.241	2.765	0.114	0.118
			LTMLE	0.900	0.140	0.151	1.837	0.327	0.356	1.923	0.220	0.240

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Abbreviations: SE: arithmetic mean of asymptotic standard errors, ESD: empirical standard deviation of estimates, % Non. Conv: % non-convergence (no solution found), JMM: joint mixed-effects model, LMM: linear mixed-effects model, FEM: fixed-effects model, LTMLE: longitudinal targeted maximum likelihood. True $g_1(a_0 = 1, a_1) - g_2(a_0 = 0, a_1) = 2.0$, $g_2(a_0, a_1) = 1$ - $g_2(a_0, a_1 = 0) = 2.0$.

Table 2

Simulation study with unmeasured confounding and exponentially distributed random effects. JMM and LMM misspecified the the random effect distribution as normal. Simulations include either random intercept only ($\text{Var}(b_{1i}) = 0.0$) or both random slope and intercept ($\text{Var}(b_{1i}) \in \{0.3, 1.0\}$).

Var(b_{1i})	α_6	α_7	Method	$\hat{g}_1(a_0 = 1) - \hat{g}_1(a_0 = 0)$			$\hat{g}_2(a_0 = 1, a_1) - \hat{g}_2(a_0 = 0, a_1)$			$\hat{g}_2(a_0, a_1 = 1) - \hat{g}_2(a_0, a_1 = 0)$			
				Estimate	SE	ESD	Estimate	SE	ESD	Estimate	SE	ESD	% Non. Conv.
0.0	0	0	JMM	1.241	0.108	0.126	2.538	0.240	0.281	2.208	0.103	0.114	0.0
			LMM	1.064	0.067	0.065	2.140	0.140	0.134	2.077	0.080	0.086	0.0
			FEM	0.927	0.064	0.059	2.167	0.154	0.142	2.383	0.091	0.094	0.0
			LTMLE	0.997	0.134	0.149	1.959	0.268	0.226	2.058	0.198	0.226	0.0
			JMM	1.037	0.102	0.111	2.080	0.212	0.232	2.046	0.091	0.086	0.0
			LMM	1.346	0.065	0.064	2.749	0.137	0.134	2.255	0.078	0.076	0.0
0.3	0	0	FEM	1.250	0.063	0.058	2.924	0.148	0.142	2.558	0.087	0.082	0.0
			LTMLE	1.485	0.130	0.160	2.877	0.238	0.276	2.258	0.159	0.192	0.0
			JMM	1.070	0.080	0.068	2.159	0.167	0.141	2.070	0.087	0.089	0.6
			LMM	1.055	0.074	0.063	2.125	0.154	0.129	2.059	0.085	0.086	0.0
			FEM	0.896	0.071	0.065	2.201	0.178	0.166	2.526	0.101	0.105	0.0
			LTMLE	1.004	0.148	0.160	2.002	0.302	0.319	2.048	0.193	0.223	0.0
1.0	0	1	JMM	1.029	0.079	0.076	2.065	0.162	0.158	2.037	0.084	0.088	0.7
			LMM	1.122	0.074	0.076	2.258	0.153	0.157	2.094	0.083	0.085	0.0
			FEM	0.999	0.071	0.081	2.481	0.179	0.210	2.587	0.101	0.112	0.0
			LTMLE	1.034	0.152	0.158	2.086	0.290	0.313	2.056	0.191	0.233	0.0
			JMM	1.042	0.083	0.078	2.092	0.171	0.160	2.031	0.089	0.084	13.3
			LMM	1.037	0.081	0.075	2.081	0.167	0.154	2.022	0.088	0.089	8.8
			FEM	0.814	0.083	0.094	2.125	0.219	0.252	2.715	0.118	0.135	0.0
			LTMLE	0.973	0.167	0.184	1.994	0.368	0.396	2.054	0.238	0.279	0.0
			JMM	0.984	0.082	0.083	1.972	0.165	0.171	1.995	0.086	0.094	23.4
			LMM	1.122	0.080	0.084	2.256	0.163	0.174	2.072	0.086	0.095	14.0
			FEM	0.902	0.083	0.102	2.387	0.221	0.262	2.731	0.118	0.128	0.0
			LTMLE	0.904	0.174	0.201	1.884	0.359	0.382	2.073	0.197	0.240	0.0

Abbreviations: SE: arithmetic mean of asymptotic standard errors, ESD: empirical standard deviation of estimates, % Non. Conv: % non-convergence (no solution found), JMM: joint mixed-effects model, LMM: linear mixed-effects model, FEM: fixed-effects model, LTMLE: longitudinal targeted maximum likelihood. True $g_1(a_0 = 1, a_1) - g_2(a_0 = 0, a_1) = 2.0$, $g_2(a_0, a_1 = 1) - g_2(a_0, a_1 = 0) = 2.0$.

Table 3

Simulation study with unmeasured confounding, nonignorable dropout, and normally distributed random effects. Simulations include either random intercept only ($\text{Var}(b_{i1}) = 0.0$) or both random slope and intercept ($\text{Var}(b_{i1}) \in \{0.3, 1.0\}$).

Var (b_{i1})	δ_6	δ_7	α_6	α_7	Method	$\hat{\beta}_1(a_0 = 1) - \hat{\beta}_1(a_0 = 0)$			$\hat{\beta}_2(a_0 = 1, a_1) - \hat{\beta}_2(a_0 = 0, a_1)$			$\hat{\beta}_2(a_0, a_1 = 1) - \hat{\beta}_2(a_0, a_1 = 0)$			% Non. Conv.
						Estimate	SE	ESD	Estimate	SE	ESD	Estimate	SE	ESD	
0.0	0	0	0	0	JMM	1.015	0.148	0.153	2.035	0.323	0.326	1.999	0.130	0.126	14.7
					LMM	0.990	0.079	0.077	1.981	0.166	0.162	1.990	0.101	0.098	0.0
					FEM	0.907	0.078	0.077	2.125	0.189	0.187	2.293	0.116	0.102	0.0
					LTMLE	0.964	0.142	0.154	1.828	0.281	0.325	1.781	0.213	0.238	0.0
					JMM	1.001	0.114	0.127	2.005	0.239	0.269	2.002	0.103	0.118	10.8
					LMM	1.345	0.076	0.066	2.749	0.163	0.141	2.226	0.096	0.102	0.0
	0.5	0	0	0	FEM	1.316	0.075	0.066	3.101	0.181	0.158	2.524	0.110	0.100	0.0
					LTMLE	1.437	0.130	0.139	2.746	0.282	0.286	2.154	0.230	0.249	0.0
					JMM	0.981	0.142	0.164	1.960	0.311	0.347	1.972	0.128	0.139	16.5
					LMM	0.980	0.081	0.084	1.954	0.170	0.169	1.977	0.104	0.099	0.0
					FEM	0.904	0.080	0.080	2.095	0.191	0.188	2.253	0.120	0.106	0.0
					LTMLE	0.950	0.141	0.148	1.776	0.280	0.287	1.770	0.220	0.255	0.0
0.3	0	0	0	0	JMM	0.965	0.117	0.136	1.929	0.245	0.286	1.966	0.107	0.129	28.7
					LMM	1.337	0.077	0.078	2.725	0.166	0.164	2.195	0.099	0.112	0.0
					FEM	1.310	0.076	0.068	3.059	0.182	0.160	2.479	0.112	0.102	0.0
					LTMLE	1.415	0.131	0.147	2.636	0.278	0.303	2.082	0.233	0.256	0.0
					JMM	1.001	0.097	0.099	2.006	0.203	0.207	1.996	0.113	0.104	4.1
					LMM	1.002	0.088	0.087	2.008	0.184	0.178	1.999	0.109	0.101	0.0
	0.5	0	0	0	FEM	1.014	0.086	0.092	2.494	0.218	0.236	2.570	0.128	0.138	0.0
					LTMLE	0.948	0.170	0.200	1.929	0.312	0.376	1.852	0.220	0.270	0.0
					JMM	0.995	0.094	0.104	1.994	0.194	0.223	2.003	0.107	0.101	2.5
					LMM	1.104	0.086	0.088	2.217	0.179	0.189	2.080	0.105	0.097	0.0
					FEM	1.164	0.084	0.088	2.889	0.215	0.235	2.688	0.125	0.108	0.0
					LTMLE	1.024	0.156	0.182	1.998	0.294	0.318	1.878	0.232	0.265	0.0
1.0	0	0	0	0	JMM	1.008	0.096	0.103	2.014	0.200	0.216	1.991	0.113	0.117	4.0
					LMM	1.009	0.088	0.087	2.011	0.183	0.181	1.995	0.109	0.111	0.0

Var (b_{il})	δ_6	δ_7	α_6	α_7	Method	$\hat{g}_1(a_0 = 1) - \hat{g}_1(a_0 = 0)$			$\hat{g}_2(a_0 = 1, a_1) - \hat{g}_2(a_0 = 0, a_1)$			$\hat{g}_2(a_0 = 1, a_1 = 1) - \hat{g}_2(a_0 = 0, a_1 = 1)$			% Non. Conv.
						Estimate	SE	ESD	Estimate	SE	ESD	Estimate	SE	ESD	
1.0	1				FEM	1.022	0.085	0.086	2.505	0.218	0.224	2.562	0.129	0.128	0.0
					LTMLE	0.948	0.160	0.184	1.850	0.282	0.304	1.848	0.217	0.271	0.0
					JMM	1.012	0.094	0.102	2.014	0.195	0.210	1.991	0.108	0.097	7.9
					LMM	1.120	0.086	0.092	2.236	0.179	0.192	2.080	0.105	0.093	0.0
					FEM	1.173	0.084	0.089	2.899	0.217	0.228	2.683	0.126	0.109	0.0
					LTMLE	1.025	0.148	0.173	1.965	0.288	0.359	1.863	0.224	0.256	0.0
	0	0			JMM	0.987	0.097	0.088	1.971	0.201	0.181	1.978	0.116	0.120	1.7
					LMM	0.986	0.095	0.084	1.968	0.196	0.173	1.978	0.115	0.124	0.0
					FEM	1.063	0.097	0.105	2.756	0.259	0.292	2.788	0.147	0.159	0.0
					LTMLE	0.927	0.197	0.231	1.883	0.369	0.454	1.880	0.242	0.328	0.0
					JMM	0.982	0.094	0.094	1.966	0.191	0.194	1.995	0.107	0.107	3.1
					LMM	1.127	0.091	0.096	2.262	0.188	0.200	2.092	0.107	0.109	0.0
0.5	0				FEM	1.264	0.095	0.101	3.326	0.256	0.274	2.864	0.141	0.139	0.0
					LTMLE	0.895	0.182	0.208	1.905	0.331	0.392	1.827	0.278	0.362	0.0
					JMM	1.002	0.098	0.106	1.993	0.202	0.217	1.997	0.116	0.119	8.2
					LMM	1.007	0.095	0.102	1.998	0.197	0.208	2.016	0.115	0.116	0.0
					FEM	1.081	0.098	0.106	2.807	0.264	0.285	2.770	0.150	0.161	0.0
					LTMLE	0.858	0.198	0.223	1.832	0.348	0.422	1.882	0.234	0.322	0.0
	1				JMM	1.014	0.093	0.076	2.012	0.187	0.158	2.001	0.106	0.104	10.4
					LMM	1.154	0.091	0.092	2.291	0.186	0.191	2.106	0.106	0.110	0.0
					FEM	1.269	0.096	0.104	3.350	0.260	0.292	2.821	0.142	0.149	0.0
					LTMLE	0.890	0.202	0.239	1.843	0.328	0.366	1.883	0.247	0.289	0.0

Abbreviations: SE: arithmetic mean of asymptotic standard errors, ESD: empirical standard deviation of estimates, % Non. Conv: % non-convergence (no solution found), JMM: joint mixed-effects model, LMM: linear mixed-effects model, FEM: fixed-effects model, LTMLE: longitudinal targeted maximum likelihood. True $g_1(a_0 = 1, a_1) - g_2(a_0 = 0, a_1) = 2.0$, $g_2(a_0, a_1) = 1$ - $g_2(a_0 = 0, a_1 = 0) = 2.0$.

Table 4

Descriptive Statistics in the InCHIANTI Study (n=1,203)

Characteristic	Visit 0	Visit 1	Visit 2	Visit 3
	Mean (SD) or Number (%)			
CES-D score	12.1 (9.0)	15.7 (8.7)	14.2 (7.7)	14.8 (8.4)
25(OH)D 20 ng/mL	368 (35.0)	615 (74.7)	371 (50.1)	
Female Sex	682 (56.7)	–	–	
Age (y)	74.9 (7.9)	–	–	
Education (y)	5.3 (3.4)	–	–	
Smoker	104 (8.6)	–	–	
Alcohol Consumption(drinks week)	7.2 (10.2)	6.5 (8.8)	5.9 (8.3)	
MMSE score	24.2 (5.3)	23.7 (6.2)	22.8 (6.8)	
Creatinine Clearance (mL/min/1.73 m ²)	71.0 (13.6)	66.9 (15.4)	69.4 (15.9)	
Body Mass Index (kg m ²)	27.5 (3.8)	26.8 (3.8)	27.1 (3.9)	
Calcium Intake (mg/day)	815 (299)	822 (304)	812 (286)	
Renal Disease	629 (52.2)	541 (57.6)	410 (48.9)	
Congestive Heart Failure	59 (4.9)	57 (6.1)	86 (10.2)	
Hypertension	525 (43.6)	283 (30.1)	257 (30.6)	
Missing Alive	0	259 (21.5)	261 (21.7)	244 (20.2)
Dead	0	143 (11.9)	298 (24.8)	435 (36.2)

Abbreviations: SD, standard deviation; ; CES-D, Center for Epidemiologic Studies Depression Scale; 25(OH)D, 25-hydroxyvitamin D; y, years; MMSE, Mini-Mental State Examination score. Missingness at visit j defined as missing CES-D at visit j or missing 25(OH)D or covariates at visit $j - 1$, for $j = 1, 2, 3$. Dashes (–) indicate time-invariant covariates assessed at visit 0.

Table 5

Estimated Causal Effect of Vitamin D on Depressive Symptoms in the InCHIANTI Study (n=1,203)

Method	Parameter	Estimate	95% CI
JMM	Lag 1	-1.37	(-2.06, -0.67)
	Lag 2	-0.31	(-1.27, 0.65)
	Lag 3	-0.12	(-0.50, 0.26)
	Lag 1,2, and 3	-1.80	(-3.42, -0.18)
LMM	Lag 1	-1.35	(-2.04, -0.66)
	Lag 2	-0.16	(-1.11, 0.80)
	Lag 3	-0.06	(-0.44, 0.31)
	Lag 1,2, and 3	-1.57	(-3.18, 0.04)
FEM	Lag 1	-1.19	(-1.88, -0.50)
	Lag 2	-0.40	(-1.35, 0.55)
	Lag 3	-0.16	(-0.54, 0.22)
	Lag 1,2, and 3	-1.75	(-3.35, -0.14)
LTMLE	Lag 1	-1.68	(-4.99, 1.64)
	Lag 2	-1.43	(-3.71, 0.84)
	Lag 3	1.29	(-0.78, 3.37)
	Lag 1,2, and 3	-1.01	(-3.00, 0.98)

Abbreviations: CI, confidence interval; JMM: joint mixed-effects model; LMM: linear mixed-effects model; FEM: fixed-effects model, LTMLE: longitudinal targeted maximum likelihood.