Causal Framework for Treatment Evaluation using Multivariate Generalized Linear Mixed-Effects Models with Longitudinal Data

Yizhen Xu, Jisoo Kim, Ami Shah, Scott Zeger

December 12, 2022

Abstract

Dynamic prediction of causal effects under different treatment regimes conditional on individual's characteristics and longitudinal history is an essential problem in precision medicine. This is challenging in practice because outcomes and treatment assignment mechanisms are unknown in observational studies, an individual's treatment efficacy is a counterfactual, and the existence of selection bias is often unavoidable.

We propose a Bayesian framework for identifying the counterfactual benefits of treatment regimes using Bayesian g-computation [26;41] with multivariate generalized linear mixed effect models. Unmeasured time-invariant factors are identified as subject-specific random effects in the joint distribution of outcomes, time-varying confounders, and treatment assignments. Existing methods mostly focus on balancing the confounder distributions of observations between different treatments. We propose a sequential ignorability assumption conditional on the treatment assignment heterogeneity. This is analogous to balancing the distribution of observable variables as well as the latent tendency toward each treatment due to unmeasured time-invariant factors.

Longitudinal causal inference; latent variable modeling; random effects models; g-computation

1 Introduction

Precision medicine [16;28] is a clinical decision-making process that uses a patient's medical history, current and previous health statuses, and observational data from a large population to make individualized treatment and care recommendations throughout the progression of a disease. For example, Wang et al. [36] predicted individual future biomarker trajectories and major clinical events for improving COVID-19 care and Coley et al. [6] utilized longitudinal biomarker measurements to improve clinical decisions about whether to remove or irradiate a patient's prostate cancer. Studying the heterogeneity in an individual's treatment effect is one of the many questions of interest in precision medicine. This involves mapping current patient information to biomarker trajectories under potential actions such as the selection and timing of therapy. We are particularly interested in using observational data to answer the causal question "what would have happened after τ days if a specific dynamic treatment regime had been implemented, given the patient's history of h days? ", where dynamic treatment regimes are defined as treatment that may change based on observed patient history. We may then determine

which treatment option is the best for a patient by assessing the average treatment effect under different regimes for a subgroup of patients who share similar characteristics and history.

Our motivating application is to study the effectiveness of an immunosuppressant medication, mycophenolate (MMF)[22;38], on scleroderma using clinically observed data from the Johns Hopkins Precision Medicine Analytics Platform (PMAP) Registry. In Scleroderma Lung Study II[34], MMF resulted in improvements in the modified Rodnan skin score (mRSS) among diffuse patients over 24 months. We focus on patients who have demonstrated tolerance to MMF, diffuse or nondiffuse, and compare it to other treatment regimens that do not include MMF in terms of effectiveness on both skin and lung measurements. In this observational study, there are multiple practical challenges: treatment assignment is not randomized based on measured factors, biomarkers are measured irregularly, missingness patterns may be informative about biomarker values, and natural heterogeneity among subjects exists beyond what the observables can explain. In order to tackle these issues, we use a Bayesian approach under the potential outcomes framework[29], which defines causal effect as a comparison of potential outcomes for the same set of subjects under different treatment regimes. The approach has the advantages of being able to handle a large number of structural missingness, incorporating Bayesian models with the flexibility to address complex data, and naturally quantifying uncertainty, all of which are important for decision making in precision medicine.

The primary factor in evaluating treatment efficacy, both in this and many other scenarios of comparing treatment regimes for precision medicine, is subject heterogeneity or unmeasured factors in treatment assignment and biomarker dynamics. Individual treatment decisions are intuitively sensitive to unmeasured variables that may confound disease progression. Often, the practitioner deciding on whether or not and when to treat a patient will have access to private signals about the patient's potential outcomes, such as frailty, willingness to be treated, and potential risk of adverse effect, etc. It is not always possible to assemble a set of observed variables that serve as a proxy for the available information from all of the signals. Unmeasured variables influence not only time-varying decisions but also biomarker progression. Heckman and Willis[12] reasoned that when unobserved permanent components exist, subjects with similar observables may have heterogenous distribution of responses, i.e. an individual's sequential responses differ systematically from the group's average behavior.

The majority of existing causal inference methods for comparing time-varying treatment assume unconfoundedness, also known as the no unmeasured confounders assumption or sequential exchangeability, i.e. the treatment assignment is independent of the potential outcomes conditional on some observed variables. The potential existence of unmeasured factors that may confound the treatment assignment and biomarker dynamics violates this fundamental assumption and thus undermines these methods, including g-estimation[26;41], structural nested models[10], history-restricted marginal structural models[21], and longitudinal targeted maximum likelihood estimation[35]. Econometric literature, on the other hand, uses unobserved effects models (UEM) or unit fixed-effects models[9;15] to eliminate time-invariant unmeasured confounding by including subject-specific intercepts and having each subject act as their own control. Imai and Kim[14] used UEM in matching to estimate contemporaneous treatment effect, i.e. comparing the outcome right before and immediately after a change in the treatment status over a short time period. The main drawback of using an UEM is that due to its assumption of strict exogeneity, it is difficult to simultaneously address biases from reverse causation and time-dependent confounding[3], which are common in the causal comparison of dynamic treatment regimes.

From a modeling perspective, we account for the unmeasured patient heterogeneity in both treatment assignment and biomarker dynamics via multivariate generalized linear mixed-effects models (MGLMM)[39;1], which allows partial identification of unobserved permanent components through re-

peated measurements for each individual in a larger population. Behavioral and social science researchers have long used mixed-effects model[2;4;20;17;24] in research involving longitudinal data . The ability of mixed-effects models to estimate subject-specific random effects allows for quantitative characterization of between-subject heterogeneity due to unobserved factors[31;5]. Furthermore, these models describe the within-subject dependence in the time-varying outcome, which improves parameter estimation efficiency. However, due to the nonlinear link functions in MGLMM, estimated parameters in the generalized model often only have causal interpretations conditional on the random effects, that is, fixed-effects coefficients no longer lead to marginal causal effect based on potential outcomes[8] even when all covariates are exogenous[40;11].

To address this issue and enable the estimation of marginal causal effect for comparing treatment regimes with MGLMM on both the population and subgroup levels, we use the g-computation algorithm, which underpins the majority of Bayesian causal inference methods. This approach directly simulates potential outcomes under a treatment sequence based on conditional distributions of time-varying confounders and outcomes estimated from the data, consistently estimating potential outcomes and thus causal effects if all the conditional distributions are correctly specified. Standard g-estimation methods lead to biased effect estimates when unmeasured confounders are present, as the unobserved potential outcomes are not missing at random. From a sensitivity analysis perspective, Yang and Lok[37] assumes a nonidentifiable bias function quantifying the impact of unmeasured confounding on the average potential outcome under structural nested mean models. Sitlani et al.[33] and Qian et al.[23] compared treatment paths that differ only at a single point in time and discussed likelihood decomposition, which supports the causal interpretation of the fixed-effects coefficients estimated from a linear mixed model, i.e. as a "blip" of a structural nested model. Shardell and Ferrucci[32] incorporated joint mixed-effects models in the g-computation algorithm to estimate the population average effect of treatment regimes over time.

In this paper, we relax the unconfoundedness assumption and provide a framework for causal comparison of treatment paths using MGLMM, which accounts for the presence of unmeasured time-invariant variables as latent subject heterogeneity in treatment assignments, longitudinal outcomes, and time-varying confounders. We aim to synthesize evidence from the population pertinent to clinical decisions of an individual and to account for the dynamic progression of the individual's trajectories, all while addressing the unobserved permanent factors in selection bias and adhering to the generic causal inference ideology of only using the past to infer on the current status. Existing works on causal inference with longitudinal data using mixed-effect models often marginalize over the latent components and identify causal estimand as a function of the treatment path, covariates, and fix-effects coefficients. While the unobserved stable trait factors influencing disease progression remain constant over time, our proposal dynamically updates the information relevant to these factors by sequentially estimating the subject-specific latent variables in the longitudinal outcome and time-varying confounder models based on a subject's accumulating observed or counterfactual history over time. In addition, we note that the distribution of treatment assignment heterogeneity is not fully identifiable under the parametric specification of the MGLMM because treatment assignment is not necessarily a recurring process. We introduce the population and subgroup distribution of the treatment assignment heterogeneity as built-in sensitivity parameters for the treatment regime comparison.

Our proposal engages the treatment assignment model as part of a larger picture to bridge the gap between the confoundedness in selection bias and the heterogeneity of patients' dynamic disease progression. The work has several advantages. First, existing ways of incorporating propensity score (PS) in Bayesian causal inference[18] include specifying outcome distribution conditional on PS[41], hav-

ing shared priors between propensity and outcome models, or using an inverse probability weighting or doubly robust estimator[30]; our method provides a new way to connect the propensity with the outcomes and time-varying confounders via the dependence structure on the subject-specific unobserved heterogeneity of the model components. Second, our method naturally incorporates unmeasured time-invariant factors via the random effects in MGLMM, for which the estimated covariances partially inform possible existence of unmeasured confounders. Third, we provide a new perspective for investigating the impact of potential unmeasured confounding by using the distribution of treatment assignment heterogeneity as the sensitivity parameters, rather than quantifying unmeasured confounders in selection bias as sensitivity parameters[27;37]. While random effects in time-varying outcome and confounder models reflect unobserved stable trait physiological factors of disease progression, treatment assignment heterogeneity is usually contextual and may be tractable based on knowledge about data collection and practice routine. As a result, the sensitivity parameters can be tailored to practitioners' needs as a controllable component to test the sensitivity of the causal estimates. And lastly, under certain circumstances, such as when treatment assignment heterogeneity is absent under the given parametric model specification or when a subgroup distribution of treatment assignment heterogeneity is assumed, our approach makes marginal subgroup treatment effect identifiable.

2 Notation and Model

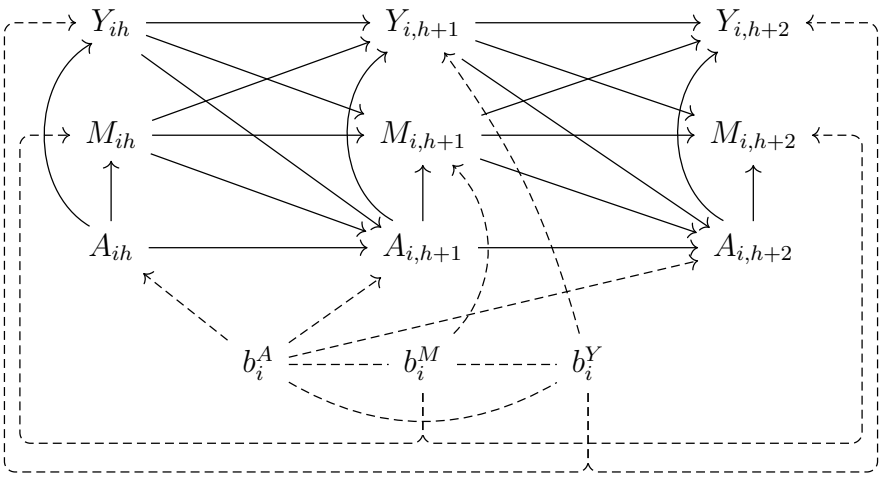

Figure 1: Directed acyclic graph (DAG) for the generalized linear mixed model displaying temporal order of the observed variables and time-invariant unmeasured heterogeneity in both treatment assignment and biomarker dynamics. Baseline characteristics V_i is excluded from the figure for simplicity.

We consider a longitudinal study that involves time-varying treatment that is sequentially randomized, and assume that time-invariant unobserved heterogeneity exists in both the biomarker dynamics and treatment assignment. This paper demonstrates the method under the assumed temporal relationship of the variables as described in Figure 1, where an arrow suggests the potential of causal relationship (single arrow) or covariance (no arrow), whereas a missing arrow implies zero influence or zero covariance. There are multivariate stochastic processes, $\{(Y_t, M_t, A_t) : t \geq 0\}$, where Y_t, M_t, and A_t represent the outcome process, time-dependent confounders, and sequential treatment, respectively. The confounders are affected by previous exposure and influence future outcomes and

treamtent assignment. Without loss of generality, we assume dichotomous treatment, $A_t \in \{0,1\}$. Let $\overline{Y}_{i,t_1:t_2}$, $\overline{M}_{i,t_1:t_2}$, and $\overline{A}_{i,t_1:t_2}$ denote the longitudinal paths observed for biomarkers, confounders, and interventions during time $t = t_1, \ldots, t_2$ for subject i, $i = 1, \ldots, N$. At any time t, practitioners decide on $A_{i,(t+1)}$ based on clinical history recorded up to time t, i.e. past treatment path $\overline{A}_{i,0:t}$ and measurement history $\mathcal{H}_{i,t+1} = (V_i, \overline{Y}_{i,0:t}, \overline{M}_{i,0:t})$, where V_i is the vector of baseline information. The updated clinical history incorporating the latest treatment decision, $(\mathcal{H}_{it}, \overline{A}_{i,0:(t+1)})$, are then the observable information in explaining the dynamics of $(Y_{i,t+1}, M_{i,t+1})$.

In this paper, we restrict the discussion to studying treatment initiations such that an initiation occurs at a single time s and we assume subjects to remain treated after time s. Without loss of generality, we consider the outcomes to be continuous and the time-dependent confounders to be the pattern of subject visits. We model the confounders as binary variables based on the missing structure of the longitudinal outcomes. We propose to use the following longitudinal multivariate generalized linear mixed model (MGLMM) for characterizing individual-level time-specific progression of biomarkers and treatment assignments. For $t = 1, \ldots, T$, the continuous outcomes have a linear mixed model specification,

$$
\begin{aligned}
Y_{it} &= \mu_{it}^Y(\mathcal{H}_{it}, \overline{A}_{i,0:t}, b_i^Y; \theta^Y) + \sigma\psi_{it}^Y, \quad \mathbb{E}(Y_{it}|\mathcal{H}_{it}, \overline{A}_{i,0:t}, b_i^Y; \theta^Y) = \lambda_Y^{-1}(\eta_{it}^Y) \\
\eta_{it}^Y &= \phi_1^Y(\mathcal{H}_{it})\beta_1^Y + \phi_2^Y(\mathcal{H}_{it})\phi_A(\overline{A}_{i,0:t})^T\beta_2^Y + \phi_3^Y(\mathcal{H}_{it})b_{i0}^Y + \phi_4^Y(\mathcal{H}_{it})\phi_A(\overline{A}_{i,0:t})^T b_{i1}^Y,
\end{aligned} \tag{1}
$$

where λ_Y is the link function, $\phi_A(\overline{A}_{i,0:t})$ may be the indicator of dosage information for person i at time t with maximum dose K, $(\mathbb{1}\{\sum_{s=1}^t A_{is} = 1\}, \ldots, \mathbb{1}\{\sum_{s=1}^t A_{is} = K\})$, $\sigma \in \theta^Y$ is the standard deviation of the outcome distribution, $\theta^Y = (\beta_1^Y, \beta_2^Y, \sigma)$ is the vector of outcome model parameters, $b_i^Y = (b_{i0}^Y, b_{i1}^Y)$ is the vector of random effects, ψ_i^Y is the stochastic randomness following a mean zero distribution, e.g. $N(0,1)$, and $\phi_2^Y(\mathcal{H}_{it}) \subseteq \phi_1^Y(\mathcal{H}_{it})$, $\phi_3^Y(\mathcal{H}_{it}) \subseteq \phi_1^Y(\mathcal{H}_{it})$, and $\phi_4^Y(\mathcal{H}_{it}) \subseteq \phi_2^Y(\mathcal{H}_{it})$. Treatment initiation is modeled as

$$
\begin{aligned}
&(A_{it} = 1|A_{i,t-1} = 0) \sim f_A(\mathcal{H}_{it}, b_i^A; \theta^A, \psi_{it}^A), \quad \mathbb{E}(A_{it}|A_{i,t-1} = 0, \mathcal{H}_{it}, b_i^A; \theta^A) = \lambda_A^{-1}(\eta_{it}^A), \\
&\eta_{it}^A = \phi_1^A(\mathcal{H}_{it})\beta_1^A + \phi_2^A(\mathcal{H}_{it})b_{i0}^A,
\end{aligned} \tag{2}
$$

where λ_A is the logit function, $\theta^A = (\beta_1^A, \beta_2^A)$, $b_i^A = b_{i0}^A$ is random effect , and $\phi_2^A(\mathcal{H}_{it}) \subseteq \phi_1^A(\mathcal{H}_{it})$. With a binary dependent variable, the randomness satisfies $\psi_{it}^A \sim U(0,1)$ such that $A_{it} = \mathbb{1}\{\psi_{it}^A \leq \mu_{it}^A\}$ when $A_{i,t-1} = 0$. For time-dependent confounders M_{it}, model specification is similar to equation (1),

$$
\begin{aligned}
M_{it} &\sim f_M(\mathcal{H}_{it}, \overline{A}_{i,0:t}, b_i^M; \theta^M, \psi_{it}^M), \quad \mathbb{E}(M_{it}|\mathcal{H}_{it}, \overline{A}_{i,0:t}, b_i^M; \theta^M) = \lambda_M^{-1}(\eta_{it}^M), \\
\eta_{it}^M &= \phi_1^M(\mathcal{H}_{it})\beta_1^M + \phi_2^M(\mathcal{H}_{it})\phi_A(\overline{A}_{i,0:t})^T\beta_2^M + \phi_3^M(\mathcal{H}_{it})b_{i0}^M + \phi_4^M(\mathcal{H}_{it})\phi_A(\overline{A}_{i,0:t})^T b_{i1}^M,
\end{aligned} \tag{3}
$$

where $\theta^M = (\beta_1^M, \beta_2^M)$, $b_i^M = (b_{i0}^M, b_{i1}^M)$ is the vector of random effects, and $\phi_2^M(\mathcal{H}_{it}) \subseteq \phi_1^M(\mathcal{H}_{it})$, $\phi_3^M(\mathcal{H}_{it}) \subseteq \phi_1^M(\mathcal{H}_{it})$, $\phi_4^M(\mathcal{H}_{it}) \subseteq \phi_2^M(\mathcal{H}_{it})$. The λ_M is the logit function in the motivating application as M_{it} represents missing indicators of the outcomes. In order for binary confounders to be identifiable, η_{it}^M has to have a parametric specification and we assume an additive model. Randomness $(\psi_{it}^Y, \psi_{it}^M, \psi_{it}^A)$ is i.i.d and independent of b_i; the same randomness vector is used in each posterior draw for each person at each time point, characterizing the stochasticity of a counterfactual realization, making the potential trajectories under different regimes comparable from controlling for the stochasticity, and is necessary for reproducibility.

The three model components, (1), (2), and (3), are connected through a correlation structure between the random effects,

$$b_i = (b_i^Y, b_i^M, b_i^A)^T \sim MVN(0, G_i),$$

where G_i is typically assumed to be the same across subjects, i.e. $b_i \sim MNV(0, G)$. These random effects are interpreted as unobserved time-invariant subject-specific heterogeneity; they are patient characteristics that influence the clinical trajectories and treatment assignment processes directly via random intercepts and indirectly via the effect of factors through random slopes. Specifically, b_i^A represents the unmeasured static heterogeneity that affects treatment assignment, such as a patient's frailty observed but not recorded in clinic. Without loss of generality, we assume that $\psi_{it}^Y \sim N(0, 1)$, identity link for λ_Y, and logit link for λ_A and λ_M for the rest of the manuscript.

3 Bayesian G-Computation with MGLMM

3.1 Causal Quantities and Target Estimand

Until now, we have focused on using MGLMM to describe the data-generating mechanism as illustrated in Figure 1. When the MGLMM is correctly specified, the posterior predictive samples of the model parameters concentrates on the true data distribution. In most cases, model parameters in MGLMM do not have a causal interpretation due to the random effects, with an exception explained in Appendix D. Shardell and Ferrucci [32] illustrated that the estimation of causal effects has to assume sequential exchangeability conditional on the random effects b_i and expressed the causal quantity as a function of the parameters under certain assumptions.

A treatment regime dynamically defines a patient's present treatment status as a function $q(\cdot)$ of the observed or counterfactual clinical history, i.e. given a past treatment path and measurement history up to time t, $(\overline{A}_{i,0:(t-1)}, \mathcal{H}_{it})$, the treatment sequence under regime q is sequentially determined by $a_t(q) = q(\overline{A}_{i,0:(t-1)}, \mathcal{H}_{it})$. For any variable X, $X(q)$ represents the value of X had the individual received treatment under regime q. We define $\overline{Y}_{i,0:t}(q)$, $\overline{M}_{i,0:t}(q)$, and $\overline{a}_{0:t}(q) = (a_1(q), \ldots, a_t(q))$ as the counterfactual longitudinal trajectories of outcomes, confounders, and treatment path under regime q, and write the counterfactual measurement history under regime q up to before time t as $\mathcal{H}_{it}(q) = \{V_i, \overline{Y}_{i,t-1}(q), \overline{M}_{i,t-1}(q)\}$.

From a modeling point of view, we recognize that the full distribution of b_i^A is not identifiable from the observed data because treatment initiation happens at most once for each subject. We create identifiability by stratifying the causal estimation on the heterogeneity in treatment assignment, b_i^A. Given a specific regime of interest, q, and the unobserved time-invariant heterogeneity, b_i^A, we aim at identifying the joint distribution of a future τ days of counterfactual trajectories conditional on observed history up to a present time h, :

$$P(\overline{Y}_{(h+1):(h+\tau)}(q), \overline{M}_{(h+1):(h+\tau)}(q) | V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A). \tag{4}$$

Based on (4) and g-computation [26], we can identify causal effects that are functions of the fix effects parameters and the time-evolving estimations of (b_i^Y, b_i^M), by integrating the expectation of the outcome over observed or counterfactual histories while fixing the treatment according to the regime of interest. For the purpose of inference in the motivating application, we set the target estimand to be the conditional subgroup effect, i.e.

$$\mathbb{E}(Y_{h+\tau}(q) | V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A), \tag{5}$$

and the causal contrast of comparing regime q_1 to q_2 is then $\mathbb{E}(Y_{h+\tau}(q_1)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A) - \mathbb{E}(Y_{h+\tau}(q_2)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$, given history information up to a time h, baseline characteristics, and the time-invariant assignment heterogeneity b_i^A. The target estimand (5) is a function of (4) as follows,

$$\begin{aligned}
&\mathbb{E}(Y_{h+\tau}(q)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A) \\
&= \int_{y_{h+\tau}} \int_{m_{h+\tau}} \cdots \int_{y_{h+1}} \int_{m_{h+1}} \\
&\quad y_\tau P(\overline{Y}_{(h+1):(h+\tau)}(q) = \overline{y}_{(h+1):(h+\tau)}, \overline{M}_{(h+1):(h+\tau)}(q) = \overline{m}_{(h+1):(h+\tau)}|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A) \\
&\qquad\qquad dm_{h+1} dy_{h+1} \ldots dm_{h+\tau} dy_{h+\tau}.
\end{aligned}$$

In the next section, we describe the subgroup causal effect estimation procedure under MGLMM, which jointly models multivariate time-varying components with link functions that may not be identity or log link, accounting for the accumulation of individual information over time via a time-evolving update of the time-invariant unobserved traits (b_i^Y, b_i^M).

3.2 Assumptions and Method

We propose a procedure for estimating the causal effect given b_i^A while accounting for the change of individual trajectory over time, which is essential for precision medicine. The proposal enables us to assess the sensitivity of the longitudinal causal effect estimation to unobserved treatment heterogeneity, while accounting for potential violation to the assumption of no unmeasured confounding. In order to show that (4) can be consistently estimated without parametric form and distributional assumptions, we make the following assumptions:

Assumption. *For $t = 0, \ldots, T$,*

1. *Consistency: $\overline{Y}_{0:t} = \overline{Y}_{0:t}(q)$ and $\overline{M}_{0:t} = \overline{M}_{0:t}(q)$ if $\overline{A}_{0:t} = \overline{a}_{0:t}(q)$;*
2. *Positivity: $P(A_{t+1} = a_{t+1}(q)|V, \overline{A}_{0:t} = \overline{a}_{0:t}(q), \overline{Y}_{0:t}, \overline{M}_{0:t}, b_i^A) > 0$ with probability 1 for $t \geq 0$;*
3. *Sequential exchangeability given b_i^A: for $\tau > 0$,*

$$\begin{aligned}
&P(\overline{Y}_{(t+1):(t+\tau)}(q), \overline{M}_{(t+1):(t+\tau)}(q)|V, A_{t+1}, \overline{A}_{0:t} = \overline{a}_{0:t}(q), \overline{Y}_{0:t}, \overline{M}_{0:t}, b_i^A) \\
&= P(\overline{Y}_{(t+1):(t+\tau)}(q), \overline{M}_{(t+1):(t+\tau)}(q)|V, \overline{A}_{0:t} = \overline{a}_{0:t}(q), \overline{Y}_{0:t}, \overline{M}_{0:t}, b_i^A).
\end{aligned}$$

The consistency assumption states that when the observed treatment path follows the hypothesized regime of interest, the observed and counterfactual biomarker dynamics are equivalent. It is important to note that the equivalence does not imply the same value, but rather the same distribution. Positivity guarantees that there is no systematic exclusion of a plausible treatment pattern over time. The classic assumption of sequential exchangeability[7] is commonly adopted in the existing literature, assuming that the observed pretreatment history can sufficiently explain the dependence between a current treatment assignment and future counterfactuals. We further elevate this assumption to condition on the unobserved time-invariant tendencies in treatment assignment, which is quantified by the random effects in model (2), b_i^A. In practice, such heterogeneity in treatment assignment is not captured by data collection and may be attributable to patients' willingness to be treated, the potential risk of adverse effects from treatment, and the clinician's perception of treatment. The conditional sequential

exchangeability must hold for identifying causal effects under the correct specification of the outcome model (1), treatment assignment model (2), and confounder model (3). Under these assumptions, (4) can be nonparametrically identified as below (see Appendix A for details),

$$
\begin{aligned}
&P(\overline{Y}_{(h+1):(h+\tau)}(q), \overline{M}_{(h+1):(h+\tau)}(q) | V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A) \\
&= \prod_{s=h}^{h+\tau-1} \int_{u_s} \int_{v_s} P(Y_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^Y = u_s) \\
&\qquad\qquad P(M_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^M = v_s) \\
&\qquad\qquad P(b_i^Y = u_s, b_i^M = v_s | V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A) du_s dv_s. \qquad (6)
\end{aligned}
$$

In Figure 2, we use a single-world intervention graph[13;25] (SWIG) to display the independencies necessary to obtain (6) and show the counterfactual dependencies that would exist if we set the treatment path to certain values. The graph is constructed by splitting the treatment nodes $\overline{A}_{i,(h+1):(h+\tau)}$ of the causal diagram in Figure 1 and replacing all descendants of the assigned treatment with their potential outcomes, marking all counterfactuals in red. The conditional sequential exchangeability assumption is demonstrated in the SWIG by d-separation between the counterfactual trajectories $(\overline{Y}_{(h+1):(h+\tau)}(q), \overline{M}_{(h+1):(h+\tau)}(q))$ and $A_{i,h+1}$ conditional on $(\overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$. If b_i^A is not controlled for, selection bias would be induced by paths $A_{i,h+1} \leftarrow b_i^A \leftrightarrow b_i^Y \rightarrow Y_{i,h+1}(q)$ and $A_{i,h+1} \leftarrow b_i^A \leftrightarrow b_i^M \rightarrow M_{i,h+1}(q)$, while stratifying by b_i^A does block these paths.

Variables inside rectangles of Figure 2 are quantities involved in (6) that are dependent on the time-invariant unobserved traits of subject i, (b_i^Y, b_i^M, b_i^A). As more samples of such variables are realized over time, (b_i^Y, b_i^M) can be partially identified under MGLMM based on information backflow from observed or counterfactual biomarker dynamics, resulting in a sequential updating of these subject-specific unobserved permanent features. Hypothesized treatment status $\overline{a}_{it}(q)$, $t \in [h+1, h+\tau]$, does not contribute to the sequential update of (b_i^Y, b_i^M), as it is a deterministic function of pretreatment history and independent of random effects conditional on pretreatment history.

Sampling from the conditional distribution of the counterfactual trajectory based on (6) involves integration over a different posterior distribution of the unobserved permanent variables (b_i^Y, b_i^M) at each time $s \in [h, h+\tau)$ due to the sequential update of these components as history information accumulates. For the Monte Carlo sampling of (b_i^Y, b_i^M) from its distribution at a time $m \in (h, h+\tau)$, $P(b_i^Y, b_i^M | V, \overline{A}_{0:h}, \overline{Y}_{0:m}, \overline{M}_{0:m}, b_i^A)$, it is conditional on the trajectories $(\overline{Y}_{0:m}, \overline{M}_{0:m})$ which, as implied by the formulation of equation (6), follows the distribution

$$
\prod_{s=h}^{m-1} P(Y_{s+1}, M_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A).
$$

In other words, trajectories $(\overline{Y}_{0:m}, \overline{M}_{0:m})$ involved in the sequential update of (b_i^Y, b_i^M) at time m has the same distribution as $(\overline{Y}_{0:h}, \overline{Y}_{(h+1):m}(q), \overline{M}_{0:h}, \overline{M}_{(h+1):m}(q))$, a mix of observed and counterfactual variables under regime q and treatment assignment heterogeneity b_i^A. MGLMM often has nonlinear link functions, so the sampling of $(b_i^Y, b_i^M) \sim P(b_i^Y, b_i^M | V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A)$ for $s \in [h, h+\tau)$ is nontrivial. We consider the following general strategy: first, calculate the Laplace approximation of the posterior distribution $(b_i^Y, b_i^M, b_i^A | V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s})$, denoted by $MVN(\hat{b}_i, V)$, and then sample the heterogeneities via the corresponding conditional distribution, $(b_i^Y, b_i^M) | b_i^A$, with b_i^A set to a certain value. The procedure is illustrated in Appendix B. The pseudocode for generating posterior samples of

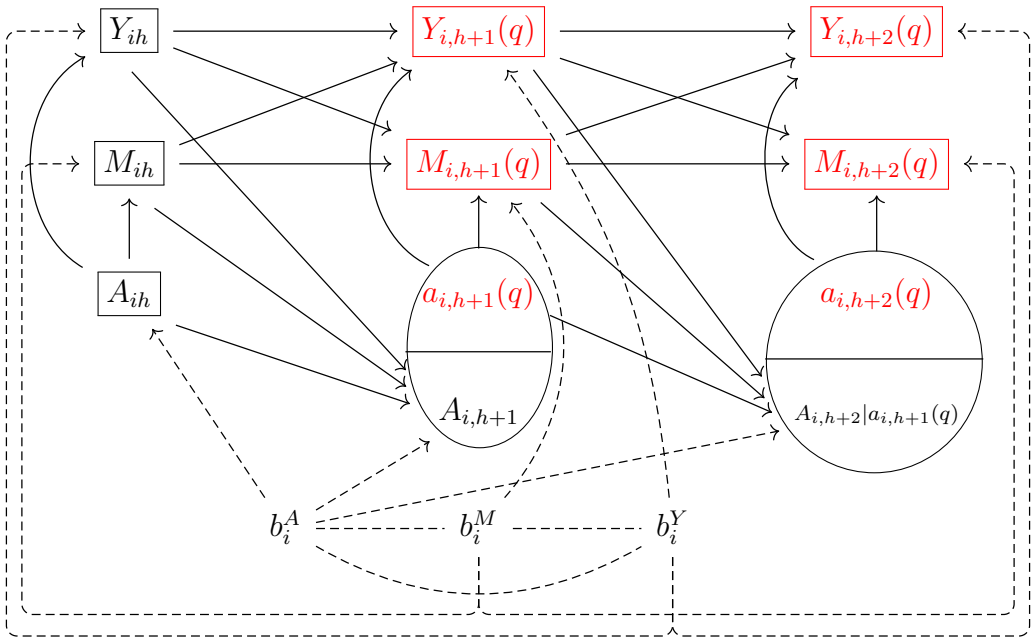

Figure 2: SWIG.

counterfactual trajectories from $P(\overline{Y}_{(h+1):(h+\tau)}(q), \overline{M}_{(h+1):(h+\tau)}(q)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$ based on (6) is provided in Appendix C.

We have been stratifying by b_i^A for the discussion so far. MGLMM specifies that (b_i^Y, b_i^M, b_i^A) jointly follows a multivariate Gaussian distribution $MVN(0, G)$, where we let v denote the variance of b_i^A. Based on model specification, the marginal population distribution of treatment assignment heterogeneity, b_i^A, is Gaussian distributed with variance v. Thus, the population ATE $\mathbb{E}(Y_{h+\tau}(q))$ can be obtained by integrating (4) over subgroup variables $(V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h})$, counterfactual history $(\overline{Y}_{(h+1):(h+\tau-1)}(q), \overline{M}_{(h+1):(h+\tau-1)}(q))$, and treatment assignment heterogeneity $b_i^A \sim N(0, v)$. Appendix B gives further details to the calculation of the mixed population ATE[19], $\widehat{\mathbb{E}}(Y_{h+\tau}(q))$, which replaces the target subgroup distribution with the corresponding empirical subgroup distribution from the data, i.e. $\widehat{P}(V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h})$. The evaluation of causal effectiveness may vary under different values of v, which represents the assumed amount of variation in treatment assignment heterogeneity among subjects. That is, the variance of the unidentifiable time-invariant quantity b_i^A serves as a sensitivity parameter in the estimation of causal effects. Specifying v creates identifiability for conditional subgroup ATE such as $\mathbb{E}(Y_{h+\tau}(q)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$ and marginal population ATE $\mathbb{E}(Y_{h+\tau}(q))$. Given the parametric assumption of MGLMM and v, distribution $P(b_i^A|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h})$ is estimable conditional on history information for the calculation of the marginal subgroup ATE, which is available as

$$\begin{aligned}
&\mathbb{E}(Y_{h+\tau}(q)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}) \\
&= \int_w \mathbb{E}(Y_{h+\tau}(q)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A = w) P(b_i^A = w|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}) dw.
\end{aligned}$$

Assuming no treatment heterogeneity under MGLMM is equivalent to setting $v = 0$, which is a sufficient but unnecessary condition for having no unmeasured confounders. In MGLMM, $\text{cov}(b_i^A, b_i^M) = \text{cov}(b_i^A, b_i^Y) = 0$ leads to no unmeasured confounders. That is, even when no unmeasured confounders

is true, MGLMM still allows unobserved factors to influence treatment assignment as long as it is not correlated with the unobserved heterogeneity in biomarker dynamics (b_i^Y, b_i^M); examples of such an unconfounding treatment assignment heterogeneity include a patient's preference for a treatment based on personal beliefs or social stigma. On the other hand, we note that the covariances $\text{cov}(b_i^A, b_i^M)$ and $\text{cov}(b_i^A, b_i^Y)$ are estimable given v, the presumed variance of b_i^A. Henceforth, our method does partially inform the possible existence of unmeasured confounders based on the estimated covariances in the MGLMM.

Next, we demonstrate that having sequential exchangeability conditional on b_i^A leads to a softer assumption than on b_i^Y. When there is no treatment assignment heterogneiety, the proposal simplifies to the standard g-computation of fitting only the outcome and confounders model using generalized linear mixed-effects model because the assignment mechanism becomes ignorable[19]. Let us consider a simplified scenario of looking at the subgroup ATE at time $h+1$ conditional on history information up to time h, assuming no time-varying confounders and no treatment assignment heterogeneity. The marginal subgroup ATE would not be identifiable under a sequential exchangeabiltiy conditional on b_i^Y,

$$Y_{h+1}(q) \perp A_{h+1} | V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y.$$

This assumption leads to direct identification of the conditional counterfactual distribution as

$$P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) = P(Y_{h+1}|V, \bar{A}_{0:(h+1)} = \overline{a}_{0:(h+1)}(q), \overline{Y}_{0:h}, b_i^Y).$$

However, the target quantity represented by $P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h})$ would not be calculable because

$$P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) = \int P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) db_i^Y$$

and the subgroup heterogeneity distribution $P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h})$ is unknown. Whereas with our proposal, we assume the following conditional sequential exchangeabiltiy,

$$Y_{h+1}(q) \perp A_{h+1} | V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^A.$$

Given the assumption of no treatment assignment heterogeneity, we know $\text{var}(b_i^A) = 0$ and consequently $\text{cov}(b_i^A, b_i^Y) = 0$, leading to $P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^A) = P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h})$. As a result, the target quantity is identifiable via (6) as

$$P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) = \int P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) db_i^Y.$$

Hence, our proposal allows the identification of subgroup causal effects when assuming no treatment assignment heterogeneity, which would not be possible if the sequential exchangeabiltiy is assumed to be conditional on b_i^A. As demonstrated earlier, the merits of using MGLMM come at a price of introducing the variance of b_i^A as a sensitivity parameter, and identifying marginal subgroup causal effects under the existence of treatment assignment heterogeneity needs to be done under additional assumption on the subgroup distribution of b_i^A, which likely requires expert knowledge.

4 Simulation

We assume that each person has two follow-up visits $T_i = 2$, with continuous biomarker Y_{it} and binary time-varying treatment A_{it} satifying

$$\begin{aligned}
Y_{it} =& \beta_0^Y + V_i\beta_1^Y + t\beta_2^Y + \sum_{k=1}^{2} 1\{\sum_{s=1}^{t} A_{i,s} = k\}\beta_{3k}^Y + Y_{i,t-1}\beta_4^Y + b_{i0}^Y + e_{it}^Y \\
=& 0.4 - 0.3V_i - 0.1t + \sum_{k=1}^{2} \frac{k}{2} \times 1\{\sum_{s=1}^{t} A_{i,s} = k\} + 0.4Y_{i,t-1} + b_{i0}^Y + e_{it}^Y \\
\text{logit}\{P(A_{it}(s) = 1 | A_{i,t-1}(s) = 0)\} & \\
=& \beta_0^A + V_i\beta_1^A + ns(t, \nu_1)\beta_2^A + ns(\tilde{Y}_{i,t-1}(s), \nu_3)\beta_4^A + b_{i0}^A \\
=& -0.1V_i - 0.5t - 0.35Y_{i,t-1} + b_i^A{}_{i0}
\end{aligned}$$

where $e_{it}^Y \sim N(0, 0.4^2)$, V_i is the baseline covariate. Write $\rho = \text{Corr}(b_i^A{}_{i0}, b_{i0}^Y)$, $s_A = \sqrt{\text{Var}(b_i^A{}_{i0})}$, $s_Y = \sqrt{\text{Var}(b_{i0}^Y)}$, and the random effects $(b_i^A{}_{i0}, b_{i0}^Y) \sim N(0, G)$, where the covariance matrix G has elements $G_{11} = s_A^2, G_{12} = G_{21} = \rho s_A s_Y$, and $G_{22} = s_Y^2$. We set $s_Y = 0.8$ and do 100 replicates for each of the 101 settings, $(s_A, \rho) \in (0, 0) \cup \{(s_A, \rho); s_A \in \{0.1, \ldots, 0.9, 1\}, \rho \in \{0, 0.1, \ldots, 0.9\}\}$. When $s_A > 0$ and $\rho < 1$, G is guaranteed to be positive definite because $\det(G) = (1 - \rho)s_A^2 > 0$. When $\rho = 0$, matrix $G = \begin{pmatrix} s_A^2 & 0 \\ 0 & s_Y^2 \end{pmatrix}$ represents the case of no unmeasured confounding. We simulate 100 datasets with sample size $n = 500$ for each parameters combination (s_A, ρ).

For each simulated replicate, we sample the posterior predictive distribution of the mixed ATE at the second visit assuming $\widehat{s_A}$ =0, 0.3, or 1, with n_{post} number of posterior draws. The combinations of simulated truth and estimation assumptions under consideration explores the following three scenarios: (1) no unmeasured confounding ($\rho = 0$), (2) unmeasured confounding exists with correctly specified models ($\rho \neq 0, \widehat{s_A} = s_A$), and (3) unmeasured confounding with a mis-specified extent of treatment assignment heterogeneity ($\rho \neq 0, \widehat{s_A} \neq s_A$).

Define $b_i = (b_{i0}^Y, b_i^A{}_{i0})$, $g(\overline{a}_t) = \mathbb{E}(Y_{it}|\overline{A}_{it} = \overline{a}_t)$, $g(\overline{a}_t|V_i) = \mathbb{E}(Y_{it}|\overline{A}_{it} = \overline{a}_t, V_i)$, and $g(\overline{a}_t|V_i, b_i^A) = \mathbb{E}(Y_{it}|\overline{A}_{it} = \overline{a}_t, V_i, b_i^A)$. At each time point,

$$\begin{aligned}
& g(1) - g(0) = \beta_{31}^Y = 0.5 \\
& g(1|V_i) - g(0|V_i) = \beta_{31}^Y \\
& g(\{a_1, a_2\}) - g(\{0, 0\}) \\
& \qquad = \sum_{k=1}^{2} \beta_{3k}^Y \times 1\{a_1 + a_2 = k\} + \beta_4^Y[g(a_1) - g(0)] \\
& \qquad = \sum_{k=1}^{2} \frac{k}{2} \times 1\{a_1 + a_2 = k\} + 0.4 \times 0.5a_1 \\
& g(\{a_1, a_2\}|V_i) - g(\{0, 0\}|V_i) \\
& \qquad = \sum_{k=1}^{2} \beta_{3k}^Y \times 1\{a_1 + a_2 = k\} + \beta_4^Y[g(a_1|V_i) - g(0|V_i)] \\
& \qquad\qquad + \mathbb{E}(b_{i0}^Y|\overline{Y}_{i1}(a_1), V_i) - \mathbb{E}(b_{i0}^Y|\overline{Y}_{i1}(0), V_i)
\end{aligned}$$

$$V_i \sim \text{Bernoulli}(0.5), Y_{i0} \sim N(0,1)$$

Figure 3 summarizes the mean squared error (MSE) of posterior mean MATE, $g(\{a_1, a_2\}) - g(\{0,0\})$, versus its simulated truth and the corresponding posterior coverage. The calculation of $g(\{a_1, a_2\}) - g(\{0,0\})$ was marginalized over $b_i^A \sim N(0, \widehat{s}_A^2)$ under $\widehat{s}_A$ being 0, 0.3, and 1 for the three columns from left to right, respectively. For each plot, the horizontal and vertical axes are the true parameters s_A and ρ used to simulate the data. Focus on $g(\{1,1\}|V_i) - g(\{0,0\}|V_i)$.

Assuming the existence of heterogeneity in the propensity of assigning treatment robustifies the estimation of MATE. That is, the MSE of treatment effect projections is similar when there is high degree of heterogeneity as assumed and when there is no or close to no heterogeneity, e.g. ρ close to zero. In contrast, if we assume there is little or no heterogeneity, e.g. $\widehat{s}_A$ close to zero, we can get small MSE when the assumed $\widehat{s}_A$ is correct, but poor estimation and coverage of the MATE when the assumption of small heterogeneity is incorrect and when there is substantial unmeasured confounding.

5 Application

V is gender, race, age at disease onset, and diffuse status. $\nu_1 = 4, \nu_2 = 4, \nu_3 = 1$. $\tilde{Y}_{i,t-1}$ is the carried forward measurement of biomarker Y at time $t-1$.

$$Y_{it}|\{M_{it}=1\} = \beta_0^Y + V_i\beta_1^Y + ns(t,\nu_1)\beta_2^Y + \sum_{k=1}^{4} 1\{\sum_{s=1}^{t} A_{i,s} = k\}\beta_{3k}^Y + ns(\tilde{Y}_{i,t-1}(s),\nu_3)\beta_4^Y + b_{i0}^Y + e_{it}^Y$$

$$\text{logit}\{P(M_{it}=1)\} = \beta_0^X + V_i\beta_1^X + ns(t,\nu_1)\beta_2^X + \sum_{k=1}^{4} 1\{\sum_{s=1}^{t} A_{i,s} = k\}\beta_{3k}^X + ns(\tilde{Y}_{i,t-1}(s),\nu_3)\beta_4^X + b_{i0}^X$$

$$\text{logit}\{P(A_{it}(s)=0|A_{i,t-1}(s)=1)\} = \beta_0^A + V_i\beta_1^A + ns(t,\nu_1)\beta_2^A + ns(\tilde{Y}_{i,t-1}(s),\nu_3)\beta_4^A + b_i^A{}_{i0}$$

$Y_{it} = \{Y_{it1}, Y_{it2}, Y_{it3}\}$ are FVC, DLCO, and quantilized MRSS skin score, respectively. $M_{it} = \{M_{it1}, M_{it2}, M_{it3}\}$ are the corresponding indicator of each Y_{itk} being updated during the k_{th} grid interval.

$$\begin{pmatrix} b_i^A{}_0 \\ b_0^M \\ b_0^Y \end{pmatrix} = \begin{pmatrix} b_i^A{}_0 \\ b_0^{M_1} \\ b_0^{M_2} \\ b_0^{M_3} \\ b_0^{Y_1} \\ b_0^{Y_2} \\ b_0^{Y_3} \end{pmatrix} \sim N(0, G)$$

where $var(b_0^{X_1}) = var(b_0^{X_2}) = var(b_0^{X_3}) = var(b_0^A) = v$.

6 Discussion

Deciding which treatment is the best for patients of a similar subtype is an important question in treating patients in clinics. Causal inference is a natural tool for answering such questions, but charac-

teristics of clinical data need to be accommodated for valid inference for evaluation of the effectiveness of treatment paths. Clinical datasets are observational, often with treatment assignments that are not randomized based on observed patient history, and measurements are irregular with potentially informative missingness from patients' visit patterns. Among these, natural heterogeneity beyond what observed variables can explain in the biomarker dynamics and treatment assignment is a primary factor to consider in comparing treatment paths. These are characteristics of typical clinical datasets. A basic function of medicine is to identify which treatment regime is the best for the person, and such a question can only be addressed by summarizing evidence from the population of patients similar to this particular person while accounting for the specific biomarker trends of this person. The statistical model used to answer this type of question is complex by nature. This paper provides a tool for regime comparison and describes the simplest possible model that accommodates these characteristics, such as nonrandom treatment assignment and patient heterogeneity.

Our proposal also provides a way of incorporating propensity scores (PS) in Bayesian causal inference. Existing ways of combining PS and outcomes models include specifying outcomes distribution based on PS, having shared parameters or priors between PS and outcome models, or using posterior-based inverse probability weighting or doubly robust estimators[19]. The method in this paper falls under the category of having shared parameters or priors between PS and outcome models, using a multivariate Gaussian latent structure to connect the two through covariance between unmeasured patient heterogeneity terms.

The proposed sequentially update subject-specific heterogeneity in biomarker dynamics as history information accumulates over time. Random effects are involved in the g-computation because for calculating subgroup causal effect. This is because the latent terms cannot be marginalized out as when calculating population effects. Such a model has a specific representation of unmeasured confounding, the degree of which is governed by the unexplained variation in treatment assignment s_A and the correlation between the treatment assignment and biomarker dynamics that operates through the correlation parameter ρ. Note that a small ρ and a large s_A, same as having a large ρ and a small s_A, would lead to heavy unmeasured confounding. The method has the potential to be extended to guide the inclusion of latent variable models in Bayesian causal inference.

References

[1] F. Achana, D. Gallacher, R. Oppong, S. Kim, S. Petrou, J. Mason, and M. Crowther. Multivariate generalized linear mixed-effects models for the analysis of clinical trial–based cost-effectiveness data. *Medical Decision Making*, 41(6):667–684, 2021.

[2] A. Agresti, J. G. Booth*, J. P. Hobert*, and B. Caffo*. Random-effects modeling of categorical response data. *Sociological Methodology*, 30(1):27–80, 2000.

[3] P. D. Allison, R. Williams, and E. Moral-Benito. Maximum likelihood for cross-lagged panel models with fixed effects. *Socius*, 3:2378023117710578, 2017.

[4] M. P. Berger and F. E. Tan. Robust designs for linear mixed effects models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(4):569–581, 2004.

[5] N. Bolger and J.-P. Laurenceau. *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford press, 2013.

[6] R. Y. Coley, A. J. Fisher, M. Mamawala, H. B. Carter, K. J. Pienta, and S. L. Zeger. A bayesian hierarchical model for prediction of latent health states from multiple data sources with application to active surveillance of prostate cancer. *Biometrics*, 73(2):625–634, 2017.

[7] S. Greenland and J. M. Robins. Identifiability, exchangeability, and epidemiological confounding. *International journal of epidemiology*, 15(3):413–419, 1986.

[8] S. Greenland, J. Pearl, and J. M. Robins. Confounding and collapsibility in causal inference. *Statistical science*, 14(1):29–46, 1999.

[9] F. I. Gunasekara, K. Richardson, K. Carter, and T. Blakely. Fixed effects analysis of repeated measures data. *International journal of epidemiology*, 43(1):264–269, 2014.

[10] J. He, A. Stephens-Shields, and M. Joffe. Structural nested mean models to estimate the effects of time-varying treatments on clustered outcomes. *The International Journal of Biostatistics*, 11 (2):203–222, 2015.

[11] P. J. Heagerty. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, 55(3):688–698, 1999.

[12] J. J. Heckman and R. J. Willis. A beta-logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy*, 85(1):27–58, 1977.

[13] M. A. Hernán and J. M. Robins. Causal inference, 2010.

[14] K. Imai and I. S. Kim. When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science*, 63(2):467–490, 2019.

[15] J. S. Kaufman. Commentary: Why are we biased against bias? *International journal of epidemiology*, 37(3):624–626, 2008.

[16] M. R. Kosorok and E. B. Laber. Precision medicine. *Annual review of statistics and its application*, 6:263–286, 2019.

[17] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[18] D. Li, S. Iddi, W. K. Thompson, M. C. Donohue, and A. D. N. Initiative. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Statistical methods in medical research*, 28(3):835–845, 2019.

[19] F. Li, P. Ding, and F. Mealli. Bayesian causal inference: A critical review. *Philosophical Transactions of the Royal Society A, Mathematical, Physical and Engineering Sciences*, 2022.

[20] T. M. Luger, J. Suls, and M. W. Vander Weg. How robust is the association between smoking and depression in adults? a meta-analysis using linear mixed-effects models. *Addictive behaviors*, 39(10):1418–1429, 2014.

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

[22] M. A. Omair, A. Alahmadi, and S. R. Johnson. Safety and effectiveness of mycophenolate in systemic sclerosis. a systematic review. *PLoS One*, 10(5):e0124205, 2015.

[23] T. Qian, P. Klasnja, and S. A. Murphy. Linear mixed models with endogenous covariates: modeling sequential treatment effects with application to a mobile health study. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 35(3):375, 2020.

[24] S. W. Raudenbush and A. S. Bryk. *Hierarchical linear models: Applications and data analysis methods*, volume 1. sage, 2002.

[25] T. S. Richardson and J. M. Robins. Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30):2013, 2013.

[26] J. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7 (9-12):1393–1512, 1986.

[27] J. M. Robins, A. Rotnitzky, and D. O. Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 1–94. Springer, 2000.

[28] A. Rosen, S. L. Zeger, et al. Precision medicine: discovering clinically relevant and mechanistically anchored disease subgroups at scale. *The Journal of clinical investigation*, 129(3):944–945, 2019.

[29] D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[30] M. E. Schnitzer, J. Sango, S. Ferreira Guerra, and M. J. Van der Laan. Data-adaptive longitudinal model selection in causal inference with collaborative targeted minimum loss-based estimation. *Biometrics*, 76(1):145–157, 2020.

[31] J. E. Schwartz and A. A. Stone. The analysis of real-time momentary data: A practical guide. *The science of real-time data capture: Self-reports in health research*, pages 76–113, 2007.

[32] M. Shardell and L. Ferrucci. Joint mixed-effects models for causal inference with longitudinal data. *Statistics in medicine*, 37(5):829–846, 2018.

[33] C. M. Sitlani, P. J. Heagerty, E. A. Blood, and T. D. Tosteson. Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Statistics in medicine*, 31(16): 1738–1760, 2012.

[34] D. P. Tashkin, M. D. Roth, P. J. Clements, D. E. Furst, D. Khanna, E. C. Kleerup, J. Goldin, E. Arriola, E. R. Volkmann, S. Kafaja, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (sls ii): a randomised controlled, double-blind, parallel group trial. *The lancet Respiratory medicine*, 4(9):708–719, 2016.

[35] M. J. Van der Laan, S. Rose, et al. *Targeted learning: causal inference for observational and experimental data*, volume 10. Springer, 2011.

[36] Z. Wang, M. G. Bowring, A. Rosen, B. Garibaldi, S. Zeger, and A. Nishimura. Learning and predicting from dynamic models for covid-19 patient monitoring. *Statistical Science*, 37(2): 251–265, 2022.

[37] S. Yang and J. J. Lok. Sensitivity analysis for unmeasured confounding in coarse structural nested mean models. *Statistica Sinica*, 28(4):1703, 2018.

[38] A. C. Zamora, P. J. Wolters, H. R. Collard, M. K. Connolly, B. M. Elicker, W. R. Webb, T. E. King Jr, and J. A. Golden. Use of mycophenolate mofetil to treat scleroderma-associated interstitial lung disease. *Respiratory medicine*, 102(1):150–155, 2008.

[39] S. L. Zeger and M. R. Karim. Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American statistical association*, 86(413):79–86, 1991.

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

[41] T. Zhou, M. R. Elliott, and R. J. Little. Penalized spline of propensity methods for treatment comparison. *Journal of the American Statistical Association*, 114(525):1–19, 2019.

Figures

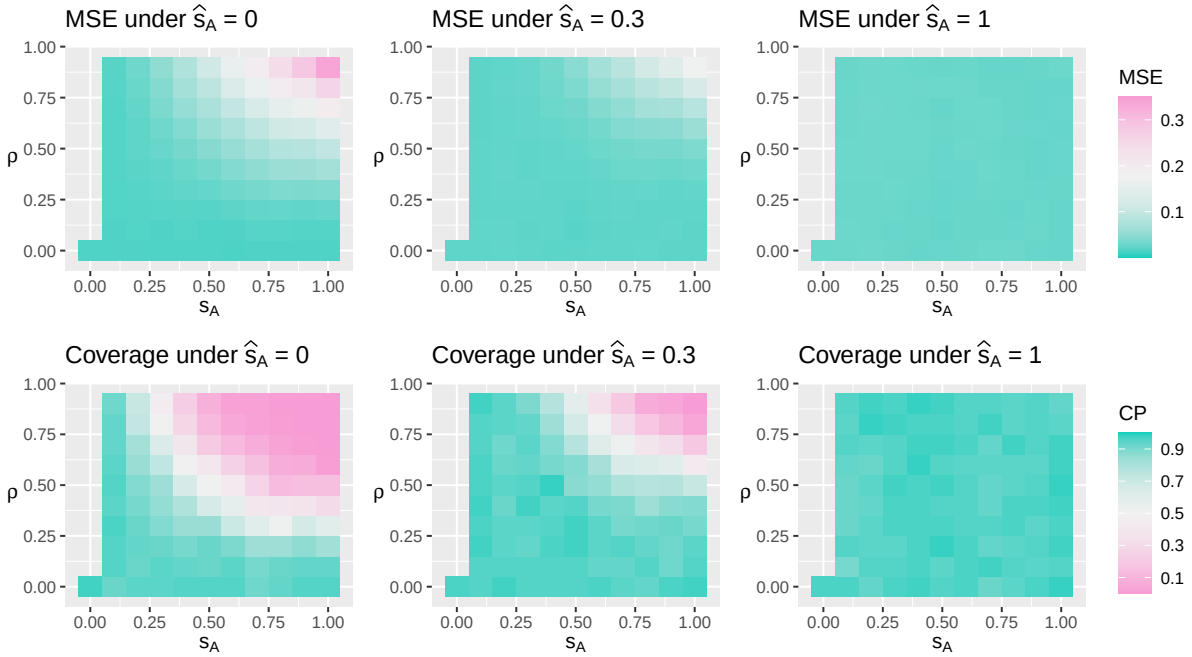

Figure 3: Under true treatment effect being 1.2 at the second time point, the figure displays mean squared error (MSE) and posterior coverage for mixed average treatment effect under different simulation truth (s_A, ρ) and assumed model parameter $\hat{s}_A$. Color green refers to better estimation, e.g. lower MSE and higher coverage probability.

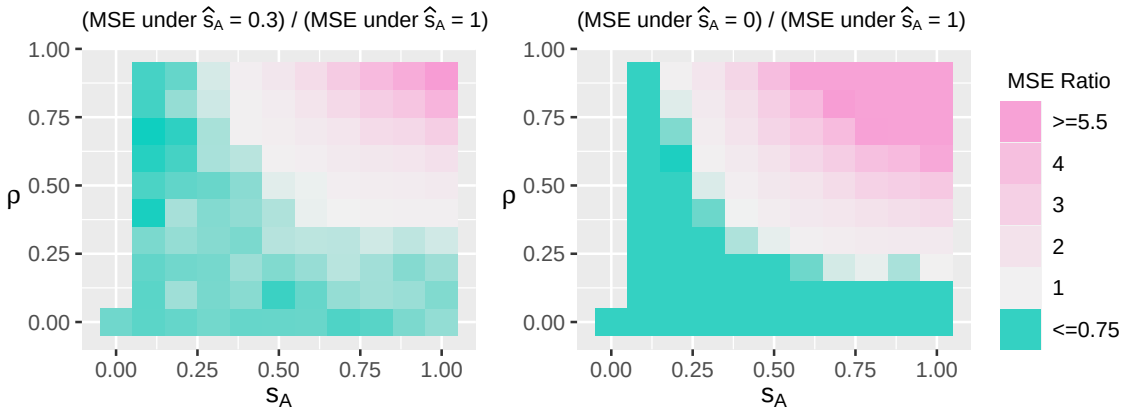

Figure 4: MSE ratio under true treatment effect being 1.2 at the second time point.

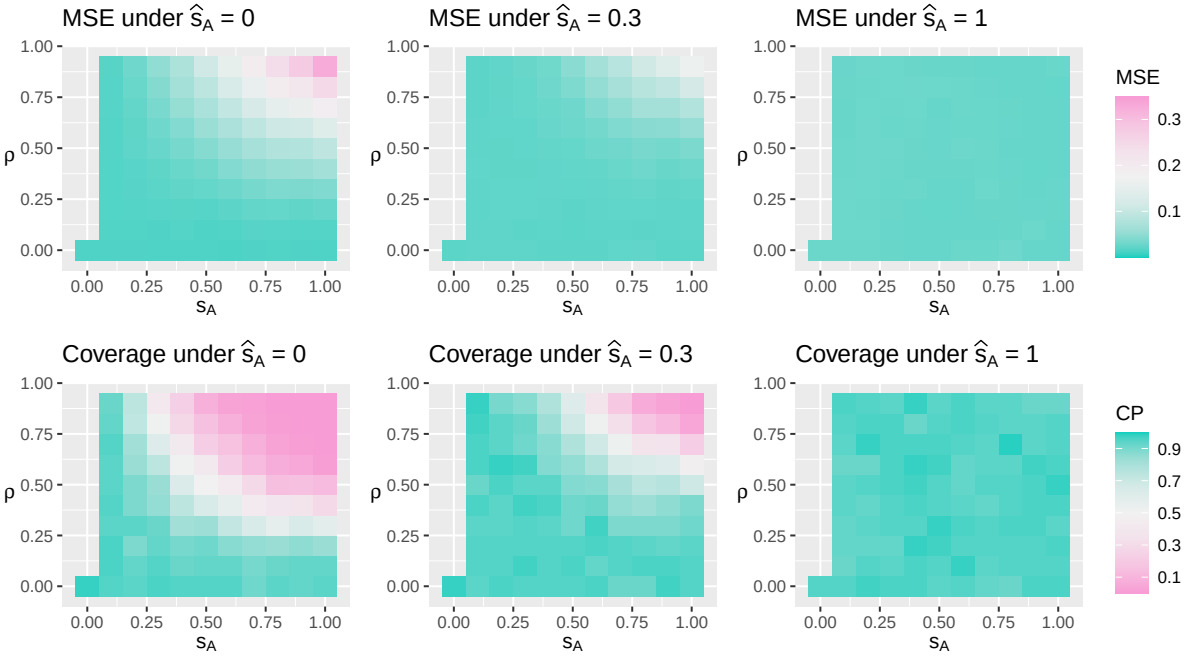

Figure 5: Under true treatment effect being 0 at the second time point, the figure displays mean squared error (MSE) and posterior coverage for mixed average treatment effect under different simulation truth (s_A, ρ) and assumed model parameter $\widehat{s}_A$. Color green refers to better estimation, e.g. lower MSE and higher coverage probability.

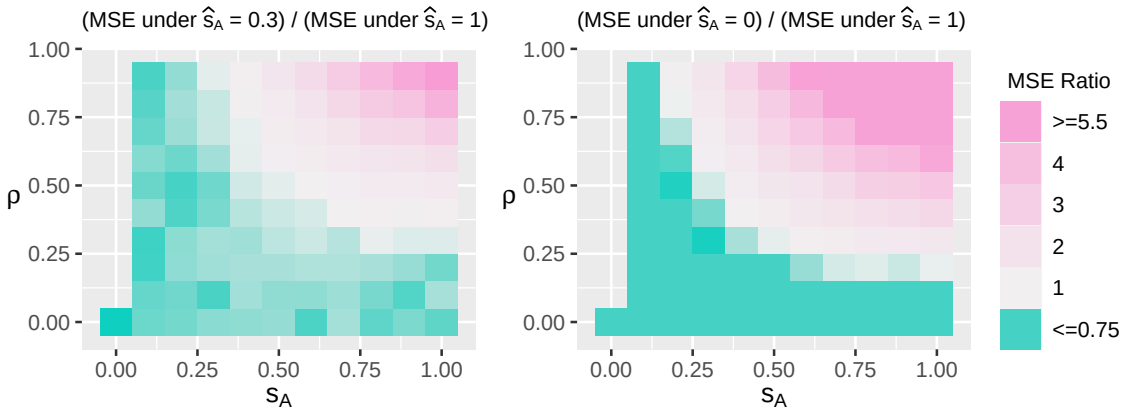

Figure 6: MSE ratio under no treatment effect.

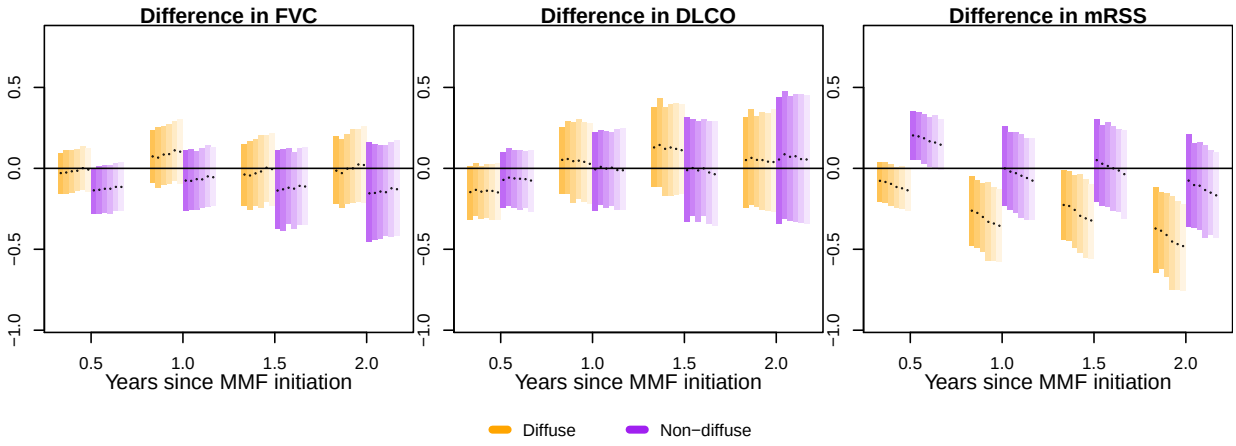

Figure 7: Application causal estimation by subgroup.

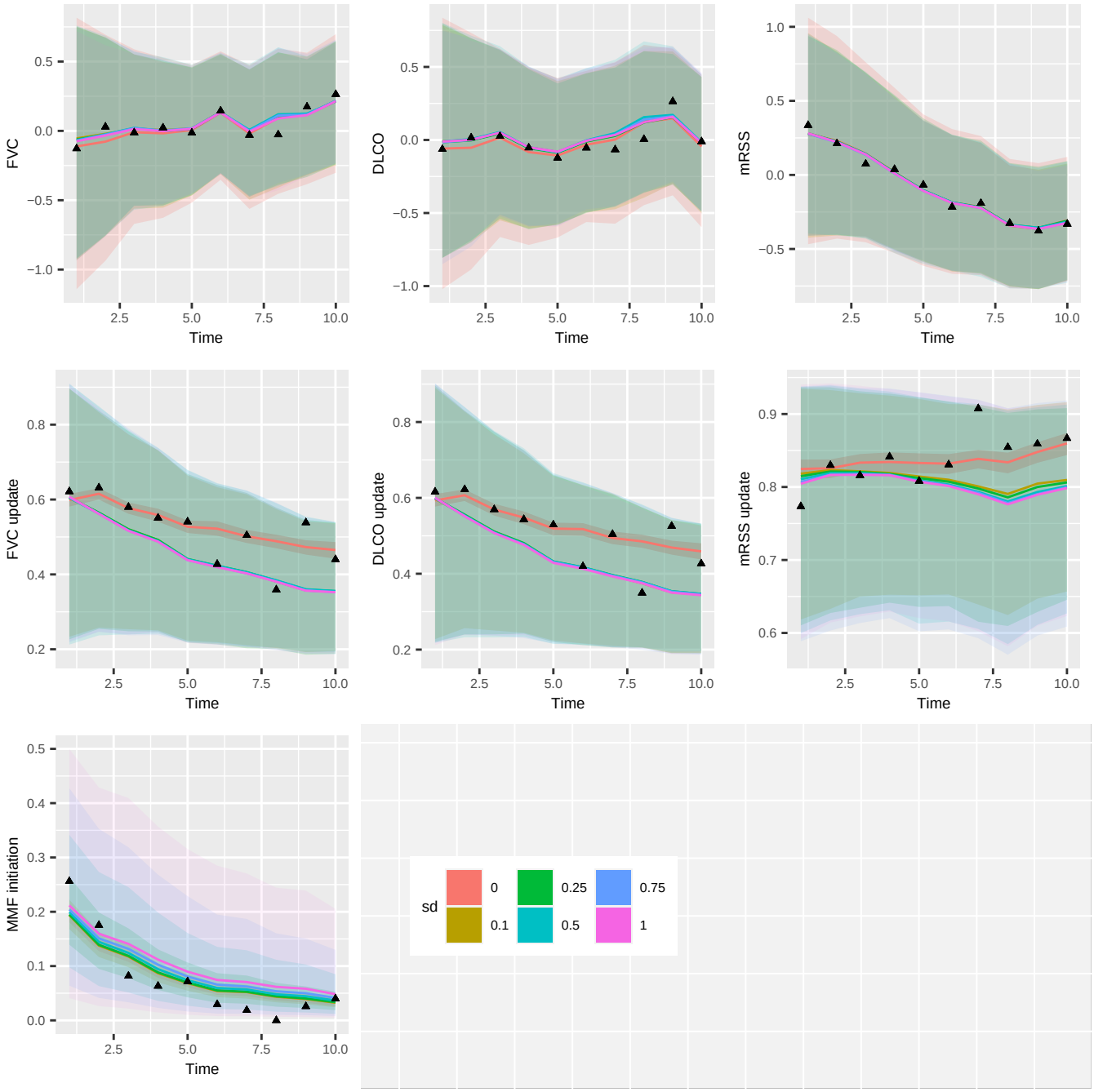

Figure 8: Application accuracy plot.

Appendices

A Identification of the G-formula

For simplicity, we ignore the subscript i for indexing subjects. Assuming $\overline{A}_{0:h} = \overline{a}_{0:h}(q)$ and time-invariant latent treatment heterogeneity $b_i^A = b_i^A$, the distribution of counterfactual trajectories for the future τ time intervals conditional on observed information up to time h can be processed as follows.

$$P(\overline{Y}_{(h+1):(h+\tau)}(q), \overline{M}_{(h+1):(h+\tau)}(q) | V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$$

by positivity and exchangeability,

$$= P(\overline{Y}_{(h+1):(h+\tau)}(q), \overline{M}_{(h+1):(h+\tau)}(q) | V, \overline{A}_{0:h}, A_{h+1} = a_{h+1}(q), \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$$

by consistency,

$$= P(Y_{h+1}, M_{h+1} | V, \overline{A}_{0:(h+1)} = \overline{a}_{0:(h+1)}(q), \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$$

$$P(\overline{Y}_{(h+2):(h+\tau)}(q), \overline{M}_{(h+2):(h+\tau)}(q) | V, \overline{A}_{0:(h+2)} = \overline{a}_{0:(h+2)}(q), \overline{Y}_{0:(h+1)}, \overline{M}_{0:(h+1)}, b_i^A)$$

by induction,

$$= \prod_{s=h}^{h+\tau-1} P(Y_{s+1}, M_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A)$$

account for and marginalize over patient heterogeneity,

$$= \prod_{s=h}^{h+\tau-1} \int_{u_s} \int_{v_s} P(Y_{s+1}, M_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^Y = u_s, b^M = v_s, b_i^A)$$
$$P(b^Y = u_s, b^M = v_s | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A) du_s dv_s$$

because counterfactual treatment path does not inform heterogeneity estimation,

$$= \prod_{s=h}^{h+\tau-1} \int_{u_s} \int_{v_s} P(Y_{s+1}, M_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^Y = u_s, b^M = v_s, b_i^A)$$
$$P(b^Y = u_s, b^M = v_s | V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A) du_s dv_s$$

by distributional assumptions illustrated in Fig 1,

$$= \prod_{s=h}^{h+\tau-1} \int_{u_s} \int_{v_s} P(Y_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^Y = u_s)$$
$$P(M_{s+1} | V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^M = v_s)$$
$$P(b^Y = u_s, b^M = v_s | V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A) du_s dv_s$$

parameterizing MGLMM as linear models, we get

$$= \prod_{s=h}^{h+\tau-1} \int_{u_s} \int_{v_s} P(Y_{s+1}|V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^Y = u_s; \beta^Y, \sigma^2)$$
$$P(M_{s+1}|V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^M = v_s; \beta^M)$$
$$P(b^Y = u_s, b^M = v_s|V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A; G) du_s dv_s$$

Given b_i^A, the g-formula for a conditional subgroup average treatment effect is defined as a conditional mean of the potential outcome at the end of follow-up at time $h + \tau$ under a user-specified regime q. It can then be derived as below,

$$\mathbb{E}(Y_{h+\tau}(q)|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$$
$$= \int_{y_\tau} y_\tau P(Y_{h+\tau}(q) = y_\tau|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A) dy_\tau$$
$$= \int_{y_{h+\tau}} \int_{m_{h+\tau}} \cdots \int_{y_{h+1}} \int_{m_{h+1}}$$
$$y_\tau P(\overline{Y}_{(h+1):(h+\tau)}(q) = \overline{y}_{(h+1):(h+\tau)}, \overline{M}_{(h+1):(h+\tau)}(q) = \overline{m}_{(h+1):(h+\tau)}|V, \overline{A}_{0:h}, \overline{Y}_{0:h}, \overline{M}_{0:h}, b_i^A)$$
$$dm_{h+1} dy_{h+1} \ldots dm_{h+\tau} dy_{h+\tau}$$
$$= \int_{y_{h+\tau}} \int_{m_{h+\tau}} \cdots \int_{y_{h+1}} \int_{m_{h+1}} y_\tau \Big\{ \prod_{s=0}^{\tau-1} \int_{u_s} \int_{v_s}$$
$$P(Y_{s+1} = y_{s+1}|V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^Y = u_s; \beta^Y, \sigma^2)$$
$$P(M_{s+1} = m_{s+1}|V, \overline{A}_{0:(s+1)} = \overline{a}_{0:(s+1)}(q), \overline{Y}_{0:s}, \overline{M}_{0:s}, b^M = v_s; \beta^M)$$
$$P(b^Y = u_s, b^M = v_s|V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A; G) du_s dv_s \Big\}$$
$$dm_{h+1} dy_{h+1} \ldots dm_{h+\tau} dy_{h+\tau}$$
$$= \int_{y_{h+\tau}} \int_{m_{h+\tau}} \int_{u_{\tau-1}} \int_{v_{\tau-1}} \cdots \int_{y_{h+1}} \int_{m_{h+1}} \int_{u_0} \int_{v_0}$$
$$y_\tau \Big\{ \prod_{s=0}^{\tau-1} P(Y_{s+1} = y_{s+1}|V, \overline{A}_{s+1} = \overline{a}_{s+1}^q, \overline{Y}_s = \overline{y}_s, \overline{M}_s = \overline{m}_s, b^Y = u_s; \beta^Y, \sigma^2)$$
$$P(M_{s+1} = m_{s+1}|V, \overline{A}_{s+1} = \overline{a}_{s+1}^q, \overline{Y}_s = \overline{y}_s, \overline{M}_s = \overline{m}_s, b^M = v_s; \beta^M)$$
$$P(b^Y = u_s, b^M = v_s|V, \overline{A}_{0:h}, \overline{Y}_{0:s}, \overline{M}_{0:s}, b_i^A; G) \Big\}$$
$$du_0 dv_0 dm_{h+1} dy_{h+1} \ldots du_{\tau-1} dv_{\tau-1} dm_{h+\tau} dy_{h+\tau}$$

The population average treatment effect conditional on b_i^A can be obtained by further integrating over the distribution of observed clinical history in the target population,

$$\mathbb{E}(Y_{h+\tau}(q)|b_i^A) = \int_v \int_{y_h} \int_{m_h} \cdots \int_{y_0} \int_{m_0} \mathbb{E}(Y_{h+\tau}(q)|V = v, \overline{A}_{0:h} = \overline{a}_{0:h}, \overline{Y}_{0:h} = \overline{y}_{0:h}, \overline{M}_{0:h} = \overline{m}_{0:h}, b_i^A)$$

$$P(V = v, \overline{A}_{0:h} = \overline{a}_{0:h}, \overline{Y}_{0:h} = \overline{y}_{0:h}, \overline{M}_{0:h} = \overline{m}_{0:h}) dm_0 dy_0 \ldots dm_h dy_h dv$$

The conditional mixed average treatment effect is computed as follows by substituting the target population distribution of the observable with the corresponding empirical distribution, $\widehat{P}(V = v, \overline{A}_{0:h} = \overline{a}_{0:h}, \overline{Y}_{0:h} = \overline{y}_{0:h}, \overline{M}_{0:h} = \overline{m}_{0:h})$.

$$\widehat{\mathbb{E}}(Y_{h+\tau}(q)|b_i^A) = \int_v \int_{y_h} \int_{m_h} \cdots \int_{y_0} \int_{m_0} \mathbb{E}(Y_{h+\tau}(q)|V = v, \overline{A}_{0:h} = \overline{a}_{0:h}, \overline{Y}_{0:h} = \overline{y}_{0:h}, \overline{M}_{0:h} = \overline{m}_{0:h}, b_i^A)$$
$$\widehat{P}(V = v, \overline{A}_{0:h} = \overline{a}_{0:h}, \overline{Y}_{0:h} = \overline{y}_{0:h}, \overline{M}_{0:h} = \overline{m}_{0:h}) dm_0 dy_0 \ldots dm_h dy_h dv.$$

Heterogeneity in treatment assignment, b_i^A, is assumed to be marginally $N(0, v)$ in the target population. The marginal mixed population average treatment effect can then be obtained by integrating b_i^A over its distribution $P(b_i^A = w)$ as $\widehat{\mathbb{E}}(Y_{h+\tau}(q)) = \int_w \widehat{\mathbb{E}}(Y_{h+\tau}(q)|b_i^A = w) P(b_i^A = w) dw$.

B Sequential Update of Random Effects

Assuming Gaussian distribution and logit model for continuous and binary variables, respectively, the structural model (**??**) can be written as follows ,

$$
\begin{aligned}
&Y_{it}|M_{it}=1 \sim \eta_{it}^{Y}+\sigma\psi^{Y},\\
&P(M_{it}=1)=\frac{\exp(\eta_{it}^{M})}{1+\exp(\eta_{it}^{M})},\\
&P(A_{it}=1|A_{i,t-1}=0)=\frac{\exp(\eta_{it}^{A})}{1+\exp(\eta_{it}^{A})},
\end{aligned}
$$

where $\psi^{Y} \sim N(0,1)$, $\eta_{it}^{Y}=\eta^{Y}(\mathcal{F}_{it},b_i^{Y};\theta^{Y})$, $\eta_{it}^{M}=\eta^{M}(\mathcal{F}_{it},b_i^{M};\theta^{M})$, and $\eta_{it}^{A}=\eta^{A}(\mathcal{F}_{it}^{A},b_i^{A};\theta^{A})$.

Sequential update for random effects is implemented for each individual, conditional on biomarker dynamics up to time t and observed treatment sequence up to time h, where $h \leq t$. For the observed trajectories of subject i, the joint likelihood is

$$
\begin{aligned}
&P(Y_{i,0:t},M_{i,0:t},A_{i,0:h}|b_i,\beta,\sigma)\\
&\propto \prod_{j=1}^{t}\left[\left(\frac{1}{\sigma}\exp\{-\frac{1}{2\sigma^2}(Y_{ij}-\eta_{ij}^{Y})^2\}\right)^{M_{ij}}\frac{\exp\{\eta_{ij}^{M}M_{ij}\}}{1+\exp(\eta_{ij}^{M})}\right]\times\prod_{j'=1}^{h}\left[\frac{\exp\{\eta_{ij'}^{A}A_{ij'}\}}{1+\exp(\eta_{ij'}^{A})}\right]^{\mathbb{1}(j'\leq s_i)},
\end{aligned}
$$

where s_i is the observed treatment initiation time for subject i, and the random effect b_i has prior

$$
P(b_i|G)\propto|G|^{-1/2}\exp(-\frac{1}{2}b_i^{T}G^{-1}b_i).
$$

The log posterior of b_i can then be written as

$$
\begin{aligned}
&\log P(b_i|Y_{i,0:t},M_{i,0:t},A_{i,0:h},\beta,\sigma,G)\\
&\propto -\frac{1}{2}b_i^{T}G^{-1}b_i+\sum_{j=1}^{t}\left\{-\frac{M_{ij}}{2\sigma^2}(Y_{ij}-\eta_{ij}^{Y})^2+\eta_{ij}^{M}M_{ij}-\log[1+\exp(\eta_{ij}^{M})]\right\}\\
&\quad+\sum_{j'=1}^{min(h,s_i)}\left\{\eta_{ij'}^{A}A_{ij'}-\log[1+\exp(\eta_{ij'}^{A})]\right\}.
\end{aligned}
$$

Using algorithms for constructing sampling chains, such as MCMC, in sampling b_i would consume a significant amount of computational resources due to the complexity of calculating counterfactual individual trajectories. We consider a Laplace approximation of the posterior distribution of b_i for an easier posterior sampling. The mean of the approximated distribution is obtained by solving the following equation for a posterior mode $\hat{b}_i=(\hat{b}_i^{Y},\hat{b}_i^{M},\hat{b}_i^{A})$,

$$
\frac{\partial}{\partial b_i}\log P(b_i|Y_{i,0:t},M_{i,0:t},A_{i,0:h},\beta,\sigma,G)\Big|_{b_i=\hat{b}_i}=0,
$$

where

$$\frac{\partial}{\partial b_i}\log P(b_i|Y_{i,0:t},M_{i,0:t},A_{i,0:h},\beta,\sigma,G) = -G^{-1}b_i + \begin{pmatrix} \sum_{j=1}^{t}\frac{M_{ij}}{\sigma_1^2}(Y_{ij}-\eta_{ij}^Y) \\ \sum_{j=1}^{t} M_{ij} - \frac{\exp(\eta_{ij}^M)}{1+\exp(\eta_{ij}^M)} \\ \sum_{j=1}^{min(h,s_i)} A_{ij} - \frac{\exp(\eta_{ij}^A)}{1+\exp(\eta_{ij}^A)} \end{pmatrix}.$$

The variance of the approximated distribution is the asymptotic variance of $\hat{b}_i$, which is the inverse of the observed Fisher information matrix defined as follows

$$V = \Bigg[-\frac{\partial^2}{\partial b_i \partial b_i^T}\log P(b_i|Y_{i,0:t},M_{i,0:t},A_{i,0:h},\beta,\sigma,G)\Big|_{b_i=\hat{b}_i}\Bigg]^{-1},$$

where

$$\begin{aligned}&\frac{\partial^2}{\partial b_i \partial b_i^T}\log P(b_i|Y_{i,0:t},M_{i,0:t},A_{i,0:h},\beta,\sigma,G)\\ =&-G^{-1} - \text{diag}\Bigg\{\frac{1}{\sigma^2}\sum_{j=1}^{t}M_{ij}, \sum_{j=1}^{t}\frac{\exp(\eta_{ij}^M)}{[1+\exp(\eta_{ij}^M)]^2}, \sum_{j=1}^{min(h,s_i)}\frac{\exp(\eta_{ij}^A)}{[1+\exp(\eta_{ij}^A)]^2}\Bigg\}.\end{aligned}$$

As a result, an approximation to the posterior distribution $P(b_i|Y_{i,0:t},M_{i,0:t},A_{i,0:h},\beta,\sigma,G)$ is the multivariate Gaussian distribution $MNV(\hat{b}_i,V)$.

Sequential update of counterfactual trajectories is also conditinoal on b_i^A being a constant, i.e. $b_i^A = c$. We sequentially update the heterogeneity in biomarker dynamics conditional on history $(Y_{i,0:t},M_{i,0:t},A_{i,0:h})$, population level estimates (β,σ,G), and $b_i^A = c$ as follows

$$(b_i^Y, b_i^M|b_i^A = c) \sim MNV(b_{\cdot|A}, V_{\cdot|A})$$

such that

$$b_{\cdot|A} = \begin{pmatrix}\hat{b}^Y \\ \hat{b}^M\end{pmatrix} + \begin{pmatrix}V^{Y,A} \\ V^{M,A}\end{pmatrix}(V^A)^{-1}(c-\hat{b}^A) \tag{7}$$

$$V_{\cdot|A} = \begin{pmatrix}V^Y & V^{Y,M} \\ & V^M\end{pmatrix} - \begin{pmatrix}V^{Y,A} \\ V^{M,A}\end{pmatrix}(V^A)^{-1}(V^{Y,A},V^{M,A}). \tag{8}$$

Suppose we are simulating the counterfactual progression of patient's longitudinal measures with treatment sequence fixed as $\overline{a}_{0:t}^q$ under regime q, where the sequence up to time h is the observed treatment, i.e. $A_{i,0:h} = \overline{a}_{0:h}^q$. If we write the third row of G^{-1} as (C_1,C_2,C_3), then the derivative entry relative to b_i^A leads to

$$\sum_{j=1}^{min(h,s_i)} a_j - \frac{\exp(X_{ij}\beta^A + b_i^A{}_{i0})}{1+\exp(X_{ij}\beta^A + b_i^A{}_{i0})} = C_1 b_i^A + C_2 b_i^M + C_3 b_i^Y, \tag{9}$$

and we can see that the specification of the counterfactual treatment sequence $\overline{a}_{(h+1):t}$ does not affect the estimation of $\hat{b}_i$. Note that $\sum_{j=1}^{min(h,s_i)} a_j$ is either 0 or 1, because the summation stops at the time of initiation. In the application, we focus on studying the effect of treatment initiation among those who were not treated before a time h, i.e. $h < s_i$ and $\sum_{j=1}^{min(h,s_i)} a_j = 0$. Hence, for the estimation of $\hat{b}_i$, equation (9) imposes condition $-\frac{\exp(X_{ij}\beta^A + b_i^A{}_{i0})}{1+\exp(X_{ij}\beta^A + b_i^A{}_{i0})} = C_1 b_i^A + C_2 b_i^M + C_3 b_i^Y$, using only treatment information before an treatment initiation.

C Pseudocode for Generating Counterfactual Trajectories

Algorithm for Dynamic Projection of Counterfactual Trajectories under MGLMM

Conditional on:

(a) observed history up to time h, $(V_i, \overline{Y}_{i,0:h}, \overline{M}_{i,0:h}, \overline{A}_{i,0:h})$

(b) posteriors of $(\theta^Y, \theta^M, \theta^A)$

(c) $var(b_i^A) = v$,

Goal: make posterior predictive inference of $(\overline{Y}_{(h+1):T}(q), \overline{M}_{(h+1):T}(q))$ under regime q.

Step 0: Initialization

(a) draw subject-specific stochastic matrices $\psi^Y, \psi^M \in \mathbb{R}^{N_{post} \times (T-h)}$, $\psi^Y \sim \mathcal{N}(0,1)$ and $\psi^M \sim \mathcal{U}(0,1)$

(b) $\mathcal{F}_{i,h+1}^{(\ell)}(q) = (V_i, \overline{Y}_{i,0:h}, \overline{M}_{i,0:h}, \overline{A}_{i,0:h}, a_{h+1}(q))$ for all ℓ

(c) draw $b_i^{A(\ell)} \sim N(0, v)$ for all ℓ

(d) $l = 0$

while $\ell < N_{post}$ **do**

for $t \in h+1, \ldots, T$ **do**

Step 1: Calculate $(\hat{b}_i^{(\ell)}(q), V_i^{(\ell)}(q))$ conditional on $(V_i, \overline{Y}_{i,0:(t-1)}^{(\ell)}(q), \overline{M}_{i,0:(t-1)}^{(\ell)}(q), \overline{A}_{i,0:h})$

Step 2: Draw $(b_i^{Y(\ell)}(q), b_i^{M(\ell)}(q)) | b_i^{A(\ell)} \sim MVN(b_{t|A}^{(\ell)}(q), V_{t|A}^{(\ell)}(q))$, where

$b_{t|A}^{(\ell)}(q)$ and $V_{t|A}^{(\ell)}(q)$ are obtained by (7) and (8), respectively.

Step 3: Update $M_{it}^{(\ell)}(q)$

let $p_{it}^{(\ell)}(q) = \text{logit}^{-1}\eta^M(\mathcal{F}_{it}^{(\ell)}(q), b_i^{M(\ell)}(q); \theta^{M(\ell)})$

draw $M_{it}^{(\ell)}(q) \sim \text{Bernoulli}(p_{it}^{(\ell)}(q))$ by setting $M_{it}^{(\ell)}(q) = \mathbb{1}\{\psi_{\ell,t-h}^M \leq p_{it}^{(\ell)}(q)\}$

Step 4: Update $Y_{it}^{(\ell)}(q)$

draw $Y_{it}^{(\ell)}(q) \sim f_Y(\eta_{it}^{Y(\ell)}(q), (\sigma^{(\ell)})^2)$ by

setting $\eta_{it}^{Y(\ell)}(q) = \eta^Y(\mathcal{F}_{it}^{(\ell)}(q), b_i^{Y(\ell)}(q); \theta^{Y(\ell)})$ and $Y_{it}^{(\ell)}(q) = \eta_{it}^{Y(\ell)}(q) + \sigma^{(\ell)}\psi_{\ell,t-h}^Y$

Step 5: Define

$$\mathcal{F}_{i,t+1}^{(\ell)}(q) = (V_i, \overline{Y}_{i,0:t}^{(\ell)}(q), \overline{M}_{i,0:t}^{(\ell)}(q), \overline{A}_{i,0:(t+1)}(q)),$$

where

$$\begin{aligned} \overline{Y}_{i,0:t}^{(\ell)}(q) &= (\overline{Y}_{i,0:h}, \overline{Y}_{i,(h+1):t}^{(\ell)}(q)) \\ \overline{M}_{i,0:t}^{(\ell)}(q) &= (\overline{M}_{i,0:h}, \overline{M}_{i,(h+1):t}^{(\ell)}(q)) \\ \overline{A}_{i,0:(t+1)}(q) &= \overline{a}_{0:(t+1)}(q) \end{aligned}$$

and the observed equals the counterfactual during the given history, i.e. $\overline{A}_{i,0:h} = \overline{a}_{0:h}(q)$.

end for

end while

Step 6: $\{\overline{Y}_{i,0:t}^{(\ell)}(q), \overline{M}_{i,0:t}^{(\ell)}(q); \ell = 1, \ldots, N_{post})\}$ are samples from the posterior predictive distribution of $(\overline{Y}_{(h+1):T}(q), \overline{M}_{(h+1):T}(q))$ under regime q.

D Connection with Structural Nested Models

Sitlani et al.[33] and Qian et al.[23] studied instantaneous treatment effect as the "blip" of a structural

nested model (SNM), using linear mixed models as the structural model and comparing treatment paths that only differ in the treatment status at a specific time m, i.e. comparing $A_m = 1$ versus $A_m = 0$ in the case of binary and monotone treatment. Our proposal, on the other hand, compares the effect of treatment paths under different regimes, i.e. $\overline{A}_{0:t}$ being $\overline{a}_{0:t}(q_1)$ versus $\overline{a}_{0:t}(q_2)$, where q_1 and q_2 are the regimes of interest. The motivating application investigates the treatment effect of taking a drug continuously over time, where the causal effect is cumulative over time and thus requires a fundamentally different characterization than a structural model approach. The instantaneous treatment effect, or the blip, can be characterized under our framework as the average causal effect comparing q_1 and q_2 where $a_t(q_1) = a_t(q_2)$ for $t \neq m$, $a_m(q_1) = 1$ and $a_m(q_2) = 0$. Specifically, assuming $\phi_A(\overline{A}_{i,0:t}) = A_{it}$, $\phi_4^Y(\mathcal{H}_{it}) = 0$, $\tau = 1$, and a linear mixed model for a continuous outcome leads to a special case in Qian et al.[23] , where we will have the instantaneous subgroup treatment effect at $h+1$ conditional on information up to time h being

$$\mathbb{E}(Y_{i,h+1}|V_i, A_{i,h+1} = 1, \overline{A}_{i,0:h}, \mathcal{H}_{ih}) - \mathbb{E}(Y_{i,h+1}|V_i, A_{i,h+1} = 0, \overline{A}_{i,0:h}, \mathcal{H}_{ih}) = \phi_2^Y(H_{it})\beta_2^Y. \quad (10)$$

Thus, the model parameter β_2^Y has a causal interpretation marginally over the subgroup defined by $(V_i, \overline{A}_{i,0:h}, \mathcal{H}_{ih})$ in this case and the MGLMM reduces to a linear structural mixed model. However, when $\phi_4^Y(\mathcal{H}_{it}) \neq 0$, equation (10) is no longer true because β_2^Y only remains with a causal interpretation conditional on b_i^Y, as showed in the conditional subgroup causal effect below,

$$\begin{aligned}\mathbb{E}(Y_{i,h+1}|V_i, A_{i,h+1} = 1, \overline{A}_{i,0:h}, \mathcal{H}_{ih}, b_i^Y) - \mathbb{E}(Y_{i,h+1}|V_i, A_{i,h+1} =& 0, \overline{A}_{i,0:h}, \mathcal{H}_{ih}, b_i^Y) \\ =& \phi_2^Y(H_{it})\beta_2^Y + \phi_4^Y(H_{it})b_{i1}^Y, \quad (11)\end{aligned}$$

and the conditional expectation $\mathbb{E}(b_i^Y|V_i, \overline{A}_{i,0:h}, \mathcal{H}_{ih})$ is not necessarily zero.

E Connection with Shardell and Ferrucci[32]

Shardell and Ferrucci[32] demonstrated longitudinal causal inference using joint mixed-effects models, assuming shared random effects between the model components for the outcome, confounders, and treatment assignment. Their model specification is similar to the MGLMM in Section 2, i.e. with $b_{i0}^A = b_{i0}^M = (b_{i0}^Y, b_{i1}^Y)$ and $\phi_4^M \equiv 0$, but distinctively different in that ϕ_2^A and ϕ_3^M are population-level coefficients instead of observed variables. Shardell and Ferrucci[32] assumed sequential exchangeability conditional on the unobserved heterogeneity in the outcome progression, i.e. (b_{i0}^Y, b_{i1}^Y), which is

assumed to be proportionate to the heterogeneity in confounders and treatment assignment. Whereas we account for unobserved time-invariant traits in treatment assignment with the random effect b_i^A, assuming that it is correlated with (b_{i0}^Y, b_{i1}^Y) and having the sequential exchangeability conditional on b_i^A instead of (b_{i0}^Y, b_{i1}^Y).

The assumption of no unmeasured confounders in the model of Shardell and Ferrucci [32] implies no treatment assignment heterogeneity. While assuming no treatment heterogeneity under MGLMM is equivalent to setting $v = 0$, which is a sufficient but unnecessary condition for having no unmeasured confounders. In MGLMM, $\text{cov}(b_i^A, b_i^M) = \text{cov}(b_i^A, b_i^Y) = 0$ leads to no unmeasured confounders. That is, even when no unmeasured confounders is true, MGLMM still allows treatment assignment heterogeneity as long as it is not correlated with the unobserved heterogeneity in biomarker dynamics (b_i^Y, b_i^M); examples of such unconfounding treatment assignment heterogeneity include a patient's preference for a treatment based on personal beliefs or social stigma. On the other hand, we note that the covariances $\text{cov}(b_i^A, b_i^M)$ and $\text{cov}(b_i^A, b_i^Y)$ are estimable given v, the presumed variance of b_i^A. Henceforth, our method does partially inform the possible existence of unmeasured confounders based on the estimated covariances in the MGLMM.

When there is no treatment assignment heterogneiety, both Shardell and Ferrucci [32] and our method simplify to the standard g-computation of fitting only the outcome and confounders model using generalized linear mixed-effects model because the assignment mechanism becomes ignorable [19]. Let us consider a simplified scenario of looking at the subgroup ATE at time $h+1$ conditional on history information up to time h, assuming no time-varying confounders and no treatment assignment heterogeneity. The subgroup ATE would not be identifiable under Shardell and Ferrucci [32]. The reason is as follows. Under their conditional sequential exchangeabiltiy assumption

$$Y_{h+1}(q) \perp A_{h+1} | V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y,$$

we can directly identify the conditional counterfactual distribution as

$$P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) = P(Y_{h+1}|V, \bar{A}_{0:(h+1)} = \overline{a}_{0:(h+1)}(q), \overline{Y}_{0:h}, b_i^Y).$$

However, the target quantity represented by $P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h})$ would not be calculable because

$$P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) = \int P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) db_i^Y$$

and the subgroup heterogeneity distribution $P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h})$ is unknown. Whereas with our proposal, we assume a different conditional sequential exchangeabiltiy assumption

$$Y_{h+1}(q) \perp A_{h+1}|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^A.$$

Given the assumption of no treatment assignment heterogeneity, we know $\text{var}(b_i^A) = 0$ and consequently $\text{cov}(b_i^A, b_i^Y) = 0$, leading to $P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^A) = P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h})$. As a result, the target quantity is identifiable via (6) as

$$P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) = \int P(Y_{h+1}(q)|V, \bar{A}_{0:h}, \bar{Y}_{0:h}, b_i^Y) P(b_i^Y|V, \bar{A}_{0:h}, \bar{Y}_{0:h}) db_i^Y.$$

Our proposal may be viewed as an extension of Shardell and Ferrucci's[32] work in the following aspects: (1) a softer assumption on the conditional sequential exchangeability, stratifying by b_i^A instead of the random effects shared across the outcome, confounders, and treatment model, (2) model specification as MGLMM, which is more generalized and has the potential to include their joint mixed-effects model as a special case, and (3) allows the identification of subgroup causal effects when assuming no treatment assignment heterogeneity. The merits of this extension come at a price of introducing the variance of b_i^A as a sensitivity parameter, and identifying subgroup causal effects under the existence of treatment assignment heterogeneity needs to be done under additional assumption on the subgroup distribution of b_i^A, which likely requires expert knowledge.