Treatment Effect Bias in Multilevel Linear Models under Time-Varying Endogenous Covariates: A New Look at Qian et al. (2020)

Research Report

Ward B. Eiling (9294163)

Supervisors: Ellen Hamaker and Jeroen Mulder

Master's degree in Methodology and Statistics for the Behavioural, Biomedical and Social Sciences

Utrecht University

January 17, 2025

Word count: 2492/2500

FETC-approved: 24-2003

Candidate journal: Psychological Methods

1 Introduction

Across a wide range of disciplines, researchers analyze clustered longitudinal, observational data to investigate prospective causal relationships between variables. When analyzing such data, psychological researchers most commonly use the multilevel linear model[1] (MLM, Bauer & Sterba, 2011), which—in the context of longitudinal data analysis—partitions observed variance into stable between-person differences and within-person fluctuations (Hamaker & Muthén, 2020). Research questions explored with the MLM often result in the availability of time invariant and/or time-varying covariates, the latter measured repeatedly over time. The inclusion of covariates is a common strategy to improve parameter precision (Boruvka et al., 2018) and address bias introduced by confounders (Robins et al., 2000). Nevertheless, this approach is not universally beneficial, as conditioning on endogenous covariates—those influenced by (prior) treatment/exposure or outcome—can create challenges for standard methods like MLMs, which implicitly assume the exogeneity of covariates (Erler et al., 2019).

Dating back to the work of Pepe and Anderson (1994) in the biostatistics literature, this assumption has been shown to be non-trivial when endogenous covariates are time-varying. In fact, their inclusion in longitudinal studies can lead to biased treatment effect estimates, an issue that, despite its significance, has received limited attention in psychological research. Building on this foundation, a recent paper by Qian et al. (2020) examined the suitability of MLM for estimating the causal effect of a time-varying treatment. Specifically, they focused on settings where the exposure is randomly assigned at each occasion within individuals. Such randomized treatments may include, for example, prompts delivered through push notifications to remind participants of mindfulness-based strategies (Nahum-Shani et al., 2021; Walton et al., 2018). While random assignment with a constant probability might seem sufficient to identify (the presence and absence of) causal effects, Qian et al. (2020) showed that model fitting issues and

[1]The MLM is known by various names in different substantive fields, including: linear mixed model, hierarchical linear model, random-effect model and mixed-effects model.

parameter bias can arise when a *time-varying endogenous covariate* is present.

However, due to a divide between the disciplines that employ the MLM, such critiques appear to have largely failed to reach the applied researcher in psychology. One specific reason might be that the technical jargon in other disciplines makes it difficult for researchers to recognize when and how these issues emerge. This report aims to explore why Qian et al. (2020) observed biased estimates of the treatment effect in certain data-generating mechanisms containing endogenous covariates, while not for others. Additionally, it seeks to explain this issue to an audience of psychologists. The study will first employ graphical diagrams to assess two criteria across various models involving an endogenous time-varying covariate and randomized treatment: (a) path diagrams to evaluate the conditional independence assumption introduced by Qian et al. (2020) and (b) directed acyclic graphs (DAGs) to assess the backdoor criterion (Pearl, 1988). Subsequently, data simulations based on Qian et al. (2020)'s original scenarios, along with additional ones, will be performed to reproduce and isolate the underlying issue and evaluate whether these criteria can predict the presence of bias in the treatment effect. The following research question will be addressed: *When does the inclusion of endogenous variables in multilevel linear models result in biased estimates of the treatment effect?*

2 Methods

In this section, we describe four generative models (GMs) that incorporate a time-varying endogenous covariate and randomized treatment, followed by the methodology used to evaluate treatment effect bias across settings.

2.1 Data Generation

We consider two GMs from Qian et al. (2020), one (GM-A) being a special case of the general model (GM-G) where bias was detected. To further isolate the source of bias, we introduce two additional special cases, GM-B and GM-C. We first describe the GM-G in detail, and then

proceed to its three special cases.

2.1.1 General Model

Following the original notation of Qian et al. (2020), the outcome of GM-G was generated as:

$$Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + \beta_1 X_{it} + b_{i2}) + \epsilon_{it+1} \tag{1}$$

where Y_{it+1} is the outcome, X_{it} the covariate, A_{it} the treatment, b_{i0} the random intercept, b_{i2} the random slope for treatment, and ϵ_{it+1} the error term. The observed variables vary across individuals i and timepoints t. Alternatively, the model can be rewritten in the multilevel notation of Raudenbush and Bryk (2002), with at the within-person level (level 1):

$$\begin{aligned} Y_{it+1} &= \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + \beta_1 X_{it} + b_{i2}) + \epsilon_{it+1} \\ &= (\alpha_0 + b_{i0}) + \alpha_1 X_{it} + (\beta_0 + b_{i2}) A_{it} + \beta_1 A_{it} X_{it} + \epsilon_{it+1} \\ &= \pi_{0i} + \pi_{1i} X_{it} + \pi_{2i} A_{it} + \pi_{3i} A_{it} X_{it} + \epsilon_{it+1}. \end{aligned}$$

and at the between-person level (level 2):

$$\begin{aligned} \pi_{0i} &= \alpha_0 + b_{i0}, \quad \text{where} \quad b_{i0} \sim \mathcal{N}(0, \sigma^2_{b0}), \\ \pi_{1i} &= \alpha_1, \\ \pi_{2i} &= \beta_0 + b_{i2}, \quad \text{where} \quad b_{i2} \sim \mathcal{N}(0, \sigma^2_{b2}), \\ \pi_{3i} &= \beta_1. \end{aligned}$$

The parameters $\alpha_0 = -2$, $\alpha_1 = -0.3$, $\beta_0 = 1$, and $\beta_1 = 0.3$ are fixed effects that are constant across individuals. Conversely, b_{i0} and b_{i2} are independent random effects that capture individual-specific deviations from population parameters α_0 and β_0 respectively. The presence of the interaction term β_1 implies treatment heterogeneity: the effect of A_{it} on Y_{it+1} depends on the value of X_{it}. The random intercept b_{i0}, random slope b_{i2} and exogenous noise ϵ_{it+1} are

assumed to be normally distributed with mean zero and variance $\sigma^2_{b0} = 4$, $\sigma^2_{b2} = 1$ and $\sigma^2_{\epsilon} = 1$, respectively.

The covariate is generated as:

$$X_{it} = \begin{cases} b_{i0} + \epsilon_{X_{it}} & \text{if } t = 1, \\ b_{i0} + Y_{it} + \epsilon_{X_{it}} & \text{if } t \geq 2, \end{cases} \quad \text{where} \quad \epsilon_{X_{it}} \sim \mathcal{N}(0, 1)$$

The treatment randomization probability is constant at $p_t = 0.5$, so $A_{it} \sim$ Bernoulli(0.5), implying that for every i and t, the probability that treatment is assigned is equivalent to a fair coinflip. Relationships between variables are illustrated in Figure 1a.

2.1.2 Special Cases

We consider three special cases of GM-G, namely GM-A, GM-B and GM-C. The relation of each special case to GM-G is summarized in Table 1.

Table 1: Summary of Differences Between Generative Models

Generative Model	Name in Qian et al. (2020)	dependency b_{i0} and X_{it}	random slope treatment b_{i2}	interaction β_1
G(eneral)	3	✓	✓	✓
A	1	×	✓	✓
B	NA	✓	×	✓
C	NA	✓	✓	×

GM-A is a special case of GM-G, where the effect of b_{i0} on X_{it} is set to zero, which implies that X_{it} is not directly determined by b_{i0} (see Figure 1b). Instead, X_{it} is given by:

$$X_{it} = \begin{cases} \epsilon_{X_{it}} & \text{if } t = 1, \\ Y_{it} + \epsilon_{X_{it}} & \text{if } t \geq 2, \end{cases} \quad \text{where} \quad \epsilon_{X_{it}} \sim \mathcal{N}(0, 1)$$

GM-B is a special case of GM-G in which b_{i2} was removed (see Figure 1c) by setting $\sigma^2_{b2} = 0$. While the within-person model is the same as GM-G, there is a slight alteration at the between-person level:

$$\pi_{2i} = \beta_0.$$

The composite model then becomes:

$$Y_{it+1} = (\alpha_0 + b_{i0}) + \alpha_1 X_{it} + \beta_0 A_{it} + \beta_1 A_{it} X_{it} + \epsilon_{it+1} \tag{2}$$

GM-C is a special case of GM-G, where we set $\beta_1 = 0$, which implies the removal of $\beta_1 A_{it} X_{it}$ (see Figure 1d) and π_{3i}. Therefore, the within-person model from GM-G changes into:

$$Y_{it+1} = \pi_{0i} + \pi_{1i} X_{it} + \pi_{2i} A_{it} + \epsilon_{it+1}.$$

The composite model becomes:

$$Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + b_{i2}) + \epsilon_{it+1}. \tag{3}$$

2.2 Data Analysis

Data generation and estimation were performed in `R`, version 4.4.2 (Team, 2024), following Qian et al. (2020)'s methodology for consistency. After data generation, analytical MLMs with restricted maximum likelihood estimation were fit using the `lme4` package (Bates et al., 2015). The MLM from Equation 1 was fit for GM-G and A, the MLM from Equation 2 for GM-B, and the MLM from Equation 3 for GM-C[2].

In the simulation study, we evaluated the bias of the analytical models across different settings by systematically varying 3 factors: Generative Model (GM = G, A, B, C); number

[2]Contrary to the data generating models, the analytical models do not model the covariate X_{it} and the treatment A_{it}.

of timepoints ($T = 10, 30$); and sample size ($N = 30, 100, 200$). By varying these factors, 24 unique settings were created, each replicated 10,000 times. Bias estimates were calculated as the difference between the mean of the estimated parameter values $\bar{\beta}_0$ (across replications) and the true treatment effect $\beta_0 = 1$.

3 Results

This section begins by formulating predictions about treatment effect bias for each GM based on the conditional independence assumption and the backdoor criterion. These predictions are compared to simulation results, which present the estimated bias across the GMs.

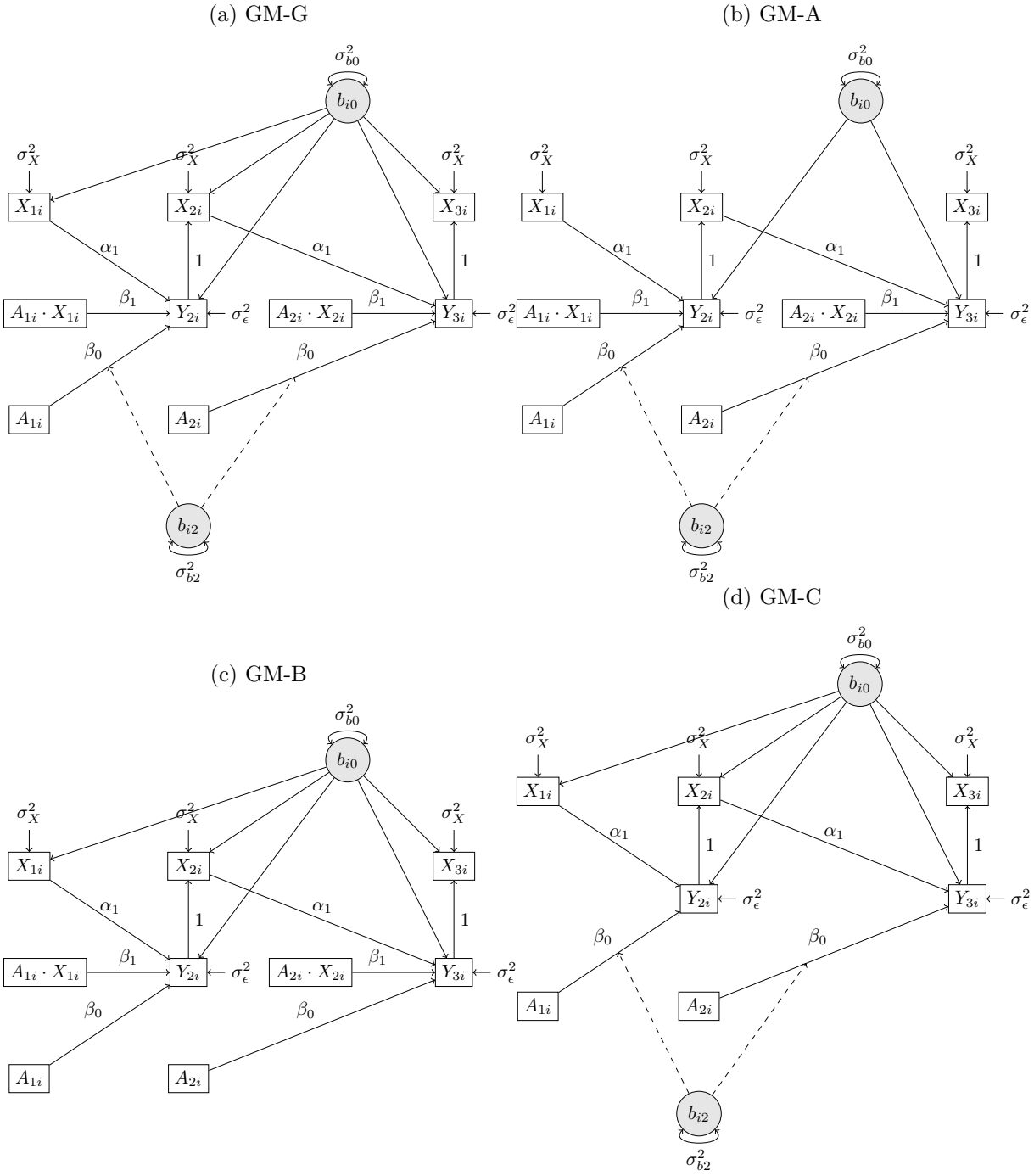
3.1 Conditional Independence and Path Diagrams

The *conditional independence assumption*, introduced by Qian et al. (2020) and based on Sitlani et al. (2012) evaluates whether treatment effect estimators are consistent and unbiased under randomized treatment assignment. This assumption states that the covariate at time t (X_{it}) should be independent of the individual's random effects (intercept b_{i0} and slope(s) b_{i1}) once we account for their history of covariates up to timepoint $t-1$ (H_{it-1}), previous treatment (A_{it-1}), and prior outcome (Y_{it}):

$$X_{it} \perp (b_{i0}, b_{i1}) \mid H_{it-1}, A_{it-1}, Y_{it}.$$

This implies that X_{it} can be influenced by prior observed variables but not directly by random effects. Verification of this assumption relies on domain knowledge (Qian et al., 2020). To illustrate its application, we constructed path diagrams (Wright, 1934) (see Figure 1).

Figure 1: Path Diagrams for Generative Models G, A, B and C (t = 1, 2, 3)



Note. Random effects are represented by grey circles, observed variables by squares and relationships across variables by arrows, where dashed lines are reserved for random slopes.

In GM-G, the X_{it} is directly influenced by b_{i0}. Consequently, conditioning on prior variables (e.g., Y_{it}), does not fully block the influence of the random effects, violating the assumption. This

aligns with the bias identified by Qian et al. (2020). In GM-A, the direct link between X_{it} and b_{i0} is removed. Although an indirect link remains between X_{it} and b_{i0} through Y_{it}, conditioning on Y_{it} "breaks the link" between X_{it} and b_{i0}, satisfying the assumption. This aligns with the absence of bias found by Qian et al. (2020). For GM-B and GM-C, the direct link between X_{it} and b_{i0} remains, as in GM-G, violating the assumption and suggesting the presence of bias.

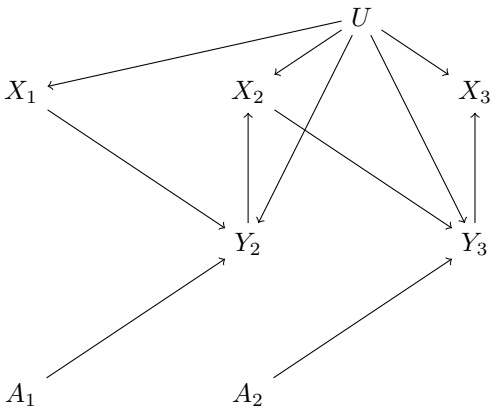
3.2 Backdoor Criterion and DAGs

The *backdoor criterion* (Pearl, 1988) can help determine which variables to control for to avoid bias in treatment effect estimates. To detect backdoor paths, DAGs (Pearl, 1995) are invaluable tools[3]. Unlike linear path diagrams (Wright, 1934) and structural equation models, DAGs are non-parametric. They encode causal assumptions about the data-generating process, with arrows representing direct causal effects that may exhibit heterogeneity across individuals or depend on other variables (effect interaction/modification) (Elwert & Winship, 2014). Consequently, DAGs do not explicitly represent random slopes and interaction effects, which precluded their use for evaluating the conditional independence assumption.

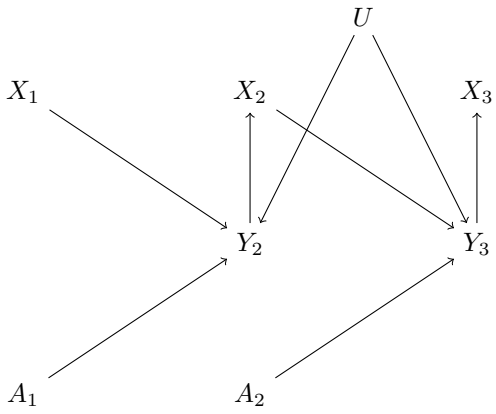
The DAGs for each GM (shown in Figure 2) confirm the absence of random slopes and fixed interaction effects, explaining why GM-G, GM-B, and GM-C are equivalent.

Figure 2: DAGs for Generative Models G, A, B and C (t = 1, 2, 3)

(a) GM-G, GM-B, GM-C



(b) GM-A



[3]An accessible introduction into DAGs and backdoor paths can be found in Rohrer (2018).

Note. The node U represents the random disturbance b_{0i} in the generative models.

Application of the backdoor criterion begins by identifying all paths that connect A_t to Y_{t+1}, excluding the direct treatment effect $A_t \rightarrow Y_{t+1}$ itself. Next, we determine whether these paths can transmit association by applying the d-separation rules[4]. A path that contains an arrow pointing to the treatment A_t is considered a *backdoor path.*

In the DAGs presented in Figure 2, no paths connect A_t to Y_{t+1} except the treatment effect $A_t \rightarrow Y_{t+1}$ itself. Thus, A_t has no parent nodes, which implies that no backdoor paths exist[5]. Therefore, X_t need not be controlled for to obtain an unbiased causal effect. However, including X_t does not introduce bias as it lies outside the pathways connecting A_t and Y_{t+1}[6].

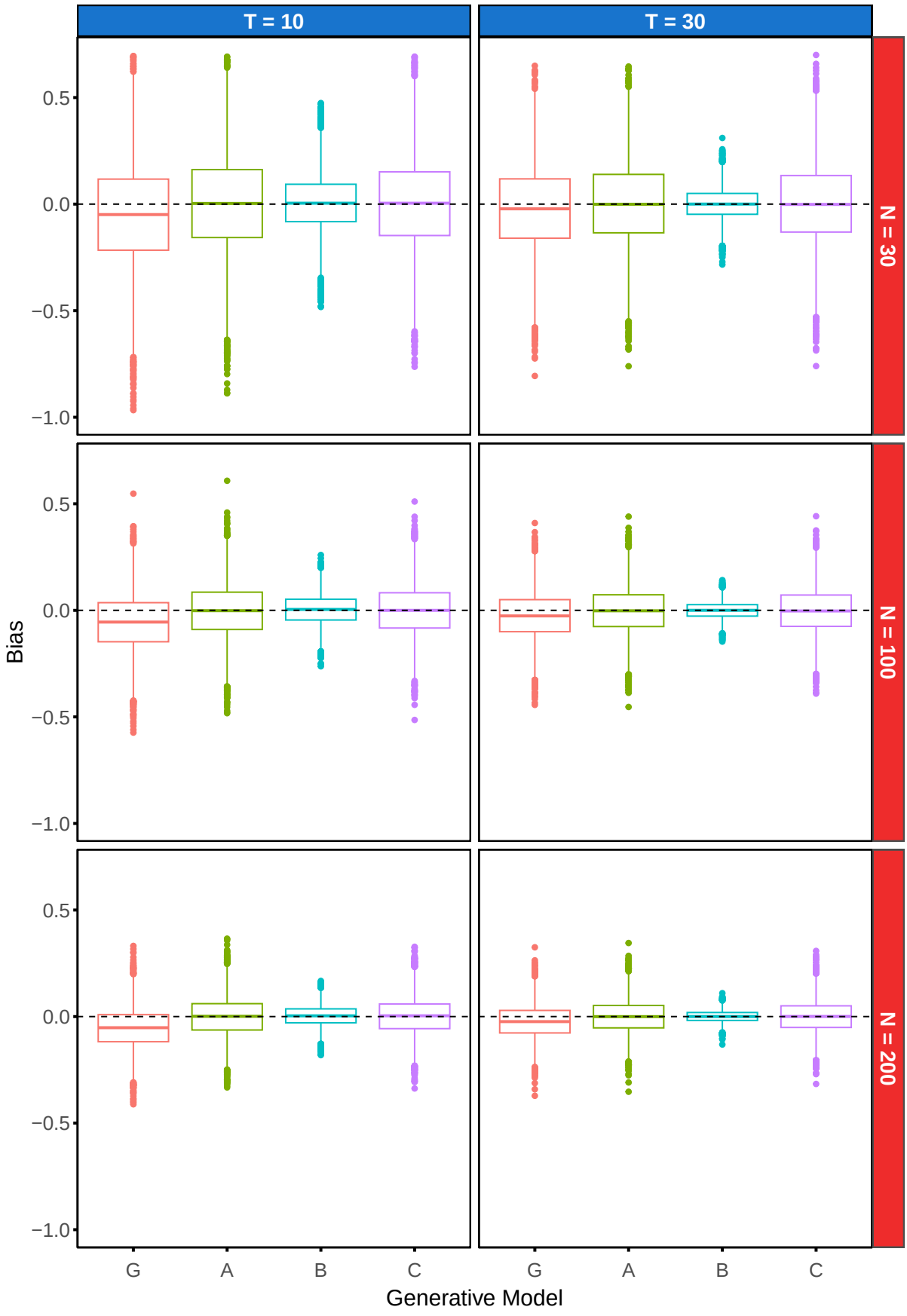
3.3 Simulation Study

Figure 3 present the bias estimates in β_0 for each of the GMs[7]. In this reproduction of Qian et al. (2020), we observed results consistent with the original study: GM-G exhibited bias ($0.023 - 0.056$ or $2.3\% - 5.6\%$), while GM-A showed none (≤ 0.002 or 0.2%). These findings align with the conditional independence assumption but contradict the backdoor criterion, which predicted no bias for any GM. As shown in Figure 3, GM-G bias decreased with increasing T but remained stable across varying N, indicating inconsistency (i.e., estimates do not converge to the true value as N increases).

Similar to GM-A, no discernible bias was observed for GM-B (≤ 0.005 or 0.5%) and GM-C (≤ 0.003 or 0.3%). These results align with the backdoor criterion, but contradict the conditional independence assumption, which predicted bias for both GMs. GM-B also exhibited much lower variability across simulations compared to the other GMs.

[4]For an accessible introduction to d-separation, see Hayduk et al. (2003).
[5]Indeed, the random assignment of the treatment ensures, by design, the absence of backdoor paths.
[6]Inclusion of the covariate may, however, increase the power to detect the treatment effect.
[7]The supplemental table with additional information can be found at the OSF repository: https://osf.io/8xawt/?view_only=aad6a13b7a4a4d36aed76ed8aac584c4

Figure 3: Estimation bias for the fixed treatment effect β_0 of each generative model for different combinations of sample size N and number of timepoints T over 10,000 simulation replications



Note. bias refers to the difference between the mean of the estimated parameter values $\bar{\beta}_0$ and the prespecified treatment effect $\beta_0 = 1$.

In summary, if the GM either does not include the direct dependency of the random intercept on the covariate (GM-A), the random slope b_{i2} (GM-B), or the interaction term β_1 (GM-C), as in GM-G, bias disappears. However, neither the backdoor criterion nor the conditional independence assumption consistently predicted treatment effect bias across models.

4 Discussion

In this report, we evaluated several GMs to investigate when endogenous time-varying covariates bias treatment effect estimates in MLMs under randomized treatment; and assessed the ability of the conditional independence assumption and backdoor criterion to predict when this bias will occur. We found biased and inconsistent estimates in the most general model (GM-G) but not in three special cases: GM-A (no direct effect of the random intercept on the covariate), GM-B (no random slope for treatment), and GM-C (no interaction between treatment and covariate). The conditional independence assumption correctly predicted bias in GM-G and GM-A, but not in GM-B and GM-C. The backdoor criterion predicted no bias in any of the GMs, contradicting the bias observed in GM-G. These findings suggest that the dependency between the covariate and treatment, the random slope for treatment, and the interaction effect are all essential for bias to occur.

This raises two questions: why was no discernible bias observed in GM-B and GM-C, as predicted by the conditional independence assumption, and can this pattern generalize beyond the current GMs? One possibility is that the interaction and random slope, both of which drive heterogeneity in the treatment effect, may need to coexist for bias to manifest. This aligns with Qian et al. (2020)'s assertion that "applying linear mixed models is problematic because potential moderators of the treatment effect are frequently endogenous" (p. 375). If treatment

effect moderation by a covariate is indeed a prerequisite for bias, it could explain the lack of bias in GM-C (no interaction). However, this does not account for the absence of bias in GM-B (no random slope). As the current study only tested a specific set of GMs, further research is needed to assess whether these findings generalize, providing practical recommendations for using MLMs with endogenous time-varying covariates.

Regarding the backdoor criterion and DAGs (Pearl, 1988), our results suggest standard non-parametric DAGs may be insufficient to identify bias in GM-G, since they do not explicitly represent random slopes and interaction effects[8]. Similar concerns about using DAGs in situations involving interaction effects have been raised (Attia et al., 2022; Weinberg, 2007). Future research could explore how extended DAG frameworks, that explicitly incorporate interaction effects, may allow the backdoor criterion to identify bias.

Another avenue for future investigation is the role of centering approaches[9]. According to Antonakis et al. (2021), the assumption of uncorrelated between random effects and level 1 covariates can be relaxed using Mundlak's contextual model[10] (Mundlak, 1978), which adds cluster means of each covariate as predictors of the random intercept. This approach, which explicitly models the source of endogeneity as suggested by Bell and Jones (2015), could shed light on the treatment effect bias present in GM-G but absent in the other GMs.

Finally, Qian et al. (2020) assumed independence between the random intercept and slope and perfect correlation between the random intercepts of the outcome and covariate. Since these assumptions may not hold in practice, investigating scenarios with different assumptions on the random effects could yield further insights.

[8]Note that the term "effect modification", while often used interchangeably with "interaction", has a distinct definition in the counterfactual framework (VanderWeele, 2009).

[9]An overview of centering approaches and their effects on interpretation in the MLM can be found in Hamaker and Muthén (2020).

[10]This is referred to as the Correlated Random Effects (CRE) approach by Wooldridge (2002).

5 References

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, *24*(2), 443–483. https://doi.org/10.1177/1094428119877457

Attia, J., Holliday, E., & Oldmeadow, C. (2022). A proposal for capturing interaction and effect modification using DAGs. *International Journal of Epidemiology*, *51*(4), 1047–1053. https://doi.org/10.1093/ije/dyac126

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 148. https://doi.org/10.18637/jss.v067.i01

Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological Methods*, *16*(4), 373–390. https://doi.org/10.1037/a0025813

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, *3*(1), 133–153. https://doi.org/10.1017/psrm.2014.7

Boruvka, A., Almirall, D., Witkiewitz, K., & Murphy, S. A. (2018). Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association*, *113*(523), 1112–1121. https://doi.org/10.1080/01621459.2017.1305274

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53. https://doi.org/10.1146/annurev-soc-071913-043455

Erler, N. S., Rizopoulos, D., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2019). Bayesian imputation of time-varying covariates in linear mixed models. *Statistical Methods in Medical Research*, *28*(2), 555–568. https://doi.org/10.1177/0962280217730851

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*(3), 365–379. https:

//doi.org/10.1037/met0000239

Hayduk, L., Cummings, G., Stratkotter, R., Nimmo, M., Grygoryev, K., Dosman, D., Gillespie, M., Pazderka-Robinson, H., & Boadu, K. (2003). Pearl's d-separation: One more step into causal thinking. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(2), 289–311. https://doi.org/10.1207/S15328007SEM1002_8

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69–85. https://doi.org/10.2307/1913646

Nahum-Shani, I., Potter, L. N., Lam, C. Y., Yap, J., Moreno, A., Stoffel, R., Wu, Z., Wan, N., Dempsey, W., Kumar, S., & al., et. (2021). The mobile assistance for regulating smoking (MARS) micro-randomized trial design protocol. *Contemporary Clinical Trials*, *110*, 106513. https://doi.org/10.1016/j.cct.2021.106513

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688. https://doi.org/10.1093/biomet/82.4.669

Pepe, M. S., & Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics - Simulation and Computation*, *23*(4), 939–951. https://doi.org/10.1080/03610919408813210

Qian, T., Klasnja, P., & Murphy, S. A. (2020). 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), 375–390. https://doi.org/10.1214/19-sts720

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). SAGE.

Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal structural models and causal

inference in epidemiology. *Epidemiology*, *11*(5), 550. https://journals.lww.com/epidem/fulltext/2000/09000/marginal_structural_models_and_causal_inference_in.11.aspx

Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Sitlani, C. M., Heagerty, P. J., Blood, E. A., & Tosteson, T. D. (2012). Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Statistics in Medicine*, *31*(16), 1738–1760. https://doi.org/10.1002/sim.4510

Team, R. C. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology*, *20*(6), 863. https://doi.org/10.1097/EDE.0b013e3181ba333c

Walton, A., Nahum-Shani, I., Crosby, L., Klasnja, P., & Murphy, S. (2018). Optimizing digital integrated care via micro-randomized trials. *Clinical Pharmacology & Therapeutics*, *104*(1), 53–58. https://doi.org/10.1002/cpt.1079

Weinberg, C. R. (2007). Commentary: Can DAGs clarify effect modification? *Epidemiology*, *18*(5), 569–572. https://www.jstor.org/stable/20486428

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

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*(3), 161–215. https://doi.org/10.1214/aoms/1177732676

6 Appendix

Table 2: Treatment effect bias for Generative Models G, A, B and C over 10,000 replications

			β_0			
GM	T	N	Bias	SD	MC-SE	SR
G	10	30	-0.052	0.247	0.002	0.999
		100	-0.056	0.135	0.001	1.000
		200	-0.054	0.094	0.001	1.000
	30	30	-0.023	0.205	0.002	0.998
		100	-0.025	0.112	0.001	0.998
		200	-0.024	0.078	0.001	0.997
A	10	30	0.000	0.238	0.002	0.999
		100	-0.002	0.130	0.001	0.999
		200	0.000	0.092	0.001	0.998
	30	30	0.000	0.202	0.002	0.998
		100	-0.002	0.110	0.001	0.997
		200	0.000	0.077	0.001	0.996
B	10	30	0.005	0.132	0.001	1.000
		100	0.004	0.071	0.001	1.000
		200	0.004	0.049	0.000	1.000
	30	30	0.001	0.073	0.001	1.000
		100	0.000	0.039	0.000	1.000
		200	0.000	0.028	0.000	1.000
C	10	30	0.002	0.220	0.002	0.999
		100	0.000	0.121	0.001	0.999
		200	0.003	0.086	0.001	1.000
	30	30	-0.001	0.194	0.002	0.998
		100	-0.002	0.107	0.001	0.998
		200	0.000	0.075	0.001	0.999

Note. GM: generative model. T: number of timepoints. N: sample size. Bias: $\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \hat{\beta}_{0i} - \beta_0$, which represents the difference between the mean of the estimated parameter values $\hat{\beta}_0$ and the prespecified treatment effect $\beta_0 = 1$. SD: $\sqrt{\frac{1}{(n_{\text{sim}}-1)} \sum_{i=1}^{n_{\text{sim}}} \left(\hat{\beta}_{0i} - \bar{\beta}_0\right)^2}$, which is the standard deviation of estimates across replications. MC-SE: $\frac{\text{SD}}{\sqrt{n_{\text{sim}}}}$, which represents the Monte Carlo SE of bias. SR: model fitting success rate.