The Dangers of Including Endogenous Time-Varying Covariates in Clustered Longitudinal Data Analysis

By: Ward B. Eiling, Supervisors: Ellen Hamaker & Jeroen Mulder

Utrecht University

Once upon a time...

> *"To our surprise, even without time-varying confounding (e.g., when the randomization probability is constant in an MRT), the inclusion of endogenous covariates in random effects models can cause bias in assessment of the treatment effects." (Qian, Klasnja, and Murphy 2020)*
> *"when using generalized estimating equations (GEE) with endogenous covariates, one should use working independence correlation structure to avoid biased estimates"(Pepe and Anderson 1994; Qian, Klasnja, and Murphy 2020)*
> *"the issue ... is important for all longitudinal data analysis methods" (Diggle 2002)*

So we should all be using GEE with working independence?

The Project

Aims:

1. Restate the issue in a psychological context

The Project

Aims:

1. Restate the issue in a psychological context
2. Provide practical recommendations

Endogenous Time-varying Covariates: What Are They?

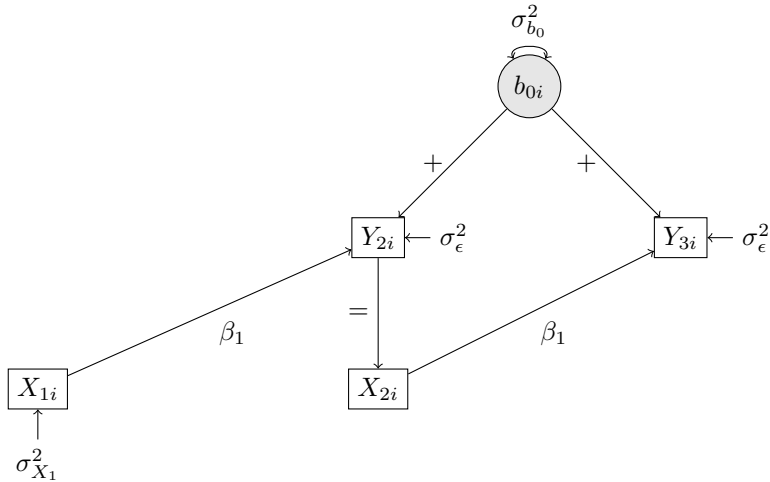


Figure 1: Path Diagram of MLM Without Treatment

Endogenous Time-varying Covariates: What Are They?

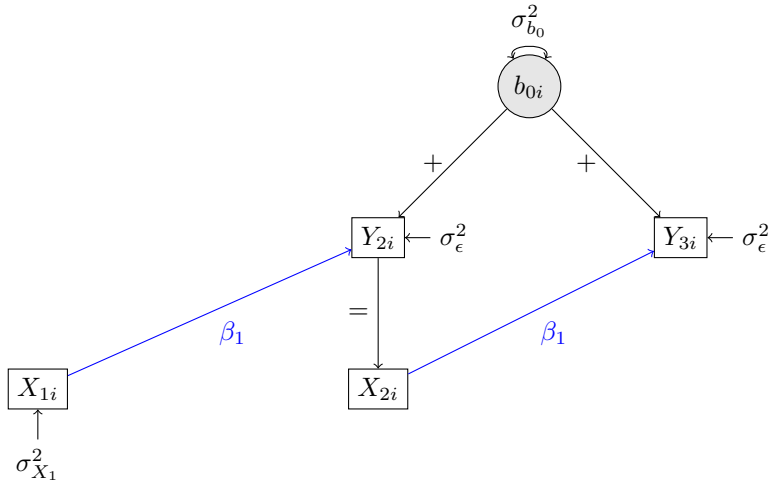
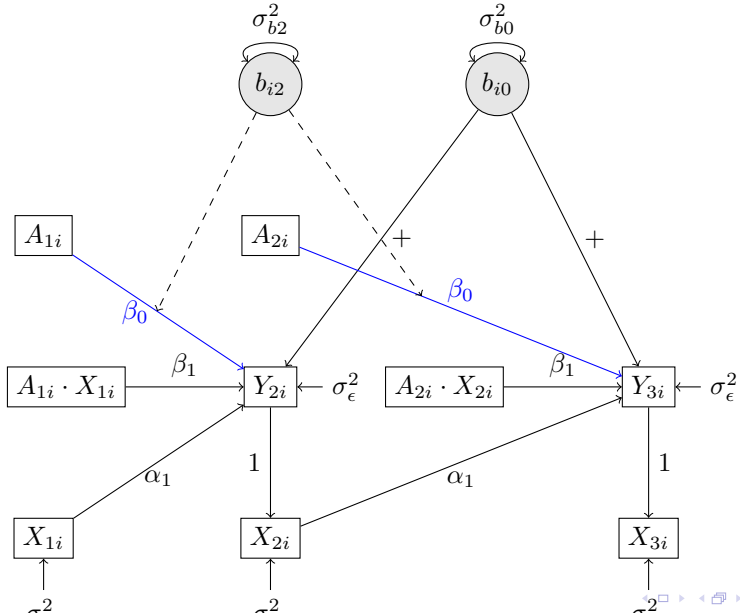
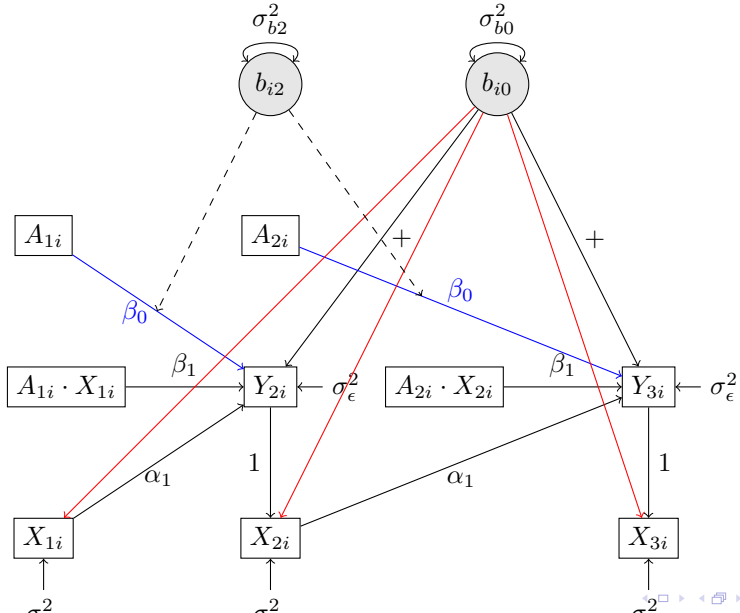


Figure 2: Path Diagram of MLM Without Treatment

Expanding to a Generative Model with Treatment



Expanding to a Generative Model with Treatment



Simulation Results

Table 1: Simulation results of Parameter Bias present in GM 1 and 3 for N = 200 over 1000 replications, for T = 10 and T = 30

GM	Characteristics	Bias (T = 10)	Bias (T = 30)
1	Includes random intercept and random slope for treatment	0.003	0.001
3	Model 1 with dependency random slope and covariate	-0.051	-0.023

Treatment Effect Bias: Can DAGs help us?

- Pearl's backdoor criterion and interaction effects?

Isolating the issue

Table 2: Simulation results of Parameter Bias present in GM 3, 3a, 3b for N = 200 over 1000 replications, for T = 10 and T = 30

GM	Characteristics	Bias (T = 10)	Bias (T = 30)
3		-0.051	-0.023
3a	Model 3 without random slope b_{2i}	0.002	-0.000
3b	Model 3 without interaction effect β_1 (between treatment and covariate interaction)	0.005	0.001

Conditional independence assumption is still violated for these models...

What is happening here?

Discussion: Takeaways and Moving Forward

- Marginal relationship primary interest?

Discussion: Takeaways and Moving Forward

- Marginal relationship primary interest?
  - GEE with independence

Discussion: Takeaways and Moving Forward

- Marginal relationship primary interest?
  - GEE with independence
- In practice, we don't know the DGM

Discussion: Takeaways and Moving Forward

- Marginal relationship primary interest?
  - GEE with independence
- In practice, we don't know the DGM
  - Theory-based decisions

Discussion: Takeaways and Moving Forward

- Marginal relationship primary interest?
  - GEE with independence
- In practice, we don't know the DGM
  - Theory-based decisions
- Next steps

Discussion: Takeaways and Moving Forward

- Marginal relationship primary interest?
  - GEE with independence
- In practice, we don't know the DGM
  - Theory-based decisions
- Next steps
  - What makes bias increase?

References

Diggle, Peter. 2002. *Analysis of Longitudinal Data*. OUP Oxford.

Pepe, Margaret Sullivan, and Garnet L Anderson. 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): 939–51. https://doi.org/10.1080/03610919408813210.

Qian, Tianchen, Predrag Klasnja, and Susan A. Murphy. 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): 375–90. https://doi.org/10.1214/19-sts720.

Appendix