Asking –and answering– causal questions using longitudinal data

Rafael Quintana

University of Kansas

Causal inference with longitudinal data

- Using longitudinal (panel) data to answer causal questions.

- $X \rightarrow Y$?

- Aims:
  1) Distinguish between three causal quantities or estimands.
  2) Clarify causal assumptions needed to identify these estimands.
  3) Review statistical models to estimate these estimands.

Definition, identification and estimation

<u>Definition</u>: specify a target causal quantity.

↓

<u>Identification</u>: assumptions to identify this quantity

↓

<u>Estimation</u>: statistical model to estimate this quantity

- "Define first, identify second, and estimate last" (Pearl, 2011)

Causal estimands in longitudinal analysis

- Specifying the estimand is crucial in longitudinal analysis:

1. There are multiple quantities we can investigate with longitudinal data.
2. Different estimands can require different assumptions and models.
3. There are many longitudinal models, but no clarity about what quantity they estimate.

- Goal: distinguish different estimands we can investigate using longitudinal data.

Specifying an estimand

- In causal inference, estimands are expressed in terms of counterfactuals.

- Example: the effect of peer victimization on internalizing behaviors.

- Notation
  - X_{it}: binary treatment variable for unit i at time t (peer victimization).
  - Y_{it}: observed value of the outcome (internalizing behaviors).
  - $Y_{it}(X_{it})$: potential outcome under the treatment status $X_{it} = x_{it}$.

- Causal effects are defined as comparisons between these potential outcomes or counterfactuals.

Three estimands in longitudinal analysis

- With cross sectional data, one can only ask:

a) Do individuals with high values of X have high values of Y?

- With longitudinal data one can ask a variety of causal questions:

a) Does X have an immediate effect on Y?
b) Is there a cumulative effect of X on Y?
c) Does X have long-term effects on Y?

- These are all relevant questions that correspond to different estimands:

a) contemporaneous, b) cumulative, and c) long-term effects.

Defining the contemporaneous effect

- The contemporaneous effect of treatment (CET) of X on Y can be defined as:

$$\tau_c = E[Y_{it}(X_{it} = 1) - Y_{it}(X_{it} = 0)]$$

- CET represents the average change in Y at time t after intervening on X at the same point in time (or immediately before, depending on time lags).

- CET provides a snapshot of the effect of X on Y at a single point in time.

- This estimand answers the question: "What are the immediate or short-run effects of victimization on internalizing behaviors?"

Estimating contemporaneous effects using FE

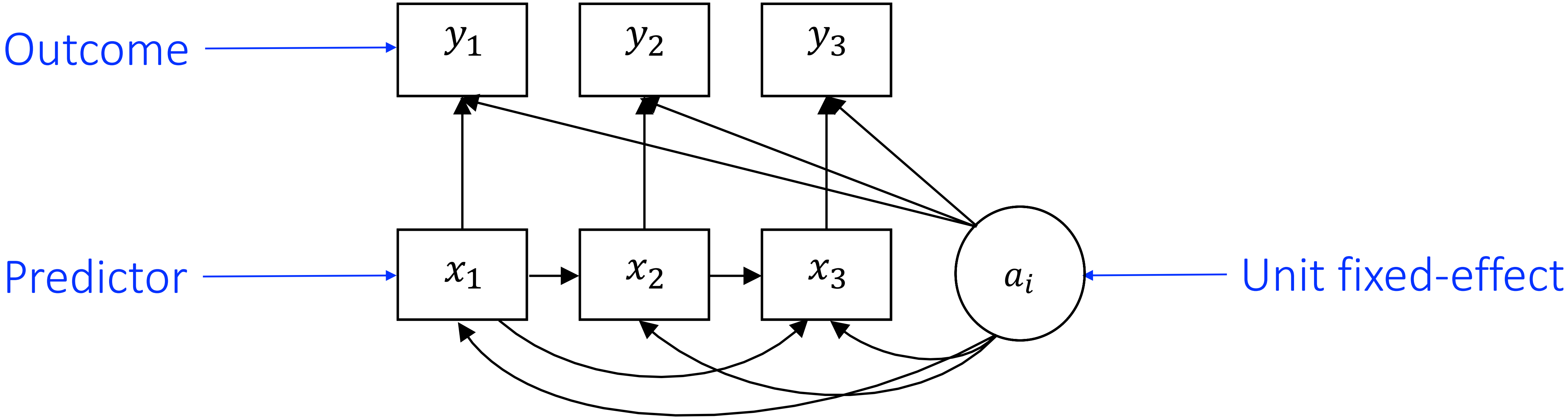
- The CET can be estimated using fixed-effects models, which control for unmeasured time-invariant confounders:

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

- α_i captures all unit-specific and time-invariant causes of Y.
- β represents the contemporaneous effect of X at time t on Y at time t.

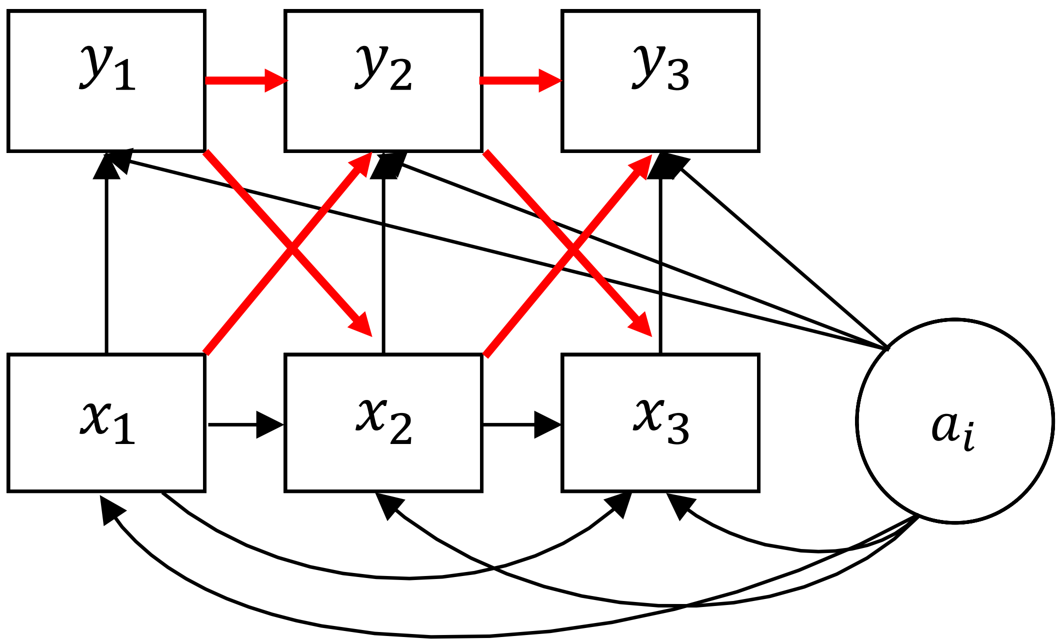
DAG-FE

- DAG corresponding to a three-wave FE model:



Identifying assumptions of FE

- In DAGs, the absence of arrows represent strong assumptions:



FE models assume:

1) no time-varying confounders

2) no reciprocal effects

3) no state effects

4) no carry-over effects

- Assumptions 3 and 4 can be relaxed by conditioning on X_{it-1} (Imai & Kim, 2019)
- Assumption 2 can be relaxed using a dynamic panel model (Allison et al., 2017)

Summary of contemporaneous effects

- Contemporaneous effects measure the shortrun effect of X on Y.
- Longitudinal research often focuses –explicitly or implicitly– on the CET.
- FE models allow us to estimate the CET by controlling for unobserved time-invariant confounders.
- Dynamic panel models allow us to control for unit effects and account for reciprocal causality.

Cumulative effects

- Contemporaneous effects focus on transient or short-lived effects.

- Yet the history of social and behavioral processes can be of critical importance.

- Example: persistent peer victimization affects mental health (e.g., Hellfeldt et al. 2018)

- The quantity of interest is not an immediate effect, but the effect of the entire history of being victimized.

Notation cumulative effects

- A cumulative effect represents the effect of the entire history of treatment.

- Notation
  - $\bar{X}_{it} = (X_{i1}, \dots, X_{it})$: the history of X_{it} until period t.
  - $Y(\bar{X}_{it} = \bar{x}_{it})$: the outcome that would occur under $\bar{x}_{it}$.

- E.g., in a four wave panel individuals who were never victimized, $\bar{x}_{it} = (0,0,0,0)$, while others who were always victimized, $\bar{x}_{it} = (1,1,1,1)$.

Defining cumulative effects

- The cumulative effect of treatment (CMET) can be defined as

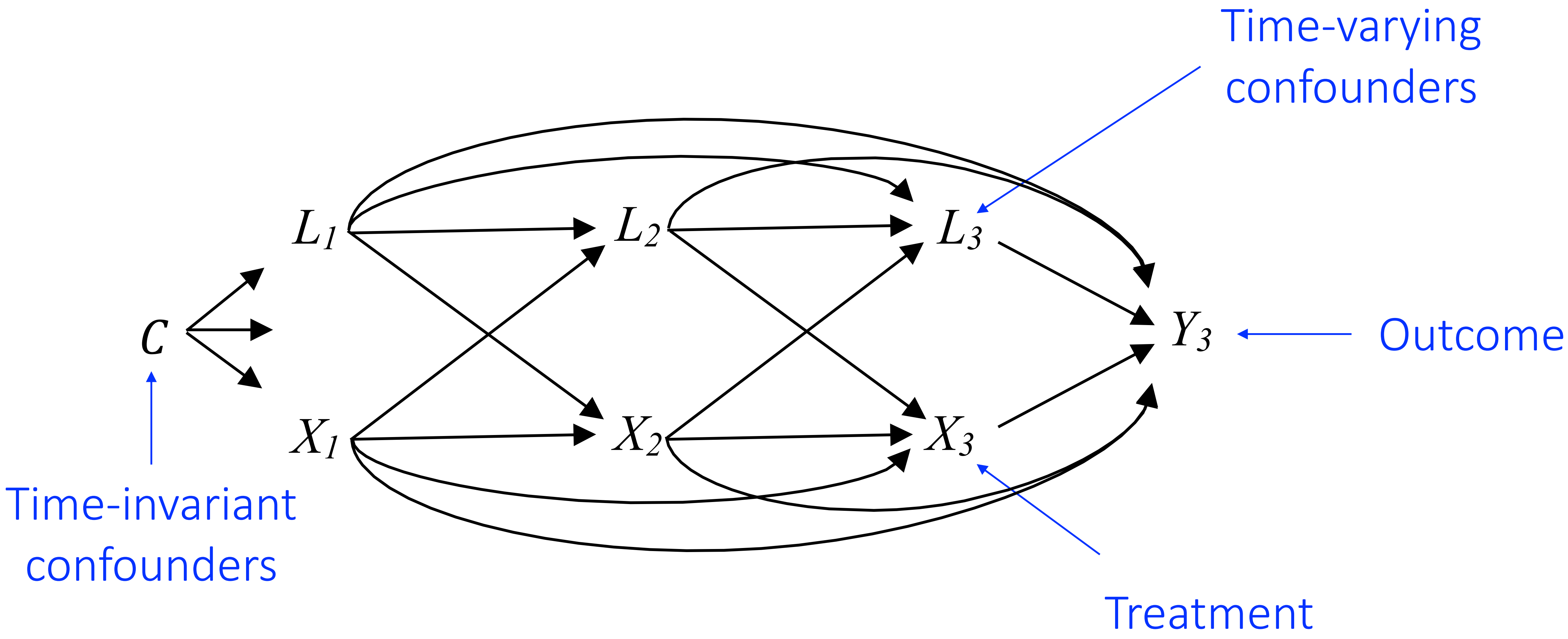
$$\tau_{cm} = E[Y_{it}(\bar{x}_{it}) - Y_{it}(\bar{x}'_{it})]$$

- where $\bar{x}_{it}$ and $\bar{x}'_{it}$ refer to particular histories of the treatment X_{it}.

- E.g., the effect of being victimized on every occasion versus not at all.

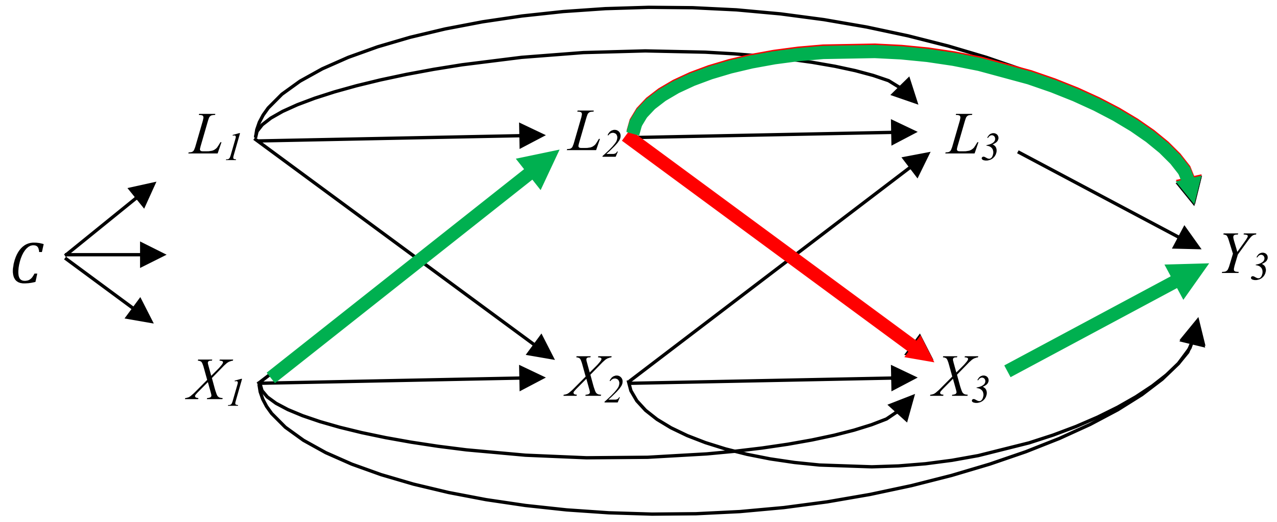
Cumulative effect DAG

- DAG where X_t affects the outcome Y_3 in all measurement occasions



Estimating cumulative effects

- Common methods (e.g., regression) cannot be used to estimated CMET.



L_t is both a confounder <u>and</u> mediator

Confounder in $X_3 \leftarrow L_2 \rightarrow Y_3$

Mediator in $X_1 \rightarrow L_2 \rightarrow Y_3$

- Controlling for L_2 introduces posttreatment bias, but failing to control for L_2 introduces confounding bias.

Estimating cumulative effects using marginal structural models (MSMs)

- Cumulative effects can be estimated with MSMs using IPTW (Robins et al. 2000)
- These models use weights to adjust for confounders, rather than including these confounders as additional regressors.
- One possibility is to estimate separate effects for each time period:

$$E[Y_{it}(\bar{x}_{it})] = \mu + \sum_{t=1}^{T} \beta_t x_{it} + \epsilon_{it}$$

- In a three-wave panel the joint effects of victimization will be $\beta_1, \beta_2, \beta_3$.

Estimating cumulative effects using marginal structural models (MSMs)

- One can also specify a model based on the cumulative amount of treatment received:

$$E[Y_{it}(\bar{x}_{it})] = \mu + \beta_1 \sum_{t=1}^{T} x_{it} + \epsilon_{it}$$

- where β_1 represents the effect of receiving one additional treatment.

- The effect of receiving treatment at all times can be estimated as $T\beta_1$.

Creating inverse probability weights

- The models are estimated using a weighted least squares regression.

- The IPTW weights are constructed as:

$$W_i = \prod_{t=1}^{T} \frac{P(X_{it} | \bar{X}_{it-1})}{P(X_{it} \mid \bar{X}_{it-1}, \bar{Y}_{it-1}, L_{it-1}, C_i)}$$

- Denominator: predicted probability of observing each individual treatment status conditional on past treatment status ($\bar{X}_{it-1}$), outcomes ($\bar{Y}_{it-1}$) and other time-varying (L_{it-1}) and time-invariant (C_i) confounders.

Summary of cumulative effects

- Cumulative effects quantify the average effect of a treatment history.
- E.g., being repeatedly victimized or not over time.
- Marginal structural models can be used to estimate cumulative effects.
- MSMs with IPTW assume no unobserved time-varying and time-invariant confounders.

Long-term effects

- Long-term effects are defined as the effect of a distant lag of treatment (which can correspond to weeks, years, decades etc.)

- E.g., research suggests that peer victimization in school has long-term effects on individuals’ mental heath (Arseneault 2017; Hellfeldt et al. 2018)

Defining long-term effects: total effect

- One can define two estimands corresponding to long-term effects.
- The first estimand is the total effect of a distant lag of treatment:

$$\tau_T = E\big[Y_{it}\big(X_{i,t-j} = 1\big) - Y_{it}\big(X_{i,t-j} = 0\big)\big]$$

Total effect DAG

- DAG representing the total effect of X on Y with pretreatment confounders (C)

$$\begin{tikzcd} C \arrow[r] \arrow[rr, bend left] & X \arrow[r] & Y \end{tikzcd}$$

Defining long-term effects: controlled direct effect

- The second estimand is the direct effect of a distant lag of X that is not mediated by future values of X.

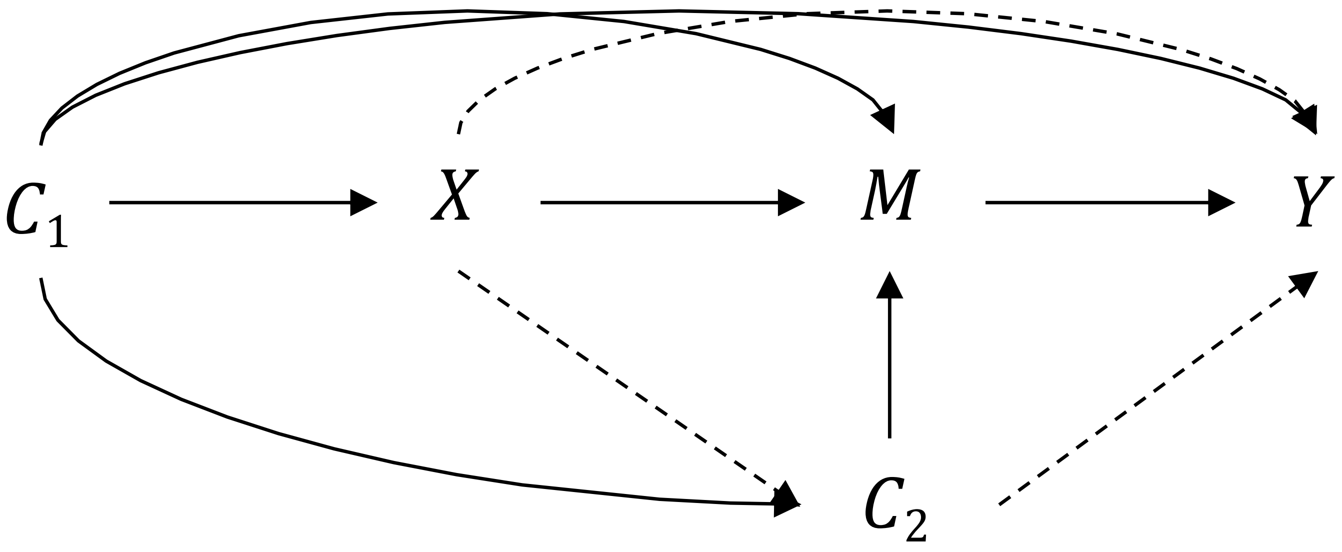
- This quantity is referred to as a "controlled direct effect" and is defined as

$$\tau_m = E\left[Y_{it}\left(X_{i,t-j} = 1, m\right) - Y_{it}\left(X_{i,t-j} = 0, m\right)\right]$$

- E.g., peer victimization has long-term effects if its influence persists even after victimization has ceased (i.e., $m = 0$).

CDE DAG

- DAG showing the effect of X on Y that is not mediated by M (dashed lines).



Estimating CDEs

- To estimate the CDE, one can implement a marginal structural model:

$$Y_{it}(X_{i,}M) = \mu + \beta_1 X + \beta_2 M + \beta_3 MX$$

- where β_1 represent the CDE.

- Inverse probability weights are used to control for confounders (VanderWeele, 2009)

Empirical illustration: Peer victimization and internalizing behaviors

- ECLS: 2011 with an analytic sample of 7,973 individuals.

- Measures of teacher-reported peer victimization and internalizing behaviors were obtained in 2^{nd}, 3^{rd}, 4^{th} and 5^{th} grade (4 waves).

- A wide range of time-varying (45) and time-invariant (15) confounders.

Results

	Contemporaneous effect		Cumulative effect		Long-term effect	
	Fixed effects model	Dynamic panel model	Multiple parameter model	Single parameter model	Total effect	Controlled direct effect
X_t	0.229	0.242	0.295	–	–	–
	(.006)	(.016)	(.018)			
X_{t-1}	–	–	0.061	–	–	–
			(.017)			
X_{t-2}	–	–	0.066	–	0.074	0.070
			(.016)		(.013)	(.029)
ΔX	–	–	–	0.139	–	–
				(.009)		

Summary

Substantive question	Causal estimand	Statistical model	Confounding assumptions
Does peer victimization have an immediate effect on mental health outcomes?	Contemporaneous treatment effect	Fixed effects or dynamic panel models	No time-varying confounders between the treatment and the outcome.
Does a history of peer victimization affect mental health outcomes?	Cumulative treatment effect	Marginal structural models	No unmeasured treatment-outcome confounders.
Does peer victimization have long-term effects on mental health outcomes?	Long-term treatment effect (total effect or controlled direct effect)	Multivariable regression or marginal structural models	Total effect: no unmeasured treatment-outcome confounders. CDE: no unmeasured treatment-outcome or mediator-outcome confounders.

Conclusion

- Empirical research should begin with a research question, which means clarifying what is the quantity of interest (estimand).
- This study distinguishes between three quantities or causal estimands that are important in social and behavioral research.
- It is essential to clearly state the estimand of interest, as these quantities can require different statistical models and identifying assumptions.
- Differentiating between these quantities can also help researchers ask sharper causal questions.

Thank you for your attention!

Rafael Quintana
quintana@ku.edu