Statistical Methods for Causal Inference in Observational and Randomized Studies

Mark J. van der Laan[1], Maya L. Petersen[1], Sherri Rose[2]

[1]University of California, Berkeley School of Public Health

[1]Johns Hopkins Bloomberg School of Public Health

laan@berkeley.edu mayaliv@berkeley.edu srose@jhsph.edu

stat.berkeley.edu/~laan/

works.bepress.com/maya_petersen/

drsherrirose.com

targetedlearningbook.com

Longitudinal Data

Day 3: Lectures 1 and 2

Outline: Longitudinal Data

1. Repeated point treatment data structures
   - Allows us to use Model, Data, Identifiability Result, and Estimators previously introduced (Day 1)
2. Estimating the effects of intervening on more than one node
   - Example parameters
     - Cumulative treatment effects
     - Right censoring
     - Longitudinal Marginal Structural Models
   - Link to Observed data and Identifiability
   - Estimators
     - Maximum Likelihood Substitution
     - Inverse Probability of Treatment Weighted

Example: Abacavir and Cardiovascular Disease

- Analysis of observational data from several cohorts suggested abacavir use associated with increased risk of myocardial infarction among treated HIV-infected population
  - Other analyses found no evidence of such an association….
- Example of a causal question: Does current use of abacavir (ABC) increase risk of myocardial infarction (MI)?

Notation for Longitudinal Data

- L(t)= covariates at time t, t=1,…,K+1
  - The time-varying equivalent of W
  - As usual, a node can be multi dimensional
- Y(t)= outcome at time t, t=1,…,K+1
  - Sometimes defined as a subset of L(t)
  - Alternative: Y measured only at the end of follow up, sometimes defined as a subset of L(K+1)
- A(t)=exposure/treatment at time t, t=1,…,K

Example: Effect of current abacavir use on MI risk

- Monthly Data (Time in month increments)
- A(t)=Indictor current abacavir use at start of month
- Y(t)=Indicator MI during month
- L(t)=Covariates in prior month
  - Other Drugs, Lipids, DM, HTN…
  - This can include summaries of patient history up to start of the month, including past CHD
- O(t)=(L(t),A(t),Y(t)), t=1,…,K

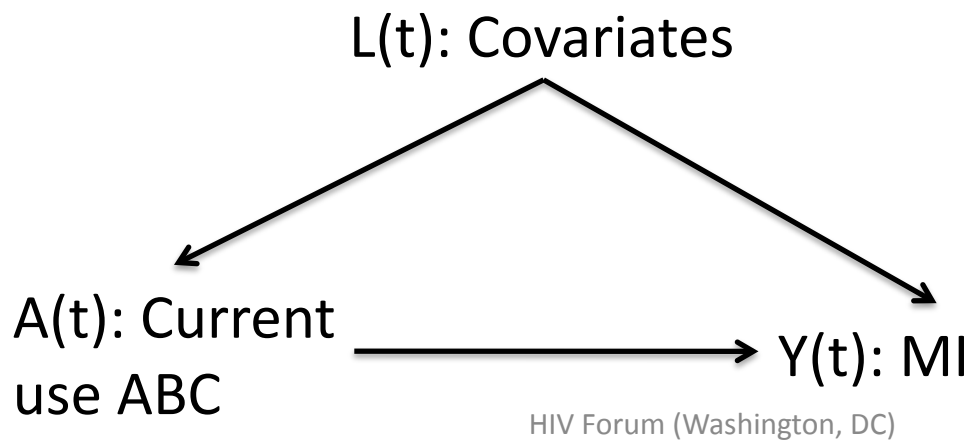
Example: Effect of current abacavir use on MI risk?

- Structural Causal Model/Graph for a single time point?

Example: Effect of current abacavir use on MI risk?

- Structural Causal Model/Graph for a single time point?

$$L(t) = f_{L(t)}(U_{L(t)})$$
$$A(t) = f_{A(t)}(L(t), U_{A(t)})$$
$$Y(t) = f_{Y(t)}(L(t), A(t), U_{Y(t)})$$



Example: Effect of current abacavir use on MI risk?

- What is the effect of current abacavir exposure on immediate risk of MI among subjects without prior history of MI?
  - Define relevant counterfactuals
  - What sort of target causal parameter might address this question?
  - What sort of an MSM might we use to define this target parameter?

Example: Effect of current abacavir use on MI risk?

- Counterfactual outcomes: $Y_{a(t)}(t)$, t=1,…,K
  - $Y_1(t)$: counterfactual MI status if used abacavir at time t
  - $Y_0(t)$: counterfactual MI status if did not use abacavir at time t

Example: Effect of current abacavir use on MI risk?

- Possible target parameters
  - $E(Y_1(t)-Y_0(t)|Y(t-1)=0)$: difference in risk of (new) CHD at time t if did vs. did not use abacavir
- Using a (working) MSM to smooth over time:

$$P(Y_{a(t)}(t) = 1|Y(t-1) = 0) = m(a(t), t|\beta)$$

  - Ex: A possible simple working MSM

$$m(a(t), t|\beta) = expit(\beta_0 + \beta_1 a(t) + \beta_2 t + \beta_3 a(t) \times t)$$

Example: Effect of current abacavir use on MI risk?

- For a given time point, the data are analogous to the (W,A,Y) data we have been discussing
  - We can consider this as a repeated point treatment data structure
- This means that you can use the TMLE package (yesterday's lab) to estimate the time point specific effect, averaged over all time points
- Cross-Validation and inference need to respect repeated measures data structure
  - Specify patient ID as unit of independence

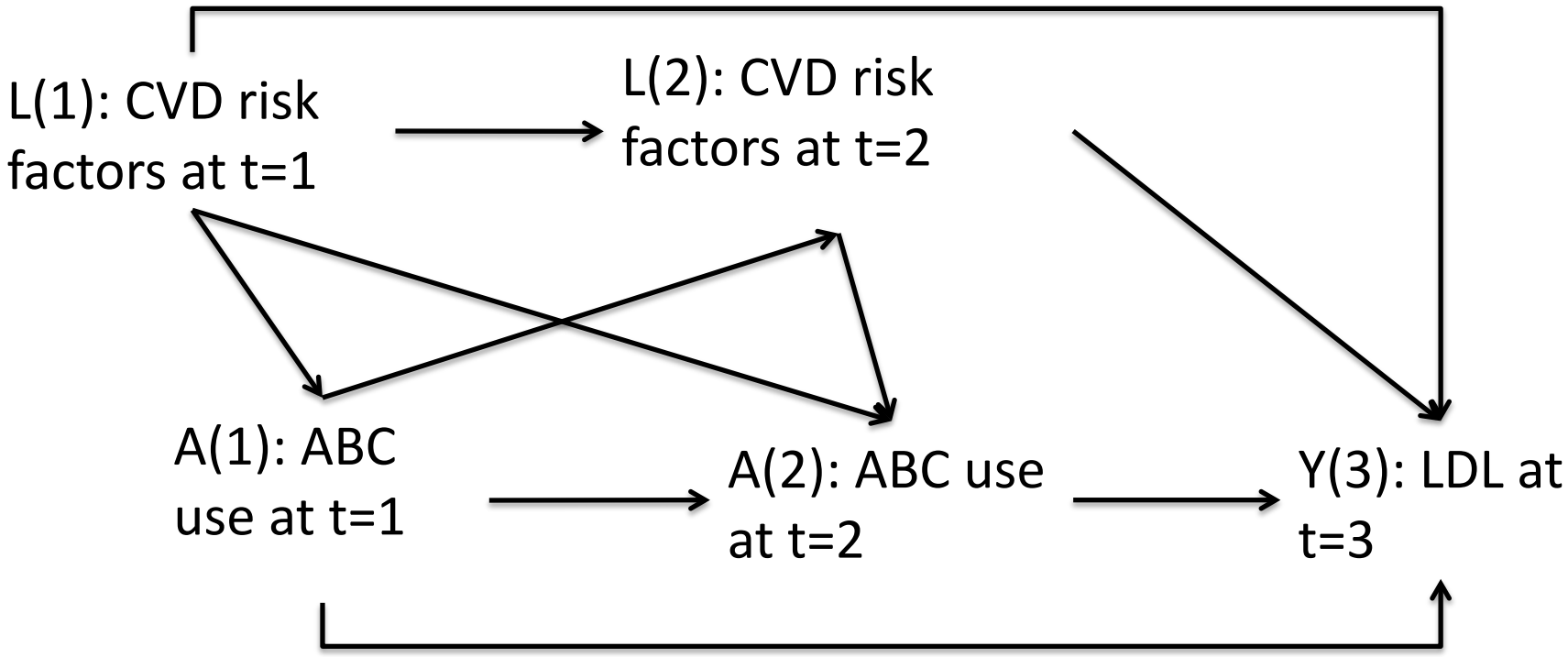
Investigating cumulative abacavir toxicity

- What if we want to know about the effects of cumulative exposure to abacavir?
  - Alternative target parameter that investigates the effect of extended abacavir use patterns?
- Need to go beyond repeated point treatment formulation
  - SCM that incorporates time-varying covariates and time-varying treatment
  - Counterfactual outcomes indexed by interventions on more than one treatment node

Abacavir Example: SCM for Longitudinal Treatments

- To simplify things, let's start with three time points, and an alternative outcome
- We measure
  - CHD risk factors (including lipids) at t=1 and t=2
  - Abacavir use at t=1 and t=2
  - Outcome= LDL cholesterol at t=3
  - Assume no deaths, censoring, or missing data for now
- How might we draw the graph? Define the SCM?

Abacavir Example: Longitudinal Causal Graph



More notation for longitudinal data

- Over-bars used to refer to the history of a variable

$$\bar{A}(t) = \{A(1), A(2), ..., A(t)\}$$
$$\bar{A} = \bar{A}(K) = \{A(1), A(2), ..., A(K)\}$$
$$\bar{L}(t) = \{L(1), ..., L(t)\}$$
$$\bar{L} = \bar{L}(K+1) = \{L(1), ..., L(K+1)\}$$
$$Y \subset L(K+1)$$

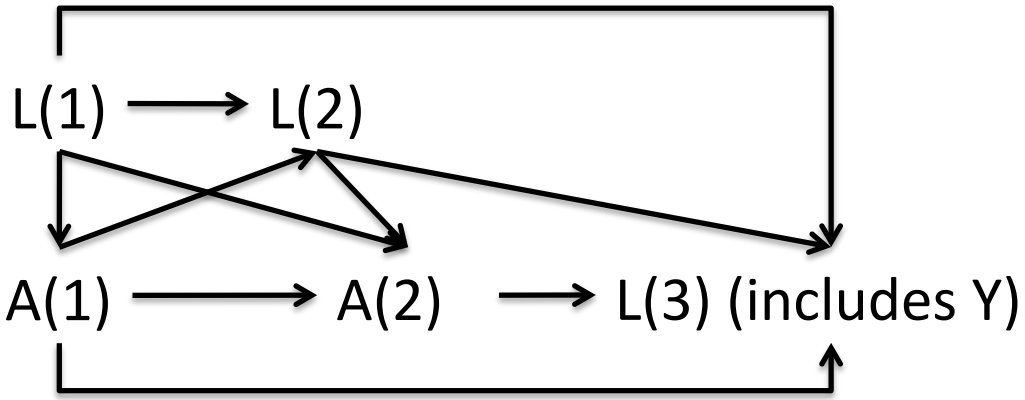
SCM for Longitudinal Data

- Notation simplifies specifying our SCM
  - Avoid writing out a separate equation for each time point
- A common SCM: Assumes each variable may be affected by all preceding variables
  - ie. Pa(X)= all variables that temporally precede X

$$L(1) = f_{L(1)}(U_{L_1})$$
$$A(1) = f_{A(1)}(L(1), U_{A(1)})$$
$$L(t) = f_{L(t)}(\bar{A}(t-1), \bar{L}(t-1), U_{L(t)}), t = 2, ..., K+1$$
$$A(t) = f_{A(t)}(\bar{A}(t-1), \bar{L}(t), U_{A(t)}), t = 2, ..., K$$

Abacavir Example: SCM (K=2)

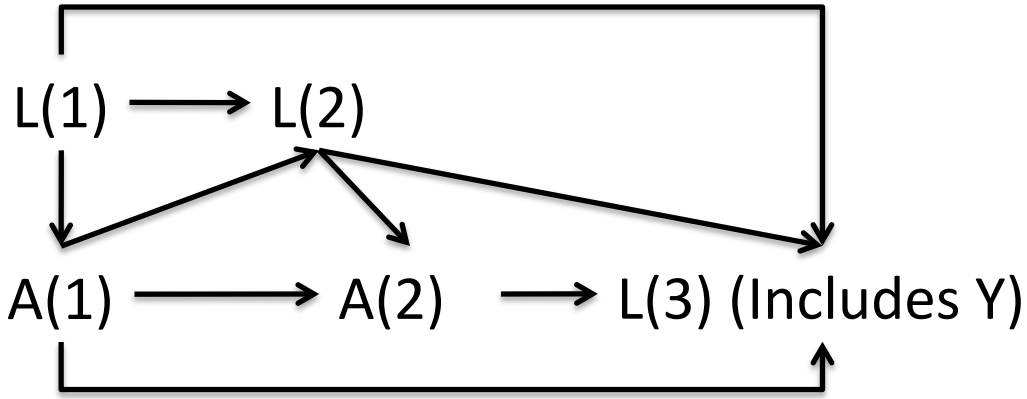
$$L(t) = f_{L(t)}(\bar{A}(t-1), \bar{L}(t-1), U_{L(t)}), t = 1, ..., K+1$$
$$A(t) = f_{A(t)}(\bar{A}(t-1), \bar{L}(t), U_{A(t)}), t = 1, ..., K$$



- What exclusion restrictions does this SCM make?

Abacavir Example: SCM (K=2)

- When might we want to modify this?
  - Example: data are from a sequentially randomized trial where abacavir use known to be assigned based on most recent CVD risk factors alone



$$L(t) = f_{L(t)}(\bar{A}(t-1), \bar{L}(t-1), U_{L(t)}), t = 1, ..., K+1$$

$$A(t) = f_{A(t)}(\bar{A}(t-1), L(t), U_{A(t)}), t = 1, ..., K$$

Counterfactuals indexed by longitudinal exposures

- Say we are interested in the difference in expected LDL at t=3 if all subjects had used abacavir at t=1 and t=2 versus if all subjects had not used abacavir at t=1 and t=2
- What intervention on the SCM could we use to define these counterfactuals?

Counterfactuals indexed by longitudinal exposures

- Original SCM

$$L(1) = f_{L(1)}(U_{L_1})$$
$$A(1) = f_{A(1)}(L(1), U_{A(1)})$$
$$L(2) = f_{L(2)}(L(1), A(1), U_{L(2)})$$
$$A(2) = f_{A(2)}(A(1), \bar{L}(2), U_{A(2)})$$
$$Y = f_Y(\bar{L}(2), \bar{A}(2), U_Y)$$

- Modified SCM, intervening on abacavir use at times 1 and 2?

Counterfactuals indexed by longitudinal exposures

- Original SCM

$$L(1) = f_{L(1)}(U_{L_1})$$
$$A(1) = f_{A(1)}(L(1), U_{A(1)})$$
$$L(2) = f_{L(2)}(L(1), A(1), U_{L(2)})$$
$$A(2) = f_{A(2)}(A(1), \bar{L}(2), U_{A(2)})$$
$$Y = f_Y(\bar{L}(2), \bar{A}(2), U_Y)$$

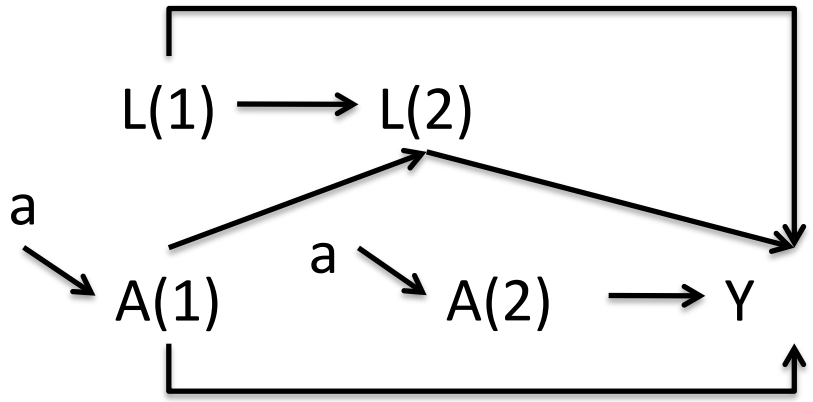
- Modified SCM, intervening on abacavir use at times 1 and 2

$$L(1) = f_{L(1)}(U_{L_1})$$
$$A(1) = a(1)$$
$$L(2) = f_{L(2)}(L(1), a(1), U_{L(2)})$$
$$A(2) = a(2)$$
$$Y = f_Y(\bar{L}(2), \bar{a}(2), U_Y)$$

Counterfactuals indexed by longitudinal exposures

- Modified SCM intervening on ABC at t=1 and t=2

$$L(1) = f_{L(1)}(U_{L_1})$$
$$A(1) = a(1)$$
$$L(2) = f_{L(2)}(L(1), a(1), U_{L(2)})$$
$$A(2) = a(2)$$
$$Y = f_Y(\bar{L}(2), \bar{a}(2), U_Y)$$



- Modified SCM used to define counterfactual outcome intervening on ABC use at two time points:

$$Y_{a(1),a(2)} = Y_{\bar{a}}$$

Intervention on counterfactual exposure history

Abacavir Example: Defining a longitudinal target parameter

- Question: How would expected LDL at t=3 have differed if all subjects had used abacavir at t=1 and t=2 versus if all subjects had not used abacavir at t=1 and t=2 ?
- How would you write the corresponding target causal parameter?

How would you write the corresponding target causal parameter?

- Denote the distribution of the corresponding counterfactual outcomes F_X

$$Y_{\bar{a}} \equiv Y_{a(1),a(2)}$$

$$\{Y_{\bar{a}} : a \in \mathcal{A}\} \sim F_X, \text{ where } \mathcal{A} = \{00, 01, 10, 11\}$$

- Example: Target causal parameter

$$E_{F_X}(Y_{11} - Y_{00})$$

- Interpretation?

Defining the target parameter intervening on multiple time points

- Example question: How would expected LDL have differed if everyone had has used abacavir throughout the study?
  - How would you write the target causal parameter?

Defining target parameters indexed by interventions on multiple time points

- Example question: How would expected LDL have differed if everyone had has used abacavir throughout the study?
  - How would you write the target causal parameter?

$$E(Y_{\bar{a}=1} - Y_{\bar{a}=0})$$

Defining target parameters using a Longitudinal Marginal Structural Model

- Example question: How does cumulative time exposed to abacavir affect LDL at the end of the study (K+1) ?
- How could your define this target causal parameter using a MSM?
- Example:

$$E(Y_{\bar{a}}) = \beta_0 + \beta_1 \sum_{t=1}^{K} a(t)$$

- An interesting working model?

Example of a Longitudinal MSM
An interesting working model?

$$E(Y_{\bar{a}}) = \beta_0 + \beta_1 \sum_{t=1}^{K} a(t)$$

1. Expected counterfactual LDL increases linearly with cumulative time spent on abacavir
2. Doesn't matter how recent this exposure is
   - Eg. expected counterfactual LDL at end of study is the same if exposed to abacavir at t=1 and 2 only, or at t=K-1 and K only

- Alternative MSM?

Survival Data

- So far, we have focused on a continuous outcome, measured at the end of the study on everybody (assumed no death or censoring/ LTFU)
- Now let's return to the original outcome: MI
  - Restrict population to those without history of MI
  - Interested in time to first MI
- T= time of first MI
- $Y(t)=I(t \leq T)$

Examples of target causal parameters with survival outcome

- Example: How would counterfactual probability of developing MI by 24 months differ under an intervention to use ABC for all 24 months *versus* to never use ABC?
- Target parameter?

Examples of target causal parameters with survival outcome

- Example: How would counterfactual probability of developing MI by 24 months differ under an intervention to use ABC for all 24 months *versus* to never use ABC?
- Target parameter?

$$E(Y_{\bar{a}=1}(t = 24mo) - Y_{\bar{a}=0}(t = 24mo))$$

Examples of target causal parameters with survival outcome

- Example: How does counterfactual (discrete) hazard of MI vary as a function of cumulative abacavir exposure since study enrollment?

- Example of MSM we could use to define the target parameter?

Examples of target causal parameters with survival outcome

- Example: How does counterfactual (discrete) hazard of MI vary as a function of cumulative abacavir exposure since study enrollment?

$$P_{F_X}(Y_{\bar{a}}(t) = 1 | Y_{\bar{a}}(t-1) = 0)$$

- Example of MSM we could use to define the target parameter?

$$logit\left(P_{F_X}(Y_{\bar{a}}(t) = 1 | Y_{\bar{a}}(t-1) = 0)\right) =$$

$$\beta_0 + \beta_1 t + \beta_2 \sum_{j=1}^{t} a(j) + \beta_3 t \times \sum_{j=1}^{t} a(j)$$

What about censoring?

- So far, we have assumed no censoring/loss to follow up
  - All subjects followed until min(K+1,T)
- In practice, of course, this is implausible
  - Abacavir example- data are gathered as part of (several) clinical cohorts
    - Patients transfer to other clinics, drop out of care…
  - Loss to follow up ubiquitous in both observational and RCT datasets

Incorporating censoring

- We can incorporate censoring in the SCM as a set of an additional X nodes in our graph (with their own structural equation)
- Define C as time when leave the cohort
  - Censoring time
- $C(t)=I(C \leq t)$
  - Indicator no longer in follow up at time t

Example of an SCM with censoring

- For example, if assume following temporal ordering for a given t: L(t), A(t), C(t), Y(t)

$$\text{For } t = 1, ..., K$$
$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), \bar{C}(t-1), U_{L(t)})$$
$$A(t) = f_{A(t)}(\bar{L}(t), \bar{A}(t-1), \bar{C}(t-1), U_{A(t)})$$
$$C(t) = f_{C(t)}(\bar{L}(t), \bar{A}(t), \bar{C}(t-1), U_{C(t)})$$
$$Y(t) = f_{Y(t)}(\bar{L}(t), \bar{A}(t), \bar{C}(t), U_{Y(t)})$$

Defining a target parameter in the presence of censoring

- Can now think of intervening not only on exposure/treatment at multiple time points, but also intervening on censoring/loss to follow up
- Example: What is the effect of cumulative abacavir exposure on hazard of MI *if all loss to follow up from the cohort had been prevented*?

Defining a target parameter in the presence of censoring

- How does the counterfactual (discrete) hazard of MI vary as a function of cumulative abacavir exposure since study enrollment?

– Counterfactuals of interest?

– Example Target parameter ?

Defining a target parameter in the presence of censoring

- Counterfactuals of interest defined by intervening on two types of nodes:
  - Exposure (abacavir use up till time t)
  - Censoring (stay in cohort up till time t)

$$Y_{\bar{a},\bar{c}=0}(t) : \bar{a} \in \mathcal{A}, t = 1, ..., K$$

$$\text{For } t = 1, ..., K$$

$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1) = \bar{a}(t-1), \bar{C}(t-1) = 0, U_{L(t)})$$

$$A(t) = a(t)$$

$$C(t) = 0$$

$$Y(t) = f_{Y(t)}(\bar{L}(t), \bar{A}(t) = \bar{a}(t), \bar{C}(t) = 0, U_{Y(t)})$$

Example of target causal parameters with survival outcome and censoring

- How does counterfactual (discrete) hazard of MI vary as a function of cumulative abacavir exposure since study enrollment?
  - Discrete counterfactual hazard:

$$P(Y_{\bar{a},\bar{c}=0}(t) = 1 | Y_{\bar{a},\bar{c}=0}(t-1) = 0)$$

  - Again, can pose a (working) MSM for how this varies as a function of time and cumulative exposure

$$P(Y_{\bar{a},\bar{c}=0}(t) = 1 | Y_{\bar{a},\bar{c}=0}(t-1) = 0) = m(\bar{a}, t | \beta)$$

Link to the Observed Data in the Presence of Censoring

- Censoring/loss to follow up determines up till what time point we observe data drawn from the underlying SCM
- Without censoring: $O = \left(\bar{L}(K), \bar{A}(K), \bar{Y}(K)\right)$
- With censoring:
  - Only observe data drawn from the underlying SCM up till the minimum of (K,C)

$$\tilde{T} = \min(K, C)$$

$$O = \left(\bar{L}(\tilde{T}), \bar{A}(\tilde{T}), \bar{C}(\tilde{T}), \bar{Y}(\tilde{T})\right)$$

Recap…

- What have we done so far?
  - Specified a longitudinal SCM
  - Defined counterfactuals and corresponding target parameters indexed by interventions on multiple nodes
- When is this applicable?
  - Effects of longitudinal exposures in general
  - Ex. Cumulative treatment effects
  - Ex. Censoring

Next…

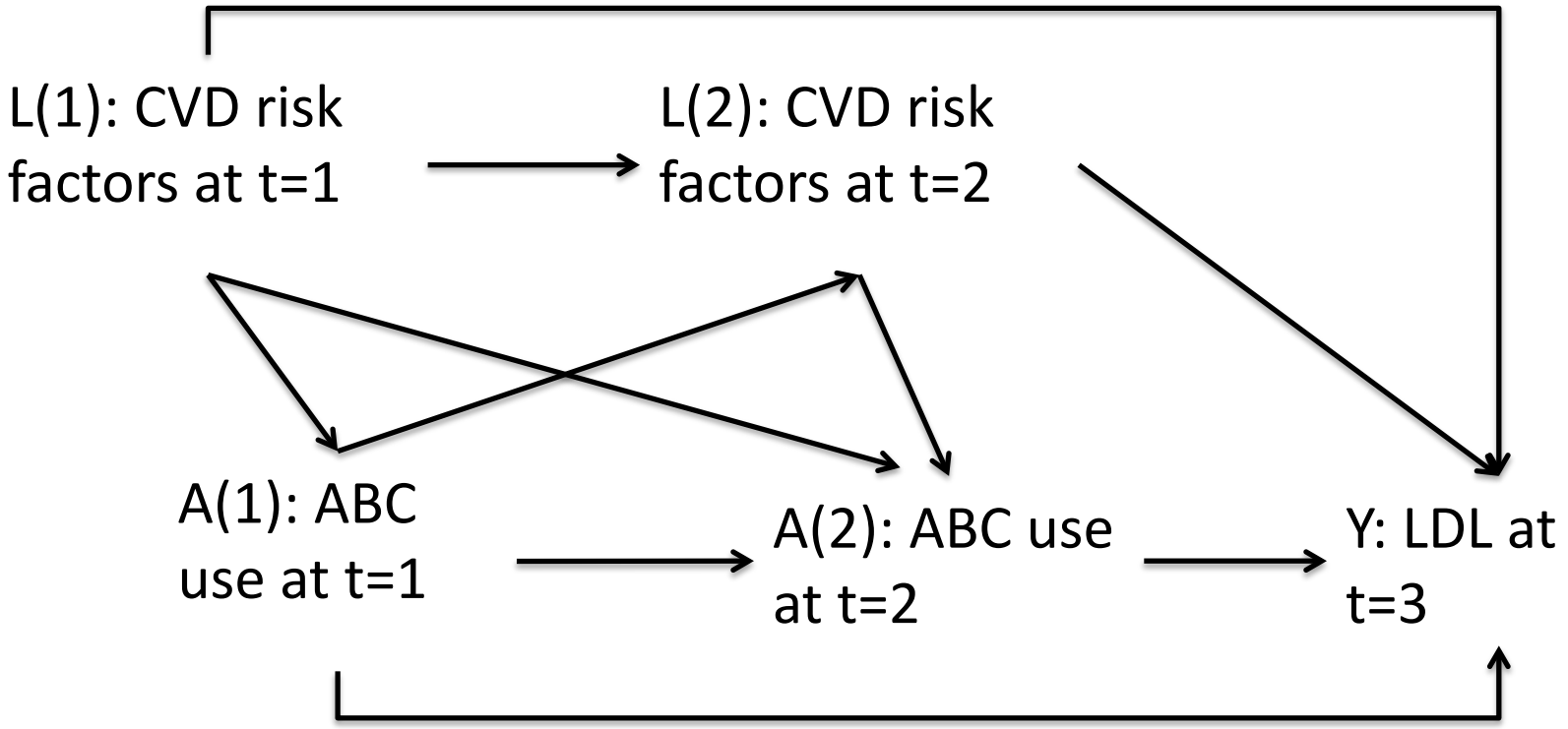
- Why all this extra notation and work?
- Identifying the effects of longitudinal exposures
  1. Requires new assumptions (beyond the standard randomization assumption or back door criterion)
  2. Results in new target statistical parameters (estimands)
  3. And thus requires new estimators

Identifiability for longitudinal exposures

- What causal assumptions are sufficient for our target longitudinal causal parameter to be identified as a parameter of the observed data distribution?

- Back to our simplified example for illustration
  - Effect of Abacavir use at t=1 and t=2 on LDL at t=3
  - Measure CHD risk factors at t=1 and t=2
  - Assume no deaths, censoring, or missing data

ABC Example: SCM/Graph



$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t}), t = 1, 2, 3$$

$$A(t) = f_{A(t)}(\bar{L}(t), \bar{A}(t-1), U_{A(t)}), t = 1, 2$$

$$Y \subset L(3)$$

ABC Example: Target Parameter and Observed Data

- Target causal parameter $E_{F_X}(Y_{\bar{a}=1} - Y_{\bar{a}=0})$
- Observed data: n i.i.d. copies of

$$O = \left(\bar{L}, \bar{Y}\right) \sim P_0$$

- Under what conditions can we write our target causal parameter as a parameter of the observed data distribution?
- We need to move beyond the simple back door criterion….

Recap: Identifiability for a Point Treatment Effect

- Assumption sufficient for identifiability of $E(Y_a)$
  - Randomization Assumption: $Y_a \perp A|W$

- Key result: Under the RA

$$P(Y = y|A = a, W = w) = P(Y_a = y|W = w)$$

  - Within strata of W, association= causation

When the randomization assumption holds given (some subset) of observed covariates W:

- The distribution of Y intervening on A is identified by the G- computation formula

$$P(Y_a = y) = \sum_w P(Y = y|A = a, W = w)P(W = w)$$

- And in particular:

$$E[Y_1 - Y_0] = \sum_w \{E[Y|A = 1, W = w] - E[Y|A = 0, W = w]\} P(W = w)$$

Recap: Identifiability for a Point Treatment Effect

- Can evaluate which covariates to include in W by using the backdoor criterion on graph
- What is the back door criterion and how do we evaluate it?

Recap: Identifiability for a Point Treatment Effect

- Can evaluate which covariates to include in W by using the backdoor criterion on graph
  - W should not be a descendent of A
  - After conditioning on W- are there any remaining unblocked backdoor paths?
  - If so, there is residual confounding-
    - There are sources of association between A and Y other then the causal association we care about

How are longitudinal parameters different?

- Our previous identifiability result relied on conditioning on some set of covariates W that were sufficient to block all back door paths from our intervention A to our outcome Y

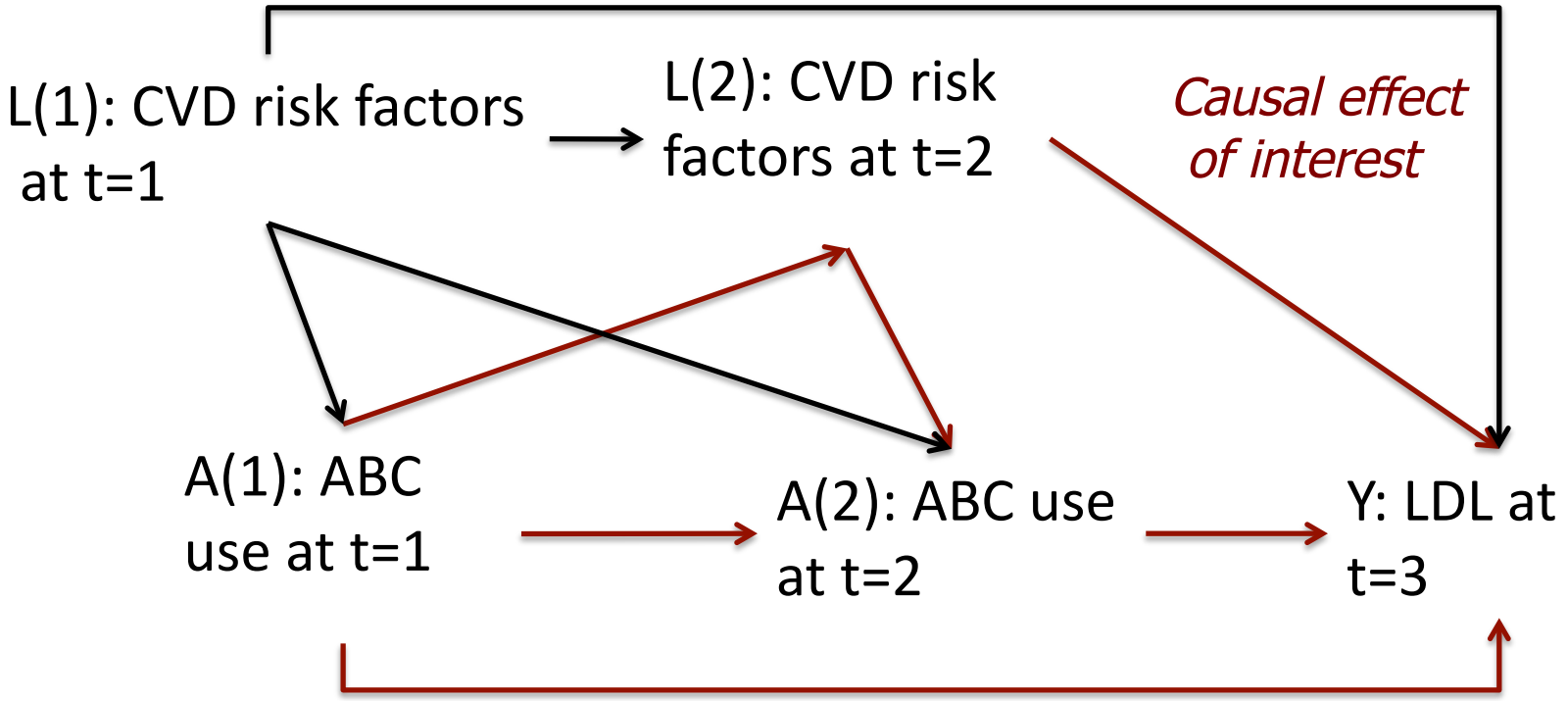
- We could not condition on descendents of A

How are longitudinal parameters different?

- When we are interested in intervening on multiple nodes, we are often in a situation where no one set of covariates that meet the back door criterion for all intervention nodes simultaneously exists

- However, the distribution of counterfactuals indexed by interventions on these multiple nodes may still be identified…

ABC Example: SCM/Graph



$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t}), t = 1, 2, 3$$

$$A(t) = f_{A(t)}(\bar{L}(t), \bar{A}(t-1), U_{A(t)}), t = 1, 2$$

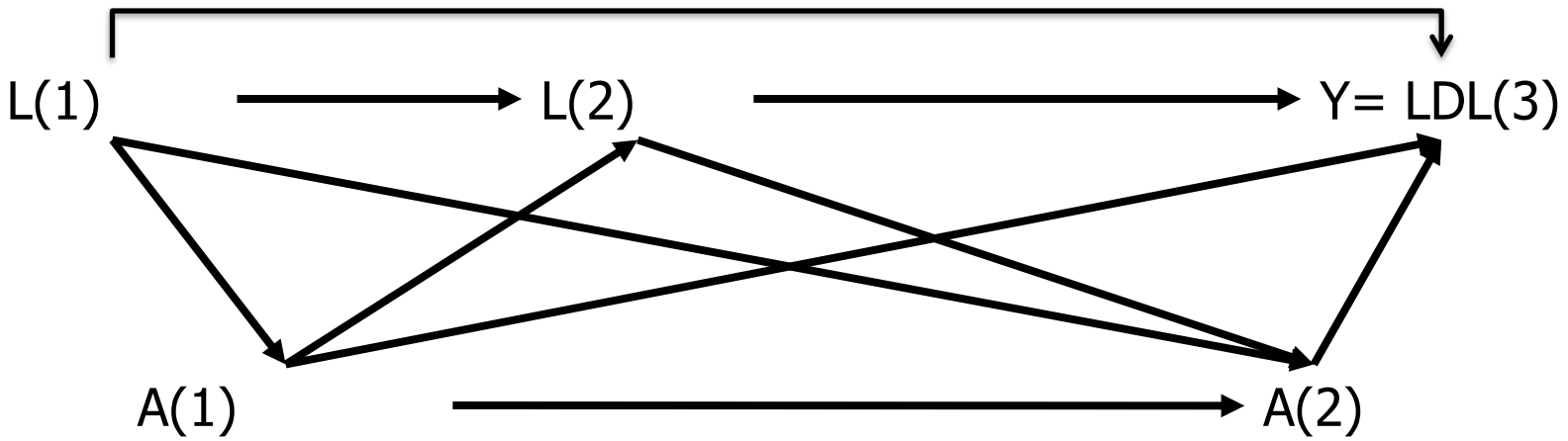
$$Y \subset L(3)$$

Why does the point treatment G-computation formula break down?

- Option #1:

$$E(Y_{11}) \stackrel{?}{=}$$

$$\sum_{l(1)} E(Y|A(1)=1, A(2)=1, L(1))P(L(1)=l(1))$$

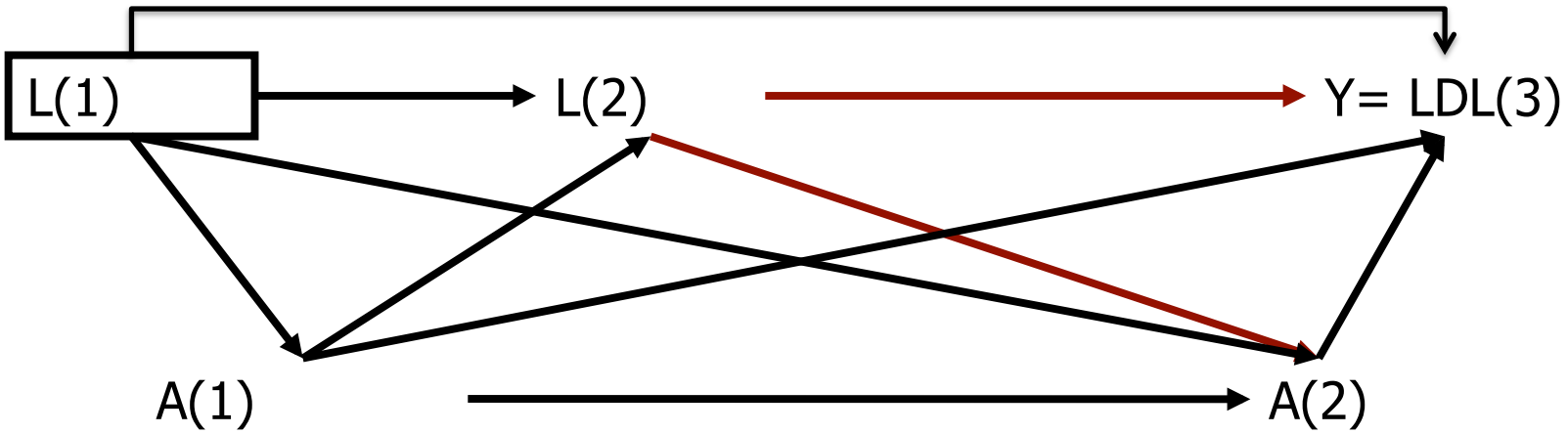


Why does the point treatment G-computation formula break down?

$$E(Y_{11}) \neq \sum_{l(1)} E(Y|A(1)=1, A(2)=1, L(1))P(L(1)=l(1))$$

- Why not?

Unblocked backdoor path

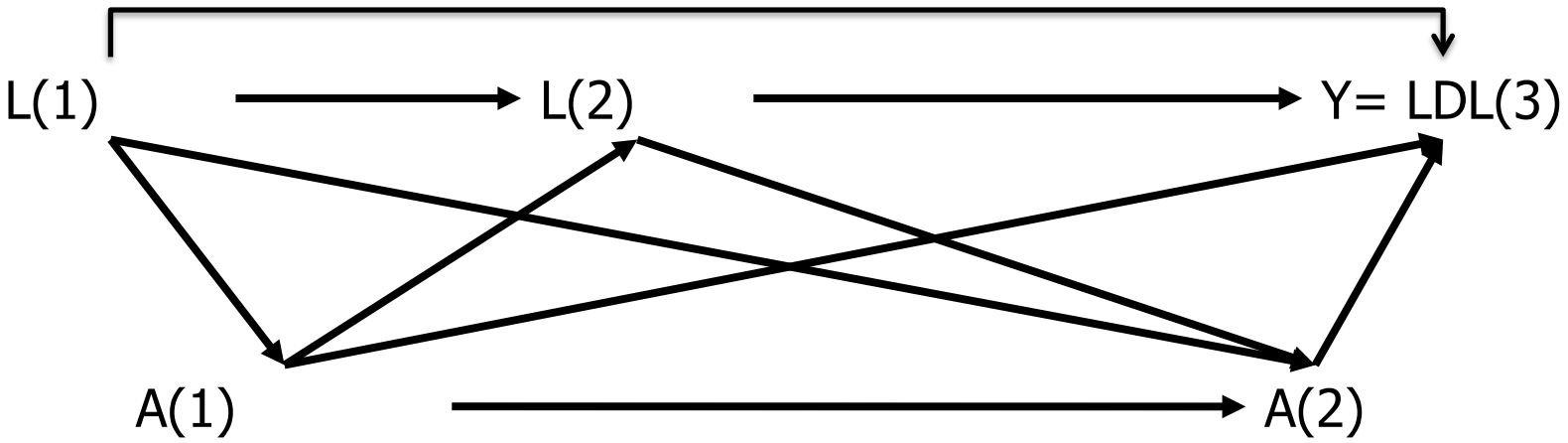


Why does the point treatment G-computation formula break down?

- Option #2:

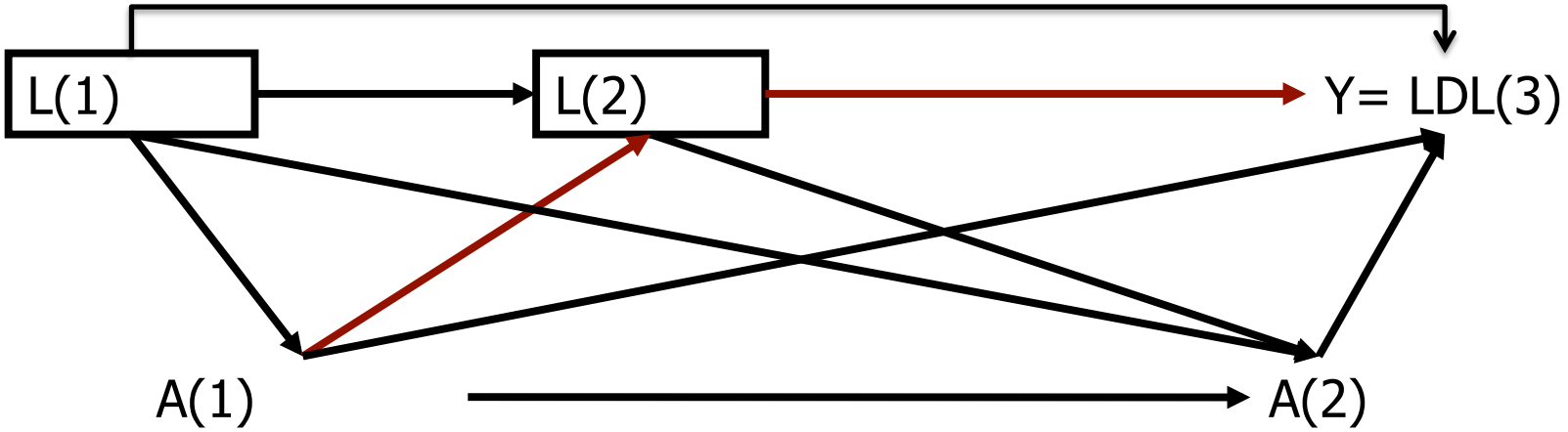
$$E(Y_{11}) \stackrel{?}{=}$$

$$\sum_{\bar{l}(1)} E(Y|A(1)=1, A(2)=1, \bar{L}=\bar{l})P(\bar{L}=\bar{l})$$



Why does the point treatment G-computation formula break down?

$$E(Y_{11}) \overset{?}{=} \sum_{\bar{l}(1)} E(Y|A(1)=1, A(2)=1, \bar{L}=\bar{l}) P(\bar{L}=\bar{l})$$



- We are losing the effect of A(1) on Y via L(2)
  - Point treatment identifiability result takes expectation with respect to the wrong distribution of covariates L

Longitudinal G computation Formula

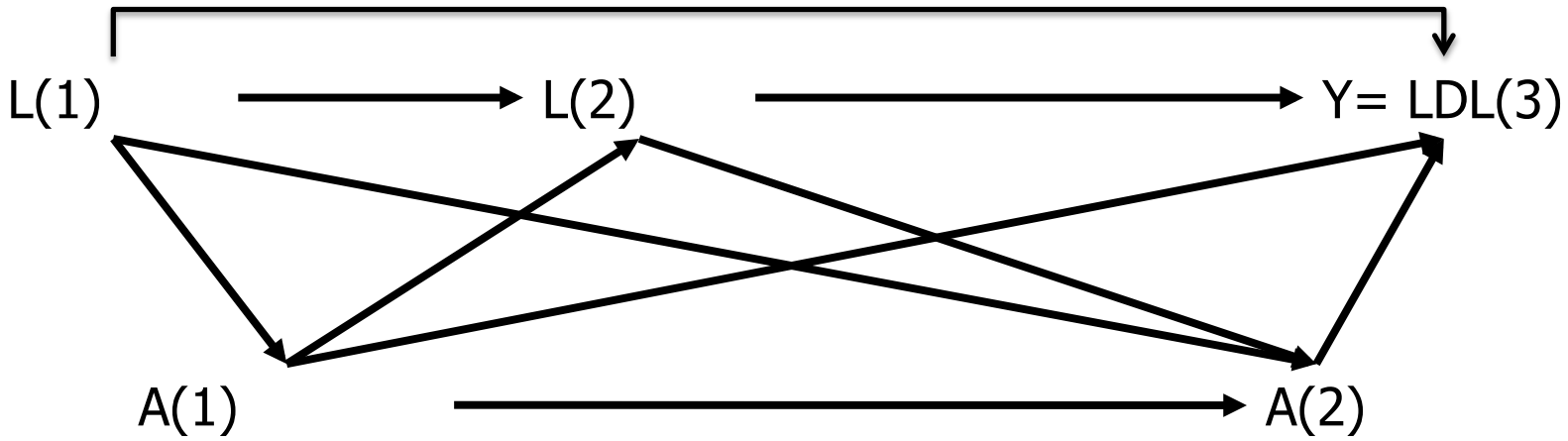
$$E(Y_{\bar{a}}) =$$

$$\sum_{\bar{l}} \left[E(Y|\bar{A}=\bar{a},\bar{L}=\bar{l}) \times \prod_{t=1}^{K} P\left(L(t)=l(t)|\bar{A}(t-1)=\bar{a}(t-1),\bar{L}(t-1)=\bar{l}(t-1)\right) \right]$$

- First term RHS: Captures only partial effect of A nodes on Y
- Second term RHS: Gives us the distribution of covariate history we need to take the expectation with respect to in order to capture effect of A nodes on Y through covariates L

- Note that, unless we make additional assumptions (such as assuming sequential randomization holds for the covariates L(t) in addition to outcome Y), neither term has causal interpretation on its own

Example: Longitudinal G-computation Formula



$$E(Y_{11}) = \sum_{l(1),l(2)} \left(\begin{array}{l} E(Y|A(1)=1, A(2)=1, L(1)=l(1), L(2)=l(2)) \times \\ P(L(2)=l(2)|A(1)=1, L(1)=l(1)) \times P(L(1)=l(1))) \end{array} \right)$$

Key insight: we don't need to adjust for everything all at once

- Instead, we can think of simulating our data sequentially from our set of structural equations
- This lets us consider the problem of identifiability sequentially
  - For each A(t) in sequence, ask if its effect on Y can be identified by conditioning on some subset of the observed past.

Identifiability for the effects of multiple interventions

- What do we need for identifiability in this case?
- Intuition: Sequentially Randomized Trial
  - At each time point, randomize A(t) within strata of (some subset of) covariates and treatment observed up until then
  - In this case, at each time point the effect of A(t) on future nodes is identified
    - We know we measured enough of the past the estimate the effect of intervening on that node
  - We can estimate the effect of setting each A(t) sequentially

Identifiability for multiple interventions

- Sequential Randomization Assumption

$$Y_{\bar{a}} \perp A(t) | \bar{L}(t) = \bar{l}(t), \bar{A}(t-1) = \bar{a}(t-1)$$
$$\text{for all } \bar{l} \text{ and } \bar{a}$$

  - If A(t) is randomly assigned at each time point, given the observed past, this will hold

- Counterpart to the Randomization Assumption for a single intervention

A graphical criterion for identifiability

- For the single time point case, the back door criterion:

1. Allowed us to evaluate if our target parameter was identified by the G computation formula
2. Helped us to decide what to condition on
   - What to include in W
   - Recall- not always a good idea to include everything

- Is there an equivalent criterion for target parameters with multiple interventions?

Identifying the effects of interventions on multiple nodes

- The sequential back door criterion (counterpart to the back door criterion for a single intervention node):

  For each intervention node A(t), every "intervention avoiding" backdoor path from A(t) to Y must be blocked by some subset of non-descendents of A(t)

  – Intervention avoiding= path with no arrows into an A node after A(t)

Sequential Back Door Criterion

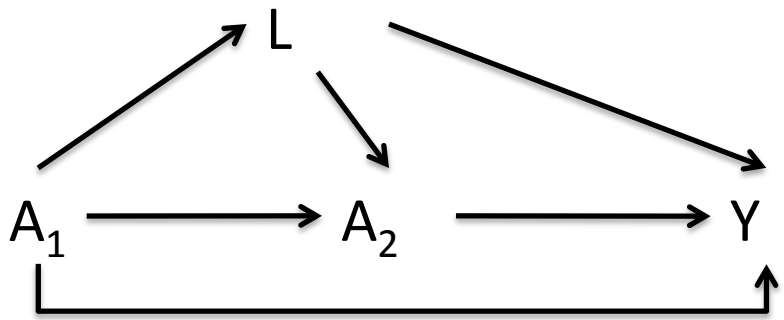
- Essentially we just want to apply the usual back door criterion, for each intervention node A(t) in series:
  1. We are looking for set of covariates (+ past treatment) that will block all back door paths from A (t) to the outcome
  2. These covariates cannot be descendents of A(t)
- Same justification: Want to remove any sources of association between each A(t) and the outcome other than those that we are interested in

Sequential Back door Criterion

- Just the standard back door criterion applied to each intervention node is sequence except
- Now it is OK if there is an unblocked back door path that goes through a future intervention node
- Why?
  - Any paths through future A nodes will already be blocked because we are intervening on them
  - We don't need to worry about blocking them

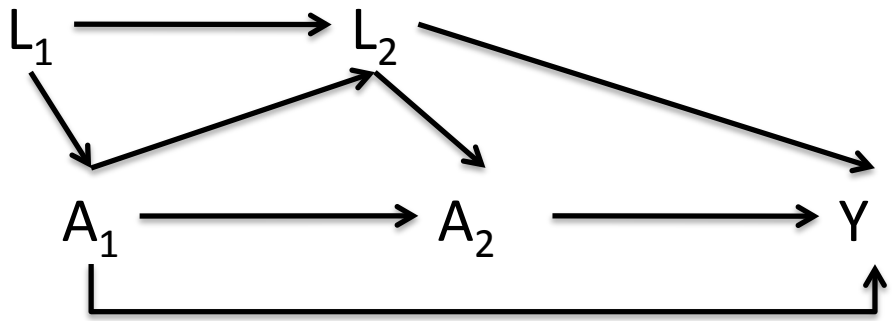
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



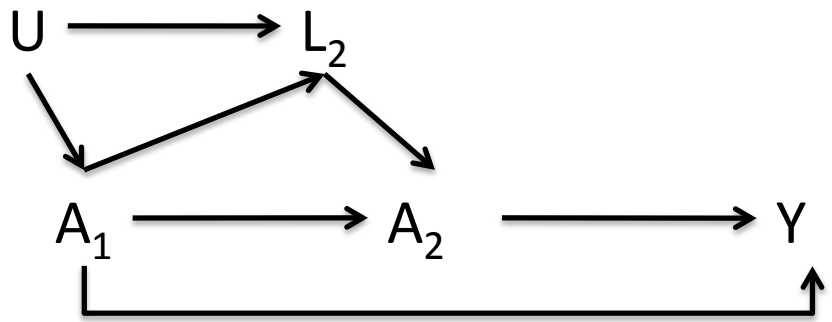
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



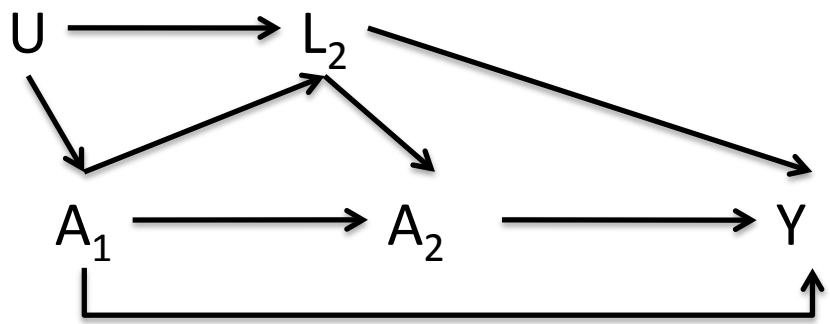
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



Identifiability Result

- Under the Sequential Randomization Assumption (i.e. if measured history sufficient to satisfy the sequential back door criterion):

$\Psi(P_{X,U,0})$: Causal Parameter of Interest

$$P(Y_{\bar{a}} = y) =$$

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y | \bar{A} = \bar{a}, \bar{L} = \bar{l}) \times \\ \prod_{t=1}^{K} P\left(L(t) = l(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)\right) \end{array} \right)$$

$\Psi(P_0)$: Target statistical parameter/estimand

Recall: Positivity Assumption

- In order for $\Psi(P_0)$ to be defined (in a non-parametric model), need each treatment history of interest to occur with some positive probability for each possible covariate history

$$\min_{a \in \mathbf{A}} g(a(t) \mid \bar{A}(t-1), \bar{L}(t)) > 0 \ a.e.$$

$\Psi(P_{X,U,0})$: Causal Parameter of Interest

$$P(Y_{\bar{a}} = y) =$$

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y | \bar{A} = \bar{a}, \bar{L} = \bar{l}) \times \\ \prod_{t=1}^{K} P\left(L(t) = l(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)\right) \end{array} \right)$$

$\Psi(P_0)$: Target statistical parameter/estimand

Classes of Estimator of the Target Parameter $\Psi(P_O)$

- Likelihood of the Observed Data

$$L(O) = \left(\prod_{t=1}^{K+1} Q\left(L(t) \mid \bar{A}(t-1), \bar{L}(t-1)\right) \prod_{t=1}^{K} g\left(A(t) \mid \bar{A}(t-1), \bar{L}(t)\right) \right)$$

Maximum Likelihood (Substitution)

Inverse Probability Weighted (Estimating Equation)

Targeted Maximum Likelihood (Substitution)
Augmented- Inverse Probability Weighted (Estimating Equation)

Efficient (in Non/Semi-Parametric Model) and Double Robust

Overview: Maximum Likelihood Substitution Estimator

$$\Psi(P_0) = \sum_{\bar{l}} \left(\begin{array}{l} P(Y = y|\bar{A} = \bar{a}, \bar{L} = \bar{l}) \times \\ \prod_{t=1}^{K} P\left(L(t) = l(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)\right) \end{array} \right)$$

- Our target statistical parameter $\Psi(P_0)$ is only a function of the *Q* factors of the observed data likelihood
  - Conditional distributions of the non-intervention covariates (including the outcome) given their parents

Overview: Maximum Likelihood Substitution Estimator

$$\Psi(P_0) =$$

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y|\bar{A} = \bar{a}, \bar{L} = \bar{l}) \times \\ \prod_{t=1}^{K} P\left(L(t) = l(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)\right) \end{array} \right)$$

1. Estimate these conditional distributions
2. Plug in the resulting estimates to get an estimate of $\Psi(P_0)$

- In practice- often use Monte Carlo simulation to average w.r.t the distribution of the covariates evaluated at the treatment history of interest

Implementation of Maximum Likelihood Substitution Estimator

1. Estimate the conditional distribution of each covariate L(t) given its parents (past covariates and treatment)

- Recall: L(t) may itself be a high dimensional vector
  - Multiple covariates measured at time point t
  - Can factorize L(t) into multiple conditional distributions
- Common approach relies on a series of parametric regression models
  - logistic regression, linear regression, etc, with parametric assumptions on the distribution of the errors

Implementation of Maximum Likelihood Substitution Estimator

- Simple ABC Example
  - Estimate the distribution of CHD risk factors at time 1 using the empirical distribution
  - Estimate the conditional distribution of risk factors at time 2 given baseline risk factors and Abacavir use at time 1
  - Estimate the conditional distribution of the outcome LDL (or, depending on the target parameter, just the expectation) given ABC use at times 1 and 2 and risk factors at times 1 and 2

Implementation of Maximum Likelihood Substitution Estimator

2. Use these estimates to "simulate counterfactual covariate histories over time" setting A(t)=a(t) for t=0,…,K
   - Draw L(1) from the empirical
   - Draw L(2) from estimate of the conditional distribution of L(2) given A(1) and L(1), setting A(1)=a(1) and L(1)= drawn value….
   - Etc.. until L(K+1)

Implementation of Maximum Likelihood Substitution Estimator

3. Repeat for each subject for each treatment history of interest to get estimate of the distribution of counterfactual outcome under that treatment history
   - Example: estimate the distribution of final LDL under intervention to always set abacavir use equal to 1 and under intervention to always set abacavir use equal to 0
   - Or under some other intervention on abacavir use
     - For example, according to a dynamic rule…

Generalizations of Maximum Likelihood Substitution Estimator

- To incorporate time-to-event outcome with right censoring:
  - Q factors of the likelihood condition on $t<T$, $C(t)=0$
  - Evaluate setting $A(t)=a(t)$ and $C(t)=0$
- To estimate parameters of working marginal structural model:
  - Simply regress simulated counterfactual outcomes on the treatment history used to generate them according to the specified marginal structural model

Limitations of Maximum Likelihood Substitution Estimator

- Point treatment: Relies on doing a good job estimating the conditional distribution of Y given A,W
- Longitudinal: Relies on doing a good job predicting the distribution of each covariate at each time point, given past covariates and past treatment/exposure

$$Q(L(t) \mid \overline{L}(t-1), \overline{A}(t-1)) : t = 1, \ldots, K+1$$

Limitations of Maximum Likelihood Substitution Estimator

- If we had sufficient knowledge to specify parametric models for the all the Q factors of the likelihood then this approach would be great
  - Just maximum likelihood estimation- efficient
- However, we essentially never have such knowledge
- Reliance on misspecified parametric models is an even bigger problem with longitudinal data

Limitations of Maximum Likelihood Substitution Estimator

- We can treat this as a series of prediction problems
  - Use loss-based learning/ cross validation/ super learner to aim for optimal estimates of each conditional distribution while respecting the non-parametric model
  - Density estimation is hard, but there are tricks we can use….
- However- even the best tools do not ensure that we will do a good job at estimating our target parameter
  - The right bias variance tradeoffs for the purposes of estimating each conditional distribution will be the wrong bias variance tradeoffs for our lower dimensional target parameter
  - Again, our causal effect estimate will be overly biased

Classes of estimator of the Target Parameter $\Psi(P_0)$

- Likelihood of the Observed Data

$$L(O) = \left(\prod_{t=1}^{K+1} Q\left(L(t) \mid \bar{A}(t-1), \bar{L}(t-1)\right) \right) \prod_{t=1}^{K} g\left(A(t) \mid \bar{A}(t-1), \bar{L}(t)\right)$$

Maximum Likelihood (Substitution)

Inverse Probability Weighted (Estimating Equation)

Targeted Maximum Likelihood (Substitution)

Augmented- Inverse Probability Weighted (Estimating Equation)

Efficient (in Non/Semi-Parametric Model) and Double Robust

Overview: Longitudinal IPTW Estimator

- The inverse probability (of treatment) weighted estimator (IPTW) provides an alternative approach
- Based on estimating the conditional distributions of the intervention nodes
  - How was the exposure assigned/censoring determined in the current data?
  - "Treatment mechanism"
- Not a substitution estimator. Instead, defined as the solution to an estimating equation

Intuition: Longitudinal IPTW Estimator

- Confounding as analogous to biased sampling
  - If exposure were randomly assigned at each time point, probability of exposure would be independent of past history
  - Instead, because exposure assignment depends on a subject's history, some covariate and exposure combinations are **over-represented** in our sample and others are **under-represented**
    - Compared to what would have been seen in a hypothetical randomized trial
- IPTW: Up-weight subjects with under-represented covariate and exposure combinations and down weight over-represented covariate and exposure combinations

Example: Intuition Behind Longitudinal IPTW Estimator

- If ABC use were randomly assigned at each time point, subjects with higher and lower CHD risk would be equally likely to be treated with ABC
- Instead, say subjects with renal disease preferentially get treated with ABC
  - Subjects with renal disease treated with ABC over-represented in our sample
    - Those subjects who have this covariate/treatment combination get smaller weights
  - Subjects with renal disease not treated with ABC under-represented in our sample
    - Those subjects who do have this covariate/treatment combination get bigger weights

Implementation: Longitudinal IPTW

1. Estimate treatment mechanism
   - Distribution of intervention nodes given the observed past for each time point *t=1,…,K:*

$$g\left(A(t) \mid \overline{A}(t-1), \overline{L}(t)\right)$$

- ABC Example: Estimate the probability of being treated with abacavir in a given month given covariate (eg CHD risk factor) and abacavir treatment history up till that month

Implementation: Longitudinal IPTW

2. For each subject and time point, estimate the predicted probability of the subject receiving his observed exposure at that time point
   - Given that subject's covariate and treatment history
   - For i=1,..,n; t=1,…,K $\hat{g}\left(A_i(t) \mid \overline{A}_i(t-1), \overline{L}_i(t)\right)$

- ABC Example:
  - For time points treated with abacavir, predicted probability of being treated given observed past
  - For time points not treated, predicted probability of not being treated given observed past

Implementation: Longitudinal IPTW

3. Estimate the predicted probability of a subject having his observed treatment history
   - Product of time point-specific predicted probabilities

$$\prod_{t=0}^{K} \hat{g}\left(A_i(t) \mid \overline{A}_i(t-1), \overline{L}_i(t)\right)$$

- Weight is inverse of this predicted probability (for subjects with observed treatment history=treatment history of interest)

$$\hat{w}_i = \frac{1}{\prod_{t=0}^{K} \hat{g}\left(A_i(t) \mid \overline{A}_i(t-1), \overline{L}_i(t)\right)}$$

Implementation: Longitudinal IPTW

4. Take weighted average of observed outcome across the population

$$\hat{E}(Y_{\bar{a}}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{I(\bar{A}_i = \bar{a})}{\prod_{t=0}^{K}\hat{g}\left(A_i(t) \mid \bar{A}_i(t-1), \bar{L}_i(t)\right)} Y_i\right)$$

- Subjects who did not receive the treatment history of interest get weights=0
- Subjects who did receive the treatment history of interest get weights inversely proportional to their predicted probability of receiving their observed treatment history given their observed past

IPTW Estimator for a Longitudinal Marginal Structural Model

- Example: target causal parameter $E(Y_{\bar{a}}) = m(\bar{a} \mid \beta)$
  - Ex: $m(\bar{a} \mid \beta) = \beta_0 + \beta_1 \sum_{t=0}^{K} a(t)$
- IPTW estimator solves the estimating equations associated with the following estimating function:

$$\frac{h(\bar{A})}{\prod_{t=1}^{K} g(A(t) \mid \bar{L}(t), \bar{A}(t-1))} \left(Y - m(\bar{A} \mid \beta) \right)$$

  - *h* is a user-supplied non-null function of treatment history
  - If we believe our MSM, choice of *h* affects efficiency, not consistency
  - If our target parameter is defined using a working MSM, choice of h defines the projection

IPTW Estimator for a Longitudinal Marginal Structural Model

- One choice of h: $h(A) = \frac{d}{d\beta} m(\bar{A}|\beta) g(\bar{A})$
  - As in point treatment case, appealing because
    - It lets us solve for β using standard software
    - If the is no confounding, it reduces to standard least squares estimator
    - Can improve efficiency by stabilizing weights
- IPTW Estimator is solution in β to :

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}(\overline{A}_i)\frac{d}{d\beta} m(\overline{A}_i \mid \beta)}{\prod_{t=1}^{K} \hat{g}(A_i(t) \mid \overline{L}_i(t), \overline{A}_i(t-1))} \left(Y_i - m(\overline{A}_i \mid \beta) \right)$$

IPTW Estimator for a Longitudinal Marginal Structural Model

- Fit weighted regression of observed outcome Y on observed treatment history according to model $m(\overline{A} \mid \beta)$
- With stabilized weights $$s\hat{w}_i = \frac{\hat{g}(\overline{A}_i)}{\prod_{t=1}^{K} \hat{g}\left(A_i(t) \mid \overline{A}_i(t-1), \overline{L}_i(t)\right)}$$
- For example of IPTW estimator of MSM parameter for time to event outcome with right censoring, see Chapter 24 in Targeted Learning Book

Recall: Positivity Assumption

- In order for $\Psi(P_0)$ to be defined (in a non-parametric model), need each treatment history of interest to occur with some positive probability for each possible covariate history

$$\min_{a \in \mathbf{A}} g(a(t) \mid \bar{A}(t-1), \bar{L}(t)) > 0 \ a.e.$$

$\Psi(P_{X,U,0})$: Causal Parameter of Interest

$$P(Y_{\bar{a}} = y) =$$

$\Psi(P_0)$: Target statistical parameter/estimand

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y | \bar{A} = \bar{a}, \bar{L} = \bar{l}) \times \\ \prod_{t=1}^{K} P\left(L(t) = l(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)\right) \end{array} \right)$$

Positivity Assumption

- Stabilized weights allow weaker ETA:

$$\max_{\bar{a}\in\mathbf{A}} \frac{g(\bar{a} \mid \bar{a}(t-1))}{\prod_{t=0}^{K} g(a(t) \mid \bar{a}(t-1), \bar{L}(t))} < \infty - \ a.\text{e.}$$

  - Relies on model $m(\bar{A} \mid \beta)$ to smooth over sparse areas of A
- But when target parameter is defined using a working MSM, use of stabilized weights change the target parameter
  - See Neugebauer&vdL 2007

Limitations of IPTW

- Inefficient
- Highly susceptible to bias arising from positivity violations/ near violations
  - In other words, tends to behave badly in the presence of strong confounding
- Have to do a good job estimating treatment/censoring mechanism
  - Again… Super Learner
  - But what covariates to include?
    - Covariates may be strong predictors of A, but not be confounders
    - At a minimum, do not blindly include all predictors of treatment assignment
  - The data adaptive fit of treatment mechanism is not targeted at the parameter of interest…

Recap: What have we done?

1. Specified causal parameters indexed by interventions on multiple nodes (and corresponding SCM)
   - Ex: longitidinal treatments
   - Ex: censoring
2. Specified classic types of observed data used to address causal queries of this nature
3. Discussed identifiability for these parameters
   - Extended assumption of no unmeasured confounding to interventions on multiple nodes
4. Introduced a new estimand: the longitudinal G-computation formula
5. Introduced two classes of estimators for this estimand

Third Class of Estimator: Double robust efficient estimators

- Implementation requires estimating both g and Q components of the likelihood
- Consistent if either is estimated consistently
- Efficient if both are estimated consistently
- A double robust estimator that is also a substitution estimator: TMLE
  - Details and real data example for longitudinal TMLE coming up next

References- Causal Inference with Longitudinal Data, Working MSM

- Petersen and van der Laan. Chapter 24. Case Study: Longitudinal HIV Cohort Data. In M.J. van der Laan and S. Rose, Targeted Learning: Causal Inference for Observational and Experimental Data. Springer, Berlin Heidelberg New York, 2011.

- Robins and Hernan. Chapter 23. Estimation of the causal effects of time-varying exposures. In. Longitudinal Data Analysis. Eds: Fitzmaurice, Davidian, Verbeke, Molenberghs. 2009. CRC Press.

- Robins. Association Causation and Marginal Structural Models. *Synthese* 1999.

- Neugebauer and van der Laan. Nonparametric causal effects based on marginal structural models. Journal of Statistical Planning and Inference. 2007. 137: 419-434.

References- Identifiability, Time-dependent confounding, Longitudinal MLE

- Pearl. Causality. Models, Reasoning, and Inference. Cambridge University Press, 2000, 2nd Ed 2009. Pp 121-126, 352-353

- Robins, Hernan, Brumback. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 2000. 11(5) 550-560

- Taubman et al. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. Intl. J. Epi. 2009. 38:1599-1611.

References- Longitudinal IPTW

- Robins, Hernan, Brumback. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 2000. 11(5) 550-560

- Bodnar et al. Marginal Structural Models for analyzing causal effects of time-dependent Treatment: An application in Perinatal Epidemiology. Epidemiology (2004). 159(10):926-934

- Cole and Hernan. Constructing Inverse Probability Weights. AJE (2008) 168:656-664.