Longitudinal Data
and
Additional Target Causal Quantities

Lecture 4

Outline: Longitudinal Data

1. Repeated point treatment data structures
2. Estimating the effects of intervening on more than one node
   - New causal quantities
     - Cumulative treatment effects
     - Longitudinal Marginal Structural Models
     - Right censoring
     - Direct Effects
     - Dynamic regimes
   - New Identifiability Result/Estimands
   - New 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
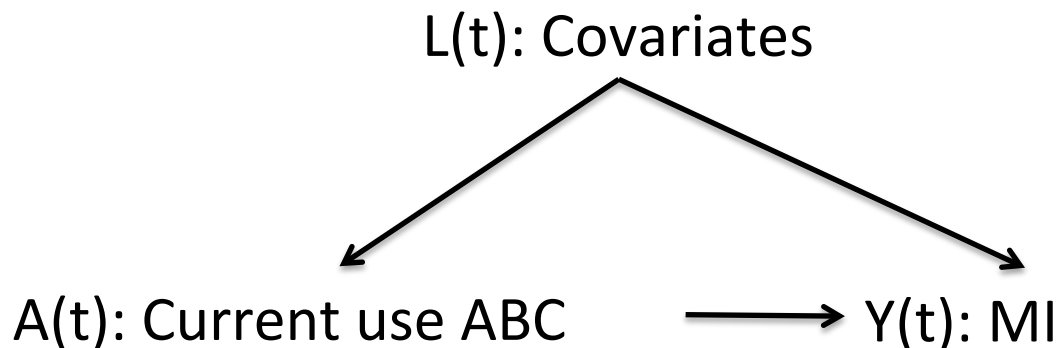- O(t)=(L(t),A(t),Y(t)), t=1,…,K

Example: Effect of current abacavir use on MI risk?

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

$$L(t) = f_{L(t)}(U_{L(t)})$$

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

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

L(t): Covariates

A(t): Current use ABC $\longrightarrow$ Y(t): MI

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
- Possible target causal quantity
  - $E(Y_1(t)-Y_0(t)|Y(t-1)=0)$
  - Difference in risk of (new) CHD at time t if did vs. did not use abacavir

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
- Allows us to use Model, Data, Identifiability Result, and Estimators previously introduced
  - Can use TMLE package 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

Cumulative effects of longitudinal treatments?

- 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

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

- A common SCM: Assumes each variable may be affected by all preceding variables
  - ie. Parents(X)= all variables that temporally precede X

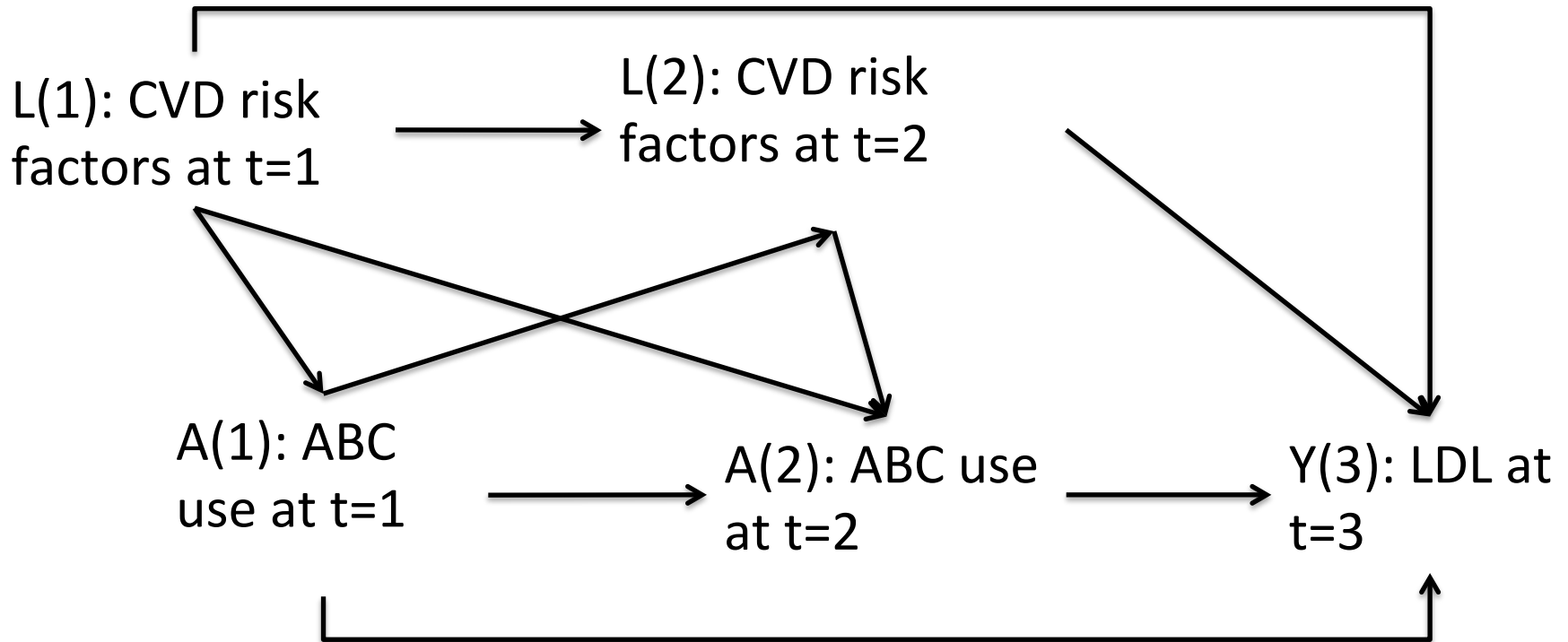$$L(1) = f_{L(1)}(U_{L_1})$$
$$A(1) = f_{A(1)}(L(1), U_{A(1)})$$
$$L(t) = f_{L(t)}(\bar{A}(t-1), \bar{L}(t-1), U_{L(t)}), t = 2, ..., K+1$$
$$A(t) = f_{A(t)}(\bar{A}(t-1), \bar{L}(t), U_{A(t)}), t = 2, ..., K$$

Simplified Abacavir Example

- Say 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
- 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

  - no subjects had used abacavir at t=1 and t=2

Abacavir Example: Longitudinal Causal Graph

L(1): CVD risk factors at t=1

L(2): CVD risk factors at t=2

A(1): ABC use at t=1

A(2): ABC use at t=2

Y(3): LDL at t=3

Counterfactuals indexed by longitudinal exposures

- Original SCM

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

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

Counterfactuals indexed by longitudinal exposures

- Original SCM

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

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

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

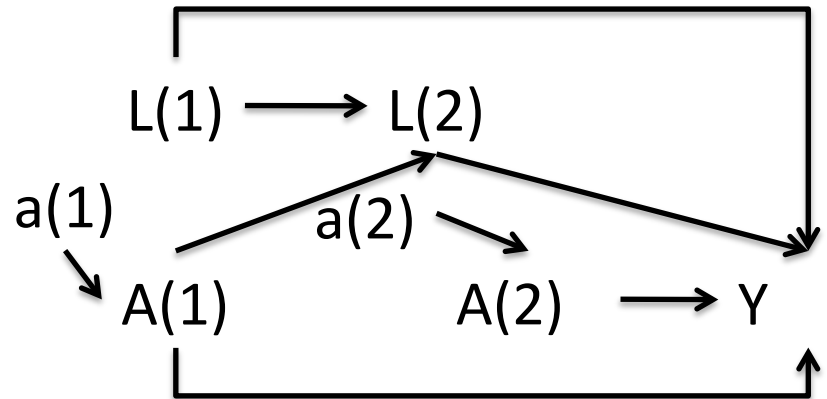# Counterfactuals indexed by longitudinal exposures

- Modified SCM/Graph

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

L(1) ⟶ L(2)

a(1)

a(2)

A(1)   A(2) ⟶ Y

- Defines counterfactual outcome intervening on ABC use at two time points:

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

Intervention on counterfactual exposure history

Example causal target quantity

- 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 counterfactual parameter

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

  - Difference in expected LDL if all subjects had versus had not used abacavir at t=1 and t=2

Defining target causal quantity using a Longitudinal Marginal Structural Model

- Example: How does cumulative time exposed to abacavir affect LDL at the end of the study?
  - Ex. Working MSM

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

- How does this effect differ depending on baseline renal function (V)?
  - Ex. Working MSM

$$E(Y_{\bar{a}}|V) = \beta_0 + \beta_1 \sum_{t=1}^{K} a(t) + \beta_2 V + \beta_3 V \times \sum_{t=1}^{K} a(t)$$

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 quantities with survival outcome

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

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

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

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

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

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 equations)

- Define C as time when leave the cohort
  - Censoring time

- $C(t)=I(C>t)$
  - Indicator still in follow up at time t

Example of an SCM with censoring

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

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 causal quantity 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 causal quantity 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)

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

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 quantities with survival outcome and censoring

- Discrete counterfactual hazard:

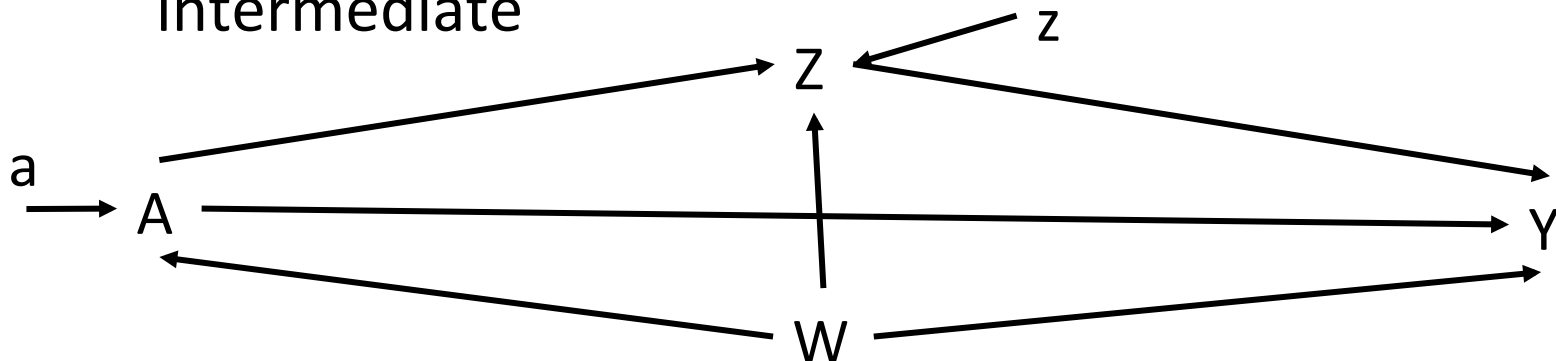$$P(Y_{\bar{a},\bar{c}=0}(t) = 1 | Y_{\bar{a},\bar{c}=0}(t-1) = 0)$$

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

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

Additional target causal quantities: Effect Mediation

- Interventions on more than one node can also be used to study effect mediation

  - Ex: How much of the effect of abacavir (A) on MI (Y) is due to changes in inflammatory biomarker (Z)?

- Define counterfactual outcome setting the levels of both treatment (A) and intermediate (Z): Y_{az}

  - By fixing level of intermediate, effect of treatment on outcome cannot be mediated via changes in intermediate

Effect Mediation Target Causal Quantities

- Controlled Direct Effect: $E(Y_{1z}-Y_{0z})$
  - Definition, identification and estimation results follow directly from those for longitudinal exposures

- Other effect mediation parameters involve nested counterfactuals

  - Z_a: counterfactual value of intermediate under treatment level a

  - Natural Direct Effect: $E(Y_{1Z_0} - Y_{0Z_0})$

  - Indirect Effect: $E(Y_{1Z_1} - Y_{1Z_0})$

Additional target causal quantities: Effects of Dynamic Regimes

- <u>Static regime</u>: Set each intervention node equal to some constant
  - Irrespective of subject characteristics
  - Ex: Always use abacavir

- <u>Dynamic regime</u>: A subject-responsive strategy for assigning treatment
  - Assign a value to each intervention node based on some known function of the observed past

Effects of Dynamic Regimes

- Ex. Dynamic regime
  - Always use abacavir **unless** a contraindication (CI) develops, in which case switch to other drug
  - Ex: set Abacavir use at time t according to rule $d_t(CI(t))$:

$$d_t(CI(t)) = 1 \text{ if } CI(t) = 0$$
$$= 0 \text{ if } CI(t) = 1$$

- Effects of dynamic regimes can be defined analogously to effects of static treatment regimens
  - Ex: $E(Y_{\bar{d}}(t) - Y_{\bar{0}}(t)),$

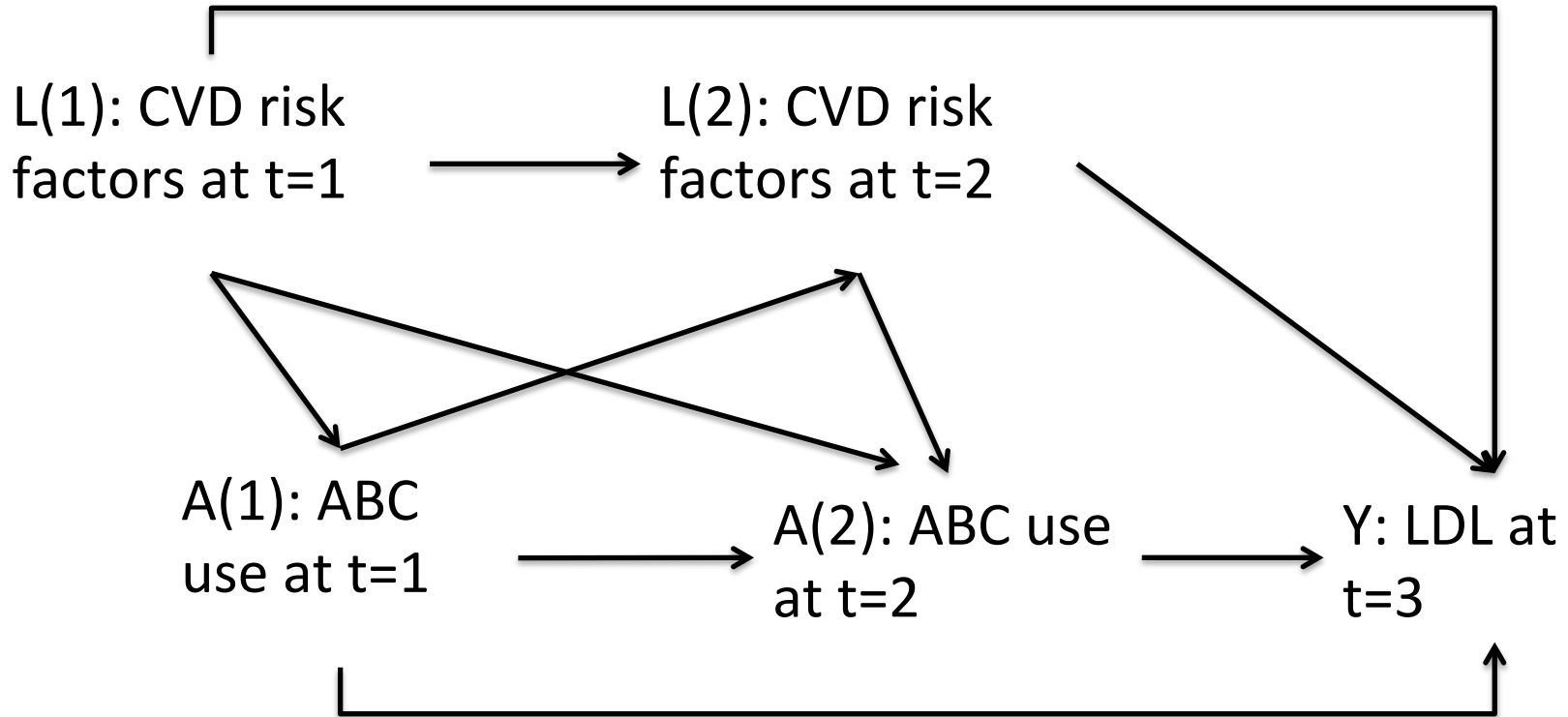$$\text{where } \bar{d} = d_1(CI(1)), d_2(CI(2)), \ldots, d_t(CI(t))$$

Dynamic Marginal Structural Models

- Dynamic regime might also be indexed by some threshold θ

  - Ex. Don't use abacavir (ie use alternative such as tenofovir) **unless renal function falls below some value** θ, in which case switch to abacavir

  - Ex. set Abacavir use according to rule d_θ(CI):
  $$d_\theta(RF(t)) = 0 \text{ if } RF(t) \geq \theta$$
  $$= 1 \text{ if } RF(t) < \theta$$

- MSM can be used to summarize how expected counterfactual outcome varies as a function of θ

  - Ex: $E(Y_{\bar{d}_\theta}) = m(\theta|\beta)$

Identifiability for longitudinal exposures

- What causal assumptions are sufficient for our target causal quantity to be identified as a parameter of the observed data distribution?
    1. Requires new assumptions (beyond the RA)
    2. Results in new target statistical parameters (estimands)
    3. And thus requires new estimators

- 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(1): CVD risk factors at t=1   L(2): CVD risk factors at t=2

A(1): ABC use at t=1   A(2): ABC use at t=2   Y: LDL at t=3

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

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

- Under what conditions can we write our target causal quantity as a parameter of the observed data distribution?

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

  - Graphical counterpart to backdoor criterion="sequential back door criterion"

    - (see eg Pearl, Causality, p. 352)
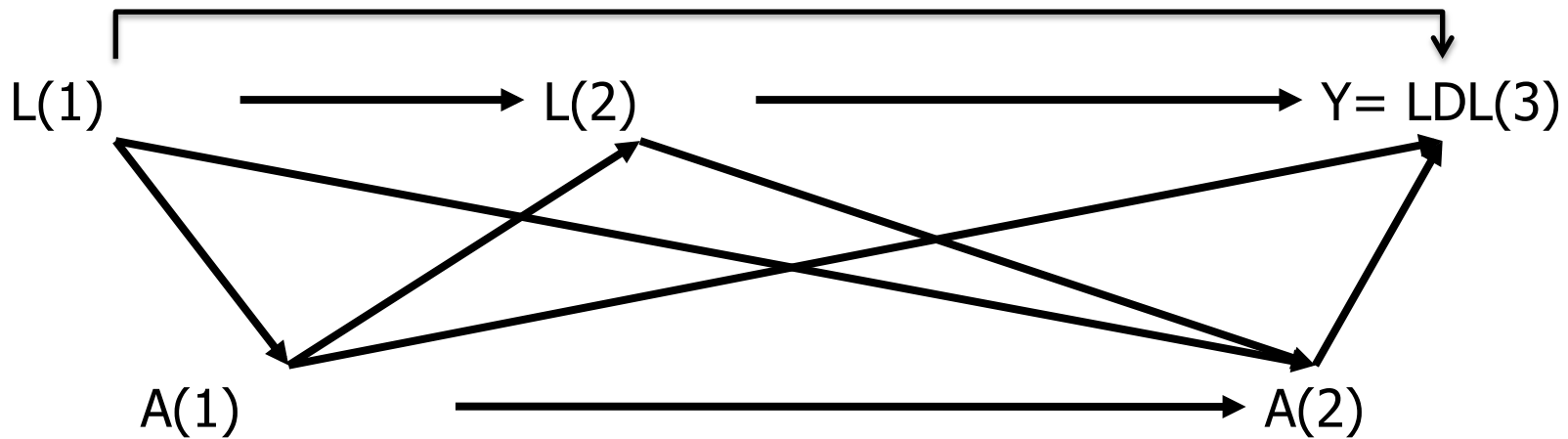
Identifiability Result

- Under the Sequential Randomization Assumption:

$\Psi^F(P_{X,U,0})$: Target causal quantity

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

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y | \bar{A} = \bar{a}, \bar{L} = \bar{l}) \\ \prod_{t=0}^{K} P(L(t) = l(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)) \end{array} \right)$$

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

Example: Longitudinal G-computation Formula

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

Positivity Assumption

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

$$\inf_{\bar{a} \in \mathcal{A}} g_0(a(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t)) > 0 - a.e.$$

- Positivity violations are common
  - Some types of patients may develop absolute indications or contraindications for some treatments
    - Ex. g(ABC(t)=1|Contraindication(t))=0
  - Can also have lack of support in finite samples due to chance

Responding to Positivity Violations

- One Option: Realistic Treatment Rules
  - Type of dynamic regime: assign treatment based on observed past

  - Positivity assumption for a dynamic regime:
  $$\inf_{\bar{d}\in\mathcal{D}} g_0(d(t)|\bar{A}(t-1)=\bar{d}(t-1),\bar{L}(t)) > 0 - a.e.$$

- Realistic rules avoid positivity violations by only assigning treatment values for which there is adequate support in the data
  - Ex: Treat with ABC <u>unless</u> a contra-indication develops
    - g(ABC(t)=d_t(Contraindication(t))|Contraindication(t))=1

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

- ## Likelihood of the Observed Data

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

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

$$\Psi(P_0) =$$

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y | \bar{A} = \bar{a}, \bar{L} = \bar{l}) \\ \prod_{t=0}^{K} P(L(t) = l(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)) \end{array} \right)$$

- 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

$$\Psi(P_0) =$$

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y | \bar{A} = \bar{a}, \bar{L} = \bar{l}) \\ \prod_{t=0}^{K} P(L(t) = l(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)) \end{array} \right)$$

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 many times 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:

  - 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 <u>each covariate</u> at <u>each time point</u>, given past covariates and past treatment/exposure

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

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….

Limitations of Maximum Likelihood Substitution Estimator

- 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 Ψ(P₀)

- ## Likelihood of the Observed Data

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

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 is 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

Intuition: Longitudinal IPTW Estimator

- IPTW:
  - Up-weight subjects with under-represented covariate and exposure combinations
  - 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 without renal disease treated with ABC under-represented in our sample
    - Those subjects who 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:*

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

- ABC Example: Estimate the probability of being treated with abacavir in a given month given covariate 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 $\quad \hat{g}\left(A_i(t) \mid \overline{A}_i(t-1), \overline{L}_i(t) \right)$

- 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 <u>history</u>

   – 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)

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

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

- Target causal quantity: $E(Y_{\bar{a}}) = m(\bar{a} \mid \beta)$

  - 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: $h(A) = \dfrac{d}{d\beta} m(\bar{A}|\beta) g(\bar{A})$

  – As in point treatment case, appealing because

    - It lets us solve for β using standard software

    - If there is no confounding, estimator reduces to standard least squares estimator

    - Can improve efficiency by stabilizing weights

- IPTW Estimator is solution in β to :

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

IPTW Estimator for a Longitudinal Marginal Structural Model

- Fit weighted regression of observed outcome Y on observed treatment history according to model

$$m(\bar{A} \mid \beta)$$

- With stabilized weights

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

- For example of IPTW estimator of MSM parameter for time to event outcome with right censoring, see Chapter 24 in Targeted Learning Book

Positivity Assumption

- Stabilized weights allow weaker ETA:

$$\sup_{\bar{a} \in \mathcal{A}} \frac{g_0(a(t)|\bar{A}(t-1) = \bar{a}(t-1))}{g_0(a(t)|\bar{A}(t-1) = \bar{a}(t-1)\bar{L}(t))} < \infty - ae$$

  - 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 changes the target parameter
  - See Neugebauer&vdL 2007

Limitations of IPTW

- Inefficient

- Susceptible to bias arising from positivity violations/near violations
  - In other words, tends to behave badly in the presence of strong confounding

Limitations of IPTW

- Have to do a good job estimating treatment/censoring mechanism
  - Again... data adaptive methods are an option
  - But...
    - 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 P(A=a|W) is not targeted at the parameter of interest...

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 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- Direct Effects/ Effect Mediation

- Robins & Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992;3:143-55

- Pearl. Direct and Indirect Effects. *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*. 2001: 411-420

- Robins Semantics of Causal DAG models and the Identification of Direct and Indirect Effects In: Highly Structured Stochastic Systems 2002

- Petersen, Sinisi, van der Laan. Estimation of Direct Causal Effects. *Epidemiology.* 2006;17: 276-284

- van der Laan & Petersen. Direct Effect Models. 2005. http:// www.bepress.edu/ucbbiostat/paper187.

References- Dynamic Regimes and Realistic Treatment Rules

- Hernan et. al. (2006) Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin. Pharm Tox*; 98:237–242

- van der Laan & Petersen. (2007) Causal Effect Models for Realistic Individualized Treatment and Intention to Treat Rules. *Int. J. Biostat. 2007* 3.1: 3.

- Robins, et. al. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Stat Med.* Oct 15;27(23):4678-721

References- Identifiability, Time-dependent confounding, Longitudinal MLE

- Pearl. <u>Causality. Models, Reasoning, and Inference.</u> 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.