Causal Effect Moderation (Modification) When Treatment or Exposure is Time-Varying

Daniel Almirall

Health Services Research in Primary Care, Durham VA MC

Dept of Biostatistics & Bioinformatics, Duke University MC

Collaborators:

Beth Ann Griffin, Rajeev Ramchand,

Andrew R. Morral, Daniel F. McCaffrey,

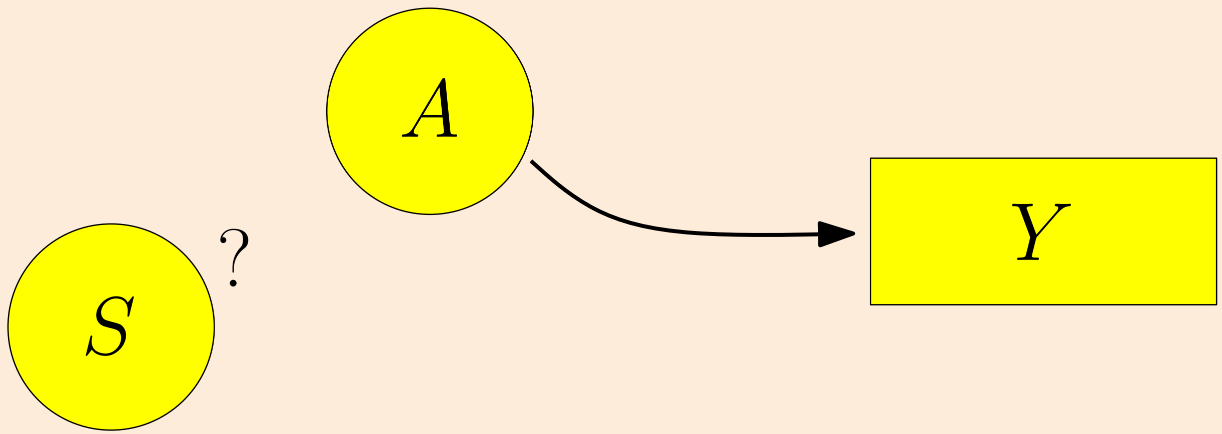
Thomas R. Ten Have, Susan A. Murphy,

September 14-15, 2009

Federal Interagency Subgroups Analysis Meeting
Washington, DC

Contents

1 Warm-up: Suppose we want $A \to Y$.



Examples

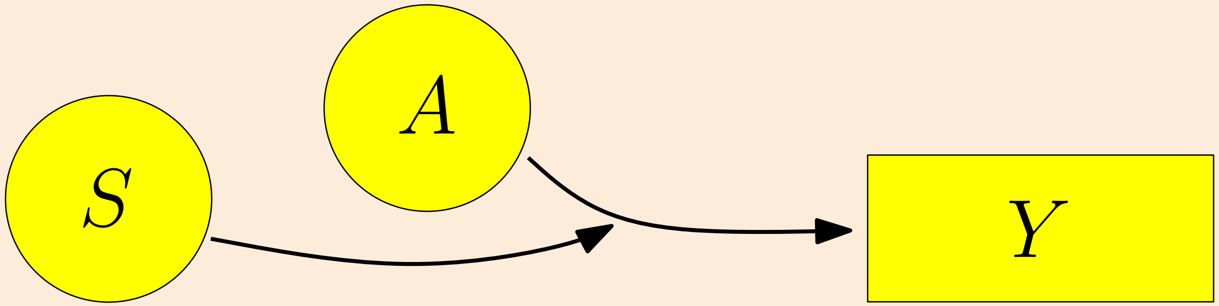
S = pre-A covt	A = txt/expsr	Y = outcome
Suicidal?	Medication?	Depression
Gender,SES	SAT Coaching?	SAT Math Score
Social Support	Inpatient vs. Outpatient	Substance Abuse

Why condition on ("adjust for") pre-exposure covariables S?

Suppose we want the effect of A on Y. Why condition on (adjust for) pre-treatment (or pre-exposure) variables S?

1. **Confounding**: S is correlated with both A and Y. In this case, S is known as a "confounder" of the effect of A on Y.
2. **Precision**: S may be a pre-treatment measure of Y, or any other variable highly correlated with Y.
3. **Missing Data**: The outcome Y is missing for some units, S and A predict missingness, and S is associated with Y.
4. **Effect Heterogeneity**: S may moderate, temper, or specify the effect of A on Y. In this case, S is known as a "moderator" of the effect of A on Y.

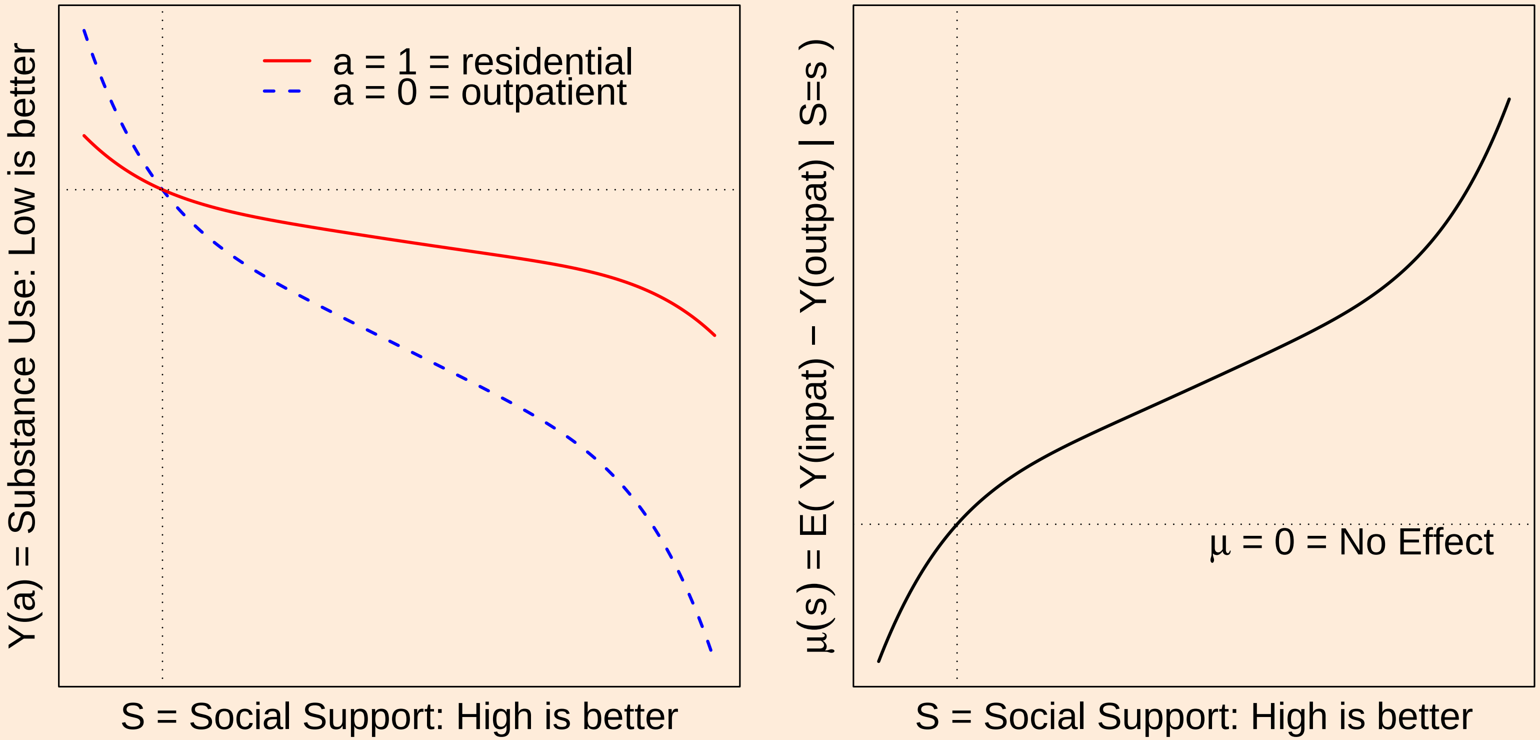
Suppose we want the effect of A on Y. Why condition on (adjust for) pre-treatment (or pre-exposure) variables S?



4. **Effect Heterogeneity**: S may moderate, temper, or specify the effect of A on Y. In this case, S is known as a "moderator" of the effect of A on Y. Formalized in next slide.

2 Effect Moderation in One Time Point

$$\mu(s, a) \equiv E(Y(a) - Y(0) \mid S = s)$$



Outpatient substance abuse treatment is better than residential treatment for individuals with higher levels of social support.

Causal Effect Moderation in Context: Relevance?

Theoretical Implication: Understanding the heterogeneity of the effects of treatments or exposures enhances our understanding of various (competing) scientific theories; and it may suggest new scientific hypotheses to be tested.

Elaboration of Yu Xie's <u>Social Grouping Principle</u>: We really want $Y_i(a) - Y_i(0) \; \forall \; i$. We settle for "groupings" of effects (here, groupings by S); $\mu(s, a)$ "comes closer" than $E(Y(a) - Y(0))$.

Practical Implication: Identifying types, or subgroups, of individuals for which treatment or exposure is not effective may suggest altering the treatment to suit the needs of those types of individuals.

** On Tailoring: Personalized Social, Behavioral, and Medical Treatments Programs **

The causal effect of interest (for most of us in this room) is

$$\mu(s, a) \equiv E(Y(a) - Y(0) \mid S = s)$$

This is the **Causal Effect Moderation Function**.

Developing **tailored treatments for personalized medicine** or **tailored social programs** is intimately tied to understanding $\mu(s, a)$.

This is, in fact, **the driving practical motivation** for what we have been working on here over the last 2 days.

** On Language: Homogeneity? **

The causal effect of interest (for most of us in this room) is

$$\mu(s,a) \equiv E(Y(a) - Y(0) \mid S = s)$$

This is the **Causal Effect Moderation Function**.

The word homogenous is misleading even if we find that S is not a moderator.

It is unlikely that the effect of treatment is homogenous (constant across the population) even if we find that the average treatment effect does not differ by S; that is, even if we find that $\mu(s,a)$ is constant in S.

Let's use the phrase **homogenous with respect to** S.

3 Mean Model in One Time Point

Decomposition of the conditional mean $E(Y(a) \mid S)$; and the prototypical linear model:

$$
\begin{aligned}
E(Y(a) \mid S = s) &= E(Y(0) \mid S = 0) \\
&\qquad + \Big(E(Y(0) \mid S = s) - E(Y(0) \mid S = 0)\Big) \\
&\qquad + E(Y(a) - Y(0) \mid S = s) \\
&= \eta_0 + \phi(s) + \mu(s, a) \\
&\overset{e.g.}{=} \eta_0 + \eta_1 s + \beta_1 a + \beta_2 a s.
\end{aligned}
$$

This is precisely what I would do, too.

4 The Time-Varying Setting

The data structure in the time-varying setting is:

S_1 $\quad a_1$ $\quad S_2(a_1)$ $\quad a_2$ $\quad \boxed{Y(a_1, a_2)}$

PROSPECT (Prevention of Suicide in Primary Care Elderly: CT)

(a_1, a_2)	Time-varying treatment pattern; a_t is binary (0,1)
$Y(a_1, a_2)$	Depression at the end of the study; continuous
S_1	Suicidal Ideation at baseline visit; continuous
$S_2(a_1)$	Suicidal Ideation at second visit; continuous

We were interested in assessing the causal effect of time-varying treatment for depression, as a function of other variables that may lessen or increase this effect (ie, effect moderation).

The Scientific Question Dictates Model Choice

This is especially important in the time-varying setting.

There are two types of scientific questions involving causal effect moderation in the time-varying setting:

Type A: What is the effect of switching off treatment for depression early versus later, as a function of **only baseline** suicidal ideation (or age, race, etc.)?

Type B: What is the effect of switching off treatment for depression early versus later, as a function of **baseline and time-varying** suicidal ideation?

What is the effect of switching off treatment for depression early versus later, as a function of baseline suicidal ideation (or age, race, etc.)?

Answering this type of question involves conditioning on baseline variables (putative moderators) thought to moderate the impact of different sequences of treatment (e.g., treatment duration) on outcomes.

Importantly, because they are collected at baseline, the putative moderators are not outcomes of prior treatment.

Marginal Structural Models are suitable for answering these types of questions.

What is the effect of switching off treatment for depression early versus later, as a function of baseline and time-varying suicidal ideation?

Answering this type of question involves conditioning on both baseline and time-varying variables thought to moderate the impact of different sequences of treatment on outcomes.

The issue here is that the intermediate time-varying moderators are themselves likely impacted by prior treatment. This has conceptual as well as statistical implications (more on this later).

Structural Nested Mean Models are suitable for answering these types of questions.

5 Robins' Marginal Structural Model

The MSM for the conditional mean of $Y(a_1, a_2)$ given S_1 is:

$$
\begin{aligned}
E\big(Y(a_1, a_2) \mid S_1\big) = E(Y(0,0) \mid S_1) \\
&+ \mathbf{E}\big(\mathbf{Y}(\mathbf{a_1}, \mathbf{0}) - \mathbf{Y}(\mathbf{0}, \mathbf{0}) \mid \mathbf{S_1}\big) \\
&+ \mathbf{E}\big(\mathbf{Y}(\mathbf{a_1}, \mathbf{a_2}) - \mathbf{Y}(\mathbf{a_1}, \mathbf{0}) \mid \mathbf{S_1}\big)
\end{aligned}
$$

$$
= \mu_0(s_1) + \boldsymbol{\mu_1}(\mathbf{s_1}, \mathbf{a_1}) + \boldsymbol{\mu_2}(\mathbf{s_1}, \mathbf{a_1}, \mathbf{a_2})
$$

$$
\begin{aligned}
&\stackrel{e.g.}{=} \beta_{01} + \beta_{02} s_1 + \boldsymbol{\beta_{10}} \mathbf{a_1} + \boldsymbol{\beta_{11}} \mathbf{a_1} \mathbf{s_1} \\
&\quad + \boldsymbol{\beta_{20}} \mathbf{a_2} + \boldsymbol{\beta_{21}} \mathbf{a_2} \mathbf{s_1}
\end{aligned}
$$

6 Robins' Structural Nested Mean Model

The SNMM for the conditional mean of $Y(a_1, a_2)$ given $\bar{S}_2(a_1)$ is:

$$
\begin{aligned}
&E\big(Y(a_1, a_2) \mid S_1, S_2(a_1)\big) \\
&\quad = E(Y(0,0)) + \Big\{E(Y(0,0) \mid S_1) - E(Y(0,0))\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big(\mathbf{Y}(\mathbf{a_1}, \mathbf{0}) - \mathbf{Y}(\mathbf{0}, \mathbf{0}) \mid \mathbf{S_1}\big)\Big\} \\
&\qquad\qquad + \Big\{E(Y(a_1, 0) \mid \bar{S}_2(a_1)) - E(Y(a_1, 0) \mid S_1)\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big(\mathbf{Y}(\mathbf{a_1}, \mathbf{a_2}) - \mathbf{Y}(\mathbf{a_1}, \mathbf{0}) \mid \bar{\mathbf{S}}_\mathbf{2}(\mathbf{a_1})\big)\Big\} \\
&\quad = \mu_0 + \epsilon_1(s_1) + \mu_\mathbf{1}(\mathbf{s_1}, \mathbf{a_1}) + \epsilon_2(\bar{s}_2, a_1) + \mu_\mathbf{2}(\bar{\mathbf{s}}_\mathbf{2}, \bar{\mathbf{a}}_\mathbf{2}) \\
&\quad \stackrel{e.g.}{=} \mu_0 + \epsilon_1(s_1) + \beta_\mathbf{10}\mathbf{a_1} + \beta_\mathbf{11}\mathbf{a_1 s_1} \\
&\qquad\qquad + \epsilon_2(\bar{s}_2, a_1) + \beta_\mathbf{20}\mathbf{a_2} + \beta_\mathbf{21}\mathbf{a_2 s_1} + \beta_\mathbf{22}\mathbf{a_2 s_2}
\end{aligned}
$$

Constraints on the Causal and Nuisance Portions

$$E\big(Y(a_1,a_2) \mid \bar{S}_2(a_1)=\bar{s}_2\big) = \mu_0 + \epsilon_1(s_1) + \boldsymbol{\mu_1(\mathbf{s_1},\mathbf{a_1})} + \epsilon_2(\bar{s}_2,a_1) + \boldsymbol{\mu_2(\bar{\mathbf{s}}_2,\bar{\mathbf{a}}_2)},$$ where

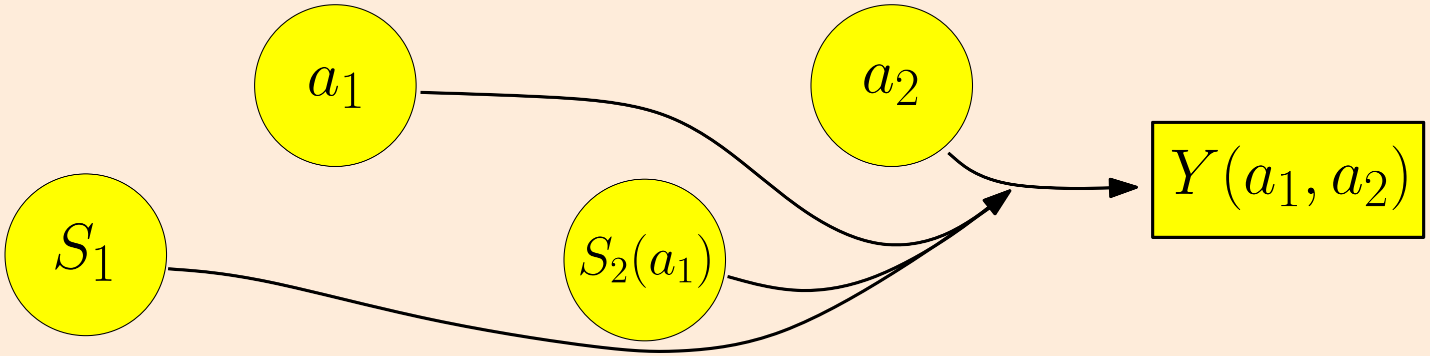
- $\mu_2(\bar{s}_2,a_2,0)=0$ and $\mu_1(s_1,0)=0$,
- $\epsilon_2(\bar{s}_2,a_1) = E(Y(a_1,0) \mid \bar{S}_2(a_1)=\bar{s}_2) - E(Y(a_1,0) \mid S_1 = s_1)$,
- $\epsilon_1(s_1) = E(Y(0,0) \mid S_1=s_1) - E(Y(0,0))$,
- $E_{S_2|S_1}(\epsilon_2(\bar{s}_2,a_1) \mid S_1=s_1)=0$, and $E_{S_1}(\epsilon_1(s_1))=0$.

The ϵ_t's make the SNMM a non-standard regression model.

Time-Varying Causal Effects of the SNMM

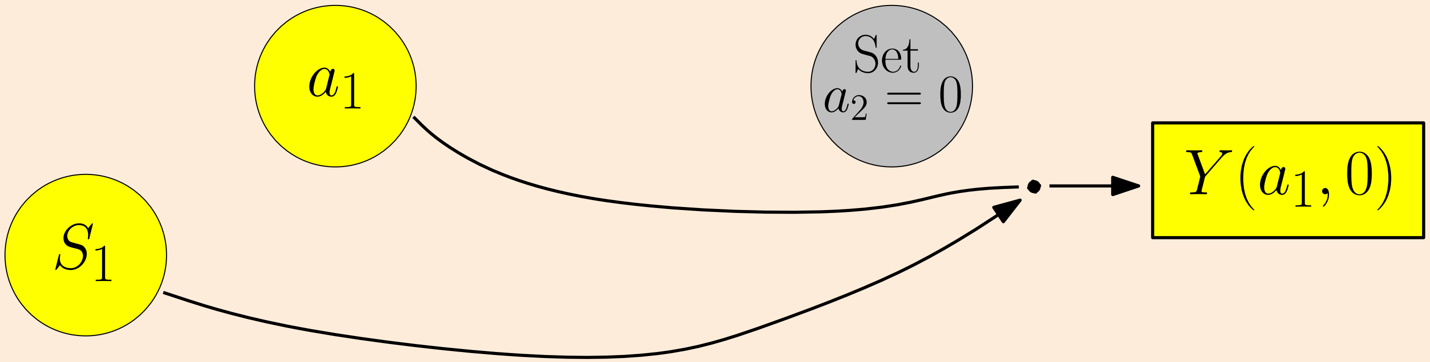
Conditional Intermediate Causal Effect at $t = 2$:

$$\mu_{\mathbf{2}}(\bar{\mathbf{s}}_{\mathbf{2}}, \bar{\mathbf{a}}_{\mathbf{2}}) = E[Y(a_1, a_2) - Y(a_1, 0) \mid S_1 = s_1, S_2(a_1) = s_2]$$



Conditional Intermediate Causal Effect at $t = 1$:

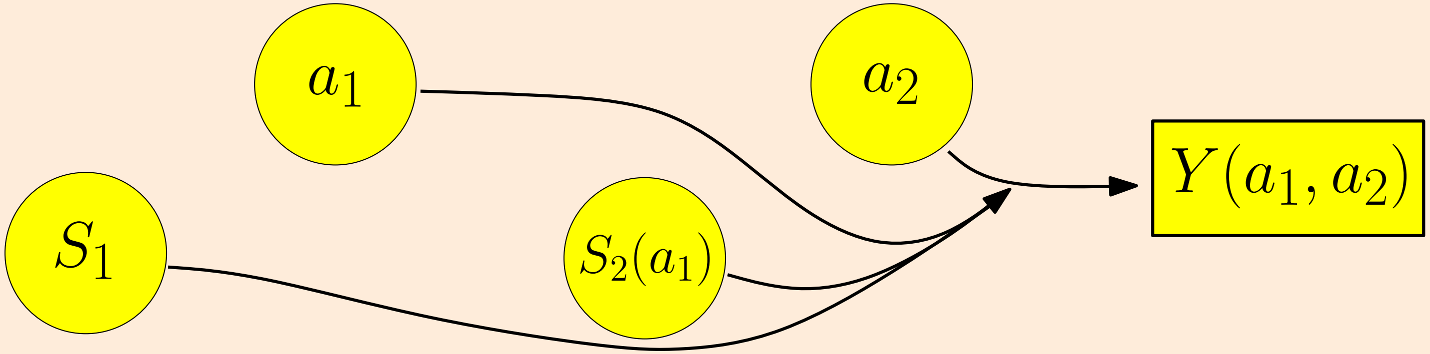
$$\mu_{\mathbf{1}}(\mathbf{s}_{\mathbf{1}}, \mathbf{a}_{\mathbf{1}}) = E[Y(a_1, 0) - Y(0, 0) \mid S_1 = s_1]$$



Time-Varying Causal Effects of the SNMM

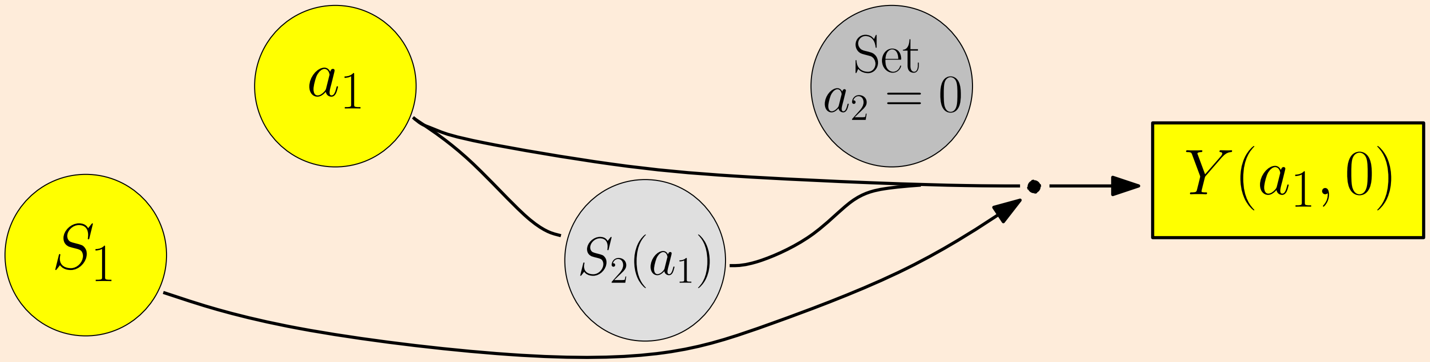
Conditional Intermediate Causal Effect at $t = 2$:

$$\mu_{\mathbf{2}}(\bar{\mathbf{s}}_{\mathbf{2}}, \bar{\mathbf{a}}_{\mathbf{2}}) = E[Y(a_1, a_2) - Y(a_1, 0) \mid S_1 = s_1, S_2(a_1) = s_2]$$



Conditional Intermediate Causal Effect at $t = 1$:

$$\mu_{\mathbf{1}}(\mathbf{s_1}, \mathbf{a_1}) = E[Y(a_1, 0) - Y(0, 0) \mid S_1 = s_1]$$



7 Estimation (in the Ole Days)

The MSM and the SNMM have helped us come a long way in clarifying and being explicit about the causal estimands of interest.

But what about estimation?

The **traditional regression estimator** is where we fit a regression of Y on S_1, A_1, S_2, A_2.

Does the traditional regression estimator work? Why or why not?

The Traditional Regression Estimator

To answer questions identified under the MSM (Type A) the scientist may be inclined to fit a regression model such as:

$$E(Y \mid \bar{S}_2 = \bar{s}_2, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_1 + \beta_{\mathbf{1}}^* \mathbf{a_1} + \beta_{\mathbf{2}}^* \mathbf{a_1 s_1} + \beta_{\mathbf{3}}^* \mathbf{a_2} + \beta_{\mathbf{4}}^* \mathbf{a_2 s_1}$$

But recognizing that intermediate response S_2 may be a **confounder** of the impact of future treatment (A_2) on outcomes (Y), the scientist may modify this model and fit:

$$E(Y \mid \bar{S}_2 = \bar{s}_2, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_1 + \beta_{\mathbf{1}}^* \mathbf{a_1} + \beta_{\mathbf{2}}^* \mathbf{a_1 s_1} + \beta_{\mathbf{3}}^* \mathbf{a_2} + \beta_{\mathbf{4}}^* \mathbf{a_2 s_1} + \eta_2 s_2$$

The Traditional Regression Estimator

Or to answer questions identified under the SNMM (Type B Questions) the scientist may be inclined to fit a regression model such as:

$$E(Y \mid \bar{S}_2 = \bar{s}_2, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_1 + \boldsymbol{\beta_1^* a_1} + \boldsymbol{\beta_2^* a_1 s_1} \\ + \eta_2 s_2 + \boldsymbol{\beta_3^* a_2} + \boldsymbol{\beta_4^* a_2 s_1} + \boldsymbol{\beta_5^* a_2 s_2}$$

In this regression, we are inclined to adjust for S_2 because it is a putative **time-varying moderator** of interest, whereas in the previous model we adjusted for it because it was a **time-varying confounder** of interest. (S_2 may be both, in fact.)

So what's wrong with the Traditional Estimator?

In any case (whether we are interested in Type A or Type B questions), we are conditioning on S_2 in the regression models and there are potential problems with doing this.

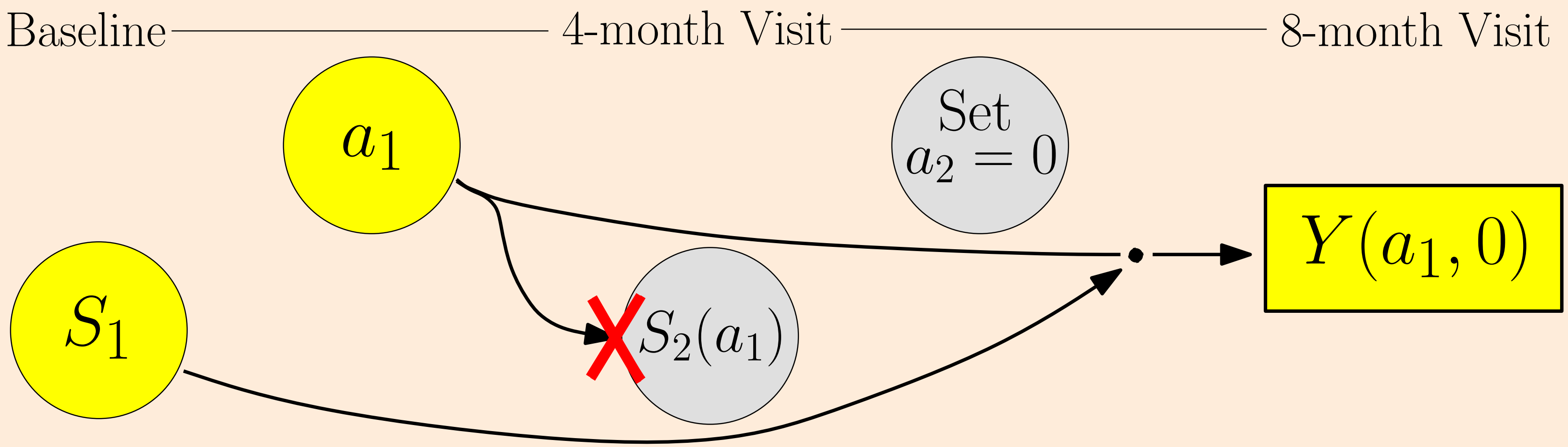
Conditioning on S_2 naively may result in bias in the estimates of the parameters of either the MSM or SNMM.

In the following slides, we offer some intuition as to why by describing at least two problems the empirical scientist may encounter?

Interestingly, these problems may occur even in the absence of time-varying confounders.

First problem with the Traditional Approach

Wrong Effect

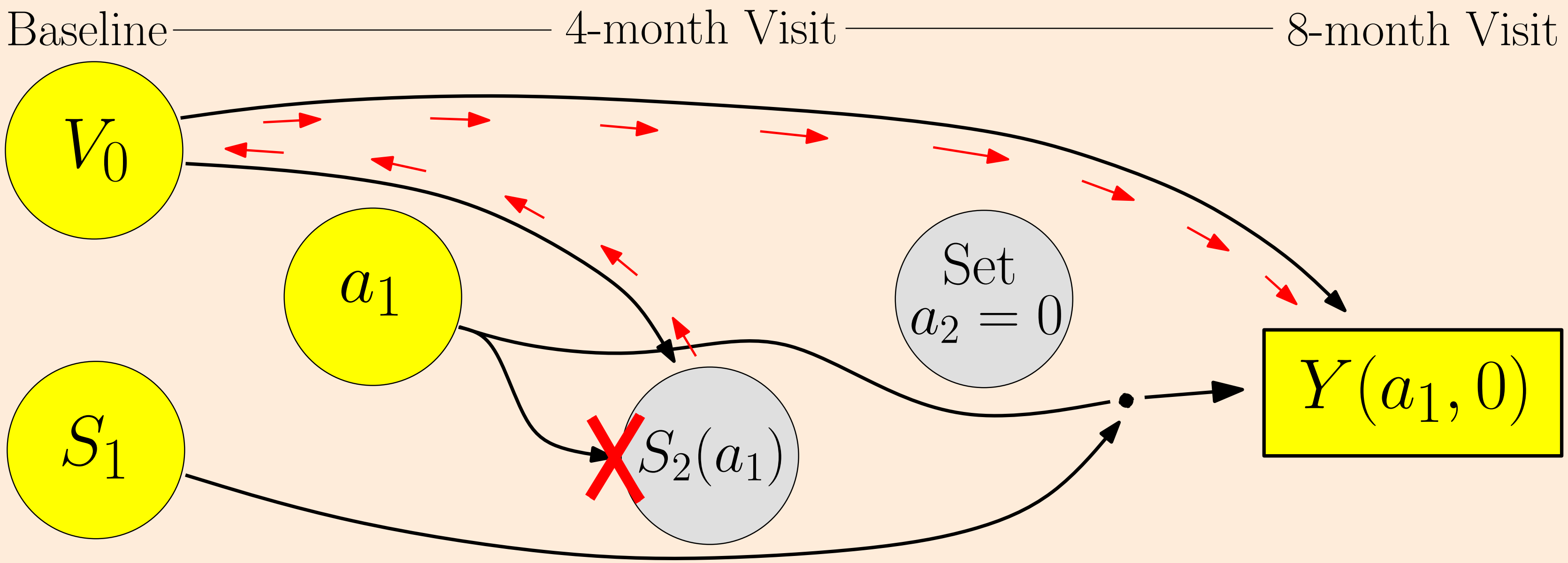


But what about the effect transmitted through $S_2(a_1)$?

The term $\beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_1}$ does not capture the "total" impact of $(a_1, 0)$ vs $(0, 0)$ on $Y(a_1, a_2)$ given values of S_1.

Second problem with the Traditional Approach

Spurious Effect



This is also known as "Berkson's paradox"; and is related to Judea Pearl's back-door criterion.

Estimation (Nowadays)

Separate estimators now exist for the MSM and SNMM that get around these two problems.

In the case of the MSM:
Inverse-probability-of-treatment-weighting allows empirical scientists to estimate the effects (including baseline causal effect moderation) of time-varying treatments without having to condition on intermediate outcomes like S_2 in the analysis model.

Estimation (Nowadays)

Separate estimators now exist for the MSM and SNMM that get around these two problems.

In the case of the SNMM: Estimators such as the G-Estimator and 2-Stage Regression Estimator exist that allow us to condition on the S_2 in a principled way such that we can assess time-varying effect moderation without the two problems described previously.

** Proposed 2-Stage Regression Estimator **

The proposed 2-Stage Estimator for the SNMM resembles the Traditional Estimator. Instead of using the Traditional Estimator

$$E(Y \mid \bar{S}_2 = \bar{s}_2, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_1 + \boldsymbol{\beta_1^* a_1} + \boldsymbol{\beta_2^* a_1 s_1} + \boldsymbol{\eta_2 s_2} + \boldsymbol{\beta_3^* a_2} + \boldsymbol{\beta_4^* a_2 s_1} + \boldsymbol{\beta_5^* a_2 s_2},$$

we use the following

$$E(Y \mid \bar{S}_2 = \bar{s}_2, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_1 + \boldsymbol{\beta_1^* a_1} + \boldsymbol{\beta_2^* a_1 s_1} + \boldsymbol{\eta_2 \big(s_2 - E(S_2 \mid A_1, S_1)\big)} + \boldsymbol{\beta_3^* a_2} + \boldsymbol{\beta_4^* a_2 s_1} + \boldsymbol{\beta_5^* a_2 s_2}.$$

We call it "2-Stage" because first we estimate $E(S_2 \mid A_1, S_1)$, then use the residual $s_2 - \widehat{E(S_2 \mid A_1, S_1)}$ in a second regression to get β's. Use sandwich/robust SEs for inference (p-vals, CIs, etc.).

8 Conclusions

1. Be explicit about the **causal effect moderation question** of interest and the appropriate model for the question:
   - Interested in effect moderation by baseline covariates? If so, then the MSM is appropriate.
   - Interested in effect moderation by time-varying covariates? If so, then the SNMM is appropriate.
2. Apply appropriate estimator matching the model of interest. Caution using the traditional regression estimator naively.
3. Plan to collect all possible time-varying confounders of treatment (or exposure) status.

9 References

Marginal Structural Models:

Robins JM, Hernàn MA, and Brumback B. **Marginal structural models and causal inference in epidemiology.** Epidemiology, 11(5):55060, 2000.

Robins JM. **Association, causation, and marginal structural models.** Synthese, 121(1):151 179, 1999.

Hernàn MA, Brumback B, Robins JM. **Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men.** Epidemiology, 11(5):561-70, 2000.

Hernàn MA, Brumback B, Robins JM. **Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments.** Journal of the American Statistical Association, 96(454), 440-448, 2001

Bray BC, Almirall D, Zimmerman RS, Lynam D, and Murphy SA. **Assessing the total effect of time-varying predictors in prevention research.** Prevention Science, 7(1):117, March 2006.

Structural Nested Mean Models:

Robins JM. **Correcting for non-compliance in randomized trials using structural nested mean models.** Communications in Statistics, Theory and Methods, 23(8):23792412, 1994.

Almirall D, Ten Have T, and Murphy SA. **Structural nested mean models for assessing time-varying effect moderation.** Biometrics, 2009 (in press, 2009).

Almirall D, Coffman CJ, Yancy Jr WS, and Murphy SA. Analysis of Observational Health- Care Data Using SAS, **Maximum Likelihood Estimation of the Structural Nested Mean Model Using SAS PROC NLP**. SAS Press (in press, 2009).

Thank you!
More Questions?