Examining moderated effects of additional adolescent substance use treatment: Structural nested mean model estimation using inverse-weighted regression-with-residuals

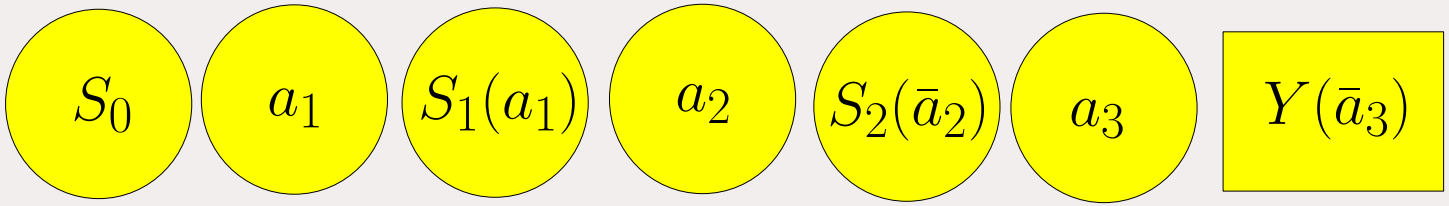
Daniel Almirall[1], Daniel F. McCaffrey[2], Beth Ann Griffin[2], Rajeev Ramchand[2], Susan A. Murphy[1]

[1]Univ of Michigan, Institute for Social Research
[2]RAND, Statistics

Institute of Mathematical Statistics
Asia Pacific Rim Meeting — July 2, 2012

1 Time-Varying Setting

The data structure in the time-varying setting is:



Motivating Example: Adolescents & Substance Use Treatment

S_0	Age, severity @ intake, contr. env in p.90
a_1	0-3mo treatment; binary, $a_1 =$ yes/no
$S_1(a_1)$	Severity @ 0-3mo
a_2	3-6mo treatment; binary, $a_2 =$ yes/no
$S_2(a_1, a_2)$	Severity @ 3-6mo
a_3	6-9mo treatment; binary, $a_3 =$ yes/no
$Y(a_1, a_2, a_3)$	Substance use frequency 9-12mo

2 What Scientific Question of Interest?

The data structure: $\{S_0, a_1, S_1(a_1), a_2, S_2(a_1, a_2), a_3, Y(a_1, a_2)\}$.

We began wondering about: Cumulative effect of treatment?

Observed treatment sequences in data are: (A_1, A_2, A_3), Rate

(0,0,0),	11%	(0,0,1),	2%
(1,0,0),	41%	(0,1,1),	2%
(1,1,0),	19%	(1,0,1),	5%
(1,1,1),	17%	(0,1,0),	2%

More specific questions emerged: What are the incremental effects of additional substance use treatment? Are these effects heterogeneous? i.e., Do they differ as a function of severity at intake and improvements over time?

3 Time-Varying Effect Moderation

The data structure: $\{S_0, a_1, S_1(a_1), a_2, S_2(a_1, a_2), a_3, Y(a_1, a_2)\}$.

Overarching question: What are the incremental effects of additional substance use treatment, as a function of severity at intake and improvements over time?

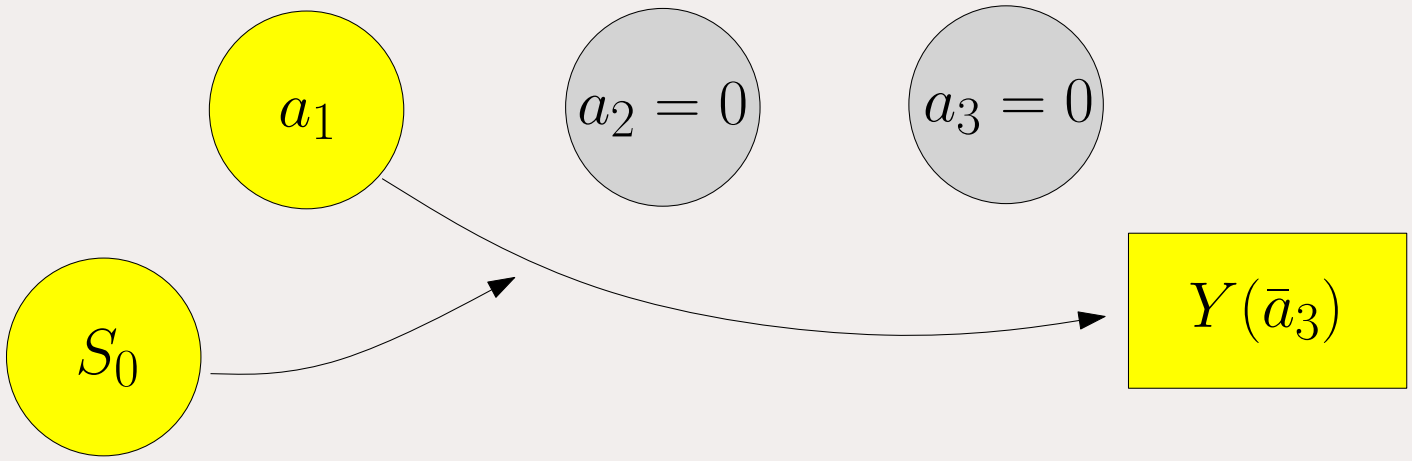
More specifically, there are 3 types of causal effects of interest:

1. **Distal moderated effect of initial treatment:** What are the effects of (1,0,0) vs (0,0,0) on Y given S_0?
2. **Medial moderated effect of cumulative treatment:** What are the effects of (1,1,0) vs (1,0,0) on Y given (S_0, S_1)?
3. **Proximal moderated effect of cumulative treatment:** What are effects of (1,1,1) vs (1,1,0) on Y given (S_0, S_1, S_2)?

What are the distal moderated effects of initial treatment?

What are the effects of (1,0,0) vs (0,0,0) on Y given S_0?

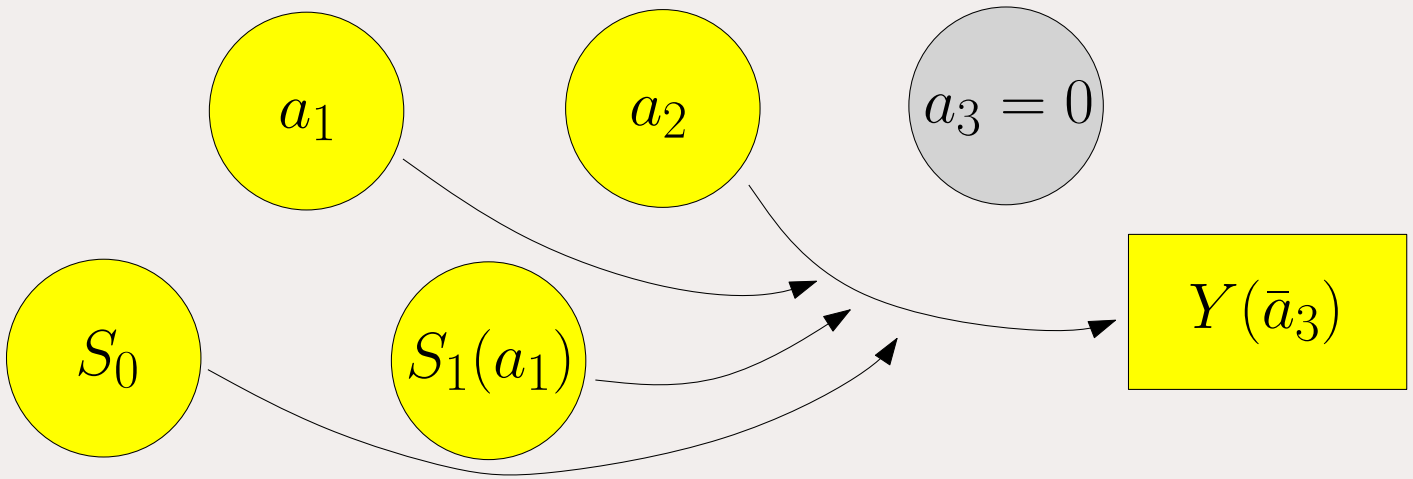
$$\mu_{\mathbf{1}} = E[Y(1,0,0) - Y(0,0,0) \mid S_0 = s_0]$$



What are the medial moderated effects of cumulative initial treatment?

What are the effects of (1,1,0) vs (1,0,0) on Y given (S_0, S_1)?

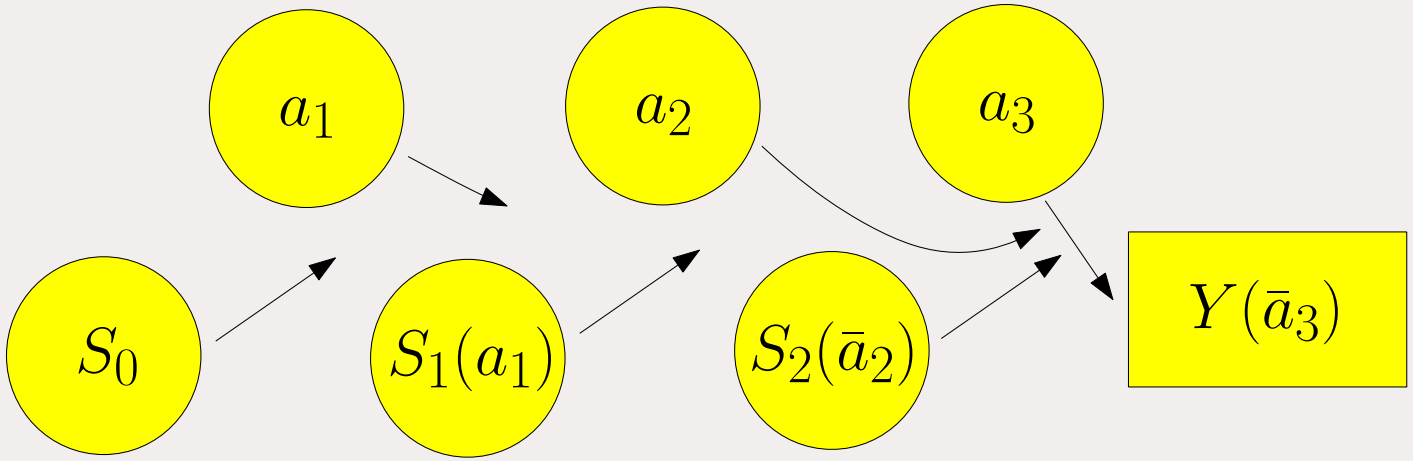
$$\mu_{\mathbf{2}} = E[Y(1,1,0) - Y(1,0,0) \mid S_0 = s_0, S_1(1) = s_1]$$



What are the proximal moderated effects of cumulative initial treatment?

What are the effects of (1,1,1) vs (1,1,0) on Y given (S_0, S_1, S_2)?

$$\mu_{\mathbf{3}} = E[Y(1,1,1) - Y(1,1,0) \mid \bar{S}_2(1,1) = \bar{s}_2]$$



4 Robins' Structural Nested Mean Model

decomposes $E(Y \mid \cdot)$ into nuisance and causal parts:

$$
\begin{aligned}
&E\big[Y(a_1, a_2) \mid S_0, S_1(a_1)\big] \\
&\quad = E[Y(0,0)] + \Big\{E[Y(0,0) \mid S_0] - E[Y(0,0)]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y(a_1, 0) - Y(0,0) \mid S_0}\big]\Big\} \\
&\qquad\qquad + \Big\{E[Y(a_1, 0) \mid \bar{S}_1(a_1)] - E[Y(a_1, 0) \mid S_0]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y(a_1, a_2) - Y(a_1, 0) \mid \bar{S}_1(a_1)}\big]\Big\} \\
&\quad = \mu_0 + \epsilon_1(s_0) + \boldsymbol{\mu_1(s_0, a_1)} + \epsilon_2(\bar{s}_1, a_1) + \boldsymbol{\mu_2(\bar{s}_1, \bar{a}_2)}
\end{aligned}
$$

Constraint: $\mu_t = 0$ when $a_t = 0$
Constraint: $E_{S_1|S_0}[\epsilon_2(\bar{s}_1, a_1) \mid S_0 = s_0] = 0$, and $E_{S_0}[\epsilon_1(s_0)] = 0$

5 Problems with Traditional Regression

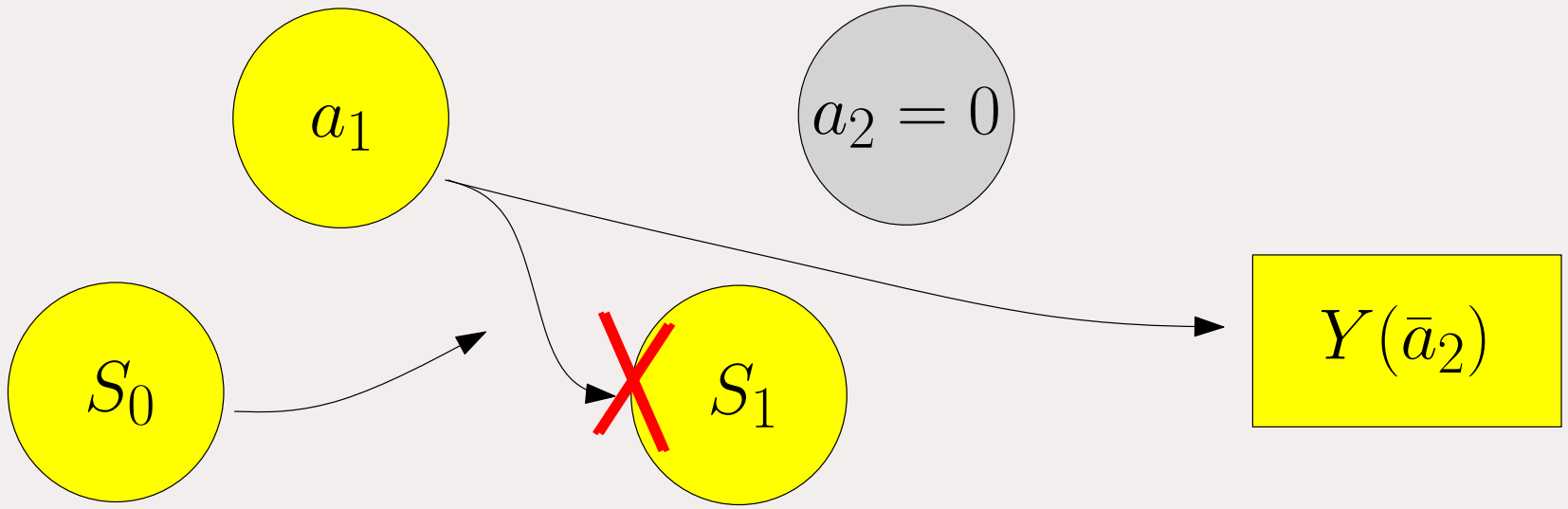
Ex: Use the **Traditional Estimator** to model the $t = 2$ SNMM as:

$$E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_0 + \beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_0} + \eta_2 s_1 + \beta_3^* \mathbf{a_2} + \beta_4^* \mathbf{a_2 s_0} + \beta_5^* \mathbf{a_2 s_1}$$

- Two problems arise from the way we condition on S_t: (1)WRONG EFFECT, (2)SPURIOUS BIAS
- One problem arises from not adjusting for time-varying confounders: (3)TIME-VARYING CONFOUNDING BIAS

First problem with the Traditional Approach

Wrong Effect

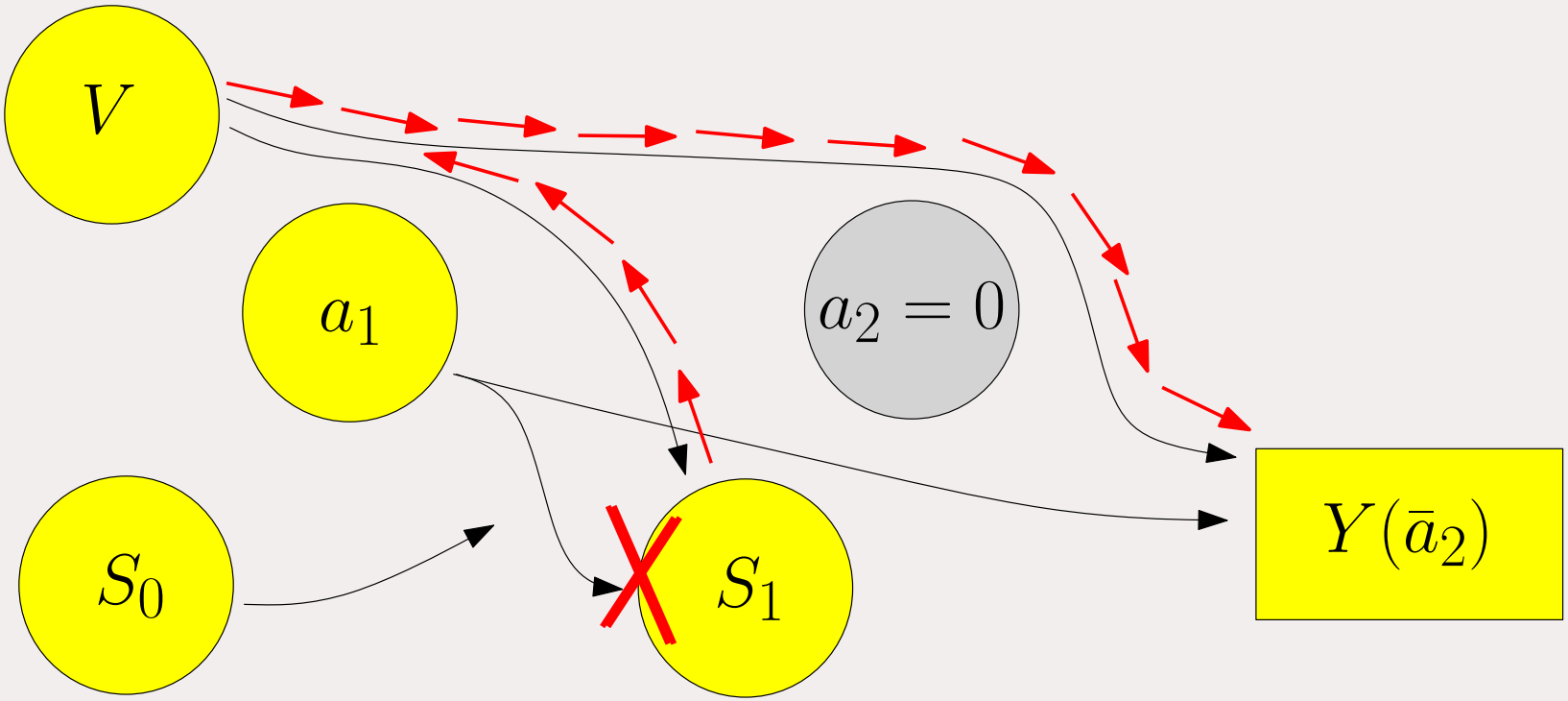


But what about the effect transmitted through $S_1(a_1)$?

So the end result is the term $\beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_0}$ does not capture the "total" impact of $(a_1, 0)$ vs $(0, 0)$ on Y given values of S_0.

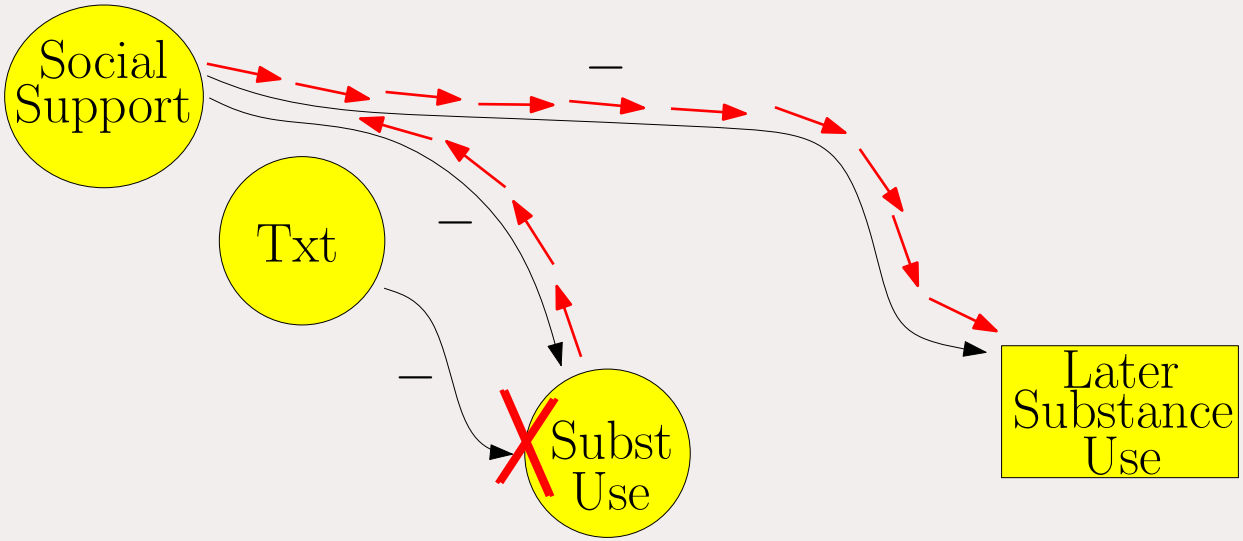
Second problem with the Traditional Approach

Spurious Bias



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

Intuition about the Spurious Bias



Imagine adolescent who is a high user despite getting treated:

Q: What does this tell you in terms of his social support?

A: There must be poor social support.

Implication: Conditional on substance use, getting treated is associated with more substance use! Bias is $-1(-)(-)(-) = +$.

Proposed Regression with Residuals Estimator

Instead of the traditional regression estimator

$$E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_0 + \beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_0} + \eta_2 \mathbf{s_1} + \beta_3^* \mathbf{a_2} + \beta_4^* \mathbf{a_2 s_0} + \beta_5^* \mathbf{a_2 s_1},$$

we use the following

$$E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_0 + \beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_0} + \eta_2 \big(\mathbf{s_1} - \mathbf{E(S_1 \mid A_1, S_0)}\big) + \beta_3^* \mathbf{a_2} + \beta_4^* \mathbf{a_2 s_0} + \beta_5^* \mathbf{a_2 s_1}.$$

We call it "regression with residuals" because first we estimate $E(S_1 \mid A_1, S_0)$, then use the residual $s_1 - \widehat{E(S_1 \mid A_1}, S_0)$ in a second regression to get β's.

Proposed Regression with Residuals Estimator

$$E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_0 + \beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_0}$$
$$+ \eta_2 \big(\mathbf{s_1} - \mathbf{E(S_1 \mid A_1, S_0)}\big) + \beta_3^* \mathbf{a_2} + \beta_4^* \mathbf{a_2 s_0} + \beta_5^* \mathbf{a_2 s_1}.$$

The proposed estimator is unbiased for the μ_t's provided:

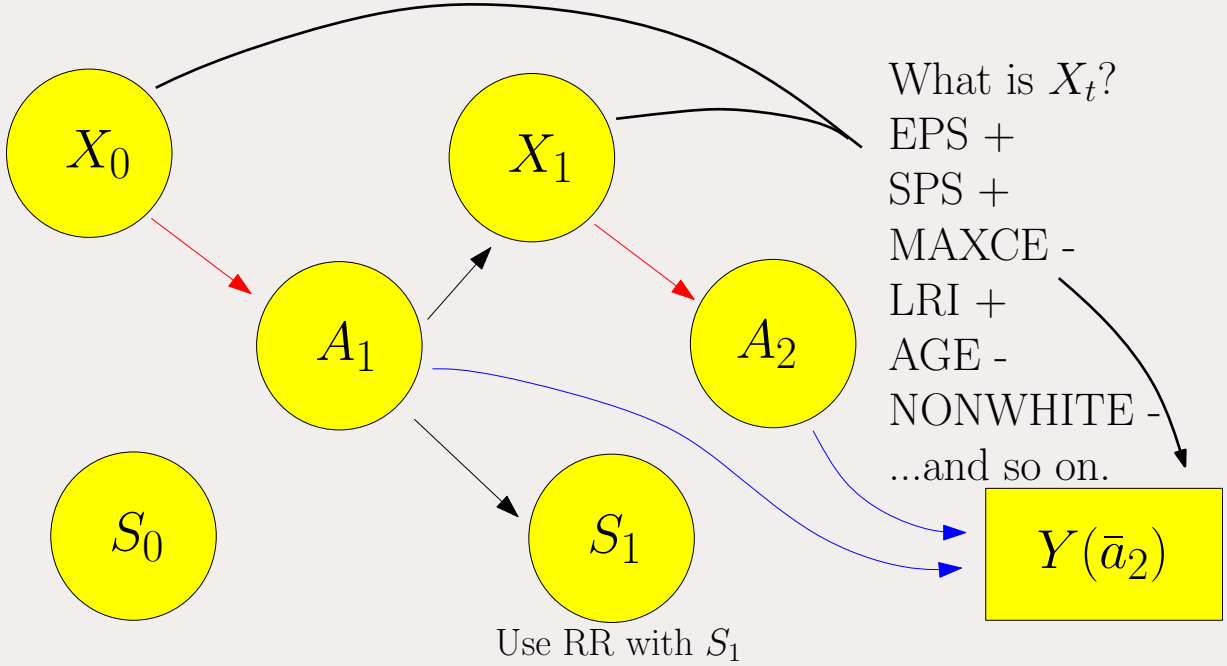
1. Correctly modeled SNMM, incl. the ϵ_t's functions.
2. $A_1 \perp \{Y(a_1, a_2)\} \mid S_0$, and
3. $A_2 \perp \{Y(a_1, a_2)\} \mid S_0, A_1, S_1$

Together, 2. and 3. is a Sequential Ignorability Assumption.

But there may be other measured time-varying confounders...

Third Problem with Traditional Approach

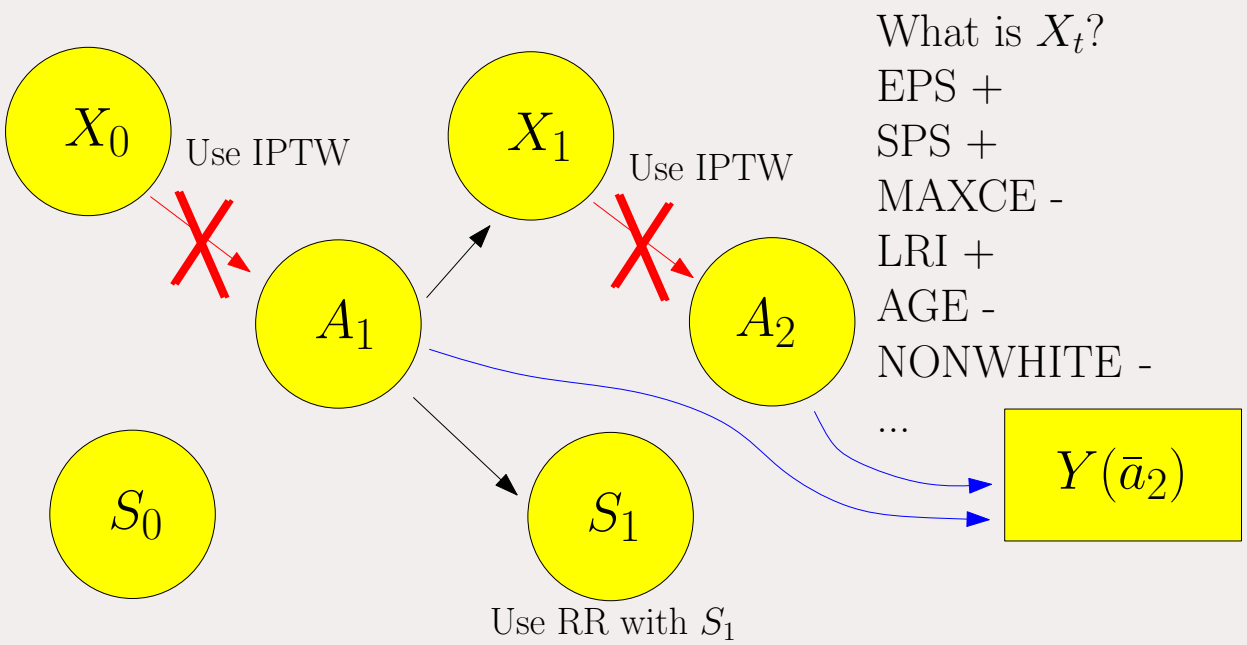
Time-varying Confounding Bias: Time-varying covariates X_t that are confounders, but not moderators of interest?



The auxiliary variables X_t may be high-dimensional.

Solution: Inverse-Probability-of-Treatment Weights

We use IPTW version of the proposed 2-Stage RR Estimator:



The proposed IPTW estimator is unbiased provided (1) correct SNMM, (2) sequential ignorability given $(\bar{S}_t, \bar{X}_t)$, (3) consistency, and (4) get the "right" weights.

The Form of the IPT Weights

$$W_1 = \frac{1}{Pr(A_1 = a_1 \mid S_0 = s_0, X_0 = x_0)}$$

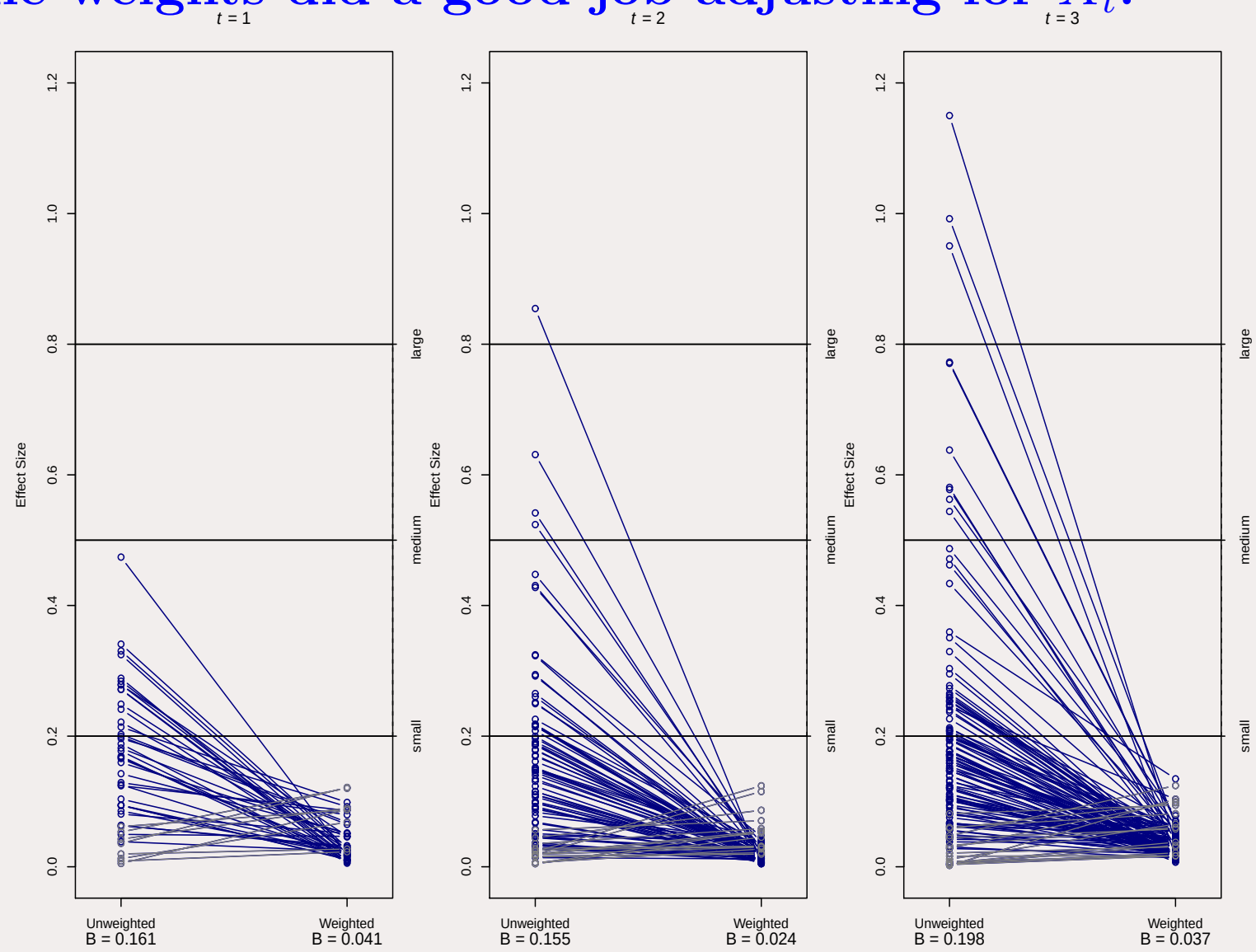
$$W_2 = \frac{1}{Pr(A_2 = a_2 \mid S_0 = s_0, X_0 = x_0, A_1 = a_1, S_1 = s_1, X_1 = x_1)}$$

- Assumes denominator probabilities are non-zero.
- We use logistic regressions to estimate the denominator probs.; models chosen to result in "best" balance.
- $W_1 \times W_2$ is used in the IPTW+RR estimator of the SNMM.
- Following Murphy, van der Laan, Robins (unpublished), we use a stabilized version where the numerator for W_t is $Pr(A_t = a_t \mid \bar{A}_{t-1}, \bar{S}_{t-1})$.

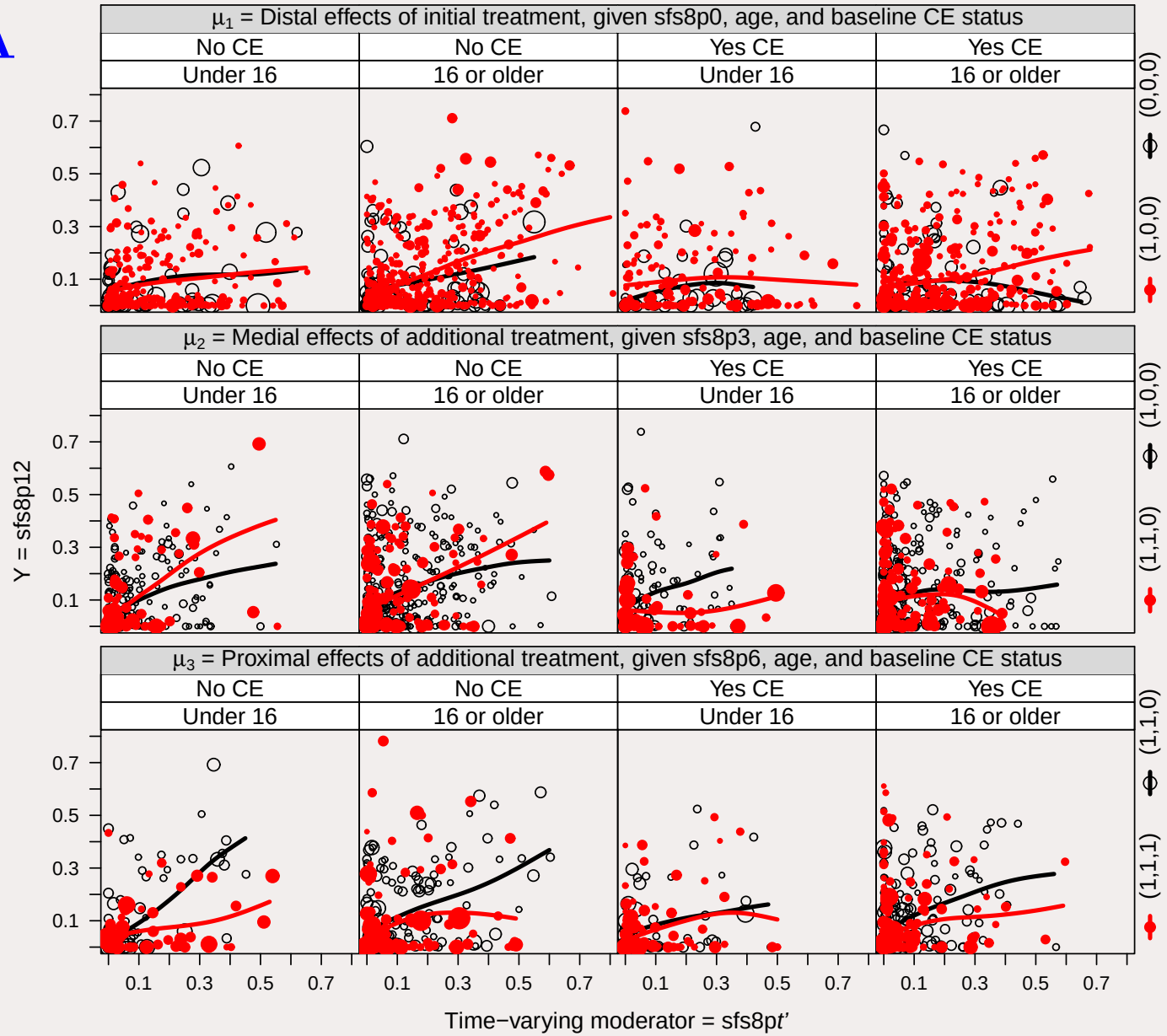
6 Data Analysis

- From US substance abuse prgms (CSAT $\subset$ SAMHSA)
- GAIN: structured clinical interview; over 100 scales/indices
- $n = 2870$ adolescents; data every 3 months for 1 year
- $\{(S_0, X_0), A_1, (S_1, X_1), A_2, (S_2, X_2), A_3, Y\}$
- $S_0 =$ hx controlled environment, age
- $S_t =$ substance frequency scale at intake, 0-3, 3-6
- $\mathbf{X}_t =$ measured time-varying confounders at intake, 0-3, 3-6
- $A_t =$ none (0) vs some txt (1=outpt, inpt, or both)
- $Y =$ substance frequency scale at 9-12mo

The weights did a good job adjusting for X_t.



EDA


μ1 = Distal effects of initial treatment, given sfs8p0, age, and baseline CE status
μ2 = Medial effects of additional treatment, given sfs8p3, age, and baseline CE status
μ3 = Proximal effects of additional treatment, given sfs8p6, age, and baseline CE status
No CE
Yes CE
Under 16
16 or older
(0,0,0)
(1,0,0)
(1,1,0)
(1,1,1)
Y = sfs8p12
Time−varying moderator = sfs8pt'
0.1
0.3
0.5
0.7

Effect Estimates from SNMM, using RR+IPTW

Contrast	Subgroup	Est.	Eff.Sz.	P-val
μ_1: **Distal**				
$(1,0,0)$ vs $(0,0,0)$	no intake sevrty, < 16yrs	-0.004	-0.03	0.74
$(1,0,0)$ vs $(0,0,0)$	hi intake sevrty, ≥ 16yrs	0.033	0.25	0.08
μ_2: **Medial**				
$(1,1,0)$ vs $(1,0,0)$	no 0-3 severity	-0.008	-0.06	0.42
$(1,1,0)$ vs $(1,0,0)$	hi 0-3 severity, yes ce	-0.048	-0.36	0.21
$(1,1,0)$ vs $(1,0,0)$	hi 0-3 severity, no ce	0.021	0.16	0.66
μ_3: **Proximal**				
$(1,1,1)$ vs $(1,1,0)$	no 6-9 severity	-0.006	-0.04	0.59
$(1,1,1)$ vs $(1,1,0)$	hi 6-9 severity	-0.168	-1.27	< 0.01
$(.,.,1)$ vs $(.,.,0)$	no 6-9 severity	0.026	0.19	0.12
$(.,.,1)$ vs $(.,.,0)$	hi 6-9 severity	-0.165	-1.24	< 0.01

Some conjectures about the substantive story

- Initial treatment alone may be iatrogenic for older kids with high severity at intake (evidence is not so strong here).
- An additional 3mos of treatment may be more helpful for kids still severe at the end of 3 months who have a hx of a controlled environment (evidence is very weak here).
- Providing full treatment is especially beneficial for the kids who are still looking bad after 6 months (evidence here is reasonably strong).
- There is not a lot of evidence for a treatment effect for kids who are not severe.

7 Summary

- **Time-varying causal effect moderation**: "What is the incremental effect of additional community-based substance use treatment, as a function of severity at intake and improvements over time?"
- Examine using Robins' Structural Nested Mean Model
- Propose wtd regression with residuals estimator for SNMM
  - Resembles traditional regression estimator; easy-to-use
  - Adjust time-varying confounders via IPT Weighting
- Standard errors: In simulation experiments, we find bootstrap SEs to be better than ASEs in small samples

Acknowledgements

NIDA Funding:

The Methodology Center at Penn State University (P50-DA-010075; PIs: Collins, Murphy & Co-I: Almirall)

RAND (R01-DA-015697; PIs: McCaffrey, Griffin & Co-I: Ramchand)

NIMH Funding:

Univ of Michigan (R01-MH-080015; PI: Murphy)

Special Help From:

Mary Ellen Slaughter, RAND

Bobby Yuen, Graduate Student, Michigan Statistics

Thank you.

Contact Information:
Daniel Almirall, Univ of Michigan, ISR
mailto:dalmiral@umich.edu, 734 936 3077

* Robins (1994), *Communications in Statistics*.

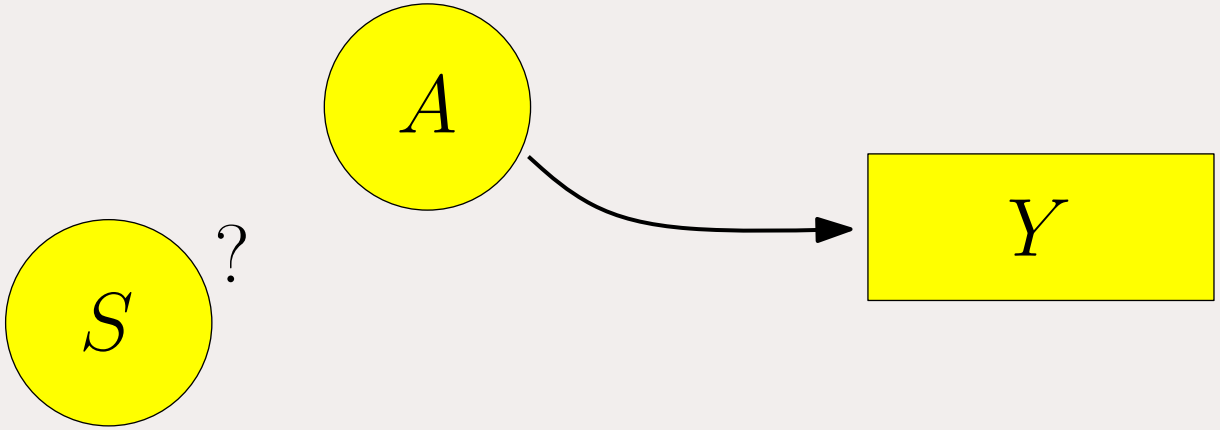
* Almirall, Tenhave, Murphy (2010), *Biometrics*.

* Almirall, McCaffrey, Ramchand, Murphy (2011), *Prevention Science*. Software to fit SNMM using RR on my website.

* Almirall, McCaffrey, Griffin, Ramchand, Murphy (to submit).

8 Back Pocket Slides

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



Examples

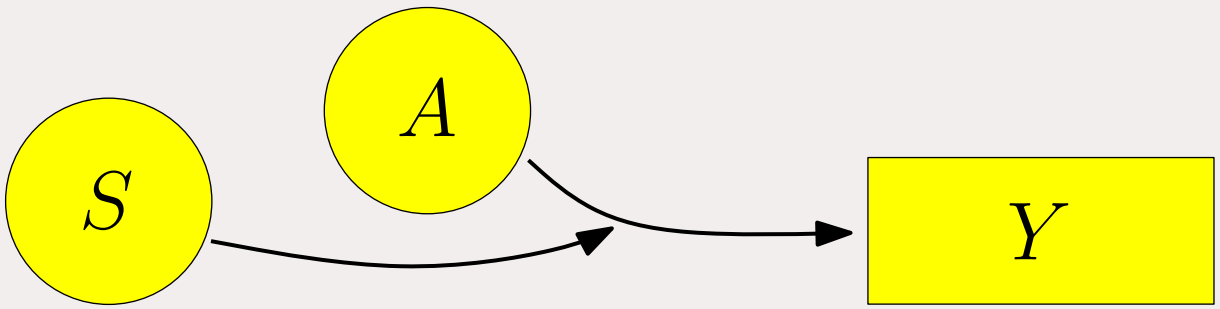
S = pre-A covt	A = txt/expsr	Y = outcome
Social Support	Inpatient vs. Outpatient	Substance Abuse

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.

Final Warm-up: Mean Model in One Time Point

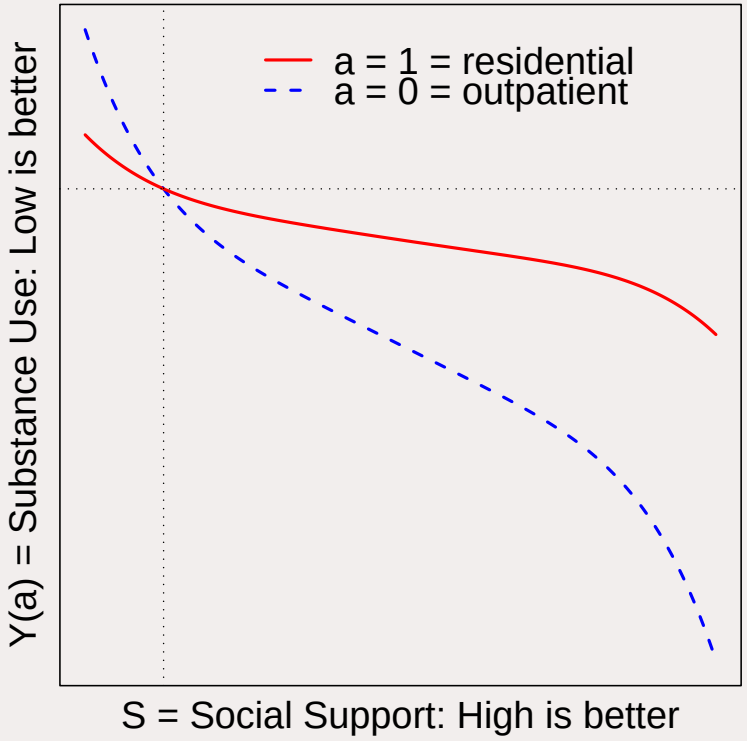
Decomposition of the conditional mean $E(Y(a) \mid S)$ and the prototypical linear model:

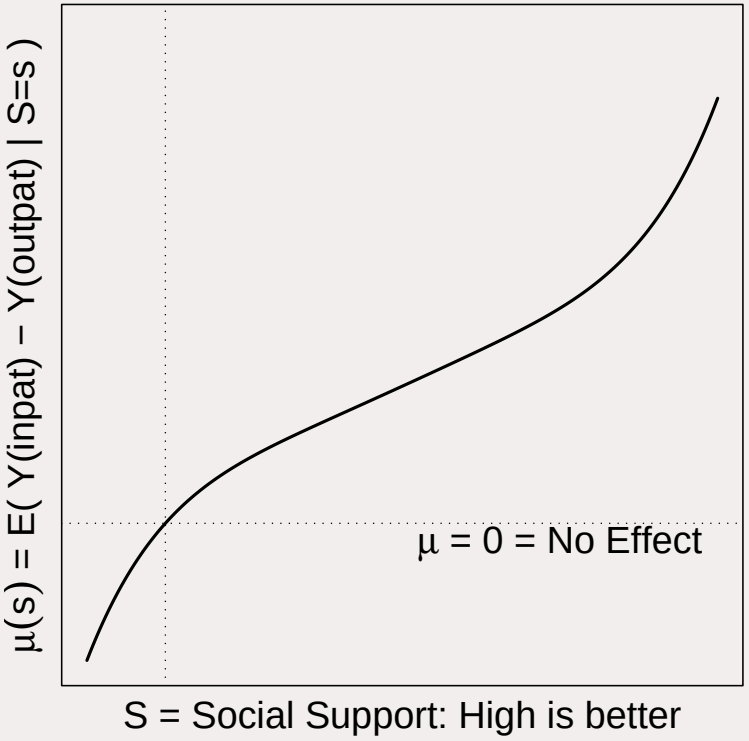
$$\begin{aligned} E(Y(a) \mid S = s) &= E(Y(0)) \\ &\quad + \Big(E(Y(0) \mid S = s) - E(Y(0))\Big) \\ &\quad + E(Y(a) - Y(0) \mid S = s) \\ &= \eta_0 + \epsilon(s) + \mu(s, a) \\ &\overset{e.g.}{=} \eta_0 + \eta_1(s - E(S)) + \beta_1 a + \beta_2 as. \end{aligned}$$

Boils down to what we always do anyway: that is, treatment $\times$ covariate interaction terms to examine effect heterogeneity.

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 treatment or exposures effects enhances our understanding of various (competing) scientific theories; and it may suggest new scientific hypotheses to be tested.

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.

Prototypical Linear Parametric Model

We use β for our causal parameters of interest:

$$\begin{aligned} E(Y(a) \mid S) &= \eta_0 + \phi(S) + \mu(S, a; \beta) \\ &= \eta_0 + \phi(S) + aH\beta \end{aligned}$$

where H is a function of S.

Sometimes we parameterize $\phi(S)$ using $\phi(S; \eta_{-0}) = G\eta_{-0}$, where G is a function of S.

Example: Let $G = (S)$ and $H = (1, S)$:

$$E(Y(a) \mid S = s) = \eta_0 + \eta_1 s + a \times (\beta_1 + \beta_2 s).$$

If a and S are binary, then this is the fully saturated model.

Estimation in One Time Point

Consider three estimators for β in $\mu(S, a; \beta)$:

1. Traditional Regression
2. Semi-parametric Estimation Method: Robins' E-Estimator
3. Inverse Probability of Treatment Weighted (IPTW) Regression

We discuss these (and more) in turn, supposing that

1. a is binary (0,1), and
2. True model for $\mu(s, a)$ is $\mu(S, a; \beta) = aH\beta$ for some H.
   Example: $H = (1, S) \Rightarrow aH\beta = a(\beta_1 + \beta_2 s)$.

An important consideration in estimation is how A comes about.

Traditional Ordinary Least Squares Regression

Recall true model: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$.

Useful when S is sole confounder, and have good model for $\phi(s)$.

Requires model for nuisance function: $\phi(S; \eta_{-0}) = G\eta_{-0}$.

Regress $Y \sim [1, G, A \times H]$ to get $(\widehat{\eta}, \widehat{\beta})$. The $\widehat{\beta}$ estimates solve

$$0 = \mathbb{P}_n \left(\left(Y - \eta_0 - G\eta_{-0} - AH\beta \right) AH^T \right).$$

$\widehat{\beta}$ unbiased for β if $\phi(S; \eta_{-0}) = G\eta_{-0}$ is true model for $\phi(s)$ and $A \perp \{Y(0), Y(1)\}$ given S.

Semi-parametric E-Estimator

Recall true model: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$.

Useful when S is sole confounder, but we have no model for $\phi(s)$.

Does NOT require model for nuisance function $\phi(s)$.

Get $\widehat{\beta}$ by solving the following estimating equations

$$0 = \mathbb{P}_n\left(\Big(Y - \widehat{b}(S;\xi) - AH\beta\Big)\Big(A - \widehat{p}(S;\alpha)\Big)H^T\right),$$

where $\widehat{b}(S;\xi)$ is a guess for $E(Y - AH\beta \mid S) = \eta_0 + \phi(S)$.

$\widehat{\beta}$ unbiased for β if $p(S;\alpha)$ is true model for $Pr(A = 1 \mid S)$, and $A \perp \{Y(0), Y(1)\}$ given S. (Discuss double-robustness.)

IPT Weighted Regression (WLS)

Recall true model: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$.

Useful when we have measured confounders V ($\supset S$).

Requires model for nuisance function: $\phi(S; \eta_{-0}) = G\eta_{-0}$.

Regress $Y \overset{\widehat{w}}{\sim} [1, G, A \times H]$ to get $(\widehat{\eta}, \widehat{\beta})$, where weights are

$$w(V, A) = A \times \frac{Pr(A = 1 \mid S)}{Pr(A = 1 \mid V)} + (1 - A) \times \frac{Pr(A = 0 \mid S)}{Pr(A = 0 \mid V)}.$$

$\widehat{\beta}$ unbiased for β if $\phi(S; \eta_{-0}) = G\eta_{-0}$ is true model for $\phi(s)$, and $A \perp \{Y(0), Y(1)\}$ given V.

Semi-parametric Regression Method (Encore)

Now, model is: $E(Y(a) \mid V) = \eta_0 + \phi^*(V) + aH\beta$.

Useful with confounders V ($\supset S$), have no model $\phi^*(V)$, and if we can assume that $V - S$ does not moderate impact of a on $Y(a)$.

Does NOT require model for nuisance function $\phi(V)$.

Get $\widehat{\beta}$ by solving the following estimating equations

$$0 = \mathbb{P}_n\left(\Big(Y - \widehat{b}(V;\xi) - AH\beta\Big)\Big(A - \widehat{p}(V;\alpha)\Big)H^T\right).$$

$\widehat{\beta}$ unbiased for β if $p(S;\alpha)$ is true model for $Pr(A = 1 \mid S)$, and $A \perp \{Y(0), Y(1)\}$ given V.

An Overview of Estimation Strategies

Model A: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$

Model B: $E(Y(a) \mid V) = \phi_0 + \phi(V) + aH\beta$

H is always a function of S Ex: $H = (1, S)$

	Model A			**Model B**
	No Confnders	S is Sole Confnder	Confnders $V^\sharp$	Modrtrs S, Confndrs V
ϕ Is Known	OLS*	OLS*	IPTW Regression*	OLS
ϕ Is Not Known	OLS (if $S \perp A$)	E-estimtr*†	IPTW E-estimtr*†	E-estimtr♯

*just discussed †need $Pr(A = 1 \mid S)$ ♯need $Pr(A = 1 \mid V)$

As a Decomposition of the Marginal Causal Effect

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

Consider the following arithmetic decomposition of the causal effect of (a_1, a_2) on Y, using the covariates $\bar{S}_2(a_1)$:

$$E\big[Y(a_1, a_2) - Y(0,0)\big] = E\Big[E\big[Y(a_1, a_2) - Y(a_1, 0) \mid \bar{S}_2(a_1)\big]\Big] + E\Big[E\big[Y(a_1, 0) - Y(0,0) \mid S_1\big]\Big].$$

The inner expectations represent the **conditional intermediate causal effects** μ_1 and μ_2, respectively.

Robins' Structural Nested Mean Model

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

$$
\begin{aligned}
&E\big[Y(a_1, a_2) \mid S_1, S_2(a_1)\big] \\
&\quad = E[Y(0,0)] + \Big\{E[Y(0,0) \mid S_1] - E[Y(0,0)]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y(a_1, 0) - Y(0, 0) \mid S_1}\big]\Big\} \\
&\qquad\qquad + \Big\{E[Y(a_1, 0) \mid \bar{S}_2(a_1)] - E[Y(a_1, 0) \mid S_1]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y(a_1, a_2) - Y(a_1, 0) \mid \bar{S}_2(a_1)}\big]\Big\} \\
&\quad = \mu_0 + \epsilon_1(s_1) + \boldsymbol{\mu_1(s_1, a_1)} + \epsilon_2(\bar{s}_2, a_1) + \boldsymbol{\mu_2(\bar{s}_2, \bar{a}_2)}
\end{aligned}
$$

SNMM Property I: $\mu_t = 0$ when $a_t = 0$.

$$
\begin{aligned}
&E\big[Y(a_1, a_2) \mid S_0, S_1(a_1)\big] \\
&\quad = E[Y(0,0)] + \Big\{E[Y(0,0) \mid S_0] - E[Y(0,0)]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y(a_1, 0) - Y(0,0) \mid S_0}\big]\Big\} \\
&\qquad\qquad + \Big\{E[Y(a_1, 0) \mid \bar{S}_1(a_1)] - E[Y(a_1, 0) \mid S_0]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y(a_1, a_2) - Y(a_1, 0) \mid \bar{S}_1(a_1)}\big]\Big\} \\
&\quad = \mu_0 + \epsilon_1(s_0) + \boldsymbol{\mu_1(s_0, a_1)} + \epsilon_2(\bar{s}_1, a_1) + \boldsymbol{\mu_2(\bar{s}_1, \bar{a}_2)}
\end{aligned}
$$

✓ $\mu_1(s_0, 0) = 0$ Ex Model: $a_1(\beta_{10} + \beta_{11} s_0)$

✓ $\mu_2(\bar{s}_1, a_2, 0) = 0$ Ex Model: $a_2(\beta_{20} + \beta_{21} s_1)$

SNMM Property II: ϵ_t's are conditional mean zero.

$$
\begin{aligned}
&E\big[Y(a_1,a_2) \mid S_0, S_1(a_1)\big] \\
&\quad = E[Y(0,0)] + \Big\{\mathbf{E[Y(0,0) \mid S_0] - E[Y(0,0)]}\Big\} \\
&\qquad\qquad + \Big\{E\big[Y(a_1,0) - Y(0,0) \mid S_0\big]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E[Y(a_1,0) \mid \bar{S}_1(a_1)] - E[Y(a_1,0) \mid S_0]}\Big\} \\
&\qquad\qquad + \Big\{E\big[Y(a_1,a_2) - Y(a_1,0) \mid \bar{S}_1(a_1)\big]\Big\} \\
&\quad = \mu_0 + \boldsymbol{\epsilon_1(s_0)} + \mu_1(s_0,a_1) + \boldsymbol{\epsilon_2(\bar{s}_1,a_1)} + \mu_2(\bar{s}_1,\bar{a}_2)
\end{aligned}
$$

✓ $E_{S_1|S_0}[\epsilon_2(\bar{s}_1,a_1) \mid S_0 = s_0] = 0$, and $E_{S_0}[\epsilon_1(s_0)] = 0$

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

So what's wrong with the Traditional Estimator?

Ex: Use the **Traditional Estimator** to model the $t = 2$ SNMM as:

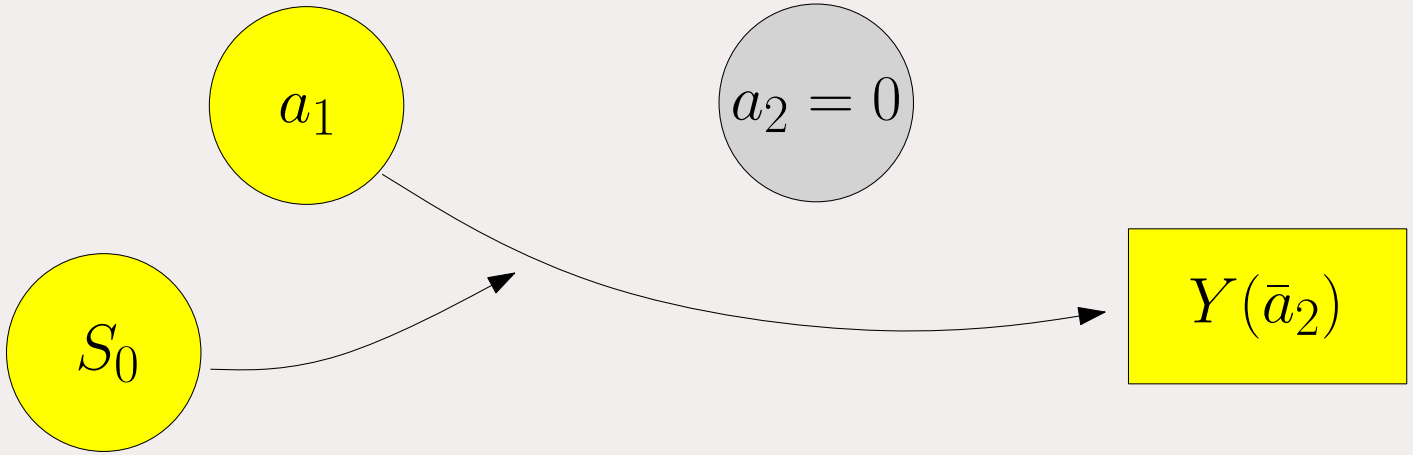
$$E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_0 + \beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_0} + \eta_2 s_1 + \beta_3^* \mathbf{a_2} + \beta_4^* \mathbf{a_2 s_0} + \beta_5^* \mathbf{a_2 s_1}$$

- Two problems arise with the interpretation of β_1^* and β_2^*.
- These two problems may occur **even when**
  - We use the correct model for the conditional mean, or
  - The sole time-varying confounder is the putative time-varying moderator S_t, or
  - There is no time-varying confounding bias at all!

Traditional approach to estimate μ_1 is problematic.

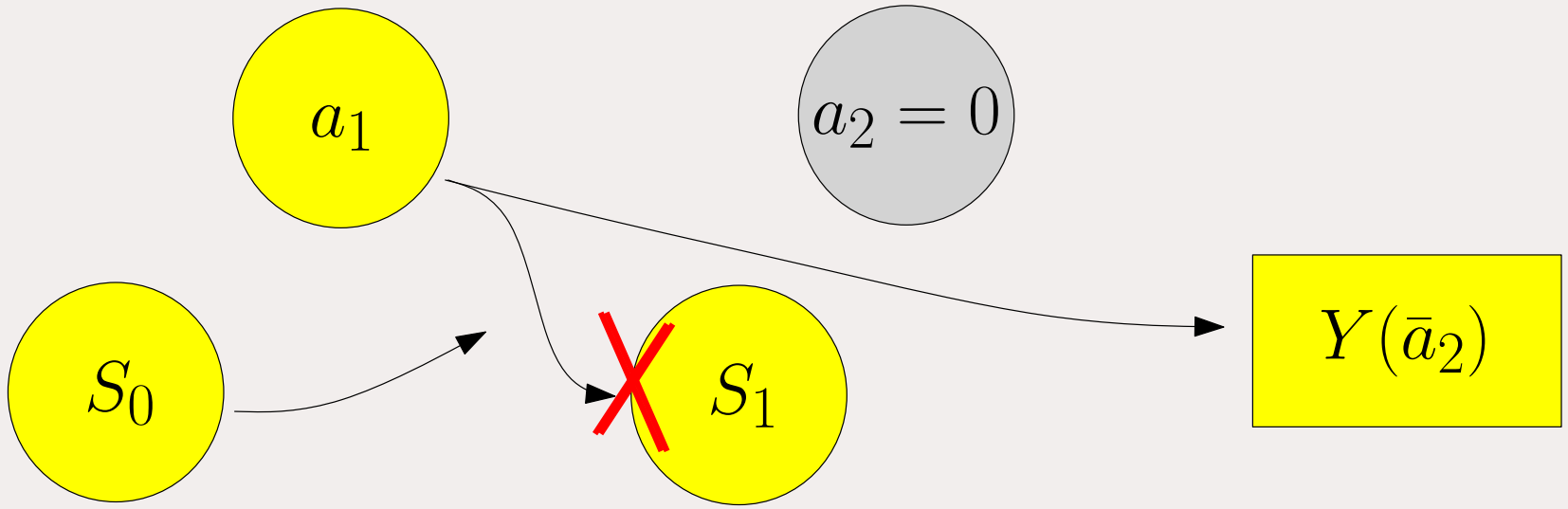
To explain what is wrong with the traditional estimator, we focus on estimating μ_1 using the traditional approach.

$$\mu_{\mathbf{1}}(\mathbf{s_0}, \mathbf{a_1}) = E[Y(a_1, 0) - Y(0, 0) \mid S_0 = s_0]$$



First problem with the Traditional Approach

Wrong Effect

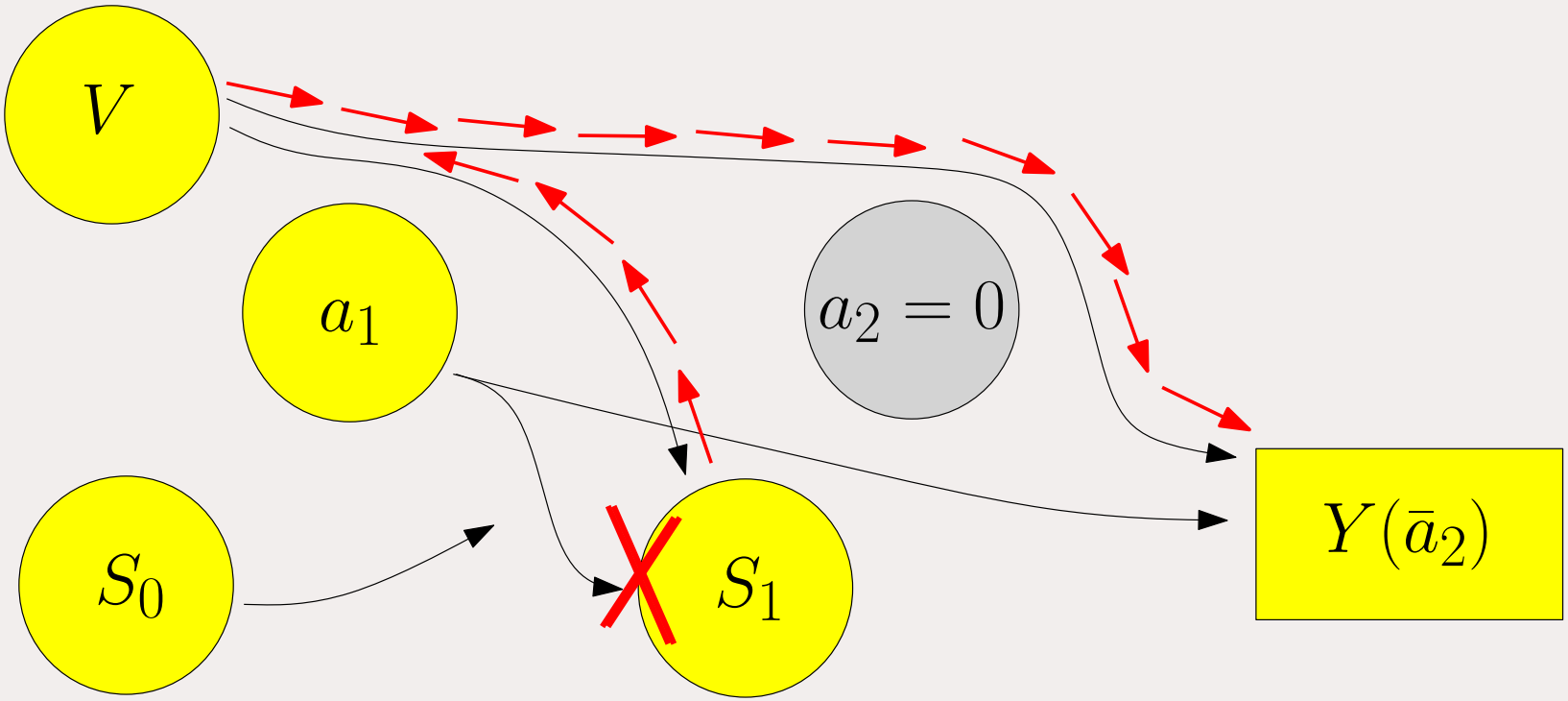


But what about the effect transmitted through $S_1(a_1)$?

So the end result is the term $\beta_1^* \mathbf{a_1} + \beta_2^* \mathbf{a_1 s_0}$ does not capture the "total" impact of $(a_1, 0)$ vs $(0, 0)$ on Y given values of S_0.

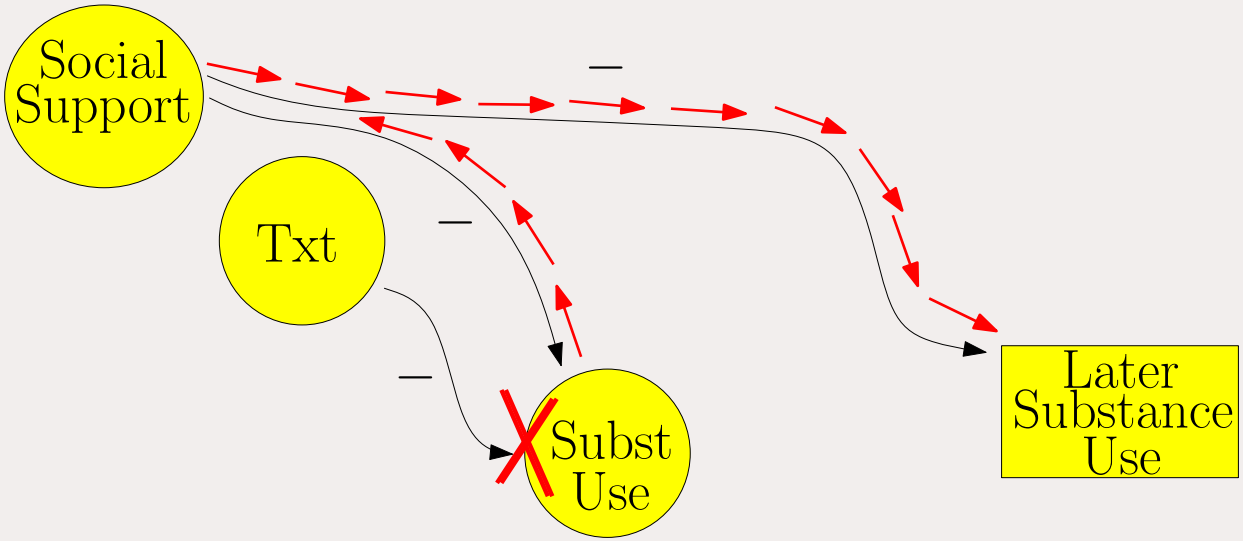
Second problem with the Traditional Approach

Spurious Bias



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

Intuition about the Spurious Bias



Imagine adolescent who is a high user despite getting treated:

Q: What does this tell you in terms of his social support?

A: There must be poor social support.

Implication: Conditional on substance use, getting treated is associated with more substance use! Bias is $-1(-)(-)(-) = +$.

The "old" warning against adjusting for post-treatment measures.

- Robins, Hernan, Cole, van der Laan, Pearl, Vanderwheele, & many others have published countless articles on elucidating this problem.
- Rosenbaum has an early article on this issue as well.
- Berkson's paradox—in the context of case-control studies using hospitalized samples—is related to this problem.
- Clinical trialists have been warning against this for a very long time! This is part of the reason why they advocate for ITT.

Proposed 2-Stage Regression Estimator

Recall that $E\big[Y(a_1, a_2) \mid \bar{S}_2(a_1) = \bar{s}_2\big] = \mu_2(\bar{s}_2, \bar{a}_2; \beta_2)$
$+ \epsilon_2(\bar{s}_2, a_1; \eta_2, \gamma_2) + \mu_1(s_1, a_1; \beta_1) + \epsilon_1(s_1; \eta_1, \gamma_1) + \mu_0$.

1. We have models for the μ's: $A_1 H_1 \beta_1$ and $A_2 H_2 \beta_2$; Set aside
2. Model $m_1(\gamma_1) = E(S_1)$, estimate γ_1 with GLM; model $m_2(s_1, a_1; \gamma_2) = E(S_2(a_1) \mid S_1 = s_1)$, estimate γ_2 with GLM
3. Construct residuals $\hat{\delta}_1 = s_1 - \hat{m}_1(\hat{\gamma}_1)$ and $\hat{\delta}_2 = s_2 - \hat{m}_2(s_1, a_1; \hat{\gamma}_2)$
4. Construct models for ϵ's: $G_1 \hat{\delta}_1 \eta_1 = G_1^* \eta_1$ and $G_2 \hat{\delta}_2 \eta_2 = G_2^* \eta_2$
5. Obtain $\hat{\beta}$ and $\hat{\eta}$ using OLS of $Y \sim [1, G_1^*, A_1 H_1, G_2^*, A_2 H_2]$

Data Descriptives: Treatment Trajectories

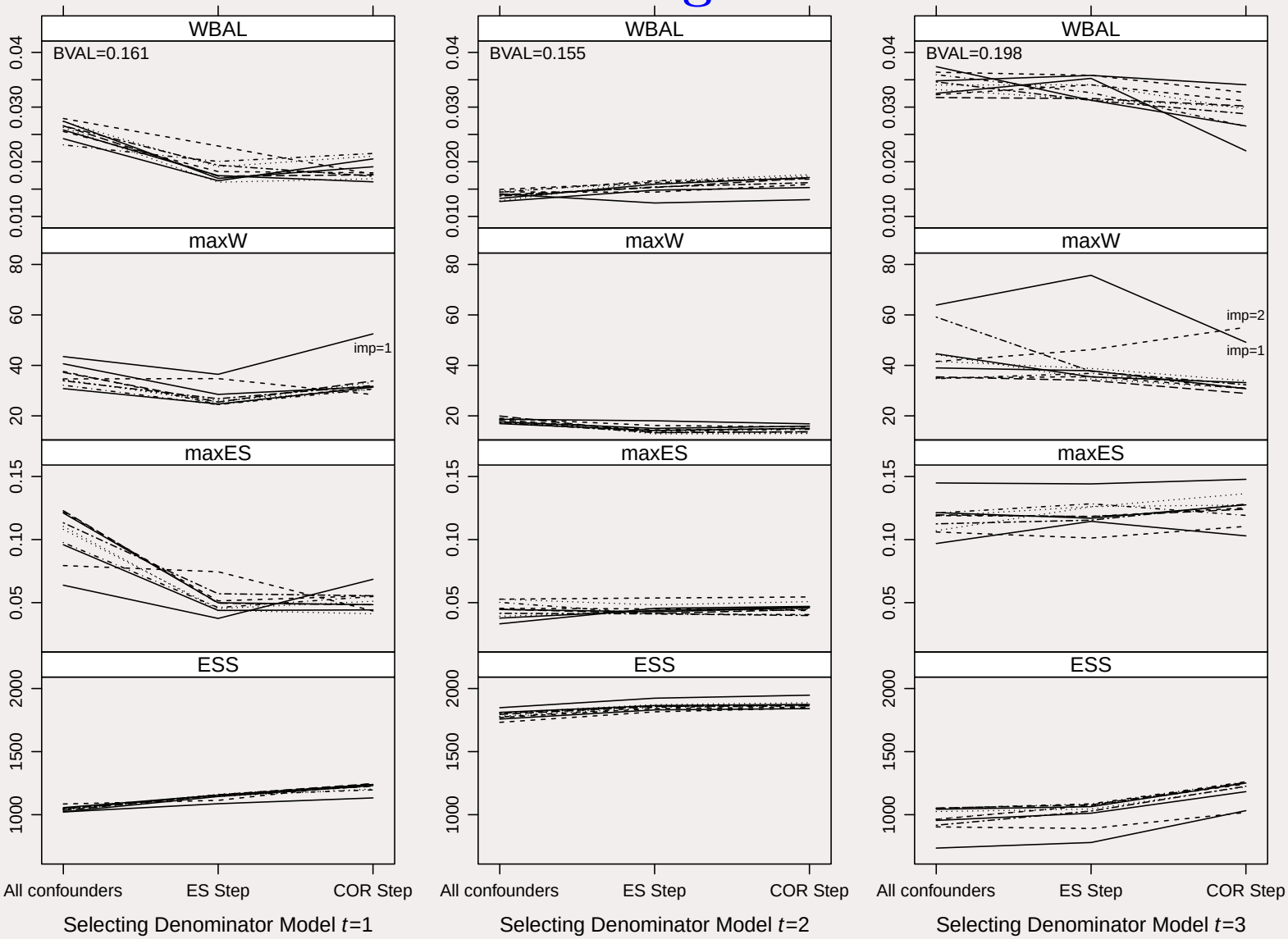
Treatment (A_1, A_2, A_3)	Frequency	Proportion
(0,0,0)	310	11%
(0,1,0)	56	2%
(1,0,0)	1184	41%
(1,1,0)	555	19%
(0,0,1)	56	2%
(0,1,1)	56	2%
(1,0,1)	153	5%
(1,1,1)	499	17%

Data Descriptives: Moderators and Outcomes

Moderators		Mean	SD	Range
S_0	sfs8p0	0.18	0.18	(0, 0.89)
	b2a	15.98	1.4	(12, 25)
	maxce0	13.95	24.6	(0, 90)
S_1	sfs8p3	0.07	0.11	(0, 0.67)
S_2	sfs8p6	0.08	0.13	(0, 0.73)

Outcome		Mean	SD	Range
Y	sfs8p12	0.09	0.13	(0, 0.78)

How did we choose our weights?



How did the weights do?

		Denom Pr.	Denominator weights		Balance		
t	No.	$(\min, \max)$	$(\min, \max)$	ESS	J	B	M
1	19	(0.20, 0.98)	(1.02, 32.83)	1214.8	46	0.041	0.13
2	45	(0.03, 0.95)	(1.03, 15.43)	1867.7	86	0.024	0.13
3	76	(0.01, 0.97)	(1.01, 37.93)	1057.0	126	0.037	0.16

Existing Semi-parametric G-Estimator

Recall our SNMM:

$$E\big[Y(a_1, a_2) \mid S_1, S_2(a_1)\big] = \mu_0 + \epsilon_1(s_1) + \beta_{\mathbf{10}}\mathbf{a_1} + \beta_{\mathbf{11}}\mathbf{a_1 s_1}$$
$$+ \epsilon_2(\bar{s}_2, a_1) + \beta_{\mathbf{20}}\mathbf{a_2} + \beta_{\mathbf{21}}\mathbf{a_2 s_1} + \beta_{\mathbf{22}}\mathbf{a_2 s_2}$$

Robins' G-Estimator models the ϵ_t's implicitly, as part of an algorithm.

It also allows for incorrect models for the ϵ_t's if models for the time-varying propensity scores—$p_t = Pr(A_t \mid \bar{S}_t, A_{t-1})$—are correctly specified. That is, if either of the p_t's or ϵ_t's are correctly specified, then the G-Estimator yields unbiased estimates of the causal β's.

Robins' Semi-parametric G-Estimator

Robins' G-Estimator is the solution to these estimating equations:

$$0 = \mathbb{P}_n \left\{ \Big(Y - A_2 H_2 \beta_2 - b_2(\bar{S}_2, A_1)\Big) \times \Big(A_2 - p_2(\bar{S}_2, A_1)\Big) \times \begin{pmatrix} 0 \\ H_2' \end{pmatrix}' \right.$$

$$\left. + \Big(Y - A_2 H_2 \beta_2 - A_1 H_1 \beta_1 - b_1(S_1)\Big) \times \Big(A_1 - p_1(S_1)\Big) \times \begin{pmatrix} H_1' \\ \Delta'(H_1) \end{pmatrix}' \right\}$$

$$\Delta(H_1) = E\left[H_2 A_2 \middle| S_1, A_1 = 1\right] - E\left[H_2 A_2 \middle| S_1, A_1 = 0\right]$$

$$b_2(\bar{S}_2, A_1) = E\left[Y - A_2 H_2 \beta_2 \mid \bar{S}_2, A_1\right]$$

$$p_2(\bar{S}_2, A_1) = Pr\left[A_2 = 1 \mid \bar{S}_2, A_1\right]$$

$$b_1(S_1) = E\left[Y - A_2 H_2 \beta_2 - A_1 H_1 \beta_1 \mid S_1\right]$$

$$p_1(S_1) = Pr\left[A_1 = 1 \mid S_2\right]$$

Bias-Variance Trade-off

This discussion assumes true models for the causal effects, the μ_ts:

Robins' G-Estimator is unbiased if either p_t or b_t are correctly specified. So-called *double-robustness* property.

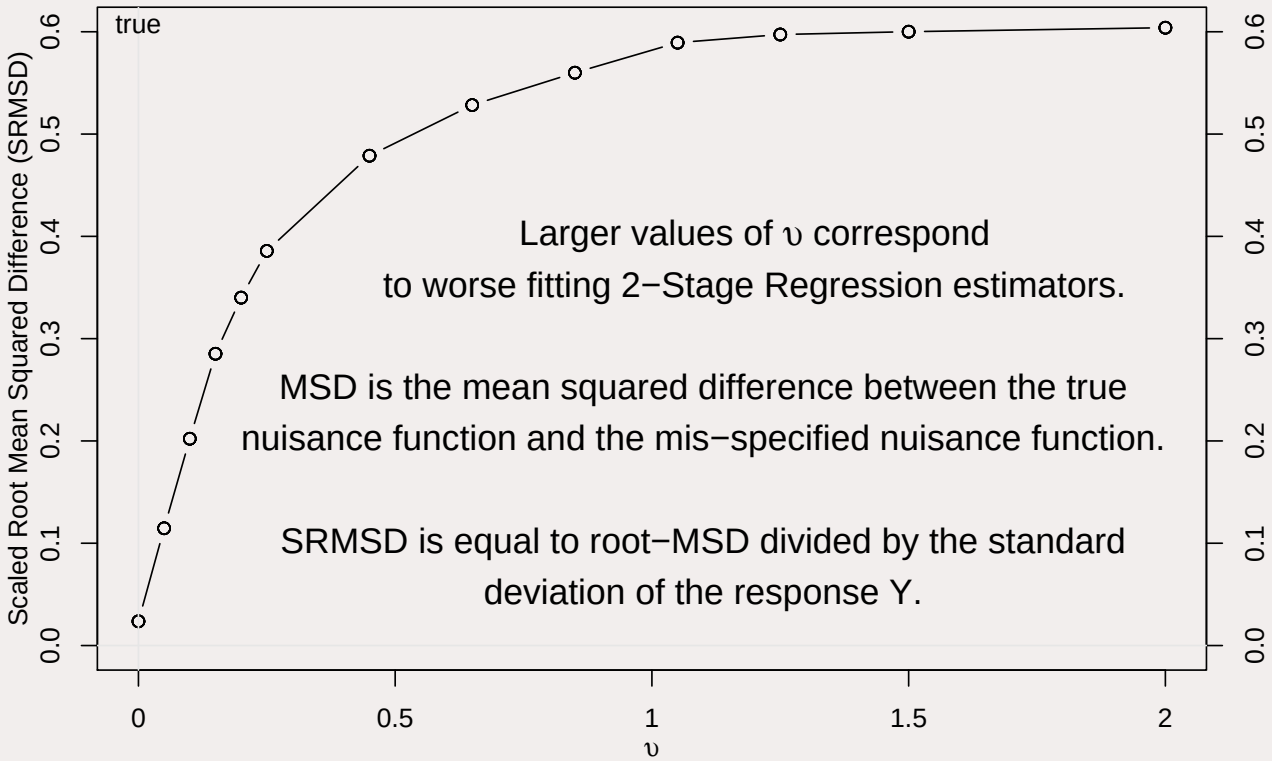
Robins' G-Estimator is semi-parametric efficient if p_t, b_t, and Δ are all correctly specified.

2-Stage Regression Estimator is unbiased only if the nuisance functions are correctly specified.

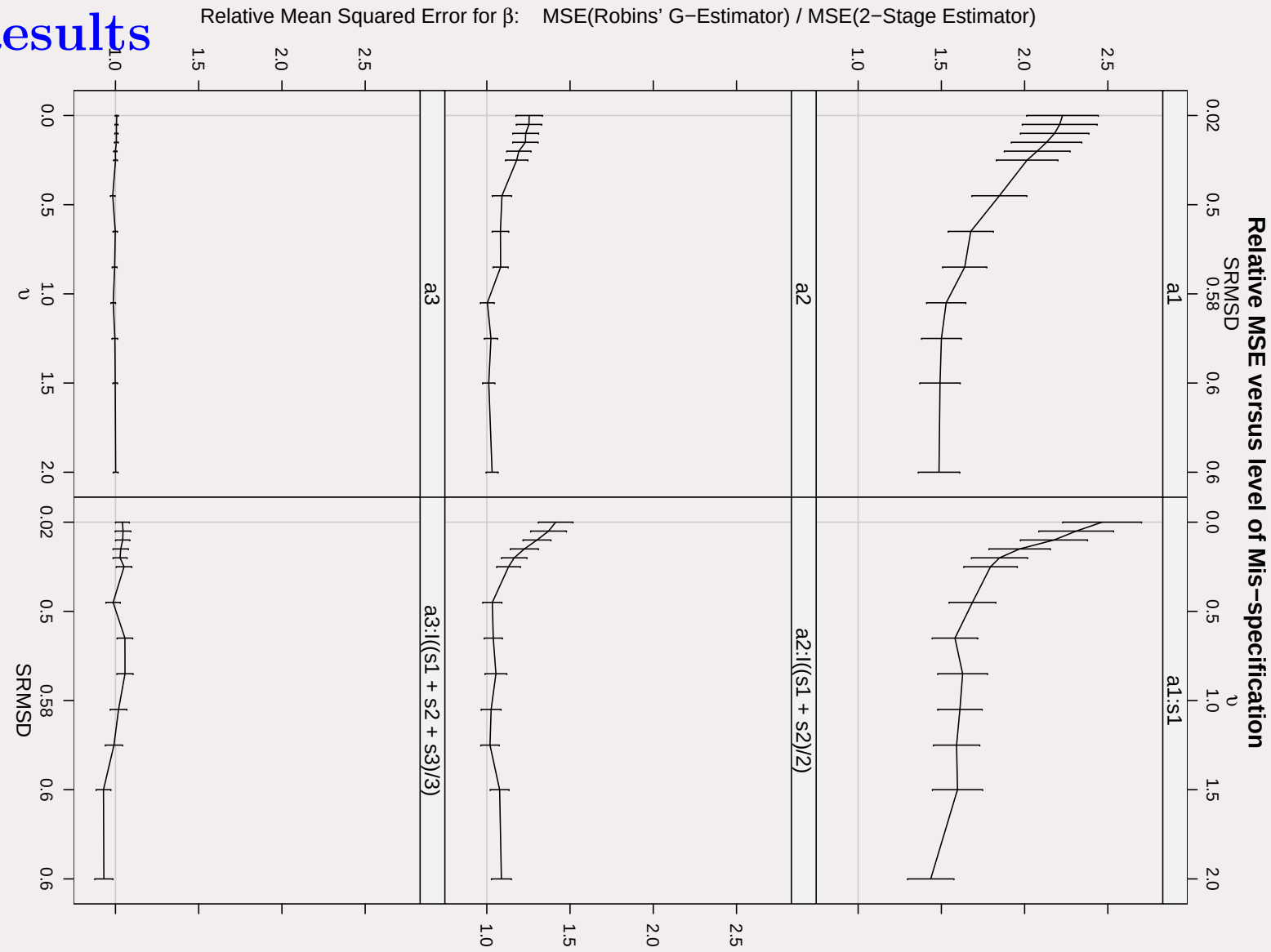
2-Stage Regression Estimator with correctly specified nuisance **is more efficient than** G-Estimator

But what happens as we mis-specify the nuisance functions?

Mis-specifying ϵ_t's using $S^* = S \times N(1, \mathbf{sd} = \nu)$



Results


Relative Mean Squared Error for β: MSE(Robins' G−Estimator) / MSE(2−Stage Estimator)
Relative MSE versus level of Mis−specification
a1
a1:s1
a2
a2:I((s1 + s2)/2)
a3
a3:I((s1 + s2 + s3)/3)
SRMSD
υ
0.02
0.5
0.58
0.6
0.0
1.0
1.5
2.0
2.5

The Generative Model in Simulations

$nits = 1000$ simulated data sets each of size $n = 500$

1. $\delta_1 \sim \widehat{res}_1$. Then $S_1 \longleftarrow 0.40 + \delta_1$.
2. $Z \longleftarrow \text{Bin}(n, p = 0.50)$. Then $A_1 \longleftarrow 0$ if $Z = 0$; otherwise

$$A_1 \longleftarrow \text{Bin}(n, p_1 = \Lambda(1.0 - 0.24 s_1))$$

3. $\delta_2 \sim \text{N}_n(0, \text{sd} = 0.75)$. Generate S_2 by setting

$$S_2 \longleftarrow 0.27 + 0.41 s_1 + 0.01 a_1 - 0.01 s_1^2 - 0.27 s_1 a_1 + \delta_2.$$

4. Set $A_2 \longleftarrow 0$ if $A_1 = 0$; otherwise

$$A_2 \longleftarrow \text{Bin}(n, p_2 = \Lambda(1.0 + 0.40 s_1 - 0.31 s_2)).$$

5. $\delta_3 \sim \mathsf{N}_n(0, \mathsf{sd} = 0.51)$. Generate S_3 by setting

$$S_3 \longleftarrow 0.17 + 0.10 s_1 - 0.25 a_1 + 0.30 s_2 - 0.75 a_2 + 0.05 s_1^2 - 0.04 s_2^2 - 0.1 a_1 s_1 + \delta_3.$$

6. Set $A_3 \longleftarrow 0$ if $A_2 = 0$; otherwise

$$A_3 \longleftarrow \mathsf{Bin}(n, p_3 = \Lambda(1.0 - 0.2 s_1 - 0.3 s_2 + 0.4 s_3)).$$

SNMM Generated as follows:

$$\begin{aligned} Y \longleftarrow\ & \mathsf{intercept} + \epsilon_1^{\mathsf{TRUE}}(s_1; \eta_1) + a_1(\beta_{1,1} + \beta_{1,2} s_1) \\ & + \epsilon_2^{\mathsf{TRUE}}(\bar{s}_2, a_1; \eta_2) + a_2(\beta_{2,1} + \beta_{2,2}(s_1 + s_2)/2) \\ & + \epsilon_3^{\mathsf{TRUE}}(\bar{s}_3, \bar{a}_2; \eta_3) + a_3(\beta_{1,1} + \beta_{3,2}(s_1 + s_2 + s_3)/3) + \delta_y, \end{aligned}$$

where

1. intercept $= 3.55$
2. $\beta_{1,1} = \beta_{2,1} = \beta_{3,2} = 0.30$,
3. $\beta_{1,2} = \beta_{2,2} = \beta_{3,1} = -0.30$,
4. δ_y is a random sample of size n from $\mathsf{N}(0, \mathsf{sd} = 0.7)$,

and where the true nuisance functions are defined as

1. $\epsilon_1^{\mathsf{TRUE}}(s_1; \eta_1) = 0.45 \times \delta_1$,
2. $\epsilon_2^{\mathsf{TRUE}}(\bar{s}_2, a_1; \eta_2) =$
   $(0.30 + 0.20 s_1 + 0.15 a_1 + 0.15 a_1 s_1 + 1.0 \sin(4.5 s_1)) \times \delta_2$,
3. $\epsilon_3^{\mathsf{TRUE}}(\bar{s}_3, \bar{a}_2; \eta_3) =$
   $(0.40 - 0.30 s_2 + 0.30 a_2 + 0.60 a_2 s_2 + 1.6 \sin(2.5 s_2)) \times \delta_3$.

Scaled Root Mean Squared Difference

This is how we measured **amount of mis-specification**:

$$\text{SRMSD}(\nu) = \sqrt{\frac{E\left(\sum_{t=3}^{K} \epsilon_t^{\text{TRUE}} - \sum_{t=1}^{K} \epsilon_t^{\nu}(\widehat{\eta}, \widehat{\gamma})\right)^2}{Var(Y)}},$$

where ν corresponds to a mis-specified 2-Stage Regression Estimator.

The expectation E and variance Var in SRMSD are over the data $D = (\bar{S}_3, \bar{A}_3, Y)$ for fixed $(\widehat{\eta}, \widehat{\gamma})$.

Calculated via Monte Carlo integration.

I claim SRMSD has an "effect-size-like" interpretation.

Confounding in PROSPECT

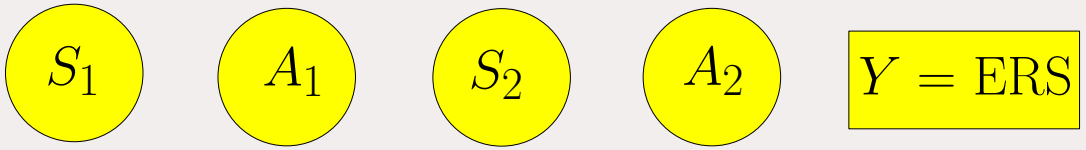
	Before Weighting			After Weighting
	Variable Name	Absolute Effect Size†	Sign	Effect Size‡
A_1 = HSANY_4				
	HAMDA_0	0.77	+	0.18
	RE_0	0.64	-	-0.05
	RE16N_0	0.66	-	-0.05
	MCS_0	0.53	-	-0.08
	MMSE2_0	0.45	+	0.22
	SSI_0	0.40	-	-0.29
A_2 = HSANY_8				
	CAD_4	0.80	+	0.27
	DYSTH_0	0.69	-	-0.66
	OPS_0	0.49	-	-0.07
	HAMDA_4	0.55	-	0.07
	CAD_0	0.51	+	0.47
	POSAF_0	0.53	+	-0.28
A_3 = HSANY_12				
	CAD_8	0.90	+	0.37
	WHITE1	0.71	+	0.08
	RP16N_0	0.60	+	0.28

Adolescent Substance Use and Community-based Treatment

- Motivating data set / application
- Collected by Center for Substance Abuse Treatment
- From a combination of major substance abuse programs
- Managed and cleaned by Chestnut Health Systems, IL (Michael Dennis)
- Global Appraisal of Individual Needs (GAIN): structured clinical interview, over 100 measures
- Full adolescent data set is $n = 6000$ and counting...

The Illustrative Data Set

- $n = 2870$ adolescents
- Interested in fitting a $K = 2$ time points SNMM

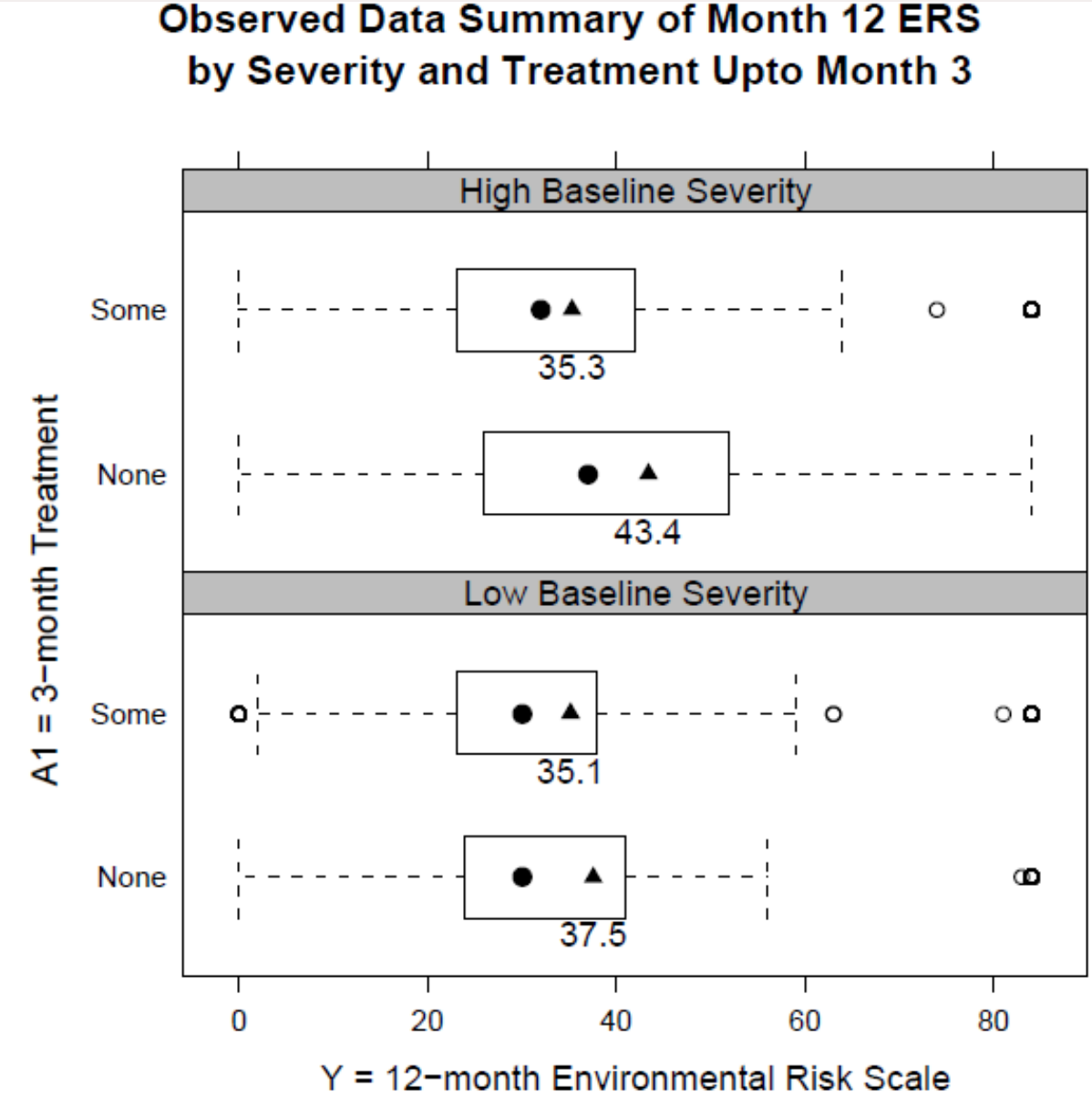


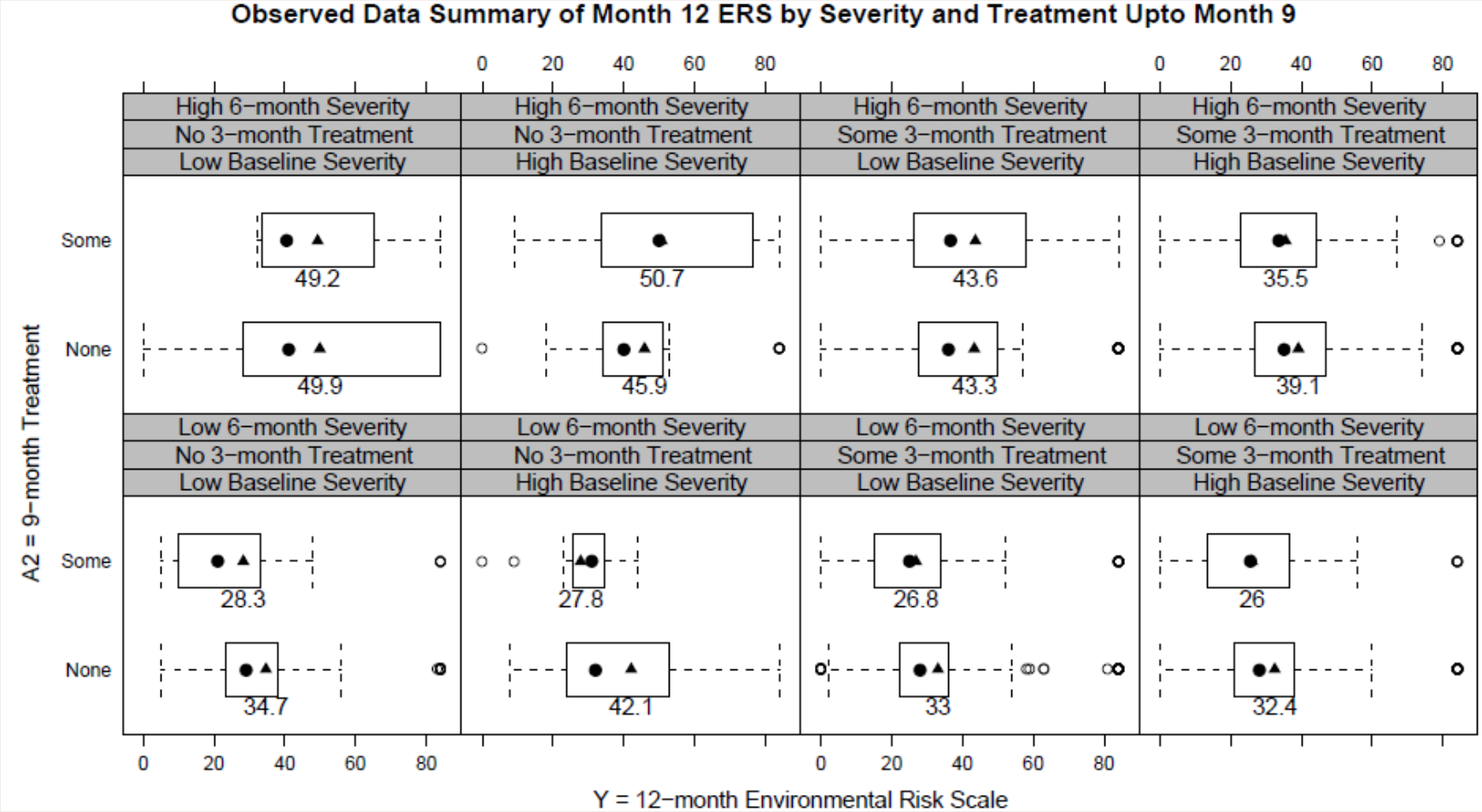
- S_1 = need0 = binary indicator of need/severity at baseline
- A_1 = anytxt3 = reported no treatment (0) versus some treatment (1=outpatient, inpatient, or both) at 3-months
- S_2 = need6 = binary indicator of need/severity at 6-months
- A_2 = anytxt9 = treatment indicator at at 9-months
- Y = ERS = Environmental Risk Scale at 12-months

What is the Scientific Question?

$\mu_1 =$ What is the effect of receiving treatment versus not at 3-months (and not receiving treatment in the future) on 12-month ERS scores, conditional on baseline severity?

$\mu_2 =$ What is the effect of receiving treatment versus not at 9-months on 12-month ERS scores, as a function of baseline severity, having received (or not) treatment at 3-months, and 6-month severity?


Observed Data Summary of Month 12 ERS
by Severity and Treatment Upto Month 3
High Baseline Severity
Low Baseline Severity
Some
None
Some
None
35.3
43.4
35.1
37.5
A1 = 3−month Treatment
0
20
40
60
80
Y = 12−month Environmental Risk Scale


Observed Data Summary of Month 12 ERS by Severity and Treatment Upto Month 9
High 6−month Severity
No 3−month Treatment
Low Baseline Severity
High 6−month Severity
No 3−month Treatment
High Baseline Severity
High 6−month Severity
Some 3−month Treatment
Low Baseline Severity
High 6−month Severity
Some 3−month Treatment
High Baseline Severity
Low 6−month Severity
No 3−month Treatment
Low Baseline Severity
Low 6−month Severity
No 3−month Treatment
High Baseline Severity
Low 6−month Severity
Some 3−month Treatment
Low Baseline Severity
Low 6−month Severity
Some 3−month Treatment
High Baseline Severity
A2 = 9−month Treatment
Some
None
49.2
49.9
50.7
45.9
43.6
43.3
35.5
39.1
28.3
34.7
27.8
42.1
26.8
33
26
32.4
0
20
40
60
80
Y = 12−month Environmental Risk Scale

Specifying the Saturated SNMM

Causal effects:

1. $\mu_1 = \text{anytxt}_3\,(\beta_{10} + \beta_{11}\ \text{need}_0)$,
2. $\mu_2 = \text{anytxt}_9(\beta_{20} + \beta_{21}\ \text{need}_0 + \beta_{22}\ \text{anytxt}_3 + \beta_{23}\ \text{need}_6 + \beta_{24}\ \text{need}_0\ \text{anytxt}_3 + \beta_{25}\ \text{need}_0\ \text{need}_6 + \beta_{26}\ \text{anytxt}_3\ \text{need}_6 + \beta_{27}\ \text{need}_0\ \text{anytxt}_3\ \text{need}_6)$

Nuisance functions:

1. $\epsilon_1 = \eta_{11} \times (\text{need}_0 - \text{Pr}(\text{need}_0 = 1))$,
2. $\epsilon_2 = (\eta_{21} + \eta_{22}\ \text{need}_0 + \eta_{23}\ \text{anytxt}_3 + \eta_{24}\ \text{need}_0\ \text{anytxt}_3) \times (\text{need}_6 - \text{Pr}(\ \text{need}_6 = 1 \mid\ \text{need}_0,\ \text{anytxt}_3))$, where

$$\text{Pr}(\text{need}_6 = 1 \mid \text{need}_0, \text{anytxt}_3) = \gamma_{20} + \gamma_{21}\ \text{need}_0 + \gamma_{22}\ \text{anytxt}_3 + \gamma_{23}\ \text{need}_0\ \text{anytxt}_3.$$

Estimates of the SNMM Using the 2-Stage Regression Estimator

	Parameters		2-Stage Estimator $\widehat{\beta}$	SE	P-val
μ_0	Int	β_{00}	39.76	1.36	< 0.01
μ_1	Int	β_{10}	-2.72	1.5	0.07
	need_0	β_{11}	-6.89	3.73	0.06
μ_2	Int	β_{20}	-6.59	4.04	0.10
	need_0	β_{21}	-2.12	6.17	0.73
	anytxt_3	β_{22}	1.13	4.20	0.78
	need_6	β_{23}	3.37	17.53	0.85
	$\text{need}_0\text{anytxt}_3$	β_{24}	4.26	6.52	0.52
	$\text{need}_0\text{need}_6$	β_{25}	5.79	20.49	0.77
	$\text{anytxt}_3\text{need}_6$	β_{26}	-0.47	17.98	0.99
	$\text{need}_0\text{anytxt}_3\text{need}_6$	β_{27}	-12.15	21.3	0.57

RR+IPTW vs RR vs TRAD

		Different Estimators		
Contrast	**Subgroup**	**RR+IPTW**	**RR**	**TRAD**
μ_1: **Distal**				
$(1,0,0)$ vs $(0,0,0)$	no intake sevrty, $<$ 16yrs	-0.004	-0.016	-0.002
$(1,0,0)$ vs $(0,0,0)$	hi intake sevrty, $\geq$ 16yrs	0.033	0.015	0.038
μ_2: **Medial**				
$(1,1,0)$ vs $(1,0,0)$	no 0-3 severity	-0.008	-0.005	0.000
$(1,1,0)$ vs $(1,0,0)$	hi 0-3 severity, yes ce	-0.048	-0.067	-0.040
$(1,1,0)$ vs $(1,0,0)$	hi 0-3 severity, no ce	0.021	-0.037	-0.010
μ_3: **Proximal**				
$(1,1,1)$ vs $(1,1,0)$	no 3-6 severity	-0.006	-0.012	-0.012
$(1,1,1)$ vs $(1,1,0)$	hi 3-6 severity	-0.168	-0.110	-0.110
$(.,.,1)$ vs $(.,.,0)$	no 3-6 severity	0.026	0.002	0.002
$(.,.,1)$ vs $(.,.,0)$	hi 3-6 severity	-0.165	-0.144	-0.144

Some conjectures about the methodological story

- Spurious bias for the distal effect is probably POS (see arguments I made earlier)
- Confounding bias for the distal effect is probably NEG (good kids get (1,0,0) and also have better/lower Y)
- Spurious and confounding bias cancel each other out and this is why we see TRAD approximately the same as RR+IPTW.
- Confounding bias for the proximal effect is probably POS (bad kids get (1,1,1) and also have worse/higher Y) and this is why we see the estimated proximal effects under RR+IPTW (vs are much stronger NEG

9 Connections with the Marginal Structural Model

- The MSM is a model for $E(Y(a_1, a_2) \mid S_0)$
- The SNMM is a model for $E(Y(a_1, a_2) \mid S_0, S_1(a_1))$
- So the law of iterated expectations gives us the MSM:

$$
\begin{aligned}
&E\big[Y(a_1, a_2) \mid S_0 = s_0\big] \\
&= E_{S_1(a_1)|S_0}\Big(E\big[Y(a_1, a_2) \mid \bar{S}_1(a_1) = \bar{s}_1\big]\Big) \\
&= \mu_0 + \epsilon_1(s_0) + \boldsymbol{\mu_1(s_0, a_1)} + E_{S_1(a_1)|S_0}\Big(\epsilon_2(\bar{s}_1, a_1) + \boldsymbol{\mu_2(\bar{s}_1, \bar{a}_2)}\Big) \\
&= \mu_0 + \epsilon_1(s_0) + \boldsymbol{\mu_1(s_0, a_1)} + E_{S_1(a_1)|S_0}\Big(\boldsymbol{\mu_2(\bar{s}_1, \bar{a}_2)}\Big)
\end{aligned}
$$

Connections with the Marginal Structural Model

- Due to linearity: If effect moderation $\exists$, we can get MSM estimates by plugging-in the estimated stage 1 regression for the time-varying moderators in the μ_t's. Think path analysis. But now we have to believe our stage 1 models are causal!
- If effect moderation $\nexists$, estimates for the μ_t's are indeed estimates for the marginal effects. Just read them off.
- Regardless, it is possible to use the RR+IPTW to get a double robust estimate of the marginal effects. Useful when we fail to balance on some covariates. To do this, (i) employ the plug-in estimator above, and (ii) don't use the numerator propensity score model in the weights.