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

Daniel Almirall\textsuperscript{1}, Daniel F. McCaffrey\textsuperscript{2}, Beth Ann Griffin\textsuperscript{2}, Rajeev Ramchand\textsuperscript{2}, Susan A. Murphy\textsuperscript{1}

\textsuperscript{1}Univ of Michigan, Institute for Social Research
\textsuperscript{2}RAND, Statistics

Michigan Department of Biostatistics, December 2011
Warm-up: Suppose we want $A \rightarrow Y$.

Examples

<table>
<thead>
<tr>
<th>$S$ = pre-$A$ covt</th>
<th>$A$ = txt/expsr</th>
<th>$Y$ = outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicidal?</td>
<td>Medication?</td>
<td>Depression</td>
</tr>
<tr>
<td>Gender, SES</td>
<td>SAT Coaching?</td>
<td>SAT Math Score</td>
</tr>
<tr>
<td>Social Support</td>
<td>Inpatient vs. Outpatient</td>
<td>Substance Abuse</td>
</tr>
</tbody>
</table>

Why condition on ("adjust for") pre-exposure covariables $S$?
Warm-up: Suppose we want $A \rightarrow Y$.

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

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

\[
\mu(s, a) \equiv E(Y(a) - Y(0) | S = s)
\]

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 Social Grouping Principle: 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. Think tailoring variables.
Mean Model in One Time Point

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

$$E(Y(a) \mid S = s) = E(Y(0))$$
$$+ \left( E(Y(0) \mid S = s) - E(Y(0)) \right)$$
$$+ E(Y(a) - Y(0) \mid S = s)$$
$$= \eta_0 + \epsilon(s) + \mu(s, a)$$

\[\text{e.g.}\]
$$= \eta_0 + \eta_1(s - E(S)) + \beta_1 a + \beta_2 as.$$

This is precisely what I would do, too.
2 Time-Varying Setting

The data structure in the time-varying setting is:

\[ S_0 \quad a_1 \quad S_1(a_1) \quad a_2 \quad S_2(\bar{a}_2) \quad a_3 \quad Y(\bar{a}_3) \]

Motivating Example: Adolescents & Substance Use Treatment

- \( S_0 \): Severity/need at intake visit; continuous
- \( a_1 \): 0-3mo treatment; binary, \( a_1 = \text{yes/no} \)
- \( S_1(a_1) \): Severity 0-3mo; continuous
- \( a_2 \): 3-6mo treatment; binary, \( a_2 = \text{yes/no} \)
- \( S_2(a_1, a_2) \): Severity 3-6mo; continuous
- \( a_3 \): 6-9mo treatment; binary, \( a_3 = \text{yes/no} \)
- \( Y(a_1, a_2, a_3) \): Substance use frequency 9-12mo; continuous
3 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

\[
\begin{align*}
(0,0,0), & \quad 11\% & (0,0,1), & \quad 2\% \\
(1,0,0), & \quad 41\% & (0,1,1), & \quad 2\% \\
(1,1,0), & \quad 19\% & (1,0,1), & \quad 5\% \\
(1,1,1), & \quad 17\% & (0,1,0), & \quad 2\%
\end{align*}
\]

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?
4 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 $S_0$?

$$\mu_1(s_0, a_1) = E[Y(a_1, 0, 0) - Y(0, 0, 0) \mid S_0 = s_0]$$
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 $(S_0, S_1)$?

$$\mu_2(\bar{s}_1, \bar{a}_2) = E[Y(a_1, a_2, 0) - Y(a_1, 0, 0) \mid S_0 = s_0, S_1(a_1) = s_1]$$
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 $(S_0, S_1, S_2)$?

$$\mu_3(\bar{s}_2, \bar{a}_3) = E[Y(a_1, a_2, a_3) - Y(a_1, a_2, 0) | \bar{S}_2(a_1, a_2) = \bar{s}_2]$$
Robins (1994, Communications in Statistics) calls the $\mu_t$’s “blip” functions. You can see why:

(1, 0, 0) vs (0, 0, 0), Distal
(1, 1, 0) vs (1, 0, 0), Medial
(1, 1, 1) vs (1, 1, 0), Proximal

For simplicity, for the next few slides we focus on just distal and proximal moderated effects. That is, the 2 time points setting.

We come back to 3 time points later in the talk, when we show the results of our data analysis.
Robins’ Structural Nested Mean Model decomposes $E(Y \mid \cdot)$ into nuisance and causal parts:

$$E[Y(a_1, a_2) \mid S_0, S_1(a_1)]$$

$$= E[Y(0, 0)] + \left\{ E[Y(0, 0) \mid S_0] - E[Y(0, 0)] \right\}$$

$$+ \left\{ E[Y(a_1, 0) - Y(0, 0) \mid S_0] \right\}$$

$$+ \left\{ E[Y(a_1, 0) \mid \tilde{S}_1(a_1)] - E[Y(a_1, 0) \mid S_0] \right\}$$

$$+ \left\{ E[Y(a_1, a_2) - Y(a_1, 0) \mid \tilde{S}_1(a_1)] \right\}$$

$$= \mu_0 + \epsilon_1(s_0) + \mu_1(s_0, a_1) + \epsilon_2(\tilde{s}_1, a_1) + \mu_2(\tilde{s}_1, \tilde{a}_2)$$

$\mu_1(s_0, a_1) =$ Moderated effects of $a_1$ on $Y(a_1, 0)$ given $S_0$

$\mu_2(\tilde{s}_1, \tilde{a}_2) =$ Moderated effects of $a_2$ on $Y(a_1, a_2)$ given $(S_0, a_1, S_1)$
Constraints on the Causal and Nuisance Portions

\[ E[Y(a_1, a_2) \mid \bar{S}_1(a_1) = \bar{s}_1] = \mu_0 + \epsilon_1(s_0) + \mu_1(s_0, a_1) \]
\[ + \epsilon_2(\bar{s}_1, a_1) + \mu_2(\bar{s}_1, \bar{a}_2), \]

where

\[ \cdot \mu_2(\bar{s}_1, a_2, 0) = 0 \text{ and } \mu_1(s_0, 0) = 0, \]
\[ \cdot \epsilon_2(\bar{s}_1, a_1) = E[Y(a_1, 0) \mid \bar{S}_1(a_1) = \bar{s}_1] - E[Y(a_1, 0) \mid S_0 = s_0], \]
\[ \cdot \epsilon_1(s_0) = E[Y(0, 0) \mid S_0 = s_0] - E[Y(0, 0)], \]
\[ \cdot E_{S_1 \mid S_0}[\epsilon_2(\bar{s}_1, a_1) \mid S_0 = s_0] = 0, \text{ and } E_{S_0}[\epsilon_1(s_0)] = 0. \]

The \( \epsilon_t \)'s make the SNMM a non-standard regression model.
6 Estimation

Almirall, Ten Have, Murphy (2010, Biometrics): We proposed a 2-stage “regression with residuals” (RR) estimator that gets around the problems of traditional regression methods. “Close” to the traditional approach.

Here: We propose an inverse-weighted regression with residuals estimator (RR+IPTW) to reduce bias in the presence of an auxiliary set of observed time-varying confounders.

But first…traditional methods may not identify the causal effects.

- Even in the absence of time-varying confounding bias.
So what’s wrong with the Traditional Estimator?

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

$$E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_0 + \beta_1^* a_1 + \beta_2^* a_1 s_0$$

$$+ \eta_2 s_1 + \beta_3^* a_2 + \beta_4^* a_2 s_0 + \beta_5^* a_2 s_1$$

- Two problems arise with the interpretation of $\beta_1^*$ and $\beta_2^*$.
- 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 $\mu_1$ is problematic.

To explain what is wrong with the traditional estimator, we focus on estimating $\mu_1$ using the traditional approach.

$$\mu_1(s_0, a_1) = E[Y(a_1, 0) - Y(0, 0) | S_0 = s_0]$$
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 $\beta_1^*a_1 + \beta_2^*a_1s_0$ 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 Effect

\[ V \]

\[ a_1 \]

\[ S_0 \]

\[ S_1 \]

\[ a_2 = 0 \]

\[ Y(\bar{a}_2) \]

This is also known as “Berkson’s paradox”; and is related to Judea Pearl’s back-door criterion and “collider 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.
The “old” problem of adjusting for post-treatment measures.

Note: 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—is related to this problem. There are likely many more...

A further side note is that this problem is central to the reason why clinical trialists have advocated for ITT over the past 40 years or so...
Proposed 2-Stage Regression Estimator

Instead of the traditional regression estimator

\[
E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta^*_0 + \eta_1 s_0 + \beta^*_1 a_1 + \beta^*_2 a_1 s_0 \\
+ \eta_2 s_1 + \beta^*_3 a_2 + \beta^*_4 a_2 s_0 + \beta^*_5 a_2 s_1,
\]

we use the following

\[
E(Y \mid \bar{S}_1 = \bar{s}_1, \bar{A}_2 = \bar{a}_2) = \beta^*_0 + \eta_1 s_0 + \beta^*_1 a_1 + \beta^*_2 a_1 s_0 \\
+ \eta_2 (s_1 - E(S_1 \mid A_1, S_0)) + \beta^*_3 a_2 + \beta^*_4 a_2 s_0 + \beta^*_5 a_2 s_1.
\]

We call it “2-stage regression with residuals” because first we estimate \(E(S_1 \mid A_1, S_0)\), then use the residual \(s_1 - E(S_1 \mid A_1, S_0)\) in a second regression to get \(\beta\)’s.
**Proposed 2-Stage Regression Estimator**

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

The proposed estimator is unbiased for the \(\mu_t\)'s provided:

1. Correctly modeled SNMM, incl. the \(\epsilon_t\)'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...**
7 Time-Varying Confounding Bias

How do we adjust for measured time-varying covariates $X_t$ that are possible confounders, but are 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:

What is $X_t$?
EPS +
SPS +
MAXCE -
LRI +
AGE -
NONWHITE -
...

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

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

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

• \(W_1\) is used to estimate the \(t = 1\) SNMM.

• \(W_1 \times W_2\) is used to estimate the \(t = 2\) SNMM.

• Observe that, unlike when IPTW is used to estimate a MSM, here we can stabilize using a time-varying covariate (namely, the putative time-varying moderator of interest \(S_t\)).
8 Data Analysis

• From US substance abuse prgms (CSAT ⊂ SAMHSA)
• Global Appraisal of Individual Needs (GAIN): structured clinical interview, over 100 measures
• \( n = 2870 \) adolescents; data every 3 months for 1 year
• \( \{(S_0, X_0), A_1, (S_1, X_1), A_2, (S_2, X_2), a_3, Y(a_1, a_2)\} \)
• \( S_t = \) substance frequency scale at intake, 0-3, 3-6
• \( X_t = \) 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
Diagnose Weights

Pre–Post Balance

B = Average Absolute Standardized Difference

Distribution of Weights

B = Average Absolute Standardized Difference

Weights

0.0 0.2 0.4 0.6 0.8

0 2 4 6 8 10 12 14

Unweighted

B = 0.1495

Weighted

B = 0.024
EDA: Distal Effect of Initial Txt

Given Baseline Severity

- Treatment Sequence (1,0,0)
- Treatment Sequence (0,0,0)
Distal Effect of Initial Treatment on 12-month SFS

Treatment sequence (1,0,0) versus (0,0,0)

Iatrogenic Effect

Beneficial Effect
EDA: Medial Effect of Cum. Txt

Given 3-month Severity

Treatment Sequence (1,1,0)
Treatment Sequence (1,0,0)
EDA: Proximal Effect of Cum. Txt

Given 6-month Severity

- Treatment Sequence (1,1,1)
- Treatment Sequence (1,1,0)
## Effect Estimates from SNMM, using RR+IPTW

<table>
<thead>
<tr>
<th>Effects</th>
<th>EST</th>
<th>EFF.SIZE</th>
<th>PVAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$\mu_1$: Distal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low intake severity, 14 yrs old</td>
<td>-0.016</td>
<td>-0.122</td>
<td>0.30</td>
</tr>
<tr>
<td>hi intake severity, 14 yrs old</td>
<td>0.017</td>
<td>0.130</td>
<td>0.50</td>
</tr>
<tr>
<td>low intake severity, 18 yrs old</td>
<td>0.014</td>
<td>0.104</td>
<td>0.47</td>
</tr>
<tr>
<td>hi intake severity, 18 yrs old</td>
<td>0.047</td>
<td>0.356</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>$\mu_2$: Medial</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low 0-3 severity, 14 yrs old</td>
<td>-0.016</td>
<td>-0.121</td>
<td>0.25</td>
</tr>
<tr>
<td>hi 0-3 severity, 14 yrs old</td>
<td>-0.027</td>
<td>-0.204</td>
<td>0.48</td>
</tr>
<tr>
<td>low 0-3 severity, 18 yrs old</td>
<td>-0.003</td>
<td>-0.022</td>
<td>0.87</td>
</tr>
<tr>
<td>hi 0-3 severity, 18 yrs old</td>
<td>-0.014</td>
<td>-0.105</td>
<td>0.69</td>
</tr>
<tr>
<td><strong>$\mu_3$: Proximal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low 6-9 severity, 14 yrs old</td>
<td>0.033</td>
<td>0.252</td>
<td>0.08</td>
</tr>
<tr>
<td>hi 6-9 severity, 14 yrs old</td>
<td>-0.144</td>
<td>-1.087</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>low 6-9 severity, 18 yrs old</td>
<td>-0.040</td>
<td>-0.305</td>
<td>0.03</td>
</tr>
<tr>
<td>hi 6-9 severity, 18 yrs old</td>
<td>-0.218</td>
<td>-1.644</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
9 Connections with the Marginal Structural Model

The (more commonly used) MSM is a model for $E(Y(a_1, a_2) \mid S_0)$ instead of a model for $E(Y(a_1, a_2) \mid S_0, S_1(a_1))$ (SNMM).

Of course, if we know the SNMM, then using the law of iterated expectations, we can get the MSM.

$$E[Y(a_1, a_2) \mid S_0 = s_0] = E_{S_1(a_1) \mid S_0} \left( E[Y(a_1, a_2) \mid \bar{S}_1(a_1) = \bar{s}_1] \right)$$

$$= \mu_0 + \epsilon_1(s_0) + \mu_1(s_0, a_1)$$

$$+ E_{S_1(a_1) \mid S_0} \left( \epsilon_2(\bar{s}_1, a_1) + \mu_2(\bar{s}_1, \bar{a}_2) \right)$$
Connections with the Marginal Structural Model

Note that if there is no effect moderation, then estimates for the $\mu_t$’s are indeed estimates for the marginal effects.

Also note that if there is no time-varying effect moderation, then it is possible to use a version of the estimator we propose as a double robust estimator of the MSM. This might be very useful in cases where we know there are certain time-varying covariates which the weights fail to balance on.
10 Conclusion

- **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
  - Percentile bootstrap CI’s provide adequate coverage in preliminary simulations
Acknowledgements

Bobby Yuen, Graduate Student, Michigan Statistics
Cha-chi Fan & Mary Ellen Slaughter, RAND

NIDA Funding:
R01-DA-015697 (PIs: McCaffrey, Griffin & Co-I: Ramchand)
The Methodology Center at Penn State University
(P50-DA-010075; PIs: Collins, Murphy & Co-I: Almirall)

NIMH Funding:
R01-MH-080015 (PI: Murphy)
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, Griffin, Ramchand, Murphy (Dec 2011?). Will deposit as Tech Report at PSU Methodology Center.
11 Extra Slides
Prototypical Linear Parametric Model

We use $\beta$ for our causal parameters of interest:

$$E(Y(a) \mid S) = \eta_0 + \phi(S) + \mu(S, a; \beta)$$
$$= \eta_0 + \phi(S) + aH\beta$$

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 $\beta$ 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 \((\hat{\eta}, \hat{\beta})\). The \( \hat{\beta} \) estimates solve

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

\( \hat{\beta} \) unbiased for \( \beta \) 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 $\hat{\beta}$ by solving the following estimating equations

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

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

$\hat{\beta}$ unbiased for $\beta$ 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 \( V \supseteq S \).

Requires model for nuisance function: \( \phi(S; \eta_0) = G\eta_0. \)

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

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

\( \hat{\beta} \) unbiased for \( \beta \) 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 \( \hat{\beta} \) by solving the following estimating equations

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

\( \hat{\beta} \) unbiased for \( \beta \) 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 \( S \) \hspace{1cm} \text{Ex: } H = (1, S) 

<table>
<thead>
<tr>
<th></th>
<th>Model A</th>
<th></th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Confnders</td>
<td>( S ) is Sole Confnder</td>
<td>Confnders ( V )</td>
</tr>
<tr>
<td>( \phi ) Known</td>
<td>OLS*</td>
<td>OLS*</td>
<td>IPTW Regression*</td>
</tr>
<tr>
<td>( \phi ) Not Known</td>
<td>OLS (if ( S \perp A ))</td>
<td>E-estimtr*†</td>
<td>IPTW E-estimtr*†</td>
</tr>
</tbody>
</table>

*just discussed \hspace{1cm} †need \( Pr(A = 1 \mid S) \) \hspace{1cm} ‡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 \(\bar{S}_2(a_1)\):

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

The inner expectations represent the conditional intermediate causal effects \(\mu_1\) and \(\mu_2\), 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:

$$E[Y(a_1, a_2) | S_1, S_2(a_1)]$$

$$= E[Y(0, 0)] + \left\{ E[Y(0, 0) | S_1] - E[Y(0, 0)] \right\}$$

$$+ \left\{ E[Y(a_1, 0) - Y(0, 0) | S_1] \right\}$$

$$+ \left\{ E[Y(a_1, 0) | \bar{S}_2(a_1)] - E[Y(a_1, 0) | S_1] \right\}$$

$$+ \left\{ E[Y(a_1, a_2) - Y(a_1, 0) | \bar{S}_2(a_1)] \right\}$$

$$= \mu_0 + \epsilon_1(s_1) + \mu_1(s_1, a_1) + \epsilon_2(\bar{s}_2, a_1) + \mu_2(\bar{s}_2, \bar{a}_2)$$
Proposed 2-Stage Regression Estimator

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

1. We have models for the $\mu$’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 $\gamma_1$ with GLM; model $m_2(s_1, a_1; \gamma_2) = E(S_2(a_1) \mid S_1 = s_1)$, estimate $\gamma_2$ 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 $\epsilon$’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 $Y \sim [1, G_1^*, A_1 H_1, G_2^*, A_2 H_2]$
**Existing Semi-parametric G-Estimator**

**Recall** our SNMM:

\[
E\left[Y(a_1, a_2) \mid S_1, S_2(a_1)\right] = \mu_0 + \epsilon_1(s_1) + \beta_{10}a_1 + \beta_{11}a_1s_1 \\
+ \epsilon_2(s_2, a_1) + \beta_{20}a_2 + \beta_{21}a_2s_1 + \beta_{22}a_2s_2
\]

Robins’ G-Estimator models the \( \epsilon_t \)'s implicitly, as part of an algorithm.

It also allows for incorrect models for the \( \epsilon_t \)'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 \( \epsilon_t \)'s are correctly specified, then the G-Estimator yields unbiased estimates of the causal \( \beta \)'s.
Robins’ Semi-parametric G-Estimator

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

\[
0 = \mathbb{P}_n \left\{ \left( Y - A_2 H_2 \beta_2 - b_2(\bar{S}_2, A_1) \right) \times \left( A_2 - p_2(\bar{S}_2, A_1) \right) \times \begin{pmatrix} 0 \\ H'_2 \end{pmatrix} \right. \\
+ \left. \left( Y - A_2 H_2 \beta_2 - A_1 H_1 \beta_1 - b_1(S_1) \right) \times \left( A_1 - p_1(S_1) \right) \times \begin{pmatrix} H'_1 \\ \Delta'(H_1) \end{pmatrix} \right\}
\]

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

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

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

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

\[
p_1(S_1) = Pr \left[ A_1 = 1 \middle| S_2 \right]
\]
Bias-Variance Trade-off

This discussion assumes true models for the causal effects, the $\mu_t$s:

**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 $\Delta$ 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 $\epsilon_t$’s using $S^* = S \times N(1, \text{sd} = \nu)$

Larger values of $\nu$ correspond to worse fitting 2-Stage Regression estimators.

MSD is the mean squared difference between the true nuisance function and the mis-specified nuisance function.

SRMSD is equal to root-MSD divided by the standard deviation of the response $Y$. 
Relative Mean Squared Error for $\beta$: $\frac{\text{MSE(Robins' G-Estimator)}}{\text{MSE(2-Stage Estimator)}}$

SRMSD

$\nu$

$0.02$ $0.5$ $0.58$ $0.6$ $0.6$
The Generative Model in Simulations

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

1. $\delta_1 \sim \hat{r}es_1$. Then $S_1 \leftarrow 0.40 + \delta_1$.

2. $Z \leftarrow \text{Bin}(n, p = 0.50)$. Then $A_1 \leftarrow 0$ if $Z = 0$; otherwise $A_1 \leftarrow \text{Bin}(n, p_1 = \Lambda(1.0 - 0.24s_1))$

3. $\delta_2 \sim N_n(0, sd = 0.75)$. Generate $S_2$ by setting $S_2 \leftarrow 0.27 + 0.41s_1 + 0.01a_1 - 0.01s_1^2 - 0.27s_1a_1 + \delta_2$.

4. Set $A_2 \leftarrow 0$ if $A_1 = 0$; otherwise $A_2 \leftarrow \text{Bin}(n, p_2 = \Lambda(1.0 + 0.40s_1 - 0.31s_2))$. 
5. \( \delta_3 \sim N_n(0, \text{sd} = 0.51) \). Generate \( S_3 \) by setting
\[
S_3 \leftarrow 0.17 + 0.10s_1 - 0.25a_1 + 0.30s_2 - 0.75a_2 + 0.05s_1^2 \\
- 0.04s_2^2 - 0.1a_1s_1 + \delta_3.
\]

6. Set \( A_3 \leftarrow 0 \) if \( A_2 = 0 \); otherwise
\[
A_3 \leftarrow \text{Bin}(n, p_3 = \Lambda(1.0 - 0.2s_1 - 0.3s_2 + 0.4s_3)).
\]

SNMM Generated as follows:
\[
Y \leftarrow \text{intercept} + \epsilon_1^{\text{TRUE}}(s_1; \eta_1) + a_1(\beta_{1,1} + \beta_{1,2}s_1) \\
+ \epsilon_2^{\text{TRUE}}(\bar{s}_2, a_1; \eta_2) + a_2(\beta_{2,1} + \beta_{2,2}(s_1 + s_2)/2) \\
+ \epsilon_3^{\text{TRUE}}(\bar{s}_3, \bar{a}_2; \eta_3) + a_3(\beta_{1,1} + \beta_{3,2}(s_1 + s_2 + s_3)/3) + \delta_y,
\]
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. $\delta_y$ is a random sample of size $n$ from $N(0, \text{sd} = 0.7),$

and where the true nuisance functions are defined as

1. $\epsilon_1^{\text{TRUE}}(s_1; \eta_1) = 0.45 \times \delta_1,$

2. $\epsilon_2^{\text{TRUE}}(\bar{s}_2, a_1; \eta_2) =
   (0.30 + 0.20s_1 + 0.15a_1 + 0.15a_1s_1 + 1.0\sin(4.5s_1)) \times \delta_2,$

3. $\epsilon_3^{\text{TRUE}}(\bar{s}_3, \bar{a}_2; \eta_3) =
   (0.40 - 0.30s_2 + 0.30a_2 + 0.60a_2s_2 + 1.6\sin(2.5s_2)) \times \delta_3.$
Scaled Root Mean Squared Difference

This is how we measured amount of mis-specification:

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

where \( \nu \) 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 \((\hat{\eta}, \hat{\gamma})\).

Calculated via Monte Carlo integration.

I claim SRMSD has an “effect-size-like” interpretation.
### Confounding in PROSPECT

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Before Weighting</th>
<th>After Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute Effect Size</td>
<td>Sign</td>
</tr>
<tr>
<td>HAMDA.0</td>
<td>0.77</td>
<td>+</td>
</tr>
<tr>
<td>RE.0</td>
<td>0.64</td>
<td>-</td>
</tr>
<tr>
<td>RE16N.0</td>
<td>0.66</td>
<td>-</td>
</tr>
<tr>
<td>MCS.0</td>
<td>0.53</td>
<td>-</td>
</tr>
<tr>
<td>MMSE2.0</td>
<td>0.45</td>
<td>+</td>
</tr>
<tr>
<td>SSI.0</td>
<td>0.40</td>
<td>-</td>
</tr>
</tbody>
</table>

\[ A_1 = \text{HSANY.4} \]

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Before Weighting</th>
<th>After Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute Effect Size</td>
<td>Sign</td>
</tr>
<tr>
<td>CAD.4</td>
<td>0.80</td>
<td>+</td>
</tr>
<tr>
<td>DYSTH.0</td>
<td>0.69</td>
<td>-</td>
</tr>
<tr>
<td>OPS.0</td>
<td>0.49</td>
<td>-</td>
</tr>
<tr>
<td>HAMDA.4</td>
<td>0.55</td>
<td>-</td>
</tr>
<tr>
<td>CAD.0</td>
<td>0.51</td>
<td>+</td>
</tr>
<tr>
<td>POSAF.0</td>
<td>0.53</td>
<td>+</td>
</tr>
</tbody>
</table>

\[ A_2 = \text{HSANY.8} \]

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Before Weighting</th>
<th>After Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute Effect Size</td>
<td>Sign</td>
</tr>
<tr>
<td>CAD.8</td>
<td>0.90</td>
<td>+</td>
</tr>
<tr>
<td>WHITE1</td>
<td>0.71</td>
<td>+</td>
</tr>
<tr>
<td>RP16N.0</td>
<td>0.60</td>
<td>+</td>
</tr>
</tbody>
</table>

\[ A_3 = \text{HSANY.12} \]
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

\[ S_1 \quad A_1 \quad S_2 \quad A_2 \quad Y = ERS \]

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

High Baseline Severity
- Some: 35.3
- None: 43.4

Low Baseline Severity
- Some: 35.1
- None: 37.5

Y = 12-month Environmental Risk Scale
Observed Data Summary of Month 12 ERS by Severity and Treatment Upto Month 9

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. \( \epsilon_1 = \eta_{11} \times (\text{need}_0 - \Pr(\text{need}_0 = 1)) , \)
2. \( \epsilon_2 = (\eta_{21} + \eta_{22} \text{need}_0 + \eta_{23} \text{anytxt}_3 + \eta_{24} \text{need}_0 \text{anytxt}_3) \times (\text{need}_6 - \Pr(\text{need}_6 = 1 | \text{need}_0, \text{anytxt}_3)) , \) where
   \( \Pr(\text{need}_6 = 1 | \text{need}_0, \text{anytxt}_3) = \gamma_{20} + \gamma_{21} \text{need}_0 + \gamma_{22} \text{anytxt}_3 + \gamma_{23} \text{need}_0 \text{anytxt}_3 . \)
### Estimates of the SNMM Using the 2-Stage Regression Estimator

<table>
<thead>
<tr>
<th>Parameters</th>
<th>2-Stage Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td></td>
</tr>
<tr>
<td>Int ( \beta_{00} )</td>
<td>39.76</td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td></td>
</tr>
<tr>
<td>Int ( \beta_{10} )</td>
<td>-2.72</td>
</tr>
<tr>
<td>need ( \beta_{11} )</td>
<td>-6.89</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td></td>
</tr>
<tr>
<td>Int ( \beta_{20} )</td>
<td>-6.59</td>
</tr>
<tr>
<td>need ( \beta_{21} )</td>
<td>-2.12</td>
</tr>
<tr>
<td>anytxt ( \beta_{22} )</td>
<td>1.13</td>
</tr>
<tr>
<td>need ( \beta_{23} )</td>
<td>3.37</td>
</tr>
<tr>
<td>need ( \beta_{24} )</td>
<td>4.26</td>
</tr>
<tr>
<td>need ( \beta_{25} )</td>
<td>5.79</td>
</tr>
<tr>
<td>need ( \beta_{26} )</td>
<td>-0.47</td>
</tr>
<tr>
<td>need ( \beta_{27} )</td>
<td>-12.15</td>
</tr>
</tbody>
</table>