Structural Nested Mean Models for Assessing Time-Varying Causal Effect Moderation

Daniel Almirall\textsuperscript{1} Thomas R. Ten Have\textsuperscript{2} Susan A. Murphy\textsuperscript{3}

\textsuperscript{1}Health Services Research in Primary Care, Durham VA MC
\textsuperscript{1}Biostatistics & Bioinformatics Department, Duke Univ MC
\textsuperscript{2}Clinical Epi & Biostatistics, Univ of Pennsylvania Medicine
\textsuperscript{3}Statistics Department & ISR, Univ of Michigan

May 21, 2009
2009 Atlantic Causal Modeling Conference
Philadelphia, Pennsylvania
**Contents**

1. Warm-up: Suppose we want $A \rightarrow Y$. 4

2. Effect Moderation in One Time Point 7

3. Mean Model in One Time Point 9

4. Time-Varying Effect Moderation 10

5. Robins’ Structural Nested Mean Model 13

6. Estimation in Time-Varying Setting 15
Contents

7 Sequential Ignorability Given $\tilde{S}_K$ 21
8 Application of the SNMM 22
9 Future Work 25
10 Extra Slides 34
1 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 covariates $S$?
Suppose we want the effect of $A$ on $Y$. Why condition on (adjust for) pre-treatment (or pre-exposure) variables $S$?

1. **Confounder**: $S$ 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 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.
3 Mean Model in One Time Point

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

\[
E(Y(a) \mid S = s) = E(Y(0) \mid S = 0)
\]

\[
+ \left( E(Y(0) \mid S = s) - E(Y(0) \mid S = 0) \right)
\]

\[
+ E(Y(a) - Y(0) \mid S = s)
\]

\[
= \eta_0 + \phi(s) + \mu(s, a)
\]

\[e.g.\]

\[
\eta_0 + \eta_1 s + \beta_1 a + \beta_2 a s.
\]

This is precisely what I would do, too.
4 Time-Varying Effect Moderation

The data structure in the time-varying setting is:

\[
S_1 \quad a_1 \quad S_2(a_1) \quad a_2 \quad Y(a_1, a_2)
\]

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

Ex: What is the effect of switching off treatment for depression early versus later, as a function of time-varying suicidal ideation?
Formal Definition of Time-Varying Causal Effects

Conditional Intermediate Causal Effect at $t = 2$:

$$\mu_2(\bar{s}_2, \bar{a}_2) = E[Y(a_1, a_2) - Y(a_1, 0) \mid S_1 = s_1, S_2(a_1) = s_2]$$

Conditional Intermediate Causal Effect at $t = 1$:

$$\mu_1(s_1, a_1) = E[Y(a_1, 0) - Y(0, 0) \mid S_1 = s_1]$$
Formal Definition of Time-Varying Causal Effects

Conditional Intermediate Causal Effect at \( t = 2 \):

\[
\mu_2(\bar{s}_2, \bar{a}_2) = E[Y(a_1, a_2) - Y(a_1, 0) \mid S_1 = s_1, S_2(a_1) = s_2]
\]

Conditional Intermediate Causal Effect at \( t = 1 \):

\[
\mu_1(s_1, a_1) = E[Y(a_1, 0) - Y(0, 0) \mid S_1 = s_1]
\]
Robins’ Structural Nested Mean Model

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

$$E[Y(a_1, a_2) \mid S_1, S_2(a_1)] = E[Y(0, 0)] + \left\{ E[Y(0, 0) \mid S_1] - E[Y(0, 0)] \right\}$$

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

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

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

$$= \mu_0 + \epsilon_1(s_1) + \mu_1(s_1, a_1) + \epsilon_2(\bar{s}_2, a_1) + \mu_2(\bar{s}_2, \bar{a}_2)$$

$$\text{e.g.} \quad \mu_0 + \epsilon_1(s_1) + \beta_{10}a_1 + \beta_{11}a_1s_1$$

$$+ \epsilon_2(\bar{s}_2, a_1) + \beta_{20}a_2 + \beta_{21}a_2s_1 + \beta_{22}a_2s_2$$
Constraints on the Causal and Nuisance Portions

\[ E[Y(a_1, a_2) \mid \bar{S}_2(a_1) = \bar{s}_2] = \mu_0 + \epsilon_1(s_1) + \mu_1(s_1, a_1) \]
\[ + \epsilon_2(\bar{s}_2, a_1) + \mu_2(\bar{s}_2, \bar{a}_2), \]

where

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

The \( \epsilon_t \)'s make the SNMM a non-standard regression model.
Recall that parametric models for our causal estimands $\mu_1$ and $\mu_2$ are based on the set of parameters $\beta = (\beta_1', \beta_2')'$.

We considered two estimators for $\beta$:

1. Proposed 2-Stage Regression Estimator
2. Robins’ Semi-parametric G-Estimator

In order to make causal inferences, both estimators rely on Robins’ **Sequential Ignorability (or Sequential Randomization) Assumption**. We discuss the two estimators in turn, but first ...
So what’s wrong with the Traditional Estimator?

An Example of the Traditional Estimator: Apply OLS with

\[
E(Y \mid \tilde{S}_2 = \tilde{s}_2, \tilde{A}_2 = \tilde{a}_2) = \beta_0^* + \eta_1s_1 + \beta_1^*a_1 + \beta_2^*a_1s_1 \\
+ \eta_2s_2 + \beta_3^*a_2 + \beta_4^*a_2s_1 + \beta_5^*a_2s_2
\]

- Possibly incorrectly specified nuisance functions.
- Two problems arise with the interpretation of \( \beta_1^* \) and \( \beta_2^* \) (i.e., parameters thought to represent \( \mu_1 \)) when using the traditional regression estimator. We describe them next.
- These problems may occur even in the absence of time-varying confounders (that is, even under Sequential Ignorability) ...
First problem with the Traditional Approach

Wrong Effect

Baseline ——— 4-month Visit ——— 8-month Visit

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

The term \( \beta_1^* a_1 + \beta_2^* a_1 s_1 \) 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

Baseline ————————— 4-month Visit ————————— 8-month Visit

$V_0$

$a_1$

$S_1$

$S_2(a_1)$

Set
$a_2 = 0$

$Y(a_1, 0)$

This is also known as “Berkson’s paradox”; and is related to Judea Pearl’s back-door criterion.
Proposed 2-Stage Regression Estimator

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

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

we use the following

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

We call it “2-Stage” because first we estimate \(E(S_2 \mid A_1, S_1)\), then use the residual \(s_2 - E(S_2 \mid A_1, S_1)\) in a second regression to get \(\beta\)’s. Use sandwich/robust SEs for inference (p-vals, CIs, etc.).
**Existing Semi-parametric G-Estimator**

**Recall** our SNMM:

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

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.
7 Sequential Ignorability Given $\bar{S}_K$

Or the absence of confounders (known, unknown, measured, or unmeasured) other than $\bar{S}_t$. Formally, for each $t = 1, 2, \ldots, K$,

$$A_t \text{ is independent of } \{Y(\bar{a}_K)\} \text{ given } (S_1, A_1, S_2, A_2, \ldots, S_t)$$
8 Application of the SNMM

\( n = 277 \) geriatric primary care patients from PROSPECT Study

\( K = 3 \): visits to clinic at baseline, 4-, 8-, and 12-months

Data structure is \( \{S_1, A_1, S_2, A_2, S_3, A_3, Y\} \)

\( S_{t1} = \) suicidal ideation, \( S_{t2} = \) intermediate depression,

\( A_t = \) adherence, \( Y = \) 12-month (end-of-study) depression

Adherence is defined as ever meeting with a health specialist

Monotonic adherence pattern: \((0, 0, 0), (1, 0, 0), (1, 1, 0), (1, 1, 1)\)

Sequential Ignorability given \( \bar{S}_3 \) is very likely violated
Models Used in the Application

Causal effects: (expect $\beta_{t0} < 0$ and $\beta_{t1} > 0$)

1. $\mu_1(S_1, a_1) = a_1 (\beta_{10} + \beta_{11}SSI_1 + \beta_{11}HAMD_1)$,
2. $\mu_2(\bar{S}_2, \bar{a}_2) = a_2 (\beta_{20} + \beta_{21}SSI_2 + \beta_{21}HAMD_2)$, and
3. $\mu_3(\bar{S}_3, \bar{a}_3) = a_3 (\beta_{30} + \beta_{31}SSI_3 + \beta_{31}HAMD_3)$.

Nuisance functions used in the 2-Stage Estimator:

1. For $t = 1, 2$, for both SSI and HAMD, we used the most parsimonious model for the $\epsilon_t$’s.
2. $\epsilon_{31}(S_{31}, \bar{S}_2, \bar{A}_2) = \left(\eta_{3,1,1} + \eta_{3,1,2}SSI_1 + \eta_{3,1,3}HAMD_1 + \eta_{3,1,4}SSI_2 + \eta_{3,1,5}HAMD_2 + \eta_{3,1,6}SSI_1 HAMD_2\right) \delta_{31}(S_{31}, \bar{S}_2, \bar{A}_2; \gamma_{31})$
3. $\epsilon_{32}(S_{32}, \bar{S}_2, \bar{A}_2) = (\eta_{3,2,1} + \eta_{3,2,2}SSI_2) \delta_{32}(S_{32}, \bar{S}_2, \bar{A}_2; \gamma_{32})$
### Results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>2-Stage Estimator</th>
<th>Robins’ G-Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>SE</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int $\beta_{10}$</td>
<td>$-0.20$</td>
<td>0.37</td>
</tr>
<tr>
<td>SSI $\beta_{11}$</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>HAMD $\beta_{12}$</td>
<td>$-0.24$</td>
<td>0.35</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int $\beta_{20}$</td>
<td>0.40</td>
<td>0.31</td>
</tr>
<tr>
<td>SSI $\beta_{21}$</td>
<td>0.22</td>
<td>0.27</td>
</tr>
<tr>
<td>HAMD $\beta_{22}$</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int $\beta_{30}$</td>
<td>$-0.27$</td>
<td>0.26</td>
</tr>
<tr>
<td>SSI $\beta_{31}$</td>
<td>$-0.12$</td>
<td>0.26</td>
</tr>
<tr>
<td>HAMD $\beta_{32}$</td>
<td>$-0.02$</td>
<td>0.19</td>
</tr>
</tbody>
</table>
9 Future Work
(a) Handling Time-varying Confounders in the SNMM

How do we handle time-varying covariates $X_t$ that are possible confounders, but are not moderators of interest?

Note: In practice, the set $X_t$ is much larger than $S_t$. 
Proposed Solution: Propensity Score Weighting

We can use IPT-Weighted versions of the proposed 2-Stage Estimator (or G-Estimator):
(b) Multi-Component MSMs and SNMMs

• Thus far, we have had 1 treatment/exposure variable per time point; and so, we have considered potential outcomes
  \[ Y(a_1, a_2) \]

• What if we had multi-component treatments/exposures at each time point?

• That is, what if \( a_t \) is now a vector of \( J \) treatment/exposure variables?

• \( Y(a_1, a_2); \) where \( a_t = (a_{t1}, \ldots, a_{tJ}) \)
Multi-Component MSMs and SNMMs: Example

\[ Y_{18}(a_1, a_2) \]
Multi-Component MSMs and SNMMs: Example

Weight at 18 Months

Baseline Macronutrient

6–Month Macronutrient
(c) SNMMs With Longitudinal Outcomes
(d) Informing RCTs for Developing Adaptive Treatment Strategies

Clinicians are becoming more interested in developing adaptive treatment strategies (ATSs), which are sequences of individually tailored decision rules that specify whether, how, and when to alter the intensity, type, or delivery of treatment at critical decision points in the medical care process.

Specialized trials are available that can be used to inform the development of ATSs.

How can/should MSMs and SNMMs be used to inform the design of these specialized trials?
Thank you!
More Questions?
Mean Model in One Time Point

Decompose the conditional mean $E(Y(a) | S)$ as follows:

$$E(Y(a) | S = s) = E(Y(0) | S = 0)$$

$$+ \left( E(Y(0) | S = s) - E(Y(0) | S = 0) \right)$$

$$+ E(Y(a) - Y(0) | S = s)$$

$$= \eta_0 + \phi(s) + \mu(s, a).$$

The intercept $\eta_0$ and the function $\phi(s)$ are non-causal. They are known as nuisance functions. $\phi(s)$ is the “associational main effect” of $S$ on $Y(0)$. 
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 \( V (\supset S) \).

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

Regress \( Y \sim [1, G, A \times H] \) to get \((\hat{\eta}, \hat{\beta})\), 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

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

\( \hat{\beta} \) 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*  
Ex: \[ H = (1, S) \]

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

*just discussed †need \(Pr(A = 1 \mid S)\) ‡need \(Pr(A = 1 \mid V)\)
Application in One Time Point

$n = 1984$ adolescents that are substance abusers.

Motivation: American Society of Addiction Medicine (ASAM) Patient Placement Criteria (PPC)

Two levels of care (LOC):

$A = 0$ outpatient,

$A = 1$ residential

Illustrate methodology using:

$S = \text{Needle Frequency Index (hi is bad)}$

$Y = \text{Substance Frequency Index (hi is bad)}$

$V = 86$ covariates to adjust for (possible confounders)
Covariate Balance Before-After Weighting

Standardized Differences Before–After

Unweighted
B = 0.3476

Weighted
B = 0.0688

B = Average Absolute Standardized Difference

P–values for No Difference Before–After

Unweighted
N = 86

Weighted
N = 11

N = Number of P–values < 0.10
Effect Moderation by $S = \text{Needle Frequency Index}$

**Unweighted Regression**

- Needle Frequency Index (nfip)
- Substance Frequency Index (sfi7pfu)
- Outpatient LOC
- Inpatient/Residential LOC

**IPT Weighted Regression**

- Needle Frequency Index (nfip)
- Outpatient LOC
- Inpatient/Residential LOC
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:

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

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

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

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

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

$$= \mu_0 + \epsilon_1(s_1) + \mu_1(s_1, a_1) + \epsilon_2(\bar{s}_2, a_1) + \mu_2(\bar{s}_2, \bar{a}_2)$$
Parameterizing the Nuisance Functions

So we must parameterize the nuisance functions correctly.

**Recall** the constraints on the nuisance functions:

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

**Example** parameterizations for the nuisance functions:

$$\epsilon_1(s_1) \overset{\text{say}}{=} \eta_{1,1}(s_1 - E(S_1))$$

$$\epsilon_2(\bar{s}_2, a_1) \overset{\text{say}}{=} \eta_{2,1}(s_2 - E(S_2(a_1) \mid S_1 = s_1))$$
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 \( Y \sim [1, G_1^*, A_1 H_1, G_2^*, A_2 H_2] \)
**Robins’ Semi-parametric G-Estimator**

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

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

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

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

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

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

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

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

\[
\frac{\text{MSE (Robins' G-Estimator)}}{\text{MSE (2-Stage Estimator)}}
\]

Diagram shows relative mean squared error versus level of mis-specification for different functions and parameters.
The Generative Model in Simulations

\( nits = 1000 \) simulated data sets each of size \( n = 500 \)

1. \( \delta_1 \sim \hat{\text{res}}_1 \). 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. \( \delta_2 \sim \text{N}_n(0, \text{sd} = 0.75) \). 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 $\mathcal{N}(0, \text{sd} = 0.7)$,

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**:

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

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</td>
<td>Effect Size‡</td>
</tr>
<tr>
<td></td>
<td>Size† Sign</td>
<td></td>
</tr>
</tbody>
</table>

### $A_1 = HSANY.4$

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Effect Size†</th>
<th>Sign</th>
<th>Effect Size‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAMDA.0</td>
<td>0.77</td>
<td>+</td>
<td>0.18</td>
</tr>
<tr>
<td>RE.0</td>
<td>0.64</td>
<td>-</td>
<td>-0.05</td>
</tr>
<tr>
<td>RE16N.0</td>
<td>0.66</td>
<td>-</td>
<td>-0.05</td>
</tr>
<tr>
<td>MCS.0</td>
<td>0.53</td>
<td>-</td>
<td>-0.08</td>
</tr>
<tr>
<td>MMSE2.0</td>
<td>0.45</td>
<td>+</td>
<td>0.22</td>
</tr>
<tr>
<td>SSI.0</td>
<td>0.40</td>
<td>-</td>
<td>-0.29</td>
</tr>
</tbody>
</table>

### $A_2 = HSANY.8$

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Effect Size†</th>
<th>Sign</th>
<th>Effect Size‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD.4</td>
<td>0.80</td>
<td>+</td>
<td>0.27</td>
</tr>
<tr>
<td>DYSTH.0</td>
<td>0.69</td>
<td>-</td>
<td>-0.66</td>
</tr>
<tr>
<td>OPS.0</td>
<td>0.49</td>
<td>-</td>
<td>-0.07</td>
</tr>
<tr>
<td>HAMDA.4</td>
<td>0.55</td>
<td>-</td>
<td>0.07</td>
</tr>
<tr>
<td>CAD.0</td>
<td>0.51</td>
<td>+</td>
<td>0.47</td>
</tr>
<tr>
<td>POSAF.0</td>
<td>0.53</td>
<td>+</td>
<td>-0.28</td>
</tr>
</tbody>
</table>

### $A_3 = HSANY.12$

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Absolute Effect Size†</th>
<th>Sign</th>
<th>Effect Size‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD.8</td>
<td>0.90</td>
<td>+</td>
<td>0.37</td>
</tr>
<tr>
<td>WHITE1</td>
<td>0.71</td>
<td>+</td>
<td>0.08</td>
</tr>
<tr>
<td>RP16N.0</td>
<td>0.60</td>
<td>+</td>
<td>0.28</td>
</tr>
</tbody>
</table>