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

Daniel Almirall¹, Daniel F. McCaffrey², Beth Ann Griffin², Rajeev Ramchand², Susan A. Murphy¹

¹Univ of Michigan, Institute for Social Research
²RAND, 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, $a_1 = \text{yes/no}$
- $S_1(a_1)$: Severity @ 0-3mo
- $a_2$: 3-6mo treatment; binary, $a_2 = \text{yes/no}$
- $S_2(a_1, a_2)$: Severity @ 3-6mo
- $a_3$: 6-9mo treatment; binary, $a_3 = \text{yes/no}$
- $Y(a_1, a_2, a_3)$: 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 $S_0$?

$$\mu_1 = E[Y(1,0,0) - Y(0,0,0) | S_0 = s_0]$$
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 $(S_0, S_1)$?

$$\mu_2 = E[Y(1, 1, 0) - Y(1, 0, 0) \mid S_0 = s_0, S_1(1) = s_1]$$
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 $(S_0, S_1, S_2)$?

$$\mu_3 = E[Y(1, 1, 1) - Y(1, 1, 0) \mid \bar{S}_2(1, 1) = \bar{s}_2]$$
4 Robins’ Structural Nested Mean Model
decomposes $E(Y \mid \cdot)$ into nuisance and causal parts:

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

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

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

$$+ \left\{ E[Y(a_1, 0) \mid S_1(a_1)] - E[Y(a_1, 0) \mid S_0] \right\}$$

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

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

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

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

- Two problems arise 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 $\beta_1^*a_1 + \beta_2^*a_1s_0$ 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 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 an 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

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

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 “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 Regression with Residuals Estimator

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

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...
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 \( (\tilde{S}_t, \tilde{X}_t) \), (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 ⊂ SAMHSA)
- GAIN: structured clinical interview; over 100 scales/indices
- \( n = 2870 \) adolescents; data every 3 months for 1 year
- \( \{(S_0, X_0), A_1, (S_1, X_1), A_2, (S_2, X_2), A_3, Y\} \)
- \( S_0 = \text{hx controlled environment, age} \)
- \( S_t = \text{substance frequency scale at intake, 0-3, 3-6} \)
- \( X_t = \text{measured time-varying confounders at intake, 0-3, 3-6} \)
- \( A_t = \text{none (0) vs some txt (1=outpt, inpt, or both)} \)
- \( Y = \text{substance frequency scale at 9-12mo} \)
The weights did a good job adjusting for $X_t$. 

<table>
<thead>
<tr>
<th></th>
<th>t = 1</th>
<th>t = 2</th>
<th>t = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unweighted</td>
<td>$B = 0.161$</td>
<td>$B = 0.155$</td>
<td>$B = 0.198$</td>
</tr>
<tr>
<td>Weighted</td>
<td>$B = 0.041$</td>
<td>$B = 0.024$</td>
<td>$B = 0.037$</td>
</tr>
</tbody>
</table>

- Small
- Medium
- Large

Effect sizes displayed in the plots indicate the magnitude of the effect after weighting.
## Data Analysis

### EDA

#### $\mu_1$ = Distal effects of initial treatment, given sfs8p0, age, and baseline CE status

<table>
<thead>
<tr>
<th></th>
<th>No CE</th>
<th>No CE</th>
<th>Yes CE</th>
<th>Yes CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 16</td>
<td>16 or older</td>
<td>Under 16</td>
<td>16 or older</td>
<td></td>
</tr>
</tbody>
</table>

#### $\mu_2$ = Medial effects of additional treatment, given sfs8p3, age, and baseline CE status

<table>
<thead>
<tr>
<th></th>
<th>No CE</th>
<th>No CE</th>
<th>Yes CE</th>
<th>Yes CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 16</td>
<td>16 or older</td>
<td>Under 16</td>
<td>16 or older</td>
<td></td>
</tr>
</tbody>
</table>

#### $\mu_3$ = Proximal effects of additional treatment, given sfs8p6, age, and baseline CE status

<table>
<thead>
<tr>
<th></th>
<th>No CE</th>
<th>No CE</th>
<th>Yes CE</th>
<th>Yes CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 16</td>
<td>16 or older</td>
<td>Under 16</td>
<td>16 or older</td>
<td></td>
</tr>
</tbody>
</table>

The figure shows scatter plots for each effect, with different markers and colors representing different categories (No CE, Yes CE) and age groups (Under 16, 16 or older).
# Effect Estimates from SNMM, using RR+IPTW

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Subgroup</th>
<th>Est.</th>
<th>Eff.Sz.</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$: Distal</td>
<td>(1, 0, 0) vs (0, 0, 0) no intake sevrt, &lt; 16yrs</td>
<td>$-0.004$</td>
<td>$-0.03$</td>
<td>$0.74$</td>
</tr>
<tr>
<td></td>
<td>(1, 0, 0) vs (0, 0, 0) hi intake sevrt, $\geq$ 16yrs</td>
<td>$0.033$</td>
<td>$0.25$</td>
<td>$0.08$</td>
</tr>
<tr>
<td>$\mu_2$: Medial</td>
<td>(1, 1, 0) vs (1, 0, 0) no 0-3 severity</td>
<td>$-0.008$</td>
<td>$-0.06$</td>
<td>$0.42$</td>
</tr>
<tr>
<td></td>
<td>(1, 1, 0) vs (1, 0, 0) hi 0-3 severity, yes ce</td>
<td>$-0.048$</td>
<td>$-0.36$</td>
<td>$0.21$</td>
</tr>
<tr>
<td></td>
<td>(1, 1, 0) vs (1, 0, 0) hi 0-3 severity, no ce</td>
<td>$0.021$</td>
<td>$0.16$</td>
<td>$0.66$</td>
</tr>
<tr>
<td>$\mu_3$: Proximal</td>
<td>(1, 1, 1) vs (1, 1, 0) no 6-9 severity</td>
<td>$-0.006$</td>
<td>$-0.04$</td>
<td>$0.59$</td>
</tr>
<tr>
<td></td>
<td>(1, 1, 1) vs (1, 1, 0) hi 6-9 severity</td>
<td>$-0.168$</td>
<td>$-1.27$</td>
<td>$&lt; 0.01$</td>
</tr>
<tr>
<td></td>
<td>(., ., 1) vs (., ., 0) no 6-9 severity</td>
<td>$0.026$</td>
<td>$0.19$</td>
<td>$0.12$</td>
</tr>
<tr>
<td></td>
<td>(., ., 1) vs (., ., 0) hi 6-9 severity</td>
<td>$-0.165$</td>
<td>$-1.24$</td>
<td>$&lt; 0.01$</td>
</tr>
</tbody>
</table>
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, Griffin, Ramchand, Murphy (to submit).
Warm-up: Suppose we want $A \rightarrow Y$.

Examples

$S = \text{pre-}A \text{ covt} \quad A = \text{txt/expsr} \quad Y = \text{outcome}$

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:

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

*e.g.* $\eta_0 + \eta_1(s - E(S)) + \beta_1 a + \beta_2 a s$.

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 $\beta$ 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 $H = (1, S)$:

$$E(Y(a) \mid S = s) = \eta_0 + \eta_1s + a \times (\beta_1 + \beta_2s).$$

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:** $H = (1,S) \Rightarrow aH\beta = a(\beta_1 + \beta_2s)$.

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 \( Y \sim \begin{bmatrix} 1, G, A \times H \end{bmatrix} \) 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 \( V \supseteq S \), 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:**  
\[ E(Y(a) | S) = \eta_0 + \phi(S) + aH\beta \]

**Model B:**  
\[ E(Y(a) | V) = \phi_0 + \phi(V) + aH\beta \]

\( H \) 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>( \phi ) Is</strong></td>
<td><strong>No Confnders</strong></td>
<td><strong>S is Sole Confnder</strong></td>
</tr>
<tr>
<td><strong>Known</strong></td>
<td>OLS*</td>
<td>OLS*</td>
</tr>
<tr>
<td><strong>( \phi ) Is</strong></td>
<td><strong>OLS (if ( S \perp A ))</strong></td>
<td><em><em>E-estimtr</em>†</em>*</td>
</tr>
<tr>
<td><strong>Not Known</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

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:

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

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

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

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

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

$$= \mu_0 + \epsilon_1(s_1) + \mu_1(s_1, a_1) + \epsilon_2(\bar{s}_2, a_1) + \mu_2(\bar{s}_2, \bar{a}_2)$$
SNMM Property I: $\mu_t = 0$ when $a_t = 0$.

\[
E[Y(a_1, a_2) \mid S_0, S_1(a_1)]
\]
\[
= E[Y(0, 0)] + \left\{ E[Y(0, 0) \mid S_0] - E[Y(0, 0)] \right\}
\]
\[
+ \left\{ E[Y(a_1, 0) - Y(0, 0) \mid S_0] \right\}
\]
\[
+ \left\{ E[Y(a_1, 0) \mid \bar{S}_1(a_1)] - E[Y(a_1, 0) \mid S_0] \right\}
\]
\[
+ \left\{ E[Y(a_1, a_2) - Y(a_1, 0) \mid \bar{S}_1(a_1)] \right\}
\]
\[
= \mu_0 + \epsilon_1(s_0) + \mu_1(s_0, a_1) + \epsilon_2(\bar{s}_1, a_1) + \mu_2(\bar{s}_1, \bar{a}_2)
\]

\checkmark \mu_1(s_0, 0) = 0 \quad \text{Ex Model: } a_1(\beta_{10} + \beta_{11}s_0)

\checkmark \mu_2(\bar{s}_1, a_2, 0) = 0 \quad \text{Ex Model: } a_2(\beta_{20} + \beta_{21}s_1)
SNMM Property II: $\epsilon_t$’s are conditional mean zero.

\[
E[Y(a_1, a_2) \mid S_0, S_1(a_1)] \\
= E[Y(0, 0)] + \left\{ E[Y(0, 0) \mid S_0] - E[Y(0, 0)] \right\} \\
+ \left\{ E[Y(a_1, 0) - Y(0, 0) \mid S_0] \right\} \\
+ \left\{ E[Y(a_1, 0) \mid \bar{S}_1(a_1)] - E[Y(a_1, 0) \mid S_0] \right\} \\
+ \left\{ E[Y(a_1, a_2) - Y(a_1, 0) \mid \bar{S}_1(a_1)] \right\} \\
= \mu_0 + \epsilon_1(s_0) + \mu_1(s_0, a_1) + \epsilon_2(\bar{s}_1, a_1) + \mu_2(\bar{s}_1, \bar{a}_2)
\]

- $E_{S_1\mid S_0}[\epsilon_2(\bar{s}_1, a_1) \mid S_0 = s_0] = 0$, and $E_{S_0}[\epsilon_1(s_0)] = 0$
- The $\epsilon_t$’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:

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

- 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 $\beta_1^* a_1 + \beta_2^* a_1 s_0$ 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”
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, Vanderweele, & 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 \( E[Y(a_1, a_2) | \bar{S}_2(a_1) = \bar{s}_2] = \mu_2(\bar{s}_2, \bar{a}_2; \beta_2) \)
\[+ \epsilon_2(\bar{s}_2, a_1; \eta_2, \gamma_2) + \mu_1(s_1, a_1; \beta_1) + \epsilon_1(s_1; \eta_1, \gamma_1) + \mu_0.\]

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
\( m_2(s_1, a_1; \gamma_2) = E(S_2(a_1) | S_1 = s_1) \), 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 \( Y \sim [1, G^*_1, A_1 H_1, G^*_2, A_2 H_2] \)
## Data Descriptives: Treatment Trajectories

<table>
<thead>
<tr>
<th>Treatment ((A_1, A_2, A_3))</th>
<th>Frequency</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0,0)</td>
<td>310</td>
<td>11%</td>
</tr>
<tr>
<td>(0,1,0)</td>
<td>56</td>
<td>2%</td>
</tr>
<tr>
<td>(1,0,0)</td>
<td>1184</td>
<td>41%</td>
</tr>
<tr>
<td>(1,1,0)</td>
<td>555</td>
<td>19%</td>
</tr>
<tr>
<td>(0,0,1)</td>
<td>56</td>
<td>2%</td>
</tr>
<tr>
<td>(0,1,1)</td>
<td>56</td>
<td>2%</td>
</tr>
<tr>
<td>(1,0,1)</td>
<td>153</td>
<td>5%</td>
</tr>
<tr>
<td>(1,1,1)</td>
<td>499</td>
<td>17%</td>
</tr>
</tbody>
</table>
# Data Descriptives: Moderators and Outcomes

<table>
<thead>
<tr>
<th>Moderators</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sfs8p0</td>
<td>0.18</td>
<td>0.18</td>
<td>(0, 0.89)</td>
</tr>
<tr>
<td>b2a</td>
<td>15.98</td>
<td>1.4</td>
<td>(12, 25)</td>
</tr>
<tr>
<td>maxce0</td>
<td>13.95</td>
<td>24.6</td>
<td>(0, 90)</td>
</tr>
<tr>
<td>$S_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sfs8p3</td>
<td>0.07</td>
<td>0.11</td>
<td>(0, 0.67)</td>
</tr>
<tr>
<td>$S_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sfs8p6</td>
<td>0.08</td>
<td>0.13</td>
<td>(0, 0.73)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sfs8p12</td>
<td>0.09</td>
<td>0.13</td>
<td>(0, 0.78)</td>
</tr>
</tbody>
</table>
How did we choose our weights?

Selecting Denominator Model

$t = 1$

$t = 2$

$t = 3$
How did the weights do?

<table>
<thead>
<tr>
<th>$t$</th>
<th>No.</th>
<th>Denom Pr. (min, max)</th>
<th>Denominator weights (min, max)</th>
<th>$ESS$</th>
<th>$J$</th>
<th>$B$</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>(0.20, 0.98)</td>
<td>(1.02, 32.83)</td>
<td>1214.8</td>
<td>46</td>
<td>0.041</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>(0.03, 0.95)</td>
<td>(1.03, 15.43)</td>
<td>1867.7</td>
<td>86</td>
<td>0.024</td>
<td>0.13</td>
</tr>
<tr>
<td>3</td>
<td>76</td>
<td>(0.01, 0.97)</td>
<td>(1.01, 37.93)</td>
<td>1057.0</td>
<td>126</td>
<td>0.037</td>
<td>0.16</td>
</tr>
</tbody>
</table>
**Existing Semi-parametric G-Estimator**

Recall our SNMM:

\[
E[Y(a_1, a_2) \mid S_1, S_2(a_1)] = \mu_0 + \epsilon_1(s_1) + \beta_{10}a_1 + \beta_{11}a_1s_1 \\
+ \epsilon_2(\bar{s}_2, a_1) + \beta_{20}a_2 + \beta_{21}a_2s_1 + \beta_{22}a_2s_2
\]

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:

\[
0 = \mathbb{P}_n \left\{ \left( Y - A_2 H_2 \beta_2 - b_2(\bar{S}_2, A_1) \right) \times \left( A_2 - p_2(\bar{S}_2, A_1) \right) \times \begin{pmatrix} 0 \\ H_2' \end{pmatrix} ^t \right. \\
+ \left. \left( Y - A_2 H_2 \beta_2 - A_1 H_1 \beta_1 - b_1(S_1) \right) \times \left( A_1 - p_1(S_1) \right) \times \begin{pmatrix} H_1' \\ \Delta'(H_1) \end{pmatrix} ^t \right\}
\]

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

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

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

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

\[
p_1(S_1) = Pr \left[ A_1 = 1 \big| S_2 \right]
\]
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 $S^* = S \times N(1, sd = \nu)$

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 $Y$. 

Scaled Root Mean Squared Difference (SRMSD)
Results

Relative Mean Squared Error for $\beta$: $\frac{\text{MSE(Robins' G-Estimator)}}{\text{MSE(2-Stage Estimator)}}$

- $a_1$
- $a_2: I((s_1 + s_2)/2)$
- $a_3: I((s_1 + s_2 + s_3)/3)$

$\nu$ SRMSD

$\nu$
The Generative Model in Simulations

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

1. $\delta_1 \sim \hat{res}_1$. 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. $\delta_2 \sim N_n(0, \text{sd} = 0.75)$. 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. $\delta_3 \sim N_n(0, sd = 0.51)$. 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**:

\[
SRMSD(\nu) = \sqrt{\frac{E \left( \sum_{t=3}^{K} \epsilon_t^{\text{TRUE}} - \sum_{t=1}^{K} \epsilon_t'(\hat{\eta}, \hat{\gamma}) \right)^2}{\text{Var}(Y)}},
\]

where \( \nu \) corresponds to a mis-specified 2-Stage Regression Estimator.

The expectation \( E \) and variance \( \text{Var} \) 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>A₁ = HSANY.4</td>
<td></td>
<td></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>
<tr>
<td>A₂ = HSANY.8</td>
<td></td>
<td></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>
<tr>
<td>A₃ = HSANY.12</td>
<td></td>
<td></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>
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

\[
\begin{align*}
S_1 & \quad A_1 & \quad S_2 & \quad A_2 & \quad Y = ERS \\
\text{need0} & \quad \text{reported no treatment (0) versus some treatment (1= outpatient, inpatient, or both) at 3-months} & \quad \text{need6} & \quad \text{treatment indicator at 9-months} & \quad \text{ERS} = \text{Environmental Risk Scale at 12-months}
\end{align*}
\]
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

A1 = 3-month Treatment

High Baseline Severity

Some

35.3

None

43.4

Low Baseline Severity

Some

35.1

None

37.5

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

\[
\Pr(\text{need}_6 = 1 \mid \text{need}_0, \text{anytxt}_3) = \gamma_{20} + \gamma_{21} \text{need}_0 + \gamma_{22} \text{anytxt}_3 + \gamma_{23} \text{need}_0 \text{anytxt}_3.
\]
Estimates of the SNMM Using the 2-Stage Regression Estimator

<table>
<thead>
<tr>
<th>Parameters</th>
<th>2-Stage Estimator</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>SE</td>
<td>P-val</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>$\beta_{00}$</td>
<td>39.76</td>
<td>1.36</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>$\beta_{10}$</td>
<td>1.36</td>
<td>1.5</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>$\beta_{11}$</td>
<td>6.89</td>
<td>3.73</td>
</tr>
<tr>
<td></td>
<td>$\beta_{20}$</td>
<td>−6.59</td>
<td>4.04</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>2.12</td>
<td>6.17</td>
</tr>
<tr>
<td></td>
<td>$\beta_{22}$</td>
<td>1.13</td>
<td>4.20</td>
</tr>
<tr>
<td></td>
<td>$\beta_{23}$</td>
<td>3.37</td>
<td>17.53</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>$\beta_{24}$</td>
<td>4.26</td>
<td>6.52</td>
</tr>
<tr>
<td></td>
<td>$\beta_{25}$</td>
<td>5.79</td>
<td>20.49</td>
</tr>
<tr>
<td></td>
<td>$\beta_{26}$</td>
<td>−0.47</td>
<td>17.98</td>
</tr>
<tr>
<td></td>
<td>$\beta_{27}$</td>
<td>−12.15</td>
<td>21.3</td>
</tr>
</tbody>
</table>
### RR+IPTW vs RR vs TRAD

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Subgroup</th>
<th>RR+IPTW</th>
<th>RR</th>
<th>TRAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$: Distal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1$ Distal</td>
<td>(1, 0, 0) vs (0, 0, 0)</td>
<td>no intake severity, &lt; 16yrs</td>
<td>-0.004</td>
<td>-0.016</td>
</tr>
<tr>
<td>$\mu_1$ Distal</td>
<td>(1, 0, 0) vs (0, 0, 0)</td>
<td>hi intake severity, ≥ 16yrs</td>
<td>0.033</td>
<td>0.015</td>
</tr>
<tr>
<td>$\mu_2$: Medial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$ Medial</td>
<td>(1, 1, 0) vs (1, 0, 0)</td>
<td>no 0-3 severity</td>
<td>-0.008</td>
<td>-0.005</td>
</tr>
<tr>
<td>$\mu_2$ Medial</td>
<td>(1, 1, 0) vs (1, 0, 0)</td>
<td>hi 0-3 severity, yes ce</td>
<td>-0.048</td>
<td>-0.067</td>
</tr>
<tr>
<td>$\mu_2$ Medial</td>
<td>(1, 1, 0) vs (1, 0, 0)</td>
<td>hi 0-3 severity, no ce</td>
<td>0.021</td>
<td>-0.037</td>
</tr>
<tr>
<td>$\mu_3$: Proximal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_3$ Proximal</td>
<td>(1, 1, 1) vs (1, 1, 0)</td>
<td>no 3-6 severity</td>
<td>-0.006</td>
<td>-0.012</td>
</tr>
<tr>
<td>$\mu_3$ Proximal</td>
<td>(1, 1, 1) vs (1, 1, 0)</td>
<td>hi 3-6 severity</td>
<td>-0.168</td>
<td>-0.110</td>
</tr>
<tr>
<td>$\mu_3$ Proximal</td>
<td>(., ., 1) vs (., ., 0)</td>
<td>no 3-6 severity</td>
<td>0.026</td>
<td>0.002</td>
</tr>
<tr>
<td>$\mu_3$ Proximal</td>
<td>(., ., 1) vs (., ., 0)</td>
<td>hi 3-6 severity</td>
<td>-0.165</td>
<td>-0.144</td>
</tr>
</tbody>
</table>
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:

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

$$
= \mu_0 + \epsilon_1(s_0) + \mu_1(s_0, a_1) + E_{S_1(a_1) \mid S_0} \left( \epsilon_2(\bar{s}_1, a_1) + \mu_2(\bar{s}_1, \bar{a}_2) \right)
$$

$$
= \mu_0 + \epsilon_1(s_0) + \mu_1(s_0, a_1) + E_{S_1(a_1) \mid S_0} \left( \mu_2(\bar{s}_1, \bar{a}_2) \right)
$$
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 $\mu_t$’s. Think path analysis. But now we have to believe our stage 1 models are causal!

• If effect moderation $\nexists$, estimates for the $\mu_t$’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.