Analysis of Micro-randomized Trials with a Continuous Outcome

Susan A. Murphy and Daniel Almirall

Module 4

May 21, 2018
Goals

• To better understand the primary and secondary aims (scientific questions) in a micro-randomized trial

• To present, illustrate and discuss an easy-to-use data analysis method for addressing such questions
Outline

• Review structure of data arising from MRT
• Review the Primary and Secondary aims in HeartSteps

• Intuition about a new, easy-to-use least squares estimator for analyzing data from a MRT
• Results of the HeartSteps MRT
  • With HeartSteps we use a special case of the estimator

• Details concerning the causal effects
• Details concerning the new, least squares estimator
Outline

• Review structure of data arising from MRT
• Review the Primary and Secondary aims in HeartSteps

• Intuition about a new, easy-to-use least squares estimator for analyzing data from a MRT
• Results of the HeartSteps MRT

• Details concerning the causal effects
• Details concerning the new, least squares estimator
Review: Data collected in a MRT

\( X_1, I_1, A_1, Y_2, \ldots, X_t, I_t, A_t, Y_{t+1}, \ldots, X_T, I_T, A_T, Y_{T+1} \)

- \( t = \) decision point (e.g., decide to send activity suggestion)
- \( X_t = \) observations collected at \( t \) (passive and active)
- \( I_t = 1 \), i.e., available individuals at \( t; = 0 \) otherwise
- \( H_t = \) history of all data up to time \( t \) decision (includes avail.)
- \( A_t = 1 \) if activity suggestion; and \( A_t = 0 \) otherwise
- Known probs. \( \rho_t(1|H_t) = \Pr(A_t = 1|H_t) = 0.4 \) in HeartSteps
- \( Y_{t+1} = \) proximal outcome (e.g., activity over next 30 min)
- \( Z_t = \) subset of \( H_t \) (or summaries of \( t \)) that are expected to be associated with \( Y_{t+1} \) (i.e., explain variance in \( Y_{t+1} \))
Review: The HeartSteps Primary Aim

- $X_t, I_t, A_t, Y_{t+1}, ...$, $X_T, I_T, A_T, Y_{T+1}$
- $H_t, A_t, Y_{t+1}$ at a decision point $t$

- $T = 42$ days * 5 times per day = 210 decision points
- $I_t = 1$, i.e., available individuals at $t$; =0 otherwise
- $A_t = 1$ if tailored activity suggestion provided; $A_t = 0$ o/w
- $Y_{t+1} =$ activity in the next 30 minutes

On average, does the tailored activity suggestion influence step count in the next 30 minutes?
- Does this effect vary over time?
In Module 2, we expressed this effect in terms of potential outcomes, and we discussed how the sequential randomizations allow us to identify the causal effect in HeartSteps in terms of observed data in the way shown above. (For general MRTs, the expression of the causal effect in terms of the observed data is slightly different.) The sequential randomizations are critical to being able to do this.
In Module 2 we wrote this causal effect in terms of potential outcomes and we discussed how the sequential randomizations allow us to write the causal effect this way, i.e., in terms of observed data.
In Module 2 we wrote this causal effect in terms of potential outcomes and we discussed how the sequential randomizations allow us to write the causal effect this way, i.e., in terms of observed data.
In Module 2 we wrote this causal effect in terms of potential outcomes and we discussed how the sequential randomizations allow us to write the causal effect this way, i.e., in terms of observed data.
Anecdote: HeartSteps Design

Originally, we planned for a test statistic that targets the quadratic alternative and randomization probability $\rho_t(1|H_t)=.4$

**But prior to starting the actual MRT** in small usability pilot designed to test the code, we learned that individuals were available at much lower rate than expected, causing us to worry about statistical power for the pre-planned hypothesis test.

As a result, we:

1. Increased randomization probability to $\rho_t(1|H_t)=.6$
2. Decided to test the null assuming a linear (rather than quadratic) alternative so as to save a degree of freedom.
# Outline

- Review structure of data arising from MRT
- Review the Primary and Secondary aims in HeartSteps

- **Intuition about a new, easy-to-use least squares estimator for analyzing data from a MRT**
- Results of the HeartSteps MRT

- Details concerning the causal effects
- Details concerning the new, least squares estimator
What analysis might we plan to do?

On average, does the tailored activity suggestion influence step count in the next 30 minutes? Does this effect vary over time (linearly)?

- $Y_{t+1}$ is subsequent activity over next 30 min.
- $A_t = 1$ if activity suggestion and 0 otherwise
- $d(t) =$ day in study: 0, 1, ..., 41
- $Z_j$ summaries formed from $t$ and past/present observations
- There are two parts to this regression...

We go through intuition that we have gained from analyses in which the treatment does not vary with time. Here the complication is that treatment is time varying. The issue is that both $S_t$ and $Z_t$ may be outcomes of past treatment.

$Z_j$ might include location, time of day, day of week, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time, $j$, so as to allow a more flexible model

$S_j$ might be a vector as well and might include features of time $S_j$ might be the output of a classifier
What analysis might we plan to do?

On average, does the tailored activity suggestion influence step count in the next 30 minutes? Does this effect vary over time (linearly)?

\[ Y_{t+1} \sim \beta_0 A_t + \alpha_1^T Z_t + \beta_1 A_t d(t) \]

- \( Y_{t+1} \) is subsequent activity over next 30 min.
- \( A_t = 1 \) if activity suggestion and 0 otherwise
- \( d(t) = \) day in study: 0, 1, ..., 41
- \( Z_t \) summaries formed from \( t \) and past/present observations
- There are two parts to this regression...

We go through intuition that we have gained from analyses in which the treatment does not vary with time. Here the complication is that treatment is time varying. The issue is that both \( S_t \) and \( Z_t \) may be outcomes of past treatment.

\( Z_j \) might include location, time of day, day of week, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time, \( j \), so as to allow a more flexible model

\( S_j \) might be a vector as well and might include features of time \( S_j \) might be the output of a classifier
These are the interpretations we know hold in cross-sectional analyses and in analyses in which the treatment does not vary with time.
**Causal Effects**

\[ Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t d(t) \]

\( \beta_0 \) is the effect, marginal over all unobserved and all observed variables \( H_\rho \) of the activity suggestion on proximal step count on the first day of the study \( d(t)=0 \)

\( \beta_1 \) describes how the proximal effect changes for each additional day in study (slope)

\( \beta_0 + \beta_1 \times 41 \) is the effect, marginal over all unobserved and all observed variables \( H_\rho \) of the activity suggestion on proximal step count on the last day of the study \( d(t)=41 \)

These are the interpretations we know hold in cross-sectional analyses and in analyses in which the treatment does not vary with time.
By **noise reduction** we mean the part of the regression that seeks to **explain residual variability** in $Y_{t+1}$ and, as a result, leads to tighter confidence intervals (smaller standard errors = more power) for the **causal $\beta$’s**

$$Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t d(t)$$

$\alpha_0 + \alpha_1^T Z_t$ is used to reduce the noise variance in $Y_{t+1}$. $Z_t$ is sometimes called a vector of control variables.

Though interesting, our primary focus is not on the $\alpha$’s.

$j$ is the decision point, I am leaving off subject $i$ subscript.

$X_j$ might include location, time of day, day of week, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time, $j$, so as to allow a more flexible model

$S_{-j}$ might be a vector as well and might include features of time
Methodological Goal

• Develop data analytic methods that are consistent with our scientific understanding of the meaning of the $\beta$'s as a causal effect

not so very independent, “independent” variables
Methodological Goal

- Develop data analytic methods that are consistent with our scientific understanding of the meaning of the $\beta$'s as a causal effect

Methodological Challenges

To develop this methodology we had to address two key challenges:

- Time-varying treatment: $A_i (i=I, \ldots, T)$
- Time-varying covariates: subsets of $H_i$ and $I_i$ that may be affected by prior treatment

not so very independent, “independent” variables
New Method

Simple method for complex data!

- **Easy**: Implement using OTC software
- **Versatile**: Able to examine a variety of causal effects
- **Efficient**: Incorporate variables $Z_t$ to reduce noise but...
- **Robust**: Not require all predictors of $Y_{t+1}$ in $Z_t$, nor require the true or good predictive model for $Y_{t+1}$
  - Critical in mHealth given the high-dimensionality of $X_t$!
  - Also critical given the limited sample size.

Boruvka, Almirall, Witkiewitz, Murphy (2016)
https://arxiv.org/abs/1601.00237
New Method

Simple method for complex data!

- For some MRTs, such as HeartSteps, the method works just as we discussed earlier with no modifications! We do this next.
- Later, we present the method more generally & discuss easy modifications needed to analyze more complex MRTs.

Boruvka, Almirall, Witkiewitz, Murphy (2016)
https://arxiv.org/abs/1601.00237
HeartSteps Primary Aim Analysis

On each of \( n=37 \) participants:

a) Activity suggestion

- Provide a suggestion with probability .6
  - a tailored walking activity suggestion (probability=.3)
  - a tailored sedentary-reducing activity suggestion (probability=.3)
- Do nothing (probability=.4)

- 5 times per day * 42 days = 210 decision points
HeartSteps Primary Aim Analysis

\[ Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0^\prime A_t \]
\[ Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \alpha_2 d_t + \beta_0 A_t + \beta_1 A_t d_t \]

- \( t=1, \ldots, T=210 \)
- \( Y_{t+1} \) = log-transformed step count in the 30 minutes after the \( t^{th} \) decision point,
- \( A_t = 1 \) if an activity suggestion is delivered at the \( t^{th} \) decision point; \( A_t = 0 \), otherwise,
- \( Z_t \) = log-transformed step count in the 30 minutes prior to the \( t^{th} \) decision point,
- \( d_t \) = days in study; takes values in \((0,1,\ldots,41)\)
n=37 participants
Caution:
Remember, that this code only works because in HeartSteps the randomizations prob was constant. Later we will show you the more general estimator.
Caution:
Remember, that this code only works because in HeartSteps the randomizations prob was constant. Later we will show you the more general estimator.
HeartSteps Primary Aim Analysis

\[ Y_{t+1} \sim \alpha_0 + \alpha_t Z_t + \beta_0 A_t, \text{ and} \]

\[ Y_{t+1} \sim \alpha_0 + \alpha_t Z_t + \alpha_2 d_t + \beta_0 A_t + \beta_1 A_t d_t \]

<table>
<thead>
<tr>
<th>Causal Effect Term</th>
<th>Estimate</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{0}^{'}, A_t )</td>
<td>( \hat{\beta}_0 = .13 )</td>
<td>(-0.01, 0.27)</td>
<td>.06</td>
</tr>
<tr>
<td>(Fit 1: average effect)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 A_t + \beta_1 A_t d_t )</td>
<td>( \hat{\beta}_0 = .51 )</td>
<td>(.20, .81)</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>(Fit 2: ( p &lt; .01 ) for the multivariate test)</td>
<td>( \hat{\beta}_1 = -.02 )</td>
<td>(-.03, -.01)</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

.13 translates into a 14% increase over no treatment in step count about 33 steps mean 30-minute step count is 253 steps

.51 translates into a 67% increase over no treatment in step count about 170 steps

Midway through study \( d_t = 20 \) this increase has reduced to 16% increase in step count
.13 translates into a 14% increase over no treatment in step count about 33 steps
mean 30-minute step count is 253 steps

.51 translates into a 67% increase over no treatment in step count about 170 steps

Midway through study d_t=20 this increase has reduced to 16% increase in step count
.13 translates into a 14% increase over no treatment in step count about 33 steps mean 30-minute step count is 253 steps

.51 translates into a 67% increase over no treatment in step count about 170 steps

Midway through study d_t=20 this increase has reduced to 16% increase in step count

Reminder: This is the result based on the pre-planned hypothesis test (a test of no effect assuming linearity in the treatment effect over day in study). These curves ought not be interpreted as “best fit data analysis”.
HeartSteps Secondary Aim Analysis

On each of $n=37$ participants:

a) Activity suggestion
   - Provide a suggestion with probability .6
     - a tailored walking activity suggestion (probability=.3)
     - a tailored sedentary-reducing activity suggestion (probability=.3)
     - Do nothing (probability=.4)

- 5 times per day * 42 days = 210 decision points
HeartSteps Secondary Aim Analysis

\[ Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t} \]

- \( A_{1t} = 1 \) if walking activity suggestion is delivered at the \( r \)th decision point; \( A_{1t} = 0 \), otherwise,
- \( A_{2t} = 1 \) if sedentary-reducing activity suggestion is delivered at the \( r \)th decision point; \( A_{2t} = 0 \), otherwise

Mean 30-minute step count is 253 steps

.21 translates into a 23% increase over no treatment in step count about 59 steps

When \( d_t \) is added to the model one finds that initially beta0 coefficient of A1t=.729 and the coefficient of A1t*d_t is -.025
Pvalues for both are .000
.729 translates into a 107% increase over no treatment in step count about 271 steps
## HeartSteps Secondary Aim Analysis

<table>
<thead>
<tr>
<th>pid</th>
<th>time ((t))</th>
<th>day ((d_i))</th>
<th>avail ((I_i))</th>
<th>push ((A_i))</th>
<th>walk ((A_{i1}))</th>
<th>sedent ((A_{i2}))</th>
<th>past30 ((Z_i=Y_i))</th>
<th>next30 ((Y_{t+1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>234</td>
<td>316</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>510</td>
<td>120</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>321</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>123</td>
<td>32</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>356</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>210</td>
<td>41</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>470</td>
<td>513</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>76</td>
<td>22</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>37</td>
<td>210</td>
<td>41</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>134</td>
<td>319</td>
</tr>
</tbody>
</table>

n=37 participants
mean 30-minute step count is 253 steps

.21 translates into a 23% increase over no treatment in step count about 59 steps

When d_t is added to the model one finds that initially beta0 coefficient of A1t=.729 and the coefficient of A1t*d_t is -.025

Pvalues for both are .000

.729 translates into a 107% increase over no treatment in step count about 271 steps
HeartSteps Secondary Aim Analysis

\[ Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t} \]

- \( A_{1t} = 1 \) if walking activity suggestion is delivered at the \( t \)th decision point; \( A_{1t} = 0 \), otherwise,
- \( A_{2t} = 1 \) if sedentary-reducing activity suggestion is delivered at the \( t \)th decision point; \( A_{2t} = 0 \), otherwise,

<table>
<thead>
<tr>
<th>Causal Effect</th>
<th>Estimate</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking vs none</td>
<td>( \hat{\beta}_0 = .21 )</td>
<td>(.04, .39)</td>
<td>.02</td>
</tr>
<tr>
<td>Sedentary-reducing vs none</td>
<td>( \hat{\beta}_1 = .03 )</td>
<td>(-.15, .21)</td>
<td>.75</td>
</tr>
<tr>
<td>Walking vs Sedentary-reducing</td>
<td>( \hat{\beta}_0 - \hat{\beta}_1 = .18 )</td>
<td>(.04, .33)</td>
<td>.01</td>
</tr>
</tbody>
</table>

Mean 30-minute step count is 253 steps

.21 translates into a 23% increase over no treatment in step count about 59 steps

When \( d_t \) is added to the model one finds that initially beta0 coefficient of A1t=.729 and the coefficient of A1t*d_t is -.025

P-values for both are .000

.729 translates into a 107% increase over no treatment in step count about 271 steps
mean 30-minute step count is 253 steps

.21 translates into a 23% increase over no treatment in step count about 59 steps

When d_t is added to the model one finds that initially beta0 coefficient of A1t=.729 and the coefficient of A1t*d_t is -.025

P values for both are .000

.729 translates into a 107% increase over no treatment in step count about 271 steps

<table>
<thead>
<tr>
<th>Causal Effect</th>
<th>Estimate</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking, day 1 (d=0)</td>
<td>$\hat{\beta}_0 = 0.729$</td>
<td>(0.42, 1.04)</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Sedty-reduc., day 1 (d=0)</td>
<td>$\hat{\beta}_1 = 0.241$</td>
<td>(-0.18, 0.66)</td>
<td>.24</td>
</tr>
<tr>
<td>Walking, slope</td>
<td>$\hat{\beta}_2 = -0.03$</td>
<td>(-0.04, -0.01)</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Sedentary-reduc., slope</td>
<td>$\hat{\beta}_3 = -0.01$</td>
<td>(-0.03, 0.01)</td>
<td>.23</td>
</tr>
</tbody>
</table>
mean 30-minute step count is 253 steps

.21 translates into a 23% increase over no treatment in step count  about 59 steps

When \( d_t \) is added to the model one finds that initially beta0 coefficient of \( A1t = .729 \) and the coefficient of \( A1t \times d_t \) is \( -.025 \)
Pvalues for both are \(.000\)
.729 translates into a 107% increase over no treatment in step count  about 271 steps
On each of $n=37$ participants:

b) Evening planning prompt, $A_t$

- **Provide a prompt with probability .5**
  - Prompt using unstructured activity planning for following day with probability=.25
  - Prompt using structured activity planning for following day with probability=.25

- **Do nothing with probability=.5

- 1 time per day * 42 days= 42 decision points
HeartSteps Secondary Aim Analysis

\[ Y_{t+1} \sim a_0 + a_1 Z_t + \beta_0 A_t \]
\[ Y_{t+1} \sim a_0 + a_1 Z_t + \alpha_2 S_t + \beta_0 A_t S_t + \beta_1 A_t (1-S_t) \]

- \( t = 1, \ldots, T = 42 \)
- \( Y_{t+1} \) = square root-transformed step count on the day after the \( t^{th} \) day,
- \( A_t = 1 \) if activity planning prompt on the evening of the \( t^{th} \) day; \( A_t = 0 \), otherwise,
- \( Z_t = \) square-root step count on the \( t^{th} \) day,
- \( S_t = 1 \) if Sunday through Thursday; \( S_t = 0 \), otherwise
1.7 translates into an increase in steps of \((1.7)^2 + 2 \times 1.7 \times \sqrt{5000}\) = 243 steps for a daily average step count of 5000 steps.

3.6 translates into a 520 increase over number of steps on weekday compared to no treatment.

No trend with time….  

<table>
<thead>
<tr>
<th>Causal Effect Term</th>
<th>Estimate</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0 A_t) (effect of planning)</td>
<td>(\hat{\beta}_0 = 1.7)</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>(\beta_0 A_t S_t + \beta_1 A_t (1-S_t)) (effect of planning for weekday ((S_t = 1)) and for weekend ((S_t = 0))</td>
<td>(\hat{\beta}_0 = 3.6) (\hat{\beta}_1 &lt; 0)</td>
<td>(.74, 6.4) ns</td>
<td>&lt;.02 ns</td>
</tr>
</tbody>
</table>
On each of $n=37$ participants:

b) Evening planning prompt
   - Provide a prompt with probability .5
     - Prompt using unstructured activity planning for following day with probability=.25
     - Prompt using structured activity planning for following day with probability=.25
   - Do nothing with probability=.5

- 1 time per day * 42 days = 42 decision points
HeartSteps Secondary Aim Analysis

\[ Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t} \]

- \( Y_{t+1} \) = square root-transformed step count on the day after the \( t \)th day,
- \( A_{1t} = 1 \) if unstructured activity planning prompt on the evening of the \( t \)th day; \( A_{1t} = 0 \), otherwise,
- \( A_{2t} = 1 \) if structured activity planning prompt on the evening of the \( t \)th day; \( A_{2t} = 0 \), otherwise,
- \( Z_t \) = square-root step count on the \( t \)th day,
HeartSteps Secondary Aim Analysis

\[ Y_{t+1} \sim a_0 + a_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t} \quad t=0, \ldots, T=41 \]

- \( A_{1t} = 1 \) if unstructured activity planning prompt on the evening of the \( t^{th} \) day; \( A_{1t} = 0 \), otherwise,
- \( A_{2t} = 1 \) if structured activity planning prompt on the evening of the \( t^{th} \) day; \( A_{2t} = 0 \), otherwise,

<table>
<thead>
<tr>
<th>Causal Effect Term</th>
<th>Estimate</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 A_{1t} + \beta_1 A_{2t} )</td>
<td>( \hat{\beta}_0 = 3.1 ) ( \beta_1 &gt; 0 )</td>
<td>(.22, 6.4) ns</td>
<td>.07 ns</td>
</tr>
</tbody>
</table>

No trend with time....
HeartSteps Secondary Aim Analysis

\[ Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \alpha_2 W_t + \beta_0 A_{1t} W_t + \beta_1 A_{1t} (1-W_t) \]
\[ + \beta_2 A_{2t} W_t + \beta_3 A_{2t} (1-W_t) \]

\( A_{1t} = 1 \) if unstructured activity planning prompt on the evening of the \( t^{th} \) day; \( A_{1t} = 0 \), otherwise,

\( W_t = 1 \) if Sunday through Thursday; \( W_t =0 \), otherwise

<table>
<thead>
<tr>
<th>Causal Effect</th>
<th>Estimate</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 A_{1t} W_t + \beta_1 A_{1t} (1-W_t) ) + ( \beta_2 A_{2t} W_t + \beta_3 A_{2t} (1-W_t) )</td>
<td>( \hat{\beta}_0 = 5.3 )</td>
<td>(2.2, 8.5)</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_1 &lt;0, \hat{\beta}_2 &gt;0, )</td>
<td>all ns</td>
<td>all ns</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_3 &lt;0 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.3 translates into an increase in steps of \((5.3)^2 + 2\times(5.3)\times\text{sqrt}(5000) = 782\) steps on weekday for an unstructured planning prompt compared to no treatment for a daily average step count of 5000 steps

No trend with time....
Outline

- Review structure of data arising from MRT
- Review the Primary and Secondary aims in HeartSteps

- Intuition about a new, easy-to-use least squares estimator for analyzing data from a MRT
- Results of the HeartSteps MRT

- Details concerning the causal effects
- Details concerning the new, least squares estimator
Review: Data collected in a MRT

\[ X_t, I_t, A_t, Y_t, \ldots, X_T, I_T, A_T, Y_{T+1} \]

\[ \leftarrow H_t, A_t, Y_{t+1} \quad \text{at a decision point } t \]

- \( t \): decision point (e.g., decide to send activity suggestion)
- \( X_t \): observations collected at \( t \) (passive and active)
- \( I_t = 1 \), i.e., available individuals at \( t \); =0 otherwise
- \( H_t \): history of all data up to time \( t \) decision (includes avail.)
- \( A_t = 1 \) if activity suggestion; and \( A_t = 0 \) otherwise
- Known probs. \( \rho_t(1|H_t) = \Pr(A_t = 1|H_t) \)
- \( Y_{t+1} \): proximal outcome
- \( Z_t \): subset of \( H_t \) (or summaries of \( t \)) that are expected to be associated with \( Y_{t+1} \) (i.e., explain variance in \( Y_{t+1} \))
Draw attention to $a_t$ and $a_{t-1}$

Draw attention to why availability might depend on prior treatment
One More Potential Outcome

- Let $S_t(\bar{a}_{t-1})$ be a vector of features of $t$ or $H_t(\bar{a}_{t-1})$, that represent States of Scientific interest
- $S_t(\bar{a}_{t-1})$ candidate time-varying moderators
- Examples:
  - Day in study, $S_t(\bar{a}_{t-1}) = d(t)$
  - Previous step count, $S_t(\bar{a}_{t-1}) = Y_t(\bar{a}_{t-1})$
  - Step count following any push over the past 24 hours
  - Empty set (i.e., no $S_t(\bar{a}_{t-1})$)

Draw attention to $a_t$ and $a_{t-1}$

Draw attention to why availability might depend on prior treatment
Marginal & Causal Effect

Causal effect at decision point $t$:

$$E[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) | I_t(\bar{A}_{t-1}) = 1, S_t(\bar{A}_{t-1})]$$

- Contrast of the average $Y_{t+1}$ under current treatment $A_t = 1$ versus the average $Y_{t+1}$ under no current treatment $A_t = 0$
- Conditional on $S_t(\bar{A}_{t-1})$, whatever you chose it to be
- Conditional on availability; only concerns the subpopulation of individuals available at decision $t$

Note how odd this causal effect is—it involves the randomization probabilities.

Availability is not equivalent to willingness to enroll. It is momentary. Willingness to enroll is up front. Our availability is closer to feasibility of trt options.

We are examining causal effects via ****excursions*** from the underlying randomization policy

$S_{-t}$ could be an empty set.

Marginal over randomization treatment policy (and effects thereof), conditional on those who have intervention on.

The group who have the intervention turned on is a selected group of people likely depending on the intervention dose they experienced up to time $j$. This intervention dose $\bar{A}_{j-1}$ may have caused burden, may have caused learning.
Note how odd this causal effect is—it involves the randomization probabilities.

Availability is not equivalent to willingness to enroll. It is momentary. Willingness to enroll is up front. Our availability is closer to feasibility of trt options.

We are examining causal effects via ***excursions*** from the underlying randomization policy.

$S_{t}$ could be an empty set.

Marginal over randomization treatment policy (and effects thereof), conditional on those who have intervention on.

The group who have the intervention turned on is a selected group of people likely depending on the intervention dose they experienced up to time $j$. This intervention dose $\bar{A}_{j-1}$ may have caused burden, may have caused learning.
Note how odd this causal effect is—it involves the randomization probabilities.

Availability is not equivalent to willingness to enroll. It is momentary. Willingness to enroll is up front. Our availability is closer to feasibility of trt options.

We are examining causal effects via **excursions** from the underlying randomization policy.

\( S_t \) could be an empty set.

Marginal over randomization treatment policy (and effects thereof), conditional on those who have intervention on.

The group who have the intervention turned on is a selected group of people likely depending on the intervention dose they experienced up to time \( j \). This intervention dose \( \bar{A}_{j-1} \) may have caused burden, may have caused learning.
Excursion effect at decision point $t$:

$$E[Y_{t+1}(\tilde{A}_{t-1}, 1) - Y_{t+1}(\tilde{A}_{t-1}, 0) \mid l_t(\tilde{A}_{t-1}) = 1, S_t(\tilde{A}_{t-1})]$$

- This causal effect is (potentially) a function of the randomization probabilities
- This is because the expectation is over the distribution of any treatments in $\tilde{A}_{t-1}$ not included in $S(\tilde{A}_{t-1})$

Note how odd this causal effect is—it involves the randomization probabilities.

Availability is not equivalent to willingness to enroll. It is momentary. Willingness to enroll is up front. Our availability is closer to feasibility of trt options

We are examining causal effects via ****excursions*** from the underlying randomization policy

$S_{\cdot t}$ could be an empty set.

Marginal over randomization treatment policy (and effects thereof), conditional on those who have intervention on.

The group who have the intervention turned on is a selected group of people likely depending on the intervention dose they experienced up to time $j$. This intervention dose $\bar{A}_{\cdot (j-1)}$ may have caused burden, may have caused learning.
Identification from the MRT Data: Micro-Randomized $A_t \rightarrow$

\[
E[Y_{t+1}(\tilde{A}_{t-1}, 1) - Y_{t+1}(\tilde{A}_{t-1}, 0) \mid I_t(\tilde{A}_{t-1}) = 1, S_t(\tilde{A}_{t-1})] = E\left[ E[Y_{t+1} \mid A_t = 1, I_t = 1, H_t] - E[Y_{t+1} \mid A_t = 0, I_t = 1, H_t] \mid I_t = 1, S_t \right] = E\left[ \frac{A_t Y_{t+1}}{\rho_t(H_t)} - \frac{(1 - A_t) Y_{t+1}}{1 - \rho_t(H_t)} \mid I_t = 1, S_t \right]
\]

($\rho_t(H_t)$ is randomization probability)

$p_{t}(H_{t})$ is randomization probability
Marginal Treatment Effect

Treatment Effect Model:

\[
E\left[ E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - E[Y_{t+1}|A_t = 0, I_t = 1, H_t]| I_t = 1, S_t \right] = S_t^T \beta
\]

- \(H_t\) is participant’s data up to and at time \(t\)
- \(S_t\) is a vector of data summaries and time, \(t\), \((S_t \subseteq H_t)\)
- \(I_t\) indicator of availability

We aim to conduct inference about \(\beta\)!

We are examining causal effects via ****excursions*** from the underlying randomization policy
Centered and Weighted Least Squares

Simple method for complex data!

- **Easy**: Implement using OTC software
- **Versatile**: Able to examine a variety of causal effects
- **Efficient**: Incorporate variables $Z_t$ to reduce noise but…
- **Robust**: Not require all predictors of $Y_{t+1}$ in $Z_t$, nor require the true or good predictive model for $Y_{t+1}$

Boruvka, Almirall, Witkiewitz, Murphy (2016)
https://arxiv.org/abs/1601.00237
Estimation

- Select probabilities: $\tilde{p}_t(s) \in (0,1)$
- Form weights: $W_t = \left( \frac{\tilde{p}_t(S_t)}{\rho_t(H_t)} \right)^{A_t} \left( \frac{1-\tilde{p}_t(S_t)}{1-\rho_t(H_t)} \right)^{1-A_t}$
- Center treatment actions: $A_t \rightarrow (A_t - \tilde{p}_t(S_t))$

Form working model for mean of $Y_{t+1}$ given $I_t=1$ and $H_t$
Estimation

- Select probabilities: $\tilde{p}_t(s) \in (0,1)$
- Form weights: $W_t = \left( \frac{\tilde{p}_t(S_t)}{\rho_t(H_t)} \right)^{A_t} \left( \frac{1-\tilde{p}_t(S_t)}{1-\rho_t(H_t)} \right)^{1-A_t}$
- Center treatment actions: $A_t \rightarrow (A_t - \tilde{p}_t(S_t))$
- Minimize:
  $$E_n \left[ \sum_{t=1}^{T} (Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t))S_t^T \beta)^2 I_t W_t \right]$$
- $E_n$ is empirical distribution over individuals.
Implementation with Known Weights

- Select probabilities: \( \tilde{p}_t(s) \in (0,1) \)
- Form weights: \( W_t = \left( \frac{\tilde{p}_t(S_t)}{\rho_t(H_t)} \right)^{A_t} \left( \frac{1-\tilde{p}_t(S_t)}{1-\rho_t(H_t)} \right)^{1-A_t} \)
- Center treatment actions: \( A_t \rightarrow (A_t - \tilde{p}_t(S_t)) \)

<table>
<thead>
<tr>
<th>pid</th>
<th>time</th>
<th>avail (a)</th>
<th>a</th>
<th>denohe (x)</th>
<th>moditr (z)</th>
<th>numprob (p)</th>
<th>denprob (q)</th>
<th>Ynext</th>
<th>O_{next}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>234</td>
<td>1</td>
<td>.2</td>
<td>.4</td>
<td>316</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>510</td>
<td>0</td>
<td>.4</td>
<td>.2</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>123</td>
<td>1</td>
<td>.2</td>
<td>.6</td>
<td>321</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>123</td>
<td>1</td>
<td>.2</td>
<td>.5</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

Form working model for mean of \( Y_{t+1} \) given \( I_t=1 \) and \( H_t \)
Form working model for mean of $Y_{t+1}$ given $I_t=1$ and $H_t$
Sidebar: The Special Case of HeartSteps

- Randomization probability: $\rho_t(H_t) = .6$
- Select: $\tilde{p}(s) = .6$ for any $S_t$
- Weights: $W_t = \left( \frac{\tilde{p}_t(S_t)}{\rho_t(H_t)} \right)^{A_t} \left( \frac{1-\tilde{p}_t(S_t)}{1-\rho_t(H_t)} \right)^{1-A_t} = 1.0$
- Center treatment: $(A_t - .6)$
- Recall primary analysis: $Y_{t+1} \sim \alpha_0 + \alpha_i Z_i + \beta_0 A_i$
- Minimize:

$$E_n \left[ \sum_{t=1}^{T}(Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t))S_t^T \beta)^2 I_t W_t \right]$$

$$= E_n \left[ \sum_{t=1}^{T}(Y_{t+1} - \alpha_0 - \alpha_i Z_t - (A_t - .6)\beta_0)^2 I_t \right]$$

$$= E_n \left[ \sum_{t=1}^{T}(Y_{t+1} - \alpha'_0 - \alpha_i Z_t - (A_t)\beta_0)^2 I_t \right]$$

Form working model for mean of $Y_{t+1}$ given $I_t = 1$ and $H_t$
This intuition is not correct because we are not assuming that the conditional mean of \( Y_{t+1} \) given \( Z_t, S_t, I_t = 1 \), has above form!
This intuition is not correct because we are not assuming that the conditional mean of $Y_{t+1}$ given $Z_t, S_t, I_t=1$ has above form!
Yes modeling the Causal Effect

\[ E_n \left[ \sum_{t=1}^{T} (Y_{t+1} - Z_t^\top \alpha - (A_t - \tilde{p}_t(S_t))S_t^\top \beta)^2 I_t W_t \right] \]

The Modeling Assumption:

\[ E[(E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - E[Y_{t+1}|A_t = 0, I_t = 1, H_t])|I_t = 1, S_t] = S_t^\top \beta_0 \]

If \( \tilde{p}_t \) depends at most on features in \( S_t \), then, under moment conditions, \( \hat{\beta} \) is consistent for \( \beta_0 \)

If in the \( W_t \) you included \( \tilde{p}_t(S_t) (1-\tilde{p}_t(S_t))^\top \mathbf{1} \) then you would be obtaining the BLP of the treatment effect.
Statistical Theory (& Standard Errors)

Under moment conditions, $\sqrt{n}(\hat{\beta} - \beta_0)$ converges to a Normal distribution with mean 0 and var-covar matrix, $(\Sigma_p)^{-1} \Sigma (\Sigma_p)^{-1}$

$$\Sigma_p = \mathbb{E}[\sum_{t=1}^{T} \tilde{p}_t(S_t)(1 - \tilde{p}_t(S_t))I_t S_t S_t^T]$$

The usual robust standard errors unless is $\tilde{p}_t(S_t)$ estimated...

$Z_t$ and $S_t$ are finite dimensional feature vectors.
More About Implementation

- No special software is necessary to obtain $\hat{\beta}$
- If $\hat{p}_t(S_t)$ is not estimated, no special software is necessary to obtain $\text{var}(\hat{\beta})$ (i.e., standard errors).
- If $\hat{p}_t(S_t)$ is estimated, we have to make a small adjustment to the usual standard errors:
  - http://www-personal.umich.edu/~dalmiral/software/mHealth/proximal-lagged-moderated-mHealth.html
  - Code includes small sample adjustments: Mancl & DeRouen (SE adjust.), Hotelling $T^2$ (df adjust.)
Gains from Randomization

- Causal inference for a marginal treatment effect
- Inference on treatment effect is robust to working model:

\[ E[W_t Y_{t+1} | I_t = 1, H_t] \approx Z_t^T \alpha \]

- \( Z_t \subseteq H_t \)
- Contrast to literature on partially linear, single index models and varying coefficient models
This “GEE-like” method can only use a working independence correlation matrix

- Estimating function is biased if off-diagonal elements in working correlation matrix: in general,

\[
E\left[ (Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t))S_t^T \beta) I_t W_t I_u W_u \sigma_{t,u} Z_u \right] \neq 0
\]

if \( u \neq t \)

Such a result is unsurprising given the bias that arises under non-independence structures in IPTW (Vansteelandt 2007; Tchetgen Tchetgen et al. 2012) or in GEEs where a time-varying response is modelled by time-varying covariates (Pepe and Anderson 1994; Schildcrout and Heagerty 2005).
Such a result is unsurprising given the bias that arises under non-independence structures in IPTW (Vansteelandt 2007; Tchetgen Tchetgen et al. 2012) or in GEEs where a time-varying response is modelled by time-varying covariates (Pepe and Anderson 1994; Schildcrout and Heagerty 2005).
Such a result is unsurprising given the bias that arises under non-independence structures in IPTW (Vansteelandt 2007; Tchetgen Tchetgen et al. 2012) or in GEEs where a time-varying response is modelled by time-varying covariates (Pepe and Anderson 1994; Schildcrout and Heagerty 2005).
Such a result is unsurprising given the bias that arises under non-independence structures in IPTW (Vansteelandt 2007; Tchetgen Tchetgen et al. 2012) or in GEEs where a time-varying response is modelled by time-varying covariates (Pepe and Anderson 1994; Schildcrout and Heagerty 2005).
Choice of Numerator of the Weights

Choice of $\tilde{p}_t(S_t)$ determines marginalization over time under model misspecification of treatment effect.

Example: $S_t = 1$, $\tilde{p}_t(S_t) = \tilde{p}$. Resulting $\hat{\beta}$ is an estimator of

$$\frac{\sum_{t=1}^{T} E[I_t] \beta_t}{\sum_{t=1}^{T} E[I_t]}$$

where

$$\beta_t = E\left[ E[Y_{t+1} | A_t = 1, I_t = 1, H_t] - E[Y_{t+1} | A_t = 0, I_t = 1, H_t | I_t = 1] \right]$$

Use when assumption does not hold.

Note that this estimand weighs all available person-times equally. If instead you would prefer $E\left[ \sum_{t=1}^{T} I_t \beta_t / \sum_{t=1}^{T} I_t \right]$ that is, you might prefer first to average within a person and then average across people, then you are really looking for a random effects model.
Discussion

• **Our goal:** Develop methodology for examining a variety of marginal causal effects (rather than a model for $Y_{t+i}$)
• Hypothesis testing versus data analysis
• Lagged effects
• Our lab at Harvard is working on a new way to conceptualize random effects modeling, and estimating BLUPs