Micro-randomized Trials in Mobile Health

S.A. Murphy
Henry Seeley White
Lecture 2
mHealth

MD2K Smoking Cessation Coach

- Wearable wrist/chest bands provide multiple physiological sensor streams...; phone records location, burden,.....
- Supportive stress-regulation interventions available on smartphone 24/7
- In which contexts should the wrist band provide supportive “cue” and smartphone activate to highlight associated support?
mHealth

HeartSteps Activity Coach

- Wearable band measures activity and sleep quality; phone sensors measure busyness of calendar, location, weather, ….

- In which contexts should smartphone ping and deliver activity ideas?
Data from wearable devices that sense and provide treatments

On each individual:

\[ O_1, A_1, Y_2, \ldots, O_t, A_t, Y_{t+1}, \ldots \]

- \( O_t \): Observations at \( t \)th decision time (high dimensional)
- \( A_t \): Action at \( t \)th decision time (treatment)
- \( Y_{t+1} \): Proximal Response (aka: Reward, Utility, Cost)
Examples

1) Decision Times (Times at which a treatment can be provided.)
   1) Regular intervals in time (e.g. every 10 minutes)
   2) At user demand

HeartSteps: Approximately every 2-2.5 hours

Smoking Cessation: Every 1 minute during 10 hour day.
Examples

2) Observations $O_t$
   1) Passively collected (via sensors)
   2) Actively collected (via self-report)

HeartSteps: classifications of activity, location, step count, busyness of calendar, usefulness ratings, adherence……..

Smoking Cessation: classifications of stress, smoking detection, mood, driving,…. 
Examples

3) Actions $A_t$
   1) Treatments that can be provided at a decision time
   2) Whether to provide a treatment

HeartSteps: Activity recommendation on phone
Smoking Cessation: Cue on wrist band
Activity
Recommendation

No Message or
Examples

4) Proximal Response (reward) $Y_{t+1}$

**HeartSteps**: Activity (step count) over next 30 minutes.

**Smoking Cessation**: Stress level over next $x$ minutes
Continually Learning Mobile Health Intervention

1) Trial Designs: Are there effects of the actions on the proximal response? experimental design

2) Data Analysis Methods for use with trial data: Are there delayed effects of the actions? Do effects vary by context, observations; predict treatment burden? causal inference

3) Learning algorithms for use with trial data: Construct a “warm-start” treatment policy. batch “control” methods

4) Online training algorithms that will result in a Personalized Continually Learning mHealth Intervention. online “control” methods
Micro-Randomized Trial

Randomize between actions at decision times →
Each person may be randomized 100’s or 1000’s of times.

- These are sequential, “full factorial,” designs.
- Design trial to detect main effects.

*Extension of A/B testing & Single Case Designs*
Micro-Randomized Trial Elements

1. **Record outcomes**
   - Distal (scientific/clinical goal) & Proximal Response

2. **Record context** (sensor & self-report data)

3. **Randomize among treatment actions at decision points**

4. **Use data after study ends to assess treatment effects, learn warm-start treatment policy**
Micro-Randomized Trial

How to justify the trial costs?

• Address a question that can be stated clearly across disciplinary boundaries and be able to provide guarantees.

• Design trial so that a variety of further interesting questions can be addressed.

First Question to Address: Do the treatment actions impact the proximal response? (aka, is there a main effect?)
Micro-Randomized Trial for HeartSteps

• 42 day trial
• Whether to provide an Activity recommendation? $A_t \in \{0, 1\}$
• Test for main effects on proximal response
• Randomization in HeartSteps

$$P[A_t = 1] = .4 \quad t = 1, \ldots, T$$
Time-varying Main Effects

Time varying potentially intensive/intrusive treatment actions $\rightarrow$ potential for accumulating habituation and burden

$\rightarrow$

In the test statistic allow the main effect of the treatment actions on proximal response to vary with time
Availability & the Treatment Effect

- Treatment actions can not be delivered at a decision time if an individual is *unavailable*.

- The effect of treatment at a decision time is the difference in proximal response between *available* individuals assigned an activity recommendation and *available* individuals who are not assigned an activity recommendation.
Availability

• Treatment actions can only be delivered at a decision time if an individual is available

• Set $I_t = 1$ if the individual is available at decision time $t$, otherwise, $I_t = 0$

• Availability is not the same as adherence.
Potential Outcomes

• Define

\[ \bar{A}_t = \{A_1, A_2, \ldots, A_t\}, \bar{a}_t = \{a_1, a_2, \ldots, a_t\} \]

• Define \( Y_{t+1}(\bar{a}_t) \) to be the observed response,
  \( Y_{t+1} \) if \( \bar{A}_t = \bar{a}_t \), e.g., \( Y_{t+1} = Y_{t+1}(\bar{A}_t) \)

• Define \( I_t(\bar{a}_{t-1}) \) to be the observed “available for treatment” indicator if \( \bar{A}_{t-1} = \bar{a}_{t-1} \)
Main Effect

• Define the main effect at time $t$ as

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

• What does this main effect mean?
Main Effect

• The randomization implies that

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

• Put

\[
\beta(t) = E[Y_{t+1}|I_t = 1, A_t = 1] - E[Y_{t+1}|I_t = 1, A_t = 0]
\]
Proposal

Determine the number of participants so that micro-randomized trial can detect a main effect of treatment on proximal response

The main effect is a time-varying main effect $\beta(t), t=1,\ldots,T$

The main effect is a causal effect.
Sample Size Calculation

• We calculate the number of subjects to test $H_0$: no effect of the action, i.e.,

$$H_0 : \beta(t) = 0, t = 1, 2, ..., T$$

• Size to detect a low dimensional, smooth alternate $H_1$.
  – Example: $H_1$: $\beta(t)$ quadratic with intercept, $\beta_0$, linear term, $\beta_1$, and quadratic term $\beta_2$ and test

$$\beta_0 = \beta_1 = \beta_2 = 0$$
Sample Size Calculation

Alternative hypothesis is low dimensional → assessment of the effect of the activity recommendation uses contrasts of between subject responses + contrasts of within subject responses.

---The required number of subjects will be small.
Test Statistic for Sample Size Calculation

Test statistic based on least squares projection of $E[Y_{t+1} | I_t = 1, A_t]$ on functions of the form

$$\gamma(t) + \beta(t)(A_t - q_t)$$

where $q_t$ is the randomization probability

$q_t = .4$ in HeartSteps

- We are not assuming this model is correct............
Test Statistic for Sample Size Calculation

• Test statistic is based on least squares fit of
  \[ \gamma(t) + \beta(t)(A_t - q_t) \] to \[ Y_{t+1} \] when \[ I_t = 1 \]

HeartSteps:

\[ \beta(t) = \beta_0 + \beta_1 \left\lfloor \frac{t-1}{5} \right\rfloor + \beta_2 \left\lfloor \frac{t-1}{5} \right\rfloor^2 \]

• You select parameterization of \( \gamma(t) \)
Alternative for Sample Size Calculation

• One calculates a sample size to detect a given alternative with a given power.

• Alternative:

\[ H_1 : \beta_i = d_i \bar{\sigma}, \ i = 0, 1, 2 \]

where \( \bar{\sigma}^2 \) is the average conditional variance.
Alternative for Sample Size Calculation

• Average conditional variance is

\[ \bar{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} E[VAR(Y_{t+1} | I_t = 1)] \]
Specify Alternative for Sample Size Calculation

• Scientist indirectly specifies standardized $d_i$’s
  – initial main effect: $d_0$,
  – average main effect over trial duration:
    \[ \frac{1}{T} \sum_{t=1}^{T} \left( d_0 + d_1 \left\lfloor \frac{t-1}{5} \right\rfloor + d_2 \left\lfloor \frac{t-1}{5} \right\rfloor^2 \right), \]
    – and day of maximal main effect: $- \frac{d_1}{2d_2}$
• We solve for $d_0$, $d_1$, $d_2$
Test Statistic for Sample Size Calculation

- Put \( Y_i = (Y_{i2}, \ldots, Y_{iT+1})^T \) for \( i^{th} \) subject

\( q+3 \) is the total number of parameters;

\( X_i \) is the associated design matrix (\( T \) by \( q+3 \))

\( N \) is sample size

Last 3 columns of \( X_i \) contain row entries:

\[
I_{it}(A_{it} - q_t), \quad I_{it}(A_{it} - q_t)\left\lfloor \frac{t-1}{5} \right\rfloor, \\
I_{it}(A_{it} - q_t)\left\lfloor \frac{t-1}{5} \right\rfloor^2
\]
Test Statistic for Sample Size Calculation

• “GEE” test statistic is

\[ N \hat{\beta}^T (K \hat{\Sigma} K^T)^{-1} \hat{\beta} = N \hat{\beta}^T (\hat{\Sigma}_\beta)^{-1} \hat{\beta} \]

where \( \hat{\Sigma} \) is the usual sandwich estimator of the variance-covariance and \( K \) is a 3 by \( 3+q \) matrix picking out columns associated with coefficients \( \beta \)
Working Assumptions for Sample Size Calculation

1) \( E(\epsilon_{it}\epsilon_{it'} \mid I_{it} = 1, I_{it'} = 1, A_{it}, A_{it'}) \) is constant.

2) \( E(\epsilon_{it} \mid I_{it} = 1, A_{it}) = 0 \)

3) \( Var(\epsilon_{i,t} \mid I_{it} = 1, A_{it}) \) is constant.

\[
\epsilon_{it} = Y_{i,t+1} - \left( \gamma(t) + \beta(t)(A_{it} - q_t) \right)
\]
Sample Size Calculation

• Under the working assumptions, $\Sigma \beta$ only depends on polynomials in $\lfloor \frac{t-1}{5} \rfloor$, the marginal distribution of $I_t$ and on the randomization probabilities.

• $\Sigma \beta$ does not depend on the form of $\gamma(t)$
Sample Size Calculation

• Under standard moment assumptions, the asymptotic distribution of the “GEE test statistic” is a Chi-Squared on 3 degrees of freedom with non-centrality parameter:

\[ Nd^T (\Sigma_\beta)^{-1} d \]

• Instead of a Chi-Squared on 3 degrees we use \( \frac{3(N-q-1)}{N-q-3} F_{3, N-q-3} \) with the same non-centrality parameter.
HeartSteps Example

- Standardized $d_i$’s
  - initial effect: $d_0 = 0$
  - output average main effect
  - day of maximal main effect: $-\frac{d_1}{2d_2} = 28$

- Model for test statistic:
  $$
  \gamma(t) + \beta(t)(A_{it} - .4), \ t = 1, \ldots, 210
  $$

where
  $$
  \gamma(t) = \gamma_0 + \gamma_1 \left[ \frac{t-1}{5} \right] + \gamma_2 \left[ \frac{t-1}{5} \right]^2
  $$
<table>
<thead>
<tr>
<th>Standardized Average Main Effect over 42 Days</th>
<th>Sample Size For 70% availability or 50% availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06 standard deviations</td>
<td>81 or 112</td>
</tr>
<tr>
<td>0.08 standard deviations</td>
<td>48 or 65</td>
</tr>
<tr>
<td>0.10 standard deviations</td>
<td>33 or 43</td>
</tr>
</tbody>
</table>
Same Test Statistic for Analysis

• “GEE” test statistic is

\[ N \hat{\beta}^T (K \hat{\Sigma} K^T)^{-1} \hat{\beta} \]

where \( K \) is a 3 by \( 3 + p \) matrix picking out columns associated with \( \beta \) coefficients

• No working assumptions
Small Sample Adjustment

• $\hat{e}_{it}$ is the $i^{th}$ subject, $t^{th}$ time point residual and
$\hat{e}_i = (\hat{e}_{i1}, \ldots, \hat{e}_{iT})^T$

• Adjusted sandwich estimator:
$\hat{\Sigma} = \hat{\sigma}^2 N \left( \sum_{i=1}^{N} X_i^T X_i \right)^{-1} \left\{ \sum_{i=1}^{N} X_i^T B_i \hat{\epsilon}_i \hat{\epsilon}_i^T B_i X_i \right\} \left( \sum_{i=1}^{N} X_i^T X_i \right)^{-1}$

$B_i = (I - H_{ii})^{-1}$
## Simulation Results

### Type 2 Error Rate (2000 data sets)

<table>
<thead>
<tr>
<th>Average Main Effect (Sample Size)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05(115)</td>
<td>0.790</td>
</tr>
<tr>
<td>0.06(81)</td>
<td>0.794</td>
</tr>
<tr>
<td>0.07(61)</td>
<td>0.800</td>
</tr>
<tr>
<td>0.08(48)</td>
<td>0.801</td>
</tr>
<tr>
<td>0.09(39)</td>
<td>0.798</td>
</tr>
<tr>
<td>0.10(33)</td>
<td>0.803</td>
</tr>
</tbody>
</table>
The micro-randomized trial is a sequential factorial trial with multiple factors, e.g. in HeartSteps:

Factor 1: Activity recommendation is randomized 5 times per day (yes/no)

Factor 2: Daily activity planning is randomized each evening (yes/no)
Experimental Design Challenges

Micro-randomized trials are a new type of factorial design

i. Time varying factors $\rightarrow$ time varying main effects, time-varying two-way interactions, different delayed effects

ii. Randomization that depends on an outcome of past actions

iii. Design studies specifically to detect interactions between factors.