Micro-randomized Trials & Mobile Health

S.A. Murphy
IST, Vienna
MD2K Smoking Cessation Coach

- Wearable bands measure activity, stress, cigarette smoking, sleep quality……..
- Smartphone provides four types of support 24/7
- Should wrist band provide supportive “cue” and smartphone activate to highlight associated support when stress reaches a criterion?
mHealth

HeartSteps Activity Coach

- Wearable bands measure activity, phone sensors measure busyness of calendar, location,.....

- Should smartphone ping and lockscreen deliver activity ideas when user is receptive and user’s calendar is not too busy?
Data from wearable devices that sense and provide treatments

\[ O_1, A_1, Y_2, \ldots, O_j, A_j, Y_{j+1}, \ldots \]

\( O_j \): Observations at j\textsuperscript{th} decision time (high dimensional)

\( A_j \) : Action at j\textsuperscript{th} decision time (treatment)

\( Y_{j+1} \) : Proximal Response (aka, Reward, Cost, Utility)
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 includes two sets of decision times
1) Momentary: Approximately every 2-2.5 hours
2) Daily: Each evening at user specified time.
Examples

2) Observations $S_j$
   1) Passively collected (location, weather, busyness of calendar, social context, activity on device, physical activity)
   2) Actively collected (self-report)

HeartSteps includes activity recognition (walking, driving, standing/sitting), weather, location, calendar, adherence, step count, whether momentary intervention is on, self-report: usefulness, burden, self-efficacy, etc.
Examples

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

HeartSteps includes two types of treatments
1) Momentary Lock Screen Recommendation
2) Daily Activity Planning
Momentary Lock Screen Recommendation

No Message or
Examples

4) Proximal Response $Y_{j+1}$

HeartSteps: Activity (step count) over next 60 minutes.
Smoking Cessation: Stress level over next $x$ minutes.
Our Group’s Scientific Goals

1) Develop trial designs/data analytics for assessing if there are proximal effects of the actions on the response.

2) Develop data analytics for assessing if there are delayed effects of the actions; assess if the effects vary by context, observations.

3) Develop data methods for constructing a treatment policy that inputs observations and delivers actions via phone.

4) Develop online training algorithms that will result in a Personalized Continually Learning mHealth Intervention
Proposed Experimental Design: Micro-Randomized Trial

Randomize between actions at decision times \( \rightarrow \) Each person may be randomized 100’s or 1000’s of times.

These are sequential, “full factorial,” designs.
Why Micro-Randomization?

• Randomization (+ representative sample) is a gold standard in providing data to assess the causal effect of an intervention option.

• Sequential randomizations will enhance replicability and effectiveness of data-based decision rules.
Micro-Randomized Trial Elements

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

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

3. Randomize among intervention options at decision points

4. Use resulting data to assess treatment effects, construct decision rules
Micro-Randomized Trial

• Focus on whether to provide a Momentary Lock Screen Recommendation, e.g.
  \[ A_j \in \{0, 1\} \]

• Randomization in HeartSteps
  \[ P[A_j = 1] = .4 \quad j = 1, \ldots, J \]
Micro-Randomized Trial

First Question to Address: Do the intervention options have an effect on the proximal response?

--Test for proximal main effects of the intervention
Micro-Randomized Trial

Time varying potentially intensive intervention delivery → potential for accumulating habituation and burden

→

Allow proximal main effects of the intervention components to vary with time
Availability & The Main Effect

- Interventions can only be delivered at a decision time if an individual is available.

- The proximal main effect of treatment at a decision time is the difference in proximal response between available individuals assigned a lock-screen message and available individuals who are not assigned a lock-screen message.
Availability

- $A_j$ is only delivered if the intervention is on at decision time $j$.

- Set $I_j = 1$ if the intervention is on at decision time $j$, otherwise $I_j = 0$. 
Potential Outcomes

- Define

$$\tilde{A}_j = \{A_1, A_2, \ldots, A_j\}, \tilde{a}_j = \{a_1, a_2, \ldots, a_j\}$$

- Define $$Y_{j+1}(\tilde{a}_j)$$ to be the observed response, $$Y_{j+1}$$ if $$\tilde{A}_j = \tilde{a}_j$$, e.g., $$Y_{j+1} = Y_{j+1}(\tilde{A}_j)$$

- Define $$I_j(\tilde{a}_{j-1})$$ to be the observed “intervention on” indicator if $$\tilde{A}_{j-1} = \tilde{a}_{j-1}$$
Proximal Main Effect

• The randomization implies that

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

$$E[Y_{j+1}|I_j = 1, A_j = 1] - E[Y_{j+1}|I_j = 1, A_j = 0]$$
Proximal Main Effect

• The Proximal Main Effect at time \( j \) is

\[
\beta(j) = E[Y_{j+1}|I_j = 1, A_j = 1] - E[Y_{j+1}|I_j = 1, A_j = 0]
\]

• What does this estimand mean?
Proposal

Design and size micro-randomized trial to detect proximal main effect of treatment

The proximal main effect is a time-varying main effect $\beta(j), j=1,\ldots,J$

The proximal main effect is a causal effect.
Sample Size Calculation

• We calculate a sample size to test:

\[ H_0 : \beta(j) = 0, j = 1, 2, \ldots, 210 \]

• Size to detect a low dimensional alternative. E.g. \( H_1 : \beta(j) \) 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

Because the alternative hypothesis is low dimensional, assessment of the effect of the lock-screen message uses not only contrasts of between person responses but also contrasts of within person responses.

--The required sample size (number of subjects) will be small.
### HeartSteps Sample Sizes

Power=.8, α=.05

<table>
<thead>
<tr>
<th>Standardized Average Proximal Effect over 42 Days</th>
<th>Sample Size For 70% availability or 50% availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>81 or 112</td>
</tr>
<tr>
<td>0.08</td>
<td>48 or 65</td>
</tr>
<tr>
<td>0.10</td>
<td>33 or 43</td>
</tr>
</tbody>
</table>
These are a new type of Factorial Design

- Time varying factors ➔ time varying main effects, time-varying two-way interactions, different delayed effects
- Better Designs?
- Design Studies to Detect Interactions Between Factors.
Steps Toward Long-Term Goal

1) Develop methods/trial designs for assessing if there are proximal effects of the actions on the response.

2) Develop data analytics for assessing if there are delayed effects of the actions; assess if the effects vary by context/observations.

3) Develop data methods, to use with batch data, for constructing a treatment policy that inputs observations and delivers actions via mobile device

4) Develop online training algorithms that will result in a “Continually Updating” Treatment Policy
**Steps Toward Long-Term Goal**

1) Develop methods/trial designs for assessing if there are proximal effects of the actions on the response.

2) Develop data analytics for assessing if there are delayed effects of the actions; assess if the effects vary by context/observations.

3) Develop data methods, to use with batch data, for constructing a treatment policy that inputs observations and delivers actions via mobile device.

4) Develop online training algorithms that will result in a “Continually Updating” Treatment Policy.
Current State

• Clinical scientists formulate mobile health intervention (e.g. treatment policy) using ideas from the literature, behavioral theory, clinical experience, observational data analyses.

• Develop analysis methods for use with data in constructing “evidence-based” treatment policies.
  -- treatment policy should be interpretable.
We aim to construct a parameterized policy, $\pi_\theta(a|s)$ that is bounded away from 0 and 1.

- Variation in actions can help retard habituation and maintain engagement.
- $\pi_\theta(a|s)$ that are continuous in $\theta$ are easier to estimate/compute.
Background

1) On each of n individuals data set contains:
   \[ S_1, A_1, Y_2, \ldots, S_T, A_T, Y_{T+1} \]
   -- \( S_t \) is a summary of \( O_1, A_1, Y_2, \ldots, Y_t, O_t \) that permits the Markovian property; a modeling assumption.

   -- \( P[A_t = a \mid S_t = s] = \mu(a \mid s) \)

2) Optimality Criterion: Average Reward for Markov Decision Process
Markov Decision Process (MDP)

Markovian Assumptions

\[ P[S_{j+1} = s'|S_1, A_1, \ldots, S_j, A_j] = P[S_{j+1} = s'|S_j, A_j] \]

and

\[ P[Y_{j+1} = r|S_1, A_1, \ldots, S_j, A_j] = P[Y_{j+1} = r|S_j, A_j] \]

Stationarity Assumptions

\[ P[S_{j+1} = s'|S_j = s, A_j = a] = p(s'|s, a) \]

and

\[ E[Y_{j+1}|S_j = s, A_j = a] = r(s, a) \]
Optimality Criterion

Average Reward, $\eta_\theta$, for policy $\pi_\theta$:

$$
\eta_\theta = \lim_{T \to \infty} \frac{1}{T} E_\theta \left[ \sum_{t=0}^{T-1} Y_{t+1} \bigg| S_0 = s \right]
$$

$$
= \sum_s d_\theta(s) \sum_a \pi_\theta(a|s) r(s, a)
$$

$E_\theta$ denotes expectation under the stationary distribution, $d_\theta$, associated with $\pi_\theta$. 
Background: Differential Value

$V_\theta$ is the Differential Value

$$V_\theta(s) = \lim_{T \to \infty} E_\theta \left[ \sum_{t=0}^{T} \left( Y_{t+1} - \eta_\theta \right) \bigg| S_0 = s \right].$$

$V_\theta(s) - V_\theta(s')$ reflects the difference in sum of centered responses accrued when starting in state $s$ as opposed to state $s'$.

$(\eta_\theta$ is the average reward)
Background: Bellman Equation

Oracle Temporal Difference:

\[ \delta_t = Y_{t+1} - \eta_\theta + V_\theta(S_{t+1}) - V_\theta(S_t) \]

Bellman Equation:

\[
E_{\theta} \left[ \delta_t \mid S_t \right] = 0
\]

\[ S_t, A_t, Y_{t+1}, S_{t+1} \]
Bellman’s equation implies that

\[ E \left[ \frac{\pi_\theta(A_t|S_t)}{\mu(A_t|S_t)} \left( Y_{t+1} - \eta + V(S_{t+1}) - V(S_t) \right) \left( \frac{1}{f(S_t)} \right) \right] \]

will be, for all \( t \), for any vector, \( f(\cdot) \), of appropriately integrable functions, and appropriate distribution expectation, \( E \), equal to 0 if \( \eta = \eta_\theta \), \( V = V_\theta \).
Estimating Function

• Construct a nonparametric model for, $V_\theta(s)$, say $f(s)^T v_\theta$, for $f(s)$ a $p$ by 1 vector of basis functions evaluated at $s$ ($p$ is large)

• Solve

$$\mathbb{P}_n \left[ \sum_{t=1}^{T} \frac{\pi_\theta(A_t|S_t)}{\mu(A_t|S_t)} \left( Y_{t+1} - \eta + f(S_{t+1})^T v - f(S_t)^T v \right) \begin{pmatrix} 1 \\ f(S_t) \end{pmatrix} \right] = 0$$
for $\hat{\eta}_\theta$, $\hat{v}_\theta$
Overview of Algorithm

• The resulting $\eta$ and $\nu$ are functions of $\theta$, denote by $\hat{\eta}_\theta$, $\hat{\nu}_\theta$
  • $\hat{\eta}_\theta$, $\hat{\nu}_\theta$ are the output of the Critic
• The Actor maximizes $\hat{\eta}_\theta$ over $\theta$ to obtain $\hat{\theta}$.
  • this will require repeated calls to the Critic
  • $\hat{\theta}$ is the output of the Actor
Actor

- The objective function for the actor is given by

\[
\hat{\eta}_\theta = \mathbb{P}_n \left[ \sum_{t=1}^{T} \frac{\pi_\theta(A_t|S_t)}{\mu(A_t|S_t)} \left( Y_{t+1} + f(S_{t+1})^T \hat{v}_\theta - f(S_t)^T \hat{v}_\theta \right) \right]
\]

- We want to construct a policy, \( \pi_\theta \) that is bounded away from 0, 1.

Binary action:

\[
\pi_\theta(a|s) = \frac{e^{\theta^T g(s) a}}{1 + e^{\theta^T g(s)}}
\]
Chance constraint on $\theta$:

$$T^{-1} \sum_{t=1}^{T} P^* \left[ p_0 \leq \pi_{\theta}(a|S_t) \leq 1 - p_0 \right] \geq 1 - \alpha$$

for all actions, $a$ and for $P^*$, a reference distribution.

- This constraint is nonconvex; we relax via Markov inequality.
Write the estimating function as,

\[
\mathbb{P}_n \left[ \sum_{t=1}^{T} \frac{\pi_t(A_t|S_t)}{\mu(A_t|S_t)} (Y_{t+1} - \eta + f(S_{t+1})^T v - f(S_t)^T v) \left( \begin{array}{c} 1 \\ f(S_t) \end{array} \right) \right] = \hat{A}_\theta \left( \begin{array}{c} \eta \\ v \end{array} \right) - \hat{b}_\theta
\]

To accommodate a large feature vector, the critic minimizes

\[
|| \hat{A}_\theta \left( \begin{array}{c} \eta \\ v \end{array} \right) - \hat{b}_\theta ||^2 + \lambda_c ||v||^2
\]

to obtain $\hat{\eta}_\theta$, $\hat{\upsilon}_\theta$
• The actor obtains $\hat{\theta}$ by maximizing

$$\hat{\eta}_\theta = \mathbb{P}_n \left[ \sum_{t=1}^{T} \frac{\pi_\theta(A_t|S_t)}{\mu(A_t|S_t)} \left( Y_{t+1} + f(S_{t+1})^T \hat{v}_\theta - f(S_t)^T \hat{v}_\theta \right) \right]$$

subject to the constraint,

$$\theta^T \Sigma_g \theta \leq \alpha \left( \log((1 - p_0)/p_0) \right)^2$$

$$\Sigma_g = T^{-1} \sum_{t=1}^{T} E^* \left[ g(S_t) g(S_t)^T \right]$$
Constructing Policies from Training Data

• We propose an off-line, off-policy actor critic algorithm for learning a treatment policy from a training set.
  – This treatment policy will be a warm-start policy for an online learning algorithm
• Any method should provide confidence intervals/permit scientists to test hypotheses.
• Computational problems……
Challenges

• How to accommodate/utilize the vast amount of missing data, some of which will be informative.
  – This must be done both for the batch, off-line setting and for online learning.

• How to reduce the amount of self-report data (there are statistical approaches to do this)

• Development of multiple risk predictors both in batch and online setting (including risk for disengagement)

• Measuring burden without causing burden.
Collaborators
Actor

• This chance constraint can be further relaxed to a convex constraint on space of $\theta$ by noting

$$1 - T^{-1} \sum_{t=1}^{T} P^* \left[ p_0 \leq \pi_\theta (a | S_t) \leq 1 - p_0 \right]$$

$$\leq \frac{\theta^T T^{-1} \sum_{t=1}^{T} E^* \left[ g(S_t) g(S_t)^T \right] \theta}{\left( \log((1 - p_0)/p_0) \right)^2}$$

• Our constraint:

$$\alpha \geq \frac{\theta^T T^{-1} \sum_{t=1}^{T} E^* \left[ g(S_t) g(S_t)^T \right] \theta}{\left( \log((1 - p_0)/p_0) \right)^2}$$
Implementation

To approximate the differential value, $V_\theta(s)$, $s=(s_1,\ldots,s_{p_1})$, we use features that are all singletons and pairwise products of piecewise linear splines in the set: $\{(s_j - c_{j,k})^+, (c_{j,k} - s_j)^+\} \ j=1,\ldots,p_1, \ k=1,\ldots,10$.

Thus the dimension of the feature vector, $f(s)$, is $\approx 600p_1^2$. 
Implementation

The class for $\pi_\theta$ consists of

$$\pi_\theta(a|s) = \frac{e^{(\theta_0 + \theta_1 g_1 + \ldots + \theta_q g_q)a}}{1 + e^{\theta_0 + \theta_1 g_1 + \ldots + \theta_q g_q}}$$

$g_j$ are features; $q$ is small— in our examples $q=3$

The constraint ($p_0=\alpha=.05$)

$$\theta^T \Sigma_g \theta \leq .43$$

$$\Sigma_g = T^{-1} \sum_{t=1}^{T} \mathbb{P}_n [g(S_t) g(S_t)^T]$$