Micro-Randomized Trials

Susan A Murphy
Outline

- **HeartSteps**
- **Elements of a Micro-Randomized Trial (MRT)**
- **Inferential Target**
- **Sample size considerations**

Using data to inform the development of technological interventions
HeartSteps Goals

• Help individuals increase—and sustain—their physical activity levels
• Increase activity by supporting *opportunistic* physical activity—activity that people can do throughout the day
Types of mHealth Intervention Components
Pull Interventions

Made available to individuals via the phone but accessed at will

- Graphs and charts for self-monitoring
- Coping strategies, educational materials
- “Help” button to receive coping support
Pull Interventions

- Allow inclusion of many components
- Put user in control of access
- Low burden

But…

- Depend on individuals to know when to access them and remember to do it
Push Interventions

Delivered based on time, individual’s context and activities

• Reminders
• Suggestions, tips, motivational messages
• Prompts to set goals, complete self-report…
• Rewards for goal attainment
Push Interventions

• Can use sensing and modeling to determine right context for delivery
• Don’t rely on individual’s awareness of times of need or remembering to access

But…

• High burden
HeartSteps Design Goal

Develop a mobile app intervention that includes the right combination of...

• pull interventions
• push components, *delivered at the right times* to encourage activity throughout the day, as context changes
HeartSteps V1

Push components:

- Actionable, context-aware suggestions for walking
- Planning of when, how, and where one will be active the next day
Suggestions tailored on:

- time of day
- weekday vs. weekend
- location
- weather

Two types of suggestions:

- to walk
- to interrupt sitting
Questions to Optimize HeartSteps’ Suggestions

• Do tailored activity suggestions have an effect at all?
• Do walking and anti-sitting suggestions work equally well?
• Does the effect of suggestions change over time? (e.g., do people get tired of them after a while?)
• When should we send suggestions for optimal effect?
  – Do they work better during certain parts of the day?
  – Do they work better when weather is good vs. bad?
• Is the suggestion effectiveness, including contexts in which they work, different for different types of people?
Outline

- HeartSteps
- Elements of a Micro-Randomized Trial (MRT)
- Inferential Target
- Sample size considerations

Using data to inform the development of technological interventions
Data from wearable devices that sense and provide treatments

• On each individual: $O_1, A_1, Y_2, \ldots, O_t, A_t, Y_{t+1}, \ldots$

• $t$: Decision point
• $O_t$: Observations at $t^{th}$ decision point
• $A_t$: Intervention option (treatment) at $t^{th}$ decision point
• $Y_{t+1}$: Proximal outcome
Micro-Randomized Trial Elements

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

HeartSteps: Approximately every 2-2.5 hours
Micro-Randomized Trial Elements

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, self report each evening……
Micro-Randomized Trial Elements

3) Intervention options $A_t$
   1) Types of treatments/engagement strategies that can be provided at a decision point, $t$
   2) Whether to provide a treatment

**HeartSteps**: tailored activity suggestion (yes/no)
Availability

- Intervention options can only be delivered at a decision point if an individual is *available*
- Availability is known prior to delivering a treatment
- Set $I_t=1$ if the individual is available at decision point $t$, otherwise, $I_t=0$

Availablity is not the same as adherence, nor is it the same as interruptibility, receptivity
Micro-Randomized Trial Elements

4) Proximal Outcome $Y_{t+1}$

**HeartSteps**: Activity (step count) over next 30 minutes.
Proximal outcomes are usually mediators thought to be critical to achieving the long-term behavior goal

1) Short term targeted behavior
   - Substance use over x hours
   - Physical activity over x minutes
   - Adherence over next hour

2) Short term risk
   - Current craving, stress

3) Engagement with mobile app/intervention burden
Micro-Randomized Trial

Randomize between intervention options at decision points ➔ Each person may be randomized 100’s or 1000’s of times.

- These are sequential, “full factorial,” across time, designs.
Micro-Randomized Trial Elements

1. Record outcomes
   – Proximal Outcome & Distal Outcome
2. Record context (sensor & self-report data)
3. Randomize among intervention options at decision points
4. Use data after study ends to assess treatment effects, develop JITAI
Why Micro-Randomization?

- Randomization + representative sample is the gold standard in providing data to assess causal effects.

- Sequential randomizations + representative sample will enhance replicability of data analyses (moderation by context and past dose).
Outline

• HeartSteps
• Elements of a Micro-Randomized Trial (MRT)
• Inferential Target
• Sample size considerations

Using data to inform the development of technological interventions
Micro-Randomized Trial

How to justify the experimental 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 intervention options differentially impact the proximal outcome? (aka, is there a main effect?)
Micro-Randomized Trial for HeartSteps

• 42 day trial
• Whether to provide a tailored activity recommendation? $A_t \in \{0, 1\}$
• Test for main effects on proximal outcome
• Randomization in HeartSteps

$$P[A_t = 1] = .4 \quad t = 1, \ldots, T = 210$$
Time-varying Main Effects

Time varying potentially intensive/intrusive treatments → potential for accumulating habituation and burden

→

In the test statistic allow the main effect of the treatments on proximal outcome to vary with time
Availability & the Treatment Effect

- Interventions can not be delivered at a decision point if an individual is *unavailable*.

- The effect of an intervention at a decision point is the difference in proximal outcome between *available* individuals assigned an activity suggestion and *available* individuals who are not assigned an activity suggestion.
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

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

• What does this main effect mean?
The randomization implies that the main effect can be written as

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

• HeartSteps
• Elements of a Micro-Randomized Trial (MRT)
• Inferential Target
• Sample size considerations

Using data to inform the development of technological interventions
MRT Sample Size

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

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 participants to test $H_0$: no effect of the intervention, i.e., 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 suggestion uses contrasts of *between subject outcomes* + contrasts of *within subject outcomes*.

--The required number of subjects will be smaller than a two group randomized trial.
Test Statistic for Sample Size Calculation

Test statistic is based on a 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

- We are not assuming this “model” is correct............
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:

randomization probability, \( q_t = 0.4 \)

\[ \beta(t) = \beta_0 + \beta_1 day_t + \beta_2 day_t^2 \]

\[ \gamma(t) = \gamma_0 + \gamma_1 day_t + \gamma_2 day_t^2 \]
Alternative

• 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.
Specify Alternative

- Specify standardized $d_i$’s by
  - initial standardized main effect: $d_0$,
  - average standardized main effect over trial duration =
    \[
    \frac{1}{T} \sum_{t=1}^{T} \left( d_0 + d_1 \text{day}_t + d_2 \text{day}_t^2 \right),
    \]
  - and day of maximal main effect: $- \frac{d_1}{2d_2}$
- We solve for $d_0$, $d_1$, $d_2$
HeartSteps

Standardized $d_i$'s

- initial effect: $d_0 = 0$

- output average standardized main effect

- day of maximal main effect: $-\frac{d_1}{2d_2} = 28$
### HeartSteps Sample Sizes

**Power=.80, False-positive error=.05**

<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>
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>
HeartSteps V1

Heartsteps MRT to Promote Physical Activity Among Sedentary People

Each day of study
Observations are continuous (except self report)
Randomizations to activity prompts occur 5x/ day at likely times for increasing physical activity

Next 30 minutes after intervention is delivered
Measured via accelerometer throughout study

Observations
- location (via GPS)
- weather (via internet)
- motion (via wrist band)
- usefulness of prompt (via user indication)
- self report of activity (via app in evenings)

Driving? NO
Driving? YES
Walking? NO
Walking? YES
R
R
R

Start Intervention
Prompt planning of next day's activity

Average 3x/day

Average 2x/day

Start Intervention
Tailored prompt to become physically active

Proximal Outcome
physical activity (steps taken)

Proximal Outcome
physical activity (steps taken)

Distal Outcome
Overall activity in the 42-day study

Following day

PI: P Klasnja
Location: University of Michigan
Funding: NHLBI/NIA R01HL125440
Planning a Micro-Randomized Trial?

Be conservative in planning the trial:

1) Under-estimate the amount of time participants are available for the intervention.
2) Under-estimate the average standardized effect
MRTs vs Other designs

- RCT
- N-of-1 Trials (& Crossover Trials)
- Factorial Designs
A randomized control trial (RCT) evaluating a JITAI compared to a suitable control.

– Assumes evidence exists to develop a high-quality JITAI including the
  • choice of tailoring variables & decision rules

– The primary aim of an RCT is to confirm the JITAI’s effectiveness compared to an alternative
  • is not well suited to constructing or optimizing a JITAI

– RCT is optimal for evaluation
MRT vs. N-of-1 Trial in Clinical Setting

N-of-1 Trials are usually multiple cross-over trials in which the order of the treatments are randomized within a person.

- RCT is too expensive or not feasible
  - Test: Is one-time treatment A better than one-time treatment B?
  - Ideally the treatments should have minimal delayed effects (minimal carryover effects) or N-of-1 design should incorporate a suitable washout period

MRT vs. N-of-1 Trial in Behavioral Science

N-of-1 Trials are usually multiple cross-over trials in which the order of the treatments are randomized within a person.

- Goal is to ascertain individual level causal effects
  - Test: Is one-time treatment A for Jane better than one-time treatment B for Jane
  - Nuanced assumptions about individual behavior based on theory are brought to bear in order to triangulate on individual level effects.

McDonald et al., 2017
MRTs vs Factorial Experiments

A factorial design

- is an experimental design involving more than one components (e.g., factors); the levels of the components can be meaningfully crossed.

A MRT

- is a special form of a factorial; components are employed sequentially in time within a person.
- components can operate at different time scales
- randomization to subsequent components in a MRT may depend on outcomes of prior components
MRTs vs Factorial Experiments

Components can be randomized at different time scales, e.g. in HeartSteps:

Factor 1: Tailored activity suggestion is randomized 5 times per day (yes/no)

Factor 2: Daily activity planning is randomized each evening (yes/no)
Micro-Randomized Trials: When are they (not) useful?

- **NOT USEFUL:** When malleable circumstances are rare: Want to learn the best type of alert to prevent suicide attempt
- **USEFUL:** When malleable circumstances change rapidly: Stress, urges to smoke, adherence, physical activity, eating
- **NOT USEFUL:** Proximal outcome cannot be feasibly assessed.
- **USEFUL:** Proximal outcome can be unobtrusively sensed or unobtrusively self-reported.
Micro-randomized Trials

A new type of factorial design

i. Time varying factors → time varying main effects, time-varying two-way interactions, different delayed effects

ii. Calculator:
https://sites.google.com/a/umich.edu/pengliao/

iii. Check out other MRTs (!!) at
https://methodology.psu.edu/ra/adap-inter/mrt-projects#proj
Collaborators!

http://people.seas.harvard.edu/~samurphy/