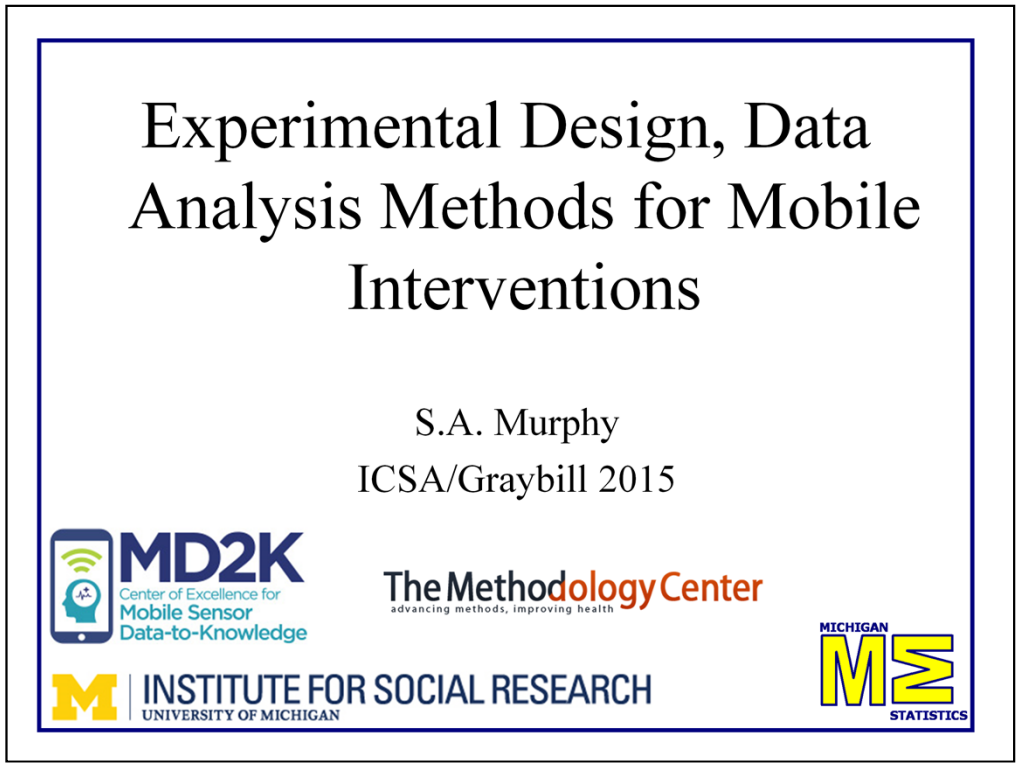


Title: Micro-randomized Trials for Just-In-Time Adaptive Intervention Development

Summary: Micro-randomized trials are trials in which individuals are randomized 100's or 1000's of times over the course of the study. The goal of these trials is to assess the impact of momentary interventions, e.g. interventions that are intended to impact behavior over small time intervals. We discuss the design and analysis of these types of trials with a focus on their use in developing JITAIs in mobile health.

mHealth

- Goal: Design a Continually Learning Mobile Health Intervention
- Example: "HeartSteps"

+

- "Micro-Randomized" Trial

2

health domains

Behavior change and maintenance of this change (exercise, healthy eating, sedentary behavior)

Self-management of a chronic disorder (Adherence to meds, adherence to self-care behaviors, mental illness, cognitive support, substance abuse)

Data from wearable devices that sense and provide treatments

$$S_1, A_1, Y_2, \ldots, S_j, A_j, Y_{j+1}, \ldots$$

S_j: State at j[th] decision time (high dimensional)

A_j : Action at j[th] decision time (treatment)

Y_{j+1} :Proximal Response (time-varying response)

3

Contents of the state are a modeling decision--State is not only what you collect between past decision time and present decision time. It includes summaries of history that are allow you to assume Markovian structure. Can include time of day or day of week and present weather. Volatility in craving over last x days

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.

4

The momentary times were selected because these times are the times at which most people are able to be active
Pre-morning commute, mid-day, mid-afternoon, evening commute, after dinner.

Another example: The phone software monitors a risk measure at regular time intervals and if the risk measures hits a criterion then a treatment is provided.

Examples

2) State S_j
   1) Passively collected (location, stress, 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.

State is not only what you collect between past decision time and present decision time. It includes summaries of history Can include time of day or day of week and present weather.

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

Some Treatment types behavioral, cognitive, motivational, social, self-monitoring, information, alerts

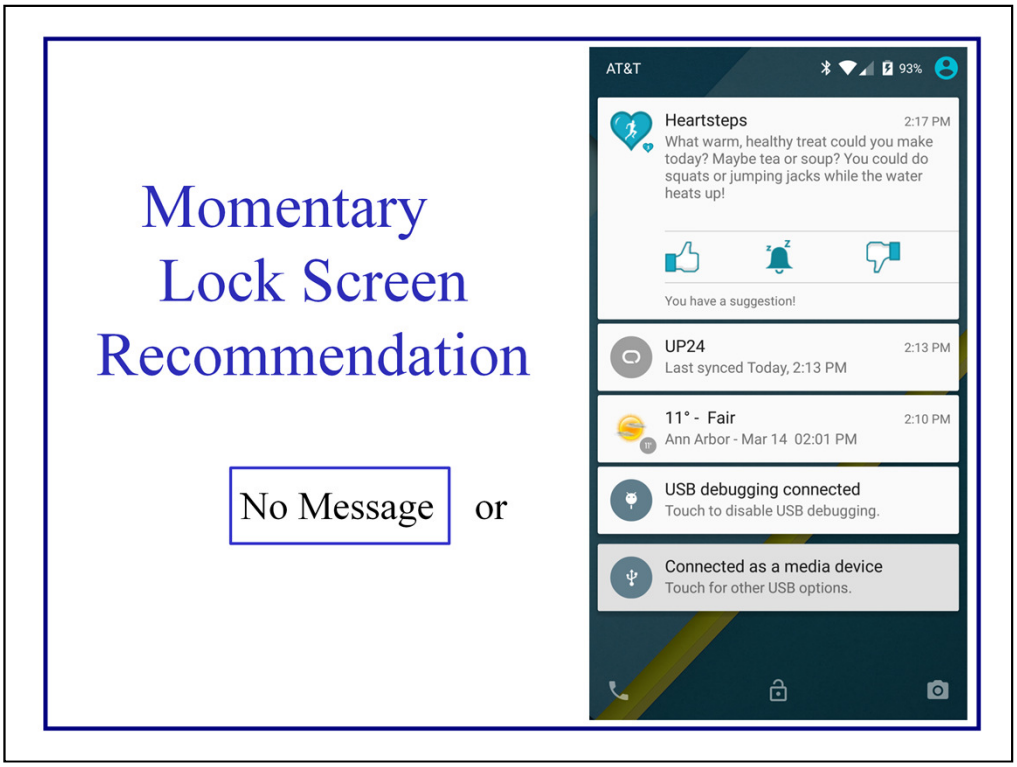
Examples

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

Some Treatment types behavioral, cognitive, motivational, social, self-monitoring, information



The location of the like button biases against the person hitting like.

The snozz button turns off the momentary lock screen recommendations for x hours.

Occurs up to 5 times per day

Examples

4) Proximal Response Y_{j+1}

HeartSteps: Activity (step count) over next 60 minutes.

Frequently the actions are primarily designed to have a near-term effect on the individual. E.g. Help then manage current craving/stress, help them manage or be aware of the impact of their social setting on their craving/stress

Our Group's Scientific Goals

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

2) Develop methods for assessing if there are delayed effects of the actions; assess if the proximal or delayed effects vary by particular state variables.

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

4) Develop online training algorithms that will result in a "Continually Updating" Treatment Policy

10

Proposed Experimental Design: 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.

11

42*5=210 times in pilot planned study 2160 decision times.

People wanting to build a dynamical system model might use "multisine excitation"

See the papers

Multisine Signals for Wireless System Test and Design, Nuno B. Carvalho, Kate A. Remley, Dominique Schreurs, and Kevin G. Gard. June 2008 ieee microwave magazine.

Evaluation of Simultaneous Multisine Excitation of the

Joined Wing SensorCraft Aeroelastic Wind Tunnel Model, Jennifer Heeg1 and Eugene Morelli2 American Institute of Aeronautics and Astronautics

Why Micro-Randomization?

- Factorial designs are the gold standard when collecting data to build a treatment involving many components
- Actions are often intended to have a proximal effect.
  - Randomization (+ representative sample) is a gold standard in providing data to assess a causal effect
- Sequential randomization will enhance quality of many interesting subsequent data analyses.

12

Justifying the Sample Size for a Micro-Randomized Trial

- Focus on whether to provide a Momentary Lock Screen Recommendation, e.g.

$$A_j \in \{0, 1\}$$

- Randomization in HeartSteps

$$P[A_j = 1] = .4 \quad j = 1, \ldots, J$$

- Size to Detect a Proximal Effect

13

planned study 2160 decision times.

J=42*5=210 momentary randomizations

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 $I_j = 0$

14

The momentary intervention can be turned off for 4 or 8 hours by the participant. The intervention is also off if the participant is classified as currently active (e.g. walking) or classified as currently driving.

Potential Outcomes

- Define

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

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

- Define $I_j(\bar{a}_{j-1})$ to be the observed "intervention on" indicator if $\bar{A}_{j-1} = \bar{a}_{j-1}$

This use of potential outcome notation is imprecise but easier to present than the more accurate contained in our paper.

Proximal Main Effect

- Define the Proximal Main Effect at time j as

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

- What does this estimand mean?

16

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.

Proximal Main Effect

- The randomization implies that

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

- Put

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

17

Assuming "consistency" THESE ARE CAUSAL EFFECTS!!!!!

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.

18

Main effects are marginal effects!

Delayed effects which are akin to higher order interactions would be investigated in secondary analyses

Test for Parametric Alternative

- We construct a test statistic for testing $H_0 : \beta(j) = 0, \forall j$
- HeartSteps parametric alternative:

$$\beta(j) = \beta_0 + \beta_1 \lfloor \tfrac{j-1}{5} \rfloor + \beta_2 \lfloor \tfrac{j-1}{5} \rfloor^2$$

$\rightarrow$ test

$$H_0 : \beta_i = 0, i = 0, 1, 2$$

19

Since the model for the proximal effect of Aj on Yj does not depend on time of day, we are averaging any variation

in proximal effect across the occasions during the day (recall we are sizing the study; a primary analysis might be a little more complex and in secondary data

analyses one would likely estimate and test if the proximal effect varies by time of day and/or varies by j,

since j denotes duration in study).

Test Statistic for Sample Size Calculation

- Test statistic based on fitted model:

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

where q_j is the randomization probability

- $q_j = .4$ in HeartSteps

20

Centering of A_j protects estimation of beta(j) from misspecification of \gamma(j). --This works for consistency of \hat\beta(j) and variance estimation in the data analysis but when we do the sample size formula we simplify the variance—the simplification assumes \gamma is correctly specified.

Test Statistic for Sample Size Calculation

- Test statistic is based on least squares fit of

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

HeartSteps:

$$\beta(j) = \beta_0 + \beta_1 \lfloor \tfrac{j-1}{5} \rfloor + \beta_2 \lfloor \tfrac{j-1}{5} \rfloor^2$$

- You select parameterization of $\gamma(j)$

21

least squares acting as if time points within a person are independent.

Alternative for Sample Size Calculation

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

- Alternative:

$$H_1 : \beta_i = d_i \sigma, i = 0, 1, 2$$

where σ^2 is the conditional variance.

22

Alternative for Sample Size Calculation

- Conditional variance is

$$\sigma^2 = VAR\left(Y_{j+1} \middle| I_j = 1, A_j\right)$$

23

Specify Alternative for Sample Size Calculation

- Scientist indirectly specifies standardized d_i's
  - initial proximal treatment effect: d_0,
  - average proximal effect over trial duration:

    $\frac{1}{J}\sum_{j=1}^{J}\left(d_0 + d_1\lfloor\frac{j-1}{5}\rfloor + d_2\lfloor\frac{j-1}{5}\rfloor^2\right)$,

  - and day of maximal proximal effect: $-\frac{d_1}{2d_2}$
- We solve for d_0, d_1, d_2

24

Specify alternative so that scientist can provide d_i's

J=5*7*6 in our study

Test Statistic for Sample Size Calculation

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

p is the total number of parameters ($p > 3$);

X_i is the associated design matrix (J by p)

N is sample size

Last 3 columns of X_i contain row entries:

$I_{ij}(A_{ij} - q_{ij}), I_{ij}(A_{ij} - q_{ij})\lfloor \frac{j-1}{5} \rfloor,$

$I_{ij}(A_{ij} - q_{ij})\lfloor \frac{j-1}{5} \rfloor^2$

least squares acting as if time points within a person are independent.

Test Statistic for Sample Size Calculation

- "GEE" test statistic is

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

where $\hat{\Sigma}$ is the usual sandwich estimator of the variance-covariance and K is 3 by p matrix picking out columns associated with coefficients β

26

Sample Size Calculation

Under standard moment assumptions, the asymptotic distribution is a Chi-Squared on 3 degrees of freedom with non-centrality parameter: $d^T(\Sigma_\beta)^{-1}d$

27

\Sigma_\beta = (Z^T Z)^1 divided by q(1-q)\tau Z is number of time points by 3. The 3 columns correspond

to the columns [A_j-.4, (A_j-.4) \lfloor\frac{j-1}{5}, (A_j-.4) \lfloor\frac{j-1}{5}^2] in the design matrix.

Working Assumptions for Sample Size Calculation

1) The model errors, $(\epsilon_{ij}, \epsilon_{ik})$ are uncorrelated with the treatments (A_{ij}, A_{ik}) given $(I_{ij} = 1, I_{ik} = 1)$

2) $E(\epsilon_{ij} \mid I_t = 1, A_t) = 0$

3) $Var(\epsilon_{ij} \mid I_t = 1, A_t)$ is constant.

$$\epsilon_{ij} = Y_{ij+1} - \Big(\gamma(j) + \beta(j)(A_{ij} - q_{ij})\Big)$$

Sample Size Calculation

- Under the working assumptions, Σ_β only depends on polynomials in $\lfloor \frac{j-1}{5} \rfloor$, the marginal distribution of I_j and on the randomization probabilities.
- Σ_β does not depend on the form of $\gamma(j)$

29

\Sigma_\beta = (Z^T Z)^1 divided by q(1-q)\tau Z is number of time points by 3. The 3 columns correspond

to the columns [A_j-.4, (A_j-.4) \lfloor\frac{j-1}{5}, (A_j-.4) \lfloor\frac{j-1}{5}^2] in the design matrix.

Sample Size Calculation

Alternative proximal main effect is parameterized →

estimation uses both between person contrasts as well as *within person contrasts* →

sample sizes are small.

30

The contrasts become within person contrasts due to the assumption of smoothness over time. If the proximal effect at each time were to be estimated separately then it would be like a two arm study at each time j.

\Sigma_\beta = (Z^T Z)^1 divided by q(1-q)\tau Z is number of time points by 3. The 3 columns correspond

to the columns [A_j-.4, (A_j-.4) \lfloor\frac{j-1}{5}, (A_j-.4) \lfloor\frac{j-1}{5}^2] in the design matrix.

Sample Size Calculation

- Instead of a Chi-Squared on 3 degrees we use $\frac{3(N-p+2)}{N-p} F_{p,N-p}$ with the same noncentrality parameter $d^T(\Sigma_\beta)^{-1}d$

31

p is the number of parameters in linear model. These are the degrees of freedom recommended by

Lloyd A. Mancl and Timothy A. DeRouen A Covariance Estimator for GEE with Improved Small-Sample Properties BIOMETRICs 57 , 126-134 March 2001

\Sigma_\beta = (Z^T Z)^1 divided by q(1-q)\tau Z is number of time points by 3. The 3 columns correspond

to the columns [A_j-.4, (A_j-.4) \lfloor\frac{j-1}{5}, (A_j-.4) \lfloor\frac{j-1}{5}^2] in the design matrix.

HeartSteps Example

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

$$\gamma(j) + \beta(j)(A_{ij} - .4), \; j = 1, \ldots, 42$$

where

$$\gamma(j) = \gamma_0 + \gamma_1 \lfloor \tfrac{j-1}{5} \rfloor + \gamma_2 \lfloor \tfrac{j-1}{5} \rfloor^2$$

32

Specify alternative so that scientist can provide d_i's

42 days in our study

Sample Sizes, Power=.8, α=.05

Standardized Average Proximal Effect $\frac{1}{J}\sum_{j=1}^{J}\left(d_0 + d_1\lfloor\frac{j-1}{5}\rfloor + d_2\lfloor\frac{j-1}{5}\rfloor^2\right)$	Sample Size For *E[I]=.7 or .5*
0.06	81 or 112
0.08	48 or 65
0.10	33 or 43

33

Average proximal effect is standardized.

#parameters=6

Meaningful increase in stepcount is 1000/day

Usual std is 2000/day

Roughly a standardized treatment effect of 200/666= .3

Primary Data Analysis

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

p is the total number of parameters ($p > 3$);
X_i is the associated design matrix (J by p)
N is sample size

Last 3 columns of X_i contain row entries:
$I_{ij}(A_{ij} - q_{ij}), I_{ij}(A_{ij} - q_{ij})\lfloor\frac{j-1}{5}\rfloor,$
$I_{ij}(A_{ij} - q_{ij})\lfloor\frac{j-1}{5}\rfloor^2$

34

GEE here is least squares acting as if time points within a person are independent.

Test Statistic

- "GEE" test statistic is

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

where K is 3 by p matrix picking out columns associated with β coefficients

35

Small Sample Adjustment

- $\hat{e}_{ij}$ is the i^{th} subject, j^{th} time point residual and $\hat{e}_i = (\hat{e}_{i1}, \ldots, \hat{e}_{iJ})^T$

- Adjusted sandwich estimator:

$$\hat{\Sigma} = \hat{\sigma}^2 N \Big(\sum_{i=1}^{N} X_i^T X_i \Big)^{-1} \Big\{ \sum_{i=1}^{N} X_i^T B_i \hat{e}_i \hat{e}_i^T B_i X_i \Big\} \Big(\sum_{i=1}^{N} X_i^T X_i \Big)^{-1}$$

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

36

GEE here is least squares acting as if time points within a person are independent.

Lloyd A. Mancl and Timothy A. DeRouen A Covariance Estimator for GEE with Improved Small-Sample Properties BIOMETRICs 57 , 126-134 March 2001

The resulting estimator of the variance of \hat\beta is consistent. We are not assuming that \gamma(j) is correct.

Simulation Results
Type 2 Error Rate (2000 data sets)

Average Proximal Effect (Sample Size)	Power
0.05(115)	0.790
0.06(81)	0.794
0.07(61)	0.800
0.08(48)	0.801
0.09(39)	0.798
0.10(33)	0.803

E[I]=0.7, no initial effect, maximal effect on day 29,randomization prob.=0.4

Simulations indicate:

Method is sensitive to

Guess of average amount of time intervention is on: 1/J \sum_{j=1}^J E[I_j]. Choose on the low side to be safe

Guess of average proximal txt effect. Choose on the low side to be safe.

Heteroscedasticity of errors variance of Y when A=1 is larger than variance of Y when A=0 is problematic

Simulations indicate robustness to

*I*_{j+1} a function of past A_j's

Guess at day of maximal proximal effect (we use different function from quadratic when this day is less than ½ of the way through the study—this is not presented here)

Non-symmetry or skewness to residual error distribution .

Positive correlated across time residual errors

Mixture of people, some of whom have the intervention turned off x % of time and some who have their intervention turned off y% of the time where overall % time turned off is .7 or .5

Scientific Goals

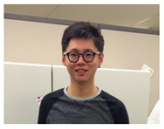
1) Develop methods/trial designs for assessing if there are proximal causal effects of the actions on the response.

2) Develop methods for assessing if there are delayed causal effects; assess if the proximal or delayed causal effects vary by particular state variables.

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

4) Develop online training algorithms that will result in a "Continually Updating" treatment policy

38



Collaborators: P. Liao, A. Lee, C. Anderson, P. Klasnja & A. Tewari


Sample size calculator at https://jisun.sinyapps.io/SampleSizeCalulator/

Email if you have questions!

samurphy@umich.edu