Micro-Randomized Trials
&
mHealth

S.A. Murphy
IMPACT
11/2014

Institute for Social Research
Survey Research Center
SRC

MICHIGAN
STATISTICS

mHealth

- Goal: Design a Continually Learning Mobile Health Intervention: “HeartSteps”

+

- “Micro-Randomized” Trial

Data from wearable devices that sense and provide treatments

$$S_1, A_1, Y_1, \ldots, S_j, A_j, Y_j, \ldots$$

S_j: State at j^{th} decision time (high dimensional)

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

Y_j : Proximal Response (time-varying response)

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) <u>Momentary</u>: Approximately every 2-2.5 hours
2) <u>Daily</u>: Each evening at user specified time.

Examples

2) State S_j
   1) Passively collected (location, weather, busyness of calendar, social context, activity on device)
   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 a decision time
   2) Whether to provide a treatment

HeartSteps includes two types of treatments

1) <u>Momentary</u> Lock Screen Recommendation
2) <u>Daily</u> Activity Planning

Examples

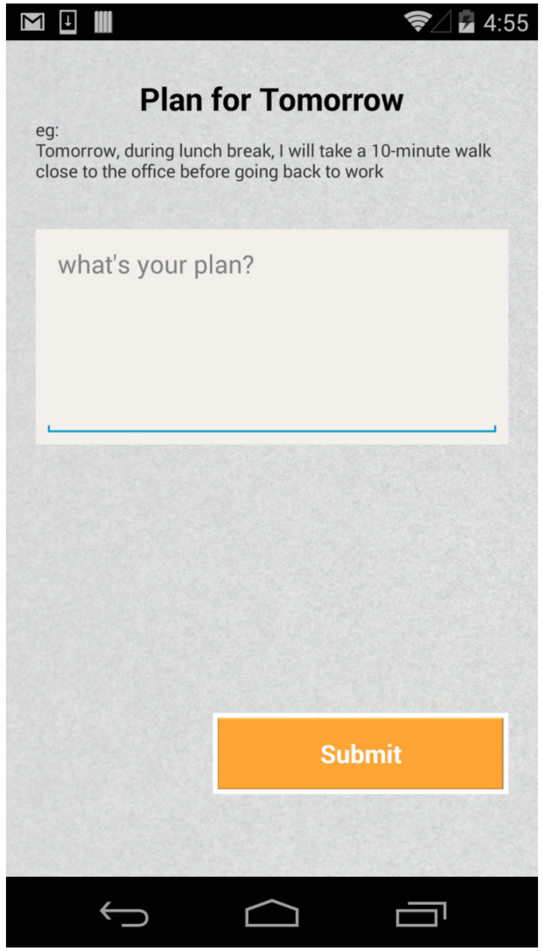
3) Actions A_j
    1) Treatments that can be provided at decision time
    2) Whether to provide a treatment

HeartSteps includes two types of treatments

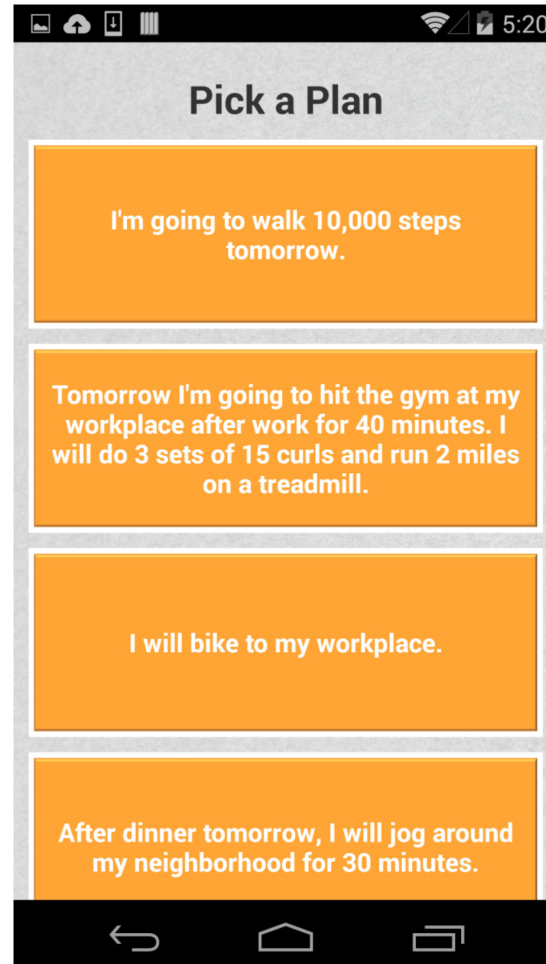
1) <u>Momentary</u> Lock Screen Recommendation
2) <u>Daily</u> Activity Planning

Daily Activity Planning

No Plan or



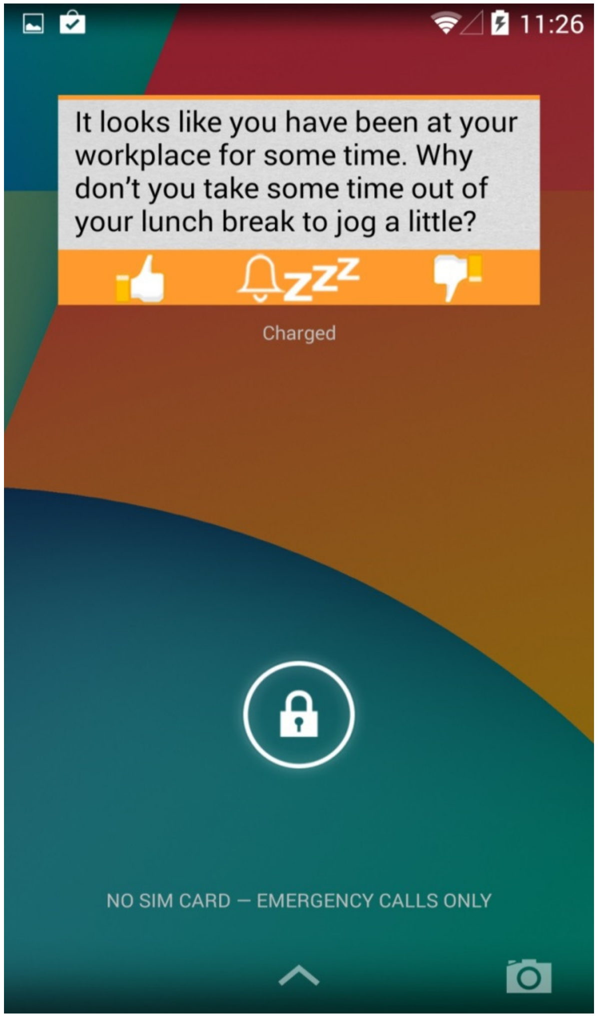
or



Momentary Lock Screen Recommendation

No Message

or



Examples

4) Proximal Response Y_j

HeartSteps: Activity (step count) over next 60 minutes between decision times or daily activity.

Our Group's 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" Personalized Treatment Policy

Today’s Focus

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” Personalized Treatment Policy

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.

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.

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

Proximal Causal Effect

- Recall that Y_j is the proximal response (activity level) recorded after action A_j

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

- Set $R_j = 1$ if the momentary intervention is on at time j, otherwise $R_j = 0$

Potential Outcomes

- Define

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

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

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

Proximal Causal Effect

- Define the Proximal Causal Effect at time j as

$$E\Big[Y_j(\bar{A}_{j-1},1) - Y_j(\bar{A}_{j-1},0)\Big|R_j(\bar{A}_{j-1}) = 1\Big]$$

- What does this estimand mean?

Proximal Causal Effect

- The randomization implies that

$$E\big[Y_j(\bar{A}_{j-1},1) - Y_j(\bar{A}_{j-1},0)\big|R_j(\bar{A}_{j-1}) = 1\big] = \\ E\big[Y_j|R_j = 1, A_j = 1\big] - E\big[Y_j|R_j = 1, A_j = 0\big]$$

- Put

$$\beta(j) = E\big[Y_j|R_j = 1, A_j = 1\big] - E\big[Y_j|R_j = 1, A_j = 0\big]$$

Proposal

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

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

Test for Sample Size Calculation

- We construct a test statistic for $H_0 : \beta(j) = 0, \forall j$
- A simple approach is parameterize

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

and test

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

Test Statistic for Sample Size Calculation

- The model

$$E[Y_j | R_j = 1, A_j] = \gamma(j) + \beta(j)(A_j - q_j)$$

where q_j is the randomization probability

- $q_j = .4$ in HeartSteps

Test Statistic for Sample Size Calculation

- Test statistic is based on "GEE" fit of

$$E[Y_j | R_j = 1, A_j] = \gamma(j) + \beta(j)(A_j - q_j)$$

where

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

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

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 residual variance.

Standardization in Sample Size Calculation

- Residual variance is

$$\sigma^2 = VAR\left(Y_j \middle| R_j = 1, A_j\right)$$

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:

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

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

Test Statistic for Sample Size Calculation

- Put $Y_i = (Y_{i1}, \ldots, Y_{iJ})^T$ 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:

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

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

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

Working Assumptions for Sample Size Calculation

1) Within subject, pairwise conditional, no correlation: the model errors, $(\epsilon_{ij}, \epsilon_{ik})$ are uncorrelated with the treatments (A_{ij}, A_{ik}) given $(R_{ij} = 1, R_{ik} = 1)$.

2) $P(R_{ij} = 1) = \tau$ a constant.

3) Model errors, ϵ_{ij}'s, have mean zero.

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

Sample Size Calculation

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

- Σ_β only depends on polynomials in $\lfloor \frac{j-1}{5} \rfloor$, the distribution of R_j and on the randomization probability.

Sample Size Calculation

- The asymptotic distribution of the test statistic does not depend on the form of $\gamma(j)$
- The asymptotic distribution <u>does</u> depend on the distribution of R_j

Sample Size Calculation

- Because proximal effects are *within person contrasts*, we expect that the sample sizes will be small.

- 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 $d^T(\Sigma_\beta)^{-1}d$

Example

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

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

where

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

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[R]=.7 or .5**
0.06	81 or 112
0.08	48 or 65
0.10	33 or 43

Primary Data Analysis

- Put $Y_i = (Y_{i1}, \ldots, Y_{iJ})^T$ 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:

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

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

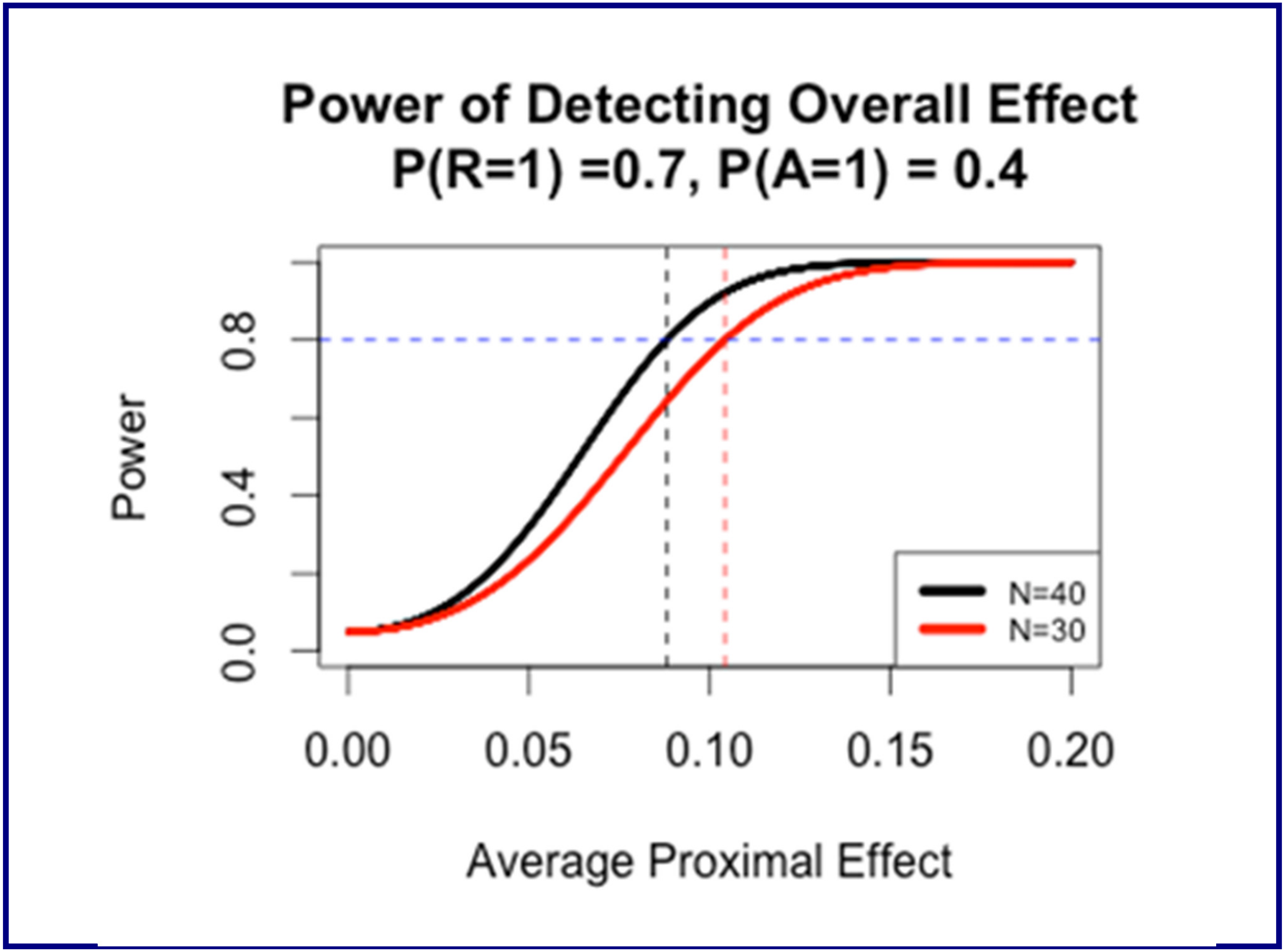
Small Sample Adjustment

- $\hat{e}_{ij}$ 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:

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

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



Power of Detecting Overall Effect
P(R=1) =0.7, P(A=1) = 0.4
Power
0.0
0.4
0.8
0.00
0.05
0.10
0.15
0.20
Average Proximal Effect
N=40
N=30

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

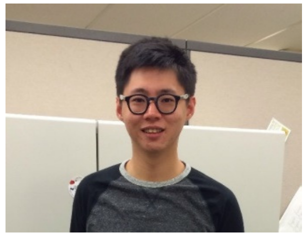
Scientific Goals

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



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


WE ARE SEEKING POSTDOCS!!!

Email if you have questions!

samurphy@umich.edu