

**Micro-randomized Trials  
for Constructing  
Mobile Health  
Interventions**

Susan A Murphy

Title: Micro-randomized Trials for Constructing Mobile Health Interventions

Abstract: Mobile devices along with wearable sensors make possible the ability to deliver treatments anytime and anywhere. These mobile interventions are being employed across a variety of health fields, including to support HIV medication adherence, encourage physical activity and healthier eating as well as to support recovery in addictions. The treatments in the mobile intervention are often time-varying and might be delivered many times across 100s or 1000s or more time points. In this tutorial we discuss a type of factorial trial design, namely the (stratified) micro-randomized trial, for use in constructing mobile interventions. We discuss primary hypotheses for these factorial designs and how to determine the sample size so as to test these hypotheses with a given power. We will also review secondary analyses that can be used to estimate and test interaction effects between time-varying context (e.g., location, current stress classification, time of day, mood and ambient noise) and time-varying treatments. These discussions will be made concrete by using three real-life micro-randomized trials (namely in physical activity, smoking cessation and encouraging adherence) to clarify ideas and analyses.

## The Dream!

“Continually Learning Mobile Health Intervention”

- Help user achieve and maintain user’s desired long term healthy behaviors
  - Provide sufficient short term reinforcement to enhance user’s ability to achieve long term goal
- The ideal mHealth intervention
  - will engage user when needed and do not intrude when not needed.
  - will adjust to unanticipated life challenges

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)

## Outline

- Introduction to mobile health
- Micro-Randomized Trials for Constructing mHealth interventions
  - Main Effects and Sample Size Calculation
- Secondary Analyses
  - Causal Treatment Effects (aka Causal Excursions)
  - Estimation Method
- Example: HeartSteps

3

## HeartSteps (PI Klasnja)

Goal: Develop an mobile activity coach for individuals who have coronary artery disease

Three iterative studies:

- 42 day micro-randomized pilot study with sedentary individuals,
- 90 day micro-randomized study,
- 365 day personalized study



A change in daily habits can significantly reduce a person's risk for heart disease, or, for patients who are already diagnosed, the chance of having an acute coronary event [2,10,84]. Cardiac rehabilitation phase II (CR)—12 to 18-week outpatient programs of supervised exercise, dietary counseling, stress management, and health education—is a central component of secondary prevention of heart disease. Reviews and meta-analyses of cardiac rehabilitation trials have shown that CR can substantially impact physiological and behavioral risk factors—blood pressure, weight, physical activity, smoking, and diet—as well as cardiac mortality and morbidity [21,56,124]. Secondary prevention remains a challenge, however. Although cardiac rehabilitation can jump-start risk reduction, the rates of maintenance of lifestyle changes after cardiac rehabilitation ends have been very low. Fewer than half of patients who complete CR still exercise or eat a low-fat diet six months later [16,105,122]. Yet, after physical activity and heart-healthy diet are discontinued, most of their risk-lowering benefits are lost within a year and a half [77]. *Behavior-change maintenance is, therefore, a key challenge of cardiac risk reduction.*

## HeartSteps



Context provided via data from:

Wearable band → activity and sleep quality;

Smartphone sensors → busyness of calendar, location, weather;

Self-report → stress, user burden

In which contexts should the smartphone provide the user with a tailored activity suggestion?

5

Activity suggestions are to help in your automated processing

To counteract this shortcoming of reflective self-regulatory processes, we will specifically target automated self-regulatory processes as a key part of our maintenance intervention. This is a novel contribution because **most mHealth interventions use reflective self-regulatory strategies almost exclusively [74]**. As we discuss in section C1, in our proposed intervention we will leverage priming [11,35], evaluative conditioning [52], and implementation intentions [41,42] to help individuals engage in opportunistic physical activity and protect planned activities from being derailed when people are stressed, tired, or otherwise depleted. Ours is the first mHealth intervention we know of that supports automated self-regulation in such a comprehensive way.

How those perceptions shift as people's goals and circumstances change over time has not been explored, however. Through the novel use of a micro-randomization methodology (see section C.3), this project will significantly advance the knowledge in this area by enabling us to analyze how the frequency and nature of different intervention components, such as prompts to self-monitor, planning activities, and activity suggestions displayed on the phone's lock screen influence, over time, not only physical activity but also user burden and engagement. Furthermore, the micro-randomization will facilitate generation and testing of theory-based hypotheses about how these effects of time-varying intervention components (e.g., whether or not to provide an activity suggestion), and their

different doses, are moderated by aspects of the patients' lifestyle and context, such as the consistency in their physical-activity routines, work schedule, etc.

## Sense2Stop



Wearable wrist/chest bands → multiple physiological and motion sensor streams (stress, event detection)

Self-report → smoking, stress, user burden

In which contexts should the wrist band provide reminder to access stress-regulation apps?

6

three available stress-regulation apps (headspace; mood-surfing; thought distancing).

Currently, AutoSense consists of an arm band with four wireless sensors and a chestband with six wireless sensors. All the ten sensors are integrated onto an embedded platform called “mote,” a tiny self-contained, battery-powered computer with a wireless radio that can host multiple sensors, collect and process data from them using customized algorithms, and communicate on secure wireless channels.

The chestband consists of 2-lead Electrocardiogram (ECG), galvanic skin response (GSR), respiratory inductive plethysmograph (RIP) band for robust measurement of respiration, skin temperature, ambient temperature, and a 3-axis accelerometer. Accelerometers help classify physical activities, estimate their intensities, and help remove motion artifacts from the measurements of ECG, RIP, and GSR. The armband consists of WrisTAS alcohol sensor from Giner Inc., accelerometers, GSR, and temperature sensors. All sensors communicate wirelessly with a smart phone. Sensors on the phone (e.g., GPS, microphone) complement those on the body. The phone also acts as a local server for heavier computation and storage.

A person is available if he/she (1) is wearing autosense (our understanding is that autosense will be worn up to 16 hours a day, participants will not wear it when they

sleep or when in the shower; if the person is not wearing Autosense, no data will be collected and recommendations will not be pushed); (2) did not receive a message in the past  $x$  minutes; and (3) is not driving a car.

## Data from wearable devices that sense and provide treatments

- On each individual:  $O_1, A_1, Y_2, \dots, O_t, A_t, Y_{t+1}, \dots$
- $t$ : Decision time
- $O_t$ : Observations at  $t^{\text{th}}$  decision point (high dimensional)
- $A_t$ : Treatment at  $t^{\text{th}}$  decision point (aka: action)
- $Y_{t+1}$ : Proximal response (aka: reward, utility, cost)

7

## Structure of Mobile Health Intervention

1) Decision Times: times,  $t$ , at which a treatment might be provided.

- 1) Regular intervals in time (e.g. every minute)
- 2) At user demand

Heart Steps: approximately every 2-2.5 hours:  
pre-morning commute, mid-day, mid-afternoon, evening commute, after dinner

8

The momentary times were selected because these times are the times at which most people tended to have the greatest within person variance in activity  
Pre-morning commute, mid-day, mid-afternoon, evening commute, after dinner.

42 day study → 210 decision points

Smoking Cessation: Every 1 minute during 10 hour day.

## Structure of Mobile Health Intervention

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

Heart Steps: classifications of activity, location, weather, step count, busyness of calendar, user burden, adherence.....

9

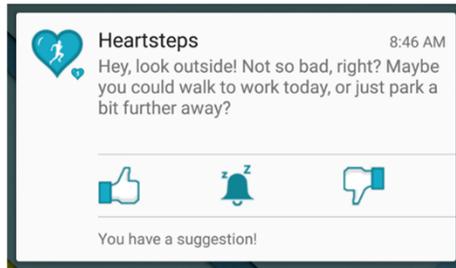
Smoking Cessation: stress, smoking detection, mood, driving,....

## Structure of Mobile Health Intervention

### 3) Treatment $A_t$

- 1) Types of treatments/engagement strategies that can be provided at a decision time,  $t$
- 2) Whether to provide a treatment

HeartSteps: tailored activity suggestion (yes/no)



Engagement strategies

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

Activity suggestions are to help in your automated processing

Smoking Cessation: Cue on wrist band

## Availability

- Treatments,  $A_p$ , can only be delivered at a decision point if an individual is *available*.
  - $O_t$  includes  $I_t=1$  if available,  $I_t=0$  if not
- Treatment effects at a decision point are conditional on availability.
- Availability is not the same as adherence!

11

Available if

1. she is not currently potentially operating a car, (unethical to deliver)
2. she is not currently walking, and (not scientific to deliver) another example is available only if currently classified as at risk.
3. Participant has turned off intervention (unethical to deliver)
3. her phone is connected to the internet. (technical concerns) we added this when we realized there was a bug in the software code that prevented intervention delivery when phone was not connected.

Availability is feasibility of trt options

Adherence (i.e. compliance) is very different from availability. Suppose a person is available at a decision point. However the phone is in their purse across the room. So they don't hear whether the phone pings/ see the lockscreen light up. This person is non-adherent at this decision point. Primary analyses will be intention-to-treat and thus will average over non-compliance.

## Structure of Mobile Health Intervention

### 4) Proximal Response $Y_{t+1}$

Heart Steps: Step count over 30 minutes following decision point,  $t$

12

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

## Continually Learning Mobile Health Intervention

- 1) Trial Designs: Are there effects of the treatment actions on the proximal response? *experimental design*
- 2) Data Analytics for use with trial data: Do effects vary by the user's internal/external context,? Are there delayed effects of the treatments? *causal inference*
- 3) Learning Algorithms for use with trial data: Construct a "warm-start" treatment policy. *batch Reinforcement Learning*
- 4) Online Algorithms that personalize and continually update the treatment policy. *online Reinforcement Learning*

13

Types of questions that may be addressed using data from a micro-randomized trial:

- 1) For a given choice of proximal outcome(e.g. momentary reward) we can eliminate (screen out) ineffective actions. For example we may find that the momentary actions have no impact
- 2) Use SNMM to consider moderation by context.

## Outline

- Introduction to mobile health
- Micro-Randomized Trials for Constructing mHealth interventions
  - Main Effects and Sample Size Calculation
- Secondary Analyses
  - Causal Treatment Effects (aka Causal Excursions)
  - Estimation Method
- Example: HeartSteps

14

## 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*

15

Combination of A/B testing and single case designs.

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

1.1.2. N-of-1 trials. At first glance, the micro-randomized trial design appears similar to the N-of-1 trial design frequently used in the behavioral sciences. However the estimand is quite different. We will, as is typical in statistical causal inference, consider average causal effects, possibly conditional on covariates. In the behavioral field N-of-1 trials are used most often to ascertain individual level causal effects [McDonald et al., 2017]. A variety of nuanced assumptions about individual behavior using behavioral science theory is brought to bear as scientists attempt to triangulate on individual level effects; see the section on “Measuring behavior over time” in McDonald et al. [2017] for a discussion. In the clinical field, N-of-1 trials were developed for settings in which scientists wish to compare the effect of one treatment versus another (treatment A versus treatment B) on an outcome but it is very expensive to recruit many participants. In

both settings a common assumption underlying the analysis of N-of-1 trials is that there are no carry-over effects. Additionally one often assumes that the treatment effect is constant over time. An excellent overview of N-of-1 designs and their use for evaluating technology based interventions is Dallery et al. [2013]. See Kravitz et al. [2014] for a review of this design in pharmacotherapy trials.

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 *iee* microwave magazine.

Evaluation of Simultaneous Multisine Excitation of the

Joined Wing SensorCraft Aeroelastic Wind Tunnel Model, Jennifer Heeg<sup>1</sup> and Eugene Morelli<sup>2</sup> American Institute of Aeronautics and Astronautics

WHAT IF ANYTHING DO WE GAIN BY THINKING OF THIS DESIGN AS A FACTORIAL DESIGN????

Units are people

What are the factors?

What about two time-varying factors (activity planning, lockscreen message). Then the main effects would be time varying. We would have different kinds of interactions as well.

Interactions between factors. The new types of effects are delayed effects of one factor on its future effects as well as delayed interactions..... How to handle nested factors?? Note that lockscreen messages on day t are nested within activity planning

So can one think also of time-varying two-way interaction between G and L (activity planning and lockscreen messages). The time scale is different for these two as lockscreen messages can occur 5 times per day whereas activity planning can only appear once per day.

\*\*\*\*\*

### Other thoughts

about whether for these small pilot studies we want to be doing full within-subject randomization as we have been planning on doing or pseudo-randomization, such as modified latin square or similar procedure, that tries to balance occurrence of meaningful events in the data, if the study is not large/long enough to ensure that this happens via full-blown randomization.

Because we have so many "factors" (e.g. treatments and some variables like day of week, time of day, etc.) all of which are sequential in time, trying to ensure balance on all factors is very complex. This midday I constructed a very complex blocking strategy. However as I think about this, I realize this is not the way to go (complexity is not good as there are often unintended consequences--I can give an example in face-to-face meeting). Rather we need to state what our big concerns are and then block to eliminate these concerns. Here is an example.

We are concerned that of the afternoon's assigned a contextual message, very few messages will be sedentary (say we are concerned that 80% will be active and only 20% will be sedentary). We want to ensure this is 50-50.

As I write this I realize that this is not the best example as we will have 30 people. So we would need to be concerned that of the person-afternoon combinations assigned a contextual message very few person-afternoon combinations will be assigned a sedentary message. A distribution that is really different from 50-50 is unlikely to occur just due to the sheer number of person-afternoon combinations. There are 1260 person-afternoons. Using our current approach to randomization we'd expect to provide a contextual message on 189 person-afternoons (this is because 1/4 of all days there are no messages and we restrict ourselves to at most 3 messages among the 5 occasions per day--this number would not be increased by blocking). Of the 189, 1/2 of the person-afternoons should be sedentary. Because 189 is so large, the number of person-afternoons that will be assigned a sedentary message will be very close to 95. There is no need to block--randomization will achieve good balance with a number as large as 189.

This example and the calculations above tell me often we are dealing with large numbers. This means that we need to have a complex concern --something like "we are concerned that by accident, there will be no messages randomly provided on saturday afternoons" Again we can calculate the expected number of saturday afternoons. We have  $30 \times 6 = 180$  person-saturday afternoons. Of this we expect that 54 person-saturday afternoons will be randomized to a message and of these 1/2, that is 27 will receive a sedentary message. Still pretty large numbers.

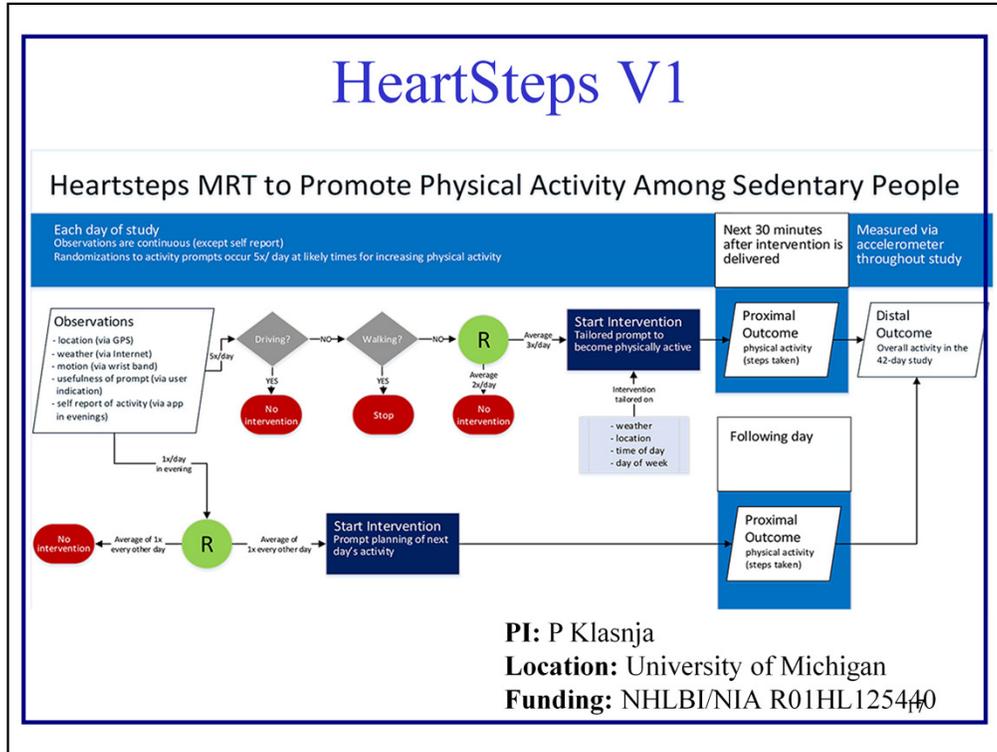
Notice that we will be pooling over people in the data analyses. Even with a complex

blocking strategy I would never expect to learn individually on people. (need an online learning type paradigm).

Can you come up with a concrete concern about balance?

## 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



This project tests the feasibility and effectiveness of providing, via a smartphone, just-in-time tailored physical activity suggestions as well as evening prompts to plan the following day's physical activity so as to help sedentary individuals increase their activity. The resulting data will be used to inform the development of a JITAI for increasing physical activity.

**PI:** Predrag Klasnja

**Location:** University of Michigan

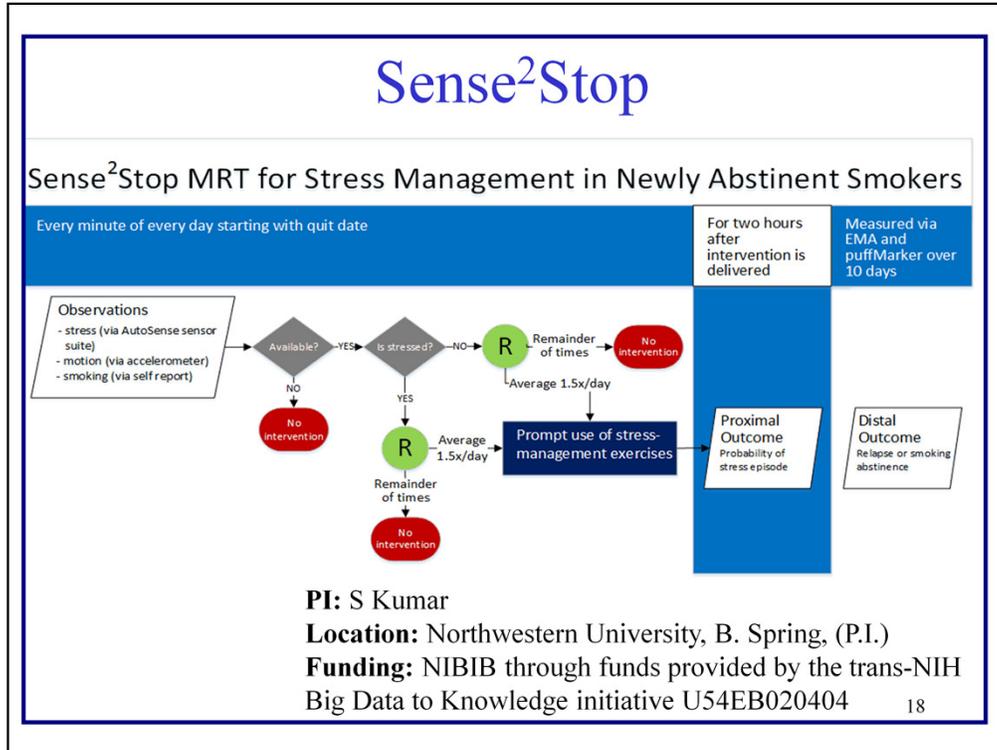
**Funding:** [NHLBI/NIA R01HL125440](https://www.clinicaltrials.gov/ct2/show/NCT03225521?titles=HeartSteps&rank=1)

Would we link the heartsteps MRT to the site

<https://www.clinicaltrials.gov/ct2/show/NCT03225521?titles=HeartSteps&rank=1>

and the MD2K smoking cessation study to

<https://www.clinicaltrials.gov/ct2/show/study/NCT03184389?recrs=a&lead=Northwestern+University&cntry1=NA%3AUS&state1=NA%3AUS%3AIL&draw=1>



This project tests the feasibility of conducting an MRT aiming to investigate whether real-time sensor-based assessments of stress are useful in optimizing the provision of just-in-time prompts to support stress-management in chronic smokers attempting to quit. The resulting data will be used to inform the development of a JITAI for smoking cessation.

**PI:** Santosh Kumar, Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K, <https://md2k.org>)

**Location:** Northwestern University, Bonnie Spring, (site P.I.)

**Funding:** NIBIB through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative ([www.bd2k.nih.gov](http://www.bd2k.nih.gov)). [U54EB020404](https://www.nih.gov/press-releases/2014/04/23/NIH-Announces-Trans-NIH-Program)

Would we link the heartsteps MRT to the site

<https://www.clinicaltrials.gov/ct2/show/NCT03225521?titles=HeartSteps&rank=1>

and the MD2K smoking cessation study to

<https://www.clinicaltrials.gov/ct2/show/study/NCT03184389?recrs=a&lead=Northwestern+University&cntry1=NA%3AUS&state1=NA%3AUS%3AIL&draw=1>

## Outline

- Introduction to mobile health
- Micro-Randomized Trials for Constructing mHealth interventions
  - [Main Effects](#) and Sample Size Calculation
- Secondary Analyses
  - Causal Treatment Effects (aka Causal Excursions)
  - Estimation Method
- Example: HeartSteps

19

## 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*?)

20

## 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

$$P[A_t = 1] = .6 \quad t = 1, \dots, T$$

21

planned study 2160 decision times.

$J=42*5=210$  momentary randomizations

## Time-varying Main Effects

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

→

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

22

## 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.

23

The momentary intervention can be turned off for 1-8 hours by the participant. The intervention is also off if the participant is currently active (e.g. walking) or if the participant may be driving a car

In the MD2K smoking study we might tentatively define availability as:

A person is available if he/she (1) is wearing autosense (our understanding is that autosense will be worn up to 12 hours a day, participants will not wear it when they sleep or when in the shower); if the person is not wearing Autosense, no data will be collected and recommendations will not be pushed); (2) did not receive a message in the past 60 minutes; and (3) is not driving a car.

## 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, nor is it the same as interruptibility, receptivity

24

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.

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.

Adherence (i.e. compliance) is very different from availability. Suppose a person is available at a decision point. However the phone is in their purse across the room. So they don't hear whether the phone pings/ see the lockscreen light up. This person is non-adherent at this decision point. Primary analyses will be intention-to-treat and thus will average over non-compliance.

## Potential Outcomes

- Define

$$\bar{A}_t = \{A_1, A_2, \dots, A_t\}, \bar{a}_t = \{a_1, a_2, \dots, a_t\}$$

- 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  $\bar{A}_{t-1} = \bar{a}_{t-1}$

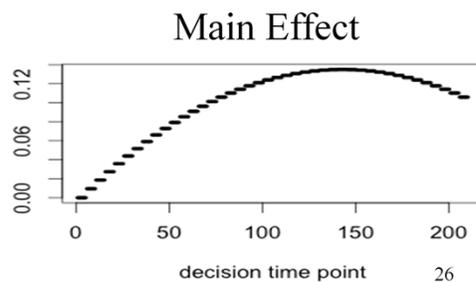
25

## 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?



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.

## 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

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

27

Assuming “consistency” THESE ARE CAUSAL EFFECTS!!!!

These equations only make sense when the randomization is constant, or time varying but only depending on an exogenous variable. If randomization is stratified based on an endogenous time varying variable then the formula changes.

## 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  $\beta(t)$ ,  $t=1, \dots, T$

The main effect is a causal effect.

28

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

## Outline

- Introduction to mobile health
- Micro-Randomized Trials for Constructing mHealth interventions
  - Main Effects and **Sample Size Calculation**
- Secondary Analyses
  - Causal Treatment Effects (aka Causal Excursions)
  - Estimation Method
- Example: HeartSteps

29

## Sample Size Calculation

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

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

- 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

$$\beta_0 = \beta_1 = \beta_2 = 0$$

30

instead of a sparsity bet, we place a smoothness bet We are not assuming that the main effect has a quadratic form

## 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.

31

The contrasts become within person contrasts due to the assumption of smoothness in average response over time (people available at neighboring decision points are similar). If the proximal effect at each time point were to be estimated separately then it would be like a two arm study at each time j.

## 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

$q_t = .6$  in HeartSteps

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

32

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 the sample size formula uses a simplified variance—the simplification assumes the working assumption that  $\gamma$  is correctly specified.

Essentially we are estimating the projection of this conditional mean on this space.

## 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:

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

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

33

Emphasize here that the goal is to construct a test statistic as opposed to estimating  $\beta(t)$ .

Since the model for the proximal effect of  $A_j$  on  $Y_j$  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).

$\gamma(t)$  can include baseline variables: gender, baseline activity level

## 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.

34

## Alternative for Sample Size Calculation

- Average conditional variance is

$$\bar{\sigma}^2 = (1/T) \sum_{t=1}^T E[VAR(Y_{t+1} | I_t = 1)]$$

35

## Specify Alternative for Sample Size Calculation

- Scientist indirectly specifies **standardized  $d_i$ 's**
  - initial main effect:  $d_0$ ,
  - average main effect over trial duration:

$$\frac{1}{T} \sum_{t=1}^T (d_0 + d_1 \lfloor \frac{t-1}{5} \rfloor + d_2 \lfloor \frac{t-1}{5} \rfloor^2),$$

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

36

Specify alternative so that scientist can provide  $d_i$ 's

## Test Statistic for Sample Size Calculation

- Put  $Y_i = (Y_{i2}, \dots, Y_{iT+1})^T$  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:

$$I_{it}(A_{it} - q_t), I_{it}(A_{it} - q_t) \lfloor \frac{t-1}{5} \rfloor, \\ I_{it}(A_{it} - q_t) \lfloor \frac{t-1}{5} \rfloor^2$$

37

## 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 3 by  $3+q$  matrix picking out columns associated with coefficients  $\beta$

38

## 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.

$$\epsilon_{it} = Y_{i,t+1} - \left( \gamma(t) + \beta(t)(A_{it} - q_t) \right)$$

These working assumptions are false!

39

## 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  $\gamma(t)$

40

$\Sigma_{\beta} = (Z^T Z)^{-1}$  divided by  $q(1-q)\tau$      $Z$  is number of time points by 3. The 3 columns correspond to the columns  $[A_{j-6}, (A_{j-6}) \lfloor \frac{j-1}{5} \rfloor, (A_{j-6}) \lfloor \frac{j-1}{5} \rfloor^2]$  in the design matrix.

## 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.

41

N is no. of subjects

q is no of control variables including intercept.

q<sub>t</sub> is randomization prob. at time t

p is the number of parameters in  $\gamma(t)$ . 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  $[A_{j-q,t}, (A_{j-q,t}) \lfloor \frac{j-1}{5} \rfloor, (A_{j-q,t}) \lfloor \frac{j-1}{5} \rfloor^2]$  in the design matrix.

## HeartSteps Example

- Standardized  $d_i$ 's
  - initial effect:  $d_0=0$
  - output average main effect
  - day of maximal main effect:  $-\frac{d_1}{2d_2} = 28$
- Projection used to form test statistic:  
$$\gamma(t) + \beta(t)(A_{it} - .6), t = 1, \dots, 210$$

where 
$$\gamma(t) = \gamma_0 + \gamma_1 \lfloor \frac{t-1}{5} \rfloor + \gamma_2 \lfloor \frac{t-1}{5} \rfloor^2$$

42

Specify alternative so that scientist can provide  $d_i$ 's

42 days in our study

**HeartSteps Sample Sizes**  
**Power=.80, False-positive error=.05**

<b>Standardized Average Main Effect over 42 Days</b>	<b>Sample Size For 70% availability or 50% availability</b>
0.06 standard deviations	81 or 112
0.08 standard deviations	48 or 65
0.10 standard deviations	33 or 43

43

Average main effect is standardized.

#parameters=6

Meaningful increase in stepcount is 1000/day

Usual std is 2000/day

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

## Same Test Statistic for Analysis

- “GEE” test statistic is

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

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

- No working assumptions

44

## Small Sample Adjustment

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

- Adjusted sandwich estimator:

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

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

45

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 Main 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 (originally we thought we would have randomization prob. =.4 but then changed this to .6 after a run in of several subjects prior to real study)

Simulations indicate:

Method is sensitive to

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

Guess of average main 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

$L_{j+1}$  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  $\frac{1}{2}$  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

## Discussion

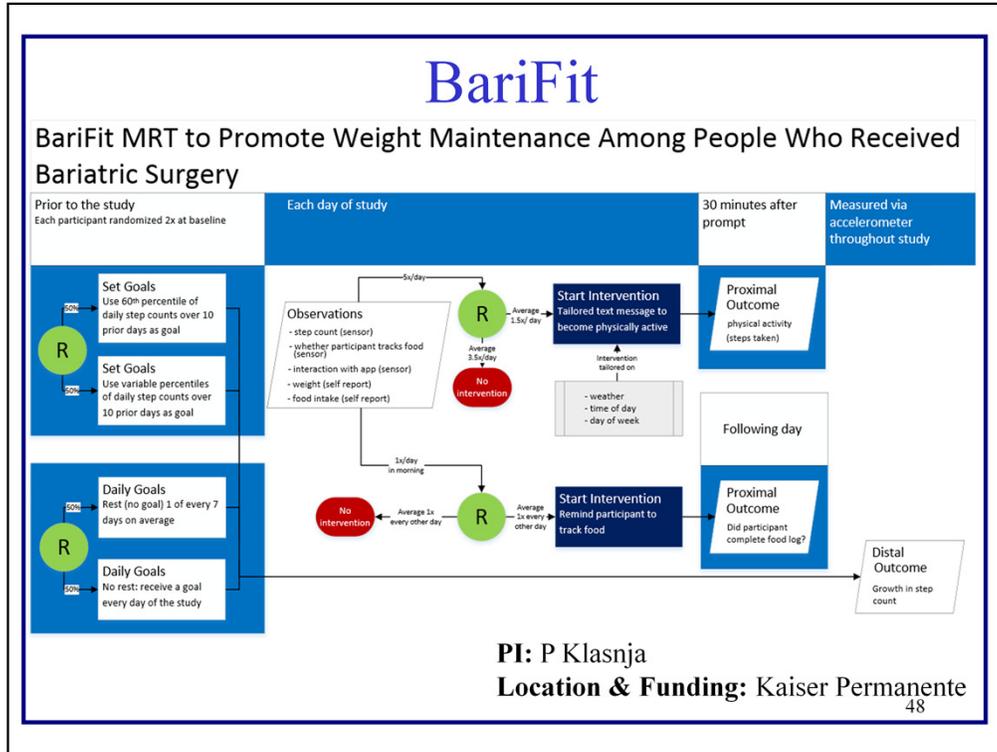
The micro-randomized trial is a sequential factorial trial with multiple factors, e.g. in HeartSteps:

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

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

47

- <https://cran.r-project.org/web/packages/MRTSampleSize/index.html>
- <https://pengliao.shinyapps.io/mrt-calculator/>



Researchers are conducting this quality-improvement MRT aiming to promote weight maintenance through increased activity and improved diet among people who received bariatric surgery. At the time it was developed, this project was novel in that it implemented separate randomizations at the start of the study, on a daily basis, and five times throughout the day.

**PI:** Pedja Klasna

**Location & Funding:** Kaiser Permanente

**50 participants in a 4-month Micro-randomized trial.**

**Before the study:** There is a 7-day baseline data collection prior to randomization. Participants will be wearing a screen-less accelerometer for a week before they start the study, to capture step count data that will be used to generate adaptive step goals (percentile values).

## Experimental Design Challenges

Micro-randomized trials are a new type of factorial design

- i. Time varying factors → time varying main effects, time-varying two-way interactions, different delayed effects
- ii. Design studies specifically to detect interactions between factors.
- iii. Randomization may depend on an outcome of past actions

49

Very different from designing a study to do system id!

## Outline

- Introduction to mobile health
- Micro-Randomized Trials for Constructing mHealth interventions
  - Main Effects and Sample Size Calculation
- Secondary Analyses
  - Causal Treatment Effects (aka Causal Excursions)
  - Estimation Method
- Example: HeartSteps

50

## Micro-Randomized Trial Data

On each of  $n$  participants and at each of  $t=1, \dots, T$  decision points:

- $O_t$  observations at decision point  $t$ ,
  - includes  $I_t=1$  if available,  $I_t=0$  if not
- $A_t=1$  if treated,  $A_t=0$  if not treated at decision  $t$ 
  - Randomized,  $p_t(H_t) = P[A_t=1 | H_t, I_t=1]$
- $Y_{t+1}$  proximal response

$H_t = \{(O_{i^p}, A_{i^p}, Y_{i+1}), i=1, \dots, t-1; O_{i^p}\}$  denotes data through  $t$

Refer to heartsteps as you discuss this slide

No subscript  $i$  denoting participant  $i$ .

Note care in subscript on  $Y$ !!!!  $Y_{j+1}$  occurs subsequent to  $A_j$  not at same time.

In heartsteps  $T=210$  in sense2stop  $T=6000$

## Conceptual Models

Generally data analysts fit a series of increasingly more complex models:

$$Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t$$

and then next,

$$Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t S_t$$

and so on...

- $Y_{t+1}$  is activity over 30 min. following  $t$
- $A_t = 1$  if activity suggestion and 0 otherwise
- $Z_t$  summaries formed from  $t$  and past/present observations
- $S_t$  potential moderator (e.g., current weather is good or not)

52

We go through intuition that we have gained from analyses in which the treatment does not vary with time. Here the complication is that treatment is time varying. The issue is that both  $S_t$  and  $Z_t$  may be outcomes of past treatment. Availability is an outcome of past treatment.

$Z_j$  might include location, time of day, day of week, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time,  $j$ , so as to allow a more flexible model

$S_j$  might be a vector as well and might include features of time  $S_j$  might be the output of a classifier

Above is conditional on availability

## Conceptual Models

Generally data analysts fit a series of increasingly more complex models:

$$Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t$$

and then next,

$$Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t S_t$$

and so on...

$\alpha_1^T Z_t$  is used to reduce the noise variance in  $Y_{t+1}$   
( $Z_t$  is sometimes called a vector of control variables)

53

t is the decision point, I am leaving off subject i subscript.

Z<sub>t</sub> might include location, time of day, day of week, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time, t, so as to allow a more flexible model

S<sub>t</sub> might be a vector as well and might include features of time

## Causal, Marginal Effects

$$Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t$$

$\beta_0$  is the effect, marginal over all observed and all unobserved variables, of the activity suggestion on subsequent activity.

$$Y_{t+1} \sim \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t S_t$$

$\beta_0 + \beta_1$  is the effect when the weather is good ( $S_t=1$ ), marginal over other observed and all unobserved variables, of the activity suggestion on subsequent activity.

54

These are the interpretations we know hold in cross-sectional analyses and in analyses in which the treatment does not vary with time.

## Goal

- Develop data analytic methods that are consistent with the scientific understanding of the meaning of the  $\beta$  coefficients
- Challenges:
  - Time-varying treatment ( $A_t, t=1, \dots, T$ )
  - “Independent” variables:  $Z_t, S_t, I_t$  that may be affected by prior treatment
- Robustly facilitate noise reduction via use of controls,  $Z_t$

55

not so very independent, “independent” variables

## Potential Outcomes

- $\bar{A}_t = \{A_1, A_2, \dots, A_t\}$  (random treatments),  
 $\bar{a}_t = \{a_1, a_2, \dots, a_t\}$  (realizations of treatments)
- $Y_{t+1}(\bar{a}_t)$  is one potential proximal response
- $I_t(\bar{a}_{t-1})$  is one potential “available for treatment” indicator
- $H_t(\bar{a}_{t-1})$  is one potential history vector
  - $S_t(\bar{a}_{t-1})$  is a vector of features of history  $H_t(\bar{a}_{t-1})$

56

Draw attention to  $a_t$  and  $a_{t-1}$

Draw attention to why availability might depend on prior treatment

## Marginal & Causal Effect

Excursion effect at decision point  $t$ :

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

- Effect is conditional on availability; only concerns the subpopulation of individuals available at decision  $t$
- Effect is marginal over any  $Y_u, u \leq t, A_u, u < t$  not in  $S_t(\bar{A}_{t-1})$ ---over all variables not in  $S_t(\bar{A}_{t-1})$ .

57

Note how odd this causal effect is—it involves the randomization probabilities.

Availability is not equivalent to willingness to enroll. It is momentary. Willingness to enroll is up front. Our availability is closer to feasibility of trt options

We are examining causal effects via **\*\*\*\*excursions\*\*\*\*** from the underlying randomization policy

$S_t$  could be an empty set.

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.

## Marginal & Causal Effect

Excursion effect at decision point  $t$ :

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

- Effect is conditional on availability; only concerns the subpopulation of individuals available at decision  $t$
- Effect is marginal over any  $Y_u, u \leq t, A_u, u < t$  not in  $S_t(\bar{A}_{t-1})$ ---over all variables not in  $S_t(\bar{A}_{t-1})$ .

58

Note how odd this causal effect is—it involves the randomization probabilities.

Availability is not equivalent to willingness to enroll. It is momentary. Willingness to enroll is up front. Our availability is closer to feasibility of trt options

We are examining causal effects via **excursions** from the underlying randomization policy

$S_t$  could be an empty set.

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.

## Consistency & Micro-Randomized $A_t \rightarrow$

$$\begin{aligned}
 & E[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) \mid I_t(\bar{A}_{t-1}) = 1, S_t(\bar{A}_{t-1})] \\
 &= \\
 & \quad E[ E[Y_{t+1} \mid A_t = 1, I_t = 1, H_t] \\
 & \quad \quad - E[Y_{t+1} \mid A_t = 0, I_t = 1, H_t] \mid I_t = 1, S_t] \\
 &= \\
 & \quad E \left[ \frac{A_t Y_{t+1}}{p_t(H_t)} - \frac{(1 - A_t) Y_{t+1}}{(1 - p_t(H_t))} \mid I_t = 1, S_t \right]
 \end{aligned}$$

$(p_t(H_t)$  is randomization probability)

$p_t(H_t)$  is randomization probability

## Marginal Treatment Effect

Treatment Effect Model:

$$E[ E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - E[Y_{t+1}|A_t = 0, I_t = 1, H_t] | I_t = 1, S_t ] = S_t^T \beta$$

$H_t$  is participant's data up to and at time  $t$

$S_t$  is a vector of data summaries and time,  $t$ , ( $S_t \subseteq H_t$ )

$I_t$  indicator of availability

We aim to conduct inference about  $\beta$ !

60

We are examining causal effects via \*\*\*\*excursions\*\*\* from the underlying randomization policy

## Outline

- Introduction to mobile health
- Micro-Randomized Trials for Constructing mHealth interventions
  - Main Effects and Sample Size Calculation
- Secondary Analyses
  - Causal Treatment Effects (aka Causal Excursions)
  - Estimation Method
- Example: HeartSteps

61

## “Centered and Weighted Least Squares Estimation”

- Simple method for complex data
- Enables unbiased inference for a causal, marginal, treatment effect (the  $\beta$ 's)
- Inference for treatment effect is not biased by how we use the controls,  $Z_t$ , to reduce the noise variance in  $Y_{t+1}$

<https://arxiv.org/abs/1601.00237><sub>62</sub>

## Estimation

- Select probabilities:  $\tilde{p}_t(s) \in (0,1)$
- Form weights:  $W_t = \left(\frac{\tilde{p}_t(S_t)}{p_t(H_t)}\right)^{A_t} \left(\frac{1-\tilde{p}_t(S_t)}{1-p_t(H_t)}\right)^{1-A_t}$
- Center treatment actions:  $A_t \rightarrow (A_t - \tilde{p}_t(S_t))$

- Minimize:

$$E_n \left[ \sum_{t=1}^T (Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t)) S_t^T \beta)^2 I_t W_t \right]$$

- $E_n$  is empirical distribution over individuals.

63

Form working model for mean of  $Y_{t+1}$  given  $I_t=1$  and  $H_t$   
 $E[Y_{t+1} | I_t = 1, H_t] \approx Z_t^T \alpha$

## Minimize

$$E_n \left[ \sum_{t=1}^T (Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t)) S_t^T \beta)^2 I_t W_t \right]$$

Good but **incorrect** intuition:

Appears to be a weighted “GEE” with a working independence correlation matrix and a centered treatment indicator,  $A_t - \tilde{p}_t(S_t)$  thus:

- $E[Y_{t+1} | A_t, I_t = 1, Z_t] = Z_t^T \alpha + (A_t - \tilde{p}_t(S_t)) S_t^T \beta$

$E_n$  is expectation with respect to empirical distribution<sup>4</sup>

This intuition is not correct because we are not assuming that the conditional mean of  $Y_{t+1}$  given  $Z_t, S_t, I_t=1$ , has above form!

## Minimize

$$E_n \left[ \sum_{t=1}^T (Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t)) S_t^T \beta)^2 I_t W_t \right]$$

Good but **incorrect** intuition:

- $E[Y_{t+1} | A_t, I_t = 1, Z_t] \neq Z_t^T \alpha + (A_t - \tilde{p}_t(S_t)) S_t^T \beta$

$E_n$  is expectation with respect to empirical distribution<sup>5</sup>

This intuition is not correct because we are not assuming that the conditional mean of  $Y_{t+1}$  given  $Z_t, S_t, I_t=1$ , has above form!

## Minimize

$$E_n \left[ \sum_{t=1}^T (Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t)) S_t^T \beta)^2 I_t W_t \right]$$

### The Modeling Assumption:

$$E[(E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - E[Y_{t+1}|A_t = 0, I_t = 1, H_t]) | I_t = 1, S_t] = S_t^T \beta_0$$

If  $\tilde{p}_t$  depends at most on features in  $S_t$ , then, under moment conditions,  $\hat{\beta}$  is consistent for  $\beta_0$

66

If in the  $W_t$  you included  $[\tilde{p}_t(S_t) (1 - \tilde{p}_t(S_t))]^{-1}$  then you would be obtaining the BLP of the treatment effect.

## Theory

Under moment conditions,  $\sqrt{n}(\hat{\beta} - \beta_0)$  converges to a Normal distribution with mean 0 and var-covar matrix,  $(\Sigma_p)^{-1} \Sigma (\Sigma_p)^{-1}$

$$\Sigma_p = E\left[\sum_{t=1}^T \tilde{p}_t(S_t)(1 - \tilde{p}_t(S_t))I_t S_t S_t^T\right]$$

$Z_t$  and  $S_t$  are finite dimensional feature vectors.

67

## Gains from Randomization

- Causal inference for a marginal treatment effect
- Inference on treatment effect is robust to working model:

$$E[Y_{t+1} | I_t = 1, H_t] \approx Z_t^T \alpha$$

- $Z_t \subseteq H_t$
- Contrast to literature on partially linear, single index models and varying coefficient models

68

## Price due to Marginal Estimand

This “GEE-like” method can only use a working independence correlation matrix

- Estimating function is biased if off-diagonal elements in working correlation matrix: in general,

$$E\left[\left(Y_{t+1} - Z_t^T \alpha - (A_t - \tilde{p}_t(S_t)) S_t^T \beta\right) I_t W_t I_u W_u \sigma_{t,u} Z_u\right] \neq 0$$

if  $u \neq t$

69

Such a result is unsurprising given the bias that arises under non-independence structures in IPTW (Vansteelandt 2007; Tchetgen Tchetgen et al. 2012) or in GEEs where a time-varying response is modelled by time-varying covariates (Pepe and Anderson 1994; Schildcrout and Heagerty 2005).

## Choice of Weights

Choice of  $\tilde{p}_t(S_t)$  determines marginalization (e.g. projection) under model misspecification of treatment effect.

**Example:**  $S_t = 1$ ,  $\tilde{p}_t(S_t) = \tilde{p}$ . Resulting  $\hat{\beta}$  is an estimator of

$$\frac{\sum_{t=1}^T E[I_t] \beta_t}{\sum_{t=1}^T E[I_t]}$$

where

$$\beta_t = E[ E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - E[Y_{t+1}|A_t = 0, I_t = 1, H_t] | I_t = 1 ]$$

70

Use when assumption does not hold.

Note that this estimand weighs all available person-times equally. If instead you would prefer  $E[\sum_{t=1}^T I_t \beta_t / \sum_{t=1}^T I_t]$  that is, you might prefer first to average within a person and then average across people, then you are really looking for a random effects model.

## Outline

- Introduction to mobile health
- Micro-Randomized Trials for Constructing mHealth interventions
  - Main Effects and Sample Size Calculation
- Secondary Analyses
  - Causal Treatment Effects (aka Causal Excursions)
  - Estimation Method
- Example: HeartSteps

71

# Heart Steps Study



On each of  $n=37$  participants:

a) Activity suggestion,  $A_t$

- **Provide a suggestion with probability .6**
  - a tailored sedentary-reducing activity suggestion (probability=.3)
  - a tailored walking activity suggestion (probability=.3)
- **Do nothing (probability=.4)**
- 5 times per day \* 42 days= 210 decision points

72

## Conceptual Models

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_t$$
$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \alpha_2 d_t + \beta_0 A_t + \beta_1 A_t d_t$$

- $t=1, \dots, T=210$
- $Y_{t+1}$  = log-transformed step count in the 30 minutes *after* the  $t^{\text{th}}$  decision point,
- $A_t = 1$  if an activity suggestion is delivered at the  $t^{\text{th}}$  decision point;  $A_t = 0$ , otherwise,
- $Z_t$  = log-transformed step count in the 30 minutes *prior* to the  $t^{\text{th}}$  decision point,
- $d_t$  = days in study; takes values in  $(0, 1, \dots, 41)$

73

## Analysis

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_t, \text{ and}$$

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \alpha_2 d_t + \beta_0 A_t + \beta_1 A_t d_t$$

Causal Effect Term	Estimate	95% CI	p-value
$\beta_0 A_t$ (effect of an activity suggestion)	$\hat{\beta}_0 = .13$	(-0.01, 0.27)	.06
$\beta_0 A_t + \beta_1 A_t d_t$ (time trend in effect of an activity suggestion)	$\hat{\beta}_0 = .51$	(.20, .81)	<.01
	$\hat{\beta}_1 = -.02$	(-.03, -.01)	<.01

74

.13 translates into a 14% increase over no treatment in step count about 33 steps  
mean 30-minute step count is 253 steps

.51 translates into a 67% increase over no treatment in step count about 170 steps

Midway through study  $d_t=20$  this increase has reduced to 16% increase in step count

# Heart Steps Study



On each of  $n=37$  participants:

a) Activity suggestion

- Provide a suggestion with probability .6
  - **a tailored walking activity suggestion (probability=.3)**
  - **a tailored sedentary-reducing activity suggestion (probability=.3)**
  - **Do nothing (probability=.4)**
- 5 times per day \* 42 days= 210 decision points

75

# Study Analysis



$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t}$$

- $A_{1t} = 1$  if walking activity suggestion is delivered at the  $t^{\text{th}}$  decision point;  $A_{1t} = 0$ , otherwise,
- $A_{2t} = 1$  if sedentary-reducing activity suggestion is delivered at the  $t^{\text{th}}$  decision point;  $A_{2t} = 0$ , otherwise,

Causal Effect	Estimate	95% CI	p-value
$\beta_0 A_{1t} + \beta_1 A_{2t}$	$\hat{\beta}_0 = .21$	(.04, .39)	.02
	$\hat{\beta}_1 > 0$	ns	ns

76

mean 30-minute step count is 253 steps

.21 translates into a 23% increase over no treatment in step count about 59 steps

When  $d_t$  is added to the model one finds that initially  $\beta_0$  coefficient of  $A_{1t} = .729$  and the coefficient of  $A_{1t} * d_t$  is  $-.025$

Pvalues for both are .000

.729 translates into a 107% increase over no treatment in step count about 271 steps

## Initial Conclusions

- The data indicates that there is a causal effect of the activity suggestion on step count in the succeeding 30 minutes.
  - This effect is primarily due to the walking activity suggestions.
  - This effect deteriorates with time
  - The walking activity suggestion initially increases step count over succeeding 30 minutes by  $\approx 271$  steps but by day 21 this increase is only  $\approx 65$  steps.

77



On each of  $n=37$  participants:

b) Evening planning prompt,  $A_t$

- **Provide a prompt with probability .5**
  - Prompt using unstructured activity planning for following day with probability=.25
  - Prompt using structured activity planning for following day with probability=.25
- **Do nothing with probability=.5**
- 1 time per day \* 42 days= 42 decision points

## Conceptual Models

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_t$$
$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \alpha_2 W_t + \beta_0 A_t W_t + \beta_1 A_t (1 - W_t)$$

- $t=1, \dots, T=42$
- $Y_{t+1}$  = square root-transformed step count on the day *after* the  $t^{\text{th}}$  day,
- $A_t = 1$  if activity planning prompt on the evening of the  $t^{\text{th}}$  day;  $A_t = 0$ , otherwise,
- $Z_t$  = square-root step count on the  $t^{\text{th}}$  day,
- $W_t = 1$  if Sunday through Thursday;  $W_t = 0$ , otherwise

79

## HeartSteps Analysis

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_t, \text{ and}$$

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \alpha_2 W_t + \beta_0 A_t W_t + \beta_1 A_t (1 - W_t)$$

Causal Effect Term	Estimate	95% CI	p-value
$\beta_0 A_t$ (effect of planning)	$\hat{\beta}_0 = 1.7$	ns	ns
$\beta_0 A_t W_t + \beta_1 A_t (1 - W_t)$ (effect of planning for weekday ( $W_t = 1$ ) and for weekend ( $W_t = 0$ ))	$\hat{\beta}_0 = 3.6$ $\hat{\beta}_1 < 0$	(.74, 6.4) ns	<.02 ns

80

1.7 translates into an increase in steps of  $(1.7)^2 + 2 * (1.7) * \sqrt{5000} = 243$  steps for a daily average step count of 5000 steps

3.6 translates into a 520 increase over number of steps on weekday compared to no treatment

No trend with time....



On each of  $n=37$  participants:

b) Evening planning prompt

- Provide a prompt with probability .5
    - **Prompt using unstructured activity planning for following day with probability=.25**
    - **Prompt using structured activity planning for following day with probability=.25**
  - **Do nothing with probability=.5**
- 1 time per day \* 42 days= 42 decision points

## Conceptual Model

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t}$$

- $Y_{t+1}$  = square root-transformed step count on the day *after* the  $t^{\text{th}}$  day,
- $A_{1t} = 1$  if *unstructured* activity planning prompt on the evening of the  $t^{\text{th}}$  day;  $A_{1t} = 0$ , otherwise,
- $A_{2t} = 1$  if *structured* activity planning prompt on the evening of the  $t^{\text{th}}$  day;  $A_{2t} = 0$ , otherwise,
- $Z_t$  = square-root step count on the  $t^{\text{th}}$  day,

82

## HeartSteps Analysis

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t} \quad t=0, \dots, T=41$$

- $A_{1t} = 1$  if *unstructured* activity planning prompt on the evening of the  $t^{\text{th}}$  day;  $A_{1t} = 0$ , otherwise,
- $A_{2t} = 1$  if *structured* activity planning prompt on the evening of the  $t^{\text{th}}$  day;  $A_{2t} = 0$ , otherwise,

Causal Effect Term	Estimate	95% CI	p-value
$\beta_0 A_{1t} + \beta_1 A_{2t}$	$\hat{\beta}_0 = 3.1$	(-.22, 6.4)	.07
	$\hat{\beta}_1 > 0$	ns	ns

83

No trend with time....

## HeartSteps Analysis

$$Y_{t+1} \sim \alpha_0 + \alpha_1 Z_t + \alpha_2 W_t + \beta_0 A_{1t} W_t + \beta_1 A_{1t} (1 - W_t) + \beta_2 A_{2t} W_t + \beta_3 A_{2t} (1 - W_t)$$

$A_{1t} = 1$  if *unstructured* activity planning prompt on the evening of the  $t^{\text{th}}$  day;  $A_{1t} = 0$ , otherwise,

$W_t = 1$  if Sunday through Thursday;  $W_t = 0$ , otherwise

Causal Effect	Estimate	95% CI	p-value
$\beta_0 A_{1t} W_t + \beta_1 A_{1t} (1 - W_t)$	$\hat{\beta}_0 = 5.3$	(2.2, 8.5)	<.01
$+ \beta_2 A_{2t} W_t + \beta_3 A_{2t} (1 - W_t)$	$\hat{\beta}_1 < 0, \hat{\beta}_2 > 0,$ $\hat{\beta}_3 < 0$	all ns	all ns

84

5.3 translates into an increase in steps of  $(5.3)^2 + 2 * (5.3) * \sqrt{5000} = 782$  steps on weekday for an unstructured planning prompt compared to no treatment for a daily average step count of 5000 steps

No trend with time....

## Initial Conclusions

- The data indicates that there is a causal effect of planning the next day's activity on on the following day's step count
  - This effect is due to the unstructured planning prompts.
  - This effect occurs primarily on weekdays
  - On weekdays the effect of an unstructured planning prompt is to increase step count on the following day by approximately 780 steps.

85

## Discussion

- Randomization enhances:
  - Causal inference based on minimal structural assumptions
- Challenge:
  - How to include random effects which reflect scientific understanding (“person-specific” effects) yet not destroy causal inference?

86

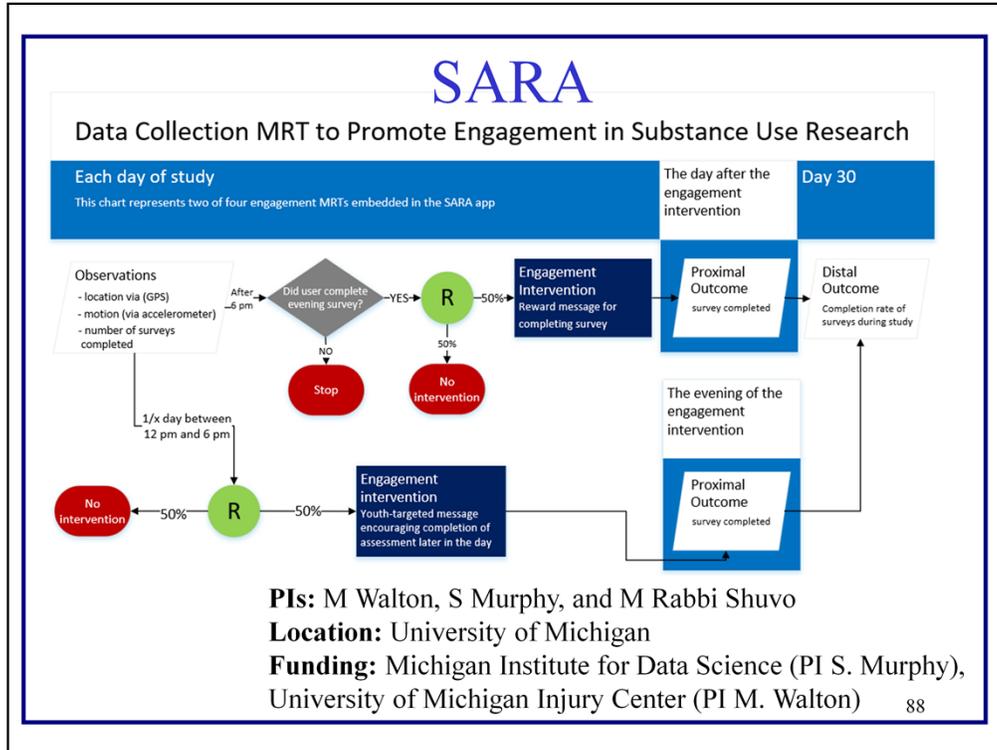
## Concluding Remarks

- Encourage your collaborators to not only register your MRT with [clinicaltrials.gov](https://clinicaltrials.gov) but also preregister with Open Science
  - Non-trivial and slightly stressful!
  - See our Open Science preregistration for SARA
    - an OK but not great start!

87

Rabbi, M., Philyaw-Kotov, M., Klasnja, P., Bonar, E., Nahum-Shani, I., Walton, M., & Murphy, S. (2017, October 18). SARA - Substance Abuse Research Assistant. Retrieved from [osf.io/vwzmd](https://osf.io/vwzmd)

Go to the files section.



The Substance Abuse Research Assistance (SARA) is an app for gathering data about substance use in high-risk populations. App developers are using an MRT to improve engagement with completion of the self-report data collection measures. At the time this summary was written, this MRT is unique in that it has an engagement component, but not a treatment one.

**PIs:** Maureen Walton, Susan Murphy, and Mashfiqui Rabbi Shuvo

**Location:** Harvard University and University of Michigan

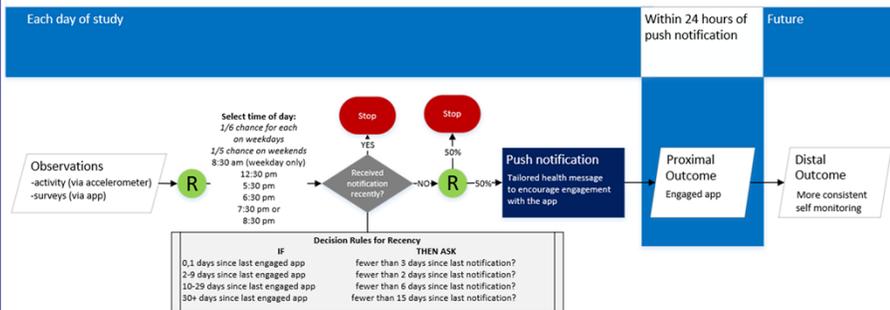
**Funding:** Michigan Institute for Data Science (PI S. Murphy), the University of Michigan Injury Center (PI M. Walton), NIDA P50 DA039838 (PI Linda Collins), NIAAA R01 AA023187 (PI S. Murphy), CDC R49 CE002099 (PI: M. Walton)

Clinical trials.gov identifier NCT03255317

Open Science

# Engagement with JOOL

## MRT to Promote Engagement with Purpose-driven Well-being App



**PI:** Victor Strecher, PhD, MPH, CEO of JOOL Health

**Location & Funding:** Ann Arbor, MI

**URL:** <https://www.joolhealth.com>

89

JOOL is a behavioral health and well-being app that is designed to help people monitor and improve their sleep, presence, activity, creativity, and eating, with the ultimate goal of helping people move closer to fulfilling their life's purpose. This MRT aims to understand whether push notifications of tailored health messages are useful in promoting engagement with the JOOL app; and, if so, when and under what circumstances they are most effective.

**PI:** Vic Strecher, PhD, MPH, CEO of JOOL Health

**Location:** Ann Arbor, Michigan

**URL:** <https://www.joolhealth.com>

## Discussion

### Problematic Analyses

- GLM & GEE analyses
- Random effects models & analyses
- Machine Learning Generalizations:
  - Partially linear, single index models & analysis
  - Varying coefficient models & analysis

*--These analyses do not take advantage of the micro-randomization. Can accidentally eliminate the advantages of randomization for estimating causal effects--*

90

In terms of the ability to obtain causal marginal effects and in terms of robustness.

### **SEMPARAMETRIC GEE ANALYSIS IN PARTIALLY LINEAR SINGLE-INDEX MODELS FOR LONGITUDINAL DATA**

BY JIA CHEN, DEGUI LI, HUA LIANG<sup>†,1</sup> AND SUOJIN WANG<sup>‡,2</sup> 2015, Vol. 43, No. 4, 1682–1715

### **New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis, Jianqing FAN and Runze LI Journal of the American Statistical Association**

**September 2004, Vol. 99,pg. 710**

### References

Intensive Longitudinal Methods by Niall Bolger and Jean-Philippe Laurenceau (2013)

Models for Intensive Longitudinal Data edited by Walls and Schaer (2006)

A Time-Varying Effect Model for Intensive Longitudinal Data by Tan et al., *Psychol Methods*. 2012 March ; 17(1): 61–77.

Dynamical systems analyses, e.g time series or pomps or mps