Micro-Randomized Trials for Developing Just-In-Time Adaptive Interventions in mHealth

S.A. Murphy
Georgia State, Psychology
12/2014

50 min.
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. A fast growing area of mHealth concerns the use of mobile devices for both collecting real-time data, for processing this data and for providing momentary interventions. We discuss the design and analysis of these types of trials.

Outline

- Adaptive Interventions and Just-in-Time Adaptive Interventions
- HeartSteps
- Micro-Randomized Trial
- Sample Size Considerations

2

Adaptive Interventions

- Intervention design that takes advantage of response heterogeneity by individualizing intervention options to the specific and changing needs of individuals
- Example: Adaptive drug court program for drug abusing offenders

Marlowe et al., 2008; 2009; 2012

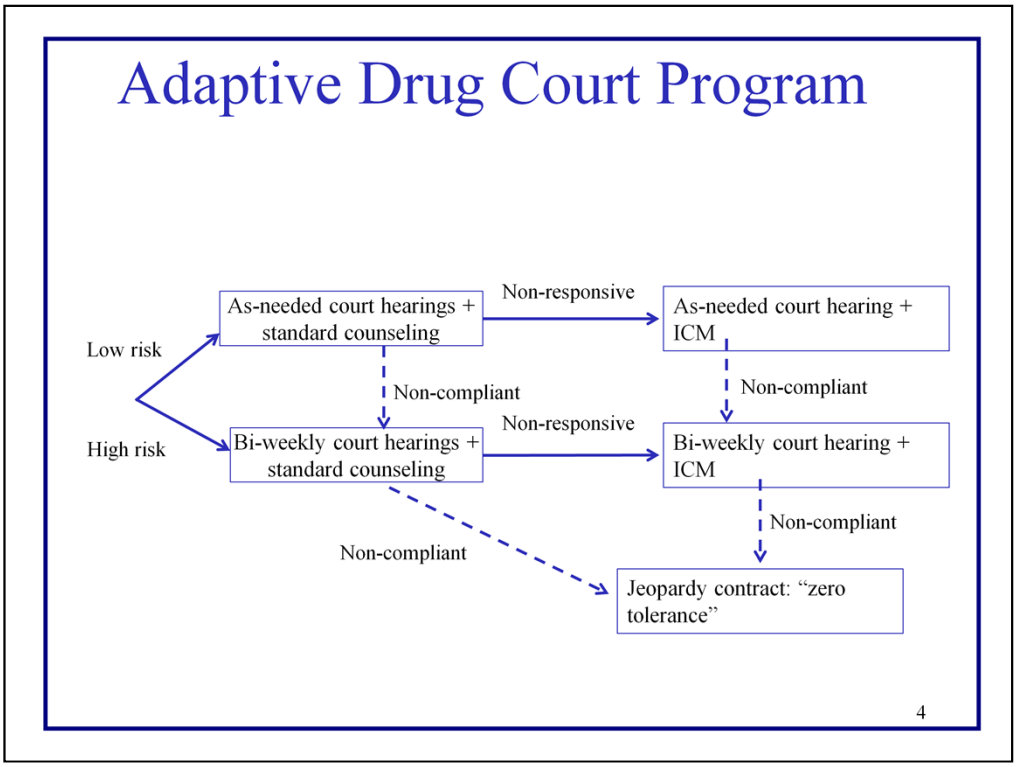
3

The idea is that the same treatment is not good for everyone, and different people need different things at different time point.

Adaptive Interventions in Drug Court: A Pilot Experiment. Criminal Justice Review 2008; 33; 343 Douglas B. Marlowe, David S. Festinger, Patricia L. Arabia, Karen L. Dugosh, Kathleen M. Benasutti, Jason R. Croft and James R. McKay

Adaptive interventions may optimize outcomes in drug courts: a pilot study. Marlowe DB, Festinger DS, Arabia PL, Dugosh KL, Benasutti KM, Croft JR. Curr Psychiatry Rep. 2009 Oct;11(5):370-6.

Adaptive Programming Improves Outcomes in Drug Court : An Experimental Trial Criminal Justice and Behavior 2012 39: 514 Douglas B. Marlowe, David S. Festinger, Karen L. Dugosh, Kathleen M. Benasutti, Gloria Fox and Jason R. Croft



Douglas B. Marlowe: developed and implemented an adaptive intervention for drug offenders

Following their initial court hearing, risk was assessed.

High risk: ASPD (Antisocial Personality Disorder, based on Diagnostic Interview: APD-DI) or history of formal drug abuse treatment otherwise low risk.

These are assessed monthly:

Noncompliance: is (1) falls below threshold for attendance in counseling sessions or (2) fails to provide 2 or more scheduled urine specimens

Nonresponsive = (1) is attending sessions and completing program requirements, and (2) is not committing new infractions, but (3) provides 2 or more drug-positive urine specimens.

If non compliance, contact with the judge is increased.

ICM– intensive clinical case management: Participants are required to meet twice weekly with an intensive clinical case manager who provides

individual substance abuse counseling with an emphasis on motivational enhancement, relapse prevention, and cognitive restructuring ("criminal thinking") techniques.

Jeopardy contract: involves "zero tolerance" for further violations of the rules of the program. Any

further violation leads to a termination hearing, also known as a show-cause hearing. At the termination hearing, the individual is terminated from the program and sentenced on the original charge or charges unless he or she can provide a good reason to be given another chance. The decision of whether or not to grant another chance is within the discretion of the judge

To graduate offender must attend 12 counseling sessions; provide 14 consecutive weekly negative drug urine specimens; remain arrest-free; obey program rules and procedures; pay 200 dollar court fee.

Adaptive Intervention: 5 Elements

The adaptation is guided by consideration of
(1) Proximal Response and Distal Outcome

The adaptation process is composed of
(2) Tailoring Variables,
(3) Decision Rules and
(4) Intervention Options

The adaptation is triggered at
(5) Decision Points

5

Monitoring, individualizing, delivering

JITAIs: Just In Time Adaptive Interventions

- A JITAI is an adaptive intervention
- That is
  - Delivered via mobile devices
  - When needed
  - Where-ever needed

(Kaplan & Stone, 2013; Spruijt-Metz & Nilsen, 2014)

The same elements that we used to describe an adaptive intervention can be used to describe JITAISs, only that now all these elements are momentary.

Dynamic Models of Behavior for Just-in-Time Adaptive Interventions

Donna Spruijt-Metz, University of Southern California Wendy Nilsen, National Institutes of Health

PERVASIVE computing, 1536-1268/14/2014 IEEE

Kaplan, R. M., & Stone, A. A. (2013). Bringing the Laboratory and Clinic to the Community: Mobile Technologies for Health Promotion and Disease Prevention a. *Annual review of psychology*, *64*, 471-498.

Riley, W. T., Rivera, D. E., Atienza, A. A., Nilsen, W., Allison, S. M., & Mermelstein, R. (2011). *Health behavior models in the age of mobile interventions: are our theories up to the task?* : Springer.

Different terms have been used in various fields to describe a JITAI, including dynamic tailoring, intelligent real-time therapy, and dynamically and

individually tailored EMI

Examples:

Intervention to reduce heavy drinking and smoking by young adults

- Participants prompted 3/day by mobile device for assessments
  - Smoking urge, affect, drinking behaviors
- Urge-management interventions delivered by the mobile device ***only*** if an individual reports an urge to smoke at an assessment.

(Witkiewitz et al., in press)

Complete assessments such as smoking urge, drinking and affective state

Witkiewitz, K., Desai, S. A., Bowen, S., Leigh, B. C., Kirouac, M., & Larimer, M. E. (in press). Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. Psychology of Addictive Behaviors.

Example

Reducing Sedentary Behavior by Office Workers

- Software on the computer measures uninterrupted computer time via mouse and keyboard activity
- Smartphone delivers message to encourage a walking activity ***only*** if 30 min. of uninterrupted computer activity occurs

(Dantzig et al., 2013)

8

Quote from paper: whenever 30 min of nearly uninterrupted computer activity was recorded, a short text message (SMS) containing a hyperlink was sent to the participant's smart phone. When participants clicked on this hyperlink, they were shown a message persuading them to be more active. Although all messages contained the same general advice, this advice was phrased in various ways, using four different persuasive strategies. The four strategies are a subset of the six social influence strategies defined by Cialdini [22].

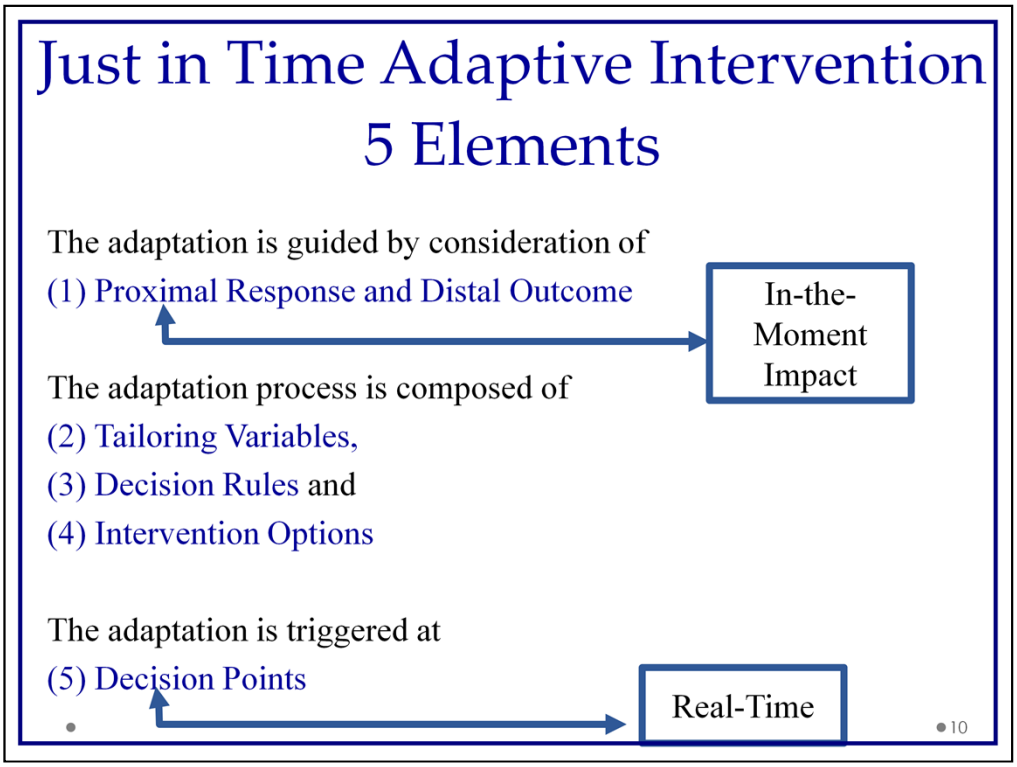
Dantzig, S., Geleijnse, G., & Halteren, A. T. (2013). Toward a persuasive mobile application to reduce sedentary behavior. *Personal and ubiquitous computing, 17*(6), 1237-1246.

Commonalities?

- The intervention package is time-varying and adaptive
- In JITAIs technology plays a critical role
  - Information can be obtained when/where needed
  - Interventions can be delivered when/where needed

The same elements that compose an adaptive intervention, also compose a JITAI.

However, in a JITAI these elements are momentary – they can occur at any moment.



Monitoring, individualizing, delivering

Distal Outcomes

Our goal is to improve a long-term, distal, outcome

- Substance use cessation; maintain increased activity level; improve functionality

To improve the distal outcome, the decision rules are formulated to target proximal responses

11

In a typical adaptive intervention that is not a JITAI (e.g., the drug court program) the outcome is long-term – program graduation at the end of study. This outcome guided the investigators when constructing the decision rule. JITAIs are also guided by a long-term outcome, because the overall purpose is to make some long-term distal impact (e.g., smoking cessation; relapse prevention; promoting a healthier lifestyle etc.). However, the decision rules in a JITAI are also guided by a momentary outcome.

This is because the aim of JITAIs is to generate some immediate impact (otherwise why would you intervene in real-time). The idea is that in real-time the participants needs something, and by providing an intervention, these momentary needs will be met.

This outcome can occur at any time, and is more proximal than the primary long-term outcome

The decision rules are constructed with the aim to manipulate a more proximal/short term.

The term momentary outcome was used by Schwartz & Stone (2007).

Schwartz, J. E., & Stone, A. A. (2007). The analysis of real-time momentary

data: A practical guide. *The science of real-time data capture: Self-reports in health research*, 76-113.

Proximal Responses

Mediators that are critical to achieving the long-term goal

1) Short term targeted behavior
   - Substance use over x hours
   - Physical activity over x minutes
   - Social interactions over next hour
2) Short term risk
   - Momentary craving, stress
3) Engagement in intervention

12

Response-based mediators: Proximal measures of distal outcome

Performance-based mediators-- momentary behavioral, biological, cognitive, or emotional processes, that are critical to address in order to achieve the distal outcome

Engagement-based mediators: concern the type and extent of engagement/adherence needed for participants to attend to, fully take advantage of and benefit from the intervention, burden

Recall that the outcome of interest, is the outcome that guides you when constructing the decision rules.

In a typical adaptive intervention that is not a JITAI (e.g., the drug court program) the outcome is long-term – program graduation at the end of study. This outcome guided the investigators when constructing the decision rule. JITAIs are also guided by a long-term outcome, because the overall purpose is to make some long-term distal impact (e.g., smoking cessation; relapse prevention; promoting a healthier lifestyle etc.). However, the decision rules in a JITAI are also guided by a momentary outcome.

This is because the aim of JITAIs is to generate some immediate impact (otherwise why would you intervene in real-time). The idea is that in real-time the participants

needs something, and by providing an intervention, these momentary needs will be met.

This outcome can occur at any time, and is more proximal than the primary long-term outcome

The decision rules are constructed with the aim to manipulate a more proximal/short term.

The term momentary outcome was used by Schwartz & Stone (2007) to dscribel the primary outcome in real-time momentary data.

Schwartz, J. E., & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. *The science of real-time data capture: Self-reports in health research*, 76-113.

Intervention options

- Intervention options:
  - Behavioral strategies, cognitive strategies, self-monitoring prompts, social linkages, motivational,…
  - When/where to provide recommendations or initiate self-monitoring prompts
  - How to provide an intervention option
  - "Provide nothing" option

- Theoretically/scientifically driven (Klein et al., 2011)

13

Intervention options in JITAIs include types of support (e.g., instrumental, emotional), sources of support (e.g., automated sources, social sources); and modes of support delivery (e.g., support provision and/or support availability).

Recommendations

Reach out recommendation (contact a friend)

Behavioral strategies (exercise; stay in locations that are supportive of change)

Cognitive strategies (relaxation; reframing)

Motivational messages (reasons for behavior change; barriers for change);

Setting goals; modifying goals

Feedback (often with visualization: fish; flower; garden)

Distractions (game, music, etc.)

Michel Klein et al. have a nice review of all the health behavior change theories used to inform EMIs

Klein, M., Mogles, N., & van Wissen, A. (2011). Why won't you do what's good for you? Using intelligent support for behavior change. In *Human Behavior Understanding* (pp. 104-115). Springer Berlin Heidelberg.

Kennedy et al., (2012) conceptualize the EMIs as active assistance.

Kennedy, C. M., Powell, J., Payne, T. H., Ainsworth, J., Boyd, A., & Buchan, I. (2012). Active assistance technology for health-related behavior change: an interdisciplinary review. *Journal of Medical Internet Research*, *14*(3).

Tailoring variables

Tailoring variables are moderators that inform which intervention option is best when, where and for whom.

- Often past proximal responses: craving, activity
- Risk & protective factors: busyness of calendar, momentary mood or stress, location, social context
- Adherence & burden

indicate risk or vulnerability. --internal risk factors , external risk factors: behaviors, social context, geographical location,
When user ignores assessment requests or ignores intervention

the JITAI is often designed to affect the distal outcome by influencing three types of pathways, corresponding to response-based proximal outcomes, performance-based proximal outcomes and engagement-based proximal outcomes. In many cases, intervention scientists might be interested in influencing the distal outcome by building a JITAI that targets all three pathways. In this case, different tailoring variables might be considered for each pathway. For example, assume that an intervention scientists is interested in developing a JITAI that will reduce smoking (i.e., the distal outcome) by (a) reducing momentary smoking urge, and momentary stress, which are two different performance-based pathways, (b) by reducing the number of cigarettes smoked per day, which is a response-based pathway; and by (c) enhancing engagement in the recommended interventions, which

is an engagement-based pathway. In this case, different information from the participant (i.e., tailoring variables) might be needed in order to make decisions that will affect each proximal outcome.

For example, to reduce stress a stress-management intervention might be recommended depending on the participant's momentary level of stress (tailoring variable #1), and to reduce smoking-urge, an urge-management intervention might be recommended depending on the participant's momentary level of smoking urge (tailoring variable #2). Additionally, to reduce the number of cigarettes smoked per day, an encouraging message might be offered at the end of each day depending on the number of cigarettes smoked during that day (tailoring variable #3). Finally, to enhance engagement in the recommended interventions (by minimizing burden), the frequency of intervention options delivered per day might be reduced depending on the number of intervention options the participant received in the prior day (tailoring variable #4). Overall, a JITAI might include multiple tailoring variables and their selection should be guided by the various proximal outcomes the JITAI is intended to achieve.

Decision Points

Typical decision points in JITAIs

- Intervals in time (every x seconds, every x minutes, every x hours)
- When user requests help (presses "help" button")

Frequency is guided by the dynamics of the tailoring variables and "momentaryness" of the intervention option effect.

15

The decision points in a JITAI are also momentary.

Recall that a decision point is the time in which we need to consider intervention options based on what we know about the patient.

In other words, it's the time in which we need to make critical decisions about the intervention options based on patient information.

the selection should be guided by what is known about the extent to which the tailoring variables are expected to change systematically over time, and whether and how change in proximal outcome can be influenced by the intervention options. If the tailoring variable is likely to change in a meaningful manner every (e.g., location), and the intervention options are thought to impact the proximal outcome in 1-2 minutes (e.g., send a warning to the participant), then there should be a decision point every minute. Recall that decision points can result in the "do nothing intervention option," hence a decision point every 3 minutes does not imply an intervention every 3 minutes.

Decision Rules

Link patient information to intervention options at decision points

- A decision rule is implemented at each decision point
- A JITAI often includes many different decision rules
- Choice is guided by an integration of empirical evidence, theory and clinical experience.

16

for every decision point, the investigator should be able to articulate the proximal outcomes and the distal outcome that is reasonable to expect if an individual with a specific value of the tailoring variable receives each of the intervention options under consideration. This has to be expressed for all values of the tailoring variable, and for each one of the proximal outcomes. This is necessary in order to identify the most effective intervention option, for a given value of the tailoring variable, in order to maximally impact the proximal outcome.

Decision Rules: Example 1

What to do when composite risk assessment at random prompt indicates risk

At self-report assessment

If composite substance abuse risk $\geq R_0$

Then, IO = {reminder to access intervention}

Else if composite substance abuse risk $< R_0$

Then, IO = {do nothing}

Intervention options

Tailoring Variable

Proximal Response: Craving

Decision Point

Decision Rules: Example 2

At 5 second intervals

If current accumulated computer activity > P_0

Then, IO= {recommend movement}

Else if current accumulated computer activity ≤ P_0

Then, IO = {do nothing}

JITAI elements

1. Outcomes
   - Distal (scientific/clinical goal) & Proximal response (guided by mediational theories pinpointing the necessary processes needed to achieve the distal outcome)
2. Intervention options
   - Guided by the proximal responses
3. Tailoring variables
   - Guided by theory concerning moderation.
4. Decision points
   - Guided by the dynamics of tailoring variable and momentary nature of the effect of the intervention option.
5. Decision rules

Outline

- Adaptive Interventions and Just-in-Time Adaptive Interventions
- HeartSteps
- Micro-Randomized Trial
- Sample Size Considerations

20

HeartSteps

- Goal: Develop a Just-in-Time Adaptive Intervention for Encouraging and Maintaining Activity

21

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)

HeartSteps

Decision times:

1) Momentary: Approximately every 2-2.5 hours
2) Daily: Each evening at user specified time.

Proximal Response:

Activity (step count) over next 60 minutes (momentary) or daily physical activity (daily).

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.

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

HeartSteps

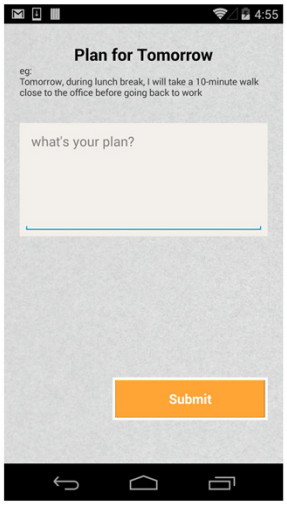
Intervention Options:

1) Whether to provide an intervention
    1) Provide Momentary Lock Screen Activity Recommendation?
    2) Provide Daily Activity Goal Setting?

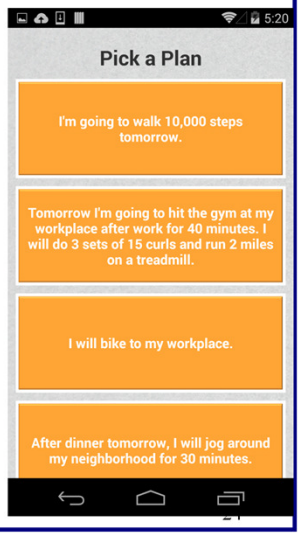
2) Type of intervention to provide if provided
    1) Structured versus unstructured daily activity goal setting

Daily Activity Planning

No Plan or

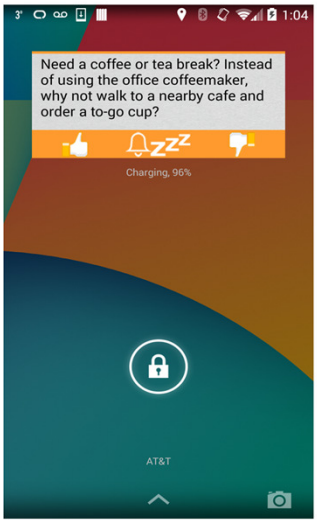


or



Momentary Lock Screen Recommendation

No Message or



25

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

the suggestion, "Need a coffee or tea break? Instead of using the office coffeemaker, why not walk to a nearby cafe and order a to-go cup?," has the following tags:

Location : work

Activity Type : sedentary, active

Time Slot : morning, lunch, afternoon, evening

Weather : outdoor

Day Type : weekday

HeartSteps

Potential Tailoring Variables:
activity recognition (walking, driving, standing/sitting), weather, location, calendar, adherence, step count, availability for momentary intervention, self-report: usefulness, burden

In Heartsteps this is step count over next hour

Outline

- Adaptive Interventions and Just-in-Time Adaptive Interventions
- HeartSteps
- Micro-Randomized Trial
- Sample Size Considerations

27

Data-based approach that can be used in concert with clinical experience, behavioral, biological theory.

Micro-Randomized Trial

A JITAI is a multi-component intervention

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

1) Optimal experimental design is a factorial design
2) Test for proximal *main effects* of the intervention components

28

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

Micro-Randomized Trial

A JITAI is composed of sequences of intervention components with the potential for accumulating habituation and burden

Allow proximal main effects of the intervention components to vary with time

29

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

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

Units are people

What are the factors?

IDEA 1>>>>> A plain vanilla version is

42*5 + 42 binary factors.

42*5 of these binary factors may be labeled as day-time-LockScreenMessage i.e. (32L,morningL) with two levels Lockscreenmessage is on or off on the morning of day 32

42 of these binary factors with labels day-goal, ie. (10G with two levels goal is yes/no on the 10th day)

This seems to be a poor way to conceptualize the factors because it does not recognize in some way the natural smoothness across days or across times within a day. Of course this recognition could instead appear in the modeling and in statements about interactions.

(in the above I am pretending that people's study entries are aligned, that is, everyone starts the study on the same day of the week. In general one would need to allow for incremental recruitment).

We would like to assume smoothness over Day and time of Day. Also the factors have natural orderings so one's understanding of interactions and whether certain higher order interactions are important must incorporate these orderings. So I might expect a two way interaction between (32L,morningL) and (32L, middayL) but not between (10L,morningL) and (32L, morningL). It is also more complicated in that due to the time structure, I might expect a two way interaction between sum of (10G, 11G, 12G, 13G, 14G) and 15G. This might be a negative interaction in the sense that if I have been assigned to many goal activities in the past week, I might ignore the goal activity assigned on day 15. So in sequential factorial designs, there can cumulative interaction effects.

IDEA 2>>>>>>>Perhaps it is better to view this as a sequence of day experiments:

On day 20, the factors are morningL, middayL, afternoonL, afterworkL, eveningL and late-eveG each with two levels. Baseline covariates for day 20 include amount of treatment received on the prior day or in the prior week.

We can consider the main effects of each of these 6 factors as well as which factors may have interactions.

Or view this as a sequence of day*time experiments:

On day 20, late-eveG, we have one factor with two levels. Baseline covariates include dose of lockscreen messages that day, amount of goal-setting over prior week, etc.

IDEA 3>>>>>The problem with the two ideas above is that main effects are defined for each day*time or each day. This is not a parsimonious way to think about main effects. I'd prefer to think of the main effect as a time varying quantity and, in the case of L, maybe time varying within a day. This is what we do in this talk.

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

average main effect of G (average across days) outcome is following day's activity level

Average main effect of L (average across days*times) outcome is next hour's activity level.

It is tempting to view L as nested within G. But G occurs after all the L's on a day. Instead think of L on present day nested within G on prior evening. So average across all prior day G=yes, obtain the average effect of L. This is a nested effect.

The above averaging can take place over the entire study duration or can be just for each week of the study or even for each day of study…..

I like idea 3.

****

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*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

Randomize between intervention options at decision times → Each person may be randomized 100's or 1000's of times.

These are sequential, full factorial, designs.

30

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

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

Units are people

What are the factors?

IDEA 1>>>>> A plain vanilla version is

42*5 + 42 binary factors.

42*5 of these binary factors may be labeled as day-time-LockScreenMessage i.e. (32L,morningL) with two levels Lockscreenmessage is on or off on the morning of day 32

42 of these binary factors with labels day-goal, ie. (10G with two levels goal is yes/no on the 10th day)

This seems to be a poor way to conceptualize the factors because it does not recognize in some way the natural smoothness across days or across times within a day. Of course this recognition could instead appear in the modeling and in statements about interactions.

(in the above I am pretending that people's study entries are aligned, that is, everyone starts the study on the same day of the week. In general one would need to allow for incremental recruitment).

We would like to assume smoothness over Day and time of Day. Also the factors have natural orderings so one's understanding of interactions and whether certain higher order interactions are important must incorporate these orderings. So I might expect a two way interaction between (32L,morningL) and (32L, middayL) but not between (10L,morningL) and (32L, morningL). It is also more complicated in that due to the time structure, I might expect a two way interaction between sum of (10G, 11G, 12G, 13G, 14G) and 15G. This might be a negative interaction in the sense that if I have been assigned to many goal activities in the past week, I might ignore the goal activity assigned on day 15. So in sequential factorial designs, there can cumulative interaction effects.

IDEA 2>>>>>>>Perhaps it is better to view this as a sequence of day experiments:

On day 20, the factors are morningL, middayL, afternoonL, afterworkL, eveningL and late-eveG each with two levels. Baseline covariates for day 20 include amount of treatment received on the prior day or in the prior week.

We can consider the main effects of each of these 6 factors as well as which factors may have interactions.

Or view this as a sequence of day*time experiments:

On day 20, late-eveG, we have one factor with two levels. Baseline covariates include dose of lockscreen messages that day, amount of goal-setting over prior week, etc.

IDEA 3>>>>>The problem with the two ideas above is that main effects are defined for each day*time or each day. This is not a parsimonious way to think about main effects. I'd prefer to think of the main effect as a time varying quantity and, in the case of L, maybe time varying within a day. This is what we do in this talk.

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

average main effect of G (average across days) outcome is following day's activity level

Average main effect of L (average across days*times) outcome is next hour's activity level.

It is tempting to view L as nested within G. But G occurs after all the L's on a day. Instead think of L on present day nested within G on prior evening. So average across all prior day G=yes, obtain the average effect of L. This is a nested effect.

The above averaging can take place over the entire study duration or can be just for each week of the study or even for each day of study…..

I like idea 3.

****

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*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?

Why Micro-Randomization?

- Randomization (+ representative sample) is a gold standard in providing data to assess the causal effect of an intervention.
- Factorial designs are the gold standard when collecting data to build a treatment involving many components.
- Sequential randomizations will enhance replicability of effectiveness of data-based decision rules.

31

HeartSteps (42 day study)

- 210 decision times for the lock-screen activity recommendations.

Lock-screen activity Recommendation?

	Randomization Probability
Yes	$^{2}/_{5}$
No	$^{3}/_{5}$

- 42 decision times for the daily activity goal-setting.

Goal Setting?

	Randomization Probability
Structured	$^{1}/_{4}$
Un-structured	$^{1}/_{4}$
No	$^{1}/_{2}$

32

Sample Size for a Micro-Randomized Trial

- Focus on whether to provide a Momentary Lock Screen Recommendation
- Size to Detect a *proximal main effect* of the Lock Screen Recommendation on Proximal Activity
- Proximal main effect may vary with time

33

Availability & The Main Effect

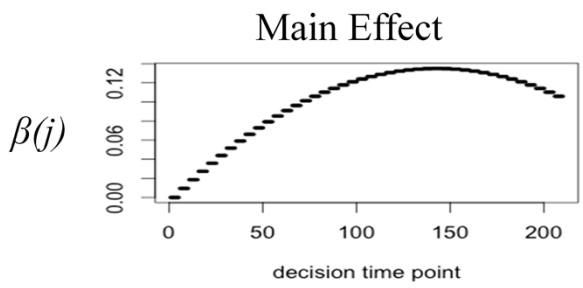
- Momentary interventions can only be delivered at a decision time if an individual is *available*.
- The proximal main effect of treatment at a decision time is the difference in proximal response between *available* individuals assigned a lock-screen message and *available* individuals who are not assigned a lock-screen message

34

The momentary intervention can be turned off for 4 or 8 hours by the participant. The intervention is also off if the participant is currently active (e.g. walking) or in a car and not engaged in a social activity on the phone.

Proximal Main Effect

Proximal main effect of lock-screen message on proximal response is likely time-varying $\beta(j)$, $j=1,\ldots,J$



35

Main effects are marginal effects!

Why would the main effect vary with time? Proximal effect varies with time (maybe diminishes due to habituation). Population of available individuals varies with time. The individuals who are available near the end of the study may be the least sensitive to the influence of the activity message.

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

Sample Size Calculation

- We calculate a sample size to test:

$$H_0 : \beta(j) = 0, j = 1, 2, ...210$$

- A simple approach is to consider $\beta(j)$ as a quadratic with intercept, β_0, linear term, β_1, and quadratic term β_2

and test $\beta_0 = \beta_1 = \beta_2 = 0$

36

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).

Sample Size Calculation

- Our test is based on standard regression.
- To calculate a sample size the scientist needs to specify a clinically/scientifically important effect to detect.

37

Sample Size Calculation

Because we assume the proximal main effect is approximately quadratic, comparisons of the effect of the lock-screen message use both *between person responses* and *within person responses*.

Thus we can expect that the required sample size (number of subjects) will be small.

38

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.

Specify Alternative for Sample Size Calculation

Scientist specifies standardized main effects

- proximal effect on first day,
- average proximal effect over trial duration
- and day of maximal proximal effect.

39

HeartSteps (42 day study)

Standardized main effects:

- initial proximal effect: 0
- average proximal effect over trial duration: ?
- day of maximal proximal effect: 28

40

Meaningful increase in stepcount is 1000/day

Usual std is 2000/day

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

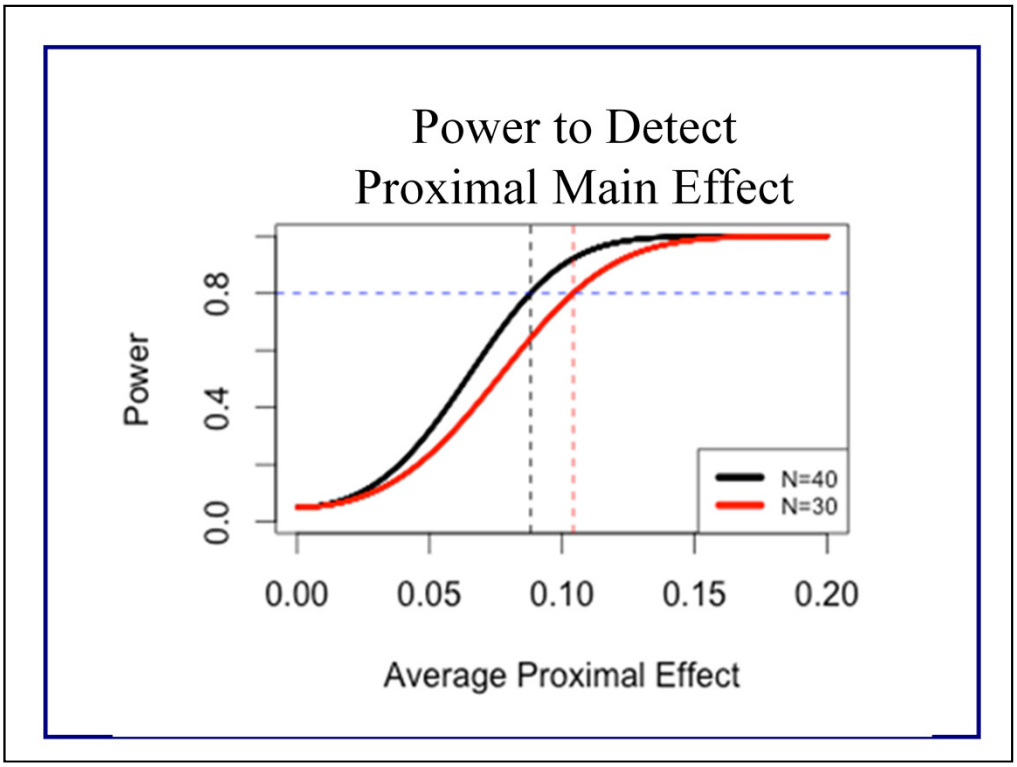
HeartSteps Sample Sizes
Power=.8, α=.05

Standardized Average Proximal Effect	Sample Size For E[R]=.7 or .5
0.06	81 or 112
0.08	48 or 65
0.10	33 or 43

41

Average proximal effect is standardized.

#parameters=6

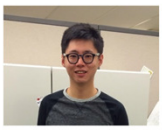

Power to Detect
Proximal Main Effect
Power
0.0
0.4
0.8
0.00
0.05
0.10
0.15
0.20
Average Proximal Effect
N=40
N=30

E[R]=.7

A Micro-Randomized Trial

1) Be conservative in planning the trial!
    1) Under-estimate the amount of time participants are available for the intervention component.
    2) Under-estimate the average proximal effect
2) Power to detect proximal main effect is robust to interactions and to delayed effects (e.g burden)
3) Secondary data analyses concern time varying effect moderation and data analyses for use in constructing data-driven decision rules for the JITAI

43



Collaborators: P. Liao, A. Lee, C. Anderson, P. Klasnja, A. Tewari & Inbal Nahum-Shani



Email if you have questions!

samurphy@umich.edu

Simulation Results
80% Power (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[R]=0.7,q1=0,q3=56,δ=0.4

Simulations indicate:

Method is sensitive to

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

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

Simulations indicate robustness to

R_{j+1} a function of past A_j's

Guess at day of maximal proximal effect (we use a simulation method 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

Heteroscedasticity of 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

Proximal Causal Effect

- Define the Proximal Causal Effect at time j as

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

- What does this estimand mean?

46

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 -- ***the people who have given consent and for whom it is ethical/appropriate to randomize*** This group of people likely depends 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.

Availability

- Define Y_j to be the proximal response (activity level in one hour following a decision time j)
- A person is *available* for an intervention at decision time j if $R_j=1$ and unavailable if $R_j=0$.
- The proximal effect at decision time j is the difference in proximal response between *available* individuals assigned a lock-screen message and *available* individuals who are not assigned a lock-screen message

47

The momentary intervention can be turned off for 4 or 8 hours by the participant. The intervention is also off if the participant is currently active (e.g. walking) or in a car and not engaged in a social activity on the phone.