Microrandomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions

Predrag Klasnja
University of Michigan

Eric B. Hekler
Arizona State University

Saul Shiffman
University of Pittsburgh

Audrey Boruvka, Daniel Almirall, Ambuj Tewari, and Susan A. Murphy
University of Michigan

Objective: This article presents an experimental design, the microrandomized trial, developed to support optimization of just-in-time adaptive interventions (JITAIs). JITAIs are mHealth technologies that aim to deliver the right intervention components at the right times and locations to optimally support individuals' health behaviors. Microrandomized trials offer a way to optimize such interventions by enabling modeling of causal effects and time-varying effect moderation for individual intervention components within a JITAI. ***Method:*** The article describes the microrandomized trial design, enumerates research questions that this experimental design can help answer, and provides an overview of the data analyses that can be used to assess the causal effects of studied intervention components and investigate time-varying moderation of those effects. ***Results:*** Microrandomized trials enable causal modeling of proximal effects of the randomized intervention components and assessment of time-varying moderation of those effects. ***Conclusion:*** Microrandomized trials can help researchers understand whether their interventions are having intended effects, when and for whom they are effective, and what factors moderate the interventions' effects, enabling creation of more effective JITAIs.

Keywords: microrandomized trials, experimental designs, mHealth, intervention optimization

With the advent of capable, affordable, and user-friendly mobile devices, mobile health (mHealth) technologies are increasingly being employed to deliver interventions to users in their natural environments. A common approach is to use the mobile device to enable users to access health interventions wherever and whenever they feel they need help. For example, a person who is trying to quit smoking might access coping strategies when she is experiencing strong cravings. While

helpful, such "pull" interventions rely on a person to be aware and motivated to request the intervention during these states of vulnerability or opportunity (Nahum-Shani, Hekler, & Spruijt-Metz, 2015).

In contrast, a "push" approach to mHealth interventions makes use of sensors, self-report, and computer algorithms to decide when an intervention is needed, and what intervention might be most appropriate. With the advent of increasingly sophisticated sensing devices (e.g., GPS) and phone-based ecological momentary assessment (EMA), it is becoming possible to deliver interventions at moments when they can most readily influence a person's behavior. This may be particularly important for promoting healthy behaviors or supporting cessation of unhealthy ones. For example, for smokers, a moment of temptation to smoke can be a turning point toward relapse or abstinence. For someone trying to increase physical activity, moments when the person can be active are similarly critical decision points when a well-timed intervention could make a difference. Thus, the potential of mHealth interventions may be best realized when they can adaptively respond to individuals' actions and states and deliver intervention options that are most needed, when and where they are most needed (Intille, 2004; Patrick, Griswold, Raab, & Intille, 2008). Such just-in-time adaptive interventions (JITAIs)—mobile applications with explicit decision rules for when to prompt users with specific intervention components (e.g., prompt a person recovering from substance use to access coping strategies when the risk score for relapse reaches a threshold value)—hold great promise for shaping healthy behavior.

Predrag Klasnja, School of Information, University of Michigan; Eric B. Hekler, School of Nutrition and Health Promotion, Arizona State University; Saul Shiffman, Department of Psychology, University of Pittsburgh; Audrey Boruvka, Department of Statistics, University of Michigan; Daniel Almirall, Institute of Social Research, University of Michigan; Ambuj Tewari and Susan A. Murphy, Department of Statistics, University of Michigan.

Correspondence concerning this article should be addressed to Predrag Klasnja, School of Information, University of Michigan, 4364 North Quad, 105 S State Street, Ann Arbor, MI 48109. E-mail: klasnja@umich.edu

Technology required for creating JITAIs has been rapidly developing over the last few years (Hekler, Klasnja, Traver, & Hendriks, 2013). Wearable sensors for detecting physical activity and physiological states are now widely available; modern smartphones can sense their users' environment in real time and mine data from their calendars, email, and other applications. EMA toolkits like PACO (http://pacoapp.com) enable timely, low-burden self-report to supplement or disambiguate passively collected information. Interventions that leverage such technologies can form a rich picture of the users' momentary context and activity. For at least some health behaviors, these developments have made it possible to detect, in real-time, moments when and where an intervention might need to be delivered.

Despite these technological advances, there are barriers to the development of effective JITAIs. Among these, limitations in experimental methodology and theories that guide intervention design are most significant. Commonly used experimental designs are not sufficient to support development of JITAIs because they do not enable researchers to determine empirically when a particular intervention component should be delivered and whether a JITAI that was delivered had the intended effect (Kumar et al., 2013). As such, researchers currently do not have the appropriate tools to gather evidence for deciding how a JITAI should be adjusted to make it more effective. In addition, some researchers have recently argued that the extent to which our behavioral theories can guide the development of JITAIs is limited (Riley et al., 2011; Spruijt-Metz & Nilsen, 2014). In particular, although intervention components included in a JITAI are often based on our best behavioral theories, with rare exceptions, these theories currently do not specify dynamics of human behavior at a granular enough level to guide the design of the decision rules that specify precisely when particular intervention components should be delivered to maximize the likelihood that they will have intended effects and to optimize the efficacy of the intervention as a whole.

In this paper, we propose an experimental design, the microrandomized trial (MRT), that overcomes limitations of current experimental methods and can supplement the use of behavioral theory to guide JITAI development. This experimental design provides empirical data for optimizing JITAIs by enabling researchers to study proximal effects of specific intervention components, changes in these effects over time, and the psychosocial or contextual factors that moderate those time-varying effects.

In a JITAI, each intervention component of interest (e.g., daily step goal in a physical-activity intervention, distraction activities in a smoking-cessation intervention) is constituted by two or more intervention options. These options may correspond to variants of an intervention component or indicate whether an intervention component is on or off. For example, there may be two options for a daily step goal, one corresponding to a user-set goal and a second corresponding to a default, prespecified goal. Similarly, we could conceptualize a distraction intervention as having two options, one in which the user is prompted to engage in a distraction activity (e.g., a video game) when she is experiencing cravings and the second in which a distraction activity is not suggested. Microrandomized trials provide data on how the effects of different intervention options—for example, whether, at a time of craving, quitters smoke less following the delivery of a distraction exercise than when a distraction exercise is not delivered—change over the course of intervention use and how time-varying contextual and psychosocial factors (e.g., the user's location or level of busyness) moderate the observed changes in intervention-component efficacy. Findings from a microrandomized trial can thus help determine decision rules for when and in what circumstances an intervention option should be delivered to optimize its efficacy and the efficacy of the intervention as a whole. In addition, for intervention components that embody theoretical constructs from behavioral science, MRTs help reveal how these theoretical constructs operate over time and in different contexts.

The rest of the paper is organized as follows: we begin by briefly reviewing currently available methods for intervention optimization and discuss why they are insufficient for JITAI development. We then describe the microrandomized trial design and discuss, via an example of a physical-activity intervention we are currently testing, questions that this experimental method enables researchers to answer. We then review important issues in the data analysis of microrandomized trials and consider the method's potential for supplementing the use of behavioral theory in guiding intervention design. We end with a discussion of key methodological and logistical considerations in designing microrandomized trials.

Current Methods for Intervention Optimization

There are three main approaches that can be currently used to optimize JITAIs. As Collins, Murphy, Nair, and Strecher (2005) note, behavioral interventions have traditionally been optimized using a series of randomized controlled trials (RCTs). As an experimental method, RCTs are designed to assess whether, on average, the intervention package as a whole had an effect on the behavior of interest. However, RCTs are not designed to investigate which components of an intervention are efficacious, when they are efficacious, or what psychosocial or contextual factors influenced their efficacy. In secondary analyses of RCT data, random assignment facilitates the assessment of treatment-effect moderation with respect to baseline characteristics (e.g., age), which provides information about static factors that influence efficacy of the intervention package (e.g., Hekler et al., 2013). However, this is insufficient for JITAI development, where we must evaluate what time-varying factors influence the efficacy of different components in order to understand when and in what contexts a particular intervention option should be delivered. If there is sufficient variability across time in receipt of a component due to nonadherence or implementation problems (e.g., participants' medication adherence goes up and down over time), then RCT data can be used to investigate time-varying moderation and effects of time-varying components. However, as is well known, in such secondary analyses, baseline randomization offers no protection against causal confounding, and results are subject to bias. Thus, important questions pertaining to JITAI optimization—when a particular intervention component should be delivered, what factors at the time of delivery affect whether the intervention component will have the desired effect, and so on—are poorly addressed by data from standard RCTs.

As an alternative to RCTs, there has been a resurgence of interest in the use of single-case experimental designs (SCEDs) to develop and evaluate mHealth interventions (Dallery, Cassidy, & Raiff, 2013; Dallery & Raiff, 2014). SCEDs enable highly efficient preliminary efficacy testing of an intervention component, since

each participant acts as his or her own control. However, in their traditional forms (i.e., reversal, multiple-baseline, and changing-criterion designs), SCEDs are of little help for determining the time or context in which a certain intervention option is most efficacious. This is because SCEDs often do not clearly articulate decision points for intervention-component delivery or systematically examine moderators of observed effects.

To overcome the limitations of RCTs in guiding intervention design, Collins and colleagues (Collins, Chakraborty, Murphy, & Strecher, 2009; Collins et al., 2005) proposed the use of factorial experiments as a part of the Multiphase Optimization Strategy (MOST) for multicomponent interventions. Traditional factorial designs can be used to assess the effects of each individual intervention component and key interactions of interest, enabling researchers to choose, based on empirical evidence, which intervention components to include in an intervention package and at what dose. However, traditional factorial designs do not allow the determination of times when it is most effective to deliver each intervention option. Nor do these designs allow researchers to investigate what time-varying factors moderate the relative effect of different time-varying intervention components. These are key questions for JITAI development. Microrandomized trials overcome these limitations of traditional factorial designs, and for JITAI development, they can be incorporated into MOST as an alternative experimental design in the early stages of intervention development.

Microrandomized Trial Design

Microrandomization involves randomly assigning an intervention option at each relevant decision point: a point in time when—based on theory, participant's past behavior, and the participant's current context—a particular intervention component might be efficacious. For a typical multicomponent intervention, multiple components can be randomized concurrently, making microrandomization a form of a sequential factorial design. Since intervention options are randomly assigned at each decision point, a study lasting several weeks or several months may randomize each person hundreds or thousands of times, depending on the frequency at which intervention components being investigated are delivered.

Microrandomized trials are well-suited for optimizing JITAIs for two main reasons. First, microrandomized trials enable assessment of the intervention's time-varying effects. Recall that randomization is used throughout the sciences because, on average, random assignment produces compositional balance in unobserved and unknown factors between the treatment conditions. Differences in outcomes between these conditions can thus be more confidently attributed to the differing treatments, enabling estimation of the causal effect of the treatment (Shadish, Cook, & Campbell, 2002). Because the intervention components being investigated are repeatedly randomized, microrandomized trials allow researchers to assess how intervention components' causal effects change over the course of the study. And second, microrandomized trials are highly efficient. Because each intervention component is repeatedly randomized for each person, effect estimations can take advantage not only of the contrasts in outcomes between people assigned to an intervention option and outcomes among people assigned to a different intervention option (or no

intervention), but also of the contrasts in outcomes between times when each person is randomized to an intervention option and times when the same person is randomized to a different intervention option. These within-subject comparisons enable microrandomized studies to require far fewer participants than traditional full factorial designs.

An Example of a Microrandomized Trial

To better understand the design and value of microrandomized trials, consider HeartSteps, an mHealth intervention for physical activity. HeartSteps consists of a wristband activity tracker that monitors users' steps throughout the day and an Android application intended to encourage walking. For the purposes of this discussion, the phone application contains two main intervention components: (a) daily activity planning and (b) contextually relevant suggestions for physical activity. The daily activity planning component is based on the concept of implementation intentions (Gollwitzer, 1999), which involves the formulation of plans that specify when, where, and how a person will engage in a goal-advancing behavior. Implementation intentions are thought to automate action initiation and have been shown to effectively support engagement in health behaviors even when the person's self-regulatory resources are depleted (Sheeran, Gollwitzer, & Bargh, 2013). The contextually relevant suggestions for physical activity are intended to be actionable in the user's current situation—time of day, location, weather, and day of the week—which (we hypothesize) should make them easier to follow. Two types of suggestions can be sent: suggestions to go for a walk and suggestions to stop being sedentary (e.g., while she is at work, the system might suggest to a user to walk over to the water cooler to get hydrated and stretch her legs). When they receive a suggestion, users can respond to it by pressing the thumbs-up or thumbs-down buttons to indicate they liked or did not like the suggestion, or by pressing "snooze" to indicate that they do not want to receive further suggestions for the next 2, 4, 8, or 12 hr.

To optimize the delivery of daily planning and contextually relevant activity suggestions, we are conducting a microrandomized trial to evaluate the effects of these two intervention components. In our trial, intervention delivery is randomized as follows. Because the planning component is a daily intervention—in the evening, individuals plan physical activities they will do the following day—every evening each person is randomized to one of two options: either prompt the person to plan her physical activity for the next day or not. We are using 50–50 randomization, which means that every evening, there is a 50% chance that a participant will be asked to plan her physical activity for the next day. There are also two main options for the contextually relevant activity suggestions: either deliver an activity suggestion or not. Further, when a person is randomized to receive an activity suggestion, the system randomizes whether to send her a suggestion to go for a walk or to move around to stop being sedentary at 50% probability. The contextually relevant suggestions can be delivered at five time points during the day—morning commute, lunch time, mid afternoon, evening commute, and after dinner. These are periods during which our prior data indicate that people have regular opportunities to be active and thus are potentially appropriate times to provide an activity suggestion. This means that a person can receive up to five contextually relevant activity suggestions in a

single day. In our microrandomized trial, whether or not a suggestion will be sent is randomized for each participant at each of these five time points each day of the study. To reduce participant burden, we use a slightly unequal randomization, randomizing suggestion delivery at 40%, meaning that, on average, a person would receive two activity suggestions per day. Subjects participate in the trial for 6 weeks (42 days), meaning that the trial will generate, per person, 42 data points for daily planning and 210 data points (42 * 5) for activity suggestions.

Distal, Proximal, and Lagged Outcomes

Microrandomized trials are designed to assess causal effects of each randomized intervention component. Defining the intervention outcomes is thus a key aspect of the design of a microrandomized trial. In HeartSteps, the ultimate goal of the intervention as a whole is to help sedentary individuals work up to and then maintain recommended levels of moderate-intensity physical activity—150 min of moderate-intensity activity per week, which is roughly equivalent to 10,000 steps of walking per day (U.S. Department of Health and Human Services, 2008). This is the intervention's desired distal outcome. But distal outcomes are (presumably) achieved through the accumulation of outcomes brought about by the repeated delivery of the application's intervention components. These proximal outcomes are the effects that the delivery of a particular intervention component is intended to have, and they often can be seen as mediators of the desired distal outcome. A first goal of a microrandomized trial is to assess whether the randomized intervention components are having their intended proximal outcomes. In the case of HeartSteps, the goal of activity suggestions is to help people walk right after they receive a suggestion. Hence, for the activity-suggestion component, a reasonable proximal outcome for activity suggestions might be the number of steps taken within 60 min of the suggestion delivery. For the daily planning component, the proximal outcome is the daily step count. Analogous to primary outcomes in RCTs, proximal outcomes may be assessed in the primary analysis of microrandomized trial data, while enabling other analyses.

Note that for both activity suggestions and daily planning, the proximal outcomes are just shorter-time-scale versions of the distal outcomes—in all three cases, the outcomes are steps. This is not always the case. In many interventions, proximal outcomes may be different from distal outcomes; here the proximal outcome might be a mediator. In a relapse-prevention intervention for substance abuse, for example, an intervention component might target social support, which has been shown to be a factor in people's ability to abstain from drug use (Havassy, Hall, & Wasserman, 1991). The proximal outcome of such a component might be the number of daily interactions with abstinence-supporting friends and family members. A microrandomized trial might focus on assessing whether the social-support intervention component is increasing these interactions while also addressing presumed mediating processes by examining if increased social interactions are subsequently associated with a reduced probability of relapse.

A microrandomized trial can investigate not only proximal but also lagged effects. Lagged effects may arise in various ways. For instance, a user of HeartSteps might receive a lunch-time suggestion to walk to a salad place a few blocks from the office, so she can get a healthy lunch and walk 1,000 steps at the same time. The user might not follow the suggestion at the time she received it, but the next day she might remember the suggestion and go to the salad bar even though she did not receive another lunch-time suggestion. An understanding of such lagged effects is critical to developing low-burden JITAIs because they may pinpoint user characteristics or contexts for which the frequency of intervention delivery can be scaled back while maintaining desired effects on the target behavior.

Randomization and Participant Availability

The core concept behind microrandomized trials involves randomizing the options for a particular intervention component each time that intervention component may be delivered. For HeartSteps' daily planning component, which asks participants to plan their activity for the next day, it is feasible and appropriate to deliver the planning activity on any evening. Thus, daily activity planning may be randomized each evening. Sometimes, however, it may be inappropriate to deliver an intervention. For HeartSteps' activity suggestions, we do not want to deliver a suggestion if a participant is already walking or, as a safety precaution, while she is driving (HeartSteps can determine the participant's current activity using Android's built-in activity recognition). When the intervention is not feasible or appropriate at a decision point, we consider the individual unavailable to receive the intervention. In such cases, randomized intervention-component delivery takes place only when the participant is available for the intervention. During unavailable times (e.g., when the person is driving a car), these intervention components would never be delivered, and the data for those decision points would include an "unavailable" indicator. As we discuss in the section on data analysis, keeping track of participant availability is important because effect estimations have to take availability into account.

Research Questions for Microrandomized Trials

We can now be more specific about the research questions that microrandomized trials can address. A microrandomized trial can help researchers answer the following questions:

- What are the proximal and lagged effects of an intervention component?
- How do the proximal and lagged effects of an intervention component change over time?
- Which factors (time-invariant or time-varying) moderate an intervention component's proximal or lagged effects?

Table 1 lists some of the research questions—concrete instantiations of the general research questions we outlined above—that we are investigating in our 6-week microrandomized trial of HeartSteps. Answers to such questions will help us assess whether the activity suggestions in HeartSteps are working as intended (i.e., if they help individuals to get steps at different times of the day) and determine the contexts in which HeartSteps should send suggestions to maximize efficacy. Knowing how suggestions affect activity and user burden, and how participants respond to them by way of changing their walking behavior based on the context of delivery (e.g., the time of the day, location, etc.), would help us formulate decision rules for a revised version of HeartSteps that would deliver relevant types of suggestions at the times and frequencies that maximize their impact on users' walking and

Table 1

Example Research Questions About Contextually Relevant Activity Suggestions That Can Be Answered Through the 6-Week Microrandomized Trial of HeartSteps

Research questions about main effects of contextually relevant activity suggestions
By how much, on average, does providing an activity suggestion increase the step count over the next 60 min relative to no activity suggestion?
By how much, on average, does providing a suggestion to get up and move around increase the step count over the next 60 min relative to a suggestion to take a walk?
How does the effect of an activity suggestion on the step count in the next 60 min change over the course of a 6-week study?
By how much, on average, does providing an activity suggestion change the step count for the remaining part of the same day? On the next day?
How does the number of availability-appropriate activity suggestions delivered in a day impact daily self-report of user burden?

Research questions about moderators of effects of activity suggestions
How does the time of day, the user's location, outside temperature, and the state of the user's calendar (free/busy) influence an activity suggestions' effect on the next 60-min step count?
How do these contextual factors affect whether the user perceives a suggestion as helpful, based on the way the user acknowledged the suggestion notification on her phone?
How does the user's baseline level of physical activity influence the effect of suggesting activity on the next 60-min step count?
How does the user's location and time of day influence efficacy of suggestions targeting disruption of sedentary behavior vs. going for a walk?
How does the influence of these contextual factors on the effect of activity suggestions change over the course of the 6-week study?

minimize impact on perceived burden. It is such optimizations of JITAIs that microrandomized trials are designed to support.

Data-Analytic Issues in Microrandomized Trials

Microrandomized trials generate intensive longitudinal data akin to data from observational EMA studies that repeatedly measure participants' behaviors, context, and psychosocial factors. As such, many of the modeling approaches used to analyze intensive longitudinal data, such as multilevel models (MLM) and the generalized estimating equation (GEE; e.g., Bolger & Laurenceau, 2013; Walls & Schafer, 2006) can also be used in microrandomized trials to assess overall treatment effects and factors that might moderate those effects.

In the following sections, we outline an analytical approach for estimating proximal effects and effect moderation from microrandomized trial data and discuss related data-analytic issues such as bias and participant availability. To illustrate the approach, we use the example of HeartSteps.

Analyses for Assessing Proximal Effects

Because microrandomized trials are sequential factorial designs, primary analyses will often focus on detecting minimum clinically significant differences in the proximal main effects (see Liao, Klasnja, Tewari, & Murphy, 2015, which includes sample size calculations for such primary analyses). Recall that a main effect for a component in a factorial design is an average effect across the effects of the other components and the interactions with the other components. The primary analyses for microrandomized trials—analyses for which ideally the trials should be powered—assess the main effect, that is, whether each randomized intervention component is showing its intended proximal effect.

Proximal effects can be estimated using standard MLM or GEE regression models. Assume that we want to assess whether HeartSteps' evening planning activity increases the next day's step count. Here, a standard MLM with an intercept and a term for evening planning can be used to estimate the proximal

effect. In this model, the estimate of the coefficient for evening planning represents the mean difference in daily step count between person days that follow the planning activity and person days when planning did not occur the previous evening, averaged across all days of the study. This model could be easily expanded to examine whether the effect of planning varies as a function of time by including a model for time (e.g., linear, quadratic, or something more flexible) in the equation. In this case, the time-varying effect of planning will be represented by the coefficients of the interactions between the term for planning and the terms representing the time function (e.g., quadratic).

Finally, as is usual in MLMs, baseline and time-varying covariates that are thought to be related to the daily step count can also be included in the model to improve power. To minimize biasing effect estimates, however, one should only include covariates in the model that are not affected by the intervention (Barber, Murphy, & Verbitsky, 2004). For example, if one wanted to include a measure of time-varying user burden as a covariate, one would only include reported burden measured on the previous day in the model. This is because burden report for the current day could be affected by whether a participant was asked to do a planning activity the prior evening. Covariates unaffected by the intervention can be freely included, however. For example, one could include weather in the model since weather is not affected by participant's planning. Similarly, participant characteristics, such as gender or age, can also be included without biasing the effect estimates.

To illustrate this discussion, equations below provide a simple model for estimating the time-varying effect of evening planning on day t on the next day's step count. For simplicity, the model includes only one contextual covariate, weather. We use the following notation: Y_{it} denotes the ith participant's proximal outcome, in this case the next day's step count; A_{it} denotes the intervention option, which is 1 if evening planning occurred on day t and 0 if it did not; W_{it} denotes weather, which is 1 for good weather and 0 for bad weather (rain or too cold or too hot to walk outside); and B_i denotes the ith participant's baseline average daily

step count. This simple model uses a quadratic function for day to model a time-varying effect of evening planning.

Level 1: Proximal step count is modeled by time, intervention, and weather.[1]

$$Y_{it} = \pi_{0i} + \pi_{1i}t + \pi_{2i}t^2 + \pi_{3i}W_{it} + \pi_{4i}A_{it} + \pi_{5i}A_{it}t + \pi_{6i}A_{it}t^2 + \varepsilon_{it}, t = 1, \ldots, 42 \quad (1)$$

Level 2: Participant differences are modeled by baseline step count.[2]

$$\pi_{ki} = \beta_{k0} + \beta_{k1}B_i + \delta_{ki}, k = 1, \ldots, 6 \quad (2)$$

Participant availability. Effect estimations for intervention components that depend on participant availability require a more nuanced concept of the proximal main effect. Here, the proximal main effect of an intervention component is conceptualized as the difference in the proximal outcome between when the treatment is delivered versus when the treatment is not delivered, but only among individuals who are available for treatment at each decision point. For such intervention components, each data row should include an availability indicator, which is 1 if the participant is available for treatment at that decision point and 0 otherwise. A model to estimate the effect of an intervention component that depends on participant availability should only include the term for the proximal effect in interaction with the availability indicator and never by itself. In a correctly specified model that includes the main effect for availability and all relevant covariates, the estimate of the coefficient of the interaction between the terms for the intervention action and the availability indicator would then represent the average proximal main effect of the intervention component among available occasions.

Note, though, that for some intervention components, the population of available individuals will likely substantially change over the course of the trial. For example, in HeartSteps, individuals for whom the intervention is working might increasingly be walking at decision times for activity suggestions. Similarly, participants who find the intervention annoying might repeatedly snooze the system and would rarely be available to receive a suggestion. For intervention components for which available population is likely to shift over the course of the trial, the proximal effect should be estimated as a time-varying effect; at each time point, the estimated effect then represents the intervention effect on the subpopulation of currently available individuals.

As a secondary analysis, microrandomized trial data can be used to examine predictors of change in availability by modeling the availability indicator as an outcome. This enables researchers to interpret for whom the effect of the intervention component is being estimated at different times over the course of the trial.

Analyses of Time-Varying Moderation

In addition to estimating proximal main effects, analyses of microrandomized-trial data are also able to assess how other time-varying factors—contextual and psychosocial factors such as participants' location, stress level, or annoyance with the intervention—moderate proximal effects of intervention components over time. For mHealth interventions, such questions about time-varying moderation are particularly important because the same individual can receive an intervention in many different contexts. Knowing how the user's context at decision time moderates the effect of an intervention component can help the researchers create decision rules that maximize the likelihood that individuals will be receptive to the intervention when it is delivered and minimize the chances of increasing user burden.

As with main effects, treatment-effect moderation can be assessed using MLM and GEE regression approaches by estimating the coefficients of the interactions of the terms for randomized intervention components and for the time-varying moderators of interest. Consider, for instance, assessing how weather influences the effect of the prior evening's planning on the daily step count. In a correctly specified model, the coefficients for the three-way interactions involving the terms for weather, evening planning, and the function for time would estimate how the dampening influence of bad weather on the effect of daily planning on the next day's step count changes over the course of the study.

Power

While a detailed discussion of sample size calculations for microrandomized trials is beyond the scope of this paper, primary analyses can be conservatively powered using regression-based techniques such as GEE. Our preliminary power calculations show that microrandomized trials are highly efficient (Liao et al., 2015). For HeartSteps activity suggestions, we make the conservative assumptions that the proximal main effect on the first study day is 0, that the maximum proximal main effect will be seen half-way through the study, and that participants will be available for an activity suggestion only 50% of the time. Our calculations show that a 6-week study would be able to detect a standardized proximal effect of 0.10 (small by Cohen's (1992) standards), with a type-I error rate of $\alpha = .05$ and 80% power using only 43 participants. If we instead assume that participants would be available for a suggestion 70% of the time, the needed sample size goes down further, to 33 participants. The high efficiency of microrandomized trials is to a great extent due to their ability to take advantage of both between-subjects and within-subject contrasts in the proximal outcome.

Microrandomized Trials and Theoretical Grounding of JITAI Design

At their best, JITAIs consist of intervention components that embody behavior-change techniques derived from our best behavioral theories (Michie et al., 2011). But while theory has been immensely helpful in guiding the design of intervention components themselves, it has been limited in its ability to guide other

---

[1] The level 1 residuals, ε_{ti}, are usually assumed to be independent, normally distributed with mean zero and constant variance, σ^2, and uncorrelated with

$$W_{it}, A_{it}, B_i, i = 1, \ldots, n; t = 1, \ldots, 42.$$

[2] The level 2 residuals, $\delta_i, i = 1, \ldots, n$, are usually assumed to be multivariate normal with mean zero, and uncorrelated with

$$\varepsilon_{it}, W_{it}, A_{it}, B_i, i = 1, \ldots, n; t = 1, \ldots, 42.$$

aspects of JITAI design. The problem is twofold. First, we currently do not have a mature body of knowledge about how to translate abstract theoretical constructs into technological behavior-change intervention components. A growing literature in human-computer interaction has shown that implementation details matter, and that decisions such as the location of self-monitoring feedback (e.g., presented within the application vs. on the phone's lock screen) or the form of the reward for goal-attainment (e.g., trophies vs. points) substantially impact user experience of an intervention and, in turn, its ability to support behavior change (Consolvo, Klasnja, McDonald, & Landay, 2014). Decisions about how to implement particular behavior-change techniques in a JITAI are thus often left to researchers' judgment, or they evolve through an iterative design process. Second, behavioral theories in their current form are often not granular enough to guide the design of decision rules for the delivery of intervention components (Riley et al., 2011; Spruijt-Metz & Nilsen, 2014). Even for theories with dynamic constructs (e.g., Witkiewitz & Marlatt, 2004), the current evidence base—often derived from studies that used temporally sparse data—lacks sufficient information about the optimal timing of intervention delivery. For example, how soon should a physical-activity goal be decreased if a participant starts to repeatedly miss the current goal? In what circumstances should individuals be asked to create implementation intentions for frequent behaviors like healthy eating?

We suggest that microrandomized trials can help build the knowledge base in both of these areas by gathering empirical evidence for principled JITAI design and evidence about the dynamics of interactions among intervention components, psychosocial processes, and contextual variables needed to refine and extend current behavioral theories.

An important role microrandomized trials can play in advancing the science of design of behavior-change technologies is in comparing different implementations of a behavior-change technique. For example, one might microrandomize different ways of providing feedback about the cigarettes that participants smoked and resisted each day, by varying, for instance, the amount of information provided in the representation or how the feedback is framed. By gathering data about the various outcomes that such feedback is intended to have—for example, increasing self-efficacy for quitting, improving participants' awareness of the contexts in which they are most likely to smoke, and the number of cigarettes smoked the following day—such studies could begin to build an evidence base about the comparative efficacy and tradeoffs of various implementations of a behavior-change technique for different health behaviors, populations, and contexts. Findings from such studies could supplement the current taxonomies in guiding the translation of behavioral theories into concrete interventions.

With regard to theory, microrandomized trials could help gather the evidence to elucidate the dynamic aspects of behavior-change processes needed to guide the design of decision rules for intervention delivery. For example, a microrandomized trial could explore how planning interacts with self-efficacy and daily routines to shape health behaviors. Such a trial might randomize whether participants plan how they will eat healthily the following day, and would measure each day self-efficacy for healthy eating, how healthily participants ate that day, and the busyness of participants' calendars. Such a trial could investigate whether plan-

ning increases healthy eating only when self-efficacy is low or whether it tends to be more helpful when participants' are unusually busy or traveling. Just as importantly, such a trial would uncover whether these relationships change over time, as participants start habituating to the planning activity or their eating patterns become more routinized. Results of such studies would help researchers better understand the dynamics of psychosocial processes such as planning, while also enabling more principled design of decision rules for interventions based on those processes. Which constructs and processes such trials should investigate will be determined by the researchers' understanding of the aspects of the current theories that need further elucidation. But insofar as theoretical constructs of interests can be measured repeatedly in a low-burden way, and can be affected through push intervention components, microrandomized trails can help elucidate their temporal dynamics.

Implementation Considerations for Microrandomized Trials

There are several important practical considerations that go into the design of microrandomized trials, such as decisions about intervention components that are randomized and the selection of proximal outcomes. We discuss these implementation issues next.

Deciding Which Components to Randomize and How

A key consideration in the design of a microrandomized trial is the decision regarding which intervention components and options to randomize. This choice should be governed by a combination of scientific, design, and usability considerations. In general, one would randomize intervention components for which there is insufficient scientific evidence for their effect on the proximal outcome (e.g., contextual activity suggestions in HeartSteps), or for which we do not sufficiently understand the dynamics of their operation over time (e.g., HeartSteps' daily planning activities). In both of these cases, a microrandomized trial would yield scientific knowledge about new intervention strategies and the psychosocial processes embodied in these strategies. Another reason to microrandomize an intervention component is to guide selection among several implementations of an intervention strategy, such as different realizations of daily goal-setting or representations of goal-attainment rewards. In this case, a microrandomized trial would take place at an early stage of intervention development to inform the design of the components that researchers are considering for the inclusion in the intervention. Finally, one could randomize among different types of interventions that can be delivered at the same decision points, such as different coping strategies for drug-use cravings. Here, a microrandomized trial would reveal for whom and in what circumstances each type of intervention strategy is most likely to be efficacious.

Another consideration is the randomization rate for different intervention components. While a 50–50 randomization provides most power, there are cases where a different randomization rate might make sense. One such scenario is when the researcher wants to concurrently determine whether an intervention strategy is efficacious and which version of that strategy works best. In this case, it can make sense to randomize nondelivery and overall

delivery of the intervention component at 50–50 to maximize the likelihood that an effect can be detected, and then break down the 50% delivery portion among different versions of the intervention strategy (e.g., randomize version 1 and version 2 at 25% each). Another scenario when a skewed randomization ratio can be useful is when the intervention component is highly burdensome. Reducing the rate from 50–50 to a lower percentage for intervention delivery, as we have done for activity suggestions in HeartSteps, can help gather data about the effects of the intervention while minimizing the risk that participants will frequently decline the intervention or abandon the system altogether due to excessive burden.

Defining Proximal Outcomes

Another key consideration involves defining proximal outcomes for randomized intervention components. For some components, proximal outcomes can be relatively clear and easily measurable. For instance, for a goal-setting component of a walking intervention, the attainment of the daily step goal is an obvious proximal outcome. For other components, no such unambiguous proximal outcomes present themselves. For example, what is the right proximal outcome for the HeartSteps' contextual activity suggestions? Is it whether or not the specific activity contained in the message was followed as suggested? Is it the absolute step count for a period after the suggestion was delivered? If so, over what period? 10 min? 30 min? An hour? Is it the change in the step count for a period following the suggestion over the step count for an equally long period prior to the suggestion? A case could be made for any of these outcomes and others in turn. When no obvious proximal outcome exists, researchers need to make a decision based on their judgment and logistical considerations (e.g., technical feasibility or burden of measuring a particular type of outcome). For HeartSteps' activity suggestions, the only way to accurately tell whether the person followed the actual suggestion that was sent to her would be to ask. This would not only introduce additional burden but also act as an intervention in its own right, confounding the effects of the suggestions themselves. For these reasons, we opted for a less direct outcome (step count over the next hour) that can be assessed passively. In addition, our chosen outcome allows the intervention to get credit for activity that might be different from the activity suggested.

When the chosen proximal outcome is not a shorter-time-scale version of the distal outcome—for example, when a component of a smoking intervention targets stress reduction and not actual smoking—that choice rests on a theory of mediation (i.e., that individuals smoke to reduce stress), which must be tested or theoretically justified. The data collected in a microrandomized trial allows for testing of such mediational relationships.

Limitations

Although they are a powerful way of optimizing JITAIs, microrandomized trials have several limitations. There are three requirements for the use of microrandomized trials for intervention optimization. First, these trials are only applicable for testing of push interventions—interventions such as reminders or prompts to set goals that are delivered to individuals based on a set of decision rules. Microrandomized trials are not suitable for testing of inter-

vention components that are made available to individuals but which individuals access at will. If a researcher wants to microrandomize a pull intervention component (e.g., graphs for providing feedback on a health behavior), that component first needs to be converted into a push intervention (e.g., a notification to access the graphs). Given that push interventions can be burdensome, however, microrandomization will be typically limited to a subset of components of an mHealth intervention. Second, given that the analyses of microrandomized trial data focus on proximal outcomes, microrandomized trials are most appropriate for testing of intervention components for which proximal outcomes can be defined in a principled way (i.e., theory-based components for which theory specifies the outcomes that the components will directly impact) and for which the proximal outcomes can be measured in a low-burden way (passively or through brief self-report) so that they can be assessed each time the intervention is randomized. Finally, as a within-subject design, microrandomized trials are not suitable for testing of interventions for very rare events, such as interventions for prevention of manic episodes in the bipolar disorder, which would be experienced by few people during a study and seldom experienced repeatedly by any one person over the course of the trial. Interventions for such rare events should be optimized through traditional cross-sectional designs.

Conclusion

JITAIs have the potential to greatly enhance how we support people's efforts to adopt and sustain healthy behaviors. The development of effective JITAIs has been stymied, however, by a lack of appropriate optimization methods. Microrandomized trials have the potential to address this limitation, enabling the creation of interventions that can effectively support individuals whenever and wherever they most need the support.

References

Barber, J. S., Murphy, S. A., & Verbitsky, N. (2004). Adjusting for time-varying confounding in survival analysis. *Sociological Methodology, 34,* 163–192. http://dx.doi.org/10.1111/j.0081-1750.2004.00151.x

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY: Guilford Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Collins, L. M., Chakraborty, B., Murphy, S. A., & Strecher, V. (2009). Comparison of a phased experimental approach and a single randomized clinical trial for developing multicomponent behavioral interventions. *Clinical Trials, 6,* 5–15. http://dx.doi.org/10.1177/1740774508100973

Collins, L. M., Murphy, S. A., Nair, V. N., & Strecher, V. J. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine, 30,* 65–73. http://dx.doi.org/10.1207/s15324796abm3001_8

Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2014). Designing for healthy lifestyles: Design considerations for mobile technologies to encourage consumer health & wellness. *Foundations and Trends in Human-Computer Interaction, 6,* 167–315.

Dallery, J., Cassidy, R. N., & Raiff, B. R. (2013). Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of Medical Internet Research, 15,* e22. http://dx.doi.org/10.2196/jmir.2227

Dallery, J., & Raiff, B. R. (2014). Optimizing behavioral health interventions with single-case designs: From development to dissemination. *Translational Behavioral Medicine, 4,* 290–303. http://dx.doi.org/10.1007/s13142-014-0258-z

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493–503. http://dx.doi.org/10.1037/0003-066X.54.7.493

Havassy, B. E., Hall, S. M., & Wasserman, D. A. (1991). Social support and relapse: Commonalities among alcoholics, opiate users, and cigarette smokers. *Addictive Behaviors, 16,* 235–246. http://dx.doi.org/10.1016/0306-4603(91)90016-B

Hekler, E. B., Buman, M. P., Otten, J., Castro, C. M., Grieco, L., Marcus, B., . . . King, A. C. (2013). Determining who responds better to a computer- vs. human-delivered physical activity intervention: Results from the community health advice by telephone (CHAT) trial. *The International Journal of Behavioral Nutrition and Physical Activity, 10,* 109. http://dx.doi.org/10.1186/1479-5868-10-109

Hekler, E. B., Klasnja, P., Traver, V., & Hendriks, M. (2013). Realizing effective behavioral management of health: The metamorphosis of behavioral science methods. *IEEE Pulse, 4,* 29–34. http://dx.doi.org/10.1109/MPUL.2013.2271681

Intille, S. S. (2004). Ubiquitous computing technology for just-in-time motivation of behavior change. *Studies in Health Technology and Informatics, 107,* 1434–1437.

Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., . . . Swendeman, D. (2013). Mobile health technology evaluation: The mHealth evidence workshop. *American Journal of Preventive Medicine, 45,* 228–236. http://dx.doi.org/10.1016/j.amepre.2013.03.017

Liao, P., Klasnja, P., Tewari, A., & Murphy, S. A. (2015). *Micro-randomized trials in mhealth (University of Michigan Tech. Rep. No. 537).* Ann Arbor: University of Michigan.

Michie, S., Ashford, S., Sniehotta, F. F., Dombrowski, S. U., Bishop, A., & French, D. P. (2011). A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. *Psychology & Health, 26,* 1479–1498. http://dx.doi.org/10.1080/08870446.2010.540664

Nahum-Shani, I., Hekler, E. B., & Spruijt-Metz, D. (2015). Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology, 34*(Suppl), 1209–1219. http://dx.doi.org/10.1037/hea0000306

Patrick, K., Griswold, W. G., Raab, F., & Intille, S. S. (2008). Health and the mobile phone. *American Journal of Preventive Medicine, 35,* 177–181. http://dx.doi.org/10.1016/j.amepre.2008.05.001

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? *Translational Behavioral Medicine, 1,* 53–71. http://dx.doi.org/10.1007/s13142-011-0021-7

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* New York, NY: Wadsworth Cengage learning.

Sheeran, P., Gollwitzer, P. M., & Bargh, J. A. (2013). Nonconscious processes and health. *Health Psychology, 32,* 460–473. http://dx.doi.org/10.1037/a0029203

Spruijt-Metz, D., & Nilsen, W. (2014). Dynamic models of behavior for just-in-time adaptive interventions. *Pervasive Computing, IEEE, 13,* 13–17. http://dx.doi.org/10.1109/MPRV.2014.46

U.S. Department of Health and Human Services. (2008). *2008 physical activity guidelines for Americans.* Washington, DC: Author.

Walls, T. A., & Schafer, J. L. (Eds.). (2006). *Models for intensive longitudinal data.* New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195173444.001.0001

Witkiewitz, K., & Marlatt, G. A. (2004). Relapse prevention for alcohol and drug problems: That was Zen, this is Tao. *American Psychologist, 59,* 224–235. http://dx.doi.org/10.1037/0003-066X.59.4.224