ann. behav. med. (2019) 53:573–582
DOI: 10.1093/abm/kay067

REGULAR ARTICLE

Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps

Predrag Klasnja, PhD[1,2] • Shawna Smith, PhD[3,4] • Nicholas J. Seewald, MS[5] • Andy Lee, BS[2] • Kelly Hall, BS[5] • Brook Luers, MS[5] • Eric B. Hekler, PhD[6] • Susan A. Murphy, PhD[7]

Published online: 5 September 2018
© Society of Behavioral Medicine 2018. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Abstract

Background HeartSteps is an mHealth intervention that encourages regular walking via activity suggestions tailored to the individuals' current context.

Purpose We conducted a micro-randomized trial (MRT) to evaluate the efficacy of HeartSteps' activity suggestions to optimize the intervention.

Methods We conducted a 6-week MRT with 44 adults. Contextually tailored suggestions could be delivered up to five times per day at user-selected times. At each of these five times, for each participant on each day of the study, HeartSteps randomized whether to provide an activity suggestion, and, if so, whether to provide a walking or an antisedentary suggestion. We used a centered and weighted least squares method to analyze the effect of suggestions on the 30-min step count following suggestion randomization.

Results Averaging over study days and types of activity suggestions, delivering a suggestion versus no suggestion increased the 30-min step count by 14% (p = .06), 35 additional steps over the 253-step average. The effect was not evenly distributed in time. Providing any type of suggestion versus no suggestion initially increased the step count by 66% (167 steps; p < .01), but this effect diminished over time. Averaging over study days, delivering a walking suggestion versus no suggestion increased the average step count by 24% (59 steps; p = .02). This increase was greater at the start of study (107% or 271 additional steps; p < .01), but decreased over time. Antisedentary suggestions had no detectable effect on the 30-min step count.

Conclusion Contextually tailored walking suggestions are a promising way of initiating bouts of walking throughout the day.

Clinical Trial information This study was registered on ClinicalTrials.gov number NCT03225521.

Keywords Mobile health • Micro-randomized trials • Physical activity • Walking • Behavior change

✉ Predrag Klasnja
klasnja.p@ghc.org

1 Kaiser Permanente Washington Health Research Institute, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, USA

2 School of Information, University of Michigan, Ann Arbor, MI, USA

3 Insitute for Social Research, University of Michigan, Ann Arbor, MI, USA

4 School of Medicine, University of Michigan, Ann Arbor, MI, USA

5 Department of Statistics, University of Michigan, Ann Arbor, MI, USA

6 School of Medicine, University of California San Diego, San Diego, CA, USA

7 Department of Statistics, Harvard University, Cambridge, MA, USA

Introduction

Recent advancements in wearable sensors and mobile phones have greatly increased the sophistication of mobile health (mHealth) interventions. It is now possible to use data about individuals' physiological states (e.g., stress), behavior (e.g., physical activity), and context (e.g., location) to deliver highly tailored interventions for a range of health behaviors, from substance use [1] to stress management [2]. Such just-in-time adaptive interventions (JITAIs) [3–5] aim to provide support for health behavior change at times when users most need the support and are most receptive to it [5]. Yet, while JITAIs hold promise for supporting health behavior change, their design involves several challenges. JITAIs often include *push* interventions, which are usually delivered as notifications on a mobile phone or a wearable (e.g., a smart watch) at times (decision points) when the

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

system has determined that a particular intervention component, such as a medication reminder or a tip for coping with alcohol cravings, could be helpful and the user would be receptive to it. Whether to deliver a notification and, if so, what type of notification to deliver are determined by JITAI *decision rules*—if-then rules or algorithms that are typically constructed based on behavioral theory and prior data.

While JITAI designers aim to provide interventions that are effective and are delivered at most helpful times for each intervention component, the available empirical evidence and clinical judgment are often not sufficient to determine in advance how exactly each component should work and when it should be delivered. But this creates a risk. If the push interventions included in a JITAI are not sufficiently helpful or are provided too frequently, they can lead to user frustration and disengagement. Users might also become habituated to the interventions [6], leading to the loss of effectiveness even for intervention components that were initially effective. And if the users become sufficiently aggravated by interventions that are being delivered too frequently or at inopportune times, they might abandon the JITAI altogether.

Optimizing [7–9] JITAIs to maximize effectiveness while minimizing user burden and risk of abandonment is thus paramount. Ideally, JITAIs should only include intervention components that are efficacious, and these components should be delivered when they are most likely to have an impact on target health behaviors and with frequency that minimizes user burden and risk of habituation. JITAI construction can, thus, be greatly facilitated by optimization studies that can evaluate the efficacy of individual intervention components to determine whether they are worth including in a JITAI and to refine the decision rules for their delivery.

Micro-randomized trials (MRTs) [10, 11] are a new experimental design developed to optimize JITAI construction. In an MRT, push intervention components that are being investigated are randomized for each participant each time the system may deliver those components. For example, in an MRT of a JITAI that includes medication reminders for a daily hypertensive medication, reminders would be randomized for each participant on each day of the study. A different component—a mindfulness reminder—would be randomized at each of its own decision points—say, three times a day.

Data from an MRT enable modeling of the causal effects of each intervention component that was micro-randomized. MRT analyses are designed to assess the causal effect of each randomized component on its *proximal* outcome: the short-term, measurable behavioral or psychosocial effect through which that component is hypothesized to mediate desired distal health outcomes. In our medication-reminder example, the proximal outcome would be whether the person took the hypertensive medication. Assessment of proximal outcomes enables researchers to determine both whether each component is working as intended and, in MRTs in which distal health outcomes are also assessed, whether the proximal and distal outcomes are related as hypothesized [12, 13]. Finally, as MRTs both randomize intervention provision and assess proximal outcomes repeatedly, MRT data can enable researchers to model how intervention effects change over time (e.g., reminders may be effective initially but their effectiveness diminishes over time). In addition, in secondary analyses, researchers can also examine how those effects are moderated by the context in which interventions are provided. For example, medication reminders may be effective when the person is at home, but not when the person is out and about. From the perspective of JITAI construction, such analyses of how intervention effects are moderated by time and context can be particularly important for optimizing rules for intervention delivery [3].

In this study, we present results from a 6-week MRT conducted to optimize HeartSteps, a novel JITAI we developed to help individuals increase their physical activity by engaging in regular bouts of activity throughout the day, primarily by walking. Walking is a moderate-intensity activity that is safe, accessible, and can be performed almost anywhere [14]. At the same time, walking is shaped by a number of contextual factors, such as weather, built environment, and available transportation options [15–17]. To encourage regular walking, HeartSteps leverages information about users' context (location, weather, time of day, and day of the week) to send contextually tailored suggestions for how users can be active in their current situation. In addition, HeartSteps supports physical activity by helping users plan when, where, and how they will be active on the following day.

To optimize HeartSteps, we conducted an MRT to assess the efficacy of HeartSteps' contextually tailored activity suggestions to determine whether this component should be included in the subsequent version of the intervention package. Specifically, we were interested in evaluating whether providing an activity suggestion tailored to the user's current context, versus not providing a suggestion, increased physical activity shortly after suggestions are delivered and, if so, whether this effect changed over time.

Methods

The HeartSteps Intervention

HeartSteps is a JITAI we are developing to help individuals increase and sustain physical activity. The initial version of HeartSteps was an Android application focused

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

on helping sedentary adults with regular daytime jobs get more activity throughout the day, primarily by walking. HeartSteps contained both *pull* and *push* intervention components. Pull components included a graph of the day's steps to enable self-monitoring and a list of previously received activity suggestions. Participants could access these interventions at any time by opening the HeartSteps application on their phones. Push components were delivered to participants via Android notifications at times determined by the intervention decision rules. Push components included contextually tailored activity suggestions and daily prompts to write implementation intentions [18, 19] for how participants will be active on the following day. The main components of HeartSteps are shown in Fig. 1. The current article focuses on HeartSteps' activity suggestions.

Activity suggestions were intended to serve as cues to action [20, 21], increasing readiness to engage in activity by providing ideas for how participants could be active in their current context. To make suggestions actionable, we tailored them to four dimensions of context: time of day, day of the week (weekday vs. weekend), location (home, work, other), and weather (suitable for walking outside, not suitable for walking outside, snow). Suggestions could be delivered at five times (decision points) each day: during morning commute, lunchtime, midafternoon, evening commute, and after dinner. These times were chosen as they are common times when individuals who work outside the home have opportunities to engage in physical activity, by, for example, parking further from the office or walking to a restaurant for lunch. Participants could configure the specific times of the five decision points based on their schedules.

HeartSteps contained two types of activity suggestions: suggestions to walk (intended to encourage bouts of 500 to 1,000 steps) and suggestions to disrupt sedentary behavior (to stand up, stretch, and/or move around). We initially included antisedentary suggestions primarily to help participants build engagement with the intervention, as they were easier to act on and could be written to be interesting and varied. We hypothesized, however, that antisedentary suggestions might still result in walking even though longer walks were not explicitly suggested. For instance, a recent study with bariatric patients [22] found that once individuals started moving, they tended to move beyond the amount of time prescribed by suggestions for brief walks. We hypothesized that HeartSteps antisedentary suggestions might have a similar effect. HeartSteps included over 500 activity suggestions developed by the research team using a combination of crowdsourcing and expert editing and review (see Table 1 for examples). The initial message set

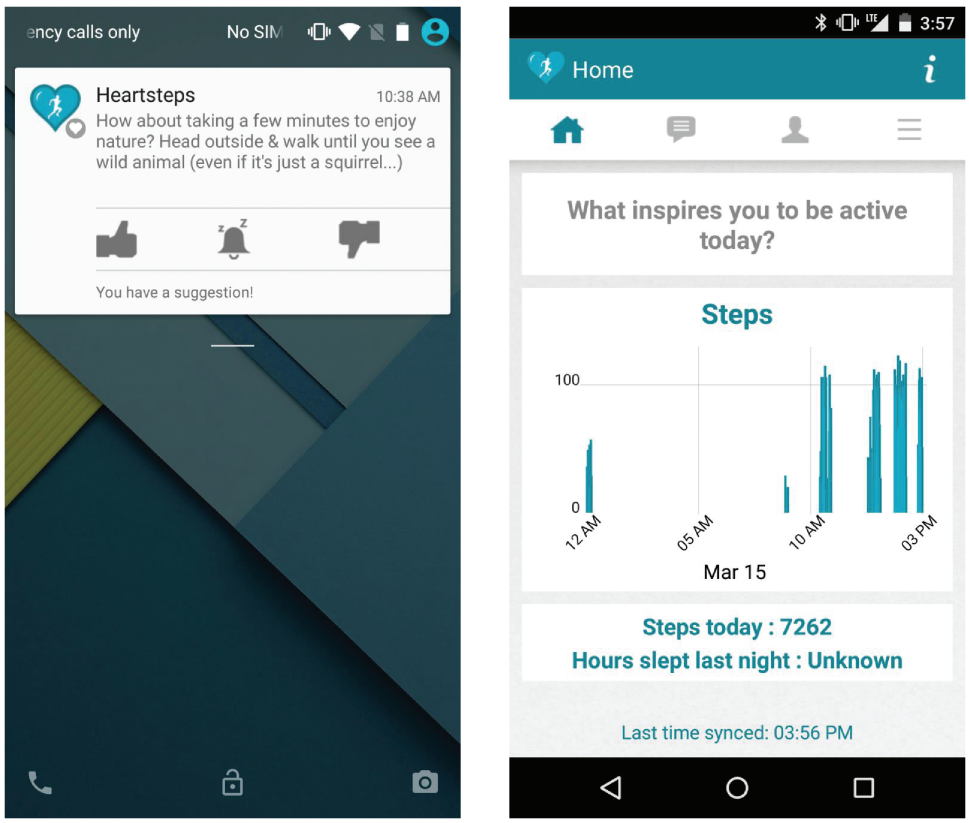

Fig. 1. Components of the HeartSteps system: an activity suggestion, delivered as a notification to the phone's lock screen (left); graph of the day's steps in the HeartSteps application (right).

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

Table 1 Examples of HeartSteps activity suggestions with their corresponding contexts

Suggestion	Type	Time of day	Day of the week	Weather	Location
Is there anything better than weekend afternoons? You could enjoy the fresh air by taking a walk around your neighborhood!	Walking	Midafternoon	Weekend	Suitable for walking outside	Home
While eating at your desk can be tempting, lunch is your time for a much-needed mental break. Can you take a few minutes to stand & stretch?	Antisedentary	Lunch	Weekdays	Any	Work
This evening would be a good time for a challenge: a walk in bad weather. Bundle up & enjoy the outdoors!	Walking	Postdinner	Weekdays Weekends	Snow	Home
Have a long conference call today? Walking in place or pacing while you talk can keep you engaged and increase your step count!	Walking	Morning Midafternoon	Weekdays	Not suitable for walking outside	Work
When's the last time you dusted? If you have 5 minutes, you could grab a feather duster & take a crack at those end tables!	Antisedentary	Morning Lunch Midafternoon Late afternoon	Weekdays Weekend	Any	Home
Have you been sitting at your desk all morning? Why not stand up & do some light stretches for 2–3 minutes? It will energize you for the rest of the day!	Antisedentary	Lunch	Weekdays	Any	Work

was produced using Mechanical Turk, by asking Turk workers to generate text messages that could be sent to people's phones to motivate them to walk (for walking suggestions) or get up and move around (for antisedentary suggestions). Workers were given tips on how to write effective health communication, developed by an experienced health communication researcher, and all resulting messages were rated and edited in the subsequent Turk tasks to improve the quality of the resulting messages. The highest rated messages were then further edited by our research team. In addition, the research team created additional suggestions to ensure that we had a sufficient number of messages for each combination of contextual factors (e.g., walking suggestions for weekend mornings when participants are at home and the weather is good).

To track participants' steps, HeartSteps used Jawbone Up Move, a small, screenless wrist-worn activity tracker with a 6-month battery life. Jawbone recorded minute-level data about participants' steps, which HeartSteps accessed via Jawbone's application programming interface.

Availability

HeartSteps would deliver activity suggestions to participants only if, at a decision point, they were considered to be *available* for treatment. To minimize burden, participants were considered to be unavailable if they were currently walking or running, or had just finished an activity bout in the previous 90 s. To mitigate safety concerns, participants were also considered to be unavailable while driving. HeartSteps determined availability at each decision point by using Android's activity recognition algorithms that automatically analyze data from the phone-based accelerometer to classify participants' current activity. Finally, because suggestion delivery required Internet connectivity, participants were considered to be unavailable when their phones were offline.

Participants

We recruited healthy sedentary adults who were interested in increasing their walking. Recruitment was conducted using fliers posted around campus, in coffee shops, and in grocery stores, and via Facebook ads. Recruitment took place from August 2015 to January 2016. Participants were eligible if they were English speakers, between 18 and 60 years of age, could walk for exercise without discomfort, and either had a full-time daytime job or a regular schedule outside the home (e.g., students). Participants needed to have a personal phone running Android 5.0 or higher or be willing to use a study-provided phone as their primary phone for

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

the duration of the study. We excluded individuals who needed medical supervision to exercise and those who were currently using an activity tracker (e.g., a Fitbit). The latter exclusion criterion was intended to manage participant burden (and, thus, adherence), as most activity-tracking applications already send users multiple notifications a day, to which HeartSteps would be adding several additional notifications. Participants were compensated up to $90 for study participation. Informed consent was obtained from all individual participants included in the study.

Study Design

HeartSteps was evaluated in a 6-week MRT [10, 11]. The study was approved by our university's institutional review board. At intake, participants were interviewed about their physical activity habits and prior experiences with mHealth interventions. Study staff then provided participants with the Jawbone tracker and installed HeartSteps either on the participant's personal phone or a study phone. In the latter case, study staff also transferred the participant's personal information (contacts, calendar, text messages, etc.) to the study phone using a commercial application for data migration between mobile phones. Participants were instructed to wear the Jawbone during waking hours, but could take it off for sleep and bathing. Participants were told they could access the pull intervention components (activity graphs, suggestion history) at will and were not required to do this with any specific frequency. As the study was intended to assess individual intervention components and not the complete intervention package, intervention delivery started immediately following the initial interview (i.e., without a baseline period). Adherence to the use of study equipment was monitored using automated scripts on the HeartSteps server, which sent participants automated emails if Jawbone data were not received for 24 hr or there was no interaction with the HeartSteps app for 48 hr. At the end of the study, participants returned the Jawbone tracker, brought back the study phone or had HeartSteps uninstalled from their phone, and were interviewed about their study experiences.

This study is registered on clinicaltrials.gov as NCT03225521. The full protocol can be obtained from authors upon request.

Randomization

The HeartSteps server micro-randomized the delivery of activity suggestions for each participant at each of the five decision points on each day of the study, conditional on participants' availability for treatment. At each available decision point, the randomization was independent of prior randomizations and the participants' responses to previously delivered activity suggestions. Given the study duration of 42 days, activity suggestions could be randomized up to 210 times for each participant. As we wanted participants to receive, on average, three activity suggestions each day with equal representation of walking and antisedentary messages, suggestion delivery at each decision point when participants were available was randomized with the following probabilities: 0.4 no suggestion, 0.3 walking suggestion, and 0.3 antisedentary suggestion.

Proximal Outcome

As activity suggestions were intended to help participants walk at the time they were delivered, we operationalized their intended proximal effect as the Jawbone-recorded step count in the 30 min following the randomization of suggestion delivery. As step count data are positively skewed, all step counts in our analyses were log-transformed. Consequently, to aid interpretability, for all analyses we report exponentiated regression coefficients.

Statistical Analyses

All analyses were performed using a centered and weighted least squares method [23]. This method estimates parameters in models for the treatment effects and allows robust (without incurring bias in the treatment effect model) inclusion of covariates to reduce noise. The method is similar to generalized estimating equation and multi-level models in that it accommodates the nested nature of the data (decision points nested within participants) and associated within-participant correlation across time in the outcome. Furthermore, the method takes advantage of the sequential randomization to estimate causal treatment effects. See the Supplementary Material for technical details. We prespecified two primary analyses [1]: overall (average, across all time points) effect of delivering a suggestion versus providing no suggestion on the subsequent 30-min step count; and [2] linear time trend of the suggestion effect, assessed by estimating the interaction between treatment effect and day in the study. For secondary analyses, we examined differences in the average and linear time trends in treatment effects separately for walking and antisedentary suggestions. To reduce noise, all analyses were adjusted for a pretreatment measure of activity, operationalized as the log of the 30-min Jawbone step count prior to the decision point. See Supplementary Material for full analysis details. Power analyses were conducted using the MRT sample size calculator [11, 24] for 80% power to detect a small effect size of 0.1, assuming a 5% type 1 error rate, and 70% availability. For these assumptions, the desired sample size was 32. As at the time of the study there was no information about the magnitudes of

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

standardized effect sizes in the MRT setting (see [25] for the subsequent work on this issue), to be conservative we powered the study using an effect size that is one-half of Cohen's small effect size of .2. Analyses were conducted using R version 3.3.2 [26].

As Jawbone Up Move records data only for minutes when an individual took steps, minute-to-minute Jawbone step counts are absent for periods both when the tracker was not worn and when the participant was sedentary. To be conservative in our analyses, we imputed zeros for all missing minutes of step count data in the outcome variable and the pretreatment covariate. For both variables, if the 30-min step count was zero, we added .5 before taking the log.

Two types of sensitivity analyses were conducted to assess robustness of findings: (i) analyzing the data from all participants who provided a minimum of 38 days of data (but might not have had full 42 days), and (ii) imputing missing Jawbone step data by using data from Google Fit, an Android application that uses the accelerometer in the phone to track steps. See Supplementary Material, Section V, for details.

Results

Participant Sample

We recruited 44 participants in the Ann Arbor, MI, metro area (see Table 2 for sample characteristics). Of these, we excluded all data from seven. Three participants were excluded because their phones were set to a non-English locale, resulting in corrupted data. Four participants dropped out early in the study (after 1, 4, 7, and 12 days) due to discomfort with using a study-provided Android phone (two participants) or unexpected need for extensive travel. Thus, 37 participants remained in the sample (see Fig. 2 for the CONSORT diagram).

Table 2 Baseline characteristics of HeartSteps participants

Variable (categorical)	*n*	%
Gender		
Female	31	70.5
Male	13	29.5
Race		
White	26	59.1
Asian	13	29.5
Black/African American	2	4.5
Other	3	6.8
Education		
Some college	8	18.2
College degree	13	29.5
Some graduate school or graduate degree	23	52.3
Married or in a domestic partnership	15	34.1
Have children	16	36.4
Used fitness app before HeartSteps	12	27.3
Used activity tracker before HeartSteps	10	22.7
Used personal phone	21	47.7
Variable (continuous)	Mean	*SD*
Age	35.9	14.7

Included Decision Points

The 37 participants generated 8,274 person-decision points over the course of the study. From these, we excluded 390 decision points that occurred while participants were traveling abroad without Internet connectivity, resulting in corrupted data, or while participants had technical issues (e.g., lost Jawbone or phone) that they reported to the study team; 340 decision points that occurred beyond the planned study end of 42 days; and four anomalous data points due to system bugs. In cases of extended (3+ days) travel or technical issues, the study team attempted to extend the study to allow a participant to complete 42 days of data collection; however, for reasons detailed in the Supplementary Material, not all participants could stay in the study for full 42 days. All 37 participants, however, had data for at least 36 study days—days when we could ascertain availability and randomize (if available)—resulting in 7,540 included decision points.

Available Decision Points

Of the 7,540 included decision points, participants were unavailable for treatment during 1,479 decision points (19.6% of the time). Thus, 6,061 available decision points were used in all analyses. Availability remained consistent, averaging around 80%, throughout the study. Across the 6,061 available decision points, HeartSteps delivered 3,594 activity suggestions (59.3%). Of these, 2,015 (33.2%) were walking suggestions and 1,579 (26%) were antisedentary suggestions. On average, participants received 2.4 suggestions each day (SD = 1.1).

Data Missingness

Missing data occurred when the Jawbone did not record step counts. Recall that Jawbone records data only for minutes when steps occurred. For 1,434 out of 6,061 available decision points (23.66%), there was no record of steps in the 30-min following the randomization. Of these, for 882 decision points (61.51%), step data exist shortly after the 30-min window, indicating sedentary behavior during the 30 min following randomization. We consider the remaining 552 missing data points (38.49%) as true missingness, some of which may be due to extended sedentary periods. As noted, a step count of 0 was imputed for all missing minutes of step count data.

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

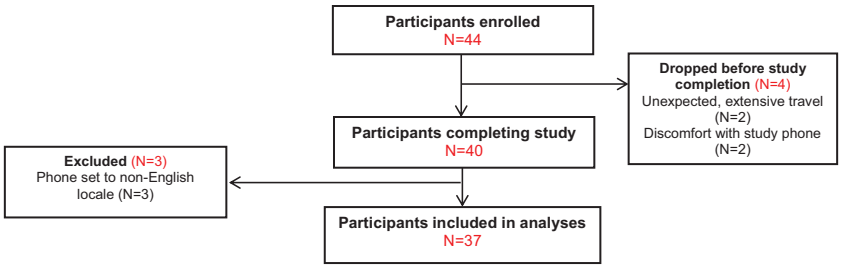

Fig. 2. CONSORT diagram.

Effects of Activity Suggestions

The average step count in the 30 min following a decision point across all 6,061 available decision points (regardless of the randomization outcome) was 253 steps ($SD = 473$). Averaging over study days and types of activity suggestions, delivering a contextually tailored activity suggestion versus providing no suggestion increased step count by 14% ($e^b = 1.14$, $p = .06$, 95% CI = .99 to 1.31), or an additional 35 steps, on average. This effect was not evenly distributed over time, however. When we estimated a linear treatment effect model with an interaction between treatment indicator and study day, we found a strong effect at the start of the study that decayed as the study progressed. Providing a suggestion versus providing no suggestion initially increased the 30-min postrandomization step count by 66%, adding 167 steps to the 253-step average on the first day of the study ($e^b = 1.66$, $p < .01$, 95% CI = 1.22 to 2.26). This effect diminished linearly over time at 2% per day ($e^b = .98$, $p < .01$, 95% CI = –.03 to –.01) and was no longer distinguishable from zero by the 28th day in the study.

Effects of Walking Versus Antisedentary Suggestions

As HeartSteps provided both walking and antisedentary suggestions, we conducted secondary analyses to examine the effects of the two types of suggestions and their time trends separately. Averaging over study days, delivering a walking suggestion versus providing no suggestion increased the average 30-min step count by 24%, a 59-step increase ($e^b = 1.24$, $p = .02$, 95% CI = 1.04 to 1.47). The effect for the antisedentary suggestions, while still positive, was smaller and not significant. These analyses indicate that the overall suggestion effects on the 30-min step count were mainly due to the walking suggestions. Time trend results reinforced this conclusion. Delivering a walking suggestion versus providing no suggestion initially more than doubled the average 30-min step count, adding 271 steps to the average count of 253 steps on the first day of the study ($e^b = 2.07$, $p < .001$, 95% CI = 1.52 to 2.82). This initial increase diminished linearly with time at 3% a day ($e^b = .97$, $p < .001$, 95% CI = –.04 to –.01) and could no longer be detected by the 29th day in the study. Neither the initial effect nor the time trend for antisedentary suggestions was significant.

Sensitivity Analyses

The results from sensitivity analyses were consistent with the above-reported results. The average across-time effect of walking suggestions retains its magnitude and statistical significance ($p < .05$) when we imputed the 552 missing data points using Google Fit, as well as when we restricted the sample to the 35 participants with at least 38 days of data. All time-trend effects remained significant and of similar magnitude across all sensitivity analyses.

Discussion

The main purpose of this MRT was to screen HeartSteps' intervention components for preliminary efficacy to decide whether they should be retained in an optimized version of the intervention. Our results indicate that contextually tailored walking suggestions are a feasible and promising intervention component for initiating short bouts of walking throughout the day. We found that averaging over the duration of the study, each suggestion results in additional 35 steps, or about 100 additional steps per day for three suggestions. Early in the study, however, this effect was larger, adding about 500 additional steps a day for three suggestions. We also found that this effect was entirely driven by walking suggestions, which, when assessed by themselves, performed better. That is, averaging over the study duration, three walking suggestions could add about 180 steps per day, and early in the study three walking suggestions could add about 800 steps per day. Given that recent meta-analyses found that physical activity interventions (both digital and in-person) have rather modest results—interventions for healthy adults on average only add 496 steps per day [27] and interventions for older adults add on average 620 steps per day [28]—our results indicate that contextually tailored activity suggestions might meaningfully enhance the effectiveness of an mHealth physical activity intervention. While they are clearly insufficient to achieve physical activity guidelines by themselves, as a component of a complex intervention, they could provide actionable, lightweight,

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

low-burden support that continuously adapts to the individuals' changing circumstances.

Contextual tailoring appears to have played a role in the efficacy of the suggestions. In the interviews, participants commented on the situational appropriateness of the suggestions and noted that the suggestions' being actionable and specific made it easier for them to engage in suggested behaviors. At the same time, our data indicate that in our initial implementation the potency of activity suggestions diminished over time and largely disappeared after 1 month. The effect attenuation is most likely due to habituation. In exit interviews, participants explicitly told us that the suggestions became boring after 2–4 weeks. This finding suggests that using contextually tailored walking suggestions in a long-term intervention would require a delivery schedule that minimizes the risk of habituation. The next version of HeartSteps will attempt to accomplish this by incorporating reinforcement learning algorithms [29] that will personalize suggestion delivery to each participant based on his or her prior responses. This approach will allow HeartSteps to learn in which contexts a particular participant is likely to respond to a suggestion, enabling the system to adapt suggestion delivery to favor contexts in which they are effective while avoiding burdening participants with suggestions at times when participants are unlikely to be receptive to them.

Our results also pose an interesting question about antisedentary suggestions. As we mentioned, we originally included these suggestions as an engagement strategy, as they were easy to follow and we could make them interesting by suggesting activities other than walking—activities such as stretching, jumping jacks, etc. Our intuition that participants would find these suggestions engaging was correct. Almost unanimously, participants told us in exit interviews that they found the antisedentary suggestions to be fun and interesting—unlike the walking suggestions, which were perceived to be more repetitive. Yet, walking suggestions increased immediate step counts while antisedentary suggestions did not. One explanation for this is that the 30-min postdecision step count was not a sensitive enough proximal outcome to detect the low-level activities suggested by antisedentary messages. Given the importance of sedentary behavior as a distinct health behavior target [30] and the high user acceptance of antisedentary messages, we are exploring sensors for use in future studies that can both accurately assess breaks in sedentary behavior and are sufficiently wearable to be deployed for several months.

It is important to note that our study was intended—as MRTs are more generally—to assess proximal effects of individual intervention components to inform intervention optimization and not the efficacy of the HeartSteps system as a whole. Thus, although our results showed that contextually tailored walking suggestions helped participants engage in short bouts of walking throughout the day, they do not tell us whether the use of HeartSteps led to an overall increase in physical activity. As is the case with other forms of optimization trials—e.g., factorial experiments used as part of the multiphase optimization strategy [8, 31, 32]—a central aim of MRTs is to collect data that can be used to optimize interventions, which can then be evaluated for efficacy as an intervention package in subsequent randomized controlled trials. Our results demonstrated feasibility and promise of contextually tailored walking suggestions, but they also indicated that the delivery schedule for the suggestions needs to be modified to maintain their efficacy over time. Consequently, we are including walking suggestions as a component of the next version of HeartSteps, while employing reinforcement learning to more intelligently manage their delivery over time.

Finally, although MRTs are primarily aimed at intervention optimization, the scientific value of MRT results such as those we presented here can go beyond their role in the optimization of a specific intervention package. Specifically, regardless of the ultimate efficacy of a particular intervention package, MRT results provide useful evidence about the tested intervention components that can inform decisions about whether and how those components are incorporated into future interventions. For instance, the results presented here provide initial evidence about the effect size for contextually tailored activity suggestions, how their effect changes over time when they are delivered two to three times a day, as well as how the effects differ for the two classes of suggestion content—suggestions to go for a short walk versus to just briefly move around. Other researchers can use this evidence to decide whether contextually tailored activity suggestions are worth incorporating into their own interventions and, if so, how they might want to design their timing and frequency to maximize their effectiveness and reduce the risk of habituation. As MRTs are conducted on contextually tailored activity suggestions included in other interventions—both our own and those of other researchers—the field will begin to accumulate a robust evidence base about how well contextually tailored suggestions work, for whom they do and do not work, when they are most effective, and what their optimal delivery schedules are. Notably, this evidence base can emerge regardless of how effective any individual intervention package that includes contextually tailored suggestions ends up being.

Limitations

One limitation of our study is that due to the need for Internet connectivity to randomize treatment, we had to exclude a small number of decision points when participants were traveling to areas with poor coverage. We are engineering the next version of HeartSteps to not have this limitation. Another limitation of the study is that the Jawbone Up Move activity tracker that we used in the

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

trial is no longer available, creating potential problems for replication of our results. However, our approach could be replicated with any activity tracker that tracks minute-level step data and that makes the data available to third-party applications via a public application programming interface. Fitbit, Garmin, and a number of other companies make products that satisfy these requirements. In fact, we ourselves used Google Fit as a secondary source of step data that we used in some of our sensitivity analyses. Replications could be similarly done with other activity trackers, as long as those devices are usable enough so participants are willing to wear them regularly. Finally, a question might arise as to whether the activity suggestions we tested are actually adding physical activity, or whether individuals might just be moving their physical activity around and walking less at other times. While this study could not examine this issue, this is primarily a question about the efficacy of the overall intervention package. Our results indicate that contextually tailored activity suggestions did what they were intended to do: They helped individuals initiate bouts of walking throughout the day. As with any other intervention component that might result in compensatory behavior, it is the task of the overall intervention package to try to manage and minimize such responses. To minimize the risk of reduced activity at other times, in the next version of HeartSteps we are incorporating intervention components that target participants' intrinsic motivation for being active, including by supporting individuals in finding activities that they enjoy or modifying current activities (e.g., walking) to make them more enjoyable (e.g., by listening to podcasts while walking). We hope that this emphasis on intrinsic motivation will reduce the potential for negative compensatory behavior.

Conclusions

Our results show that contextually tailored activity suggestions may be worthy of inclusion in a physical activity intervention. Additional analyses will help us optimize the JITAI decision rules for suggestion delivery, so that they are sent at times and in contexts that maximize their effectiveness and minimize habituation and user burden.

Supplementary Material

Supplementary material is available at *Annals of Behavioral Medicine* online.

Acknowledgments We thank our students for their extensive testing of the HeartSteps system, without which this study would not have been possible. This work was generously supported by the following grants from the National Institutes of Health: R01HL12544001 (PI: P. Klasnja), R01AA023187 (PI: S.A. Murphy), P50DA039838 (Site PI: S.A. Murphy), and U54EB020404 (Site PI: S.A. Murphy). Study sponsors had no role in the study design, collection, analysis, and interpretation of data, or writing of the report. The research presented here is solely that of the authors and does not reflect views of the NIH.

Compliance With Ethical Standards

Authors' Statement of Conflict of Interest and Adherence to Ethical Standards The authors declare that they have no conflicts of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The reported study was approved by the University of Michigan Institutional Review Board. All participants completed a written informed consent to take part in the study.

References

1. Gustafson DH, McTavish FM, Chih MY, et al. A smartphone application to support recovery from alcoholism: A randomized clinical trial. *JAMA Psychiatry*. 2014;71:566–572.
2. Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: Towards a gold standard for continuous stress assessment in the mobile environment. *Proc ACM Int Conf Ubiquitous Comput*. 2015;2015:493–504.
3. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann Behav Med*. 2018;52:446–462.
4. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Comput*. 2014;13:13–17.
5. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychol*. 2015;34S:1209–1219.
6. Berlyne DE. Novelty, complexity, and hedonic value. *Atten Percept Psychophys*. 1970;8:279–286.
7. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. *Ann Behav Med*. 2005;30:65–73.
8. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. *Am J Prev Med*. 2007;32:S112–S118.
9. Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. Evaluating individual intervention components: Making decisions based on the results of a factorial screening experiment. *Transl Behav Med*. 2014;4:238–251.
10. Klasnja P, Hekler EB, Shiffman S, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychol*. 2015;34S:1220–1228.
11. Liao P, Klasnja P, Tewari A, Murphy SA. Sample size calculations for micro-randomized trials in mHealth. *Stat Med*. 2016;35:1944–1971.
12. Riddle M; Science of Behavior Change Working Group. News from the NIH: Using an experimental medicine

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020

approach to facilitate translational research. *Transl Behav Med*. 2015;5:486–488.
13. Nielsen L, Riddle M, King JW, et al.; NIH Science of Behavior Change Implementation Team. The NIH science of behavior change program: Transforming the science through a focus on mechanisms of change. *Behav Res Ther*. 2018;101:3–11.
14. Lee IM, Buchner DM. The importance of walking to public health. *Med Sci Sports Exerc*. 2008;40:S512–S518.
15. Böcker L, Dijst M, Prillwitz J. Impact of everyday weather on individual daily travel behaviours in perspective: A literature review. *Transport Reviews*. 2013;33:71–91.
16. Owen N, Humpel N, Leslie E, Bauman A, Sallis JF. Understanding environmental influences on walking; Review and research agenda. *Am J Prev Med*. 2004;27:67–76.
17. Saelens BE, Handy SL. Built environment correlates of walking: A review. *Med Sci Sports Exerc*. 2008;40(7 suppl):550–566.
18. Gollwitzer PM, Fujita K, Oettingen G. Planning and the implementation of goals. In: Vohs KD, Baumeister RF, eds. *Handbook of Self-Regulation: Research, Theory, and Applications*. New York, NY: The Guilford Press; 2004:211–228.
19. Gollwitzer PM. Implementation intentions: Strong effects of simple plans. *Am Psychol*. 1999;54:493–503.
20. Becker MH. *The Health Belief Model and Personal Health Behavior*. San Francisco, CL: Society for Public Health Education; 1974.
21. Maiman LA, Becker MH. The health belief model: Origins and correlates in psychological theory. *Health Educ Behav*. 1974;2:336–353.
22. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: Implications for JITAI optimization. *Health Psychol*. 2015;34S:1261–1267.
23. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. *J Am Stat Assoc*. 2017. https://amstat.tandfonline.com/action/showCitFormats?doi=10.1080%2F01621459.2017.1305274
24. A sample size calculator for micro-randomized trials. https://link.springer.com/article/10.1007/s11121-017-0862-5. Accessed August 5, 2018.
25. Luers B, Klasnja P, Murphy S. Standardized effect sizes for preventive mobile health interventions in micro-randomized trials. *Prev Sci*. 2018.
26. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*; 2017. http://www.R-project.org/.
27. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: Meta-analysis of outcomes. *Am J Public Health*. 2011;101:751–758.
28. Chase JA. Interventions to increase physical activity among older adults: A meta-analysis. *Gerontologist*. 2015;55:706–718.
29. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press; 1998.
30. Owen N, Salmon J, Trost S, et al. Sedentary behaviour and health: Strengthening the evidence base. *J Sci Med Sport*. 2014;18:e131.
31. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: Efficient tools for evaluation of intervention components. *Am J Prev Med*. 2014;47:498–504.
32. Chakraborty B, Collins LM, Strecher VJ, Murphy SA. Developing multicomponent interventions using fractional factorial designs. *Stat Med*. 2009;28:2687–2708.

Downloaded from https://academic.oup.com/abm/article-abstract/53/6/573/5091257 by Harvard Library user on 13 February 2020