Optimizing Digital Integrated Care via Micro-Randomized Trials

Ashley Walton 1[1] ; Billie Nahum-Shani 2[2]; Lori Crosby 3 [3,4], Predrag Klasnja 4[5] Susan Murphy 5[1]

[1] Harvard University, Department of Statistics, Boston, MA.
[2] University of Michigan, Institute for Social Research, Ann Arbor, MI.
[3] Cincinnati Children's Hospital Medical Center, Cincinnati, OH.
[4] University of Cincinnati, Department of Psychology, Cincinnati, OH.
[5] University of Michigan, School of Information, Ann Arbor, MI.

Corresponding author: Ashley Walton, ashley_walton@fas.harvard.edu

Science Ctr. #706

1 Oxford St., Cambridge, MA 02138

Tel: (513) 406-7584

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

ABSTRACT

Mobile Health (mHealth) interventions are a promising tool in providing digitally mediated integrative care. They can extend care outside of the clinic by providing reminders to take medications, assisting in managing symptoms, and supporting healthy behaviors including physical activity, healthy eating and stress management. mHealth interventions can adapt the delivery of care across time in order to optimize treatment effectiveness. Yet there exists limited empirical evidence useful to the development of adaptive mHealth interventions. This paper describes a new randomized trial design, the Micro-Randomized Trial (MRT), for informing the development of mHealth interventions. We provide examples of scientific questions important to the development of a mHealth intervention, and describe how these questions can be answered using an MRT.

INTRODUCTION

There is an increasing interest in the use of mobile devices to provide digitally mediated integrated healthcare. This is because many digital technologies such as smartphones and wearable trackers are deeply integrated into everyday life and thus might be a low-barrier way to improve medical care. For example, digital technologies could help address adherence to prescribed medications, as well as provide support for self-care behaviors such as engaging in sufficient physical activity. Though not considered as a replacement for face-to-face interactions, mobile devices are thought to have the potential to extend the reach of healthcare, making treatments more accessible and scalable[1,2]. Mobile health (mHealth) interventions have been developed, tested and evaluated in various health domains, including weight management[3], smoking cessation[4], and mental health[5]. It has also become possible to seek approval for mHealth interventions from the U.S. Food and Drug

Administration (FDA), where recently the FDA approved a prescription-only "digital therapeutic" for substance abuse[6].

In mHealth, treatments, which are typically in the form of reminders, messages, and supportive content, can either be "pushed" or "pulled." Pull treatments are initiated by the individual; for example in FOCUS[7], a mHealth intervention that provides illness management support to individuals with schizophrenia, patients can access a collection of self-management strategies (e.g., focusing on medication adherence, sleep, or mood-regulation) on their mobile device, at any time they choose. Pull treatments require that individuals are able to recognize when they need support, and are motivated to seek out that support. Alternatively, push treatments are initiated by the mHealth intervention. Treatment is intended to be delivered to the individual at moments when it is best poised to be effective; when treatment is most likely to prevent unwanted behaviors or support the achievement of desired health behaviors. Adapting the timing and content of the push treatments in order to optimize effectiveness is of great importance, because push treatments interrupt the individual, and can potentially lead to burden[8-10] and disengagement[11-13]. There is limited empirical evidence available to inform the development of these kind of intervention components[14-17]. Thus, it is particularly important to assess when to deliver a push treatment.

One method for understanding when both push and pull treatments are effective, is to use a newer type of trial design, the *Micro-Randomized Trial (MRT)*. MRTs are currently in use to inform the development of mHealth interventions including interventions that target weight management, physical activity, smoking cessation[18] and patient engagement[19] (see the MRT website for more details)[20]. This paper describes how MRTs can be used to assess the effectiveness of treatments in mHealth interventions; here we focus on push treatments. We discuss the scientific questions that data from a MRT is able to address in reference to MRTs currently in the field. We then provide an overview of the MRT and define key terms, and

then we walk through an example how data is collected and analyzed to address a scientific question regarding the delivery of a mHealth treatment from a current MRT. In conclusion, we compare the MRT to N-of-1 and single case experimental designs, as well as randomized clinical trials.

CRITICAL QUESTIONS IN THE DEVELOPMENT OF AN MHEALTH INTERVENTION

There are many different scientific questions that might need to be addressed when developing an mHealth intervention. Discussed below is a small selection of possible research questions, where MRTs may focus on one or many of the questions below (see Klasjna et al., 2015 for more examples)[21].

First, a researcher needs to know if it is worthwhile to deliver a treatment at all. For example, suppose a researcher wants to know if sending a text message to a patient's phone to remind them to take their medication increases medication adherence on that day. Daily medication adherence could be measured using an electronic pill bottle that records when the bottle is opened. Whether or how often the pill bottle was opened after sending a reminder, compared to whether or how often the pill bottle was opened when a reminder was not sent, will provide information about the effect the reminder has on adherence. Consider another example from the MRT of the application HeartSteps[21-24], a mHealth intervention designed to increase physical activity in sedentary adults. In this intervention the smartphone may push activity suggestions tailored to the time of day, the individuals' current location, and current weather conditions. This tailoring is designed to make the activity suggestions actionable in the individual's current context. A first question to address is whether pushing the tailored activity suggestions results in an increase in participants' step count in the thirty minutes following the delivery of suggestion, compared to not pushing a suggestion. Addressing these

types of questions provides a first indication of whether or not the push treatment should be included in the mHealth intervention.

There are often multiple different kinds of treatments that could be delivered as part of a mHealth intervention, and a researcher might want to know which of them is more effective. Consider a mHealth intervention to target adherence to a preventative medication, where the lack of short-term benefits may lead to negative beliefs about the medication's efficacy. In behavioral research there is evidence that making the long-term benefits of a decision more salient in the moment can prevent individuals from over-valuing the short-term impacts of treatment[25-27]. Thus, a researcher might want to compare the effect on daily adherence of a generic reminder message with a reminder message that includes content referencing long-term benefits of taking the medication. Alternately, consider a mobile intervention designed to optimize the delivery of prompts (via push notifications) to improve self-monitoring of dietary intake among obese adults. Here an important question would be whether an evening notification that prompts participants to plan for how they will reach their calorie goal the following day, has a different effect if this evening planning demands a high level of effort (high burden) or a low level of effort (low burden) on the part of the participant. To address this question, success achieving calorie goals when participants are prompted to engage in "structured," low burden, evening planning (e.g. choosing a plan from a drop-down menu) can be compared to success achieving calorie goals when participants are prompted to engage in "unstructured," higher burden evening planning (e.g. writing out their plan in a provided text box). Here addressing these types of questions can be used to inform decisions about what type of push treatments to include as part of a mHealth intervention.

The significance of mHealth interventions, as previously discussed, is partially attributed to the ability to integrate them into individuals' everyday lives. As a result, researchers may need to understand how the success of treatments depends upon

characteristics of the contexts in which they are delivered. The Sense2Stop MRT is investigating whether the effect of a stress-regulation prompt to support smoking cessation is more effective when the individual is detected as being stressed, compared to when they are not stressed[18]. Furthermore, suppose a researcher is developing a mHealth intervention for individuals with Sickle Cell Disease (SCD) to help them remember to drink water. Hydration has been shown to be a critical determinant of pain experience and health-related outcomes in SCD[28-30]. Evidence suggests that there are certain conditions under which staying hydrated is of particular importance, such as change in temperature[31] or recent levels of physical activity[32]. Therefore, a researcher might ask how these factors influence the effectiveness of a hydration reminder to increase the number of ounces of water consumed by the patient in the following hour. The contextual factors that relate to the effectiveness of treatment might also include individual's responses to past treatments. In the example of hydration reminders for patients with SCD, a researcher might ask if the effect of delivering a reminder varies by the ounces of water patients consumed in response to previous reminders. Addressing these types of questions related to under what conditions it is most useful to provide treatments can also be used to develop a mHealth intervention.

mHealth interventions can be designed to provide long-term treatment over the course of weeks, months or even years. Therefore, it is important to understand how the effects of different treatments will change as the mHealth intervention progresses. For example, a researcher might ask if the effect of a daily medication reminder on adherence dissipates with time due to habituation to the messages[33,34]. It may also be important to understand the best way to sequence the delivery of different treatments. Consider again the evening planning prompts aimed at increasing self-monitoring of dietary intake, where individuals plan how to meet their calorie goal for the next day. There may be scientific questions concerning how the effectiveness of the evening planning prompt might be impacted by the delivery of other

treatments that are provided as part of the same mHealth intervention, such as a mid-day prompt to reminding participants to monitor their dietary intake. Researchers might want to know whether the unstructured (higher-burden) evening planning is more effective (compared to the lower-burden structured planning) if the mid-day reminder was framed in a way that emphasized the individual's core values (i.e. "value-based"). This is a scientific question concerning the optimal sequencing of these two treatments, specifically if value-based reminders build the foundation for the effectiveness of unstructured planning by increasing motivation to invest effort in achieving weight loss goals. Or alternatively, there might be important questions about mHealth interventions that involve sequences of treatments focusing on different aspects of a disorder. Consider the development of a mHealth intervention for smoking cessation that incorporates mindfulness exercises as well as financial incentives to support an individual in abstaining from smoking. It may be that financial incentives are most useful early in the quit attempt whereas prompts to remind individuals to practice mindfulness exercises might be most useful later in recovery. In this case, data to address which treatment should be delivered and when, would be useful in designing a mHealth intervention that is intended to be used over a longer time period. Addressing these kinds of questions can inform when during the course of a mHealth intervention delivering certain treatments will be most effective.

In summary, answering questions regarding what kind of treatment, when, and for whom it is most beneficial, is critical to the design of mHealth interventions. This is particularly important if the mHealth intervention is to include push components as these can be burdensome, leading to disengagement[8-13]. We now describe how the MRT design and how it can provide data to answer such questions.

THE MICRO-RANDOMIZED TRIAL

MRTs can be used to inform the development of a mHealth intervention in which *treatment components* (e.g., treatment pushes such as activity suggestions), may be delivered multiple times during the day or week, and are designed to impact a long-term outcome by impacting a near-term *proximal outcome.* For example, as discussed earlier, HeartSteps is a mHealth intervention in which activity suggestions may be delivered 5 times per day, aiming to help sedentary adults achieve and maintain recommended levels of moderate-intensity physical activity (distal outcome) by increasing an individual's physical activity in the thirty minutes following the delivery of an activity suggestion (proximal outcome)[21-24]. A MRT can be used to inform the construction of an effective mHealth intervention by providing data to address questions regarding the assessment of whether the treatments will impact the proximal outcome(s), how treatment effects vary over time, and how to effectively sequence treatments.

The scientific questions outlined above can be addressed with a MRT through the *randomization* of individuals to *treatment options* each time a treatment component might be delivered. This *micro*-randomization may result in 100's or even 1000's of randomizations per individual during a study. For example, in the HeartSteps MRT, participants were randomized to three treatment options five times per day: an anti-sedentary activity suggestion, a walking activity suggestion or no suggestion. The activity suggestions were designed to have a momentary effect on step count, thus the proximal outcome is total step count in the thirty minutes following delivery of the suggestion. In this example, the set of different treatment options includes not delivering treatment (i.e., not delivering an activity suggestion to reduce burden). Alternatively, the set of treatment options within a component can include only multiple active treatment options, where for example every evening an individual is randomized to always receive one of two types of evening planning prompts

(structured vs. unstructured). The randomization provides data that can be used to estimate the causal effect of treatment components like the activity suggestions because it ensures that known and unknown factors that may influence a participant's response to a suggestion are distributed evenly between the treatment options.

In a MRT, participants are randomized between treatment options at each *decision point*, which are the moments in time at which a treatment component might be delivered. How often participants receive a particular treatment at each decision point is determined by the *randomization probabilities* defined for each treatment option. The timing of decision points, as well as the probability of a participant receiving each of the treatment options at each decision point, are motivated by scientific and practical considerations concerning when and how often providing treatment is most likely to be effective. Consider the timing of decision points for the activity suggestions in the HeartSteps MRT, described above. Prior studies indicated that sedentary individuals with standard, weekday employment tend to have the highest within-person variance in step count at five times: pre-morning commute, mid-day, mid-afternoon, evening commute, and after dinner[21]. This higher variance indicates that the individuals might be more responsive to an activity suggestion during these times. Thus, these five times each day were used as the decision points in HeartSteps for the activity suggestion component. In order to reduce burden and prevent disengagement with the intervention, researchers wanted participants to only receive an activity suggestion at approximately two of the five decision points each day. Researchers also wanted participants to have an equal chance of receiving the walking activity suggestion and the anti-sedentary activity suggestion. Therefore, at each of the five decision points, participants had a .3 probability of receiving an anti-sedentary activity suggestion, a .3 probability of receiving a walking activity suggestion, and a .4 probability of receiving no suggestion. In other cases, an MRT might be designed to have more decision points where there is a lower probability of

being assigned to treatment, or less decision points with a higher probability of participants being assigned to treatment. These choices are always driven by scientific evidence related to when and how often a treatment is likely to have the most significant impact on the intended proximal outcome.

A more complex MRT might provide data concerning multiple treatment components. There can be different decision points for each treatment component. For example, the BariFit MRT (see the fourth diagram on the MRT website)[20], includes an "activity goal schedule" component in addition to the "activity suggestion" component. As in HeartSteps, the "activity suggestion" component in BariFit includes three treatment options (delivering one of two types of tailored activity suggestions, and not delivering an activity suggestion), which are assigned at five different decision points each day. However, the "activity goal schedule" treatment component includes two options: daily activity goal with no rest days, vs. daily activity goals including rest days. The treatment options for this component are assigned to individuals at the beginning of the study; that is, there is only one decision point for the activity goal schedule component, where each individual is randomized to receive either type of goal schedule for the duration of the study. An MRT can be designed to estimate the effectiveness of multiple different intervention components, where randomization and delivery occurs at different points in the study.

In addition to addressing scientific questions concerning what treatment options are generally worthwhile, randomization also assists in ascertaining if certain conditions influence the effectiveness of treatment components. For example, MRT data can be used to investigate if an activity suggestion delivered as part of the HeartSteps intervention is more successful at increasing physical activity and is perceived as less burdensome under certain circumstances, such as when an individual is less busy (e.g., based on his/her entries in their smartphone

calendar) than otherwise. Therefore, MRTs assist researchers in deciding when it is most worthwhile to interrupt the individual to provide treatment.

The randomization in MRTs can be used to investigate whether and how the effect of treatments may vary over time. For example, data from a MRT can be used to address questions such as whether the effect of delivering an activity suggestion dissipates with time. Data from a MRT can also inform the sequencing of treatment components. Consider again our example of an MRT that aims to increase self-monitoring of dietary intake among obese adults, where participants are randomized every evening to the two types of planning prompts: structured (i.e., low burden evening planning) vs. unstructured (higher burden evening planning). This MRT could also include a mid-day randomization to reminders that encourage self-monitoring of dietary intake, where every day the individual is randomized to either (1) a generic reminder; (2) a value-based reminder; or (3) no reminder. This data can be used to investigate whether the unstructured (higher-burden) evening planning is more effective (compared to the lower-burden structured planning) if a value-based reminder was delivered at mid-day on the same day. This scientific question concerns the sequence of treatment options, and is motivated by the conjecture that mid-day value-based reminders may build the foundation for the effectiveness of the unstructured evening planning prompt by increasing motivation to invest effort in achieving weight loss goals.

We have described how an MRT can be used to address questions to inform the development of mHealth interventions, questions including what treatments are effective, and how their effectiveness may vary with respect to context of delivery, time, and sequencing of treatment. The scientific questions in the current MRTs we have referenced are addressed using a generalization of the common generalized estimating equation (GEE) analysis method to test for causal effects of treatment components, and moderation effects of time-varying factors[35-37]. For researchers interested in running an MRT there are open source tools

available for power planning and sample size calculations[38-40]. Additionally, we walk through an example below of how the data generated by an MRT can be used to answer scientific questions about the delivery of mHealth treatments.

EVALUATING THE EFFECTIVENESS OF DAILY ACTIVITY PLANNING IN HEARTSTEPS

In order to demonstrate how an MRT can be used to generate data for the optimization of mobile health interventions, we use an example from the HeartSteps MRT. The goal of HeartSteps was to develop an mHealth intervention to increase physical activity levels in sedentary adults. Accordingly, the MRT is designed to answer scientific questions with respect to two intervention components: activity suggestions, and daily activity planning, including whether or not these interventions are effective, in what contexts delivering these interventions is beneficial, and how the effectiveness of these interventions might change over time. For simplicity, we will focus our example on the daily activity planning. First, we will describe the daily planning intervention component delivered as part of the HeartSteps MRT, how it was implemented, and the data the MRT generated. Then, we describe how this data would be analyzed in order to estimate the treatment effect of the daily activity planning intervention component, as well as moderation effects.

As discussed above, an MRT provides data to inform the development of mobile health interventions by randomizing participants to different treatment options at multiple decisions points for the duration of a study, usually across a period of weeks or months. In the HeartSteps MRT, the daily activity planning was an intervention component where participants were randomized to receive one of the two treatment options at a daily decision point every evening: (1) a prompt to make a plan for how they were going to be active the next day or (2) no prompt. Participants had a 50% chance of receiving either of these

treatment options at every daily decision point. This intervention component was motivated by behavioral science suggesting that specifying intentions for how to implement behavioral goals increases the ease of engaging in that behavior[42]. Thus, the proximal outcome for the daily planning intervention component was the total number of steps taken the day following being randomized to receive one of these two treatment options. This proximal outcome was measured with a wristband activity tracker that participants wore for the duration of the study that monitored their steps throughout the day. In the HeartSteps MRT, 37 individuals were randomized with a .5 probability to either receive daily activity planning or no activity planning, once a day, for 42 days. Therefore, the data generated by the HeartSteps MRT with respect to the daily activity planning comprised of daily step counts for 37 individuals across the 42 day study, which can be used to estimate the effect of receiving a daily activity planning prompt on step counts for the following day.

Data from a MRT can be analyzed with a generalization of regression specifically developed to ensure unbiased estimates of causal effects of time-varying treatments. Proximal effects are defined in terms of the potential outcomes, and estimated using a GEE approach with an independent correlation matrix (see Boruvka et al. 2016 for more details)[35]. These analyses pool time-varying, longitudinal data across all study participants. Now consider the daily activity planning in Heartsteps. Here, the MRT is designed to investigate whether there is an effect of providing a daily activity planning prompt vs. no prompt on next day step count. Let t_1 denote time points at which an individual may or may not receive a daily planning prompt; t_1 ranges from 1 to 42 (42 days*1 decision point per day). The focus is on the marginal effect of delivering an activity planning prompt (vs. no prompt) at t_1 on step count at the next day t_{+1} (proximal effect). To test whether there is an average activity planning prompt effect on next-day step count at any point during the study, we can use a model that includes an intercept and an indicator variable for activity planning prompt (=1)

vs. no prompt (=-1). In order to investigate potential moderators of this treatment effect (e.g., day of the study at t_1, weather at t_1) the model above can be extended to include a covariate-by-prompt interaction term. Estimating the proximal effect of the daily activity planning prompt on participants' step count the next day provides information about whether or not it should be included as part of an mHealth intervention to increase physical activity in sedentary adults. Investigating the moderation effects of factors such as day of study and weather conditions can provide guidance for when and in what contexts delivering the daily activity planning prompt would be most effective.

DISCUSSION

We discussed the use of an MRT in developing mHealth interventions to provide digitally integrated care. We discussed scientific questions relevant to the delivery of this care, and described how the data generated by a MRT can be used to answer these kinds of questions in order to optimize the effectiveness of treatments delivered as part of mHealth interventions. We conclude with a discussion of how the MRT compares to other types of trial designs.

MRTs may appear very similar to N-of-1 studies or Single Case Experimental Designs (SCED) used in the behavioral sciences[42]. In both of these latter designs a participant undergoes a series of intervention episodes; in each episode a different intervention package may be (randomly) assigned. These designs are very similar, but each are appropriate for different scientific goals. Traditionally N-of-1 studies are used to estimate individual-level causal effects of different interventions[43], and are best suited to assess intervention effects in settings in which individuals can serve as their own control, and when it is too expensive or too difficult to recruit a large number of participants[44-46]. In these cases, individuals can serve as their own control if there are no carry-over effects of treatments and the "wash out period"

between treatment episodes is sufficiently long as to allow the effect of the prior treatment to dissipate. Here the scientific goal is the same as a classical randomized trial: "Is intervention package (A) is better than another intervention package (B)".

In comparison, the MRT is designed to provide data to address dynamic questions about sequences of individual treatment components that are part of an intervention package. For example, a researcher would use an MRT if they are interested in the carry-over effects, specifically how delivering one treatment directly before another type of treatment may change the effectiveness of the intervention. MRTs are also most appropriate in settings in which time under treatment may be a significant factor in understanding the effect of subsequent treatment, and where developing a mHealth intervention requires understanding how treatment may need to be adapted across time in response to changes to incoming information about the individual's current state. It is important to note that MRTs and N-of-1 studies can appear similar, particularly in cases where multiple N-of-1 studies are combined[47-49]. We expect that in the future there will be increasing convergence between MRTs and N-of-1 studies, particularly in the case of evaluating technology-based interventions (see Dallery et al., 2013 for a review)[50].

A researcher would not choose an MRT design if they wanted to assess an intervention package as a whole. The MRT cannot replace the standard randomized clinical trial in which a mHealth intervention might be contrasted with standard care or a suitable control in terms of the distal outcome. Rather the MRT is akin to a factorial design used to optimize and build multi-component interventions, where they are best suited to address questions regarding the effectiveness of the individual treatment components that make up a mHealth intervention[51]. This allows for the discarding of less effective, more burdensome components, as well as an understanding of when and in what contexts delivering different treatments will have the greatest impact. After an MRT provides guidance for how to optimize the components in an

intervention package, an evaluation on the basis of the distal outcome would occur in a future

randomized clinical trial.

REFERENCES

1. Klasnja P., Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. *Journal of biomedical informatics*. **45**, 184-198 (2012).

2. Kumar S., *et al*. Center of excellence for mobile sensor data-to-knowledge (MD2K). *Journal of the American Medical Informatics Associatio*n. **22**, 1137-1142 (2015).

3. Kozak A.T., *et al*. Technology-based interventions for weight management: current randomized controlled trial evidence and future directions. *Journal of behavioral medicine*. **40**, 99-111 (2017).

4. Haskins B.L., Lesperance D., Gibbons P., Boudreaux E.D. A systematic review of smartphone applications for smoking cessation. *Translational Behavioral Medicine*. **19**, 1-8 (2017).

5. Ben-Zeev D., Badiyani N. Mobile Health for Illness Management. In *Wellbeing, Recovery and Mental Health*. (eds. Slade, M., Oades, S., & Jarden, A.) 147-156 (Cambridge University Press, New York, 2017).

6. Waltz E. First Prescription App for Substance Abuse Approved by FDA. <https://spectrum.ieee.org/the-human-os/biomedical/devices/first-prescription-app-for-substance-abuse-approved-by-fda> (2017). Accessed 1 December 2017.

7. Ben-Zeev D., Brenner C.J., Begale M., Duffecy J., Mohr D.C., Mueser K.T. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophrenia bulletin*. **40**, 1244-1253 (2014).

8. Ho J., Intille S.S. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 909-918 (2005).

9. Nahum-Shani I., Hekler E.B., Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology*. **34**, 1209-1219 (2015).

10. Nahum-Shani I., Smith S.N., Tewari A., Witkiewitz K., Collins L.M., Spring B., Murphy S. Just in time adaptive interventions (jitais): An organizing framework for ongoing health behavior support. *Methodology Center technical report*. 14-126 (2014).

11. Eysenbach G. The law of attrition. Journal of medical Internet research. **7**, (2005).

12. Clawson J., Pater J.A., Miller A.D., Mynatt E.D., Mamykina L.. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 647-658 (2015).

13. Wagner III B., *et al*. e wrapper: operationalizing engagement strategies in mHealth. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 790-798 (2017).

14. Michie S., Yardley L., West R., Patrick K., Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *Journal of medical Internet research*. e232 (2017).

15. Murray E., Hekler E.B., Andersson G., et al. Evaluating digital health interventions: key questions and approaches. *Am J Prev Med*. 2016.

16. Free C., *et al*. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS medicine*. **10**, (2013).

17. Nilsen W., *et al*.. Advancing the science of mHealth. *Journal of health communication*. **17**, 5-10 (2012).

18. Study Record: Sense2Stop:Mobile Sensor Data to Knowledge. <https://clinicaltrials.gov/ct2/show/NCT03184389> (2017). Accessed 1 December 2017.

19. Rabbi M., *et al*. SARA: a mobile app to engage users in health data collection. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 781-789 (2017).

20. Projects Using MRTs. <https://methodology.psu.edu/ra/adap-inter/mrt-projects#proj> (2017). Accessed 1 December 2017.

21. Klasnja P., *et al*. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*. **34**, 1220- 1228 (2015).

22. Smith S.N., *et al*. Design Lessons from a Micro-Randomized Pilot Study in Mobile Health. In *Mobile Health*. 59-82 (Cham, Springer, 2017).

23. Dempsey W., Liao P., Klasnja P., Nahum-Shani I., Murphy S.A. Randomised trials for the Fitbit generation. *Significance.* **12**, 20-23 (2015).

24. Liao P., Klasnja P., Tewari A., Murphy S.A. Micro-randomized trials in mHealth. arXiv preprint arXiv:1504.00238. (2015).

25. Radu P.T., Yi R., Bickel W.K., Gross J.J., McClure S.M. A mechanism for reducing delay discounting by altering temporal attention. *Journal of the experimental analysis of behavior*. **96**, 363-385 (2011).

26. Kaplan B.A., Reed D.D., Jarmolowicz D.P. Effects of episodic future thinking on discounting: Personalized age-progressed pictures improve risky long-term health decisions. *Journal of applied behavior analysis*. **49**, 148-169 (2016).

27. Bickel W.K., Miller M.L., Yi R., Kowal B.P., Lindquist D.M., Pitcock J.A. Behavioral and neuroeconomics of drug addiction: competing neural systems and temporal discounting processes. *Drug and alcohol dependence*. **90**, 85-91 (2007).

28. National Institutes of Health, National Heart, Lung, and Blood Institute, Division of Blood Diseases and Resources: The Management of Sickle Cell Disease. <https://www.nhlbi.nih.gov/files/docs/guidelines/sc_mngt.pdf> (2002). Accessed 1 December 2017.

29. Okomo U., Meremikwu M.M. Fluid replacement therapy for acute episodes of pain in people with sickle cell disease. *Cochrane Database of Systematic Reviews*. (2007).

30. Karlson C.W., Baker A.M., Bromberg M.H., David Elkin T., Majumdar S., Palermo T.M. Daily pain, physical activity, and home fluid intake in pediatric sickle cell disease. *Journal of pediatric psychology*. **42**, 335-44 (2016).

31. Tewari S., Brousse V., Piel F.B., Menzel S., Rees D.C. Environmental determinants of severity in sickle cell disease. *Haematologica*. **100**, 1108-1116 (2015).

32. Connes P., Machado R., Hue O., Reid H. Exercise limitation, exercise testing and exercise recommendations in sickle cell anemia. *Clinical hemorheology and microcirculation*. **49**, 151-163 (2011).

33. Rankin C.H., *et al*. Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiology of learning and memory*. **92**, 135-138 (2009).

34. Shaw R.J., *et al*. Development of a theoretically driven mHealth text messaging application for sustaining recent weight loss. *JMIR mHealth and uHealth*. **1**, (2013).

35. Boruvka A., Almirall D., Witkiewitz K., Murphy S.A. Assessing Time-Varying Causal Effect Moderation in Mobile Health. *Journal of the American Statistical Association*. (just accepted).

36. SAS Institute Inc. 2017. SAS/STAT® 14.3 User's Guide, The GEE Procedure. Cary, NC: SAS Institute Inc.

37. Carey, V. (2015). Generalized Estimation Equation solver. R package version 4.13-19. https://CRAN.R-project.org/package=gee.

38. Liao P., Klasnja P., Tewari A., Murphy S.A.3 Sample size calculations for micro-randomized trials in mHealth. *Statistics in medicine*. **35**, 1944-1971 (2016).

39. Seewald, N. J., Sun, J., & Liao, P. (2016). MRT-SS Calculator: An R Shiny Application for Sample Size Calculation in Micro-Randomized Trials. arXiv preprint arXiv:1609.00695.

40. Liying H., Nicholas J. Seewald, Peng Liao and Ji Sun (2017). MRTSampleSize: A Sample Size Calculator for Micro-Randomized Trials. R package version 0.1.0. https://CRAN.R-project.org/package=MRTSampleSize.

41. Gollwitzer P.M. Implementation intentions: Strong effects of simple plans. *American Psychologist*. **54**, 93–503 (1999).

42. Tate, R. L., & Perdices, M. N-of-1 Trials in the Behavioral Sciences. In *The essential guide to N-of-1 trials in health* (eds. Nikles, J. & Mitchell G.) 19-41 (Springer, Netherlands, 2015).

43. McDonald, S., *et al.* The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. *Health Psychology Rev.*, **0**, 1–17 (2017).

44. Guyatt, G.H., *et al.* The n-of-1 randomized controlled trial: clinical usefulness: our three-year experience. *Annals of Internal Medicine*. **112**, 293-299 (1990).

45. Scuffham, P.A., *et al.* Using N-of-1 trials to improve patient management and save costs. *Journal of general internal medicine*. **9**, 906-913 (2010).

46. R.L. Kravitz, N. Duan, eds, and the DEcIDE Methods Center N-of 1 Guidance Panel. Design and implementation of n-of-1 trials: A users guide. *AHRQ Publication*, 13(14), January 2014. URLhttp://www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm.

47. Zucker D.R., Schmid C.H., McIntosh M.W., D'Agostino R.B., Selker H.P., Lau J. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *Journal of clinical epidemiology*, **50**, 401-410 (1997).

48. Zucker D.R., Ruthazer R., Schmid C.H. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *Journal of clinical epidemiology*, **63**, 1312-1323 (2010).

49. Nikles J., *et al*. Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: the case of palliative care. Journal of clinical epidemiology. **64**, 471-480 (2011).

50. Dallery, J., Cassidy, R. N., & Raiff, B. R. Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of medical Internet research*, **15**, (2013).

51. Collins, L. M. *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST)*. (Springer, New York, 2018).