

Randomised trials for the Fitbit generation

Data from activity trackers and mobile phones can be used to craft personalised health interventions. But measuring the efficacy of these “treatments” requires a rethink of the traditional randomised trial.

By **Walter Dempsey, Peng Liao, Pedja Klasnja, Inbal Nahum-Shani**
and **Susan A. Murphy**



MiteA/iStock/Thinkstock

Are you one of the millions of people who count how many steps they take each day, how much time they spend exerting themselves, or how often they fidget while asleep? If you are, then congratulations: you might not have known it, but you are part of a movement – the quantified-self movement – and all that data you are generating is rich with potential to improve human health and well-being.

The rise of wearable activity sensors, such as Fitbit, Fuelband and Jawbone, has generated a lot of public excitement as well as interest from the scientific community. Behavioural scientists in particular are enthused about the opportunities these devices provide not only to monitor activity, but also to structure behavioural interventions to alter patterns of activity and to help people lead healthier lives.

Consider the example of HeartSteps, a mobile health application that seeks to reduce sedentary behaviour. The app is installed on a smartphone which, when paired with a Jawbone device, monitors data such as steps taken per minute, a user's current activity and location, weather conditions, time of day and day of the week. HeartSteps then uses this data to come up with suggestions for physical activity which are adapted to the user's current situation.

Like many mobile health interventions, HeartSteps attempts to support behaviour change by providing users with choices that could have a positive impact on health, such as deciding to go for a walk during the lunch hour, or standing up while conducting conference calls. But as with many other types of treatment, it is unclear if the treatments proposed by HeartSteps actually produce the intended response.

Furthermore, there may be differences in treatment effects, both between individuals and within the same individual over time. If this "heterogeneity" exists, then, to maximise the impact of mobile health interventions, it may be desirable to adapt interventions so that they are only suggested at times and within contexts where they are likely to have the hoped-for effect.

In some cases we can identify these contexts using current theory, clinical expertise and initial consultations with users. In less clear-cut cases, evidence, in the form of data, should be brought to bear. But the traditional randomised trial designs that produce data for other areas of health and

medicine – whether for assessing treatment effects or adapting interventions – are not well suited to mobile health. Perhaps the "micro-randomised trial" can fill this knowledge gap?

A new trial design

Micro-randomised trials are trials in which participants are randomly assigned a treatment from the set of possible treatment actions at several times throughout the day. Thus each participant may be randomised hundreds or thousands of times over the course of a study. This is very different than a traditional randomised trial, in which participants are randomised once to one of a handful of treatment groups.

In the HeartSteps example, we want to know whether providing an activity suggestion increases the near-term physical activity of a user compared to those who do not receive an activity suggestion. We are also interested in learning whether the activity suggestions are more effective when the weather is good compared to when the weather is poor, or whether the effect of the activity suggestion is influenced by the busyness of the user's calendar. The micro-randomised trial can provide data for investigating these and other questions.

HeartSteps is a 42-day trial leading to 210 decision points per participant (or five decisions per participant per day). At each decision point a set of treatment actions is possible. These can include the type of treatment to provide, as well as whether to provide treatment at all. In general the set of potential treatment actions may depend on the state of the individual as well as the previously assigned treatments. For example, the set of treatments at once-a-day decision times may contain two alternative daily-step goal treatments – one a fixed 10 000-steps-a-day goal, the other an adaptive goal depending on previous activity levels. Treatments may be delivered via the smartphone, as in Figure 1, or via other wearable devices such as a wristband.

The treatment actions are often designed to have near-term effects on a longitudinal response, which we call the "proximal response". In the HeartSteps study, the proximal response is a participant's step count in the hour following the intervention.

Interventions are randomised, but the randomisation probabilities may not

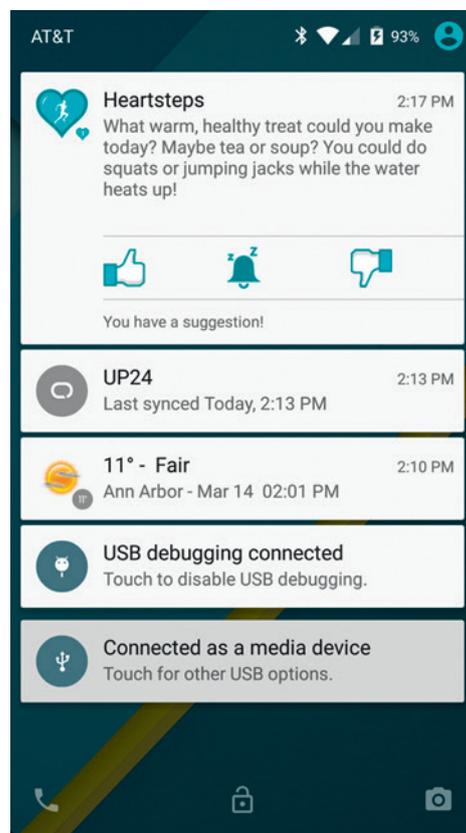


Figure 1. Example activity message from HeartSteps

be uniform across the set of treatment actions. Furthermore, the randomisation can vary within a person. For example, the randomisation probabilities may differ by how busy the individual's calendar is over the next hour.

In the HeartSteps example, the probability of providing an activity suggestion is set to 40% so that individuals are provided an average of two interventions per day. That level was decided based on a range of considerations, including the risk of treatment burden leading to dropout, the intrusiveness of interventions and concerns that users would begin to ignore activity suggestions through habituation.

Finally, as is the case for most clinical trials, the sample size for a micro-randomised trial is determined to guarantee a specified power to test one or at most a few very specific hypotheses.¹ However, the resulting data can be used to investigate a variety of questions useful in designing a mobile intervention.

Statistical challenges

In this era of personalised medicine and complex mobile health interventions, the

Alternatives to micro-randomisation

An alternative to the micro-randomised trial design is the single-case design used in the behavioural sciences. These trials usually only involve a small number of participants (fewer than 10) and the data analyses focus on the examination of visual trends for each participant separately.

In these trials, each participant is subject to periods of treatment interspersed with periods of no treatment. For example, during periods when a participant is on treatment, one might expect the response to be generally higher than the response during the time periods in which the participant is off treatment.

An excellent overview of single-case designs and their use for evaluating technology-based interventions is Dallery *et al.*⁶ This paper illustrates the visual analyses that would be conducted on each participant's data. A critical assumption is that the effect of the treatment is only temporary (no carry-over effect) so that each participant can act as his own control.

Additionally, one generally assumes that the effect of a treatment is constant over time. In settings in which treatments are expected to have sufficiently strong effects so as to overwhelm the within-person variability in response, these designs provide a nice alternative to the micro-randomised trial design.

Kratochwill and Levin consider a variety of ways to introduce randomisation into the single-case design, including randomising the order of treatment/baseline phases.⁷ The micro-randomised trial design can be viewed as a statistical generalisation of the single-case design, both in terms of involving multiple participants and including more traditional statistical analyses with sample size formula for the number of participants.

micro-randomised trial provides a new experimental paradigm. But even though this trial design is very new, there is already at least one completed trial,² while we (the authors) are involved in both the HeartSteps trial as well as a trial on smoking cessation that was scheduled to begin in late autumn 2015. There are also numerous trials in the proposal stages.

It is a positive start, though more work is necessary in the area of experimental design for mobile health. Interesting challenges include how best to design a clinical trial that ensures a given power to detect interactions between different intervention factors. Many questions involving causal inference arise in mobile health, and so causal inference methods must be extended to mobile health and micro-randomised trials in particular.

Some immediate needs are data analysis methods that can model and assess the existence of delayed effects of interventions. For example, an activity suggestion may not induce the participant to be physically active in the next hour, but it may inspire that person to be more active over the course of the next day.

Second, in many mobile health settings, the proximal response is not the primary clinical outcome. A mobile health intervention to tackle smoking might, for example, include treatments that are designed to help a person

manage momentary stress or momentary urges to smoke. So here the proximal response is stress or urge to smoke, yet the clinical outcome is time to relapse to smoking.

Natural methods from causal inference concern the use of surrogate outcomes or mediation analyses. These causal inference methods need to be generalised to the mobile health setting, as do methods from the field of missing data, because missing data is rampant in mobile health trials due to slippage of wearable devices, problems with wireless connectivity and user non-adherence.

In addition, there are a multitude of secondary analyses that are possible using data arising from a micro-randomised trial. Beyond delayed effects, the examination of effect moderation – testing for the treatment effect conditional on covariate history – is of interest. Mobile health data can also be used to predict latent states, such as underlying stress levels given associated sensor measurements. Latent state models may be useful in the prediction of time-to-event outcomes, such as time until next cigarette and drug lapse or relapse more generally.

Finally, the data is useful for building optimal mobile health treatment policies for future patients, by estimating the best treatment action to take given a patient's observed history. In HeartSteps, for example,

the data can help determine the timing of activity suggestions to maximise the patient's long-term physical activity.

Individual experience

Secondary analyses are also of interest because of concerns that there are systematic differences, both between individuals and within individuals, in response to treatment actions in the mobile health setting. This is the type of concern that motivates research in personalised and precision medicine. It is unclear if the response variability between individuals and within individuals can be captured by covariate information, but given the sequential interlacing of time-varying covariates with randomised treatment actions, we believe that the micro-randomised trial provides a unique opportunity to investigate treatment effect differences. This point deserves significant emphasis as science begins to tackle problems in maximising not only population health, but also the health of each particular individual.

A further interesting challenge in mobile health is whether and how any mobile intervention can be optimised in real time – that is, as the individual experiences the treatments and responses are recorded. Multi-armed bandit and contextual bandit learning algorithms provide a promising approach to optimisation. These learning algorithms use sequential randomisations that are increasingly biased towards the treatment that appears, according to past data, most effective. They have been popularised in the online advertising space by Google and Facebook.

A critical assumption underlying bandit algorithms is that the current treatment (or ad placement) will not impact the type/context of the subsequent participant. This assumption is likely to be valid in many advertising applications as the randomisation of 'treatments' is sequential and between participants as they visit the website, as opposed to within a participant. Similarly in mobile health, if the influence that a treatment has on participant learning, treatment fatigue and non-adherence is low, then the above assumption may be approximately correct. Rabbi *et al.*, as well as members of our team, are currently investigating the extent to which bandit algorithms might be used to optimise mobile interventions in real time.^{3,4}

The dawn of data-based mobile health

Mobile devices and wearable activity sensors hold the promise of providing low-cost supportive behavioural interventions and thus lowering medical costs, particularly for those individuals struggling with chronic conditions. It is no wonder that behavioural scientists are excited by the potential and have begun to incorporate these new methods in their work.

However, as Stephen Senn warned in 2001 with respect to the advance of genomic data in the health fields, the potential of new technologies may not be realised.⁵ We, however, have high hopes for this technology, though we caution against the slow pace of development of corresponding statistical methods for the rigorous collection and study of data in mobile health.

The micro-randomised trial described here is just a first step towards designing experiments in mobile health. These methods should provide a rigorous statistical

framework and will hopefully lead to effective mobile health interventions.

References

1. Liao, P., Klasnja, P., Tewari, A. and Murphy, S.A. (2015) Micro-randomized trials in mHealth. arXiv:1504.00238 [stat.ME].
2. Beckjord, E., Shiffman, S., Cardy, A., Bovbjerg, D., Siewiorek, D., Jindal, G., Smith, V., Klein, D. and Praduman, J. (2015) Testing a context-aware, evidence-based, just-in-time adaptive intervention for smoking cessation. Lecture to Society for Behavioral Medicine Annual Meeting, San Antonio, TX, 25 August.
3. Rabbi, M., Pfammatter, A., Zhang, M., Spring, B. and Choudhury, T. (2015) Automated personalized feedback for physical activity and dietary behavior change with mobile phones: A randomized controlled trial on adults. *Journal of Medical Internet Research mHealth and uHealth*, 3(2), e42.
4. Lei, H., Tewari, A. and Murphy, S. (2014) An actor-critic contextual bandit algorithm for personalized interventions using mobile devices. *Advances in Neural Information Processing Systems*, 27.
5. Senn, S. (2001) Individual therapy: New

dawn or false dawn? *Drug Information Journal*, 35, 1479–1494.

6. Dallery, J., Cassidy, R. and Raiff, B. (2013) Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of Medical Internet Research*, 15(2), e22.

7. Kratochwill, T. and Levin, J. (2010) Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, 15, 124–144.

Walter Dempsey is a postdoctoral research fellow at the University of Michigan, Department of Statistics

Peng Liao is a graduate student at the University of Michigan, Department of Statistics

Pedja Klasnja is assistant professor of information, School of Information, and assistant professor of health behavior and health education, School of Public Health at the University of Michigan

Inbal Nahum-Shani is research assistant professor at the Survey Research Center, Institute for Social Research, University of Michigan

Susan A. Murphy is H.E. Robbins Distinguished University Professor of Statistics, professor of psychiatry, and research professor, Institute for Social Research at the University of Michigan

