

# Moving Toward Rigorous Evaluation of Mobile Health Interventions

Kristin A. Linn

*Abstract.* Qian, Klasnja and Murphy provide an assumption that allows for unbiased estimation of treatment effects in microrandomized trials when the data are modeled using linear mixed models with endogenous covariates. In this discussion, the validity of the assumption in the context of the HeartSteps microrandomized trial is reassessed. The utility of a marginal interpretation, versus the proposed conditional-on-the-random-effect interpretation, is also discussed.

*Key words and phrases:* Linear mixed model, endogenous covariates, microrandomized trial, causal inference.

Smartphones and other smart devices have become ubiquitous in everyday life throughout the world and have arguably increased human knowledge, efficiency and productivity. Simultaneously, these products provide access to a proliferating host of distractions that encourage avoidance and procrastination. Smart devices can collect, store and visualize large amounts of personal user data on daily activities and health metrics, with the promise that this comprehensive, salient information may facilitate user behavioral changes that lead to improved individual health. However, to deliver on this promise and meaningfully impact population health, strategies for sustaining user motivation and healthy behaviors over time must be developed and rigorously tested (Patel, Asch and Volpp, 2015). Examples of strategies to maintain user engagement include sending push notifications with reminders or personalized activity suggestions as exemplified in Qian, Klasnja and Murphy’s paper; incorporating gamification elements designed using behavioral economic principles (Miller, Cafazzo and Seto, 2016, Cotton and Patel, 2019); and distributing monetary incentives or other types of rewards, possibly awarded via lotteries. Tailoring these and other interventions based on time-varying user characteristics will likely result in better health outcomes in a target population compared to one-size-fits-all approaches.

There is a need for rigorous methodology to evaluate mobile health interventions and inform public health programs and policies, and this need extends well beyond academic research and the realm of public health. The

tech industry is increasingly leveraging the promise of mobile health apps and their data to attract new users and motivate engagement with mobile platforms. A number of established companies already exist in this space, including household names such as Nike, FitBit and Garmin. Strava is an example of an activity tracking app with an integrated social networking component. Noom is a newer company that sells app-based diet programs and costs clients a minimum of \$59 USD per month. The company has a special program that targets diabetes prevention and was deemed successful in a pilot study, but the study did not include a randomized comparator and focused on weight loss and engagement as the primary outcomes rather than HbA1c (Michaelides et al., 2016). Noom published a similar study focused on hypertension prevention (Toro-Ramos et al., 2017). The company partners with insurance companies and employers to provide its services to a variety of clients. As mobile health services continue to flood the market, information about their effectiveness over time will be useful to potential clients and business partners.

Using the HeartSteps microrandomized trial (MRT) as an illustrative example, Qian, Klasnja and Murphy provide an accessible introduction to some of the challenges in the analysis of data from MRTs with a focus on handling endogenous covariates and interpreting estimates of time-varying intervention effects. In their illustrative data analysis, the authors argue that the conditional independence assumption (equation (10)) holds for each of the three time-varying covariates in their model: location ( $X_{it,2}$ ), 30-minute step count prior to time  $t$  ( $X_{it,3}$ ), and the unavailability indicator at time  $t$  ( $I_{it}$ ). In particular, they argue that the time-varying location and unavailability instances are likely to be conditionally independent of

---

Kristin A. Linn is Assistant Professor of Biostatistics, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Drive, Philadelphia, Pennsylvania 19104, USA (e-mail: [klinn@upenn.edu](mailto:klinn@upenn.edu)).

the subject-specific random effects because of the enrollment criterion that required participants to either be employed full-time or a student. However, scrutiny of the authors' given justification may be warranted.

Potentially problematic are latent factors such as levels of depression and anxiety that could be associated with a participant's time-varying location and activity levels (De Moor et al., 2006). For example, participants with higher levels of depression or anxiety may tend to remain at home for longer periods of time, skip work or class irregularly, or avoid social functions that might involve walking or other physical activity. Furthermore, exercise has been shown to have a small effect on depression symptoms (Cooney et al., 2013) such that the treatment effectiveness could influence this latent factor over time. Similarly, inherent introverted or extroverted qualities may influence a participant's time-varying location. Extroverted participants with larger social networks may be more likely to walk to after-work social functions or recruit friends to support them in achieving their activity goals, which may subsequently impact step counts.

Even in settings where the conditional independence assumption can be carefully justified, a marginal effect estimate may be preferred. For example, large corporate employers might be interested in developing push notifications, nudges or incentives to increase the activity levels of their employees. In this example, management would likely be interested in population-level intervention effects on activity, weight loss, alertness, job performance and overall health. Based on the results of a pilot or full-scale program, they might be able to make a deal with an employee benefits provider to reduce the company's contributions to the plan. In addition, there might be ethical and privacy concerns with respect to estimating employee-specific, conditional effects if there is any chance this information could influence performance evaluations, employment contract renewals, raises, promotions and so forth.

The authors' methodological contribution opens an avenue for analyzing data from MRTs using existing software, which is a critical step toward greater adoption of these types of studies. Ease of implementation also opens the door for quick application of the proposed estimation approach by nonstatisticians or time-deprived statisticians who may not take time to fully consider whether the critical assumptions seem reasonable. Of course, this is not unique to the authors' contribution in this work but rather

is always a concern when a statistical method relies on a set of assumptions, whether they are able to be checked empirically or not. I congratulate Qian, Klasnja and Murphy on their notable contribution to the literature in an area of high contemporary interest.

The innovation of MRTs is in some ways analogous to the emergence of electronic medical records (EMRs) in the late 1900s. EMRs were more practical and scalable than hard copy charts and promised to bring about more organized, comprehensive and higher-quality medical care. At the same time, EMR data present a number of statistical challenges including, but certainly not limited to, selection bias and measurement error. While MRTs are better suited for developing impactful, scalable, time-varying interventions in the real-world compared to single-randomization studies in well-controlled research environments, analysis of data from MRTs is more complex and requires great care. This paper is an important step toward thoughtful, rigorous and widely accessible analysis of data from such designs. I have no doubt these authors and other thought leaders in the field will continue to provide research innovations that improve mobile health and help it deliver on its promise of user health and well-being.

## REFERENCES

- COONEY, G. M., DWAN, K., GREIG, C. A., LAWLOR, D. A., RIMER, J., WAUGH, F. R., MCMURDO, M. and MEAD, G. E. (2013). Exercise for depression. *Cochrane Database Syst. Rev.* **9** CD004366.
- COTTON, V. and PATEL, M. S. (2019). Gamification use and design in popular health and fitness mobile applications. *Am. J. Health Promot.* **33** 448–451. <https://doi.org/10.1177/0890117118790394>
- DE MOOR, M. H., BEEM, A., STUBBE, J. H., BOOMSMA, D. I. and DE GEUS, E. J. (2006). Regular exercise, anxiety, depression and personality: A population-based study. *Prev. Med.* **42** 273–279.
- MICHAELIDES, A., RABY, C., WOOD, M., FARR, K. and TORO-RAMOS, T. (2016). Weight loss efficacy of a novel mobile diabetes prevention program delivery platform with human coaching. *BMJ Open Diabetes Res. Care* **4** e000264.
- MILLER, A. S., CAFAZZO, J. A. and SETO, E. (2016). A game plan: Gamification design principles in mHealth applications for chronic disease management. *Health Inform. J.* **22** 184–193.
- PATEL, M. S., ASCH, D. A. and VOLPP, K. G. (2015). Wearable devices as facilitators, not drivers, of health behavior change. *JAMA* **313** 459–460. <https://doi.org/10.1001/jama.2014.14781>
- TORO-RAMOS, T., KIM, Y., WOOD, M., RAJDA, J., NIEJADLIK, K., HONCZ, J., MARRERO, D., FAWER, A. and MICHAELIDES, A. (2017). Efficacy of a mobile hypertension prevention delivery platform with human coaching. *J. Hum. Hypertens.* **31** 795–800.