Are we making the most of ecological momentary assessment data? A comment on Richardson, Fuller-Tyszkiewicz, O'Donnell, Ling, & Staiger, 2017

Special Issue on Advanced Analytic and Statistical Methods in Health Psychology

Kathryn Lynn Modecki
School of Applied Psychology, Griffith University
Menzies Health Institute Queensland; Queensland Australia
School of Psychology & Exercise Science, Murdoch University; Western Australia, Australia

Gina L. Mazza
Department of Psychology
Arizona State University; Tempe, AZ

Health Psychology Review, 11(3). 295-297.

https://doi.org/10.1080/17437199.2017.1347513

Without a doubt, smartphones in our pockets have revolutionized day-to-day life. This constant connection, and health psychologists' associated ability to capture repeated, fine-grained assessments of people's behaviors, experiences, and contexts via ecological momentary assessment (EMA) provides an unmatched opportunity for conducting original research. Yet this opportunity brings with it its own tension (Baraniuk, 2014)—Are we making the most of it?

This tension between recognizing that technology has dramatically altered our prospects for amassing health psychology data, and acknowledging our current limits in exploiting these possibilities motivates our commentary. Here we highlight four critical points which are nicely underscored by the authors in their recommendation of a novel method for analyzing EMA data. To each of these points of agreement, we add accompanying caveats and recommendations based on noted limitations and constraints within the field.

Four points of agreement (with a few caveats)

First, we appreciate a key theme featured by the authors—the potential for leveraging secondary analyses from hard-won longitudinal data. In this case, the focus is on intensive longitudinal data, with many repeated assessments nested within people. Apps and widespread ownership of smartphones have made collecting such EMA data relatively easy, but collecting useful, reliable, and valid EMA data from populations of interest is not (Trull & Ebner-Priemer, 2009; Uink, 2017). Rather, doing so remains a challenge—with enrollment, compliance, and engagement under-discussed but essential elements to garnering valid ambulatory assessments of emotion, decisions, health, and well-being (Hamaker & Wichers, 2017; Wichers et al., 2011). Thus, our endorsement of the use of secondary analysis of EMA data comes with recognition of its practical challenges. For those instances in which relevant, representative data are available, scholars should by all means be mining these to glean what we can about health behaviors and how best to intervene to enhance them.

Second, regression trees can readily accommodate complex relations between the predictors and outcome; this represents a major advantage when analyzing EMA data. Regression trees are nonparametric, meaning they do not rely on distributional assumptions or posit that the association between each predictor and the outcome is linear. Briefly, regression tree analysis uses recursive partitioning to sequentially split the sample into sets of observations with similar scores on the outcome of interest. Because splits are conditional on previous splits, regression trees may represent interactions (Doove, van Buuren, & Dusseldorp, 2014). Statistically, this is an important advantage of regression trees given that estimating and probing interactions has received relatively little attention in multilevel modelling (Enders & Tofighi, 2007), which is typically used to analyze EMA data. Conceptually, this is an important advantage because real-world behaviors may be best envisioned as resulting from person-context-behavior interactions (Ciampi, Chang, Hogg, & McKinney, 1987; Wilhelm & Grossman, 2010). Indeed, single variables can only explain so much variability in any given health behavior. Acknowledging this complexity that exists, with interrelations among real-world constructs, is constructive for the field. That said, although the potential exists within regression trees to identify meaningful subgroups of people, or behaviors, or situations, which are defined by specific combinations of predictors, it is debatable as to whether regression trees are likely to do so, particularly within a single sample. Given that regression tree analysis is an exploratory, data-driven technique, any application to clinical practice or intervention design would need to be extraordinarily cautious (Marzano et al., 2015). Instead, exploratory techniques such as regression trees are useful for developing theories which can then be tested in new samples using theory-driven techniques.

Third, EMA tends to generate an overwhelming amount of data (Stone & Shiffman, 2002), rendering the availability of a data-reduction technique attractive. Recall that EMA datasets typically consist of numerous constructs, measured at many time points within each person, among samples consisting of 100 or more. The possibility of reducing these thousands of data points to a series of binary decision points suitable for diagnosis or prediction is intuitively appealing. At the same time, a single regression tree is highly sensitive to its first split and is prone to instability even with minor data perturbations. While the authors encourage us to consider what we may find with data-driven techniques, we would caution that it is equally important to consider what we may miss. Key variables that are correlated with the first splitting variable but are not chosen may mistakenly appear unimportant.

Further, regression trees are subject to overfitting, which then limits the generalizability of results. Thus, parsing down into manageable parcels of information is useful, but only to the extent that resultant parcels have meaning and relate in informative ways to focal constructs. Consequently, future applications should consider bagging (bootstrap aggregation), random forests, or ensemble methods as possible alternatives to single regression trees (Hajjem, Bellavance, & Larocque, 2014). Although such results are more difficult to interpret because they are not based on a single regression tree, bagging, random forests, and ensemble methods help to determine variable importance and typically improve prediction. This tradeoff between improved prediction and performance on the one hand, and accessibility and ease of use on the other, characterizes many of the challenges common to present-day health research.

Fourth, in EMA research, the ability to visually summarize results is attractive and would be practically useful. Regression tree analysis affords this possibility and, in theory, visualization might help bridge EMA research with hands-on work in the field. That said, regression trees may not necessarily be the choice method for doing so. Although a single regression tree provides an easy-to-follow visual representation of the results, single regression trees are rarely used in other research arenas due to the issues raised above.

Reducing the tension between possibility and limits

The widespread proliferation of smartphones creates unmatched opportunities for generating and testing cutting-edge questions in health psychology. Enhanced by readily-available apps, smartphone-based EMA provides health psychologists with the ability to capture individuals' health habits, decisions, behaviors, and associated affect and contexts (Ginexi Riley, Atienza, & Mabry, 2014). Yet challenges associated with collecting high-quality ambulatory assessment data and hurdles tied to data management and analysis, may dictate the pace with which health psychologists are prepared to leverage EMA's advantages. Consequently, exploratory techniques such as regression trees should arguably be listed among a diverse repertoire of potential methods implemented during secondary data analysis—supplementing rather than replacing testing of theorized models. A number of valuable premises underlie the proposal of regression tree analysis as a fruitful methodology for analyzing datasets derived from EMA. Here we highlighted four, with accompanying suggestions to more fully leverage the dynamic "in situ" data afforded by EMA.

References

Baraniuk.C. (2013, January 14). The promise of technology [Blog post]. Retrieved from http://www.themachinestarts.com/read/2013-01-the-promise-of-technology

Ciampi, A., Chang, C. H., Hogg, S., & McKinney, S. (1987). Recursive partition: A versatile method for exploratory-data analysis in biostatistics. In *Biostatistics* (pp. 23-50). Springer Netherlands.

Doove, L. L., van Buuren, S., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics and Data Analysis, 72*, 92-104. http://dx.doi.org/10.1016/j.csda.2013.10.025

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological methods*, *12*(2), 121.

Hajjem, A., Bellavance, F., & Larocque, D. (2014). Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, *84*(6), 1313-1328.

Ginexi, E. M., Riley, W., Atienza, A. A., & Mabry, P. L. (2014). The promise of intensive longitudinal data capture for behavioral health research. Nicotine & Tobacco Research, *16*, S73-75.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10-15.

Marzano, L., Bardill, A., Fields, B., Herd, K., Veale, D., Grey, N., & Moran, P. (2015). The application of mHealth to mental health. *The Lancet Psychiatry*, *2*(10), 942-948.

Neubert, S., Arndt, D., Thurow, K., & Stoll, R. (2010). Mobile real-time data acquisition system for application in preventive medicine. *Telemedicine and e-Health*, *16*(4), 504-509.

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24*(3), 236-243.

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. *Psychological Assessment, 21*(4), 457-462.

Uink, B (2017, March). Steps toward an inclusive research methodology for "at risk" youth. Paper presented at the AIATSIS National Indigenous Research Conference, Canberra, ACT. Australia.

Wichers, M., Simons, C. J. P., Kramer, I. M. A., Hartmann, J. A., Lothmann, C., Myin-Germeys, I., ... & van Os, J. (2011). Momentary assessment technology as a tool to help patients with depression help themselves. *Acta Psychiatrica Scandinavica*, *124*(4), 262-272.