



Smartphone-based passive assessment of mobility in depression: Challenges and opportunities



Brenna N. Renn^{a,*}, Abhishek Pratap^{b,c}, David C. Atkins^a, Sean D. Mooney^b, Patricia A. Areán^a

^a Department of Psychiatry & Behavioral Sciences, University of Washington, 1959 NE Pacific St. Box 356560, Seattle, WA, 98195, USA

^b Department of Biomedical Informatics and Medical Education, University of Washington, 1959 NE Pacific St. Box 358047, Seattle, WA, 98195, USA

^c Sage Bionetworks, 1100 Fairview Ave N, M1-C119, Seattle, WA, 98115, USA

ARTICLE INFO

Keywords:

Depression
Smartphone
Mobile health (mHealth)
Passive data collection
Mobility
Physical activity

ABSTRACT

Advances in technology have ushered in exciting potential for smartphone sensors to inform mental health care. This commentary addresses the practical challenges of collecting smartphone-based physical activity data. Using data ($N = 353$) from a large scale, fully remote randomized clinical trial for depression, we discuss findings and limitations associated with using passively collected mobility data to make inferences about depressive symptom severity. We highlight a range of issues in associating mobility data with mental health symptoms, including a high degree of variability, data featurization, granularity, and sparsity. Given the considerable efforts toward leveraging technology in mental health care, it is important to consider these challenges to optimize assessment and guide best practices.

Clinical Trials.gov identifier: NCT01808976.

Smartphones, which are mobile phones with advanced hardware and software computing capabilities, are ubiquitous—77% of Americans own such a device (Pew Research Center, 2017). Of profound clinical importance is their ability to passively collect phone sensor and usage data with the potential to digitally phenotype a person's behavior and make inferences about one's mental health. Both major depressive disorder and subthreshold depressive symptomatology are relatively common mental health conditions that impart a high level of disability and burden (Center for Behavioral Health Statistics and Quality, 2016; Karsten et al., 2011). Moreover, depression often goes undetected and thus untreated (Cepoiu et al., 2008). There is evidence that moderate-to-vigorous intensity physical activity may be associated with lower odds of depression in national samples (Vallance et al., 2011); however, less is known about patterns of physical activity among depressed adults, particularly in uncontrolled settings. Using ubiquitous smartphone technology to monitor physical activity and passively detect changes in mobility may facilitate recognition and timely treatment of mood disorders by capturing depression symptoms such as lassitude, anhedonia, psychomotor retardation, and overall reduced activity engagement. For example, accelerometry provides information about physical activity and sleep; global positioning system (GPS) information offers even more details about the degree of movement and the context and variety of activity. These passively collected features are less intrusive and more precise than traditional self-report

approaches, and may provide clinical information important for determining a person's well-being.

While still in the proof-of-principle stage, pioneering studies provide initial evidence of the feasibility and potential efficacy of using such passively collected physical activity data for clinical inferences. However, there is yet no clear consensus or guideline for using smartphone technology to track mobility and other daily physical activity. Moreover, emotional states and other symptoms of clinical disorders such as depression are more distal from the features normally used in passive data collection (Mohr, Zhang, & Schueller, 2017). Put simply, passive sensor data of movement and location must first be translated into meaningful features of physical activity, such as location, activity type, and movement intensity. These data then become proxies, or behavioral markers, of motor activity, fatigue, and reduced engagement in activity. Finally, in order to make inferences about mental health, these behavioral markers must accurately reflect an underlying clinical state (e.g., depression).

Despite the emerging interest in these sensors, there are few demonstrations of using such sensors to passively collect physical activity data and translate it into meaningful mental health metrics. Work by Ben-Zeev, Scherer, Wang, Xie, and Campbell (2015) used smartphone sensor technology to detect geospatial, kinesthetic, sleep, and social (i.e., proximity to human speech) variables among 47 undergraduate and graduate college students in the US over a 10-week period.

* Corresponding author.

E-mail addresses: bnrenn@uw.edu (B.N. Renn), apratap@uw.edu (A. Pratap), datkins@uw.edu (D.C. Atkins), sdmooney@uw.edu (S.D. Mooney), parean@uw.edu (P.A. Areán).

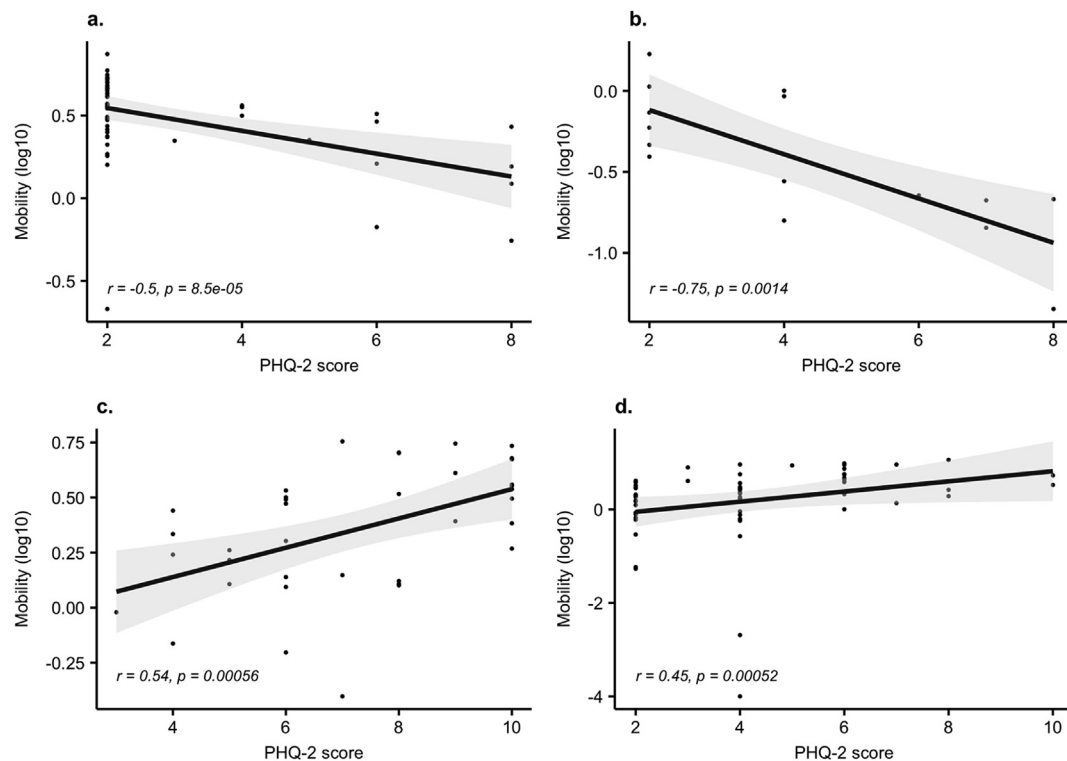


Fig. 1. Scatter plot between mobility (log10 transformed) and daily PHQ-2 score across the study period of 12 weeks for four individuals.

Increased GPS-derived activity was associated with fewer depressive symptoms in this sample; however, there was no association between depression and kinesthetic activity (derived from embedded accelerometers). Although this small study used a minimally depressed sample of university students, they demonstrated proof-of-concept of an innovative “digital trace” (p. 224) of passively collected behavioral data with minimal burden to participants.

Saeb et al. (2015) used an open-source Android mobile phone application to gather sensor data to identify depressive symptom severity among 28 US adults aged 19–58 years of age. Their app collected data on the location and distance traveled by the participant, in addition to phone usage characteristics such as duration, frequency, and temporal information of phone usage accumulated by the participants over two weeks of study participation. The authors could accurately classify someone with depressive symptoms 86.5% of the time based on self-reported depressive symptoms. While this was an important preliminary study, data sparsity limited key analyses to 18 individuals with GPS location data. Furthermore, this study did not assess physical activity *per se*; they measured overall distance accumulated by a participant regardless of mode of transportation.

Taken together, this emerging literature base suggests that passive smartphone sensors may offer an unobtrusive collection of behavioral markers, such as physical activity, associated with mental health. Nonetheless, issues of data precision, sparsity, and adherence are implementation barriers. Moreover, the majority of work using phone sensor data to date has been on small samples, predominately comprised of college students. Evolving technological advances may offset some of these issues, yet the feasibility and effectiveness of such technology needs to be extended to larger and more representative samples. Although the potential for real-time assessment of mental health states are exciting, there are few reports on the practical challenges of collecting smartphone-based data. The purpose of this commentary is to present a proof-of-principle demonstration linking passively collected mobility data with mood from the largest remote study of depression to date. We discuss the challenges and opportunities illustrated by this case in the broader context of using passively collected GPS-derived

mobility data to draw inferences about mental health.

1. Case example: the Brighten study

To illustrate the challenges and potential applications of capturing mobility using mobile phones, we present secondary data analysis of the Brighten study, which has been described in detail elsewhere (Anguera, Jordan, Castaneda, Gazzaley, & Areán, 2016; Areán et al., 2016). In brief, mobility data was passively collected using smartphone-based sensors from 600 adults in the US aged 18 years and older. All participants endorsed symptoms of depression (Patient Health Questionnaire [PHQ-9; Kroenke, Spitzer, & Williams, 2001] ≥ 5), provided informed consent, and were enrolled in a randomized controlled trial that delivered both assessments and treatment entirely through mobile devices. A GPS-derived mobility estimate and self-reported depressive symptoms were collected using a publicly available passive data-tracking mobile application developed by Ginger.io™. Mobility was defined as the approximate distance in miles (determined by location data) covered by the user by foot on a particular day. The application developer used a proprietary algorithm to infer mobility, such that raw data only contributed to mobility scores if movement speed was ≤ 3.0 mph. Participants self-reported their depressive symptoms using two items (PHQ-2; Kroenke, Spitzer, & Williams, 2003) to assess frequency of depressed mood and anhedonia. The response scale was modified for daily tracking, offering a 5-point rating scale (1 = *not at all*; 5 = *most of the day*), with total possible scores ranging from 2 to 10. To best understand the relation between daily mobility and depressive symptoms, we restricted our analysis to individuals who contributed at least two weeks of both mobility and depressive ratings ($n = 353$).

There was heterogeneity in daily mobility ($M = 1.56$, $SD = 1.59$) and PHQ-2 scores ($M = 4.31$, $SD = 2.23$). Using Spearman correlations on individual participant data, only 8.4% of the sample ($n = 30$) evidenced a statistically significant association between mobility and depressive symptoms ($p < .05$); once corrected for multiple testing, this correlation was statistically significant for only one participant ($r_s = -0.50$, $p < .001$). We did not find GPS-derived mobility to be a

valid surrogate for predicting daily variation in depressive symptoms at the cohort level (R^2 values: $M = -0.07$, $SD = 0.04$, using an ensemble-based random forest regression). To assess the robustness of the predictions we used a repeated sampling approach (50 random training/test data splits), where each sample included a 70/30 split of training and test data. To illustrate the individual level heterogeneity in associating depressive symptoms with mobility, we present data from four representative participant profiles in Fig. 1. The top two panels (1a, 1b) display two participants with inverse correlations between depressive symptoms and mobility; the bottom right panel illustrates no association (1d). Finally, the bottom left panel (1c) highlights a significant, albeit small, positive association between PHQ-2 scores and mobility. Such variations in mobility as it relates to daily PHQ-2 scores demonstrate the challenge of using digitally-captured physical activity as a predictor or behavioral marker of depressive symptoms at a cohort level.

However, before discounting smartphone-based mobility as an indicator of mental health status, we highlight several, more fundamental challenges in data collection and processing that could inform the development and design of new research studies in this domain:

- 1 **Data Featurization:** This refers to the process of converting raw data (i.e., GPS data) into “features,” or meaningful summaries, of the data. Often, raw data is of minimal value by itself. In the case of GPS-derived mobility, features are extracted from raw data and may refer to information about how quickly the person is moving and where they are located. To infer meaningful predictive patterns from the novel data generated by smartphones, the next steps are to generate suitable features from the raw data and translate these into meaningful data for analysis. For example, location may be used to infer whether they are engaging in occupational, hedonic, or social activity. We used a third party application to collect mobility data from the smartphone, which provided pre-processed, per-day mobility estimates, thus limiting featurization. The absence of raw data restricted our data analytical and feature engineering options.
- 2 **Data Granularity:** Granularity is the frequency with which data is collected. For example, a daily GPS signal can be batched into daily mobility or collect per hour (or even per minute). This level of “granularity,” or specificity, could potentially assist with a better understanding of patterns in the data and associations between variables. Because of proprietary software restrictions, we were not able to gather granular data to observe within-day trends to study the relation between circadian rhythm to daily fluctuations in PHQ-2 scores.
- 3 **Data Sparsity:** The issue of sparsity refers to the missing data inherent in linking passively collected features with volitional participant responses (i.e., completion of the PHQ-2). Although we passively captured mobility features (to minimize participant burden), assessment of depressive symptoms required daily input by the participant via the PHQ-2. Insufficient daily data led to some participants being dropped from our analysis, and may be less representative of those with motivational or attentional barriers (that is, those who did not complete daily PHQ-2 ratings or forgot to carry their phone with them).

2. Challenges of smartphone data collection

Despite the growing excitement in the use of smartphone technology to increase access to care and yield robust, timely, and accurate patient data, the infancy of this field begets complications and challenges. As noted above, the sparsity of data collected by smartphone sensors is a hurdle for repurposing these devices for mental health care. Indeed, Saeb et al. (2015) noted that 30% of their original sample did not have sufficient GPS and/or phone usage data available for analysis. Similarly, over half (50.8%) of the 126 adult participants recruited for a mobile sensing study for depression in Germany and Switzerland

uninstalled the app within the first 2 weeks of the study (Wahle, Kowatsch, Fleisch, Rufer, & Weidt, 2016). On a practical level, the collection of data using smartphone sensors requires that people adhere to carrying their phone and keeping it charged. Given the relatively small sample sizes in the extant literature, these missing data are not inconsequential and give pause to researchers attempting to implement these tools.

Precision of measurement is also necessary when using these technologies to advance diagnosis of mental health diagnoses. The reality of using embedded sensor technology and commercial apps is that many of the estimation formulas are proprietary, yielding questions about the exact construct being assessed and the precise method of calculation. For example, the GPS data captured by smartphone technology is not yet entirely accurate. Moreover, we relied on data from proprietary algorithms developed by the app developer; without input from our research team, the data constituting “mobility” was capped at 3.0 mph. Clearly, this feature did not capture mobility in the form of fast walking, jogging, running, or cycling. Since most trials have repurposed existing commercial platforms without research-quality raw data, replication of findings is limited. Finally, issues of privacy and transparency are of concern to consumers, clinicians, and researchers. An industry survey of digital health technologies underscored this very issue, such that only 18% of respondents expressed willingness to share their health data with technology companies (Gandhi & Wang, 2015). Although consumers were much more likely to agree to sharing data with healthcare providers, they ultimately felt that they should be in control of their data. As this field progresses, discussions of privacy and transparency are paramount to promote trustworthy applications while balancing the needs of diverse stakeholder and consumer interests (Torous & Roberts, 2017).

3. Discussion

Among our large sample of adults with depression, passively collected mobility was poorly associated with daily ratings of depressive symptom severity. The high degree of intra- and inter-individual variability in mobility data highlights the limitations of inferring daily mental health symptomatology from mobility data at a population level. In particular, our experience in the [name blinded for review] trial uncovered critical issues of featurization, granularity, and data sparsity of the GPS data feeds. Harnessing the potential of such sensor data to predict mental health outcomes will require further refinement to ensure that the signal for from mobility is indeed meaningful. Leveraging technology will also require successfully engaging consumers to use these devices, especially for active tasks such as self-reporting and other assessments. However, smartphone technology has unique advantages because of the proliferation of such devices. Passive data collection techniques using smartphone sensors are feasible, require little effort of user, and leverage behaviors the consumer has already adopted (e.g., charging the phone regularly, rather than expecting an individual to remember to charge a novel device). Future work will need to rectify these limitations with the great potential of such technology for physical activity and mental health research.

There is growing interest in using passively collected mobility information to infer behavior related to depression, with the intent of unobtrusively detecting and monitoring symptoms, predicting symptom severity, and tailoring intervention delivery. The smaller studies to date (Ben-Zeev et al., 2015; Saeb et al., 2015; Wahle, Kowatsch, Fleisch, Rufer, & Weidt, 2016) have demonstrated proof-of-concept of this technology but have not conclusively demonstrated an association between passive mobile data and severity of depressive symptoms. Our examination of a large, nationally recruited sample of individuals with depressive symptoms did not show a meaningful relationship between GPS-derived mobility and daily mood. Our analyses of the [trial name blinded for review] data demonstrated marked heterogeneity in the GPS-derived mobility feature, both across and within individuals,

making it difficult to link to daily ratings of depressive symptoms. In addition to the proprietary calculation of mobility in our study, we believe there might be other patterns and noise affecting the accurate quantification of mobility. Indeed, such analyses infer a direct relationship between phone movement and person movement, when in fact this is complicated by people's idiosyncratic use of their phones. For example, leaving a cell phone stationary during the day could falsely indicate sedentary lifestyle, signaling a false alarm. Similarly, passive sensing using smartphone technology remains difficult for water-based activities (e.g., swimming, water aerobics) or contact sports, both instances where users are not likely to have their devices on their person. Smart data-driven sensor fusion techniques such as using other cellphone parameters (e.g., number of apps used, calls, and messaging history) can extract signal from noise in mobility data. We also suggest that future studies aiming to capture sensor-based data from smartphones use an application that allows for raw sensor data export. Although raw data requires more effort to engineer new features, we believe it provides an opportunity to explore novel hypotheses such as those relating mood to location, time of day and weather patterns. Case in point—although we used a publicly available app, the mobility feature did not accurately capture physical activity, as speeds above 3.0 mph were not included. Researchers using raw data likely have more scientific oversight to methodological and analytic decisions than those relying on proprietary calculations.

4. Conclusion

Smartphone-based technology holds potential for unobtrusive assessment of depression and other mental and physical health disorders. The ability to capture continuous behavioral patterns renders this technology more sensitive to daily change than would be possible with clinic-based assessments. Such data could detect behavioral states signifying risk or worsening of depression (e.g., social withdrawal, staying at home), and thus improve the identification and measurement-based care of depression. The potential of such technology to increase access to mental and physical health care has not been lost on academia (Siwicki, 2016) or industry (Apple Inc., 2017). Nonetheless, the current example, based on a large remote trial for depression, highlighted issues of featurization, granularity, and data sparsity, which limit our ability to understand fully the meaning of passively collected mobility data in this sample. The effectiveness of this, or any, technology will rely on how seamlessly this information is interpreted and transmitted to healthcare providers and systems. After all, optimal care depends upon accurate and timely detection of problems. Ultimately, the goal of connected mental healthcare will be to motivate and sustain action in accordance with depression treatment or other interventions for health. Thus, the nascence of mHealth holds great promise, but also great challenges.

Funding

This work was supported by the National Institute of Mental Health [grant numbers R34MH100466 and T32MH073553].

Disclosures

Dr. Areán provides consultation to Verily Life Sciences and Akili Interactive. Dr. Atkins is a co-founder with equity stake in a technology company, *Lyssn.io*, focused on tools to support training, supervision, and quality assurance of psychotherapy and counseling. The other authors have no disclosures to report.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.mhpa.2018.04.003>.

References

- Anguera, J. A., Jordan, J. T., Castaneda, D., Gazzaley, A., & Areán, P. A. (2016). Conducting a fully mobile and randomised clinical trial for depression: Access, engagement and expense. *BMJ Innovations*, 2(1), 14–21. <http://dx.doi.org/10.1136/bmjinnov-2015-000098>.
- Apple Inc (2017). *ResearchKit and CareKit*. Retrieved from <https://www.apple.com/researchkit/>.
- Areán, P. A., Hallgren, K. A., Jordan, J. T., Gazzaley, A., Atkins, D. C., Heagerty, P. J., & Anguera, J. A. (2016). The use and effectiveness of mobile apps for depression: Results from a fully remote clinical trial. *Journal of Medical Internet Research*, 18(12), e330. <http://dx.doi.org/10.2196/jmir.6482>.
- Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal*, 38(3), 218–226. <http://dx.doi.org/10.1037/prj0000130>.
- Center for Behavioral Health Statistics and Quality (2016). *2015 National survey on drug use and health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- Cepoiu, M., McCusker, J., Cole, M. G., Sewitch, M., Belzile, E., & Ciampi, A. (2008). Recognition of depression by non-psychiatric physicians: A systematic literature review and meta-analysis. *Journal of General Internal Medicine*, 23(1), 25–36. <http://dx.doi.org/10.1007/s11606-007-0428-5>.
- Gandhi, M., & Wang, T. (2015). *Digital health consumer adoption*. Retrieved from <https://rockhealth.com/reports/digital-health-consumer-adoption-2015/>.
- Karsten, J., Hartman, C. A., Smit, J. H., Zitman, F. G., Beekman, A. T., Cuijpers, P., et al. (2011). Psychiatric history and subthreshold symptoms as predictors of the occurrence of depressive or anxiety disorder within 2 years. *British Journal of Psychiatry*, 198(3), 206–212. <http://dx.doi.org/10.1192/bjp.bp.110.080572>.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical Care*, 41(11), 1284–1292.
- Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, 13, 23–47.
- Pew Research Center (2017). *Mobile fact sheet*. Retrieved from <http://www.pewinternet.org/fact-sheet/mobile/>.
- Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., et al. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research*, 17(7), e175. <http://dx.doi.org/10.2196/jmir.4273>.
- Siwicki, B. (2016). *Duke liberates Epic EHR data with Apple HealthKit and FHIR*. Retrieved from <http://www.healthcareitnews.com/news/duke-liberates-epic-ehr-data-apple-healthkit-and-fhir>.
- Torous, J., & Roberts, L. W. (2017). Needed innovation in digital health and smartphone applications for mental health: Transparency and trust. *JAMA Psychiatry*. <http://dx.doi.org/10.1001/jamapsychiatry.2017.0262>.
- Vallance, J. K., Winkler, E. A., Gardiner, P. A., Healy, G. N., Lynch, B. M., & Owen, N. (2011). Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005–2006). *Preventive Medicine*, 53(4–5), 284–288.
- Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M., & Weidt, S. (2016). Mobile sensing and support for people with depression: A pilot trial in the wild. *JMIR mHealth and uHealth*, 4(3), e111.