



Published in final edited form as:

*Psychiatr Rehabil J.* 2015 September ; 38(3): 218–226. doi:10.1037/prj0000130.

## Next-Generation Psychiatric Assessment: Using Smartphone Sensors to Monitor Behavior and Mental Health

Dror Ben-Zeev, PhD<sup>1</sup>, Emily A. Scherer, PhD<sup>2</sup>, Rui Wang, MS<sup>3</sup>, Haiyi Xie, PhD<sup>1</sup>, and Andrew T. Campbell, PhD<sup>3</sup>

<sup>1</sup>Dartmouth Psychiatric Research Center, Department of Psychiatry, Geisel School of Medicine at Dartmouth

<sup>2</sup>Division of Biostatistics, Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth

<sup>3</sup>Department of Computer Science, Dartmouth College

### Abstract

**Objective**—Optimal mental health care is dependent upon sensitive and early detection of mental health problems. The current study introduces a state-of-the-art method for remote behavioral monitoring that transports assessment out of the clinic and into the environments in which individuals negotiate their daily lives. The objective of this study was examine whether the information captured with multi-modal smartphone sensors can serve as behavioral markers for one’s mental health. We hypothesized that: a) unobtrusively collected smartphone sensor data would be associated with individuals’ daily levels of stress, and b) sensor data would be associated with changes in depression, stress, and subjective loneliness over time.

**Methods**—A total of 47 young adults (age range: 19–30 y.o.) were recruited for the study. Individuals were enrolled as a single cohort and participated in the study over a 10-week period. Participants were provided with smartphones embedded with a range of sensors and software that enabled continuous tracking of their geospatial activity (using GPS and WiFi), kinesthetic activity (using multi-axial accelerometers), sleep duration (modeled using device use data, accelerometer inferences, ambient sound features, and ambient light levels), and time spent proximal to human speech (i.e., speech duration using microphone and speech detection algorithms). Participants completed daily ratings of stress, as well as pre/post measures of depression (Patient Health Questionnaire-9), stress (Perceived Stress Scale), and loneliness (Revised UCLA Loneliness Scale).

**Results**—Mixed-effects linear modeling showed that sensor-derived geospatial activity ( $p < .05$ ), sleep duration ( $p < .05$ ), and variability in geospatial activity ( $p < .05$ ), were associated with daily stress levels. Penalized functional regression showed associations between changes in depression and sensor-derived speech duration ( $p < .05$ ), geospatial activity ( $p < .05$ ), and sleep duration ( $p < .05$ ). Changes in loneliness were associated with sensor-derived kinesthetic activity ( $p < .01$ ).

**Conclusions and implications for practice**—Smartphones can be harnessed as instruments for unobtrusive monitoring of several behavioral indicators of mental health. Creative leveraging of smartphone sensing will create novel opportunities for close-to-invisible psychiatric assessment at a scale and efficiency that far exceed what is currently feasible with existing assessment technologies.

---

## Introduction

Similarly to physical health, mental health is fluid. Psychiatric conditions that are considered “chronic” are not necessarily static; even in the context of severe psychopathology, people will experience periods when their illness is in full or partial remission interspersed between episodes of symptom exacerbation and functional impairment (Goldberg & Harrow, 2011; Strauss et al., 2010). On the other end of the spectrum, people who would largely be considered “healthy” may have periods when, due to internal factors, situational conditions, or a combination of both, they experience mental health difficulties.

Early identification of warning signs that are associated with increased risk for mental health difficulties could facilitate time-sensitive interventions (Calear & Christensen, 2010; Kearns et al., 2012; Stafford et al., 2013). Traditional methods of psychiatric assessment such as clinical interviews and self-report measures are limited in their capacity to accomplish this goal. First, they typically rely on individuals’ retrospective summaries of their experiences over weeks, months, or even years. Consequently, they are susceptible to recall inaccuracies or reporting biases (Ben-Zeev & Young, 2011; de Beurs et al., 1992; Gloster et al., 2008). Second, they take place in treatment center or hospital settings and are administered by mental health professionals-- a clinical context that is substantially different from one’s usual environment. Thus, they have limited ecological validity (Trull & Ebner-Priemer, 2012) and may be prone to response biases that are linked with one’s motivation to receive treatment (e.g. hyper-endorsing problems) or avoid care (e.g. minimizing symptoms). Finally, assessments often take place after mental health problems or functional impairment have already reached a level of severity that warrants clinical attention, and are more difficult to treat. More nuanced detection of behavioral indicators that may be associated with impending problems may lead to earlier interventions, perhaps altering one’s trajectory towards harder to treat conditions down the line (Komatsu et al., 2013; Morriss et al., 2007).

Smartphones (i.e. mobile phones with computational capacities) may offer us unique opportunities to bypass many of the limitations associated with traditional assessment techniques (Proudfoot, 2013; Luxton et al., 2011). Over fifty percent of adults in the United States already own smartphones (Smith, 2013) and there is accumulating evidence that many people with significant mental health conditions are also part of this group (Ben-Zeev et al., 2013a; Carras et al., 2014; Torous et al., in press). Contemporary smartphones have a host of embedded sensors (e.g., light sensor, Global Positioning System [GPS], accelerometers, microphone) that enable their diverse functions (e.g., calling, navigation, gaming). We propose to repurpose these sensors so that in addition to their originally intended functions, they are also leveraged to passively track behavior without response, manipulation, or data entry-burden placed on the user. Thus, while smartphone owners are carrying the device for

personal use, the embedded sensors could potentially collect information that may be pertinent to their mental health.

Research using smartphone sensors in mental health research is in its infancy, and it is unclear whether the types of behavioral data collected with sensors could serve as meaningful indicators of functioning or mental health. To address this question, we developed unique smartphone software that enabled us to collect multi-modal sensor data from individual users over ten weeks. Participants provided self-report stress ratings daily on the device, and completed measures of stress, depression, and loneliness before and after the sensor data collection period. Assuming internal mental states have measureable behavioral implications, we hypothesized that smartphone sensor data would be associated with individuals' daily self-reported stress ratings. We also hypothesized that sensor data would be associated with changes in participants' mental health over the ten week study period.

## Method

### Participants

The study was approved by the Committee for the Protection of Human Subjects at Dartmouth. The study sample consisted of 47 participants (64% undergraduate students and 36% graduate students) recruited through class announcements describing the study. Average age was 22.5 (range: 19–30). These participants were 79% male, 49% Asian, 47% white, and 4% black/African-American. After complete description of the study to participants, written informed consent was obtained.

### Procedures

Individuals were enrolled as a single cohort and participated in the study over a 10-week period in the spring of 2013. Participants completed measures of stress, depression, and loneliness before and after the smartphone data collection period. The study software was developed for smartphones using the Android 4.0 Operating System or higher. These devices have several embedded sensors including microphone, Global Positioning System (GPS), Wireless Fidelity (WiFi) receiver, multi-axial accelerometers, and light sensor. All participants were offered a study smartphone and instructed to carry it with them throughout the data collection period, charging it in the room when they sleep. Sensor data were collected continuously and did not require participant activation. Prompts to complete self-report stress ratings would appear daily on the smartphone touchscreen. Participants were incentivized to adhere to the study protocol at several stages: at week 3 and week 6, five *JawBone Up* wristband sensors were raffled off among the top 15 data collectors. At week 10, Ten *Google Nexus* Smartphones were raffled off among the top 30 data collectors. All participants who completed the study received a t-shirt.

### Measures

**Smartphone Sensing**—Each smartphone was installed with software that enables unobtrusive sensor data collection. Evolving versions of this software were described elsewhere (Berke et al., 2011; Chen et al., 2013; Lane et al., 2011; Lu et al., 2010). The system combines both pre-timed and behaviorally-triggered sensor activation for data

collection (see Wang et al., for technical description and discussion of preliminary findings in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing).

**Speech Duration:** The microphone is activated every 2 minutes to capture ambient sound. If the smartphone detects speech, it remains active for the duration of the conversation. To protect participant privacy, we used a speech detection system that does not record raw audio, but instead, destructively processes the data in real-time to extract and store features that are useful to infer the presence of human speech, but not enough to reconstruct conversation content. This approach has been shown to have 85.3% accuracy in detecting proximal human speech (Lane et al., 2011). Speech Duration was calculated as the total minutes the participant was proximal to human speech daily.

**Geospatial Activity:** The embedded GPS and WiFi receivers were activated every 10 minutes. GPS calculates the participants' location by receiving signals sent by satellites orbiting the earth. The WiFi receiver on the phone scans WiFi access points in its vicinity and records the identification/signal strength for each. The University campus, as well as the entire town center which include all dorm areas, are covered by WiFi. We collected the locations of all WiFi access points in the network and the WiFi scan logs to determine whether a participant was in a specific building. Both GPS and WiFi localization run independently. If GPS-derived location was unavailable, the system used WiFi location as the entry. Once a GPS or WiFi signal was received, the participant's location was geo/time-stamped. The next time the positioning system was activated, it recorded the participant's location. Our software calculated the distance between the previous and current location. Geospatial Activity was calculated as the total distance covered daily. Data were log-transformed due to extremely skewed distribution.

**Kinesthetic Activity:** Multi-axial accelerometers embedded in the smartphone allow the device to detect movement. Our software was developed to identify and classify human activity (e.g. walking, running) based on accelerometer streams (Lane et al., 2011; Lu et al., 2010). The system culls accelerometer data continuously, and an activity rating is generated every 2 seconds. For each 10-minute period, the system calculates the ratio of moving vs. stationary ratings. If the ratio is greater than a threshold of 0.5, the period is labeled "active". Research with an earlier version of this software has shown it is accurate in inferring walking (90.3%), running (98.1%), and being stationary (94.3%) (Lane et al., 2011). Kinesthetic Activity was calculated as the summation of all "active" periods, daily.

**Sleep Duration:** The system exploits smartphone use data (i.e. device "lock" duration), accelerometer inferences (i.e. stationery time), sound features (i.e. ambient silence), and light levels (i.e. ambient darkness) to approximate the amount of time each participant was sleeping. The model is calculated as a linear combination of these four factors. Research using an earlier version of this technique has shown that sleep duration can be inferred within +/- 42 minutes of self-reported sleep duration (Chen et al., 2013). Sleep Duration was calculated as the approximated amount of time each participant slept the previous night.

## Self-report

**Daily Stress:** Individuals were asked to complete self-assessments of momentary stress daily, five days a week, using the smartphone touch screen (Average weekly response rate: 4.92 days a week). EMA was signal-contingent and administered at randomly selected time points during the day. The statement: “Right now, I am...” appeared on the top of screen, and underneath were five response options (1- “Feeling great” to 5- “Stressed out”). Daily Stress scores were calculated as the stress ratings for each individual, daily.

Data from the smartphones (i.e. sensor streams, self-report) were uploaded automatically to a secure study server when the user recharged the device (See Figure 1).

**Mental Health**—Participants completed measures of stress, depression, and loneliness before and after the sensor data collection period.

**Stress:** Participants completed the Perceived Stress Scale (PSS) (Cohen et al., 1983). The 10-item PSS measures how often situations in one’s life were perceived as stressful during the last month. Items were designed to tap how unpredictable, uncontrollable, and overloaded respondents experience their lives. Higher scores suggest greater levels of stress. Among college students, the PSS has been shown to have good internal consistency (coefficient  $\alpha=.85$ ) and test-retest reliability ( $r = .85$ ). The correlation of PSS scores with physical symptoms of stress is .65 (Cohen et al., 1983).

**Depression:** Participants completed the nine item depression scale from the Patient Health Questionnaire (PHQ) (Spitzer, Kroenke, & Williams 1999). The PHQ-9 is a widely-used self-report measure of depressed symptomology over the last two weeks (Kroenke, Spitzer, Williams, 2001). Higher scores suggest greater severity. The PHQ-9 has good internal consistency ( $\alpha=.89$ ) and test re-test reliability ( $r = .84$ ). The correlation of PHQ-9 scores with a measure of mental health that is associated with depression is .73 (Kroenke, Spitzer, Williams, 2001).

**Loneliness:** Participants completed the Revised UCLA Loneliness Scale (Russel et al., 1980). This 20-item self-rated instrument measures one’s feelings of loneliness and social isolation. Participants are asked to indicate how often they feel the way described in a series of statements (e.g. “There is no one I can turn to”, “I feel isolated from others”). Higher scores indicate greater subjective feelings of loneliness. The measure has high internal consistency ( $\alpha=.94$ ), and scores have been shown to be significantly associated with the amount of time individuals are alone each day ( $r =.41$ ), the number of close friends they have ( $r = -.44$ ), and self-labeled loneliness ( $r =.70$ ) (Russel et al., 1980).

## Overview of Analyses

The objective of the first analysis was to model the relationship between Daily Stress (outcome) and several covariates derived from smartphone sensing— Geospatial Activity, Kinesthetic Activity, Speech Duration, and Sleep Duration. Data were aggregated daily and resulted in a maximum of 65 possible data points. We conducted mixed-effects linear modeling to account for the repeated measures design, the time-varying nature of the

variables, and missing data. Time was coded as 0,1,2...N (with each point representing a day), and treated as continuous as we fitted parametric curves in the model. For the analyses, time was centered (i.e., the intercept was shifted to the middle).

The second analysis was conducted to evaluate the relationship between sensor data, and pre/post changes in participants' mental health. In this analysis, we are also able to find out during which period of time in the study the sensor data is most associated with individual-level changes in outcome. To do so, we used penalized functional regression (Goldsmith, Bobb, Crainiceanu, Caffo, Reich, 2011). This multi-step technique can use intensive repeated-measure variables as predictors and relate them, as a whole, to individual-level outcome measures. First, a smooth non-parametric function is fit to each individual's sensor variable using principle component basis splines. Thus, sensor data is described as a function rather than a series of data points and does not require that each participant is measured at the same time points. Next, a set of functional regression coefficients is fit to this functional representation of sensor data (via a truncated power series spline) to predict changes in PSS, PHQ-9, and UCLA Loneliness Scale scores. This is akin to estimating a set of weights for each section of the predictor function in relation to the outcome. Finally, a permutation test evaluated the significance of the set of functional regression parameters in predicting the individual outcome. The model was fit repeatedly, each time with different random pairing of each outcome value with a set of functional predictor values, and an F-statistic is calculated for the model. The proportion of times an F-statistic more extreme than the one obtained from the original dataset was obtained gives the permutation p-value which is used to quantify the likelihood that the observed result was obtained by chance if there was no association between the outcome and the functional predictor. Model fit and permutation tests were performed via the *funreg* package in R (Dziak & Shiyko, 2014) which makes the Goldsmith et al. technique available for intensively collected data. The benefit of penalized functional regression in the sensor data context is the ability to use all available sensor data rather than summarizing (e.g. via averaging across time points) and the ability to estimate a flexible, time-varying relationship between the sensor variable and the outcome. In this way, penalized regression allows differential weighting of observations in the study based on which time periods tend to be most related to the outcome. Thus, in addition to an indication of whether or not a sensor variable was related to an outcome, we were able to obtain an estimate (via regression coefficients) of which period of time in the study is most associated with individual-level change in outcome.

## Results

Of 37 participants who completed PHQ-9 ratings at the end of the data collection period, 11% endorsed no depressive symptoms, 38% reported minimal depression, 32% reported mild depression, 8% reported moderate depression, 6% moderately severe depression, and 5% reported severe depression. Averaged EMA daily stress ratings were significantly correlated with scores on a valid measure of stress (PSS) completed at the end of the study ( $r = 0.41, p < 0.01$ ).

### Relationship between Smartphone Sensor Data and Average Daily Stress

Three mixed-effects models were fitted to the data: one with original covariates, a second with both mean scores (average score for each individual across days) and deviation scores (difference between each individual's daily score and his/her average) of the time-varying covariates, and a third model only with deviation scores. For models with time-varying covariates, statisticians recommend evaluating both between-subject and within-subject effects, unless both types of effects were similar (model fit would not be improved) (Hedeker & Gibbons, 2006). Our results indicated that the model with mean scores and deviation scores of covariates resulted in poorer fit, so it was dropped. The model with deviation scores had better fit, and we were also interested in understanding the relationship between "deviation" (variability) of the covariates and outcome. We report on the results from these two models in Table 1.

For both models, Speech Duration, Geospatial Activity, Kinesthetic Activity, and Sleep Duration were included as covariates. "Time" was also included to capture the natural evolution of Daily Stress over the data collection period. Daily Stress levels showed a curvilinear pattern, so both linear time and quadratic time components were included in the model. Random variation terms (i.e. random intercept and random slope of time) were included to capture both between-individual difference and variations in change in Daily Stress. The results of Model 1 indicate that both Geospatial Activity (Estimate:  $-0.09$ , SE:  $0.04$ ,  $p=0.0278$ ) and Sleep Duration (Estimate:  $-0.03$ , SE:  $0.02$ ,  $p=0.0493$ ) were inversely associated with Daily Stress. The results of Model 2 show inverse relationships with Daily Stress: deviation from one's average Geospatial Activity was significantly negatively associated with Daily Stress (Estimate:  $-0.08$ , SE:  $0.04$ ,  $p=0.048$ ), and deviation from one's average Speech Duration had marginally significant negative association with Daily Stress (Estimate:  $-0.03$ , SE:  $0.02$ ,  $p=.0873$ ).

In both models, quadratic time components were significant, which indicates that as time progressed, Daily Stress increased but in a decelerating rate, or the trend gradually reversed itself. This may be linked with students gradually growing accustomed to academic pressures as the term progresses, experiencing them as less subjectively stressful over time. Intercept variance and slope variance for linear time component were significant, and marginally significant, respectively, in both models. This indicates that participants varied on average level of stress and on linear change in stress. Stress ratings and sensor-derived data as a function of time are presented in Figure 2.

### Relationship between Smartphone Sensor Data and Pre/Post Changes in Mental Health

Significant pre-post changes were seen in mental health measures PHQ-9, PSS, and UCLA Loneliness Scale ( $p<0.0001$ ,  $p=0.003$ , and  $p=0.02$ , respectively) when evaluated with a Wilcoxon signed rank test. Change scores were calculated as the post-measurement minus the pre-measurement. A higher value for each of the measures represents poorer mental health. Therefore, the smaller the value of the change score, the better one's mental health status; a highly negative change score is an indicator of greater improvement in mental health. For each measure, the median change scores were  $1.0$ ,  $0.5$ , and  $-0.5$ , respectively, indicating overall worsening of depression (PHQ-9) and stress (PSS), and improvement in

subjective loneliness (UCLA Loneliness Scale) over the study period. Several sensor variables were found to be significantly associated with changes in mental health scores. In the penalized functional regression analysis, speech duration was significantly associated with changes in PHQ-9 over the course of the study ( $p=0.048$ ). The regression coefficients indicate that earlier in the study (approximately days 10–30), increased Speech Duration was associated with better PHQ-9 change scores (smaller valued), while in the later days of the study (approximately days for 40–58) the direction of the association changed such that increased Speech Duration was associated with worse PHQ-9 change scores (larger valued). A significant relationship was also found between Geospatial Activity and changes in PHQ-9 ( $p=0.022$ ). Specifically, earlier in the study (approximately days 15–20) increased Geospatial Activity was associated with better PHQ-9 change scores. Finally, sleep duration was also significantly associated with PHQ-9 ( $p=0.028$ ) with longer sleep duration experienced during the early days of the study (days < 45) predictive of worse PHQ-9 change scores. In contrast, during the later days of the study (days > 45), increased Sleep Durations were predictive of better PHQ-9 change scores. No significant association was found between Kinesthetic Activity and PHQ-9 change scores. Kinesthetic Activity was significantly related to UCLA Loneliness Scale change scores ( $p=0.002$ ) with higher values later in the study (approximately days > 50) predictive of better UCLA Loneliness Scale change scores. No significant associations were found between UCLA Loneliness Scale change scores and Sleep Duration, Geospatial Activity, or Speech Duration. No significant associations were found between change in PSS scores and any sensor variables. Each of the significant relationships between sensor data and mental health change scores are illustrated with the plots of the functional regression coefficient plotted along with 95% confidence intervals (Figure 3. panels a-d), where positive regression coefficients indicate a prediction of worse mental health change scores and negative regression coefficients indicate a relationship to better mental health change scores. Confidence intervals indicate periods of time where the functional regression coefficient is significantly different from 0, while the overall permutation test p-value indicates the significance of the relationship between the functional regression coefficient across the entire time-period and the value of the change score. Insert Figure 3 here

## Discussion

This proof-of-concept study showed that smartphones can be used as instruments for unobtrusive collection of behavioral markers that are associated with daily stress and changes in mental health. Sensor-derived Geospatial Activity, Sleep Duration, and variability in Geospatial Activity were associated with self-reported daily stress levels. Variability in the amount of time people spent proximal to human speech was marginally associated with daily stress. Speech Duration, Geospatial Activity, and Sleep Duration were associated with changes in participants' depression levels over the course of the study. Kinesthetic activity was associated with changes in their subjective loneliness. Importantly, sensor data were collected passively and processed automatically on the device, with minimal user burden.

Previous research has already identified relationships between activity, sleep, social context, and mental health (e.g., Asmundson et al., 2013; Hawley & Cacioppo, 2010; Kyung Lee &



Douglass 2010; Mammen & Faulkner, 2013). The innovation of the current study lies in its demonstration that the digital traces of these behaviors, as captured with smartphone sensors and software, may serve as indicators of one's daily stress and changes in mental health status over time. Of particular significance is that this can be accomplished passively, with minimal burden to individuals (i.e. no need to engage in periodic interviews, complete self-reports, or set foot in a clinic). Future research should examine whether sensor data captured by smartphones can be used to predict impending problems in clinical populations (e.g., relapse in people with schizophrenia, manic episodes in people with bipolar disorder).

This study has several limitations. Participants were all younger adults, and may be less representative of the broader population in terms of their demographic characteristics, education, level of functioning, or willingness to engage in smartphone monitoring. However, the participants were not monolithic in terms of their behaviors and mental health status; some were very active while others were not, some reported no mental health difficulties while others endorsed severe symptoms of depression. Future research should examine feasibility in older individuals and people with identified mental health conditions. The study was conducted at a university campus, a setting that might be different from other environments (e.g. more wireless access points, structured social gatherings). Additional research is needed to examine the feasibility of continuous smartphone data collection in diverse environments. The smartphone speech detection system may not be fully capable of differentiating live human speech from radio or television-generated audio, making it difficult to determine with complete accuracy whether a participant was actually in a social context. Our modelling strategy assumed that sleep typically takes place in darkened environments, as indicated by the smartphone light sensor. However, this approach does not account for naps throughout the day, or individuals who prefer to sleep with the light on. In future research, it would be advantageous to have participants describe their typical sleep environment in order to calibrate more individualized models (e.g., should ambient light influence the sleep duration estimate or not).

### **Conclusions and Implications for Practice**

Mobile Health (mHealth) research using sophisticated technologies in uncontrolled environments entails a host of conceptual and methodological challenges (Ben-Zeev et al., 2014a), but the potential benefits seem to far outweigh the costs (Brian & Ben-Zeev, 2014; Luxton et al., 2011; Proudfoot, 2013). As technology evolves, it is exciting to envision a future in which individuals who could benefit from additional support have personalized sensor-enabled data collection systems that are installed on their smartphones and calibrated to their individual needs. Different people will likely have different behavioral indicators of mental health difficulties (e.g. inactivity/frenetic activity, insomnia/hypersomnia). Data captured with smartphone sensing techniques may be used to trigger time-sensitive follow-up including feedback reports processed and displayed on the device, or mobile health (mHealth) interventions that target problem areas (e.g. irregular sleep patterns) delivered directly to the individual (e.g., Ben-Zeev et al., 2014b; Burns et al., 2011). Moreover, smartphone users who are receiving mental health services may elect to authorize transmission of sensor summary reports seamlessly to their mental health providers to inform ongoing care.

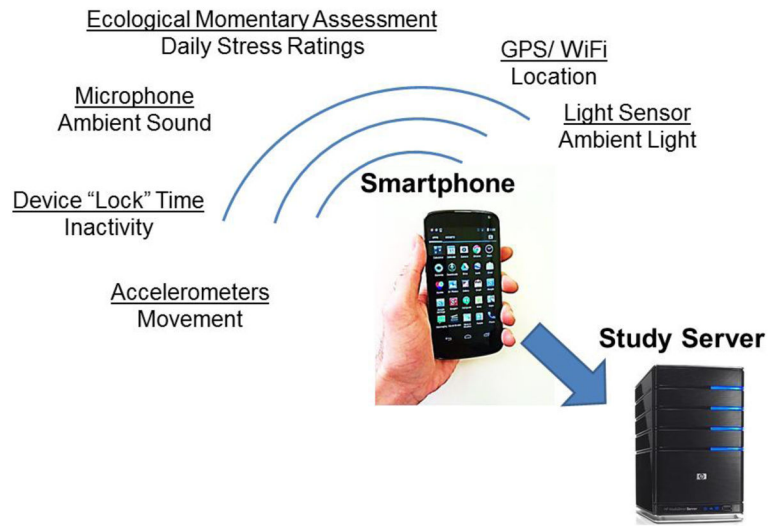
Smartphone technology and telecommunication infrastructure is spreading globally. While the technological sophistication and capabilities of smartphones continue to increase, the costs of devices and data plans continue to drop. We demonstrated that these widely-available instruments can be successfully harnessed to capture behavioral information that is relevant to changes in mental health and functioning. In the future, innovative interdisciplinary research combining behavioral, computational, and engineering sciences will create smartphone sensing techniques that will enable close to invisible mental health monitoring at a scale and efficiency that far exceed what is currently feasible with existing assessment technology. Passive smartphone sensing may be insufficient to make definitive predictions regarding one's mental health status without consideration of a host of additional variables (e.g., environmental, intrapsychic, contextual) that cannot be captured by phone sensors alone. Integration of other digital sources of information such as personal social media posts and crowd-sourced information about one's immediate environment into predictive models may also strengthen the role of assistive technology in quantified self-tracking (Swan, 2009).

## References

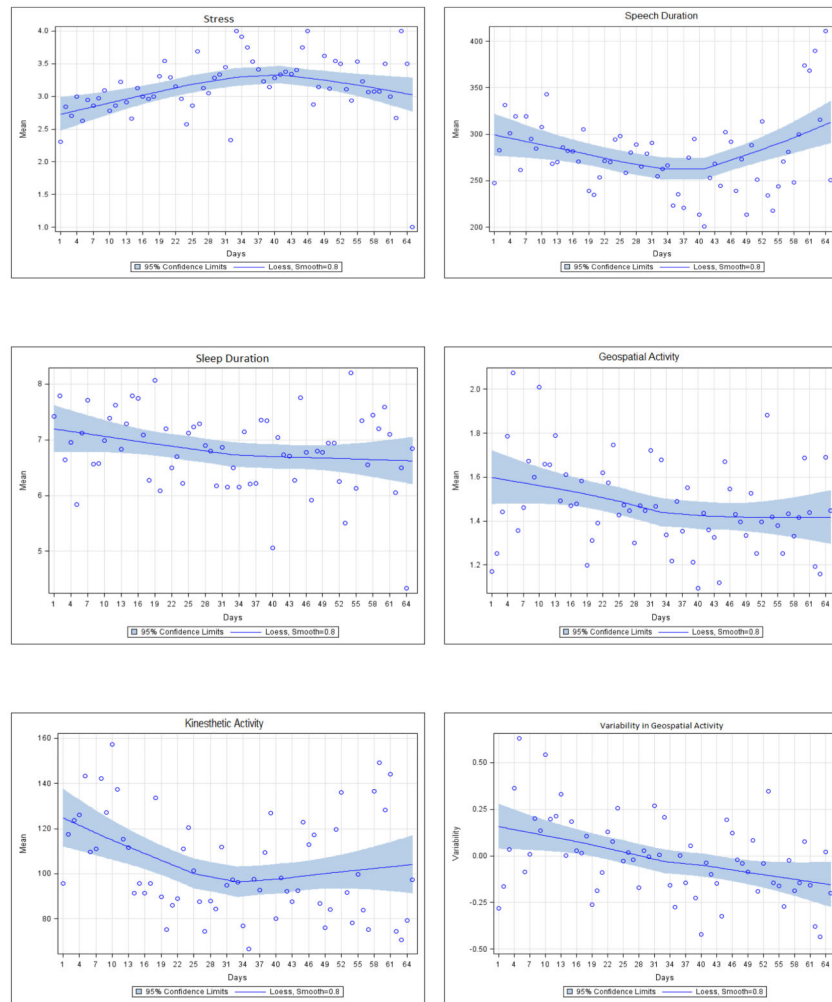
- aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*. 2012; 32:510–523. [PubMed: 22721999]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5. Arlington, VA: American Psychiatric Publishing; 2013.
- Asmundson GJ, Fetzner MG, Deboer LB, Powers MB, Otto MW, Smits JA. Let's get physical: a contemporary review of the anxiolytic effects of exercise for anxiety and its disorders. *Depression and Anxiety*. 2013; 30:362–373. [PubMed: 23300122]
- Brian RM, Ben-Zeev D. Mobile Health (mHealth) for Mental Health in Asia: Objectives, Strategies, and Limitations. *Asian Journal of Psychiatry*. 2014; 10:96–100. [PubMed: 25042960]
- Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophrenia Bulletin*. 2014b; 40 (6):1244–1253. [PubMed: 24609454]
- Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: Opportunities for future services. *Administration and Policy in Mental Health and Mental Health Services Research*. 2013; 40:340–343. [PubMed: 22648635]
- Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. *Administration and Policy in Mental Health and Mental Health Services Research*. 2014a 10.1007/s10488-014-0556-2.
- Ben-Zeev D, Young MA. Accuracy of hospitalized depressed patients' and healthy controls' retrospective symptom reports: an experience sampling study. *Journal of Nervous and Mental Disease*. 2010; 198:280–285. [PubMed: 20386257]
- Berke EM, Choudhury T, Ali S, Rabbi M. Objective measurement of sociability and activity: mobile sensing in the community. *Annals of Family Medicine*. 2011; 9:344–350. [PubMed: 21747106]
- Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, Mohr DC. Harnessing context sensing to develop a mobile intervention for depression. *Journal of Medical Internet Research*. 2011; 13:e55. [PubMed: 21840837]
- Calear AL, Christensen H. Systematic review of school-based prevention and early intervention programs for depression. *Journal of Adolescence*. 2010; 33:429–438. [PubMed: 19647310]
- Carras MC, Mojtabei R, Furr-Holden CD, Eaton W, Cullen B. Use of mobile phones, computers and internet among clients of an inner-city community psychiatric clinic. *Journal of Psychiatric Practice*. 2014; 20(2):94–103. [PubMed: 24638044]

- Chen, Z.; Lin, M.; Chen, F.; Lane, MD.; Cardone, G.; Wang, R.; Li, T.; Chen, Y.; Choudhury, T.; Campbell, AT. Unobtrusive sleep monitoring using Smartphones. 7th International ICST Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health); Venice. 2013.
- Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of Health and Social Behavior*. 1983; 24:386–396.
- de Beurs E, Lange A, Van Dyck R. Self-monitoring of panic attacks and retrospective estimates of panic: discordant findings. *Behaviour Research and Therapy*. 1992; 30:411–413. [PubMed: 1352100]
- Dziak, J.; Shiyko, M. funreg: Functional Regression for Irregularly Timed Data (R package). 2014. <http://cran.r-project.org/web/packages/funreg/funreg.pdf>
- Germain A. Sleep disturbances as the hallmark of PTSD: where are we now? *American Journal of Psychiatry*. 2013; 170:372–382. [PubMed: 23223954]
- Gloster AT, Richard DC, Himle J, et al. Accuracy of retrospective memory and covariation estimation in patients with obsessive-compulsive disorder. *Behaviour Research and Therapy*. 2008; 46:642–655. [PubMed: 18417100]
- Golberg JF, Harrow MA. 15-year prospective follow-up of bipolar affective disorders: comparisons with unipolar nonpsychotic depression. *Bipolar Disorders*. 2011; 13:155–163. [PubMed: 21443569]
- Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized Functional Regression. *Journal of Computational and Graphical Statistics*. 2011; 20:830–851. [PubMed: 22368438]
- Hawkey LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*. 2010; 40:218–227. [PubMed: 20652462]
- Hedeker, D.; Gibbons, R. Longitudinal data analysis. NJ: Wiley-Interscience; 2006.
- Hidaka BH. Depression as a disease of modernity: explanations for increasing prevalence. *Journal of Affective Disorders*. 2012; 140:205–214. [PubMed: 22244375]
- Kearns MC, Ressler KJ, Zatzick D, Rothbaum BO. Early interventions for PTSD: a review. *Depression and Anxiety*. 2012; 29:833–842. [PubMed: 22941845]
- Komatsu H, Sekine Y, Okamura N, Kanahara N, Okita K, Matsubara S, Hirata T, Komiyama T, Watanabe H, Minabe Y, Iyo M. Effectiveness of Information Technology Aided Relapse Prevention Programme in Schizophrenia excluding the effect of user adherence: a randomized controlled trial. *Schizophrenia Research*. 2013; 150:240–244. [PubMed: 23998952]
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*. 2001; 16:606–613. [PubMed: 11556941]
- Kyung Lee E, Douglass AB. Sleep in psychiatric disorders: where are we now? *Canadian Journal of Psychiatry*. 2010; 55:403–412.
- Lane, N.; Choudhury, T.; Campbell, AT.; Mohammad, M.; Lin, M.; Yang, X.; Doryab, A.; Lu, H.; Ali, S.; Berke, E. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. 5th International ICST Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health); Dublin. 2011.
- Lu, H.; Yang, J.; Liu, Z.; Lane, DL.; Choudhury, T.; Campbell, AT. The Jigsaw continuous sensing engine for mobile phone applications. Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems; New York: ACM; 2010.
- Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: Integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice*. 2011; 42:505–512.
- Morriss RK, Faizal MA, Jones AP, Williamson PR, Bolton C, McCarthy JP. Interventions for helping people recognise early signs of recurrence in bipolar disorder. *Cochrane Database of Systematic Reviews*. 2007; 1:CD004854. [PubMed: 17253526]
- Proudfoot J. The future is in our hands: The role of mobile phones in the prevention and management of mental disorders. *Australian & New Zealand Journal of Psychiatry*. 2013; 47:111–113. [PubMed: 23382507]
- Proudfoot J, Clarke J, Birch MR, Whitton AE, Parker G, Manicavasagar V, Harrison V, Christensen H, Hadzi-Pavlovic D. Impact of a mobile phone and web program on symptom and functional

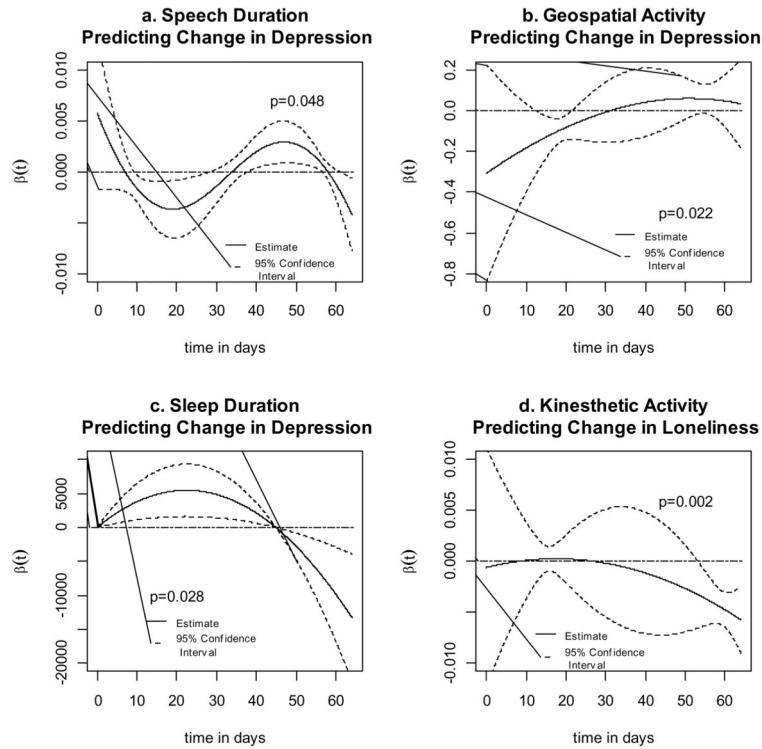
- outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. *BMC Psychiatry*. 2013; 13:312. [PubMed: 24237617]
- Russell D, Peplau LA, Cutrona CE. The Revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*. 1980; 39:472–480. [PubMed: 7431205]
- Shiffman S, Stone AA, Hufford MR. Ecological Momentary Assessment. *Annual Review in Clinical Psychology*. 2008; 4:1–32.
- Smith, A. Pew internet and American life project: Americans and their cell phones. 2011. Retrieved August 15, 2014 from: <http://www.pewinternet.org/2011/08/15/americans-and-their-cell-phones/>
- Smith, A. Pew internet and American life project: Smartphone Ownership 2013. 2013. Retrieved December 10, 2013 from: [http://pewinternet.org/~media/Files/Reports/2013/PIP\\_Smartphone\\_adoption\\_2013\\_PDF.pdf](http://pewinternet.org/~media/Files/Reports/2013/PIP_Smartphone_adoption_2013_PDF.pdf)
- So SH, Peters ER, Swendsen J, Garety PA, Kapur S. Detecting improvements in acute psychotic symptoms using experience sampling methodology. *Psychiatry Research*. 2013; 210:82–88. [PubMed: 23849758]
- Spitzer R, Kroenke K, Williams JB. Validation and utility of a self-report Version of PRIME-MD: The PHQ Primary Care Study. *JAMA*. 1999; 282:1737–1744. [PubMed: 10568646]
- Stafford MR, Jackson H, Mayo-Wilson E, Morrison AP, Kendall T. Early interventions to prevent psychosis: systematic review and meta-analysis. *BMJ*. 2013; 346:f185. [PubMed: 23335473]
- Strauss GP, Harrow M, Grossman LS, Rosen C. Periods of recovery in deficit syndrome schizophrenia: A 20-year multi-follow-up longitudinal study. *Schizophrenia Bulletin*. 2010; 36:788–799. [PubMed: 19095758]
- Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *International journal of environmental research and public health*. 2009; 6:492–525. [PubMed: 19440396]
- Torous J, Chan S, Tan S, Behrens J, Mathew I, Hinton L, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions: A survey study of four geographically distinct clinics. *Journal of Medical Internet Research*. in press.
- Trull TJ, Ebner-Priemer U. Ambulatory assessment. *Annual Review of Clinical Psychology*. 2013; 9:151–176.
- Wang, R.; Chen, F.; Chen, Z.; Li, T.; Harari, G.; Tignor, S.; Zhou, X.; Ben-Zeev, D.; Campbell, AT. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones. *Proceedings of 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014)*; Seattle. 2014.



**Figure 1.**  
Overview of Smartphone Data Collection



**Figure 2.** Self-Reported Stress Ratings and Smartphone Sensor Behavioral Data as a Function of Time in Study (N=47).

**Figure 3.**

Estimated functional coefficients and 95% confidence intervals from penalized functional regression models. Positive regression coefficients indicate a positive relationship between the sensor predictor variable and the individual-level outcome change score and negative regression coefficients indicate an inverse relationship. Confidence intervals indicate periods of time where the functional regression coefficient is significantly different from 0, while the overall permutation test p-value indicates the significance of the relationship between the functional regression coefficient across the entire time-period and the value of the change score.

**Table 1**

Mixed-Effects Linear Models for the Relationship Between Sensor Data and Daily Stress (N=47)

<b>Model 1: Predictors in original measurement scale</b>			
<b>Covariate</b>	<b>Estimate</b>	<b>Standard error</b>	<b>p-value</b>
<b>Fixed effects</b>			
Intercept	3.93230	0.19830	<.0001
Time (linear)	0.00218	0.00358	0.5470
Time <sup>2</sup> (quadratic)	-0.00071	0.00014	<.0001
Kinesthetic Activity	-0.00031	0.00061	0.6103
Speech Duration	-0.00041	0.00027	0.1260
Sleep Duration	-0.03721	0.01889	0.0493
Geospatial Activity	-0.08954	0.04060	0.0278
<b>Random effects</b>			
Intercept variance	0.29870	0.09239	0.0006
Time (linear) Variance	0.00012	0.00008	0.0701
<b>Model fit statistics</b>			
-2LL = 1490.0, AIC = 1498.0 & BIC = 1505.3			
<b>Model 2: Predictors as deviation scores from individual means</b>			
<b>Covariate</b>	<b>Estimate</b>	<b>Standard error</b>	<b>p-value</b>
<b>Fixed effects</b>			
Intercept	3.38300	0.10790	<.0001
Time (linear)	0.00203	0.00360	0.5769
Time <sup>2</sup> (quadratic)	-0.00070	0.00014	<.0001
Kinesthetic Activity	-0.00062	0.00063	0.3239
Speech Duration	-0.00047	0.00028	0.0873
Sleep Duration	-0.03152	0.01931	0.1033
Geospatial Activity	-0.08237	0.04155	0.0480
<b>Random Effects</b>			
Intercept variance	0.30570	0.09151	0.0004
Time (linear) Variance	0.00012	0.00008	0.0666
<b>Model fit statistics</b>			
-2LL = 1489.8.0, AIC = 1497.8 & BIC = 1505.1			