Graduate Student Journal of Psychology
2018, Vol. 17

Remote Behavioral Sampling for Psychological Assessment: Using Interactive Technologies to Detect Depression

Jacob Edward Thomas, M.A. and Randall Richardson-Vejlgaard, Ph.D.
Teachers College, Columbia University

The US National Comorbidity Survey indicates that up to 47% of Major Depressive Disorder (MDD) patients receive treatment predominately in a primary care setting. The screening methods primary care physicians use to monitor treatment response offer low time-sensitivity, and thus hinder the provider's ability to adjust or augment treatment effectively. Contemporary technologies that directly capture behavioral data may offer objective insight into an individual's mental health status, with little burden. The current study tracks 12 participants over two weeks to evaluate the convergent validity of a clinician administered measure of depression with physical activity and sleep measured by an actigraph, linguistic style present in Facebook content, and repeated brief assessments of mood delivered via a smartphone. Feasibility analysis showed that 91.7% of participants completed the study, with 83.6% of targeted data captured. Correlations were found between depression symptoms and physical activity (r = -.501, p = .058), restful sleep (r = -.536, p = .045), ecologically assessed affect (r = -.582, p = .03), and 3 linguistic styles present in social media content ('future focused' (r = -.539, p = .043), 'feeling' (r = .532, p = .046), and 'masculine' (r = .590, p = .028)). The study suggests that data collected from interactive technologies offers an easily accessible inventory of patient's behaviors which is highly correlated with traditional MDD assessment. Depressed patients who are not seeing a specialist regularly are at high risk for nonadherence with medication during the latency period between antidepressant initiation and MDD remission. A monitoring platform utilizing interactive technologies would be useful in identifying patients who need treatment.

Introduction

Currently, 10% of the American population is clinically depressed, and nearly 20% will experience a depressive episode at some point in their life (Kessler, Berglund, Demler, et al., 2005). Epidemiologists estimate that Major Depressive Disorder (MDD) will be the leading cause of life years lost due to ill-health, disability, or early death by 2030 (Mathers, Fat, & Boerma, 2008). This disorder is debilitating and recurrent. Longitudinal economic analyses indicate that being depressed leads to immense individual burden, including diminished educational accomplishments, annual family income reduction of 20%, seven fewer weeks worked per year, and an 11% decrease in probability of getting married (Smith, & Smith, 2010). Cumulatively, depression is estimated to cause an $83 billion annual American economic loss (Greenberg, Kessler, Birnbaum, et al., 2003); almost triple what the National Institutes of Health (NIH) invests in all medical research each year (NIH, 2015). As evidenced by these staggering numbers, depression is one of the most prevalent and destructive psychological forces in present society.

Up to 47% of MDD patients receive treatment predominately in a primary care setting (Kessler, Berglund, Demler, et al., 2003). Healthcare providers often monitor antidepressant treatment response using self-report questionnaires such as the Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996) or the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001). These screening tools are only validated to be sensitive to changes over two-week intervals (Lowe, Kroenke, Herzog, & Grafe, 2004; Beck, et al., 1996; Kroenke, et al., 2001). With an average onset of antidepressant action in pharmacological intervention of 20 days (Stassen, Delini-Stula, & Angst, 1993), systematically administered self-report screeners therefore fail to identify many patients' antidepressant response trajectories until day 28 of treatment. Clinical improvement during the first month is one of

Please address correspondence to: jet2173@tc.columbia.edu

the strongest predictors of long-term stability in MDD, and mounting evidence suggests that dosage modification and treatment augmentation are key in achieving this (Trivedi, Fava, Wisniewski, et al., 2006). In order to more effectively modulate depressive symptoms during this critical window, primary care providers are in need of a novel and more time-sensitive screening method that allows them to identify patients refractory to treatment for earlier outreach (see Figure 1).

The Present Study

Recent technological advances have ushered in a new era of data availability and access. Americans are densely interconnected with online social networks and carry a powerful computer capable of internet access in their pockets. Polls indicate that 65% of American adults use social networking sites (Pew Research Center, 2015A), and nearly 70% own a smartphone (Pew Research Center, 2015B). Wearable devices, such as fitness trackers and smart watches, provide users with real-time behavioral data. Ten million of such devices are sold annually, and this industry is trending rapidly (Wei, 2014). Hidden within the way we interact with our technology is a map of our lives. Our watches track the way we move, our phones geo-tag our locations, our cameras record what we see, and our online habits offer an abstract narrative of our thoughts. All of this data is recorded, stored, time stamped, and easily transmitted.

The present study seeks to use data captured by interactive technologies to make inferences about depression symptoms. First, discrete behavioral domains that are passively monitored by existing technology are identified, and then the evidence linking these domains to depression is reviewed. Next, a platform that integrates these domains into a data collection system intended to remotely monitor the presence of depression-related behaviors is deployed in a two-week study of 12 participants to test (1) the platform's feasibility for implementation, and (2) the convergent validity of data generated in each domain with a clinician administered assessment of MDD. In closing, discussion is presented on how the platform can impact antidepressant treatment response monitoring in primary care, as well as limitations to the present study, and directions for future research.

Measurable Behavioral Domains and Their Association with Depression

Population-based studies assessing physical activity levels indicate that people who are more active are less likely to be depressed (Stathopoulou, Powers, Berry, Smits, & Otto, 2006). Meta-analyses of intervention studies using exercise as a treatment for depression show moderate to large effect sizes consistent across multiple participant variables and exercise types (Cooney, Dawn, Greg, et al., 2013). Furthermore, MDD patients

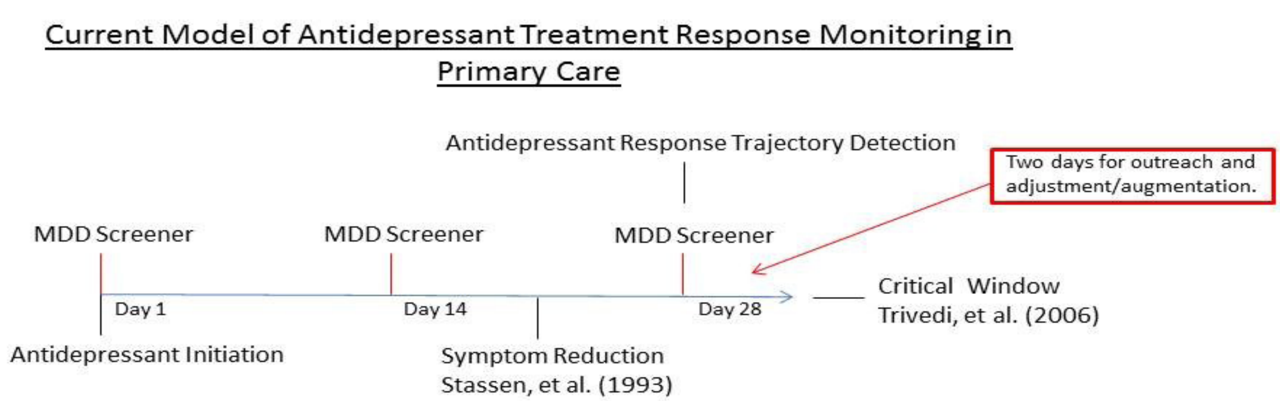

Figure 1. The Current Model of Antidepressant Treatment Response Monitoring in Primary Care.

often report increases in desire and ability to exercise as symptoms remit (Otto, Church, Craft, et al, 2007). Actigraphs present in wrist-worn technology provide accurate reports of physical activity, often reporting step counts and distance traversed (Tudor-Locke, Williams, Reis, Pluto, 2002). Discreet by nature, these devices automatically pair with user's smartphone to upload data to a target server, which is accessible remotely.

Up to 90% of depressed individuals report sleep disturbances, commonly including insomnia, frequent nighttime wakefulness, hypersomnia, and recurrent daily fatigue despite sleeping throughout the night (Franzen, & Buysse, 2008). A good indicator of MDD symptom improvement is a reduction in these issues (Keller, 2003). Wrist-worn actigraphs provide accurate assessments of sleep duration and quality, with outputs highly correlated to metrics of traditional laboratory sleep studies (Ferguson, Rowlands, Olds, & Maher, 2015; Slater, Botsis, Walsh, et al., 2015). Similar to physical activity monitoring, sleep data is passively collected and remotely accessible after syncing with a smartphone.

As social media usage has grown in recent years, social scientists have utilized the forum to generate new findings. Researchers at Stanford University have shown that status updates accurately reflect dimensions of life satisfaction (Lui, Tov, Kosinski, Stillwell, & Qui, 2015). An Italian team has demonstrated that emotionally positive or negative linguistic styles in Facebook content is closely related to emotional well-being (Settanni, & Marengo, 2015). Positive psychosocial dimensions such as these have a long history in depression research, universally having strong discriminant validity with depression (Pavot, & Diener, 1993). Users generate their social media content naturalistically, curating it for remote access on their social media profile.

The majority of screening and diagnostic traditions conceptualize depression as the persistent experience of symptoms (e.g., American Psychiatric Association, 2013). A proposed way to record the experience of symptoms as they occur is by repeatedly sampling subjects' current behaviors and experiences in real time, in their natural environment—a strategy known as Ecological Momentary Assessment (EMA; Shiffman, Stone, & Hufford, 2008). EMA mitigates recall biases, which greatly affect the accuracy of retrospective reports (Porta, Greenland, Hernan, dos Santon, & Last, 2014; Hunt, Auriemma, & Cashaw, 2003). With the exponential growth of smartphone use, EMA becomes highly feasible, as assessments can be sent to subjects' cell phones instantaneously. If the assessment is designed in an extremely brief and intuitive manner, capturing this data can become an unintrusive part of one's day, similar to sliding a finger across a screen to unlock a smartphone or replying to a short text message.

The Proposed Platform & Research Objectives

Physical activity, sleep, social media interaction, and moment-to-moment experiences are all intricately linked to depression. These data are captured with minimal burden using existing technologies. We propose a data collection system that combines physical activity and sleep variables from an actigraph, linguistic style present in Facebook posts, and repeated brief assessments of mood delivered via a smartphone into a single, multimodal diagnostic tool. The study that follows compares outputs from the platform to a clinician administered MDD assessment.

We have two primary aims. First, to assess the feasibility of implementing the platform by monitoring adherence and attrition. Adherence reports will be based on percentage of targeted data collected: daily outputs of physical activity, nightly outputs of sleep duration, presence of Facebook content, and number of responses to EMA. Second, to test the convergent validity of each behavioral output generated by the platform with MDD symptom ratings using bivariate Pearson Correlation tests.

We hypothesize that lower physical activity levels, less sleep, and lower EMA reported mood will be associated with higher MDD symptom rating. Additionally, we hypothesize that by mining Facebook content, significant correlations between the content's linguistic styles and MDD symptom rating can be extrapolated.

Methods

Participants

Participants (N = 12) were four men (mean age = 29.5; range = 24–34) and eight women (mean age = 24.4; range = 22–28) who were recruited through advertisements posted online and on the Teachers College, Columbia University campus. Nine of the participants were students, four were professionals.

Of the 12 participants, one was lost to follow-up, and thus, excluded from the final analysis.

Inclusion criteria were as follows: (1) men and women aged 18–65; (2) has a Facebook profile and is a self-reported "active" user, defined as making at least 1–2 interactions every few days; (3) possesses a smartphone that is activated on a cellular network; (4) willing and able to provide informed consent; (5) willing and able to complete a follow-up assessment two weeks after consent.

Exclusion criteria were as follows: (1) substantial physical injury or disability that prevents the participant from exercising should they wish to; (2) English illiteracy; (3) prior diagnosis of mental retardation or acute psychosis.

Materials

Actigraph. Participants wore a Misfit Flash™ (by Misfit Inc., Burlingame, CA) for the duration of the study. This device utilizes an accelerometer to continuously monitor wrist movement. A six-month battery life and an open-source application program interface make this device ideal for use in research. The device is inexpensive, costing less than $25. Data is output into five interval time-stamped variables: steps per day, distance traversed, total sleep duration, restful sleep duration, and light sleep duration.

Facebook web application (Facebook, 2016). Participants granted the investigators access to all public activity conducted on their Facebook account for the duration of the study. All content generated by the user that was viewable by their social network was recorded. Thus, private messages were not included. When web articles, videos, songs, or other content not explicitly composed by the participant were posted, the material's title and subtitle was captured as relevant data. This decision was made because the title and subtitle is viewable directly by the participant's social network without being redirected to a different webpage. That is, the title and subtitle are expressed in full and may capture something about what the participant was doing or thinking at the time of the post. All content was compiled into a single text document for each participant, which was then analyzed for linguistic style.

Linguistic Inquiry and Word Count 2015 (LIWC) software. LIWC is a text analysis program that outputs percentages of words in a given document which fall into over 80 linguistic categories (Pennebaker, Francis, & Booth, 2001). The program allows for users to develop custom dictionaries tailored to their interest, or use the internal dictionary which was built from decades of the developer, Dr. James Pennebaker's, own work on psycholinguistics. Linguistic style has been investigated extensively in social science research. It can be utilized to explore a vast array of constructs from emotional tone to intelligence (Kess, 1992). Pennebaker's internal dictionary was utilized for the present analysis.

Google Voice web application. All EMA correspondence was conducted through the Google Voice text messaging service. Google Voice is a secure web-based telephone service offered to Google users (About Google Voice, 2016). Transmissions are composed on the user's computer and forwarded to a target cellular number. Responses are received by the user's Google account.

Measures

Beck Depression Inventory-II (BDI). The BDI (Beck, et al., 1996) was used to provide a stratified depression rating, wherein higher scores indicate more severe depression symptoms. This scale has been extensively used in research and repeatedly shown to have strong validity and reliability in multiple populations, with a recent comprehensive review reporting concurrent validity to other depression scales ranging from $r = 0.66–0.86$, and reliability ranging from $\alpha = 0.83–0.96$ (Wang, & Gorenstein, 2013). The BDI was administered during the follow-up assessment.

The Affect Grid. The Affect Grid is a single item measurement of current affective state, constructed on two continuous dimensions of hedonic valence (level of pleasure) and arousal (see Figure 2; Russell, Weiss, & Mendelsohn, 1989). From these two dimensions, higher-order emotional states can be inferred. For example, high arousal with pleasant feelings is often conceptualized as happy or excited. The measure has been independently validated for accuracy on the hedonic valence arousal dimensions (Killgore, 1998). The Affect Grid was utilized as the EMA variable in the present study. Preliminary analysis has shown the Affect Grid to be sensitive to emotionally salient events, and thus a prime measure for use in EMA (Renaud, & Thomas 2016). The assessment was administered twice daily via text message. Participants received a picture

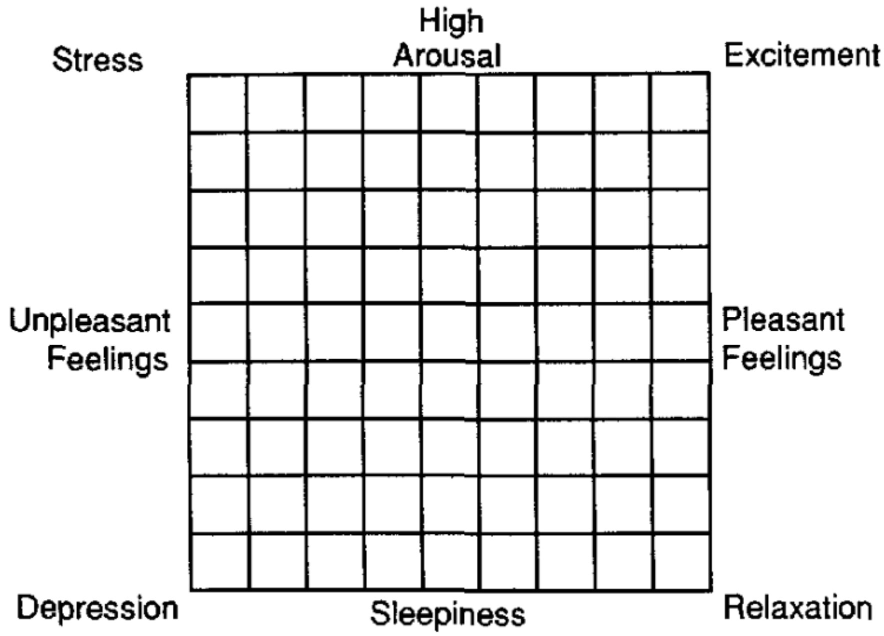

Figure 2. The Affect Grid (Russel, et al., 1989). The x-axis is hedonic valence. The y-axis is Level of arousal. The four corners display emotional landmarks.

of the grid with numbers 1–9 plotted on the X and Y axis. Coordinates (X,Y) of the response were recorded. Times of assessments were determined randomly, between 9am and 9pm.

Procedure

Each participant provided informed consent, as approved by the Institutional Review Board at Teachers College, Columbia University. Participants completed a basic contact/demographic information sheet and were fitted with a Misfit Flash™ device. They were instructed to keep the device on for the remainder of the study, including while they sleep. If the participant needed to remove the device (e.g., for comfort while bathing), they were instructed to replace it on their wrist as soon as possible. They were then trained to complete the Affect Grid assessment (see Appendix A). Participants were told that they would receive two text messages per day that contained a picture of the Affect Grid with numbers 1–9 plotted on the X and Y axis. They were instructed to respond to each text message with coordinates of their square's location (e.g., Feelings-5, Arousal-6) as soon as they safely could. A two-week follow-up appointment was then scheduled. For the subsequent two weeks, participant's physical activity and sleep was continuously sampled, and they received two text message EMA's each day. When the participant returned to the lab, the Misfit Flash data was collected, and the BDI assessment was administered. Participants were then asked to temporarily log onto their Facebook profile on a study computer in order to extract their Facebook data.

Data Analysis

Descriptive statistics were used to generate adherence reports. Daily physical activity and sleep variables were averaged over two-weeks to provide discrete steps-per-day and sleep-duration values. Hedonic valence and arousal ratings on the affect grid were each averaged over 26 assessments to generate two discrete global ratings. LIWC generated content percentages of 44 linguistic variables (see Appendix B). For tests comparing these behavioral domain outputs to BDI score Pearson product-moment correlation coefficients were calculated. All analyses were carried out on IBM SPSS (2011), and G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007).

Results

Feasibility

The observation period lasted an average of 14 days (range = 13–15). Attrition was low, as 91% of the participants completed the study protocol through to the follow-up. Physical activity data was captured for 86.4% of all days observed; sleep data was captured 79.3% of all nights. All participants received one EMA on intake and follow-up days, and two EMAs on all other days of observation, up to 14 days total. On average, 70.6% of the EMAs were responded to. Facebook content composed during the observation period consisted of an average of 189 words (range = 14–468, $SD = 142.47$). Adherence was high—cumulatively 83.6% of targeted data was collected. Participants reported minimal interference with daily life and had generally positive commentary about the design. A systematic review of previous studies using experience sampling methods comparable to ours reported similar participant compliance rates, with an average of 73–90% of targeted data acquired, and an 82–86% completion rate (Csikszentmihalyi, & Larson, 2014). The platform has shown to be highly feasible.

Convergent Validity

Indeed, evidence suggests that physical activity, sleep, social media activity, and momentary affect change concurrently with MDD status. Theoretically, as depression severity increases, physical activity levels and sleep quality should decrease. As psychologically positive social media content dissolves, and psychologically negative social-media content surfaces, depression severity is expected to increase. As momentary negative affect is more frequently reported, depression severity is anticipated to increase as well. Here we report if the behavioral data collected by the platform is reflective of these theories.

Physical activity. A bivariate correlation was conducted comparing average steps per day and BDI score. A moderate inverse correlation was very near significant ($r = -.501$, $p = .058$), suggesting a trend of higher levels of physical activity being associated with lower levels of depression. In agreement with the anticipated association, physical activity levels measured by actigraph, and BDI score validly converge.

Sleep. Bivariate correlations were conducted independently comparing minutes of total sleep, minutes of light sleep, and minutes of restful sleep to BDI score. Restful sleep duration and BDI score were found to be moderately inversely correlated ($r = -.536$, $p = .045$). Getting more restful sleep was associated with lower levels of depression. In agreement with the anticipated association, sleep quality measured by actigraph, and BDI score validly converge.

Facebook content. Bivariate correlations were conducted independently comparing each of the 44 linguistic variables mined from LIWC with BDI score. A moderate correlation with BDI was found in the percentage of masculine linguistic content (e.g., boy, man, he) ($r = .590$, $p = .028$). Composing more masculine words during the observation period was associated with higher levels of depression. A moderate inverse correlation with BDI was also found in the percentage of future-tense linguistics (e.g., will, soon, going to) ($r = -.539$, $p = .043$). Composing more future focused words during the observation period was associated with lower levels of depression. A third moderate correlation with BDI was found in the presence of perceptual-processing through 'feeling' linguistics, including emotion-focused words (e.g., feel, sad, angry) ($r = .532$, $p = .046$). Composing more words in the 'feeling' category during the observation period was associated with higher levels of depression.

Surprisingly, explicitly positive or negative linguistic content (e.g., positive emotional tone and negative emotional tone) was unrelated to BDI score. Instead, more dynamic associations were found. Masculine content being associated with depression is perplexing. It may be an artifact that emerged, in part, because of the 2016 Presidential Election. These data were collected from February–May of 2016, during which a blitz of negative content circulated social media concerning "him"—Donald Trump—which may be a confounding variable. The relationships between future-tense linguistics and depression, and feeling linguists and depression are not surprising. Rumination, particularly about the past, is a common attribute in the Self-Regulatory Executive Function (S-REF) model of depression (Wells, & Matthews, 1996). Participants with more symptoms of depression may have used Facebook as a vehicle to express their ruminative or retrospective thoughts. The correlation we found is reflective of the S-REF model. Furthermore, Billings, & Moos (1984) demonstrated that recruiting social resources for emotional-discharge—that is, expressing your emotional problems to others—is a common coping strategy used by individuals with MDD. Participants may have utilized their social networks on Facebook to engage in emotional-discharge. In line with these theories, two of the linguistic variables that emerged from the behavioral data that the platform collected validly converge with BDI score.

EMA Affect Grid. Bivariate correlations were conducted independently comparing global hedonic valence, and arousal to BDI score. A moderate inverse correlation was found between global hedonic valence and BDI score ($r = -.582, p = .03$). Lower average ratings of hedonic valence—that is, repeatedly reporting unpleasant feelings—was associated with higher BDI scores. In agreement with the anticipated association, ecologically assessed negative affect and BDI score validly converge.

Ancillary Analysis

Given the good convergent validity of each of our behavioral domains with depression rating, we attempted a multiple linear regression to predict BDI score based on average daily steps, average nightly minutes of restful sleep, global hedonic valence, and percentage of future-focused, feeling, and masculine linguistics in Facebook content. The underpowered model (β-1 = .46) was unable to achieve significance ($F(6,4) = 1.450$, $p = .375$, $R^2 = .685$). With $R^2 = .685$, $\alpha = .05$, and six predictor variables, the model needs $n = 18$ to achieve >95% power.

Discussion

Using data generated by interactive technologies, the platform was able to identify six discrete behavioral measures that have good convergent validity with the BDI for assessing depression symptoms (five were significant, one was trending towards significance). If deployed in a primary care setting, providers would be able to remotely monitor patients from the onset of their prescribed antidepressant regimen. The presented evidence suggests that the platform is capable of detecting behaviors reminiscent of remission, or more importantly, lack thereof, in real time. Thus, the platform can supply healthcare providers with valuable information that helps identify patients who require medication adjustment or augmentation. With further research to achieve more statistical power, the regression model tested in the ancillary analysis has the potential to have unprecedented predictive ability of depression rating, explaining 68.5% of felt symptom variance.

Limitations

We believe the platform is in its infancy. We have presented evidence that suggests it can validly monitor depression symptoms; however, we have not yet developed the software infrastructure needed to deploy the platform on a large scale. To achieve large scale deployment, we must collaborate with software developers to automate the platform's back-end data aggregation methods.

Our analyses were inherently limited by our sample size, and BDI range (0–15). In order to strengthen the validity of the platform, our findings must be replicated with a larger sample size, and an expanded BDI range, which requires studying more severely depressed individuals. As this range and the sample size grows, it is imperative to be iterative in re-mining LIWC and re-testing other available variables for new correlations.

Ethical considerations are important when providing clinicians with access to personal information that is not traditionally considered related to healthcare. Some patients may be uncomfortable with having their behaviors tracked in the manner that the platform does. As we further develop the platform, all back-end software development will need to be secured to federally mandated privacy standards for patient data, and all identifiable data will need to be classified as protected health information. Additionally, patients will need to provide consent prior to enrolling in the platform.

Cost of implementation is another potential limiting factor. Although designed to consider cost, some patients will inevitably have economic restrictions that limit their enrollment in the platform. The actigraph we utilized retails for $19.99. It may be possible to mitigate this expense by capitalizing on select health insurance benefits (e.g., Humana, 2016) which provide activity trackers to customers at no cost; however, this option is not available to everyone. Perhaps the most substantial limiting economic factor is that enrolment requires easy access to a computer or smartphone for social media participation, and the EMA requires a smartphone. A shrinking minority of Americans either don't use social media, or don't possess a smartphone and hence, the platform will be unavailable to these individuals.

The presented study only establishes preliminary validity of the constructs that the platform monitors. This does not necessarily mean that the platform will assist real-life identification of patients in need of treatment adjustment/augmentation. Testing this requires a randomized controlled field implementation study comparing the platform versus a sham monitoring system. All of the presented findings must be considered alongside these limitations. Substantial further research and development is needed.

References

About Google Voice. (2016) Retrieved March 20, 2016, from https://support.google.com/voice/answer/115061?hl=en

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the beck depression inventory-II.

Billings, A. G., & Moos, R. H. (1984). Coping, stress, and social resources among adults with unipolar depression. *Journal of Personality and Social Psychology*, *46*(4), 877.

Budget | National Institutes of Health (NIH). (2015, October 14). Retrieved November 10, 2015, from http://www.nih.gov/about-nih/what-we-do/budget#note

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35–54). Springer Netherlands.

Cooney, G. M., Dwan, K., Greig, C. A., Lawlor, D. A., Rimer, J., Waugh, F. R., . . . & Mead, G. E. (2013). Exercise for depression. *The Cochrane Library*.

Facebook – Company Info. (2016). Retrieved March 20, 2016, from http://newsroom.fb.com/company-info/

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act*, *12*(1), 42.

Franzen, P. L., & Buysse, D. J. (2008). Sleep disturbances and depression: risk relationships for subsequent depression and therapeutic implications. *Dialogues in Clinical Neuroscience*, *10*(4), 473.

Greenberg, P. E., Kessler, R. C., Birnbaum, H. G., Leong, S. A., Lowe, S. W., Berglund, P. A., et al. (2003). The economic burden of depression in the United States: How did it change between 1990 and 2000? *Journal of Clinical Psychiatry*, *64*, 1465–1475.

Humama – New Fitness Device Experience. (2016). Retrieved December 07, 2016, from https://www.humana.com/learning-center/health-and-wellbeing/fitness-and-exercise/devices

Hunt, M., Auriemma, J., & Cashaw, A. C. (2003). Self-report bias and underreporting of depression on the BDI-II. *Journal of Personality Assessment*, *80*(1), 26–30.

Keller, M. B. (2003). Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. *Jama*, *289*(23), 3152–3160.

Kess, J. F. (1992). *Psycholinguistics: Psychology, Linguistics, and the Study of Natural Language* (Vol. 86). John Benjamins Publishing.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., . . . & Wang, P. S. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Jama*, *289*(23), 3095–3105.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 602.

Killgore, W. D. S. (1998). The Affect Grid: a moderately valid, nonspecific measure of pleasure and arousal. *Psychological Reports*, *83*(2), 639–642.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. *Journal of General Internal Medicine*, *16*(9), 606–613.

Liu, P., Tov, W., Kosinski, M., Stillwell, D. J., & Qiu, L. (2015). Do Facebook Status Updates Reflect

Subjective Well-Being?. *Cyberpsychology, Behavior, and Social Networking*, *18*(7), 373–379.

Löwe, B., Kroenke, K., Herzog, W., & Gräfe, K. (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of Affective Disorders*, *81*(1), 61–66.

Mathers, C., Fat, D. M., & Boerma, J. T. (2008). The global burden of disease: 2004 update. World Health Organization.

Otto, M. W., Church, T. S., Craft, L. L., Greer, T. L., Smits, J. A., & Trivedi, M. H. (2007). Exercise for mood and anxiety disorders. *Journal of Clinical Psychiatry*, *68*(5), 669.

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment*, *5*(2), 164.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*, 2001.

Pew Research Center – Social Media Usage: 2005–2015. (2015A). Retrieved March 20, 2016, from http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/

Pew Research Center – US Smartphone Use in 2015. (2015B). Retrieved March 20, 2016, from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Porta, M., Greenland, S., Hernán, M., dos Santos Silva, I., & Last, J. M. (2014). *A Dictionary of Epidemiology*. Oxford University Press, USA.

Renaud, A.M., & Thomas, J.E. (2016, April). *Detecting Emotionally Salient Events in Real Time: Validating a single item scale for ecological momentary assessment*. Poster session presented at the Psychology at Teachers College Columbia University Conference, New York City, NY.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493.

Settanni, M., & Marengo, D. (2015). Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology*, *6*.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32.

Slater, J. A., Botsis, T., Walsh, J., King, S., Straker, L. M., & Eastwood, P. R. (2015). Assessing sleep using hip and wrist actigraphy. *Sleep and Biological Rhythms*, *13*(2), 172–180.

Smith, J. P., & Smith, G. C. (2010). Long-term economic costs of psychological problems during childhood. *Social Science & Medicine*, *71*(1), 110–115.

SPSS, I. (2011). IBM SPSS statistics base 20. *Chicago, IL: SPSS Inc.*

Stassen, H. H., Delini-Stula, A., & Angst, J. (1993). Time course of improvement under antidepressant treatment: a survival-analytical approach. *European Neuropsychopharmacology*, *3*(2), 127–135.

Stathopoulou, G., Powers, M. B., Berry, A. C., Smits, J. A., & Otto, M. W. (2006). Exercise interventions for mental health: a quantitative and qualitative review. *Clinical Psychology: Science and Practice*, *13*(2), 179–193.

Trivedi, M. H., Fava, M., Wisniewski, S. R., Thase, M. E., Quitkin, F., Warden, D., . . . & Luther, J. F. (2006). Medication augmentation after the failure of SSRIs for depression. *New England Journal of Medicine*, *354*(12), 1243–1252.

Tudor-Locke, C., Williams, J. E., Reis, J. P., & Pluto, D. (2002). Utility of pedometers for assessing physical activity. *Sports Medicine*, *32*(12), 795–808.

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Revista Brasileira de Psiquiatria*, *35*(4), 416–431.

Wei, J. (2014). How Wearables Intersect with the Cloud and the Internet of Things: Considerations for the developers of wearables. *Consumer Electronics Magazine, IEEE*, *3*(3), 53–56

Wells, A., & Matthews, G. (1996). Modelling cognition in emotional disorder: The S-REF model. *Behaviour Research and Therapy*, *34*(11), 881–888.

Appendix A

Affect Grid Training

With the grid in front of them, participants were read the following passage:

> This is the Affect Grid. It is intended to capture how you feel in a given moment. First you orient yourself with the horizontal axis. On the left side of the grid represents unpleasant feelings, on the right are pleasant feelings, and in the middle is neutral. After you have located your place on this axis, you then adjust either up or down on the vertical axis according to your current energy level. At the top represents high arousal or high energy, and the bottom represents sleepiness or low arousal. Again, the middle is neutral. In the four corners you will find landmarks to help you navigate the grid. These landmarks are emotions that many people conceptualize certain couplings of feelings and arousal. For example, someone with unpleasant feelings and high arousal may view that state as stress. Inversely, pleasant feelings with low arousal may feel like relaxation. These landmarks are there to help you navigate the grid; however, you may conceptualize the feelings differently than other people. That is ok. Please more so focus on your location along the horizontal and vertical axis. When you are ready, please indicate so by touching the center of the square associated with where you are located.

Appendix B

List of LIWC Variables Analyzed

Summary Dimensions

Word count
Analytic
Clout
Authentic
Tone
Words per sentence
Words longer than six letters
Dictionary word count

Perceptual Processes

See
Hear
Feel

Biological Processes

Body
Health
Sexual
Ingest

Drives

Affiliation
Achieve
Power
Reward
Risk

Time Orientation

Past Focus
Present Focus
Future Focus

Affect

Positive Emotional Tone
Negative Emotional Tone
Relativity:
Motion
Space
Time

Social

Family
Friend
Feminine
Masculine

Cognitive Processes

Insight
Casual
Discrepancies
Tentative
Certainty
Differentiation

Personal Concerns

Work
Leisure
Home
Money
Religion
Death