

J Med Internet Res. 2018 Jul; 20(7): e241.

PMCID: PMC6076371

Published online 2018 Jul 20. doi: 10.2196/jmir.9775; 10.2196/jmir.9775

PMID: [30030209](#)

Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A BiAffect Digital Phenotyping Study

Monitoring Editor: Gunther Eysenbach

Reviewed by Diego Hidalgo-Mazzei, Ines Hungerbuehler, Darius Adam Rohani, and Christopher Karr

[John Zulueta](#), MD,¹ [Andrea Piscitello](#), MS,¹ [Mladen Rasic](#), BA,¹ [Rebecca Easter](#), BA,¹ [Pallavi Babu](#), BA,² [Scott A Langenecker](#), PhD,¹ [Melvin McInnis](#), MD,² [Olusola Ajilore](#), MD, PhD,¹ [Peter C Nelson](#), PhD,¹ [Kelly Ryan](#), PhD,² and [Alex Leow](#), MD, PhD^{✉1}

¹ University of Illinois at Chicago, Chicago, IL, United States

² University of Michigan, Ann Arbor, MI, United States

Alex Leow, University of Illinois at Chicago, 1601 W Taylor St, Chicago, IL, 60612, United States, Phone: 1 312 996 7383, Fax: 1 312 996 9517, Email: aleow@psych.uic.edu.

[✉]Corresponding author.

Corresponding Author: Alex Leow aleow@psych.uic.edu

Received 2018 Jan 4; Revisions requested 2018 Mar 24; Revised 2018 May 13; Accepted 2018 May 29.

Copyright ©John Zulueta, Andrea Piscitello, Mladen Rasic, Rebecca Easter, Pallavi Babu, Scott A Langenecker, Melvin McInnis, Olusola Ajilore, Peter C Nelson, Kelly Ryan, Alex Leow. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.07.2018.

This is an open-access article distributed under the terms of the Creative Commons Attribution License

(<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Abstract

Background

Mood disorders are common and associated with significant morbidity and mortality. Better tools are needed for their diagnosis and treatment. Deeper phenotypic understanding of these disorders is integral to the development of such tools. This study is the first effort to use passively collected mobile phone keyboard activity to build deep digital phenotypes of depression and mania.

Objective

The objective of our study was to investigate the relationship between mobile phone keyboard activity and mood disturbance in subjects with bipolar disorders and to demonstrate the feasibility of using passively collected mobile phone keyboard metadata features to predict manic and depressive signs and symptoms as measured via clinician-adminis-

tered rating scales.

Methods

Using a within-subject design of 8 weeks, subjects were provided a mobile phone loaded with a customized keyboard that passively collected keystroke metadata. Subjects were administered the Hamilton Depression Rating Scale (HDRS) and Young Mania Rating Scale (YMRS) weekly. Linear mixed-effects models were created to predict HDRS and YMRS scores. The total number of keystrokes was 626,641, with a weekly average of 9791 (7861), and that of accelerometer readings was 6,660,890, with a weekly average 104,076 (68,912).

Results

A statistically significant mixed-effects regression model for the prediction of HDRS-17 item scores was created: conditional $R^2=.63$, $P=.01$. A mixed-effects regression model for YMRS scores showed the variance accounted for by random effect was zero, and so an ordinary least squares linear regression model was created: $R^2=.34$, $P=.001$. Multiple significant variables were demonstrated for each measure.

Conclusions

Mood states in bipolar disorder appear to correlate with specific changes in mobile phone usage. The creation of these models provides evidence for the feasibility of using passively collected keyboard metadata to detect and monitor mood disturbances.

Keywords: digital phenotype, mHealth, ecological momentary assessment, keystroke dynamics, bipolar disorder, depression, mania, mobile phone

Introduction

The burden of mental illness is high. It has been estimated that mental illness accounts for 32% of years lived with disability around the world [1]. Bipolar disorder is a serious mental illness characterized by recurrent episodes of depression and mood elevation [2] and is associated with high rates of functional impairment, decreased quality of life, and increased rates of mortality from comorbid medical conditions [3]. Given these costs, it is imperative that we deepen our understanding of this disorder to promote accurate diagnosis and effective treatment.

The ubiquity of mobile phones, smartphones in particular, presents a new opportunity in the study of mental illness. An estimated 64% of adults in the United States own a mobile phone and use it for a variety of tasks, including phone calls, Web browsing, and social media; however, the most widely and frequently used feature on mobile phones is short message service text messaging [4]. These devices can be employed as platforms for the unobtrusive collection of myriad data that can be used in the study of psychopathology. Ecological momentary assessment is a methodology that aims to collect data using repeated measures in real time (or near real time), in people's natural environment [5]. When applied to the use of digital technologies such as mobile phones, this methodology can be used to create digital phenotypes defined as the set of observable behaviors resulting from the interaction between human disease and people's use of digital technologies [6].

Because recurring mood episodes are a defining characteristic of bipolar disorder, we posited that it is an ideal illness for a pilot study investigating the relationship between mobile phone keyboard activity and the correlates of these episodes, such as changes in cognitive function, psychomotor activity, social behavior, and diurnal activity patterns.

We elected to focus on keystroke dynamics because features using text input (eg, texting and Web browsing) are among the most commonly used features in mobile phones and because we hypothesized that keystroke dynamics provide a sufficiently dense space from which to extract relevant features that could be used to predict the severity of depression and mania.

Methods

Participants

Study subjects were members of the Prechter Longitudinal Study of Bipolar Disorder, a naturalistic, longitudinal study based in the University of Michigan [7]. This cohort includes subjects with bipolar disorder, other psychiatric illnesses, and healthy controls; however, only those with bipolar disorder were recruited into this study. Subjects were recruited into this study by email or phone invitation. The inclusion criteria included being a current Android mobile phone user, asserting familiarity with the Android operating system, having no gross impairments in fine motor abilities, sufficient vision to use a mobile phone keyboard, and self-reporting of frequent mood fluctuations or having longitudinal data from the longitudinal study suggesting that they experience frequent mood symptoms (ie, endorsed frequent mood symptoms on bimonthly self-report measures of mood or categorized as rapid cycling).

We initially included 19 subjects with a bipolar spectrum disorder as per the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (Text Revision) criteria [8] (11 with bipolar I, 7 with bipolar II, and 1 with bipolar not otherwise specified); of these subjects, 1 never activated the app and 2 deleted the app early in the study. Of the remaining 16 subjects, participation varied in terms of the number of weeks that had any keyboard activity, with an average of 4.69 (3.05) weeks. Because of concerns about adherence, data analysis was restricted to subjects who provided data for at least 4 weeks. This resulted in 9 subjects: 5 with bipolar I and 4 with bipolar II. Of these, 8 subjects met the criteria for rapid cycling (ie, 4 or more mood episodes per year), and all subjects with bipolar II had recurrent depressive episodes. Of these 9 subjects, 7 showed keyboard activity for at least 6 weeks. The total usable data from these subjects included 626,641 keystrokes and 6,660,890 accelerometer readings.

Mobile Keyboard

A custom keyboard called “BiAffect” was developed for the Android operating system that replaced the default keyboard and collected metadata consisting of keystroke entry date and time and accelerometer displacement. It uploaded these data using secure encrypted protocols to the study server hosted at the University of Illinois at Chicago. Accelerometer data collection was initiated by keystroke entry and continued for 5 seconds afterward. Individual character data outside of the backspace key and space bar were not collected, anonymizing the entry. The keyboard was designed to appear similar to the standard Android keyboard ([Figure 1](#)).

Data Collection

For 8 weeks, subjects were provided a Samsung Galaxy Note 4 smartphone that they were instructed to use as their primary phone during the study period. Subjects were encouraged to use their current phone number and subscriber identification module card; with the exception of 1 subject, all subjects did so. During the study period, trained staff at the University of Michigan administered the Structure Interview Guide for the Hamilton Depression Rating Scale (HDRS) [9] and Young Mania Rating Scale (YMRS) [10] once a week via phone interviews.

Statistical Analyses

Subject demographics are described in [Table 1](#). The YMRS results showed a right-tailed skew ($\gamma_1=1.14$) [[11](#)], so a log transformation was performed on the YMRS scores by taking the natural log of the sum of the YMRS scores and 1 ($\gamma_1=-0.44$).

In order to identify the possible relationships between subject demographics and phone usage, Spearman correlations were calculated between subjects' total key counts and their age and education.

Mixed-effects linear models were created correlating keyboard metadata collected from the week prior to the administration of the HDRS (17-item) and YMRS mood rating scores. Missing data were handled with pairwise deletion. Features extracted from the metadata were modeled as fixed effects. Observations were grouped by subject, with each subject having his or her own random intercept for his or her mood ratings.

Overall significance was assessed by using likelihood ratio tests comparing the null models that consisted of just the subject-level effect with full models consisting of the subject-level effect and metadata features. Because the mixed-effects model for the YMRS scores showed that the random effect was accounting for none of the variance of the YMRS scores, a fixed-effects ordinary least squares model was created instead (mixed-effects model log likelihood -64.621 , Akaike Information Criterion 149.24, Bayesian Information Criterion 170.83; fixed-effects ordinary least squares model: log likelihood -64.621 , Akaike Information Criterion 147.24, Bayesian Information Criterion 166.67). For the HDRS model, conditional and marginal R^2 values were calculated using the method specified by Nakagawa and Schielzeth [[12](#)], as implemented in the R package `piecewiseSEM` [[13](#)]. Using this method, the conditional R^2 is equal to the proportion of variance explained by both the fixed and random effects, and the marginal R^2 is equal to the proportion of the variance explained by the fixed effects alone. The P values of the model coefficients were calculated using Wald chi-square tests, as implemented in the R package `car` [[14](#)] for the HDRS model. For the YMRS model, overall significance was tested using an F-test and individual coefficient significance was determined with t-tests.

The fixed-effect variables included the average interkey typing delay, the average accelerometer displacement, the backspace and autocorrect rates (ie, the total number of each divided by the total number of keystrokes), the average length of each typing session in seconds, the total number of typing sessions, and the cosine similarity between each week's keypress activity and the total keypress activity of the study period (described further below). All aggregate variables were calculated for the week preceding each mood assessment. A session was defined as beginning with a keypress that occurs after 5 or more seconds have elapsed since the last keypress and ending when 5 or more seconds have elapsed between keypresses.

Models were created using the software package `lme4` [[15](#)] for the R software environment version 3.3.3 [[16](#)].

Predictor Variables

The predictor variables were chosen based on the hypothesis that they map to key cognitive and behavioral domains affected by mania and depression. [Table 2](#) provides definitions of each variable, and each domain and their corresponding variables are discussed in turn below.

Psychomotor Activity As per the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), changes in psychomotor functioning are criteria for both major depressive and manic episodes [[2](#)]. Psychomotor activity is also a component of the clinician's ratings within HDRS and YMRS. We hypothesized that psychomotor activity (agitation and retardation) manifests in the accelerometer displacement and the average interkey delay. We predicted that increasing levels of psychomotor agitation lead to subjects holding their phones less stably, thus resulting in higher accelerometer displacement values. In the case of average interkey delay, it can be argued that increased levels of psychomotor agitation could lead to either a lower or higher delay. In the case of the former, higher levels of agitation

would lead to a general speeding up of behaviors, including typing; however, it is also possible that while more agitation may lead to an increase in the amount of activity, the ability to effectively type will be impaired, leading to a higher interkey delay and possibly more use of backspace and autocorrect. In contrast, psychomotor retardation was hypothesized to manifest as a higher average delay.

Social Activity The BiAffect app did not capture the context of keyboard activity; however, we hypothesized that increases in keyboard activity are likely associated with increased social activity consisting of both texting and social media usage and that more activity would be associated with higher YMRS scores and lower HDRS scores. There are mixed data on the role of social media use and depression, with some studies showing decreases [17] and others reporting increases in social media usage in both high school [18] and college [19] students.

Cognition Impairments in attention and concentration are seen in both depressive and manic episodes, as described in the DSM-5 and previous studies [2]. Impulsivity and deficits in error correction have also been identified as features seen in manic episodes [20]. Variables that characterize concentration and cognition were hypothesized to include the average interkey delay, the backspace rate, and the autocorrect rate. It was hypothesized that increased backspace rates indicated increased error correction and increased autocorrect rates indicated decreased error detection. Impaired concentration was hypothesized to manifest as increased interkey delay.

Diurnal Activity Patterns Changes in sleep patterns are characteristic of both depressive and manic episodes. In the case of depression, this may take the form of insomnia or hypersomnia, whereas in the case of mania, there is typically a decreased need for sleep [2]. We expected that such changes in sleeping patterns would manifest as changes in phone typing activity. To characterize such changes, we created a cosine-based similarity feature of keypress activity. Cosine-based similarity is a frequently used technique in the field of machine learning and predictive algorithms to characterize the similarity between entities [21,22]. In our implementation, the distribution of keypress activity for a given week was defined as vector of 24 dimensions, with each dimension corresponding to an hour of the day. The value of the vector in each dimension was set equal to the number of keypresses in that hour. We then calculated the cosine of the angle between each week's vector and the vector representing activity for the entire study period. In this way, the more dissimilar a given week's pattern of activity was compared to the total activity, the lower the value of the cosine would be. It was hypothesized that more dissimilar weeks would correspond to higher HDRS and YMRS scores.

Results

Predictor Variable Summary Statistics

Summary statistics for each predictor variable are presented in [Table 3](#).

Total Key Press Activity and Subject Demographics

No statistically significant correlations were found between total key counts and subjects' age ($S=139.16$, $P=.68$) and education levels ($S=144.41$, $P=.60$).

Prediction of Depression Symptoms

Likelihood ratio testing comparing the null model that consisted of just the subject-level random effect to the full model showed that the full model had superior fit ($\chi^2_7=17.6$, $P=.01$; see [Tables 4](#) and [5](#)). The marginal R^2 (ie, the proportion of the variance explained by the metadata features) was 0.41, and the conditional R^2 (ie, the proportion of the vari-

ance explained by both the subject-level effect and the metadata features) was .63. Accelerometer displacement ($P=.002$), average interkey delay ($P=.02$), session count ($P=.003$), and the autocorrect rate ($P=.004$) were found to be positively correlated with the HDRS scores.

Prediction of Hypomania or Mania Symptoms

A multiple linear regression model was created that accounted for 34% of the variance of the natural logarithm of YMRS scores (multiple $R^2=.34$, $F_{7,56}=4.08$, root mean square error=.66, $P=.001$; [Table 5](#)). Accelerometer displacement ($P=.003$) was found to be positively correlated with YMRS scores, and the backspace rate ($P=.01$) was found to be negatively correlated.

Discussion

Principal Findings

Using only passively collected metadata, keystroke activity predicted both depressive and manic symptoms. The model to predict depression scores demonstrated greater explanatory capacity as shown by the larger proportion of variance explained by the model and the larger number of significant predictors.

Psychomotor Activity

Increased accelerometer activity was found to be positively correlated with both depression and mania scores. One possible explanation for the positive correlation with both scores is that the subjects in our study had more mildly agitated or irritable forms of depression or depression with mixed features rather than forms exhibiting psychomotor retardation.

Social Activity

In contrast to our hypothesis that decreased sessions would be predictive of higher depression, the overall number of sessions was actually positively correlated with depression. This may be a reflection of the dynamic between loneliness and withdrawal. Sessions from a phone can be seen as lower risk and can also include passive use of social media, such as viewing but not posting, enabling a feeling of connection and withdrawal. At least one study has demonstrated an association between increased usage of the internet more generally and depressive symptoms [23]. It is also worth noting that while the session count was positively correlated, the average session length was negatively correlated (although this predictor did not reach statistical significance, $P=.15$), suggesting that patterns of activity may be more relevant than the overall volume of activity.

Cognition

Impairments in executive function have been demonstrated more in individuals with bipolar disorder in depressed, manic, and euthymic states than in healthy controls [20], although it has also been shown that executive functioning may be especially impacted during manic states [24,25]. Interestingly, our depression and mania symptom models diverge in their relationships with respect to what we theorized would be the key features related to cognition: backspace and autocorrect rates. The increase in autocorrect rate with depression symptoms seems relatively straightforward. Here, the ability to concentrate becomes impaired in more depressed states, and therefore, the rate of typing errors increases. What is less clear is why the backspace rate would be negatively correlated with mania symptoms without a

concomitant positive correlation with the autocorrect rate. One possibility is that the lower backspace usage seen with higher mania scores reflects a phenomenon of less self-monitoring or impaired response inhibition with errors. Those with elevated mania do not trigger the autocorrect mechanism because their inputs are generally correctly spelled but often grammatically or semantically inappropriate words, fitting the profile of someone who keeps deleting what they type because it was impulsively entered.

Diurnal Activity Patterns

Because sleep disturbance is such a prominent aspect of mood disturbance, we were surprised that measurements that aimed to reflect diurnal variations in activity were not predictive of depressive or mania symptoms. With the assumption that the distribution for the entire observation period would approximate the subject's baseline, we expected that lower values of similarity would be correlated with higher depression and mania scores. The cosine similarity values did not reach statistical significance in both models. One possible explanation for this is that the period of observation was not long enough to establish actual baselines in the sense of encompassing activity through a variety of mood states, including euthymia, and that the distribution for the entire observation period for many subjects corresponded to a single mood state. Another important consideration is that while diurnal patterns of phone activity may be related to sleep, they are not identical.

Limitations and Future Directions

The limitations of this study include its sample size (relative to the model's complexity), sample characteristics that are probably not representative of a general population (ie, mostly women who have a high frequency of episodes), and the constraint of having subjects using study-issued phones. A larger study in which participants use their own phones is warranted in order to determine the generalizability of these findings. More data may also enable the creation of more sophisticated models with higher rates of prediction accuracy and reliability.

Unfortunately, there were fewer predictors of mania scores, and overall, this prediction was less accurate. Prediction of acute changes in mania may have stronger clinical implications, given the reduced tendency to seek treatment in mania generally. We suspect that primary reasons for the decreased prediction of mania are that our sample contained generally low mania scores and that both mania and hypomania elevations are often short and sporadically observed relative to longer and more stable episodes of depression. Rather than demonstrating correlates of mania per se, the mania model presented here might represent correlates of mixed or agitated depression.

Comparison with Prior Work

Prior studies have investigated the potential utility of various aspects of mobile phone activity as a means to diagnose mood states. Early studies focused on demonstrating the practicality of collecting self-reports of mood using mobile phones from patients [26,27]. While this approach may increase the facility with which such data are collected, it is still subject to the biases associated with self-reported data, potentially leading to spurious results [28]. More recent studies have focused on the validation of passive data collection methods and yielded encouraging results. Passive data features that have been demonstrated to correlate with mood ratings include physical movement [29,30], amount of phone usage [30], and frequency of calls and text messages with personal contacts [31].

The use of keystroke dynamics as a means to detect the emotion or mood of users is an active area of research in the field of affective computing, with most studies to date investigating the use of desktop keyboards [32]; however, there have been at least two studies that have examined the use of mobile phone keyboard dynamics as means to recognize user emotion. The first study was a 2-week pilot study based on the activity of a single user on Twitter, wherein the user was instructed to write a Tweet whenever he or she experienced certain emotions and to record the emotion from

a preset selection of options. Using a Bayesian Network classifier, the investigators were able to achieve an overall classification accuracy of 67.52%, with the most important feature being typing speed [33]. The second study consisted of a larger sample of 22 subjects and was conducted over 3 weeks. It also presented users with a preset selection of options for emotions; although, in contrast to the first study, keyboard activity was recorded over all applications and the users were prompted to input their emotional state on a regular basis. Using a random forest model, the investigators were able to achieve an average classification accuracy of 84%, with the most important typing dynamic feature being typing speed [34].

Although the aforementioned studies measuring mobile phone keystroke dynamics sought to predict emotion rather than mood, we find the relative importance of typing speed as an important feature across their studies as well as our own to be of note. To the best of our knowledge, our study is the first effort to use passively collected mobile phone keyboard metadata features to predict mood disturbances in a clinical sample using clinically relevant measures.

Conclusions

Passively collected mobile phone keystroke dynamics may be a useful and important method to identify incipient mood processes in persons with bipolar disorder. The facility with which such data may be used to infer the presence and severity of mood disturbances may enable clinical providers to intervene earlier in their patients' mood episodes, as well as increase the number of patients a single provider can effectively manage. Models such as those presented here may also lead to a deeper understanding of these disorders by revealing novel behavioral traits associated with them.

Acknowledgments

The authors thank the staff and participants of the Heinz C Prechter Research Program. This work was funded by the Mood Challenge, a New Venture Fund program funded by the Robert Wood Johnson Foundation.

Abbreviations

DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
HDRS	Hamilton Depression Rating Scale
YMRS	Young Mania Rating Scale
SMS	short messaging service

Footnotes

Contributed by

Authors' Contributions: AL had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest: MM is listed as an inventor on US Patent US9685174 (mood monitoring of bipolar disorder using speech analysis).

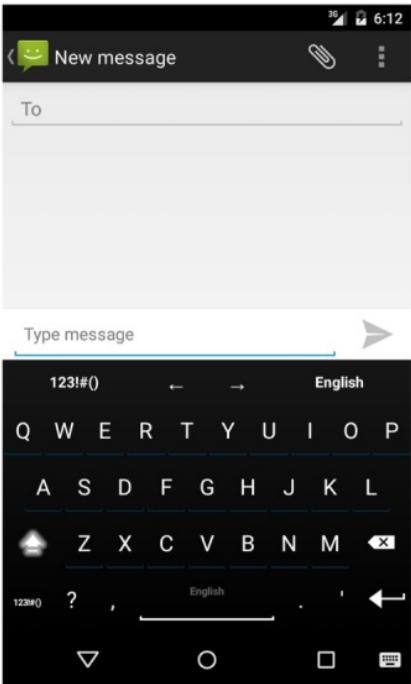
References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. *Lancet Psychiatry*. 2016 Feb;3(2):171–8. doi: 10.1016/S2215-0366(15)00505-2.S2215-0366(15)00505-2 [PubMed: 26851330] [CrossRef: 10.1016/S2215-0366(15)00505-2]
2. American Psychiatric Association *Internet*. [2018-07-03]. Diagnostic and statistical manual of mental disorders (5th edition) <https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596> .
3. Sajatovic M. Bipolar disorder: disease burden. *Am J Manag Care*. 2005 Jun;11(3 Suppl):S80–4. <http://www.ajmc.com/pubMed.php?pii=2889.2889> [PubMed: 16097718]
4. *Pew Research Center*. US Smartphone Use in 2015 <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>
5. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1–32. [PubMed: 18509902]
6. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol*. 2015 May 12;33(5):462–3. doi: 10.1038/nbt.3223.nbt.3223 [PubMed: 25965751] [CrossRef: 10.1038/nbt.3223]
7. McInnis Melvin G, Assari Shervin, Kamali Masoud, Ryan Kelly, Langenecker Scott A, Saunders Erika F H, Versha Kritika, Evans Simon, O'Shea K Sue, Mower Provost Emily, Marshall David, Forger Daniel, Deldin Patricia, Zoellner Sebastian, Prechter Bipolar Clinical Research Collaborative Cohort profile: the Heinz C. Prechter longitudinal study of bipolar disorder. *Int J Epidemiol*. 2018 Feb 01;47(1):28–28n. doi: 10.1093/ije/dyx229. <http://europepmc.org/abstract/MED/29211851> .4683653 [PMCID: PMC5837550] [PubMed: 29211851] [CrossRef: 10.1093/ije/dyx229]
8. American Psychiatric Association *Diagnostic and statistical manual of mental disorders, fourth edition, text revision*. 2000. [2018-07-08]. <https://dsm.psychiatryonline.org/doi/pdf/10.1176/appi.books.9780890420249.dsm-iv-tr> .
9. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960 Feb;23:56–62. <http://jnnp.bmj.com/cgi/pmidlookup?view=long&pmid=14399272> . [PMCID: PMC495331] [PubMed: 14399272]
10. Young R, Biggs J, Ziegler V, Meyer D. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry*. 1978 Nov;133:429–35. [PubMed: 728692]
11. Doane D, Seward L. Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education*. 2017 Aug 29;19(2):A. doi: 10.1080/10691898.2011.11889611. <https://www.tandfonline.com/doi/full/10.1080/10691898.2011.11889611> . [CrossRef: 10.1080/10691898.2011.11889611]
12. Nakagawa S, Schielzeth H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol Evol*. 2012 Dec 03;4(2):133–142. doi: 10.1111/j.2041-210x.2012.00261.x. <http://doi.wiley.com/10.1111/j.2041-210x.2012.00261.x> . [CrossRef: 10.1111/j.2041-210x.2012.00261.x]
13. Lefcheck J. piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods Ecol Evol*. 2016;7(5):579. doi: 10.1111/2041-210X.12512. [CrossRef: 10.1111/2041-210X.12512]
14. Fox J, Weisberg S. *An R Companion to Applied Regression Internet. Second Edition*. Thousand Oaks, CA: Sage; 2011.
15. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 2015;67(1):1. doi: 10.18637/jss.v067.i01. <http://www.jstatsoft.org/v67/i01/> [CrossRef: 10.18637/jss.v067.i01]
16. R Core Team . *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017.
17. De Chourdry M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. International Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, MA. 2013. <https://www.microsoft.com/en-us/research/publication/predicting-depression-via-social-media/>

18. Pantic I, Damjanovic A, Todorovic J, Topalovic D, Bojovic-Jovic D, Ristic S, Pantic S. Association between online social networking and depression in high school students: behavioral physiology viewpoint. *Psychiatr Danub*. 2012 Mar;24(1):90–3. http://www.hdbp.org/psychiatria_danubina/pdf/dnb_vol24_no1/dnb_vol24_no1_90.pdf . [PubMed: 22447092]
19. Steers M, Wickham R, Acitelli L. Seeing Everyone Else's Highlight Reels: How Facebook Usage is Linked to Depressive Symptoms. *Journal of Social and Clinical Psychology*. 2014 Oct;33(8):701–731. doi: 10.1521/jscp.2014.33.8.701. doi: 10.1521/jscp.2014.33.8.701. [CrossRef: 10.1521/jscp.2014.33.8.701] [CrossRef: 10.1521/jscp.2014.33.8.701]
20. Martínez-Arán Anabel, Vieta E, Reinares M, Colom F, Torrent C, Sánchez-Moreno Jose, Benabarre A, Goikolea JM, Comes M, Salamero M. Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder. *Am J Psychiatry*. 2004 Feb;161(2):262–70. doi: 10.1176/appi.ajp.161.2.262. [PubMed: 14754775] [CrossRef: 10.1176/appi.ajp.161.2.262]
21. Huang A. Similarity Measures for Text Document Clustering. New Zealand Computer Science Research Student Conference 2008; 2008; New Zealand. 2008. http://nzcsrsc08.canterbury.ac.nz/site/proceedings/Individual_Papers/pg049_Similarity_Measures_for_Text_Document_Clustering.pdf .
22. Singhal A. Modern information retrieval: A brief overview. *IEEE Data Eng Bull*. 2001;24(4):43. <http://singhal.info/ieee2001.pdf> .
23. Katalapudi R, Chellappan S, Montgomery F, Wunsch D, Lutzen K. Associating Internet Usage with Depressive Behavior Among College Students. *IEEE Technol. Soc. Mag*. 2012 Dec;31(4):73–80. doi: 10.1109/MTS.2012.2225462. <http://ieeexplore.ieee.org/document/6387969/> [CrossRef: 10.1109/MTS.2012.2225462]
24. Dixon T, Kravariti E, Frith C, Murray RM, McGuire PK. Effect of symptoms on executive function in bipolar illness. *Psychol Med*. 2004 Jul;34(5):811–21. [PubMed: 15500302]
25. Ryan KA, Vederman AC, McFadden EM, Weldon AL, Kamali M, Langenecker SA, McInnis MG. Differential executive functioning performance by phase of bipolar disorder. *Bipolar Disord*. 2012 Aug;14(5):527–36. doi: 10.1111/j.1399-5618.2012.01032.x. <http://europepmc.org/abstract/MED/22834461> . [PMCID: PMC3773478] [PubMed: 22834461] [CrossRef: 10.1111/j.1399-5618.2012.01032.x]
26. Courvoisier DS, Eid M, Lischetzke T, Schreiber WH. Psychometric properties of a computerized mobile phone method for assessing mood in daily life. *Emotion*. 2010 Feb;10(1):115–24. doi: 10.1037/a0017813.2010-01983-003 [PubMed: 20141308] [CrossRef: 10.1037/a0017813]
27. Kuntsche E, Labhart F. Using Personal Cell Phones for Ecological Momentary Assessment. *European Psychologist*. 2013 Jan;18(1):3–11. doi: 10.1027/1016-9040/a000127. <http://econtent.hogrefe.com/doi/abs/10.1027/1016-9040/a000127> . [PMCID: PMC3839619] [PubMed: 24285917] [CrossRef: 10.1027/1016-9040/a000127]
28. Bauhoff S. Systematic self-report bias in health data: impact on estimating cross-sectional and treatment effects. *Health Serv Outcomes Res Method*. 2011 Feb 9;11(1-2):44–53. doi: 10.1007/s10742-011-0069-3. [CrossRef: 10.1007/s10742-011-0069-3]
29. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. *Psychiatry Res*. 2014 Jun 30;217(1-2):124–7. doi: 10.1016/j.psychres.2014.03.009.S0165-1781(14)00187-5 [PubMed: 24679993] [CrossRef: 10.1016/j.psychres.2014.03.009]
30. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, Mohr DC. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J Med Internet Res*. 2015 Jul 15;17(7):e175. doi: 10.2196/jmir.4273. <http://www.jmir.org/2015/7/e175/> v17i7e175 [PMCID: PMC4526997] [PubMed: 26180009] [CrossRef: 10.2196/jmir.4273]
31. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile Phone-Based Unobtrusive Ecological Momentary Assessment of Day-to-Day Mood: An Explorative Study. *J Med Internet Res*. 2016 Mar 29;18(3):e72. doi: 10.2196/jmir.5505. <http://www.jmir.org/2016/3/e72/> v18i3e72 [PMCID: PMC4829730] [PubMed: 27025287] [CrossRef: 10.2196/jmir.5505]
32. Kolakowska A. A review of emotion recognition methods based on keystroke dynamics and mouse movements. 6th International Conference on Human System Interactions (HSI); 2013; Sopot, Poland. 2013. [CrossRef: 10.1109/HSI.2013.6577879]
33. Lee H, Choi YS, Lee S, Park IP. Towards Unobtrusive Emotion Recognition for Affective Social Communication. 2012 IEEE Consumer Communications and Networking Conference (CCNC); 2012; Las Vegas, NV, USA. 2012.

34. Ghosh S, Ganguly N, Mitra B, De P. TapSense: Combining Self-Report Patterns and Typing Characteristics for Smartphone based Emotion Detection. International Conference on Human-Computer Interaction with Mobile Devices and Services; 2017; Vienna, Austria. 2017. <http://dl.acm.org/citation.cfm?doid=3098279.3098564> .

Figure 1



Screenshot of the BiAffect keyboard (keyboard design derived from AnySoftKeyboard by Menny Evan-Danan and licensed under Apache License 2.0.).

Table 1

Subject characteristics.

Characteristics	Value
Age in years, mean (SD)	48.67 (9.63)
Female gender, n (%)	8 (89)
Years of education, mean (SD)	14.00 (2.12)
Diagnosis, n (%)	
	Bipolar I
	Bipolar I with rapid cycling
	Bipolar II, recurrent depressive episodes, with rapid cycling
Number of keystrokes, mean (SD)	69,627 (57,477)
Number of accelerometer readings, mean (SD)	740,099 (47,165)
Weeks of activity, mean (SD)	7.70 (0.70)
Initial HDRS ^a -17 item, mean (SD)	11.90 (6.17)
Final HDRS-17 item, mean (SD)	11.11 (5.49)
Initial YMRS ^b , mean (SD)	7.56 (5.00)
Final YMRS, mean (SD)	6.67 (4.03)

^aHDRS: Hamilton Depression Rating Scale.^bYMRS: Young Mania Rating Scale.**Table 2**

Predictor variable definitions.

Predictor variable	Definition
Average interkey delay	The average time between keystrokes measured in seconds
Backspace ratio	Number of backspace keypresses divided by total keypresses
Autocorrect rate	Number of autocorrect events divided by total keypresses
Circadian baseline similarity	The cosine-based similarity between the hourly distribution of keypresses/week and the hourly distribution of keypresses/week
Average accelerometer displacement	Square root of sum of squares of accelerometer displacement along each coordinate (x, y, z) as $\sqrt{x^2+y^2+z^2}$
Average session length	Length of sessions in seconds averaged over the week
Session count	Number of sessions: A session begins when a keypress is initiated and ≥ 5 s has elapsed since the last key was pressed.

Table 3

Variable summary statistics.

Statistics	Mean (SD)	Min	Max	Average number of observations
Average accelerometer displacement (m/s ²)	9.56 (0.22)	9.07	9.99	104,076
Average interkey delay (s)	0.69 (0.36)	0.30	1.95	9780
Backspace rate	0.093 (0.050)	0.0070	0.27	982 ^a
Autocorrect rate	0.10 (0.033)	0.011	0.14	1021 ^a
Average session length (s)	19.63 (6.12)	6.90	31.51	—
Session count	241.13 (159.33)	9	804	—
Circadian baseline similarity	0.77 (0.17)	0.14	0.96	—
HDRS ^b 17-item	11.83 (6.29)	0	25	—
YMRS ^c	5.64 (4.87)	0	20	—
Natural log of YMRS	1.60 (0.82)	0	3.045	—

^aAverage number of backspace or autocorrect events.^bHDRS: Hamilton Depression Rating Scale.^cYMRS: Young Mania Rating Scale.**Table 4**

Fixed effects estimates of regression models.

Scaled predictors	17-item Hamilton Depression Rating Scale		Natural log of You
	Linear mixed-effects (95% CI)	<i>P</i> value	Ordinary least squares
Average accelerometer displacement	3.20 (1.20 to 5.21)	.0017	0.39 (0.15 to 0.64)
Average interkey delay	2.88 (0.42 to 5.35)	.022	0.13 (−0.19 to 0.44)
Backspace ratio	−0.01 (−1.53 to 1.52)	.99	−0.30 (−0.53 to −0.07)
Autocorrect rate	2.67 (0.87 to 4.47)	.0036	0.06 (−0.17 to 0.29)
Average session length	−1.16 (−2.71 to 0.39)	.14	−0.04 (−0.24 to 0.16)
Session count	2.18 (0.77 to 3.56)	.0025	−0.04 (−0.28 to 0.20)
Circadian baseline similarity	0.34 (−1.07 to 1.75)	.64	0.03 (−0.22 to 0.27)
Constant	11.77 (9.80 to 13.74)	<.001	1.60 (1.43 to 1.78)

Table 5

Summary of regression results.

Scaled predictors	17-item Hamilton Depression Rating Scale	Natural log of Young
	Linear mixed-effects	Ordinary least squares
Observations	64	64
Multiple R^2	—	.34
Adjusted R^2	—	.26
Conditional R^2	.63	—
Marginal R^2	.41	—
Log likelihood	-179.65	—
Residual standard error	—	.71 ^a
Chi-square statistic or F statistic	17.6 ^b	4.1 ^c

^a $df=56$ ^b $\chi^2_7, P=.014$.^c $F_{7,56}, P=.0011$.