Check for updates

OPEN ACCESS

EDITED BY Anthony Ahmed, Cornell University, United States

REVIEWED BY Ellen E. Lee, University of California, San Diego, United States Matthew J. Hoptman, Nathan S. Kline Institute for Psychiatric Research, United States Tyler Pia,

Cornell University, United States

*CORRESPONDENCE Olusola Ajilore Moajilore@uic.edu

RECEIVED 09 May 2024 ACCEPTED 03 February 2025 PUBLISHED 06 March 2025

CITATION

Ajilore O, Bark JS, Demos AP, Zulueta J, Stange J, Duffecy J, Hussain F, Langenecker SA, Nelson P, Ryan K, McInnis MG and Leow A (2025) Assessment of cognitive function in bipolar disorder with passive smartphone keystroke metadata: a BiAffect digital phenotyping study. *Front. Psychiatry* 16:1430303. doi: 10.3389/fpsyt.2025.1430303

COPYRIGHT

© 2025 Ajilore, Bark, Demos, Zulueta, Stange, Duffecy, Hussain, Langenecker, Nelson, Ryan, McInnis and Leow. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Assessment of cognitive function in bipolar disorder with passive smartphone keystroke metadata: a BiAffect digital phenotyping study

Olusola Ajilore^{1*}, John S. Bark¹, Alexander P. Demos², John Zulueta¹, Jonathan Stange³, Jennifer Duffecy¹, Faraz Hussain¹, Scott A. Langenecker⁴, Peter Nelson⁵, Kelly Ryan⁶, Melvin G. McInnis⁶ and Alex Leow^{1,5}

¹Department of Psychiatry, University of Illinois College of Medicine, Chicago, IL, United States, ²Department of Psychology, University of Illinois Chicago, Chicago, IL, United States, ³Department of Psychology and Psychiatry and the Behavioral Sciences, University of Southern California, Los Angeles, CA, United States, ⁴Department of Psychiatry, The Ohio State University, Columbus, OH, United States, ⁵Department of Bioengineering, University of Illinois Chicago, Chicago, IL, United States, ⁶Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States

Background: Cognitive dysfunction in bipolar disorder persists in the euthymic state and has been shown to be associated with a number of negative sequelae including treatment resistance and increased risk of relapse. There has been recent attention on digital phenotyping and passive sensing through smart, connected devices to probe cognition in real-world settings. BiAffect is a custom-built smartphone keyboard that captures keystroke metadata ('how you type, not what you type'). In previous studies, our group has demonstrated that BiAffect-derived keystroke metadata is associated with cognitive domains like processing speed. For the present study, we hypothesized that typing metadata would be significantly associated with executive function and planning.

Methods: 18 participants with bipolar disorder and 12 healthy comparison participants from the Prechter Longitudinal Study of Bipolar Disorder at the University of Michigan were provided a mobile phone with a customized keyboard that passively collected keystroke metadata. Participants also completed a neuropsychological battery including the Tower of London task. Irregularities in typing and times to make a move on the Tower of London task were compared using sample and Shannon entropy, respectively.

Results: Participants with bipolar disorder had significant increases in entropy in typing (p = .005, d = -1.28) and entropy of Tower of London move times (p = .029, d = -.84). Furthermore, typing entropy was significantly associated with irregularity in Tower of London moves in participants (r = .59, p = .006), as well as variability of clinician-rated depressive symptoms and self-rated impulsive actions and feelings.

Conclusions: This pilot study demonstrates that passive, unobtrusive smartphone keystroke metadata can be used to probe cognitive function and dysfunction in bipolar disorder, revealing multi-scalar behavioral features accessible through digital assays

KEYWORDS

digital biomarkers, bipolar disorder, executive function, smartphone, digital phenotyping

Introduction

Cognitive dysfunction in bipolar disorder has been shown to be present even in the euthymic state (1). Cognitive dysfunction in bipolar disorder has been associated with poor treatment response, increased risk of relapse (2), and comorbid substance abuse (3). In bipolar depressive disorders, patients demonstrate executive function deficits such as impaired response inhibition and difficulties with set-shifting. These cognitive deficits are often associated with illness course and poor treatment response. In addition, the further implications of cognitive dysfunction have been demonstrated in first episode bipolar patients where better performance on frontal/executive tasks was significantly correlated with faster time to recovery (4) and work status (5).

The literature on cognitive dysfunction demonstrates that not only can these deficits occur during an acute episode and impact treatment response but can also occur during euthymic states and predict symptom recurrence and relapse. Furthermore, these deficits broadly have a negative impact of emotion regulation across the spectrum of mood disorders, as well as in patients with bipolar disorder (6-8). Additionally, targeting cognitive dysfunction can augment current treatment modalities and prevent future mood episodes (9). Because of this, monitoring cognitive performance in patients with mood disorders has tremendous clinical implications. However, traditional cognitive assessment can be expensive, time-consuming, and temporally inexact, thus limiting its use and widespread implementation. These limitations demonstrate the need for passive, unobtrusive, objective assessment of cognitive function in the context of mood disorders. There has been recent attention on digital phenotyping and passive sensing through smart, connected devices to probe cognition in real-world settings. Our passive sensing, BiAffect, is a custom-built smartphone keyboard that captures keystroke metadata ('how you type, not what you type') (10). In previous studies, our group has demonstrated that BiAffect-derived keystroke metadata is associated with mood symptom severity and can be used prospectively to predict changes in mood in patients with bipolar disorder (11). For the present study, we hypothesized that typing metadata would be significantly associated with executive function measured with traditional cognitive testing in participants with bipolar disorder.

In order to compare typing metadata to traditional cognitive testing, we employed a dynamical systems approach to time-series derived from typing metadata and cognitive performance by utilizing sample entropy analysis (12, 13). Sample entropy analysis has been successfully applied to analyzing time series data from several clinical domains, including but not limited to bipolar disorder to predict changes in mood self-report (4), motor activity (7), as well as motor activity in depression (8), and a range of the other physiological disorders such as heart disease (5). In taking this dynamical systems approach, we were also interested in examining how typing dynamics relates to variability of depressive symptoms and impulsivity, as our previous study demonstrated that instability of mood ratings were associated with worsening depressive symptoms (11).

For the present study, we hypothesized that if irregularity of move times on the Tower of London task is reflective of poor planning or impulsivity, participants with bipolar disorder will demonstrate higher levels of entropy in move times compared to healthy comparison participants. Furthermore, we hypothesized that higher entropy of move times on the Tower of London task would be associated with higher levels of entropy in typing dynamics for participants with bipolar disorder (Figure 1).

Materials and methods

Participant recruitment

18 participants with bipolar disorder (12 with bipolar I, 6 with bipolar II) and 12 healthy comparison participants from the Prechter Longitudinal Study of Bipolar Disorder at the University of Michigan (14) completed informed consent in accordance with the Declaration of Helsinki and were provided a mobile phone with the BiAffect app. Information regarding data collection from this cohort have been described in more detail in previous publications (10, 11, 15).

Cognitive assessment

Participants completed a custom made cognitive task on the smartphone modeled after the Tower of London task (16), a test developed to measure planning and problem solving. In the custom smartphone task, participants were to replicate the top configuration of colored, stacked balls by making the bottom



pegboard look the same (Figure 2). The top pegboard was predetermined (includes multiple versions) and the bottom pegboard is manipulated by the participant. The bottom configuration of balls are touch-sensitive so that the participant can move/drag one ball from one peg and drop/place onto another peg. The pegs vary in height so that the tallest peg on the left can hold up to three balls, the middle can hold up to two balls, and the short peg can only hold one ball. For each trial, there is a preset configuration for the bottom configuration for all moves. There are set rules for moving the balls: 1) can only move one ball at a time; 2) cannot put more balls on a pole than can fit; 3) cannot move a bottom ball without moving the ball on top first. Participants are recorded on how long it takes them to complete the bottom configuration to match the top and the number of moves they make (total time to complete configuration, the time it takes move by move, and number of moves it takes to match the top



FIGURE 2

Tower of London task. Instructions that were given before starting the task "You must make the bottom board match the top board by moving the balls. You touch a ball with your finger and drag it to the new location. You are not to place more balls on a peg than it will hold; You can place 3 balls on the far-left peg, 2 balls on the middle, and only 1 ball on the right peg. You can only move one ball at a time. You must move the ball on top in order to move the ones below it. Tap BEGIN and complete the task as quickly as you can and in as few of moves as you can." If participants did not finish the task within 2 minutes, they received this prompt "That was a good try. Let's go to the next one.".

configuration. The minimum number of moves for each problem was preset at 5 and 6, defined to be a "moderate" difficulty level, each one to be given randomly in the morning ecological momentary assessment (EMA) session and in the evening EMA session. For example, they were given a 5- and a 6-move problem in the morning. We administered two problems so that the task could remain brief (as opposed to up to 30 moves of increasing difficulty, ranging from 10 moves to 32 moves on variant versions of the Tower of London) and include moderate difficulty problems to obtain a range of move times and errors. A total of 11 unique configurations of 5-move problems and 11 unique configurations of 6-move problems were randomly assigned so that no configuration was repeated until all versions were completed, thus minimizing any learning effects. Thus, participants completed a total of 22 problems across all sessions.

BiAffect and ecological momentary assessment

Participants used BiAffect-equipped Android smartphones and completed a 17-day baseline EMA period during which they were prompted, once daily, to rate their mood, energy, rapidity of thinking, and impulsive feelings and actions using a visual analog scale of 0 to 100 (17, 18). Three participants had less than 17 days collected (11, 15 and 16 days collected). Simultaneously, BiAffect unobtrusively collected typing kinematics metadata as participants interacted with their phones as usual, allowing us to extract average typing speed (measured using time since last key or interkey delay) within each day in their natural environments. Participants also completed a baseline clinician-rated Hamilton Rating Scale for Depression (HAM-D) (19) at study entry. Variability of EMA and HAM-D measures were obtained using the standard deviation.

Entropy analysis

To calculate entropy for the Tower of London task, Shannon entropy (20) was calculated on move times as a function of trial order across all sessions. Shannon entropy is a measure of uncertainty or randomness in a complex system. Shannon entropy was selected due to Tower of London move times being measured on the order of seconds and the length of the total timeseries. For calculation of entropy from typing data, sample entropy (21) was calculated based on the interkey delay as a function of keypress order across all keypresses with interkey delays less than 96 seconds (which represents delays between letters, words, sentences, and conversations).

Statistical analysis

t-tests were used to analyze group differences for continuous variables and the chi-square test was used for categorial variables. Pearson's correlations were used to analyze the association of typing entropy and Tower of London entropy, as well as typing entropy and self-rated EMA measures.

Results

Participant characteristics

Clinical, demographic and Tower of London performance data are summarized in Table 1. This cohort was derived from the larger Prechter Longitudinal Study which was 85% Caucasian with bipolar disorder onset at the average age of 17 years (14). There was no significant difference in age, sex, education, or IQ (all p > .05). On the Tower of London task, participants with bipolar disorder took longer to complete the task and had more total moves but these differences were not statistically significant (p = .16 and p = .21, respectively) after controlling for age and education.

Entropy analysis

Participants with bipolar disorder had significant increases in entropy of interkey delay times (typing entropy) (p = .005, d = -1.28) and entropy of Tower of London move times (p = .029, d = -.84) (Figure 3). Typing entropy was significantly correlated with Tower of London entropy (Figure 4, r = .59, p = .006), variability in depressive symptoms as measured by the HAM-D (r = .57, p = .009), and variability in self-rated impulsive actions (r = .52, p = .04) and feelings (r = .59, p = .02). There were no significant associations of typing entropy with the average values of EMA ratings of mood, energy, impulsive actions/feelings or racing thoughts. There was a significant association of Tower of London entropy with variability of self-rated impulsive actions (r = .43, p = .03) but no significant associations with variability of impulsive feelings or average levels of impulsive feelings or actions.

Discussion

In this study, we used an innovative approach leveraging entropy to measure poor planning and impulsivity actively (Tower of London) and passively (typing), associated with cognitive dysfunction in the context of bipolar disorder. We found that Tower of London move time entropy was significantly higher in participants with bipolar disorder compared to controls TABLE 1 Clinical, cognitive, and demographic participant characteristics.

	Control (n =12)	Bipolar (n=18)	
Age [range]	44.9 (9.0) [30 - 61]	47.4 (10.6) [31 - 63]	t (40) = 1.1 p = .28
Sex (M/F)	4/8	6/12	$\begin{array}{l} \chi 2 = 0.1 \\ P = .92 \end{array}$
Education (years)	16 (1.3)	15.8 (1.9)	t(40) = .70 p = .49
WAIS-IQ	108.8 (11.4)	108.5 (7.4)	t(29) =05 p = .97
Mean HAM-D	0.7 (0.9)	10.7 (5.4)	t(29) = 6.5 p <.001
TOL duration (seconds)	15.7 (9.0)	18.6 (11.3)	t(2242) = -4.5 p <.001
Total TOL legal moves	7.9 (3.3)	8.4 (3.9)	t(2242) = -2.9 p <.001

All data are represented as means (standard deviations). HAM-D, Hamilton Rating Scale for Depression; TOL, Tower of London.

and this was also reflected in significantly higher multiscale entropy in typing speed in participants with bipolar participants. Typing entropy was also positively associated with variability in depressive symptoms and self-rated impulsivity. While our application of entropy measures to cognitive performance metrics (like move times), our findings are consistent with previous studies that have linked lower entropy in motoric behavior to better planning. Interestingly in a review of research on skilled climbing expertise, the authors found that higher entropy was associated with higher climb times and reduced experience and skill with climbing (22). Similarly, it has been shown that higher entropy on a motor preservation task was associated with worse cognitive performance in adolescents with high-risk for psychosis (23).

With regards to the relationship between typing entropy and variability of mood, Stange et al. demonstrated that instability of typing speed and instability of mood both predicted future elevations of depressive symptoms in this same cohort (11). This highlights the potential of these instability metrics being used prospectively to prevent future mood episodes.

Implications for emotion regulation

We demonstrated that typing entropy was associated with the instability of self-reported impulsivity and mood symptoms. This is reflective of the relationship between impulsivity and emotion dysregulation seen not only in bipolar disorder but transdiagnostically as well. In a study of participants with borderline personality disorder, it was shown that greater levels of emotional dysregulation were associated with higher entropy of choice patterns on a strategic decision-making task (24). Additionally, in alignment with this paper's findings, a population-based study by Wen and colleagues demonstrated passive detection of impulsivity using the entropy of the smartphone activities (call logs, battery charging and screen status) (25). They found that motor impulsivity was positively correlated with entropy of the screen checking suggestive of poor planning. Of note,



impulsivity is one of the core emotion regulation deficits demonstrated in bipolar disorder (26). With regards to mood symptoms and emotion regulation, it has been shown that variability of cognitive performance was associated with negative affect (27).



Limitations

Our study was limited by a small sample size and a short study duration. Additionally, the small sample precluded analysis of potential confounders like digital literacy, medication usage and familiarity with the use of smartphone. Another limitation is inherent to mobile cognitive assessments, namely the lack of standardization and control of the testing environment. These potential confounders are being addressed in our ongoing BiAffect studies where we are collecting this important contextual information (ClinicalTrials.Gov NCT04358900).

Conclusions

In summary, our pilot study demonstrates that passive, unobtrusive smartphone keystroke metadata can be used to probe cognitive function and dysfunction in bipolar disorder, revealing multi-scalar behavioral features accessible through digital assays. Future work is needed in larger samples to validate these findings and to determine whether this type of passive digital biomarker data can be used as an early warning sign to enable just-in-time adaptive interventions and prevent future mood episodes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by University of Illinois at Chicago Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

OA: Conceptualization, Writing – original draft, Writing – review & editing. JB: Formal analysis, Writing – review & editing. AD: Conceptualization, Data curation, Formal analysis, Software, Writing – review & editing. JZ: Writing – review & editing. JS: Writing – review & editing. JD: Writing – review & editing. FH: Data curation, Writing – review & editing. SL: Conceptualization, Writing – review & editing. PN: Writing – review & editing. KR: Data curation, Writing – review & editing. MM: Funding acquisition, Writing – review & editing. AL: Conceptualization, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by the National Institutes of Mental Health: 1R01MH120168 and the Robert Wood Johnson Foundation.

Conflict of interest

Authors OA and AL are the co-founders of KeyWise AI.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Ajilore O, Vizueta N, Walshaw P, Zhan L, Leow A, Altshuler LL. Connectome signatures of neurocognitive abnormalities in euthymic bipolar I disorder. *J Psychiatr Res.* (2015) 68:37–44. doi: 10.1016/j.jpsychires.2015.05.017

2. Baune BT, Malhi GS. A review on the impact of cognitive dysfunction on social, occupational, and general functional outcomes in bipolar disorder. *Bipolar Disord*. (2015) 17 Suppl 2:41–55. doi: 10.1111/bdi.12341

3. Marshall DF, Walker SJ, Ryan KA, Kamali M, Saunders EFH, Weldon AL, et al. Greater executive and visual memory dysfunction in comorbid bipolar disorder and substance use disorder. *Psychiatry Res.* (2012) 200:252–7. doi: 10.1016/j.psychres.2012.06.013

4. Gruber SA, Rosso IM, Yurgelun-Todd D. Neuropsychological performance predicts clinical recovery in bipolar patients. *J Affect Disord.* (2008) 105:253–60. doi: 10.1016/j.jad.2007.04.014

5. Ryan KA, Vederman AC, Kamali M, Marshall D, Weldon AL, McInnis MG, et al. Emotion perception and executive functioning predict work status in euthymic bipolar disorder. *Psychiatry Res.* (2013) 210:472–8. doi: 10.1016/j.psychres.2013.06.031

6. Kjærstad HL, de Siqueira Rotenberg L, Macoveanu J, Coello K, Faurholt-Jepsen M, Bjertrup AJ, et al. Stable neural underpinnings of emotional cognition subgroups in patients newly diagnosed with bipolar disorder: A prospective fMRI study. *Bipolar Disord.* (2024) 26:556–69. doi: 10.1111/bdi.13444

7. Green MJ, Cahill CM, Malhi GS. The cognitive and neurophysiological basis of emotion dysregulation in bipolar disorder. *J Affect Disord.* (2007) 103:29-42. doi: 10.1016/j.jad.2007.01.024

8. Afriyie-Agyemang Y, Bertocci MA, Iyengar S, Stiffler RS, Bonar LK, Aslam HA, et al. Lifetime depression and mania/hypomania risk predicted by neural markers in three independent young adult samples during working memory and emotional regulation. *Mol Psychiatry*. (2024) 30(3):870–80. doi: 10.1038/s41380-024-02702-6

9. Tsapekos D, Strawbridge R, Cella M, Young AH, Wykes T. Does cognitive improvement translate into functional changes? Exploring the transfer mechanisms of cognitive remediation therapy for euthymic people with bipolar disorder. *psychol Med.* (2021) 53:936–44. doi: 10.1017/S0033291721002336

10. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: A biAffect digital phenotyping study. *J Med Internet Res.* (2018) 20:e241. doi: 10.2196/jmir.9775

11. Stange JP, Zulueta J, Langenecker SA, Ryan KA, Piscitello A, Duffecy J, et al. Let your fingers do the talking: Passive typing instability predicts future mood outcomes. *Bipolar Disord*. (2018) 20:285–8. doi: 10.1111/bdi.2018.20.issue-3

12. Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol.* (2002) 283: R789–97. doi: 10.1152/ajpregu.00069.2002

13. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. (2000) 101:E215–20. doi: 10.1161/01.CIR.101.23.e215

14. McInnis MG, Assari S, Kamali M, Ryan K, Langenecker SA, Saunders EFH, et al. Cohort profile: the heinz C. Prechter longitudinal study of bipolar disorder. *Int J Epidemiol.* (2018) 47:28–28n. doi: 10.1093/ije/dyx229

15. Ross MK, Demos AP, Zulueta J, Piscitello A, Langenecker SA, McInnis M, et al. Naturalistic smartphone keyboard typing reflects processing speed and executive function. *Brain Behav.* (2021) 11:e2363. doi: 10.1002/brb3.v11.11

16. Shallice T. Specific impairments of planning. Philos Trans R Soc London B Biol Sci. (1982) 298:199–209. doi: 10.1098/rstb.1982.0082

17. Schwartz S, Schultz S, Reider A, Saunders EF. Daily mood monitoring of symptoms using smartphones in bipolar disorder: A pilot study assessing the feasibility of ecological momentary assessment. *J Affect Disord.* (2016) 191:88–93. doi: 10.1016/j.jad.2015.11.013

18. Ryan KA, Babu P, Easter R, Saunders E, Lee AJ, Klasnja P, et al. A smartphone app to monitor mood symptoms in bipolar disorder: development and usability study. *JMIR Ment Health.* (2020) 7:e19476. doi: 10.2196/19476

19. Hamilton M. A rating scale for depression. J Neurol Neurosurgery Psychiatry. (1960) 23:56. doi: 10.1136/jnnp.23.1.56

20. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* (1948) 27:379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

21. Richman JS, Lake DE, Moorman JR. Sample entropy. In: *Methods in enzymology*. Cambridge, MA, United States: Elsevier (2004). p. 172–84.

22. Orth D, Kerr G, Davids K, Seifert L. Analysis of relations between spatiotemporal movement regulation and performance of discrete actions reveals functionality in skilled climbing. *Front Psychol.* (2017) 8:1744. doi: 10.3389/fpsyg.2017.01744

23. Schreiber H, Stolz-Born G, Heinrich H, Kornhuber HH, Born J. Attention, cognition, and motor perseveration in adolescents at genetic risk for schizophrenia and control subjects. *Psychiatry Res.* (1992) 44:125–40. doi: 10.1016/0165-1781(92)90047-7

24. Parr AC, Calancie OG, Coe BC, Khalid-Khan S, Munoz DP. Impulsivity and emotional dysregulation predict choice behavior during a mixed-strategy game in

adolescents with borderline personality disorder. Front Neurosci. (2022) 15. doi: 10.3389/fnins.2021.667399

25. Wen H, Sobolev M, Vitale R, Kizer J, Pollak JP, Muench F, et al. mPulse mobile sensing model for passive detection of impulsive behavior: exploratory prediction study. *JMIR Ment Health*. (2021) 8:e25019. doi: 10.2196/25019

26. Miola A, Cattarinussi G, Antiga G, Caiolo S, Solmi M, Sambataro F. Difficulties in emotion regulation in bipolar disorder: A systematic review and meta-analysis. J Affect Disord. (2022) 302:352–60. doi: 10.1016/j.jad.2022.01.102

27. Brose A, Schmiedek F, Lövdén M, Lindenberger U. Daily variability in working memory is coupled with negative affect: The role of attention and motivation. *Emotion*. (2012) 12:605–17. doi: 10.1037/a0024436