



The Effects of Bipolar Disorder Risk on a Mobile Phone Keystroke Dynamics Based Biomarker of Brain Age

John Zulueta¹, Alexander Pantelis Demos², Claudia Vesel³, Mindy Ross⁴, Andrea Piscitello¹, Faraz Hussain¹, Scott A. Langenecker⁵, Melvin McInnis⁶, Peter Nelson⁷, Kelly Ryan⁶, Alex Leow^{1,3} and Olusola Ajilore^{1*}

¹ Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, United States, ² Department of Psychology, University of Illinois at Chicago, Chicago, IL, United States, ³ Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, United States, ⁴ Graduate College, University of Illinois at Chicago, Chicago, IL, United States, ⁵ Department of Psychiatry, The University of Utah, Salt Lake City, UT, United States, ⁶ Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States, ⁷ College of Engineering, University of Illinois at Chicago, Chicago, IL, United States

OPEN ACCESS

Edited by:

Ellen E. Lee,
University of California, San Diego,
United States

Reviewed by:

Rao Kosagisharaf,
Instituto de Investigaciones Científicas
y Servicios de Alta
Tecnología, Panama
Andrea Iaboni,
University Health Network, Canada

*Correspondence:

Olusola Ajilore
oajilore@uic.edu

Specialty section:

This article was submitted to
Aging Psychiatry,
a section of the journal
Frontiers in Psychiatry

Received: 09 July 2021

Accepted: 19 November 2021

Published: 22 December 2021

Citation:

Zulueta J, Demos AP, Vesel C,
Ross M, Piscitello A, Hussain F,
Langenecker SA, McInnis M, Nelson P,
Ryan K, Leow A and Ajilore O (2021)
The Effects of Bipolar Disorder Risk on
a Mobile Phone Keystroke Dynamics
Based Biomarker of Brain Age.
Front. Psychiatry 12:739022.
doi: 10.3389/fpsy.2021.739022

Background: Research by our group and others have demonstrated the feasibility of using mobile phone derived metadata to model mood and cognition. Given the effects of age and mood on cognitive performance, it was hypothesized that using such data a model could be built to predict chronological age and that differences between predicted age and actual age could be a marker of pathology.

Methods: These data were collected via the ongoing BiAffect study. Participants complete the Mood Disorders Questionnaire (MDQ), a screening questionnaire for bipolar disorder, and self-reported their birth year. Data were split into training and validation sets. Features derived from the smartphone kinematics were used to train random forest regression models to predict age. Prediction errors were compared between participants screening positive and negative on the MDQ.

Results: Three hundred forty-four participants had analyzable data of which 227 had positive screens for bipolar disorder and 117 had negative screens. The absolute prediction error tended to be lower for participants with positive screens (median 4.50 years) than those with negative screens (median 7.92 years) ($W = 508, p = 0.0049$). The raw prediction error tended to be lower for participants with negative screens (median = -5.95 years) than those with positive screens (median = 0.55 years) ($W = 1,037, p = 0.037$).

Conclusions: The tendency to underestimate the chronological age of participants screening negative for bipolar disorder compared to those screening positive is consistent with the finding that bipolar disorder may be associated with brain changes that could reflect pathological aging. This interesting result could also reflect that those who screen negative for bipolar disorder and who engaged in the study were more likely to have higher premorbid functioning. This work demonstrates that age-related changes may be detected via a passive smartphone kinematics based digital biomarker.

Keywords: digital biomarkers, bipolar disorder, brain age estimation, smartphone, digital phenotyping

INTRODUCTION

The development of biomarkers has long been a goal for psychiatry with the hope that these biomarkers may be able to facilitate early detection, diagnosis, and treatment selection—moving the field closer to a paradigm of precision medicine (1, 2). Aging is a heterogeneous process associated with increased risk of morbidity and mortality. It has been proposed that differences between biological age and chronological age may be indicative of pathology, and various phenomena have been investigated as potential aging biomarkers including telomere length (3), DNA methylation (4), and features derived from neuroimaging (5, 6).

In previous work, our group identified age associated effects on smartphone typing kinematics—specifically enhancement of the difference between midday typing speed and typing speed at the beginning and end of the day (7). These kinematic data were collected *via* the BiAffect platform which collects such data passively as participants use their smartphones in their day-to-day routines thus enabling the creation of ecologically valid and temporally associated markers of cognitive performance (8).

Bipolar disorder is a psychiatric disorder characterized by recurrent episodes of mood disturbances. It is associated with cognitive deficits during mood episodes, some of which remain during euthymia (9, 10). It has also been proposed that bipolar disorder may exacerbate age associated neuropathologic processes in a phenomenon termed neuroprogression (11, 12). In this study we investigated the hypothesis that cognitive changes associated with the disorder would be detectable *via* changes in typing kinematics. To do this, we leveraged the BiAffect platform taking advantage of the open enrollment of the project to obtain a large, heterogeneous sample. Rather than utilizing binary self-report of diagnosis to distinguish between healthy controls and participants with a bipolar spectrum disorder, we categorized participants based on screening status on the Mood Disorders Questionnaire (MDQ), a screening instrument for bipolar disorder (13) using standard cut-off scores with sensitivity of 61% and specificity of 88% (14). We then examined differences in the performance characteristics of age prediction between the groups to investigate smartphone kinematic based age prediction's utility as a digital biomarker.

METHODS

Data for this study was collected as part of the open science BiAffect project. This study began in March 2018 with enrollment open to all adults in the United States with an iOS based smartphone that supports the BiAffect app. As of the time of the writing of this manuscript, the study is ongoing. Its protocol has been approved by the University of Illinois at Chicago Institutional Review Board.

The data for this study was collected from March 2018 to February 2021. Subject enrollment and data collection were all performed within the BiAffect app. The app includes modules for participants to complete questionnaires and perform tasks designed to measure aspects of cognitive performance such as response inhibition, set shifting, and reaction time. Its core technology is a custom built keyboard designed to replace the

default keyboard. This keyboard collects typing related metadata including the type of keypress event (alphanumeric, backspace, autocorrection, etc.) and the associated timestamp. It does not collect the actual alphanumeric content. These data are then securely uploaded to the study server.

Participants completed the mood disorders questionnaire (MDQ) a screening instrument for bipolar disorder (13). The MDQs were typically completed at study entry. The performance characteristics of this instrument vary based on the setting but in general has been estimated to have relatively high specificity of 88% and adequate sensitivity of 61% (14) with a cut score of ≥ 7 . Participants also provided self-reports of whether they have a diagnosis of a bipolar spectrum disorder, their birth year, and their gender. Given the fact that MDQ performance characteristics are better characterized than the reliability of self-reported bipolar disorder diagnosis and given the high rates of participants not disclosing their diagnosis status, it was decided to utilize MDQ status as a feature of interest in our analysis.

Analysis was restricted to participants who had provided at least 12 weeks worth of typing data. This was determined by calculating the median number of keystrokes per day across the entire sample and then filtering accordingly.

Data Processing and Feature Engineering

Each subject's typing data was tokenized into sessions by grouping together consecutive keystroke events which have differences in timestamps of < 5 seconds. Metrics were calculated for each session and then summarized for each subject. For sample entropy calculations the following parameters were used: $m = 2$, $r = 0.2 * \text{the standard deviation}$ and $\tau = 1$. **Table 1** includes a description of these features. This data processing was performed via the pandas package, version 1.2.4 (15) for Python (Version 3.8.3).

Model Training and Assessment

Data were split into training and validation sets (75:25). Because of the collinearity among many of the features and the relative robustness of random forest models to collinearity (16), random forest models were used. Random forest models consist of a collection of decision trees whose individual predictions are aggregated to make a single prediction. They are a popular analytic tool in bioinformatics given their ability to model complex interactions (17). Two random forest regression models were trained using the caret and randomForest packages for R (18, 19). The *mtry* parameter determines the number of features that will be available for use when splitting nodes during the training of the model's decision trees. The *mtry* parameter was selected via a grid search of values ranging from 1 to 30 features using 10-fold cross-validation with 3 repeats. The *mtry* value which minimized the Root Mean Square Error (RMSE) was selected as the value used in the final models. The models were constructed in a stepwise fashion with the first model including only typing related features, and the second model included all features from the first with the addition of gender and MDQ screening status.

Each model's performance was assessed using the validation set to calculate RMSE, Breiman's pseudo R-squared, and median

TABLE 1 | Model features.

| Feature | Description |
|--|---|
| Mean_keypresses_per_session | Mean number of keypresses per session |
| Median_keypresses_per_session | Median number of keypresses per session |
| Standard_deviation_keypress_per_session | Standard deviation of keypresses per session |
| Median_absolute_deviation_keypress_per_session | Median absolute deviation of keypresses per session |
| Mean_interkey_time_mean | Mean of mean of interkey times per session |
| Median_interkey_time_mean | Median of mean interkey times per session |
| Standard_deviation_interkey_time_mean | Standard deviation of mean interkey times per session |
| Median_absolute_deviation_interkey_time_mean | Median absolute deviation of mean interkey times per session |
| Mean_interkey_time_median | Mean of median of interkey times per session |
| Median_interkey_time_median | Median of median interkey times per session |
| Standard_deviation_interkey_time_median | Standard deviation of median interkey times per session |
| Median_absolute_deviation_interkey_time_median | Median absolute deviation of median interkey times per session |
| Mean_autocorrect_rate | Mean autocorrect rate per session (# of autocorrect events / total # of keystrokes per session) |
| Median_autocorrect_rate | Median autocorrect rate per session (# of autocorrect events / total # of keystrokes per session) |
| Standard_deviation_autocorrect_rate | Standard deviation of autocorrect rate per session (# of autocorrect events / total # of keystrokes per session) |
| Median_absolute_deviation_autocorrect_rate | Median absolute deviation of autocorrect rate per session (# of autocorrect events / total # of keystrokes per session) |
| Mean_backspace_rate | Mean backspace rate per session (# of backspace events / total # of keystrokes per session) |
| Median_backspace_rate | Median backspace rate per session (# of backspace events / total # of keystrokes per session) |
| Standard_deviation_backspace_rate | Standard deviation of backspace rate per session (# of backspace events / total # of keystrokes per session) |
| Median_absolute_deviation_backspace_rate | Median absolute deviation of backspace rate per session (# of backspace events / total # of keystrokes per session) |
| Mean_session_length | Mean length of sessions in seconds |
| Median_session_length | Median length of sessions in seconds |
| Standard_deviation_session_length | Standard deviation of length of sessions in seconds |
| Median_absolute_deviation_session_length | Median absolute deviation of length of sessions in seconds |
| Sample_entropy_keypress | Sample entropy of # of keypresses per sessions |
| Sample_entropy_interkey_time_mean | Sample entropy of mean interkey times per session |
| Sample_entropy_interkey_time_median | Sample entropy of median interkey times per session |
| Sample_entropy_autocorrect_rate | Sample entropy of autocorrect rate per session |
| Sample_entropy_backspace_rate | Sample entropy of backspace rate per session |
| Sample_entropy_session_length | Sample entropy of session length in seconds |

absolute error. Differences in model performance testing were assessed using paired Wilcoxon tests of their absolute errors. Feature importance was assessed using out-of-bag changes in Mean Square Error (MSE). Accumulated Local Effects plots (ALE Plots) (20) were constructed for features which appeared important or interesting. These plots allow the visualization of the effect of individual features and the interaction of two features on the model's prediction. They are especially useful when features may be correlated. Differences within model performance between participants based on MDQ screen status were assessed using Wilcoxon tests comparing raw prediction error scores and absolute prediction error scores. All tests were two sided with a significance level of 0.05. Family-wise error rates were controlled using the Holm-Bonferroni method. All statistical testing was performed in R (Version 4.0.0).

RESULTS

A total of 344 participants met criteria for inclusion in this analysis: 117 with negative MDQ screens and 227 with positive

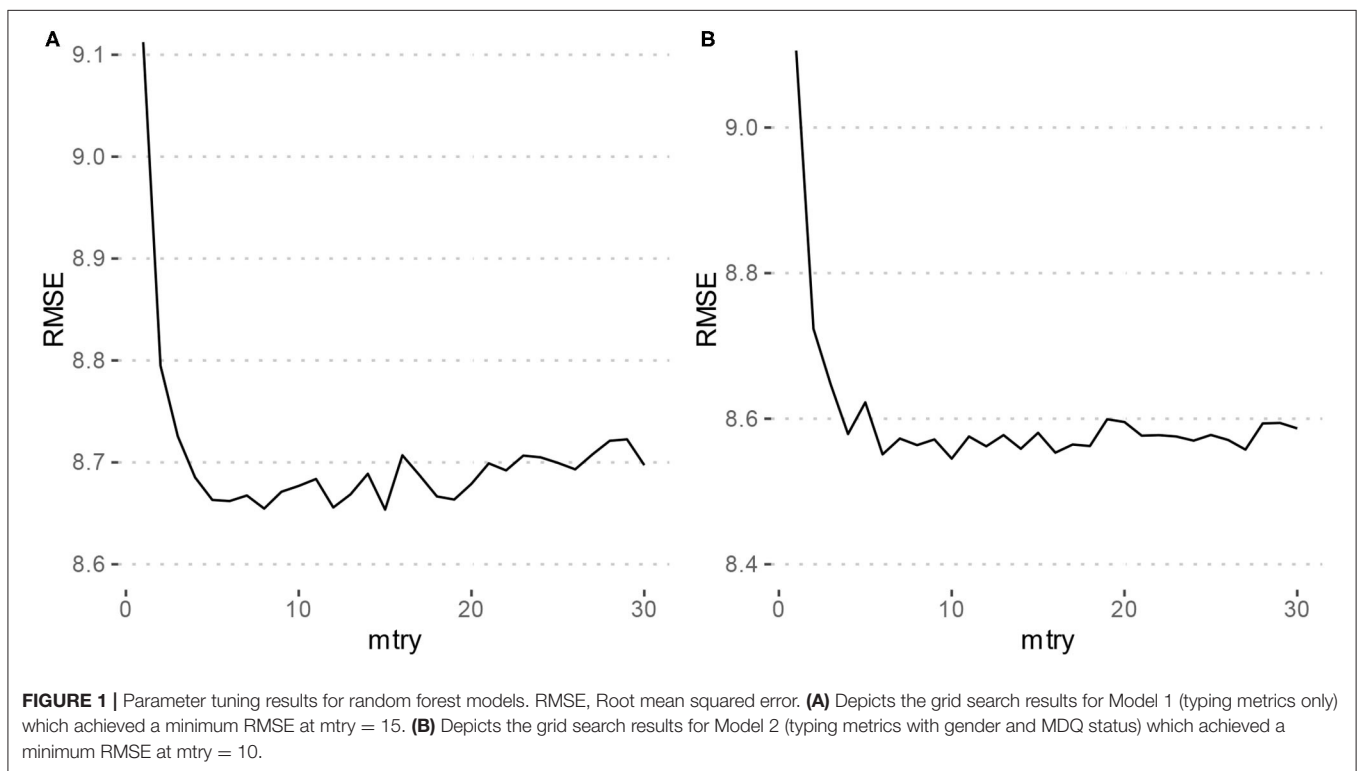
screens. As summarized in **Table 2**, the group with positive screens tended to have fewer males (Fisher's Exact $p = 0.0042$) and be younger than the positive screen group ($W = 15,887, p = 0.0028$). Compared to participants with positive MDQ screens, participants with negative screens had a lower rate of reporting a diagnosis of bipolar disorder, a higher rate of reporting no history of bipolar disorder, and also provided no diagnosis history at a lower rate (all comparisons Fisher's Exact $p < 0.001$). The participants with negative screens tended to have lower MDQ scores compared to those with positive screens ($W = 23,322, p < 0.001$) and have a greater total number of keypresses ($W = 15,098, p = 0.037$).

Using the criterion of minimizing RMSE for tuning the mtry parameter of the models, Model 1 which only included the typing metrics had an mtry = 15, and Model 2 which included the features of Model 1 as well as gender and MDQ status had an mtry = 10. **Figure 1** depicts the tuning results.

Using the training set, both Model 1 and Model 2 had an RMSE of 8.7 years. The Breiman Pseudo R-squared values were 0.56 and 0.57 for Models 1 and 2, respectively. The performance

TABLE 2 | Subject characteristics.

| | MDQ negative | MDQ positive | |
|---|-----------------|-----------------|--------------------------|
| # of participants | 117 | 227 | |
| % not male | 60% | 75% | $p = 0.0042$ |
| Age in years, mean (sd) | 41 (16) | 35 (11) | |
| Age in years, median (mad) | 39 (16) | 33 (12) | $W = 15,887, p = 0.0028$ |
| Age (min, max) | (20, 88) | (18, 70) | |
| Self-reports history of diagnosis with bipolar spectrum disorder | 24 (21%) | 115 (51%) | $p < 0.001$ |
| Self-reports no history of diagnosis with bipolar spectrum disorder | 66 (56%) | 29 (13%) | $p < 0.001$ |
| Does not provide any information regarding diagnosis of bipolar spectrum disorder | 27 (23%) | 83 (37%) | $p < 0.001$ |
| MDQ score, mean (sd) | 6 (4) | 12 (1) | $W = 23,322, p < 0.001$ |
| Total keypresses, mean (sd) | 37,027 (87,464) | 36,381 (71,262) | |
| Total keypresses, median (mad) | 7,600 (7,465) | 12,043 (12,682) | $W = 15,098, p = 0.037$ |



of these models using the validation set are described in **Table 3**. Using the validation set, Model 1 had an RMSE of 9.7, and Model 2 had an RMSE of 9.5. Breiman's Pseudo R-squared was 0.42 and 0.44 for Models 1 and 2, respectively. Model 1 had a median absolute error of 5.9 and Model 2 had a median absolute error of 5.5. This difference was not statistically significant ($V = 2,109, p = 0.21$).

Given the trend toward improved performance with the inclusion of gender and MDQ status as model features, analysis of feature importance and differences in prediction by MDQ status are presented only for Model 2.

Feature importance is depicted in **Figure 2**. Features whose exclusion from the model results in a larger increase in Mean Squared Error are considered more important. While this allows us to understand the relative importance of the features, it does not provide information on how each feature's value is associated with age. One method that allows for the examination of these relationships in random forest models is an ALE plot. Given that many of the most important features are different summaries of the same essential feature (e.g., interkey time), in plots A–D of **Figure 3** we present the ALE plots of four of the most important features: the median of mean interkey times, the mean session

TABLE 3 | Model performance comparison using the validation dataset.

| | Model 1 (only typing metrics) | Model 2 (typing metrics with MDQ status and gender) | |
|----------------------------|-------------------------------|---|-----------------------|
| RMSE | 9.7 | 9.5 | |
| Breiman's Pseudo R-Squared | 0.42 | 0.44 | |
| Median absolute error | 5.9 | 5.5 | $V = 2,109, p = 0.21$ |

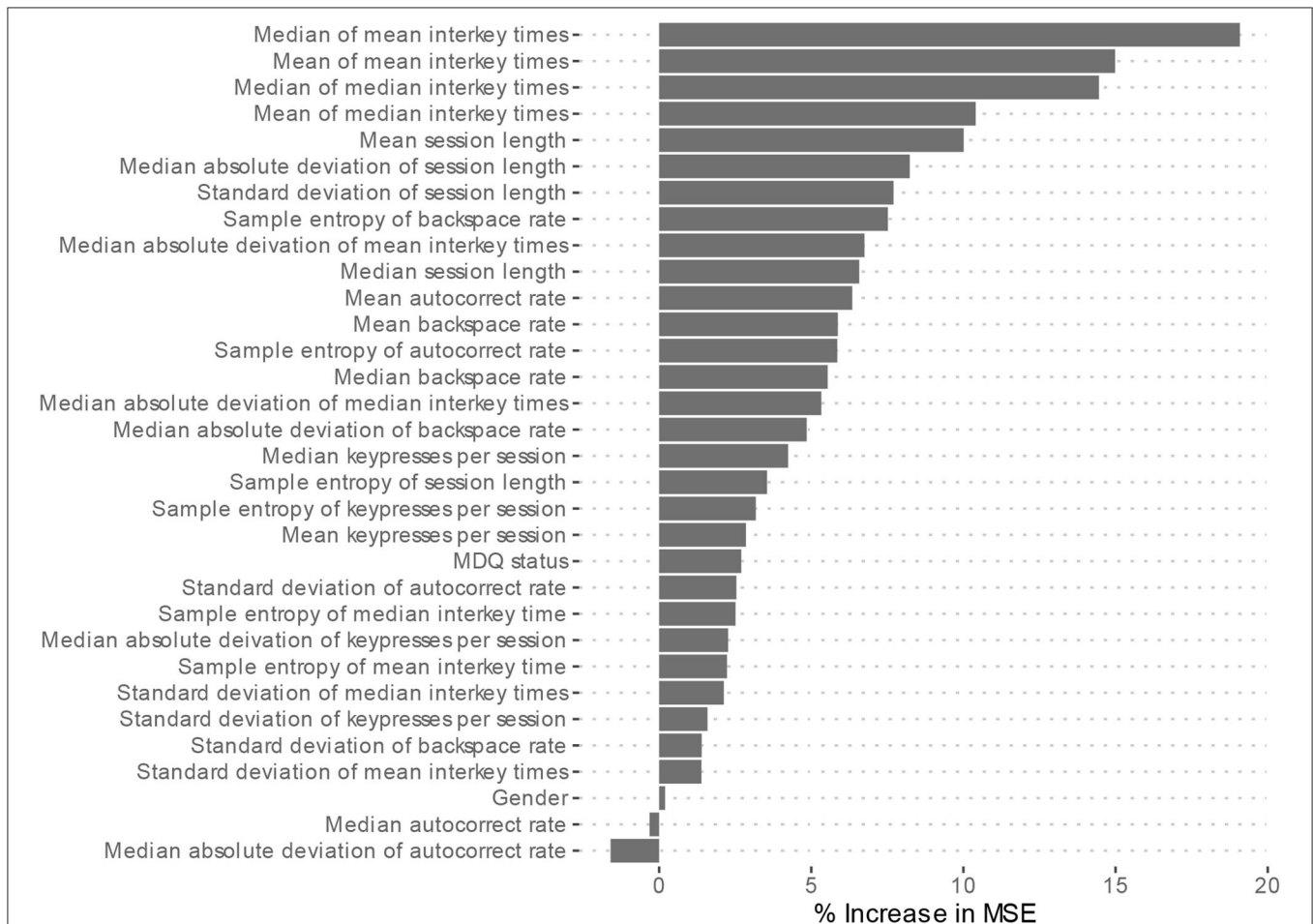


FIGURE 2 | Model 2 Feature importance. MSE, Mean square error. Higher increases in MSE indicated increased importance of the feature in predicting age but do not indicate directionality of the relationship.

length, the sample entropy of the backspace rate, and the mean backspace rate. Based on these plots, increased interkey time and session length are both generally associated with increased age; whereas, increased sample entropy of the backspace rate is associated with younger age, and the association between age and the mean backspace rate is not monotonic. Plots E and F of **Figure 3** depict the interaction between the median of mean interkey times and the mean session length and between the mean backspace rate and the sample entropy of the backspace rate, respectively. In these plots we see that the existence and directionality of linear trends between the predicted age and these features depend on the range of a second associated feature highlighting the complexity of the relationship between typing behaviors and predicted age.

The raw prediction error for age, i.e., how many years over or under the model predicted from the correct age, tended to be lower for participants with a negative screen (median = -5.95) than those with positive screens (median = 0.55) ($W = 1,037, p = 0.037$). The absolute prediction error, which measures the absolute deviation from the correct age, tended to be lower for participants with a positive screen (median = 4.50) than those with negative screens (median = 7.92) ($W = 508, p = 0.0049$). These comparisons are depicted with boxplots in **Figure 4**. The significant difference in absolute prediction error between the groups suggests the existence of an intrinsic difference between the groups in terms of how each group's typing behaviors relate to age, and the significant difference in raw errors specifically demonstrates that participants with a negative MDQ screen tend

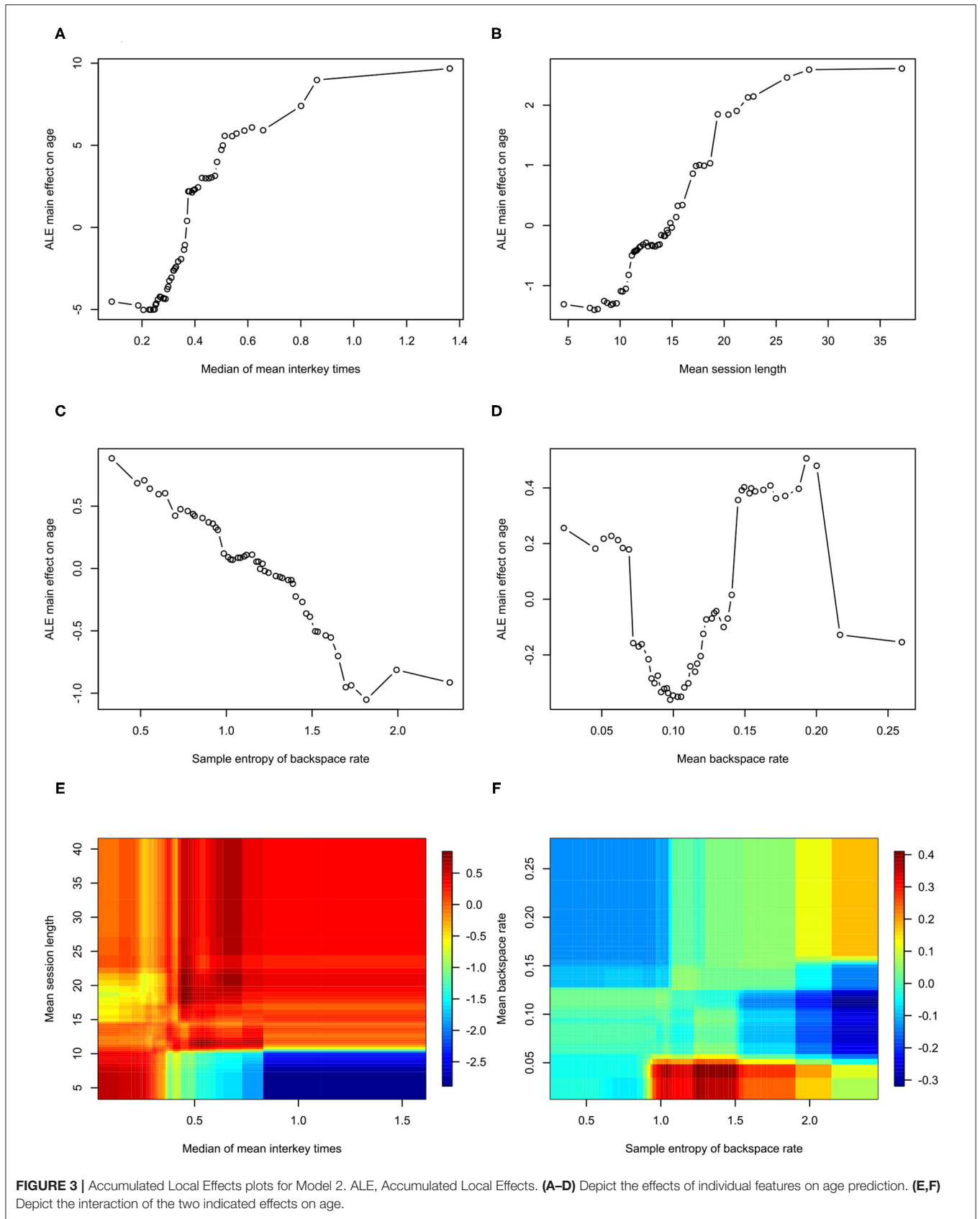
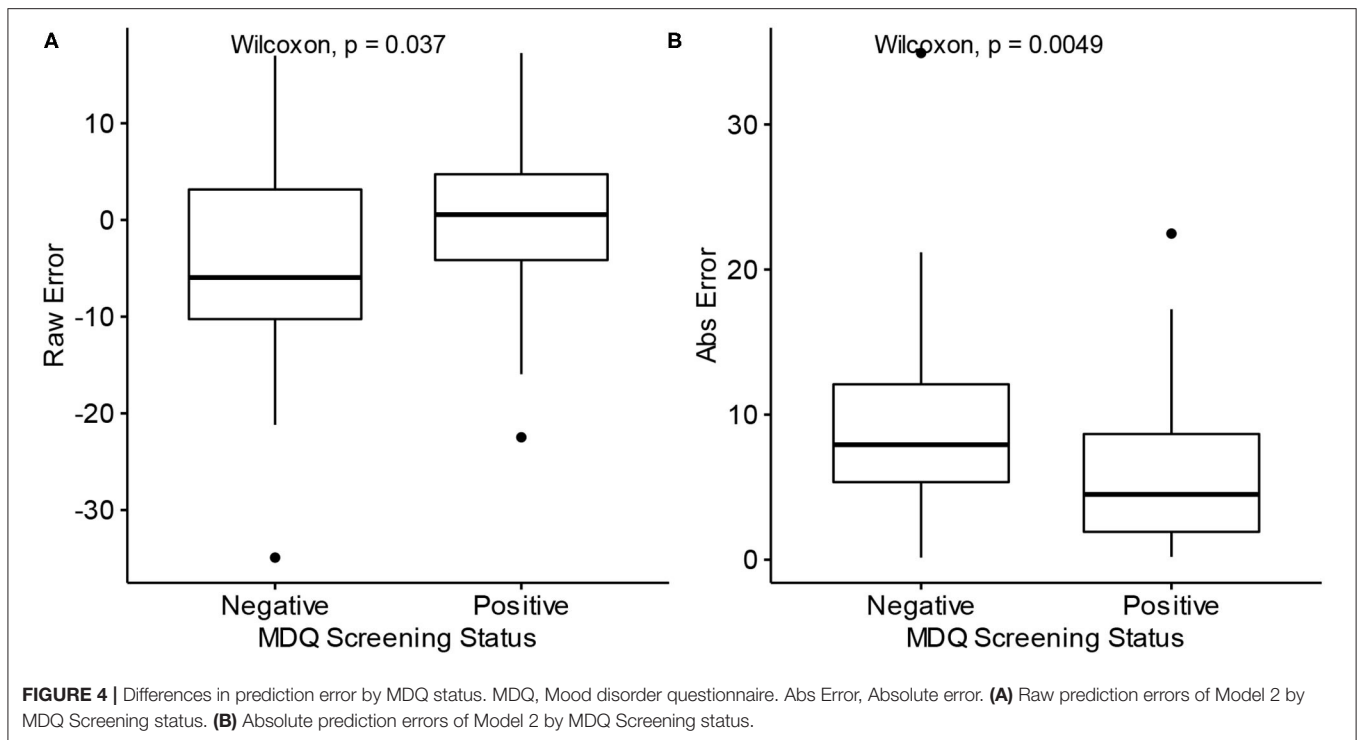


FIGURE 3 | Accumulated Local Effects plots for Model 2. ALE, Accumulated Local Effects. **(A–D)** Depict the effects of individual features on age prediction. **(E,F)** Depict the interaction of the two indicated effects on age.



to be predicted as younger may be consistent with the theory of bipolar disorder's association with neuroprogression.

DISCUSSION

Objective biomarkers of psychiatric pathology have the potential to transform the practice of psychiatry by providing clinicians more precise and reliable data to inform treatment decisions. In this study, we investigate the possibility of creating such a biomarker by using passively collected keyboard dynamics metadata derived from smartphone usage. This method of collection has the advantages of enabling high frequency sampling and perhaps even more importantly enabling such sampling to take place in people's normal day-to-day lives.

While there have been several studies which have used mobile phone derived metadata to predict demographic features such as age, these studies have tended to use features such as the number of calls and length of calls, the number of text messages, time of day of usage, and metrics derived from the networks of interactions (i.e., calls/text messages) between the users and other people (21–23). Further, these studies appear to have been focused on the utility of such methods for marketing applications, and in keeping with that aim, they used binned age groups and measured model performance based on correct classification of users to those groups. This makes comparing performance between these models and our regression oriented models difficult.

Although models of biological age are typically constructed by using a cohort of “healthy” participants to train a model which is then applied to participants with pathology, for this

study we trained our model on both healthy and non-healthy participants in order to make maximum use of the available data. This is a limitation that we plan to address in future studies via larger, more well-characterized samples. The significant difference in prediction accuracy between participants with positive and negative MDQ screens suggests that there may be some intrinsic difference in the pattern of typing between participants who are likely at elevated risk of having bipolar disorder compared to those without such risk. That participants with positive screens have a lower absolute prediction error may be a consequence of the fact that the training sample consisted mostly of participants with positive screens; however, it may also be consistent with the emerging finding that psychiatric disorders may be characterized by a decrease in complexity and variability of behavior, which makes the brain less adaptable to a constantly changing environment (24). Even if the difference is primarily driven by the imbalance between participants with positive and negative MDQ screens, the very fact that such a difference exists is notable in that it suggests that the psychiatric pathology associated with a positive screen produces detectable changes in mobile phone typing kinematics. That there was also a tendency for the model to underpredict the age of participants with negative MDQ screens is consistent with the concept that pathology is associated with the phenomenon of biological age exceeding chronological age (5).

With this study design, we were not able to include in our models other potential covariates of cognitive and motor performance which could affect age prediction errors. Such factors could include co-morbidities, medication status, specific psychiatric diagnoses and severity, and general facility with

phone usage. Given the tendency for bipolar disorder to be associated with a host of other co-morbidities (25), the between group differences found between participants with positive and negative MDQ screens could be due to bipolar disorder itself or some other combination of co-morbidities that is associated with people who are at increased risk of bipolar disorder that alters typing behaviors. Without such information, we cannot rule out the possibility that the differences in prediction accuracy between the participants with positive and negative MDQ screens are due to a disproportionate allocation of these characteristics which are not associated with the pathology associated with a positive MDQ screen. In future studies, we plan to collect data such as diagnoses, severity, and treatment status in order to create models which will properly attribute and quantify the effect of these variables on brain age biomarkers.

Examining which features are most important in predicting age, the most important features are measures of typing speed and the length of typing session. This is consistent with previous studies that have found correlations between age and typing performance (26–28). Plots A and B of **Figure 3** demonstrate that both interkey time and session length tend to be positively correlated with age – older age is associated with slower speed and longer session; however, in examining the interaction of these two features depicted in plot D of **Figure 3**, we see that for sessions under 10 s in length, there is actually a negative correlation between interkey time and age. One possible explanation is that in these short sessions the interkey time represents the pauses that occur in a rapid exchange of text messages with another person, and that for longer sessions the interkey time represents the pauses that occur in the composition of a longer body of text.

The relatively high importance of the sample entropy of backspace rates across sessions is an intriguing finding. This feature theoretically measures the complexity of participant backspace use. Based on its ALE plot, plot E of **Figure 3**, it appears to generally negatively correlated with age; however, notably if we examine the interaction of the sample entropy of the rate with the overall mean rate depicted in plot D of **Figure 3**, we see that at relatively low overall usage of backspace increased entropy is associated with younger age, but that at relatively high usage increased entropy is associated with older age. Several studies have examined how measures of complexity like entropy can be applied to functional imaging derived brain networks and how complexity changes with aging (29–31). With data we are collecting in one of our current studies we will be able to examine how neuroimaging derived measures of brain complexity are associated with the complexity of signals derived from the kinematics of smartphone usage (32).

Limitations

This study is limited by the fact that all subjective data were provided via self-report. A sample which consists of participants with clinically confirmed diagnoses would yield greater insight into the utility of digital biomarkers such as those described here in characterizing psychiatric disorders. Along those same lines, objective neuropsychological performance data would also have helped better contextualize our findings with existing literature in bipolar disorder. Although we included age and gender as

features in our model, a sample in which age and gender distributions are equivalent across the case and control groups would likely yield more robust findings.

Future Directions

Future directions of this research include attempting to replicate these findings in a more well-characterized sample and determining how this marker compares to other biomarkers such as neuroimaging based markers and other digital biomarkers. We also aim to investigate how differences in handedness (one-handed vs. two-handed typing) and distinguishing between different types of keystroke transitions (e.g., alphanumeric to alphanumeric vs. alphanumeric to backspace) might yield better performing models. Another potential line of investigation is determining whether differences between predicted age and chronological age by our model are associated with state level phenomena such as the severity of mood episodes.

CONCLUSION

Passively collected typing kinematics can be used to estimate biomarkers of brain age. The differences we found in this study between the performance of such a biomarker in participants with and without a positive screen for bipolar disorder—i.e., the tendency to underestimate the age of healthy participants—suggest that this biomarker may also be a useful marker of pathology. Further investigation to refine the model and determine its relation to other markers of pathology such as neuroimaging and neuropsychological testing is warranted.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Illinois at Chicago Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JZ did the analysis and wrote the manuscript. All other authors contributed data and edited the manuscript.

FUNDING

This study was partially funded by Mood Challenge for Research kit (AL and PN) and 1R01MH120168 (OA and AL).

ACKNOWLEDGMENTS

The BiAffect team would like to thank all the citizen scientists who have enrolled in our study and contributed their data.

REFERENCES

- McGorry PD. Early clinical phenotypes, clinical staging, and strategic biomarker research: building blocks for personalized psychiatry. *Biol Psychiatry*. (2013) 74:394–5. doi: 10.1016/j.biopsych.2013.07.004
- Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF, Berk M. The new field of “precision psychiatry.” *BMC Med*. (2017) 15:1–7. doi: 10.1186/s12916-017-0849-x
- Mather KA, Jorm AF, Parslow RA, Christensen H. Is telomere length a biomarker of aging? A review. *J Gerontol A Biol Sci Med Sci*. (2011) 66:202–13. doi: 10.1093/geron/gdq180
- Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol*. (2019) 20:1–24. doi: 10.1186/s13059-019-1824-y
- Cole JH, Marioni RE, Harris SE, Deary IJ. Brain age and other bodily ‘ages’: implications for neuropsychiatry. *Molecular Psychiatry*. (2019) 24:266–81. Available from: doi: 10.1038/s41380-018-0098-1
- Cole JH, Franke K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci*. (2017) 40:681–90. doi: 10.1016/j.tins.2017.10.001
- Vesel C, Rashidisabet H, Zulueta J, Stange JP, Duffecy J, Hussain F, et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: a biaffect iOS study. *J Am Med Informatics Assoc*. (2020) 27:1007–18. doi: 10.1093/jamia/ocaa057
- Biaffect. *The First Study on Mood and Cognition Using Mobile Typing Kinematic*. Available online at: <https://www.biaffect.com/> (accessed June 26, 2021).
- Ryan KA, Vederman AC, McFadden EM, Weldon AL, Kamali M, Langenecker SA, et al. Differential executive functioning performance by phase of bipolar disorder. *Bipolar Disord*. (2012) 14:527–36. doi: 10.1111/j.1399-5618.2012.01032.x
- Langenecker SA, Saunders EFH, Kade AM, Ransom MT, McInnis MG. Intermediate: cognitive phenotypes in bipolar disorder. *J Affect Disord*. (2010) 122:285–93. doi: 10.1016/j.jad.2009.08.018
- Eclarinal E, Ajilore O. Neurobiology of older age bipolar disorder. In: *Bipolar Disorder in Older Age Patients*. Cham: Springer International Publishing (2017) p. 43–56. doi: 10.1007/978-3-319-48912-4_3
- Berk M, Kapczynski F, Andrezza AC, Dean OM, Giorlando F, Maes M, et al. Pathways underlying neuroprogression in bipolar disorder: focus on inflammation, oxidative stress and neurotrophic factors. *Neurosci Biobehav Rev*. (2011) 35:804–17. doi: 10.1016/j.neubiorev.2010.10.001
- Hirschfeld RMA, Williams JBW, Spitzer RL, Calabrese JR, Flynn L, Keck J, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the mood disorder questionnaire. *Am J Psychiatry*. (2000) 157:1873–5. doi: 10.1176/appi.ajp.157.11.1873
- Zimmerman M, Galione JN. Screening for bipolar disorder with the mood disorders questionnaire: a review. *Harv Rev Psychiatry*. (2011) 19:219–28. doi: 10.3109/10673229.2011.614101
- McKinney W. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*. Austin, TX (2010). p. 56–61. doi: 10.25080/Majora-92bf1922-00a
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop)*. (2013) 36:27–46. doi: 10.1111/j.1600-0587.2012.07348.x
- Boulesteix A-L, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip Rev Data Min Knowl Discov*. (2012) 2:493–507. doi: 10.1002/widm.1072
- Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. *Caret: Classification and Regression Training*. R package version 6.0–21. CRAN Wien, Austria (2014)
- Liaw A, Wiener M. Classification and regression by random forest. *R News*. (2002) 3:18–22.
- Apley DW, Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. *J R Stat Soc Ser B Stat Methodol*. (2020) 82:1059–86. doi: 10.1111/rssb.12377
- Dong Y, Yang Y, Tang J, Chawla N V. Inferring user demographics and social strategies in mobile social networks. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min*. (2014) 14:15–24. doi: 10.1145/2623330.2623703
- Jahani E, Sundsøy P, Bjelland J, Bengtsson L, Sandy PA, De Montjoye Y-A. Improving official statistics in emerging markets using machine learning and mobile phone data. *EPJ Data Sci*. (2017) 6:3. doi: 10.1140/epjds/s13688-017-0099-3
- Al-Zuabi IM, Jafar A, Aljoumaa K. Predicting customer’s gender and age depending on mobile phone data. *J Big Data*. (2019) 6:1–16. doi: 10.1186/s40537-019-0180-9
- Yang AC, Tsai SJ. Is mental illness complex? From behavior to brain. *Prog Neuro-Psychopharmacol Biol Psychiatry*. (2013) 45:253–7. doi: 10.1016/j.pnpbp.2012.09.015
- Crump C, Sundquist K, Winkleby MA, Sundquist J. Comorbidities and mortality in bipolar disorder: a swedish national cohort study. *JAMA Psychiatry*. (2013) 70:931–9. doi: 10.1001/jamapsychiatry.2013.1394
- Czaja SJ, Sharit J. Age differences in the performance of computer-based work. *Psychol Aging*. (1993) 8:59–67. doi: 10.1037/0882-7974.8.1.59
- Kalman YM, Kavé G, Umanski D. Writing in a digital world: self-correction while typing in younger and older adults. *Int J Environ Res Public Health*. (2015) 12:12723–34. doi: 10.3390/ijerph121012723
- Chen R, Jankovic F, Marinsek N, Foschini L, Kourtis L, Signorini A, et al. Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams world from consumer-grade multimodal sensor streams. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage, AK (2019). p. 2145–55. doi: 10.1145/3292500.3330690
- Zhou J, Lo OY, Halko MA, Harrison R, Lipsitz LA, Manor B. The functional implications and modifiability of resting-state brain network complexity in older adults. *Neurosci Lett*. (2020) 720:134775. doi: 10.1016/j.neulet.2020.134775
- Sokunbi MO, Cameron GG, Ahearn TS, Murray AD, Staff RT. Fuzzy approximate entropy analysis of resting state fMRI signal complexity across the adult life span. *Med Eng Phys*. (2015) 37:1082–90. doi: 10.1016/j.medengphy.2015.09.001
- Liu CY, Krishnan AP, Yan L, Smith RX, Kilroy E, Alger JR, et al. Complexity and synchronicity of resting state blood oxygenation level-dependent (BOLD) functional MRI in normal aging and cognitive decline. *J Magn Reson Imaging*. (2013) 38:36–45. doi: 10.1002/jmri.23961
- Unobtrusive Monitoring of Affective Symptoms and Cognition Using Keyboard Dynamics (UnMASCK)*. Available online at: <https://clinicaltrials.gov/ct2/show/NCT04358900> (accessed May 20, 2021).

Conflict of Interest: OA is a co-founder of KeyWise AI. He also serves on the advisory boards of Embodied Labs and Blueprint Health. AL is an advisor for Buoy health and a consultant for Otsuka USA and ATAI Life Sciences, in addition to being a Co-founder 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.

Copyright © 2021 Zulueta, Demos, Vesel, Ross, Piscitello, Hussain, Langenecker, McInnis, Nelson, Ryan, Leow and Ajilore. 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.