

DR. JONATHAN P. STANGE (Orcid ID : 0000-0003-3718-5945)

DR. KELLY A. RYAN (Orcid ID : 0000-0001-5032-3949)

DR. OLUSOLA AJILORE (Orcid ID : 0000-0003-0737-0437)

Article type : Clinical Curiosity

Running Head: LET YOUR FINGERS DO THE TALKING

**Let Your Fingers Do the Talking:
Passive Typing Instability Predicts Future Mood Outcomes**

Jonathan P. Stange¹, John Zulueta¹, Scott A. Langenecker¹, Kelly A. Ryan², Andrea Piscitello²,
Jenna Duffecy¹, Melvin G. McInnis², Pete Nelson¹, Olusola Ajilore¹, Alex Leow¹

¹ University of Illinois at Chicago, Chicago, IL, USA

² University of Michigan Medical Center, Ann Arbor, MI, USA

BiAffect is supported by The Mood Challenge for ResearchKit, a New Venture Fund program funded by the Robert Wood Johnson Foundation. Jonathan P. Stange was supported by grant 1K23MH112769-01A1 from NIMH. In the past 3 years Melvin G. McInnis has consulted for Janssen and Otsuka Pharmaceuticals; he is listed as an inventor on US patent US 9685174, Mood

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/bdi.12637](https://doi.org/10.1111/bdi.12637)

This article is protected by copyright. All rights reserved

monitoring of bipolar disorder using speech analysis, held by the Regents of the University of Michigan.

Correspondence: Alex Leow, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA; Email: alexfeuille@gmail.com; Jonathan P. Stange, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA; Email: jstange@uic.edu

Author Manuscript

Key Message

Smartphone innovations have opened new frontiers in the assessment of disease processes. Greater day-to-day instability in actively-reported mood and passively-recorded typing kinematics across two weeks predicted a poorer prospective course of depression and mania. This demonstrates the feasibility and utility of digital phenotyping in detecting individual differences in disease course.

Author Manuscript

Introduction

Advances in smartphone technology allow for the assessment of disease processes in real time outside of the lab or clinic, providing novel information with greater ecological validity – that is, how disease influences patients in the “real world.” Ecological momentary assessment (EMA), which involves prompting patients to answer questions about their current state throughout their daily lives, is one technique that can be administered via smartphone. This approach has the ability to identify contextual information influencing disease processes that may not be available when patients visit the lab or clinic¹. Moreover, additional information can be passively recorded, such as metrics of keyboard typing kinematics. These metrics represent a novel category of “digital phenotyping” data streams that involve the passive measurement of behavior from smartphone sensors and keyboard interaction^{2,3}. These real-time metrics hold great promise as thousands of data points per day can be collected with little to no additional effort from the user. Thus, smartphone metrics have important potential for the identification of periods of risk, transition, and translation to real-time interventions¹⁻³.

Active EMA approaches can be used to assess intra-individual variability in self-reported mood⁴. Mood instability (MI), defined as large, frequent fluctuations in mood over time⁴, is one process that is relevant for understanding vulnerability to persistent mood dysfunction in mood disorders⁵. Yet, no studies to our knowledge have extended this work to examine instability in real-time, passively-recorded digital phenotyping metrics. These passively-recorded metrics place less burden on the user and thus may be more feasible for the long-term monitoring of clinically-relevant features of behavior.

We hypothesized that day-to-day instability in actively-reported mood and passively-recorded instability in typing speed would predict a poorer prospective symptom course, in a pilot study of individuals with bipolar disorder. We also examined how many days of observation would be necessary to obtain a reliable estimate of each person’s measures of instability.

Case Presentation

Participants were 18 adult participants in the Prechter Longitudinal Study of Bipolar Disorder, as part of the BiAffect-PRIORI consortium, who had familiarity with the Android operating system. Participants completed a baseline period of approximately two weeks in which EMA was conducted using the BiAffect app on Android smartphones. Participants were prompted semi-randomly, once daily, to respond to five questions about their current affective state (mood,

energy, rapidity of thinking, impulsive actions, and impulsive feelings), which were rated on scales ranging from 0 (minimum) to 100 (maximum).

Simultaneously, BiAffect unobtrusively collected typing kinematics metadata, evaluating *how* a person typed (on the keyboard) rather than *what* was typed, as subjects interacted with their phones in a usual manner. This allowed us to extract individuals' typing speed within each day in their natural environments. To do this, we computed the daily mean of the time since the last keystroke, based on an average of 1,371 keystrokes per day per person.

For each participant, instability metrics were created for each EMA affective rating and for daily typing speed, using the root mean square successive difference (rMSSD). This metric evaluates variability in consecutively-recorded daily ratings and encapsulates both the *size* and the *temporal order* of within-subject fluctuations, in contrast with other indices of variability such as the standard deviation, which ignore potentially-meaningful information based on the order in which change occurs⁴.

Following the EMA period, participants were followed for up to 8 weeks. At weekly intervals, they completed clinician interviews by phone to assess their current symptoms of depression (Hamilton Depression Rating Scale, HAM-D) and mania (Young Mania Rating Scale, YMRS). These measures also were completed prior to the EMA period to evaluate baseline mood state.

Multilevel models were estimated using Mplus to examine the extent to which indices of instability in active and passive assessments would be predictive of levels of future mood states (random intercepts of HAM-D or YMRS ratings across the prospective period; $N=120$ observations). To conduct a conservative approach of hypotheses (and relative clinical merit), we tested the predictive utility of instability metrics after entering the daily variables used in the instability computations (e.g., average daily EMA mood rating), and after entering the influence of baseline mood symptoms into the models (Figure 1).

As hypothesized, greater instability of mood during the baseline EMA period significantly predicted elevated prospective symptoms of depression ($t=2.91, p<.005$; Figure 2a) and mania ($t=2.72, p<.01$). Instability of energy predicted elevated prospective symptoms of mania ($t=2.20, p=.03$), but not depression ($t=1.78, p=.08$). Instability in other affective variables (rapidity of thinking, impulsive feelings and actions) did not significantly predict future symptom course

($p > .10$). In addition, instability of typing speed predicted elevated future symptoms of depression ($t = 2.18, p = .03$; Figure 2b), but not mania ($t = 0.73, p = .46$).

Given that active and passive indices of instability each predicted future mood symptoms, we conducted bootstrapped mediation analyses. These analyses indicated that mood instability mediated the relationship between instability of typing speed and prospective symptoms of depression and mania (indirect effect $p < .05$), suggesting shared contributions of active and passive metrics of instability in the prediction of illness course. Conversely, instability of typing speed did not mediate the relationship between mood instability and depression or mania.

Finally, sensitivity analyses suggested that evaluating instability across as little as 5 to 7 days is optimal. Instability metrics over 5-7 days provided similar information and had comparable predictive utility in our models (Figure 2c,d).

Discussion

The results of this preliminary investigation support the feasibility and potential utility of active and passive indices of instability in mood and typing kinematics in the prediction of prospective illness course among adults with bipolar disorder. These findings suggest that the use of EMA and related smartphone technologies may be useful for identifying patients who have impending worsening in their patterns of mood and behavior, and who may be in need of more timely, alternative, or intensive treatments.

When examined as predictors separately, passive assessments of typing instability accounted for a similar amount of variance in future mood symptoms compared to actively-reported EMA mood instability. Importantly, this suggests that comparable information about risk for mood changes may be provided by passive observation of keyboard dynamics, which are unobtrusive, minimally invasive, and can be evaluated indefinitely without any additional need for self-reflection on the part of participants. This finding highlights the great potential that passive smartphone assessment holds for taking a major leap forward in advancing patient-centered assessment and intervention in psychiatry. It also is important to note that although subjects in the present study were highly compliant over a period of two weeks, not all patients may be as compliant for extended periods of time⁴. Whereas actively-reported EMA responses are subject to these problems, passively-recorded keyboard dynamics can be recorded indefinitely over time, allowing for potentially greater clinical utility.

Despite predicting a similar amount of variance, it is worth commenting that in a combined statistical model, self-reported EMA mood instability was superior to instability of passively-recorded typing speed in predicting future clinician-reported mood symptoms. It may be that EMA reports and clinician-rated mood scores are methodologically more similar as both rely on patients' awareness of their mood, thus ultimately are dependent on their own account. In contrast, passive assessments are not dependent on patient awareness, and thus may capture additional variance. For example, passive assessments of typing instability could index related latent domains of clinical relevance, such as homeostasis of motor planning and execution circuits, of which a person may be altogether unaware. Thus, the example of typing instability highlights the potential benefits of smartphone technology, and digital phenotyping approaches more broadly, in the future of psychiatric assessment and intervention.

Furthermore, these data suggest that using as few as 5-7 days of data provides estimates of instability that are as reliable, and provide similar predictive utility, as when assessing instability over longer periods of time. It may be that evaluating instability across a full week provides a sufficiently-representative sample of day-to-day variability across the week. Such metrics could be used to assess for and manage intervals in between treatment visits.

In conclusion, we provide an example of how passive and active assessments of affective and behavioral instability may provide clinical utility for understanding risk for negative outcomes in bipolar disorder⁵. These methods highlight a possible pathway for evaluating latent cognitive domains or clinical classifications that might lead to the development of novel "brain fitness" indicators that have actionable clinical utility. These approaches also offer possibilities for real-time digital phenotyping in disorders in which changes in the character of mood, behavior, and thought might translate into variations in how individuals interact with their phones¹⁻³. This work could then be tailored to specific individuals using a real-time, person-centered approach to detecting vulnerability points when intervention might be most useful.

References

1. Malhi GS, Hamilton A, Morris G, Mannie Z, Das P, Outhred T. The promise of digital mood tracking technologies: are we heading on the right track? *Evidence-based mental health*. 2017 Nov 1;20(4):102-7
2. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA*. 2017 Oct 3;318(13):1215-6.
3. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*. 2016 Jun;41(7):1691-6.
4. Ebner-Priemer UW, Eid M, Kleindienst N, Stabenow S, Trull TJ. Analytic strategies for understanding affective (in) stability and other dynamic processes in psychopathology. *J Abnormal Psychol*. 2009 Feb;118(1):195-202.
5. Stange JP, Sylvia LG, da Silva Magalhães PV, Miklowitz DJ, Otto MW, Frank E, Yim C, Berk M, Dougherty DD, Nierenberg AA, Deckersbach T. Affective instability and the course of bipolar depression: results from the STEP-BD randomised controlled trial of psychosocial treatment. *Br J Psychiatry*. 2016 Apr 1;208(4):352-8.

a)

b)

c)

Figure 1. Example of one participant's assessments of (a) actively-recorded daily EMA mood ratings, (b) passively-recorded daily average typing speed, and (c) baseline and prospective symptoms of depression across several weeks of follow-up.

Author Manuscript

a) b)

c)

d)

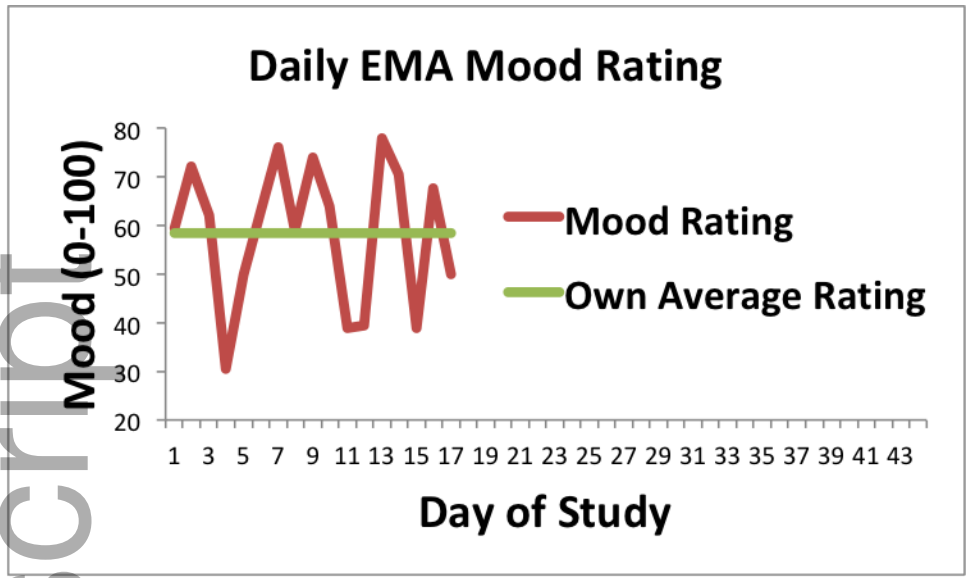
Figure 2.

A&B. Associations between subject-level prospective depression (Ham-D) symptom values predicted by multilevel instability models, and actual mean levels of prospective symptoms. Predictors and covariates in these models included the given instability metric (e.g., EMA mood instability), average level of the variable used in the instability metric (e.g., average daily EMA mood rating), and baseline mood symptoms (baseline Ham-D when predicting prospective Ham-D; baseline YMRS when predicting prospective YMRS). TLK = time since last key (typing speed).

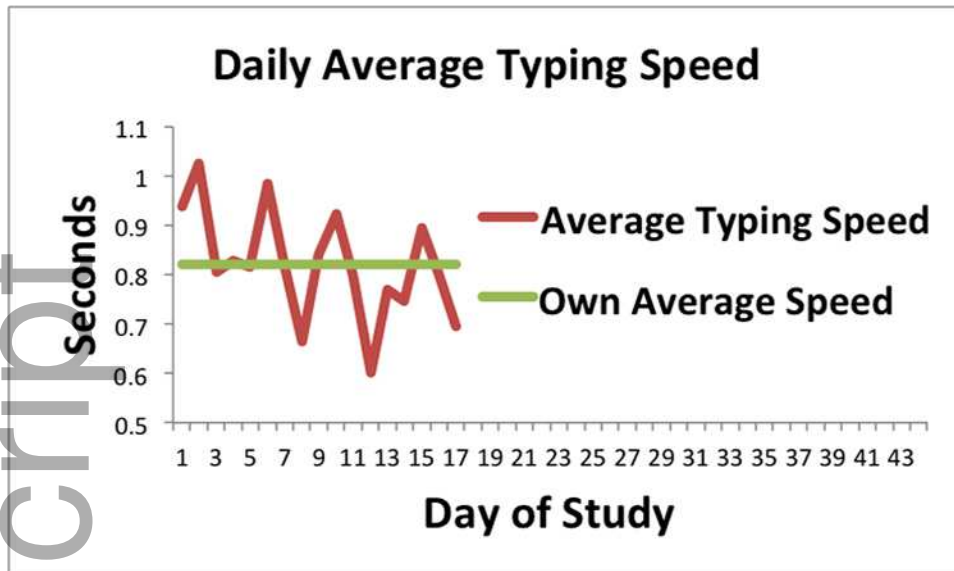
C. Reliability of active and passive assessments of instability based on number of days of assessment: amount of variance shared when using 11-day versus 3-to-10-day windows for computing instability.

D. Predictive utility of active and passive assessments of instability based on number of days of assessment: amount of variance in random intercepts of prospective symptoms predicted in multilevel models by active and passive indices of instability.

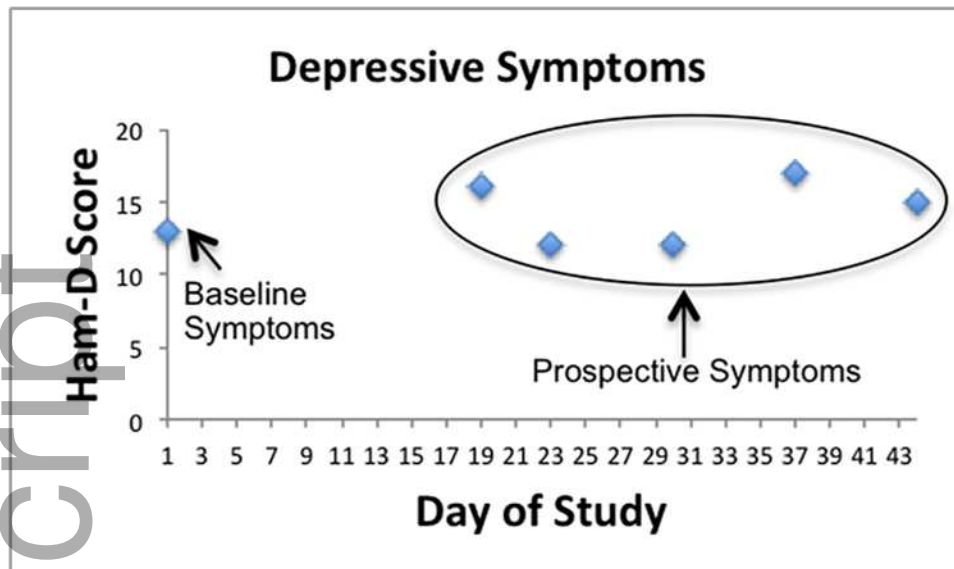
Author Manuscript



bdi_12637_f1a.png

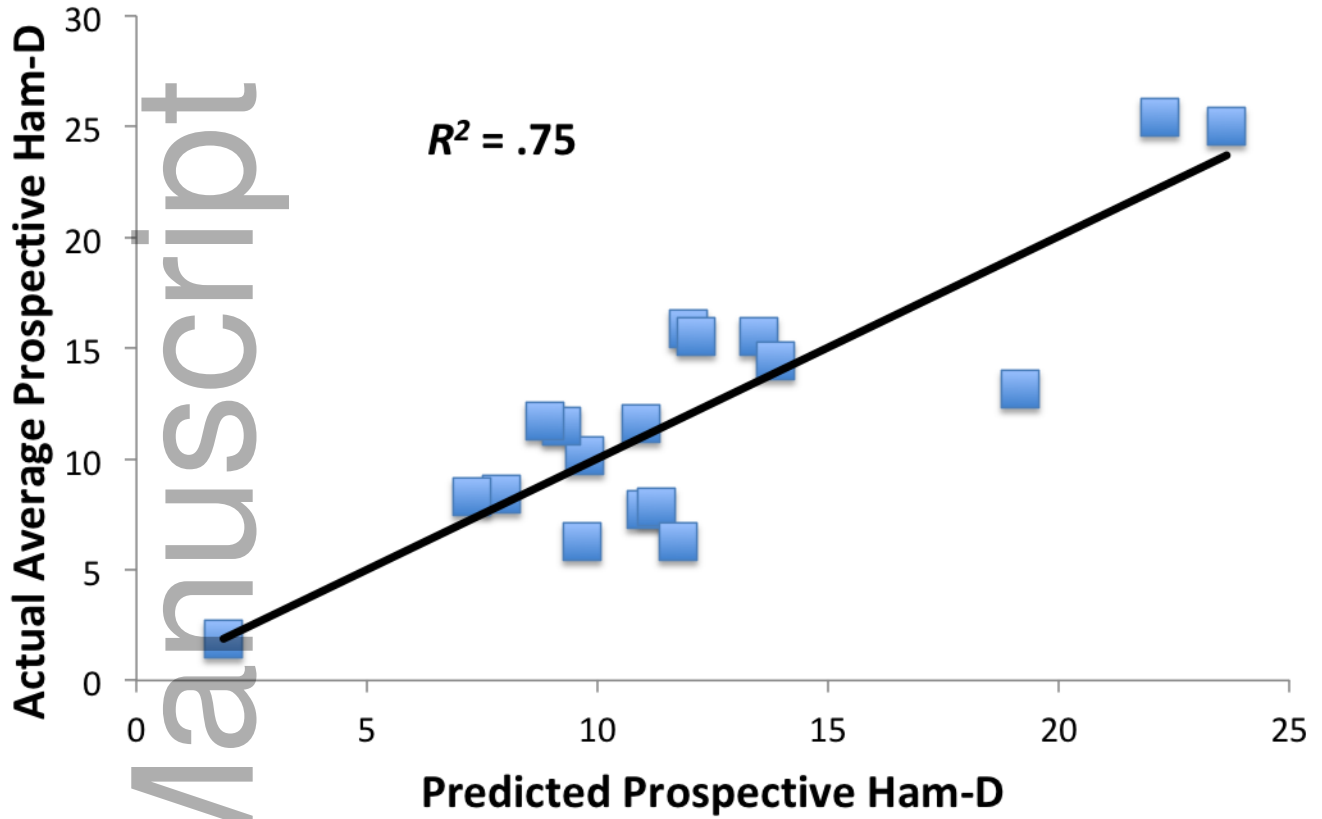


bdi_12637_f1b.png



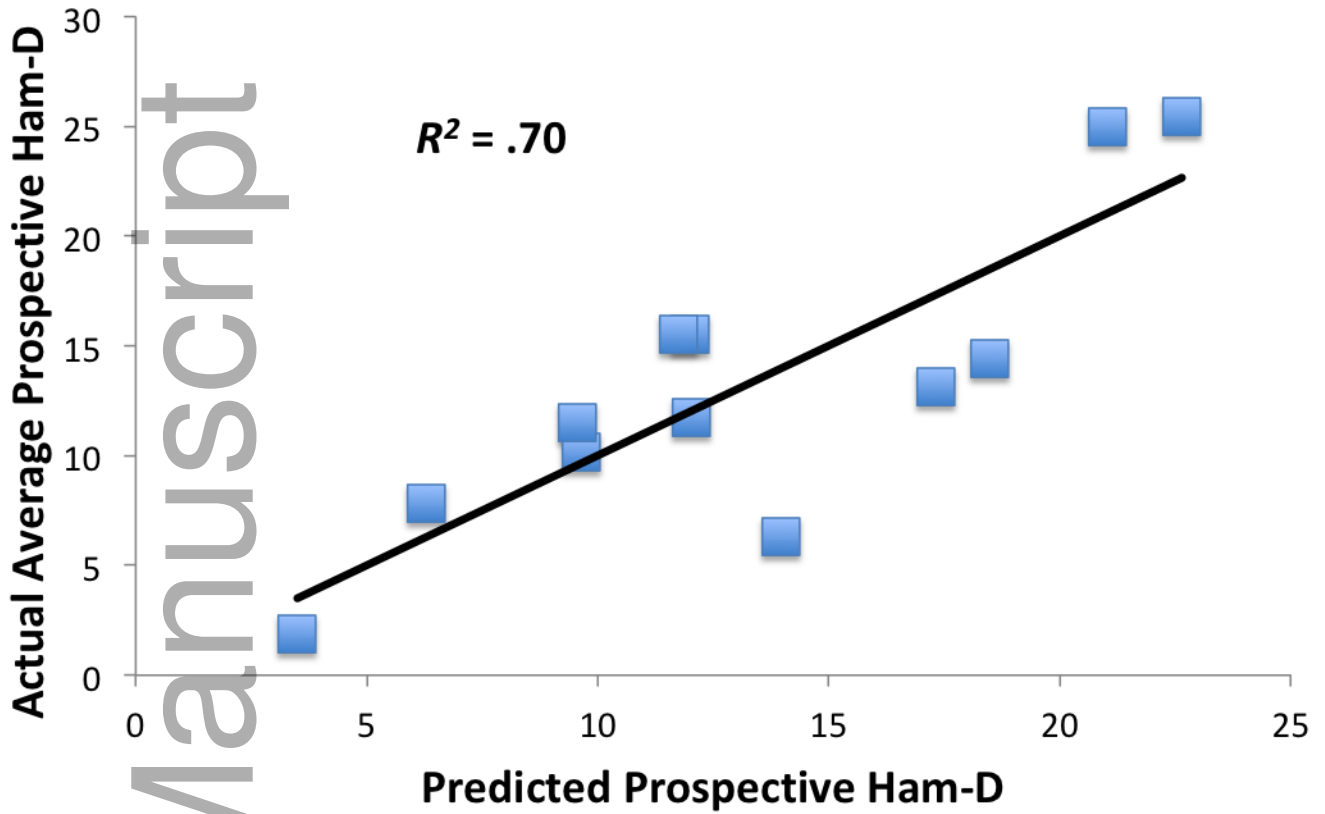
bdi_12637_f1c.png

Instability of Daily Mood Predicting Future Depression Levels

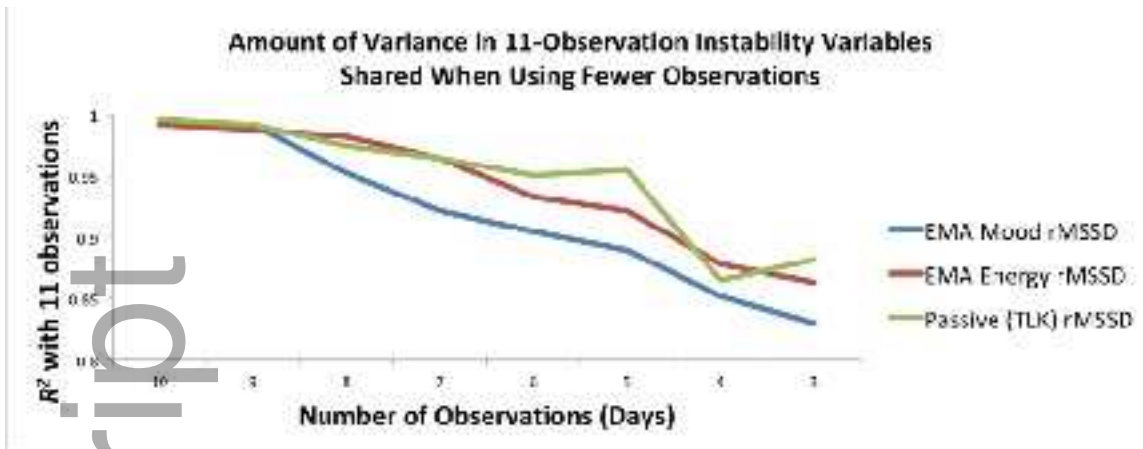


bdi_12637_f2a.png

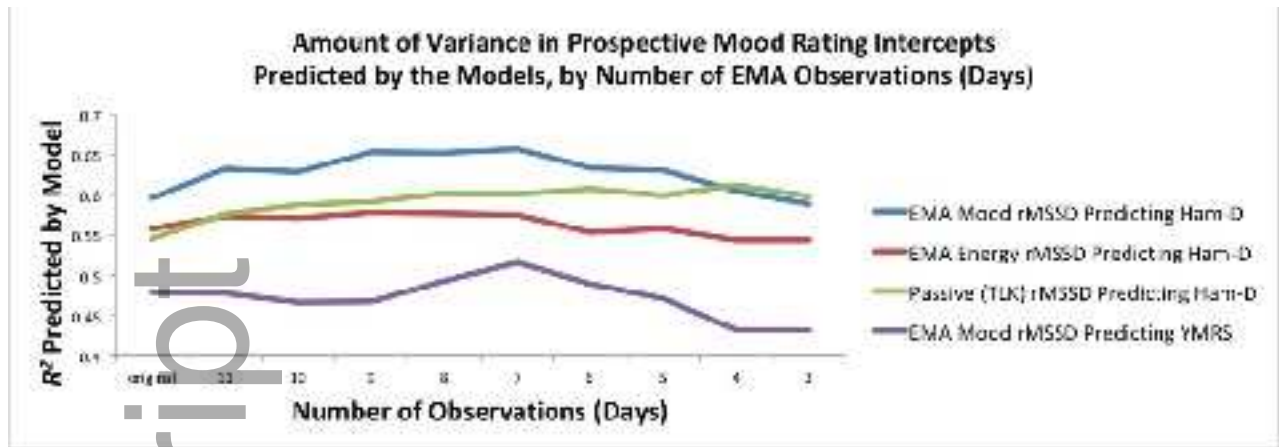
Instability of Daily Typing Speed Predicting Future Depression Levels



bdi_12637_f2b.png



bdi_12637_f2c.png



bdi_12637_f2d.png