Using Digital Phenotyping to Accurately Detect Depression Severity

Nicholas C. Jacobson, MS, Hilary Weingarden, PhD, and Sabine Wilhelm, PhD

Abstract: Development of digital biomarkers holds promise for enabling scalable, time-sensitive, and cost-effective strategies to monitor symptom severity among those with major depressive disorder (MDD). The current study examined the use of passive movement and light data from wearable devices to assess depression severity in 15 patients with MDD. Using over 1 week of movement data, we were able to significantly assess depression severity with high precision for self-reported ($r = 0.855$; 95% confidence interval [CI], 0.610–0.950; $p = 4.95 \times 10^{-5}$) and clinician-rated ($r = 0.604$; 95% CI, 0.133–0.894; $p = 0.017$) symptom severity. Pending replication, the present data suggest that the use of passive wearable sensors to inform healthcare decisions holds considerable promise.

Key Words: Digital phenotyping, major depression, sensors

(*J Nerv Ment Dis* 2019;207: 893–896)

Regularly monitoring patients' psychiatric symptoms is essential for detecting clinical deterioration and initiating timely intervention. Traditionally, symptom tracking has required repeated completion of clinician-administered or self-report assessments during in-person visits—a time-consuming, costly, and burdensome process (Insel, 2017). In recent years, researchers have increasingly called for the development of digital phenotyping tools as low-burden, scalable aids to symptom monitoring (Insel, 2017; Onnela and Rauch, 2016; Torous et al., 2016). Digital phenotyping is the "moment-by-moment quantification of the individual-level human phenotype in situ, using data from personal digital devices such as smartphones" (Onnela and Rauch, 2016). That is, digital phenotyping involves the use of unobtrusive, technology-based sensors to continuously collect objective and temporally sensitive data about individuals, to make health inferences. Given the widespread adoption of smartphones, now owned by 77% of the US adult population (Pew Research Center, 2018), there is strong potential for traditional clinical assessments to be meaningfully supplemented by this low-burden, context-rich digital approach (Torous et al., 2016). Use of digital phenotyping to remotely monitor patients' symptom severity may be especially beneficial for disorders with an episodic course, such as major depressive disorder (MDD; American Psychiatric Association, 2013). Specifically, digital phenotyping may enable early detection and intervention of sudden increases in severity, so that clinical deterioration and sequential comorbidity may be prevented (Jacobson and Newman, 2017).

To date, growing research has focused on digital phenotyping for the diagnostic assessment of mood disorders (Jacobson et al., 2019), obtaining strong accuracy in diagnosing major depressive episodes using GPS and call log data (for a review, see Bourla et al., 2018a). However, less work has focused on honing passive assessment of depression severity (Mohr et al., 2017). Moreover, previous studies examining digital phenotyping of depression severity have yielded mixed results (Rohani et al., 2018), and most have yielded small associations between passive data and self-reported or clinician-assessed mood (Pratap et al., 2018). Thus, there is a need for additional work focused on digital phenotyping for the assessment of depression severity.

To build on this preliminary yet promising research, we analyzed public-use movement and light exposure data (Tonon et al., 2017) over 1 week in 15 patients with clinician-diagnosed MDD. (Note that was not registered as a clinical study, and instead represents a secondary objective of the data.) Researchers have proposed that movement and light data could be especially relevant for detecting features of mood disorders, as these data may correspond with behavioral patterns such as disturbances in energy, psychomotor agitation or retardation, and sleep disturbances that characterize mood episodes (Krane-Gartiser et al., 2017; Tonon et al., 2017). We hypothesized that these data could be used to accurately assess current depression severity.

METHODS

Participants

Fifteen outpatients (87% female, $M_{Age} = 47.60$, $SD_{Age} = 10.45$) were recruited from the Mood Disorders Program of Porto Alegre Clinical Hospital. Based on the Hamilton Depression Inventory (HAM-D), symptom severity was very severe in most patients (66%). All patients were taking either selective serotonin reuptake inhibitor or tricyclic antidepressants. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving patients were approved by the Research Ethics Committee of the Porto Alegre Clinical Hospital (HCPA, protocol number 11–0456 GPPG/HCPA) (Tonon et al., 2017). Written informed consent was obtained from all patients following a complete description of the study (Tonon et al., 2017).

Clinical Assessments

Psychiatrists served as independent evaluators and confirmed the diagnosis of MDD using the Mini-International Neuropsychiatric Interview (M.I.N.I. 5.0), Brazilian Portuguese Version. The self-report Beck Depression Inventory-II (BDI-II) and clinician-administered HAM-D were used to assess depression severity.

Actigraphy

Similar to an actigraphy study from a large nationally representative sample (Loprinzi and Cardinal, 2011), patients wore actigraphs (Actiwatch-L) for 1 week, which recorded continuous movements (≥0.01 g) and ambient light exposure in lux every 15 minutes.

Analyses

We created digital biomarkers by applying square root, square, and log transformations to light and actigraphy data (*e.g.*, oscillations and peak values during daytime and nighttime intervals; Tonon et al., 2017). We created biomarkers without reliance on study-specific information (*e.g.*, not tied to the particular day within the study), so that features can be generalized to new samples. Subsequently, we used the

Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts.
Send reprint requests to Nicholas C. Jacobson, MS, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge St, Suite 2000, Boston, MA 02114. E-mail: ncjacobson@mgh.harvard.edu.

ISSN: 0022-3018/19/20710–0893
DOI: 10.1097/NMD.0000000000001042

extracted digital biomarkers to assess BDI-II and HAM-D scores via extreme gradient boosting ("xgboost"), a high-precision, tree-based boosting algorithm, which is a form of machine learning. We applied leave-one-out cross-validation (LOOCV) and a permutation test to control for overfitting (Friedman et al., 2001; Ojala and Garriga, 2010). Primary outcomes were the correlations between predicted and observed depression severity scores. Note that as a sensitivity test we also computed the partial correlation coefficients after controlling for sex and age (the only demographic/patient characteristics available).

RESULTS

The correlation between predicted and observed depression severity values was very strong using observed BDI-II scores (r = 0.855; 95% confidence interval [CI], 0.610–0.950; $p = 4.95 \times 10^{-5}$), and it was strong using observed HAM-D scores (r = 0.604; 95% CI, 0.133–0.894; p = 0.017; see Fig. 1 and Fig. 2). Results remained significant after controlling for the false discovery rate using a permutation test. After controlling for sex and age, the association of the predicted and observed depression severity values remained very strong using observed BDI-II scores (r = 0.907; 95% CI, 0.737–0.969; $p = 1.87 \times 10^{-5}$), and remained strong using observed HAM-D scores (r = 0.577; 95% CI, 0.092–0.841; p = 0.039).

DISCUSSION

The present research demonstrates that passive movement and light data collected for 1 week in a sample of outpatients with MDD can be used to accurately assess both self-reported and clinician-rated depression severity. In fact, although most prior research has obtained null to weak relationships between digital biomarkers and depression severity (Pratap et al., 2018), the present statistical model had strong ability to detect depressive symptom severity.

Developing the ability to remotely assess depression severity with high temporal resolution and low patient burden has strong potential to enhance traditional assessment. For example, standard in-person assessments can be supplemented by passive assessment between visits, potentially allowing for longer time to pass between in-person visits with a clinician. Passive monitoring can also enable clinicians or connected smartphone-based treatments to intervene with just-in-time support, if patients' symptoms worsen between visits. In particular, such tools could be beneficial for risk management (Torous et al., 2018). Given the early scientific foundation for digital phenotyping tools and the lack of regulatory oversight, Torous and Roberts (2017) suggest that, at present, digital phenotyping tools should enhance existing clinical methods rather than replacing them entirely. Moreover, research by Bourla et al. (2018b) shows that psychiatrists have hesitation about the potential utility, usability, reliability, and risk (in particular, legal risk) related to digital phenotyping tools. Thus, in order for these tools to have maximal impact, barriers to mental health provider adoption must be addressed (*e.g.*, with education).

In addition to the strong correlations obtained, it is notable that the current sample spanned patients with high symptom severity, suggesting that using digital phenotyping to detect symptom severity in depression may be helpful even when severity is particularly high. The present study also had several limitations that should be considered. First, movement and light data were collected for a short period. Although it is notable that robust results can be obtained in 1 week, data collection over longer periods will likely yield more reliable findings. Second, because all patients were medicated and the medication type was not provided in the public use data, we were unable to account for the impact of medication status on psychomotor movements. In

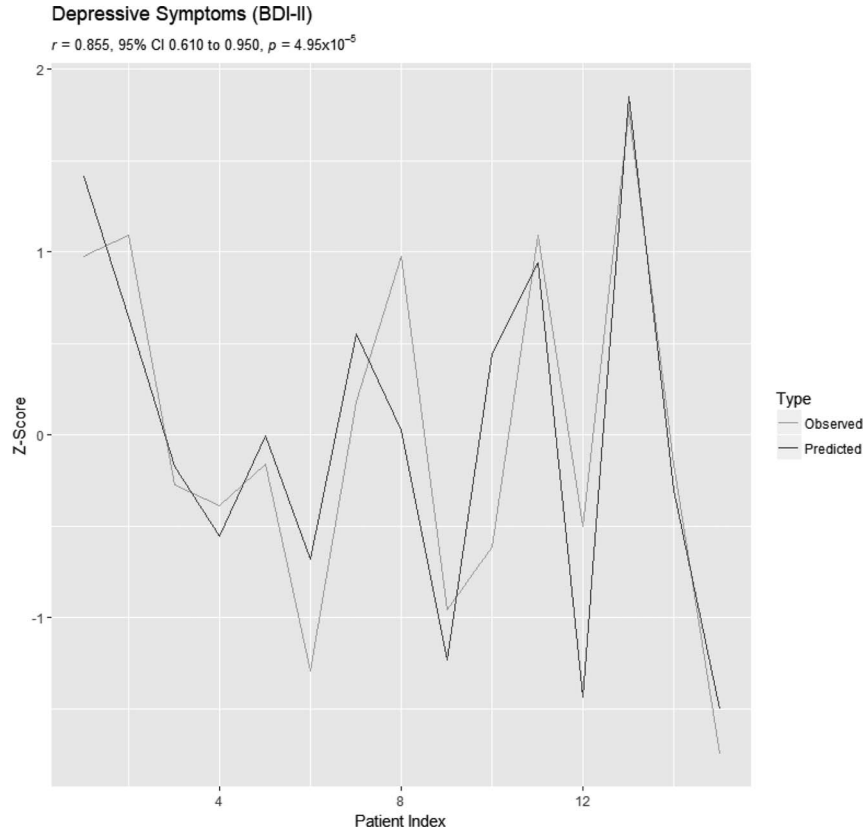

FIGURE 1. This graph depicts the observed and predicted *z*-scores from the BDI-II.

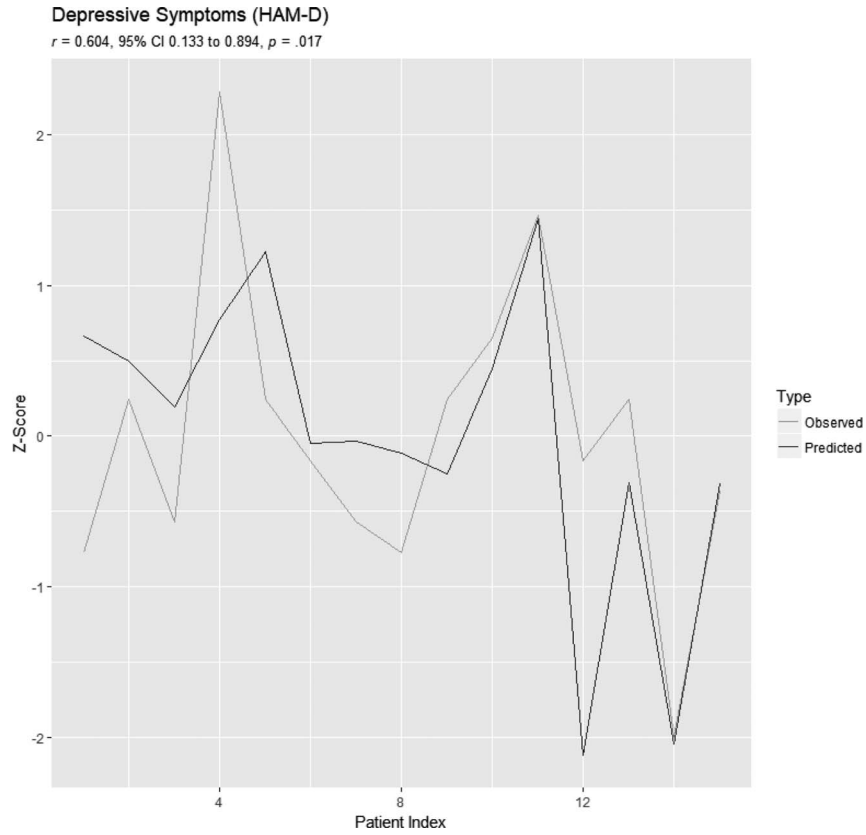

FIGURE 2. This graph depicts the observed and predicted *z*-scores from the HAM-D.

addition, our sample was small. Although the current application is limited to 15 outpatients, a large number of data points were utilized to build digital biomarkers incorporated into statistical models. Moreover, the use of LOOCV and permutation tests suggests that results are not due simply to overfitting of data (Friedman et al., 2001; Ojala and Garriga, 2010). Taken together, results are promising, yet preliminary, and require replication in a larger, independent sample.

CONCLUSIONS

Initial results suggest that incorporating continuous, objective, and unobtrusive measures of depression severity into patient management may reduce clinician reliance on higher burden (*e.g.*, clinician-administered) assessments. Digital phenotyping may in turn enable scalable, unobtrusive, early detection of risk and just-in-time interventions aimed at lowering the impact of depression.

DISCLOSURES

Mr. Jacobson is the owner of a free application published on the Google Play Store entitled "Mood Triggers." He does not receive any direct or indirect revenue from his ownership of the application (i.e., the application is free, there are no advertisements, and the data are only being used for research purposes). Drs. Weingarden and Wilhelm have received salary support from Telefonica Alpha, Inc. Dr. Wilhelm is a presenter for the Massachusetts General Hospital Psychiatry Academy in educational programs supported through independent medical education grants from pharmaceutical companies. Dr. Wihelm has received royalties from Elsevier Publications, Guilford Publications, New Harbinger Publications, and Oxford University Press. Dr. Wilhelm has also received speaking honoraria from various academic institutions and foundations, including the International Obsessive Compulsive Disorder Foundation and the Tourette Association of America. In addition, she received payment from the Association for Behavioral and Cognitive Therapies for her role as Associate Editor for the Behavior Therapy journal, as well as from John Wiley & Sons, Inc. for her role as Associate Editor for the journal Depression & Anxiety.

REFERENCES

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

Bourla A, Ferreri F, Ogorzelec L, Guinchard C, Mouchabac S (2018a) Assessment of mood disorders by passive data gathering: The concept of digital phenotype versus psychiatrist's professional culture. *Encephale*. 44:168–175.

Bourla A, Ferreri F, Ogorzelec L, Peretti CS, Guinchard C, Mouchabac S (2018b) Psychiatrists' attitudes toward disruptive new technologies: Mixed-methods study. *JMIR Ment Health*. 5:e10240.

Friedman J, Hastie T, Tibshirani R (2001) *The elements of statistical learning* (Vol 1). New York: Springer Series in Statistics.

Insel TR (2017) Digital phenotyping: Technology for a new science of behavior. *JAMA*. 318:1215–1216.

Jacobson NC, Newman MG (2017) Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. *Psychol Bull*. 143: 1155–1200.

Jacobson NC, Weingarden H, Wilhelm S (2019) Digital biomarkers of mood disorders and symptom change. *Digital Medicine*. 2:3.

Krane-Gartiser K, Vaaler AE, Fasmer OB, Sørensen K, Morken G, Scott J (2017) Variability of activity patterns across mood disorders and time of day. *BMC Psychiatry*. 17:404.

Loprinzi PD, Cardinal BJ (2011) Association between objectively-measured physical activity and sleep, NHANES 2005–2006. *Ment Health and Phys Act*. 4:65–69.

Mohr DC, Zhang M, Schueller SM (2017) Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annu Rev Clin Psychol*. 13:23–47.

Ojala M, Garriga GC (2010) Permutation tests for studying classifier performance. *J Mach Learn Res*. 11:1833–1863.

Onnela JP, Rauch SL (2016) Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*. 41:1691–1696.

Pew Research Center (2018) Mobile Fact Sheet. Available at: https://www.pewinternet.org/fact-sheet/mobile/. Accessed May 23, 2019.

Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, Areán PA (2018) The accuracy of passive phone sensors in predicting daily mood. *Depress Anxiety*. 36:72–81.

Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE (2018) Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: Systematic review. *JMIR Mhealth Uhealth*. 6:e165.

Tonon AC, Fuchs DFP, Barbosa Gomes W, Levandovski R, Pio de Almeida Fleck M, Hidalgo MPL, da Silva Alencastro L (2017) Nocturnal motor activity and light exposure: Objective actigraphy-based marks of melancholic and non-melancholic depressive disorder. Brief report. *Psychiatry Res*. 258:587–590.

Torous J, Kiang MV, Lorme J, Onnela JP (2016) New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*. 3:e16.

Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, Firth J (2018) Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. *Curr Psychiatry Rep*. 20:51.

Torous J, Roberts LW (2017) The ethical use of mobile health technology in clinical psychiatry. *J Nerv Ment Dis*. 205:4–8.