REVIEW ARTICLE

WEARABLE DIGITAL HEALTH TECHNOLOGIES IN MEDICINE

Stephen H. Friend, M.D., Ph.D., *Guest Editor*, Geoffrey S. Ginsburg, M.D., Ph.D., *Guest Editor*, Rosalind W. Picard, Sc.D., *Guest Editor*, and Jeffrey M. Drazen, M.D., *Editor*

Wearable Technology in Clinical Practice for Depressive Disorder

Szymon Fedor, Ph.D., Robert Lewis, M.Sc., Paola Pedrelli, Ph.D., David Mischoulon, M.D., Ph.D., Joshua Curtiss, Ph.D., and Rosalind W. Picard, Sc.D.

From the MIT Media Lab, Massachusetts Institute of Technology, Cambridge (S.F., R.L., R.W.P.), and the Depression Clinical and Research Program, Massachusetts General Hospital (P.P., D.M., J.C.), and the Applied Psychology Department, Northeastern University (J.C.), Boston — all in Massachusetts. Dr. Fedor can be contacted at sfedor@mit.edu or at MIT Media Lab, E14-348b, 75 Amherst St., Cambridge, MA 02139.

N Engl J Med 2023;389:2457-66.
DOI: 10.1056/NEJMra2215898

Joe, who has received a diagnosis of major depressive disorder, is meeting every 2 weeks with his psychiatrist, Sandy.

"How have you been feeling since we last met?" asks Sandy.

"Much better," says Joe. "I've been much more active and social, and I'm sleeping great!"

"That's wonderful," says Sandy. "But…I think your wearable must be broken. The data from it look very irregular for your sleep these past 2 weeks."

"Oh," says Joe, "it's not broken. Actually, now that you mention it, my sleep has been really messed up. I slept well only yesterday."

"Well," asks Sandy, "should we talk more about how we can improve your sleep?"

This conversation is based on a real patient–therapist interaction. In this case, the data from wearable technology served as a prompt to obtain details of the patient's life that might have otherwise been missed. Traditional clinical assessments depend on patient recall. Although such recall can include important factors that wearable technology (often termed "wearables") do not detect, such as patients' reports of distress, the assessments by wearables of longitudinal data from daily life may augment methods of monitoring and treating depression, providing objective complements to subjective information from patients.

Perspective on Current Clinical Practice

Treatment of depression is based on evaluation for the presence and severity of symptoms. Depression is mostly diagnosed and monitored by means of interview-based assessments or self-report scales, which include questions about current circumstances, personal history, family history, and the frequency of symptoms attributable to depression.[1]

Much of this information is subjective (i.e., based on the patient's report) and shaped by the clinician's judgment. Hence, it is prone to various biases that may affect the quality of care. Researchers have suggested various explanations for these biases.[2-4]

Self-report scales have been used to remotely monitor patients' symptoms between clinical visits. If these instruments (e.g., the Patient Health Questionnaire 9 [PHQ-9], which scores depression on a scale from 0 to 27, with higher numbers indicating more severe symptoms) are administered regularly, they can reduce retrospective bias.[5,6] However, the burden on the patient can limit how frequently questionnaires are collected. Indeed, adherence to questionnaires is reduced with an increase in the frequency or period of administration.[7]

WEARABLES IN THE DIAGNOSIS, MONITORING, AND TREATMENT OF DEPRESSION

Wearables enable passive monitoring (i.e., monitoring without active input) of behavioral and physiological factors. The greater frequency and objectivity of wearable measurements help overcome some of the limitations of adherence and bias when depression is monitored exclusively with self-reports.

Longitudinal measurements from wearables — especially when coupled with contextual data from smartphones — have the potential to augment existing clinical decision-making processes in psychiatry.[8-11] Complementing patient self-reports, these measurements increase the information available for differential diagnosis or treatment planning.

Wearables support personalized approaches to depression treatment. Longitudinal measurements help establish patient-specific behavioral and physiological baselines and support identification of personalized factors that result in deviations from these norms. A clinician can use this idiographic (within-person) information along with more general, nomothetic (between-person) evidence from randomized, controlled trials (RCTs) to monitor condition changes and adjust the treatment plan for each patient. Furthermore, these measurements may enable patient subtyping[12] on the basis of similar behavioral or physiological types, which can subsequently be associated with treatment outcomes.

Data from wearables can help clinicians compile a more comprehensive view of a patient's condition and identify potential targets for intervention, but the data do not dictate what to target. Clinical judgment is still essential for making appropriate treatment decisions. For example, Joe's wearable indicated some disruptions in his sleep. Treating sleep in persons with depression can significantly reduce their depression.[13] On the basis of this information, Joe's clinician may decide to make sleep the focus of treatment. However, if Joe is a new parent, this focus may be suboptimal. Thus, clinical judgment, which includes an understanding of each patient's context, is required when wearable-provided information is used.

In Joe's case, the objective sleep patterns helped the doctor to characterize Joe's condition more accurately. When presented with these measurements, Joe and his doctor agreed that he needed further help managing his sleep. This vignette also illustrates how objective data can engage patients, improve their understanding of their condition, and potentially motivate them to try a treatment. Patients who have been undergoing a long-term treatment may also be engaged and motivated by wearables. For example, objective trends in sleep and activity data may illustrate clinical improvement, and viewing these trends with their physicians may encourage patients to persevere with treatment.[14]

By combining sensor measurements with self-reports, clinicians may reduce inaccuracies due to biases. Data from wearable devices may allow clinicians to monitor changes in behavior and symptoms more accurately between visits, and such monitoring may indicate whether a patient has had a response to medication, has adhered to therapy, or has had a relapse. For example, data indicating an effect on sleep may guide a change in medications. Although it might seem obvious that this information can improve outcomes, there are no rigorous RCTs confirming whether improved access to data will indeed produce better clinical care and outcomes.

Access to interpretable data from a patient's wearable may affect how clinicians choose to structure their time with the patient. The initial psychiatric evaluation may take up to 2 hours.[15] Part of this time is spent asking about symptoms over the previous weeks. However, information about some symptoms of depression, including sleep patterns and activity levels, could be gathered from the patients' wearable data and made available to the clinician at the first visit. These data could be entered into clinical records in a fully automated or semiautomated way. (An example of a semiautomated approach would be that the longitudinal graphs are generated automatically, and then a specially trained team member, called a digital navigator,[16] summarizes the data, enabling the clinician to quickly see the data that are most relevant and actionable.) In this proposed scenario, clinicians could use their limited time with patients more efficiently, gathering insights not provided by a wearable, thus forming a better therapeutic relationship and delivering personalized treatment. Furthermore, recently introduced Current Procedural Terminology (CPT) codes for remote patient monitoring with wearables (e.g., codes 99453,

99454, and 99457) may provide new reimbursement streams for clinicians.[17]

MODEL OF DIGITAL MENTAL HEALTH MONITORING

Wearable systems that have been evaluated for monitoring depression usually include several layered components, as shown in Figure 1. The purpose is to transform raw measurements from the sensors into meaningful information to support clinical decisions. Raw measurements from wearables are usually received by a synchronized smartphone that provides additional context (e.g., location, app usage, and website activity). Many signals of interest for monitoring behaviors while treating depression can be tracked in real-world settings with the use of current wearable technology; these signals include acceleration, location, screen-based activity, app usage, wearable-based payments, and logs of smartphone calls and messages.[18,19] Several of these signals may be directly of interest in personalizing treatment (e.g., visualizing and discussing social or spending behaviors as part of cognitive–behavioral therapy [CBT]).

In addition, wearables can collect longitudinal physiological measurements, such as electrodermal activity, usually measured as skin conductance and cardiac signals (e.g., pulse rate variability) through photoplethysmography. These measurements can provide insights into changes in a patient's circadian rhythm or autonomic nervous system activity (e.g., due to stress or autonomic dysfunction).[20,21] Historically, studies of these physiological signals in depression have been limited to brief measurements during a laboratory visit (e.g., showing electrodermal hypoactivity in depression),[22] which can be confounded by short-term events such as a stressful commute. Comfortably worn autonomic sensors that reliably collect these data in daily life, while minimizing artifacts, have recently been cleared by the Food and Drug Administration (FDA) and can be used to collect longitudinal data.[23] Reliable physiological information may advance the management and understanding of depression by providing greater insight into the dynamics of psychophysiological factors in patients' daily lives. However, such wearable-based physiological measures have not yet been tested in controlled trials to confirm this potential clinical benefit.

PROCESSING OF DATA FROM WEARABLES USED TO MONITOR DEPRESSION

Raw measurements need to be cleaned before the data are processed. Cleaning consists of removing the samples recorded when the device was not worn, filtering out bad-quality signals (e.g., those containing motion artifacts), and reducing noise in the data.[24]

After cleaning, raw signals are processed with feature extraction algorithms to obtain meaningful data, called low-level features. For example, the wristband accelerometer measures linear acceleration of the wrist.[25] This signal can be processed to extract and estimate more interpretable metrics such as movement intensity (e.g., for assessment of psychomotor impairment and sleep).[26] Given their high dimensionality and complexity, low-level features comprising long and frequently sampled time series of multimodal measurements are often too granular to directly inform clinical decisions. However, ensuring the integrity of these low-level features is critical for providing the quality of data needed for higher-level insights.

Low-level features are joined together in various ways and then aggregated to represent clinically meaningful metrics, such as daily total sleep time[27] or daily vigorous activity.[28] These metrics, termed high-level information, are grouped by construct into clinically meaningful categories (e.g., activity, sleep, stress, and communication). Estimates of stress can be computed from a combination of physiological measurements, such as electrodermal activity or pulse rate variability, and contextual information, such as motion, temperature, and location.[24] Communication patterns can be measured from smartphone activity metrics, including the number of initiated and answered calls and the number of sent and received messages.[10] Aggregation (e.g., averaging measurements from the minute level to the day level) may result in a feature such as the number of sleep hours per day. Alternatively, aggregation may combine low-level features (e.g., sleep and pulse rate, for a sleep and context baseline). The plots in Figures 2, 3, and 4 show high-level information regarding activity levels and sleep patterns, aligned with depression levels from a PHQ-9 modified to measure symptoms of depression over the previous 24 hours. Having these

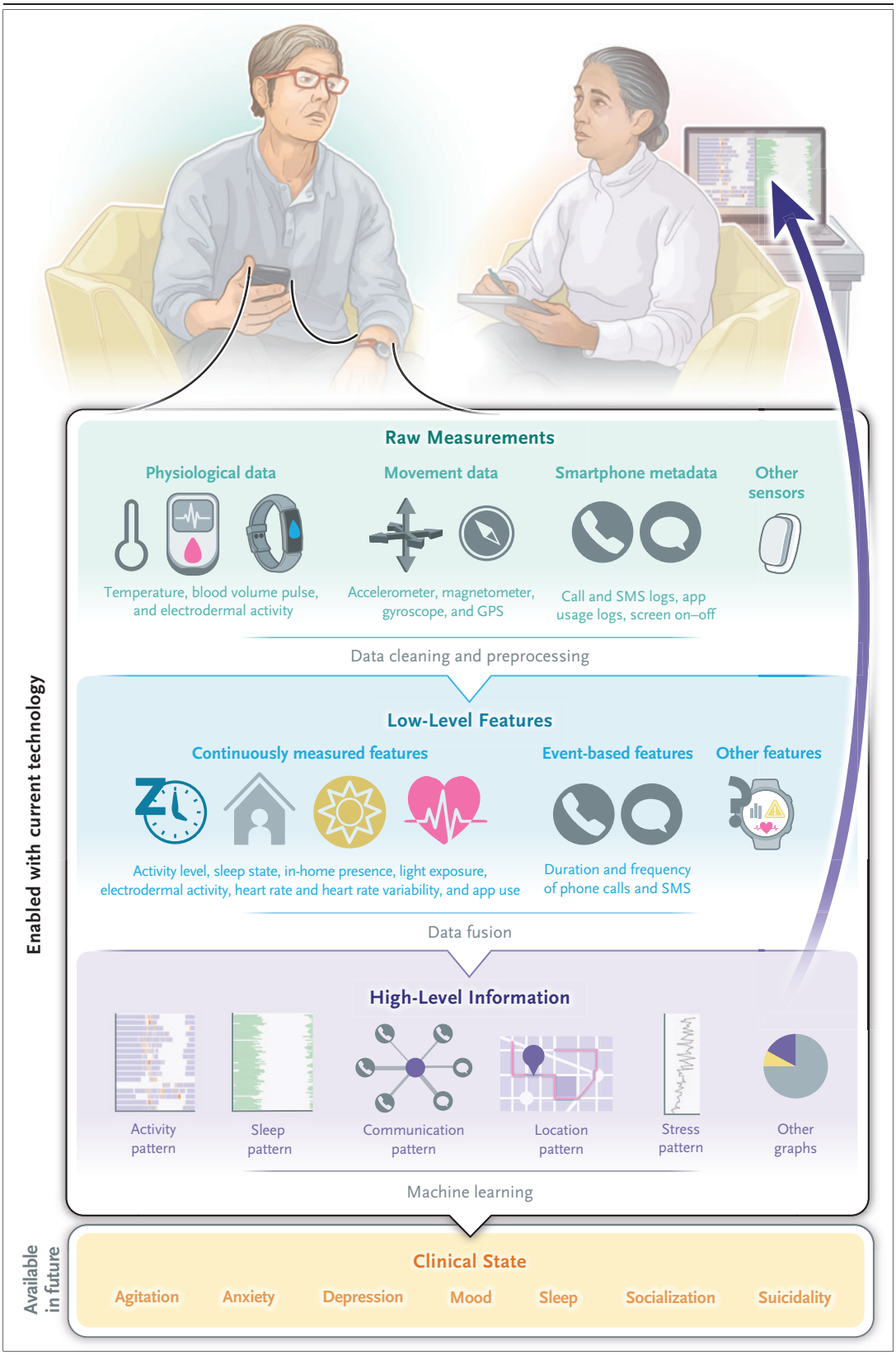
Enabled with current technology
Raw Measurements
Physiological data
Movement data
Smartphone metadata
Other sensors
Temperature, blood volume pulse, and electrodermal activity
Accelerometer, magnetometer, gyroscope, and GPS
Call and SMS logs, app usage logs, screen on–off
Data cleaning and preprocessing
Low-Level Features
Continuously measured features
Event-based features
Other features
Activity level, sleep state, in-home presence, light exposure, electrodermal activity, heart rate and heart rate variability, and app use
Duration and frequency of phone calls and SMS
Data fusion
High-Level Information
Activity pattern
Sleep pattern
Communication pattern
Location pattern
Stress pattern
Other graphs
Machine learning
Available in future
Clinical State
Agitation
Anxiety
Depression
Mood
Sleep
Socialization
Suicidality

Figure 1 (facing page). Layers of Data Provided by Wearable Devices.

The top layer of data (green) comprises raw measurements from the wearable sensors. After the data are cleaned, low-level features, shown in the second layer (blue), are obtained. Combinations of these features, usually over a period of days, and with consideration of contexts such as location and time of day, reveal high-level patterns of potential clinical relevance, shown in the third layer of data (purple). The data in the bottom layer (gold) indicate the clinical states related to depressive symptoms. Current technology enables access to only the first three layers of data. Ongoing research focused on building predictive machine-learning algorithms is likely to make the fourth layer of information available in the near future. GPS denotes global positioning system, and SMS short message service.

types of information side by side gives the clinician a more holistic view of the patient's health, which may be useful.

The rapid progress in machine learning on large data sets has led to optimism that algorithms may someday be able to automatically estimate the values of clinical states related to major depressive disorder.[36-38] For example, an algorithm could take raw measurements from wearables together with other contextual data (e.g., weather, time of day, and other information shared by the patient) as inputs and produce a score for the severity of depressive symptoms. Such scores are not yet validated for clinical use. Their validation requires large data sets, with accurately labeled data from diverse patient populations and diagnoses, to achieve accuracies that are useful clinically.

Substantial progress has been made in automatically estimating some of the high-level states, such as sleep. However, other states, such as suicidality, do not yet have reliable estimators and remain active research areas.[49] More generally, further improvements in the accuracy, explainability, and generalizability of these methods are required before they are used in clinical practice. Furthermore, machine-learning algorithms can amplify biases in the data they are trained on, which may not, for example, contain data from patients with diverse demographic characteristics. Care must be taken to detect and correct these algorithmic biases and to make them transparent when they are present.[50] Therefore, large studies, with suitably diverse patient populations, are required before these machine-learning algorithms can be evaluated for inclusion in clinical practice.

A notable challenge in processing wearable data is patient heterogeneity. The same value from a particular sensor may mean different things for different people. For instance, sedentary people and athletes usually have very different resting pulse rates. Data normalization methods are usually required for making comparisons across different people. Normalization is sometimes also required for comparisons within the same person when large changes have occurred that can affect measurement baselines. For example, a patient may have a reduced level of activity that is the result of a physical injury rather than depression.

A CASE STUDY OF WEARABLE DATA FROM A REAL PATIENT

Consider the wristband data for Maria (Figs. 2, 3, and 4), a 21-year-old woman with a diagnosis of major depressive disorder who started taking estrogen for depression 2 weeks before commencing wearable data collection in a study. The patient's name and identifying details have been changed to protect her privacy. Figure 2 shows 16 days of her activity patterns, from a low level of activity to a high level, as processed from her wristband accelerometer.[29] The arrows in Figure 2A point to a pattern of regular moderate activity, which occurs at weekly intervals, during weekends, and around the morning hours. The arrows in Figure 2B point to days with less than 5 total minutes of moderate or vigorous activity. The juxtaposed patterns in Figure 2B and 2C show no visible correlation between activity level and depressive symptoms over this short period. However, a view of the same data over a period of 6 weeks (Fig. 3) shows that Maria's activity pattern has become sustained and is associated with a decrease in her depressive symptoms. This correlation of change in self-reported depression and activity level is consistent with, but does not imply, causation.

Figure 4 shows Maria's sleep patterns. Because Figure 4A includes regions of missing data, we cannot conclusively say that her sleep is as regular as it looks. Nevertheless, the data over a 6-week period (Fig. 4B) suggest an improvement in Maria's hypersomnia, with more normal levels of sleep (7 to 8 hours per night). CBT for

insomnia or hypersomnia (CBT-I) can significantly reduce depressive symptoms in many patients.[30,31] Thus, reviewing charts like these may be useful for either identifying the need for a sleep therapy or tracking adherence and outcomes once sleep therapy has been started.

Although it may be tempting to look for simple connections between daily self-reported symptoms and wearable data that are generalized consistently across patients, these connections are often hard to find. In an analysis of multiple studies[18] that were based on between-group (or between-patient) comparisons, De Angel et al. found that daily total sleep time showed mixed directionality of significance. This finding is to be expected, given that some people with depression are known to sleep less than nondepressed people, whereas others sleep more.[32] De Angel et al. noted that among 12 studies in which "activity levels as a feature within physical activity" were measured as an index of depression, 3 of the 12 showed that this feature from wearable or smartphone data was closely correlated with the severity of depression. This finding indicates that the psychomotor retardation observed in depression can be captured by objective measures from

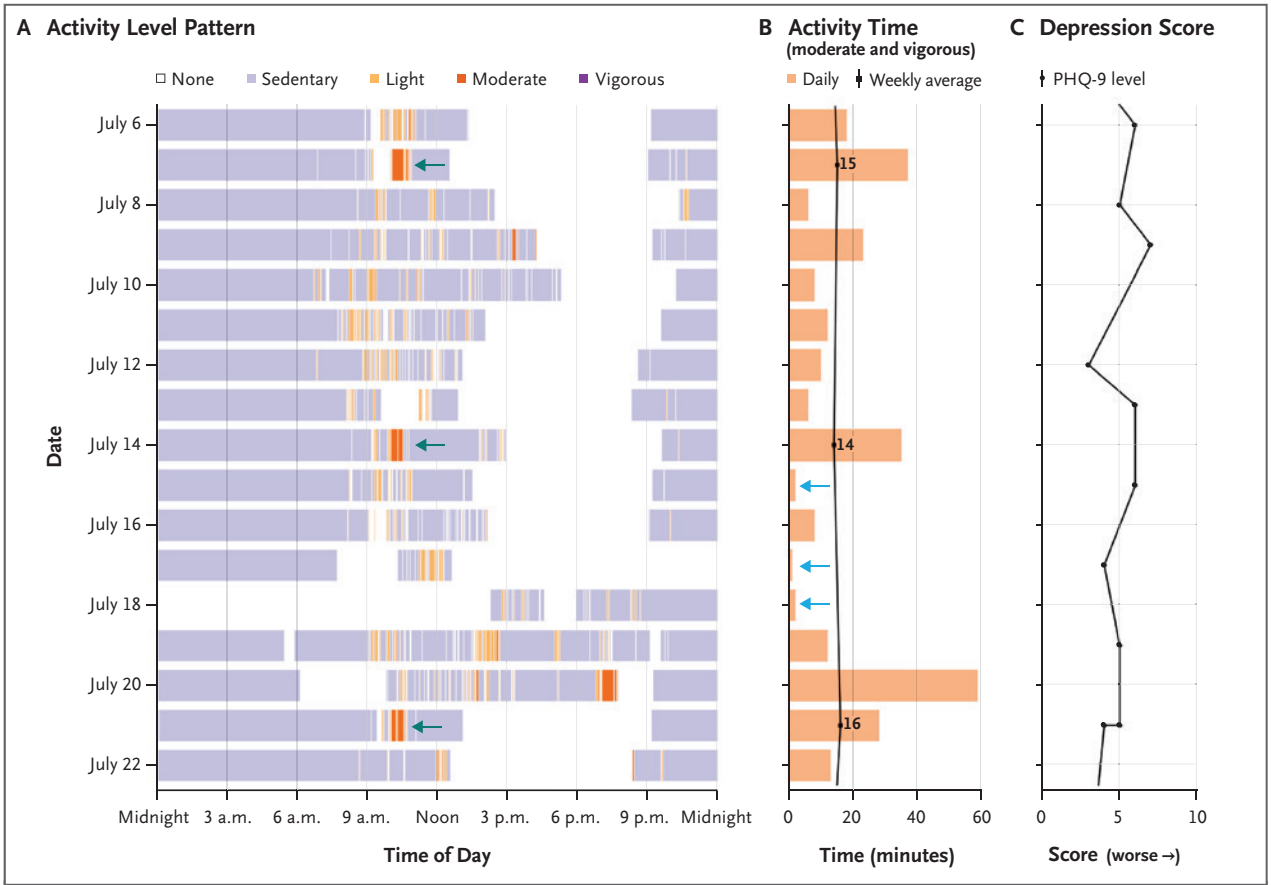

Figure 2. Activity and Depression Data for 16 Days in a Real Patient.

Panel A shows the number of times each day when the patient was sedentary or engaged in light, moderate, or vigorous activity, as measured by an accelerometer worn on the wrist. White areas indicate that the sensor was not being worn or was not collecting data. Panel B shows the daily total minutes of moderate or vigorous activity and the average for the week (black numbers). Panel C shows the level of depression as measured by scores on a modified version of the Patient Health Questionnaire 9 (PHQ-9), based on symptoms from the previous 24 hours. PHQ-9 scores range from 0 to 27, with higher numbers indicating more severe symptoms of depression. Scores of 5, 10, 15, and 20 represent thresholds for mild, moderate, moderately severe, and severe depressive symptoms, respectively. A 5-point change is clinically significant. The activity data in Panel A show that the patient was exercising moderately every 7 days (arrows) and also had other periods of intense activity. On certain days, as shown in Panel B, the patient spent less than 5 minutes in an active state (arrows). Over the 16-day period, there is no visible correlation between active time and PHQ-9 score.

wearables in certain scenarios. Since mental health disorders are often highly idiosyncratic, however, RCTs that seek causal factors across a group may not have findings that apply to each individual patient.[33]

In contrast, wearables enable high-frequency, longitudinal monitoring of individual patients, and the data can be used as another way to gather evidence about related symptoms and treatment effectiveness through N-of-1 methods.[34] These methods may, for example, look for patient-specific (or within-patient) correlates among objective wearable metrics, treatment decisions (e.g., medication or psychotherapy recommended), and desired outcomes (e.g., overall depression levels). Although idiographic, N-of-1 evidence-gathering methods also have limitations, we support their use alongside conventional RCT sources of nomothetic evidence in depression-related clinical practice, since the objective longitudinal data may produce findings that align more closely with the patient's experience.

LIMITATIONS TO BE ADDRESSED

Multiple obstacles currently curtail the adoption of wearable technology in clinical practice. One limitation is the clinical interpretability of outputs. Often, these frequently sampled, longitudinal measurements are high-dimensional and interact with each other in a complex manner. Their association with depression is heterogeneous and

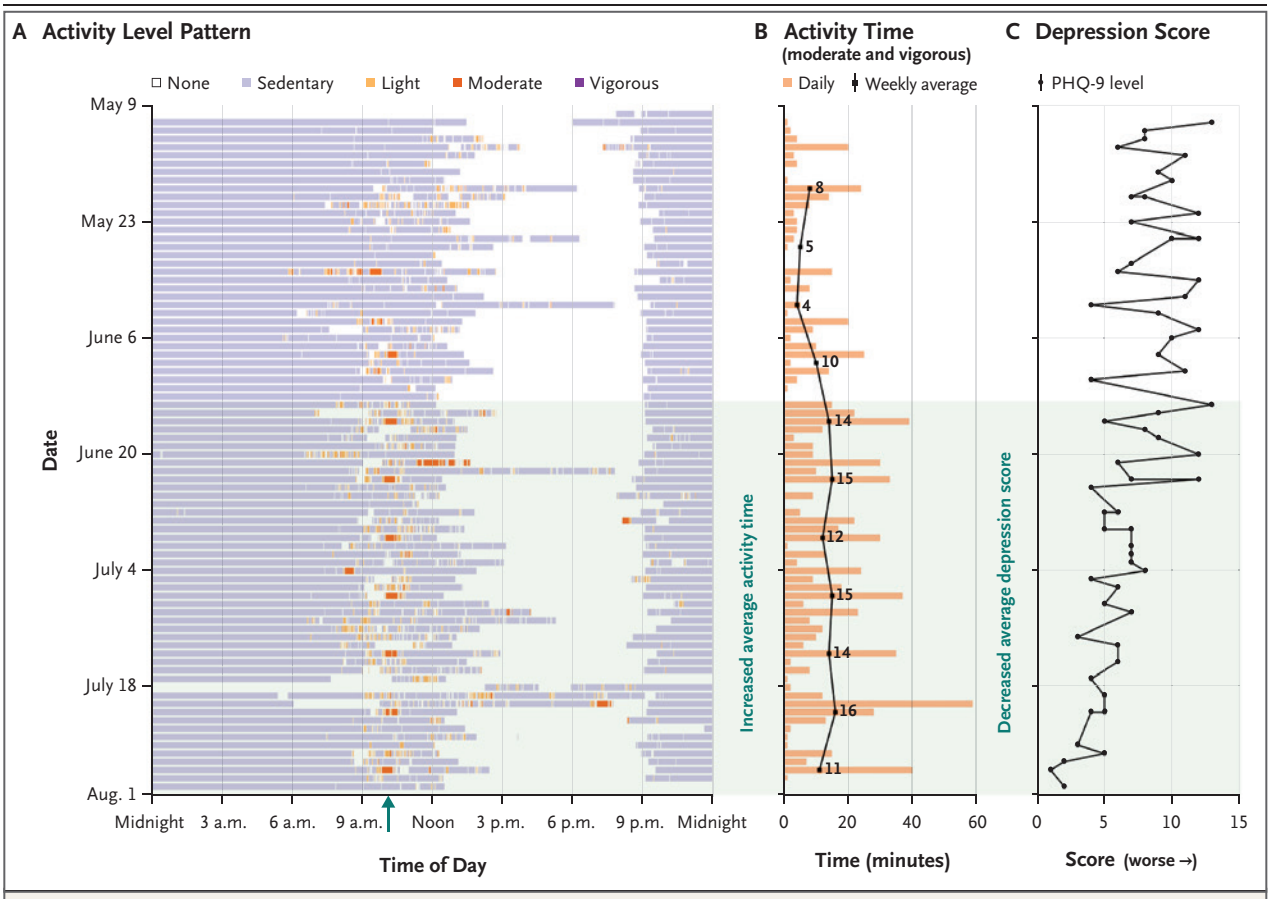

Figure 3. Activity and Depression Data over a 6-Week Period in the Same Patient.

Six weeks of raw activity data (Panel A) and activity averages (Panel B) are aligned with patient-reported PHQ-9 scores (Panel C), showing that the patient's activity levels were associated with her reduced depression scores. At the beginning, the patient had a low level of daily activity, which increased in the second part of treatment (green shading in the three panels). The increase in activity was confirmed on a plot depicting weekly average activity time (black numbers in Panel B). The average weekly minutes spent at moderate or vigorous activity levels tripled over the 6-week period. The patient also started to exercise regularly in the second week (arrow in Panel A). The patient's depression started to decrease in the second half of the recorded period.

can be difficult to interpret. Several strategies, as yet untested, may improve the usefulness of wearable measurements. First, high-level features should be codesigned with clinicians to ensure clinical meaningfulness. For example, the intensity of movement can be presented as different metrics such as the Euclidean norm (i.e., the square root of the sum of the squares of the three-axis accelerometer values) or the average time spent at a vigorous activity level.[35] The latter is likely to be more relevant to a clinician, who may recommend an increased duration of vigorous activity and then monitor whether the patient follows the recommendation and has clinical improvement. Second, visualization tools showing the longitudinal pattern of clinically relevant information should be developed. For example, clinicians may be interested in patterns of a patient's activity and how these change across visits and treatments.

Adherence is often a challenge for longitudinal monitoring. In our studies, in which participants were asked to wear sensors on both wrists continuously for up to 8 weeks, adherence was about 70%.[36,39] In another study, in which participants were monitored with a wearable for up to 24 months, adherence was not affected by the severity of depression.[40] However, research participants are often remunerated for wearing sensors, and such financial incentives are not part of clinical practice. Although the acceptability of wearables has been shown for various patient populations, including patients with severe psychopathology,[41] further research is required in

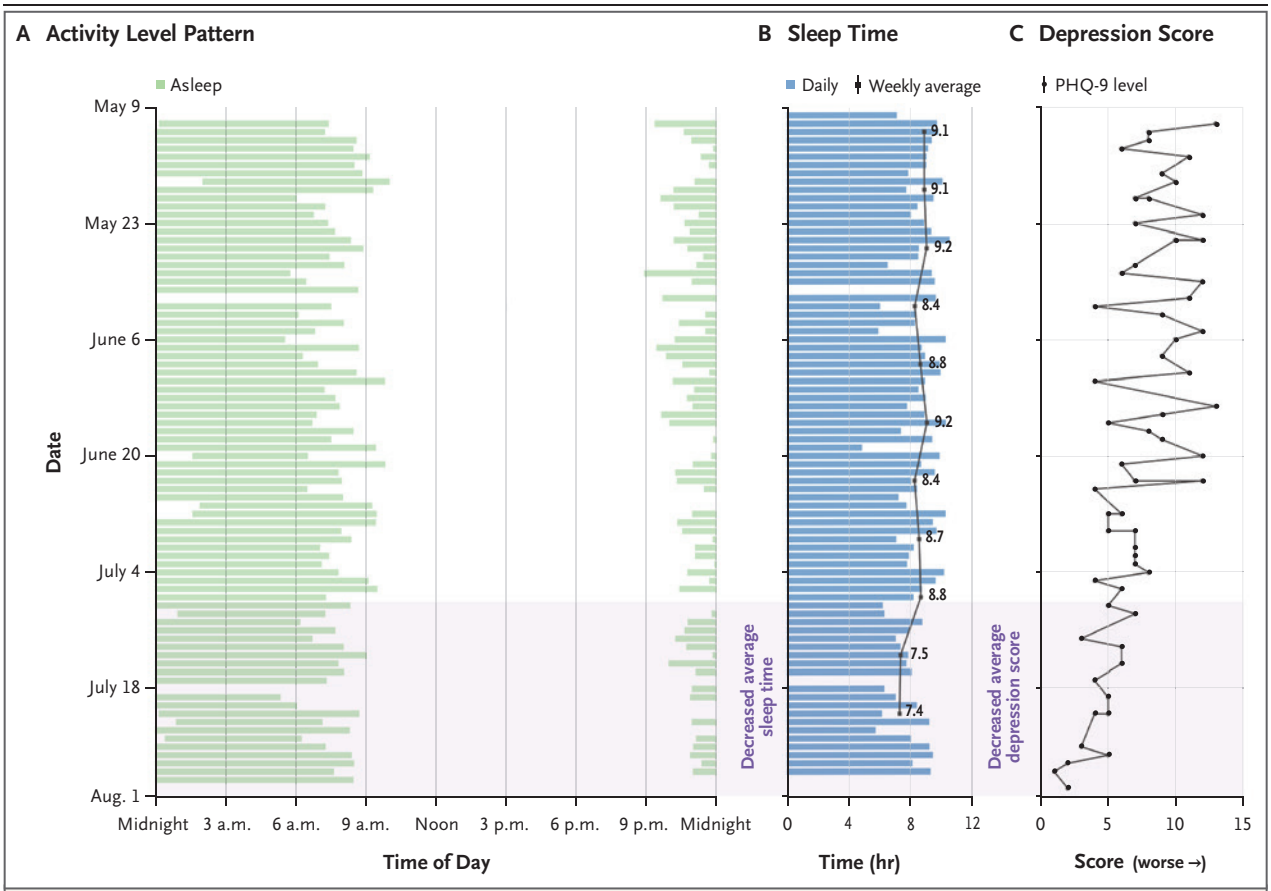

Figure 4. Activity Level, Sleep Time, and Depression Score over a 6-Week Period in the Same Patient.

Sleep patterns, estimated from wrist accelerometer data, are aligned with PHQ-9 scores, showing whether the patient's sleep changes were associated with her reduced depression scores. Panel A shows times when the patient was asleep. Panel B shows daily total and weekly average sleep hours. At the beginning of treatment, Maria slept more than 10 hours on certain nights, but she reduced her daily sleep time to about 7 hours in the last 2 weeks of recorded data. Her weekly average sleep time dropped from 9.1 to 7.4 hours, as shown in Panel B (black numbers). Panel C shows the patient-reported daily PHQ-9 scores.

order to determine how to increase wearable adherence in clinical practice.[42] In particular, access to the patient's wearable data and an opportunity to review the data with the clinician have been shown to increase the patient's engagement in treatment.[43]

Another limitation of wearables that has led to the delay in their deployment in psychiatry is that wearables often provide only high-level information, such as daily sleep time, without the manufacturer's disclosure of the algorithms or raw data used to arrive at this output. The lack of access to high-quality raw data from some wearables prevents systematic evaluation by the scientific community of the accuracy and robustness of algorithms used to estimate low-level features by these devices. Another problem is that manufacturers of wearables often provide data processing as part of the device's function but may not notify clinicians or researchers if they update their processing algorithms, and these updates can change study results.[44] Consequently, clinical studies should prioritize the use of devices that provide raw data and that disclose the algorithms used to generate high-level information. Use of such devices aids the development and validation of digital mental health systems.

ETHICS IN THE USE OF WEARABLE DEVICES FOR DEPRESSION

The ethical use of wearables in clinical practice requires careful consideration, and safeguards should be established to minimize risks.[45] In clinical practice, it seems reasonable to limit use to devices that have been cleared by the FDA or similar respected regulatory organizations. Such devices will have undergone quality checks with respect to any data listed in their FDA-cleared indications for use, as well as biocompatibility, safety, usability, and cybersecurity testing that is usually not required for consumer devices. Other important factors to consider include technical support for wearers, researchers, and clinicians, together with battery life, cost, size, interoperability with different smartphone platforms, and the stigma that may be associated with the appearance of visible devices.[46]

Data fairness, security, and privacy are of paramount importance, since wearables measure many health-related signals that are considered to be sensitive. Care must be taken to ensure that wearable data remain secure (e.g., with encryption, deidentification, and dedicated personnel who manage the information and monitor it for security breaches). Moreover, work is required to ensure that wearables provide equitable results across heterogeneous populations. The FDA acknowledged that pulse oximetry may be less accurate for persons with dark skin pigmentation.[47] The main sensor (photoplethysmography) used in wearable pulse rate analysis can also have this problem.[48] Finally, informed consent and transparency are critical for the ethical use of wearables. Patients have the right to choose whether they will use a wearable in their treatment and to know how their data will be used should they decide in favor of a wearable. All these factors are essential for the future use of wearables in the clinical treatment of depression and for conducting the large studies needed to increase our understanding of wearable data and their potential for improving patient outcomes.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Justin Baker for his shared insights and Amit Chopra for his critical review of an early draft of the manuscript and helpful suggestions.

REFERENCES

1. Fortney JC, Unützer J, Wrenn G, et al. A tipping point for measurement-based care. Psychiatr Serv 2017;68:179-88.
2. Falvey JE. Science or serendipity? What we know about clinical judgment in mental health. J Ment Health Couns 1992; 14:458-70.
3. Gambrill ED. Critical thinking in clinical practice: improving the accuracy of judgments and decisions about clients. San Francisco: Jossey-Bass, 1990.
4. Rabinowitz J, Efron NJ. Diagnosis, dogmatism, and rationality. J Ment Health Couns 1997;19:40-56.
5. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16:606-13.
6. Torous J, Staples P, Shanahan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8.
7. Colombo D, Fernández-Álvarez J, Patané A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019;8:465.
8. Barron D. Reading our minds: the rise of big data psychiatry. New York: Columbia Global Reports, 2021.
9. Baker JT, Germine LT, Ressler KJ, Rauch SL, Carlezon WA Jr. Digital devices and continuous telemetry: opportunities for aligning psychiatry and neuroscience. Neuropsychopharmacology 2018;43:2499-503.
10. Onnela J-P, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016;41:1691-6.

11. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017;13:23-47.
12. Kleiman EM, Turner BJ, Fedor S, et al. Digital phenotyping of suicidal thoughts. Depress Anxiety 2018;35:601-8.
13. Cheng P, Luik AI, Fellman-Couture C, et al. Efficacy of digital CBT for insomnia to reduce depression across demographic groups: a randomized trial. Psychol Med 2019;49:491-500.
14. Kang HS, Exworthy M. Wearing the future — wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022;10(7):e35684.
15. Blackburn TP. Depressive disorders: treatment failures and poor prognosis over the last 50 years. Pharmacol Res Perspect 2019;7(3):e00472.
16. Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. Acta Psychiatr Scand 2020;141:350-5.
17. Tang M, Nakamoto CH, Stern AD, Mehrotra A. Trends in remote patient monitoring use in traditional medicare. JAMA Intern Med 2022;182:1005-6.
18. De Angel V, Lewis S, White K, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022;5:3.
19. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018;9:743.
20. Germain A, Kupfer DJ. Circadian rhythm disturbances in depression. Hum Psychopharmacol 2008;23:571-85.
21. Sgoifo A, Carnevali L, Alfonso MdeL, Amore M. Autonomic dysfunction and heart rate variability in depression. Stress 2015;18:343-52.
22. Sarchiapone M, Gramaglia C, Iosue M, et al. The association between electrodermal activity (EDA), depression and suicidal behaviour: a systematic review and narrative synthesis. BMC Psychiatry 2018;18:22.
23. Lee JJ. K221282: Empatica health monitoring platform. Silver Spring, MD: Food and Drug Administration, November 20, 2022 (https://www.accessdata.fda.gov/cdrh_docs/pdf22/K221282.pdf).
24. Smets E, Rios Velazquez E, Schiavone G, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1:67.
25. Godfrey A, Conway R, Meagher D, ÓLaighin G. Direct measurement of human movement by accelerometry. Med Eng Phys 2008;30:1364-86.
26. Perry B, Herrington W, Goldsack JC, et al. Use of mobile devices to measure outcomes in clinical research, 2010-2016: a systematic literature review. Digit Biomark 2018;2:11-30.
27. Smith MT, McCrae CS, Cheung J, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2018;14:1231-7.
28. Pearce M, Garcia L, Abbas A, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry 2022;79:550-9.
29. Migueles JH, Cadenas-Sanchez C, Rowlands AV, et al. Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. Sci Rep 2019;9:18235.
30. Cunningham JEA, Shapiro CM. Cognitive Behavioural Therapy for Insomnia (CBT-I) to treat depression: a systematic review. J Psychosom Res 2018;106:1-12.
31. Ong JC, Dawson SC, Mundt JM, Moore C. Developing a cognitive behavioral therapy for hypersomnia using telehealth: a feasibility study. J Clin Sleep Med 2020;16:2047-62.
32. Geoffroy PA, Hoertel N, Etain B, et al. Insomnia and hypersomnia in major depressive episode: prevalence, sociodemographic characteristics and psychiatric comorbidity in a population-based study. J Affect Disord 2018;226:132-41.
33. Tanious R, Onghena P. Randomized single-case experimental designs in healthcare research: what, why, and how? Healthcare (Basel) 2019;7:143.
34. Barlow DH, Nock M, Hersen M. Single case experimental designs: strategies for studying behavior for change. 3rd ed. Boston: Pearson, 2009.
35. van Hees VT, Gorzelniak L, Dean León EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013;8(4):e61691.
36. Pedrelli P, Fedor S, Ghandeharioun A, et al. Monitoring changes in depression severity using wearable and mobile sensors. Front Psychiatry 2020;11:584711.
37. Taylor S, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affect Comput 2020;11:200-13.
38. Moshe I, Terhorst Y, Opoku Asare K, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247.
39. Massachusetts Institute of Technology. Sensor-based characterization of depression study (SENSCODE) (https://senscode.media.mit.edu/).
40. Matcham F, Leightley D, Siddi S, et al. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022;22:136.
41. Kleiman E, Millner AJ, Joyce VW, Nash CC, Buonopane RJ, Nock MK. Feasibility and acceptability of using wearable physiological monitors with suicidal adolescent inpatients. July 13, 2019 (http://preprints.jmir.org/preprint/13725/accepted). preprint.
42. Simblett S, Matcham F, Siddi S, et al. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 2019;7(1):e11325.
43. Omoloja A, Vundavalli S. Patient generated health data: benefits and challenges. Curr Probl Pediatr Adolesc Health Care 2021;51:101103.
44. Onnela J-P. Exporting the same data from a wearable twice doesn't give you the same data (http://www.beiwe.org/exporting-the-same-data-from-a-wearable-twice-doesnt-give-you-the-same-data/).
45. Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. NPJ Digit Med 2018;1:68.
46. Polhemus AM, Novák J, Ferrao J, et al. Human-centered design strategies for device selection in mHealth programs: development of a novel framework and case study. JMIR Mhealth Uhealth 2020;8(5):e16043.
47. Food and Drug Administration. Pulse oximeter accuracy and limitations: FDA safety communication. November 7, 2022 (https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication).
48. Colvonen PJ, DeYoung PN, Bosompra NA, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. Sleep 2020;43(10):zsaa159.
49. Kleiman EM, Bentley KH, Maimone JS, et al. Can passive measurement of physiological distress help better predict suicidal thinking? Transl Psychiatry 2021;11:611.
50. Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? AMA J Ethics 2019;21(2):E167-E179.