

Building the Thermometer for Mental Health

By Joshua J. Chauvin, Ph.D., and Thomas R. Insel, M.D.



Source/Shutterstock

Editor's Note: Millions of people suffer from serious mental illness, but very few receive consistent coordinated care. Since leaving his post in 2015 after 13 years as director of the National Institute of Mental Health, co-author Tom Insel has been on a mission to use technology (such as mining your smartphone) to better understand your state of mind and treat depression, schizophrenia, and other disorders. Insel and co-author Joshua Chauvin, part of the team at a healthcare innovation company, examine the potential and pitfalls of this next digital frontier.

Imagine that you visit your physician complaining of a fever and, rather than taking out a thermometer, they begin hovering their “educated hands” over you. Gradually, they press down against your arm to gain a full impression of your skin’s temperature and the “deeper seated combustions.” Removing their hand, they look closely at your appearance and pronounce their assessment: you do, in fact, have a fever. You might (justifiably) be dubious.

Thankfully, clinicians today have an inexpensive, ubiquitous tool to measure a patient’s temperature. But it wasn’t always this way.

¹ The first person to devise a thermometer, an instrument to track temperature, was an Italian physician, Santorio Santorio, who in the early 17th century described a device for measuring the expansion of water or alcohol with heat. It was a century later that Fahrenheit, a Polish-born Dutchman who was both a physicist and a glass blower, used mercury instead of water or alcohol and created the temperature-measuring scale that we continue to use today. But this device could not be used in clinical practice. The early versions were cumbersome and, for over a century after Fahrenheit, no one knew how to connect the measurement of body temperature to the state of disease. Indeed, until late in the 19th century, the physician’s hand was the standard medical instrument for detecting a fever.

Lack of Objective Measurement in Psychiatry

Just as the thermometer provided a standardized, objective measurement for detecting fever, tools to quantify health and disease parameters have transformed medicine in almost every major disease area—electrocardiograms for heart disease, blood glucose for diabetes, and, recently, genetic diagnostic tests for cancer. But when it comes to brain health, and in the case of mental illness especially, progress has been uneven. Although direct brain imaging instruments exist, most (MRI, PET, MEG) are expensive, inaccessible to many, rarely useful for deciding the treatment of an individual patient, and time-intensive to administer. While they can identify brain lesions in multiple sclerosis or dementia, they are less useful in mental disorders. This lack of measurement matters because, to borrow a truism from business, “we don’t manage well what we don’t measure well.”

In the absence of reliable instruments, clinicians treating mental illness use indirect, intuition-based measures. The DSM-5, the principal schema for classifying mental disorders, requires clinicians to form diagnoses based on their subjective judgments and arbitrary cut-off points. For instance, the diagnosis of major depressive disorder requires five of eight features (such as diminished interest or pleasure in activities, feelings of worthlessness, or diminished ability to think or concentrate) to persist for two weeks—based on patient or family reports.

During treatment, similarly, patients are not routinely monitored with objective assessments. Some clinicians employ self-report questionnaires, like the PHQ-9 (a nine-item scale to rate depressive symptoms), but these scales are only modestly correlated with ratings of trained observers.² The self-report information is, of course, important to monitor but, like reports of chest pain or headache, usually proves insufficient. (That is, they have relatively low inter-rater reliability, do not assess patients in real world settings, and often cannot reliably attribute the effects of a given intervention). Moreover, only 18 percent of US psychiatrists and 11 percent of psychologists routinely use symptom-rating scales or Patient-Reported Outcome Measures (PROMs) to monitor patient improvement.^{3,4} Thus, for the vast majority of patients with a mental illness, measurement often comes down to “How are you feeling?” during sporadic, brief visits in primary care. This is a bit like our example of the doctor trying to determine whether your temperature is increasing using only his or her hands and clinical experience to guide them. Or treating hypertension without a blood pressure cuff or diabetes without a glucometer.

There has been a push toward using “measurement-based care” that relies on standard rating scales and patient-reported outcomes, and good evidence that it can improve clinical outcomes.⁵ But this approach has its limitations. Practically speaking, standard assessments can be difficult to implement, held back by a lack of financial support and limited personnel to administer the tests. They increase paperwork, which can burden stretched clinicians.⁶ Perhaps most problematic, these measurement tools are necessarily brief and can capture only a narrow spectrum of a patient’s overall state (e.g., general depression symptoms). And since they are administered infrequently, usually in the clinic, they of necessity collect one-time, or “snapshot,” impressions of a person’s mental health.⁷

How can we move beyond this state of affairs?

Let's imagine the ideal form of measurement for mental health. In addition to being objective, it would be continuous (assessing symptoms frequently) and precise (both sensitive and specific) and collected in the "real world" (outside the context of the clinical encounter). It should give clinicians access to summarized and up-to-date patient data (e.g. on symptom severity), easily interpretable to provide meaningful, clinically actionable information.^{8,9} Such information would enable clinicians to measure response to treatment in real-time on an ongoing basis and to adjust treatment plans based on the patient's preference and response. Finally, to be effective and scale to global populations, the measurement should be passive—done without asking individuals to change their behavior or do anything on top of what they are already doing. Taken together, the combination of attributes would help to ensure early and timely intervention. Instead of the current model of care, which is largely reactive (administered when someone is presenting symptoms), better measurement that is objective, continuous, and passive can move the health system towards more proactive and preventive interventions.

The Hope: Advent of Digital Phenotyping

If such a tool sounds implausible, consider for a moment recent advances in information technology and data science.

In 2011, the World Health Organization stated: "The use of mobile and wireless technologies to support the achievement of health objectives (mHealth) has the potential to transform the face of health service delivery across the globe."¹⁰ Since then, smartphone subscriptions have increased more than five-fold (from 856 million to more than five billion today), with projections to reach nearly seven billion by 2022.¹¹ There's also been astonishing growth in broadband access, even in areas without easy access to clean water.¹²

Over the same period, there have been significant advances in data science too, including the advent of machine learning, which can find patterns in large data sets that were not evident using conventional statistical approaches. These developments are already transforming healthcare: diagnostic testing is beginning to incorporate a form of machine learning called neural networks,¹³ healthcare systems are taking advantage of machine learning to help triage and streamline patients

through services,¹⁴ and predictive modeling—the analysis of past and current data to forecast outcome—is using electronic health records to drive personalized medicine and improve healthcare quality.¹⁵

The increasing ubiquity of smartphones and advent of technologies, such as home devices (Amazon Echo and Google Home) and wearables (FitBit, Apple Watch), that can act as a reliable source of measurement, combined with advances in analyzing continuous data, presents us for the first time with an opportunity to monitor brain function at population scale. This approach, called *digital phenotyping*, is a two-step process that works by applying machine learning to data collected from digital devices such as wearables and smartphones.¹⁶ Obtaining the signals from the phone or wearable device is the first step. Making sense of these signals by finding the patterns that correlate with clinical state is often the more difficult second step. To find patterns in complex data most researchers have used machine learning, a powerful statistical approach that can extract predictive features from large data sets.¹⁷ Machine learning is a rapidly evolving field which promises to improve our ability to find clinically-relevant signals with each iteration. With it, information from sensors (e.g. physical activity, location, heartrate), keyboard interactions, and other features such as voice and speech can be analyzed to provide insight into changes in a person’s behavior, their psychological state, and cognitive function. The approach can even provide predictors of risk.

Examples of using activity monitors and motion sensors to monitor the behavior of patients with mental illness — what we are now calling the digital phenotype — have been around since at least the 1980s.^{18,19} Today, studies continue to demonstrate that patterns of activity and geolocation can herald mania or depression²⁰ and sleep actigraphy can predict suicidal ideation;²¹ moreover, other biosensors have shown that heart rate variability can help predict Post Traumatic Stress Disorder diagnosis²² and speech and voice, which can reveal important aspects of our emotional, social and psychological worlds, may be able to provide insight into depression.²³

While signals from actigraphy and voice have proven to be predictive, they are also noisy and *nosey*. One particularly promising approach to developing digital phenotypes of cognition that might help to move the field beyond these concerns involves data from human-computer interaction (HCI). HCI-based digital biomarkers can be generated from passively-collected, content-free interactions, like

typing and scrolling patterns on a smartphone, measuring the latency between space and character in a text or the interval between scroll and a click. This approach was originally developed in cybersecurity to track hackers with what was called “digital fingerprinting.” (Based on an individual’s pattern of activity, every individual who spends time online leaves a unique trace, which can be used to create identifiers for individual users—hence the notion of a digital *fingerprint*.) Applying this concept to mental health, scientists have developed digital *biomarkers* that strongly correlate with performance on traditional cognitive tests and with mood ratings.²⁴ With the average user’s output of over 2,600 smartphone touches a day,²⁵ these ubiquitous computer interactions can reveal a lot about how we think and how we feel and when combined with other measures like sleep, activity, and speech, create a digital phenotype.

Supplementing clinical impressions and subjective, episodic assessment, digital phenotyping offers an opportunity to move towards objective, measurement-based care. For psychiatry, it could bring brain health measures to the population, and with it the ability to target care and intervention to high-risk patients, extending independence and improving productivity. Applications could include screening, early detection, disease monitoring, precise diagnosis, and a new care model based upon these.

The Challenges

Let’s go back to the history of the thermometer for a moment. It wasn’t because thermometers didn’t exist that 19th century physicians were reluctant to change practice. As we have seen, they had in some form been available for 200 years. Nor was it because physicians weren’t aware that temperature was related to illness—that had been known since Hippocrates 2,000 years before. So, what was it that held the field back?

While physicians had a reliable instrument for measuring body heat, they didn’t know what a normal temperature range was. It was only with the discoveries made by Carl Wunderlich (1868), a psychiatrist who collated nearly 100,000 observations, that data could define normal and abnormal body temperature. At that point, with the clinical utility of the thermometer evident, it was routinely adopted in clinical practice as part of a complete medical evaluation. Temperature could be used as a biomarker for disease.²⁶

To gain widespread clinical use, digital phenotyping will need to overcome similar challenges, and a few contemporary hurdles as well.

As with body temperature, digital phenotyping needs to be tested in large, diverse populations to identify the digital biomarkers that matter. This means validating digital parameters against standard (if imperfect) measures of cognition and mood to determine which, if any, reliably give accurate, actionable data. The good news? There are already many ongoing large-scale clinical trials helping to validate this technology, and so far, the results have been promising.

But the clinical use of digital phenotyping presents ethical, legal, and social questions that the thermometer did not.²⁷ And there's a gap between demonstrating clinical value and achieving public trust: patients must be able to balance the benefits against real or perceived risks. No doubt building an evidence-base will be an essential step in this direction, but it will not be enough.

With the recent “teclash” against giant technology companies—consider the stir caused when Cambridge Analytica misused personal data²⁸ and the ongoing wave of negative news coverage for Facebook²⁹—such acceptance will require more than compliance with healthcare and privacy regulations. Besides protecting user data, digital tools must offer transparency and informed consent, and when there are questions of malpractice, users must be able to hold designers, providers, companies, or otherwise, accountable.³⁰ To pre-empt ethical transgressions, and build trust with patients, active engagement with users in the development of new technologies and careful consideration of users concerns is essential. Tech companies must also consider limiting the range of data they are collecting and consider the potential invasiveness of their approach. The content-free digital phenotyping provided by human-computer interactions described above, for example, is likely to be more acceptable than approaches drawing upon personally identifiable information like voice or location.

Even with scientific backing and public trust, adoption and acceptance—by patients, clinicians, and healthcare systems—still presents significant challenges. Patients must *want* to engage with the new digital health tools. Of the more than 300,000 digital health apps currently on the market, a mere 41

account for the bulk of all downloads, while 85 percent have fewer than 5,000 installs.³¹ Clinicians must likewise be won over. According to an American Medical Association survey, current levels of use of digital health tools by clinicians remain low, with only 26 percent currently using patient engagement technologies (i.e., solutions for chronic conditions designed to promote patient wellness and active participation in their care, for example, through promoting adherence to treatment) and 13 percent using remote patient monitoring technologies designed for daily measurement.³² Among other reasons, clinicians reject these emerging tools because they're disruptive, time-consuming, unvalidated, and costly to use. For health systems to adopt digital health tools more broadly, progress must include better curation and evaluation of apps. This includes: establishing best practices around privacy and security; getting patients and providers to recognize the value; establishing regulatory guidelines and reimbursement models for payments; and making it easy for clinicians to integrate new technologies into their practice.³³

Finally, we must recognize that digital phenotyping is only one piece of the puzzle. Improved health outcomes require more than detection: if the smartphone becomes a digital smoke alarm, how do we put out the fire? For mental health, many of the best treatments involve communication, skill building, and a therapeutic relationship. All of these can be done on a phone, allowing a "closed loop" approach to mental healthcare, where digital phenotyping identifies a need and the treatment is delivered immediately by a remote clinician. The same phone can also monitor the impact of the treatment, making measurement-based care for an individual with depression or psychosis the equivalent of both the thermometer and the antibiotic for a patient with a fever.

The pioneering German psychiatrist Carl Wunderlich is said to have commented that "a physician who carried on his profession without employing the thermometer was like a blind man endeavoring to distinguish colors by feeling."³⁴ The same might one day be said about clinicians who don't adopt more objective measures of brain health. Whether digital phenotyping or some other method takes precedence, it is clear that practice as usual is no longer an option if we are to improve outcomes for people with mental illness.

Bios

Joshua J. Chauvin, Ph.D., is a partnerships manager for [Mindstrong Health](#), a start-up looking to transform the future of mental health care by developing better ways to measure and detect (passively) changes in mental health state. As partnerships manager, Chauvin is working to expand the company's European reach from the United Kingdom. He has previously worked as a research consultant at the Canadian Mental Health Association and was involved in a number of anti-stigmatization campaigns in Ontario, Canada. He has also chaired the Mind Your Head Campaign in Oxford and is the co-founder of the [It Gets Brighter Campaign](#), which shares and collects video messages of hope for young people struggling with mental health issues. Chauvin completed a DPhil in Experimental Psychology at the University of Oxford as both a Rhodes Scholar and a [Canadian Centennial Scholar](#), is an associate fellow at the [Evidence-Based Practice Unit](#), an advisor to [Jack.org](#), and an Honorary Senior Research Associate at University College, London.

Thomas R. Insel, M.D., a psychiatrist and neuroscientist, is a co-founder and president of Mindstrong Health, a start-up looking to transform the future of mental health care by developing better ways to measure and detect (passively) changes in mental health state. From 2002-2015, Insel served as director of the National Institute of Mental Health (NIMH), the component of the National Institutes of Health (NIH) committed to research on mental disorders. Prior to serving as NIMH director, Insel was Professor of Psychiatry at Emory University, where he was founding director of the Center for Behavioral Neuroscience in Atlanta. Most recently (2015–2017), he led the mental health team at Verily (formerly Google Life Sciences) in San Francisco. Insel is a member of the National Academy of Medicine and has received numerous national and international awards, including honorary degrees in the US and Europe.

Financial Disclosure

References

¹ Haller Jr, J. S. (1985). Medical thermometry--a short history. *Western journal of medicine*, 142(1), 108-116.

-
- ² Yonkers, K.A., & Samson, J. (2000). Mood disorder measures. In American Psychiatric Association Task Force for the Handbook of Psychiatric Measures (ed). *Handbook of Psychiatric Measures*. American Psychiatric Association, Washington DC, pp 515-548.
- ³ Zimmerman, M., & McGlinchey, J. B. (2008). Why don't psychiatrists use scales to measure outcome when treating depressed patients? *Journal of Clinical Psychiatry*, *69*, 1916–1919.
- ⁴ Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *17*(1), 25-32.
- ⁵ Fortney J., Sladek, R. Unützer J. (2015). Fixing behavioral healthcare in America. *The Kennedy Forum*, Washington D.C.
- ⁶ Hatfield, D. R., & Ogles, B. M. (2007). Why some clinicians use outcome measures and others do not. *Administration and policy in mental health and mental health services research*, *34*(3), 283-291.
- ⁷ Hirschtritt, M. E., & Insel, T. R. (2018). Digital Technologies in Psychiatry: Present and Future. *Focus*, *16*(3), 251-258.
- ⁸ Fortney J., Sladek, R. Unützer, J. (2015). Fixing behavioral healthcare in America. *The Kennedy Forum*, Washington D.C.
- ⁹ Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2016). A tipping point for measurement-based care. *Psychiatric Services*, *68*(2), 179-188.
- ¹⁰ Kay, M., Santos, J., & Takane, M. (2011). mHealth: New horizons for health through mobile technologies. *World Health Organization*, *64*(7), 66-71.
- ¹¹ Ericsson Mobility Visualizer. (2018, June 12). Retrieved September 16, 2018, from <https://www.ericsson.com/en/mobility-report/mobility-visualizer>
- ¹² Insel, T. R. (2018). Digital phenotyping: a global tool for psychiatry. *World Psychiatry*, *17*(3), 276-277.
- ¹³ Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Kim, R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, *316*(22), 2402-2410.
- ¹⁴ How we're helping today. (n.d.). Retrieved September 16, 2018, from <https://deepmind.com/applied/deepmind-health/working-partners/how-were-helping-today/>
- ¹⁵ Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Sundberg, P. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, *1*(1), 18.

-
- ¹⁶ Insel, T. R. (2017). Digital phenotyping: technology for a new science of behavior. *JAMA*, *318*(13), 1215-1216.
- ¹⁷ Bzdok, D., & Meyer-Lindenberg, A. (2017). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.
- ¹⁸ Porrino, L. J., Rapoport, J. L., Behar, D., Sceery, W., Ismond, D. R., & Bunney, W. E. (1983). A naturalistic assessment of the motor activity of hyperactive boys: I. Comparison with normal controls. *Archives of General Psychiatry*, *40*(6), 681-687.
- ¹⁹ Wolff, E. A., Putnam, F. W., & Post, R. M. (1985). Motor activity and affective illness: the relationship of amplitude and temporal distribution to changes in affective state. *Archives of General Psychiatry*, *42*(3), 288-294.
- ²⁰ Reinertsen, E., & Clifford, G. D. (2018). A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. *Physiological measurement*, *39*(5), 05TR01.
- ²¹ Bernert, R. A., Luckenbaugh, D. A., Duncan, W. C., Iwata, N. G., Ballard, E. D., & Zarate, C. A. (2017). Sleep architecture parameters as a putative biomarker of suicidal ideation in treatment-resistant depression. *Journal of affective disorders*, *208*, 309-315.
- ²² Reinertsen, E., Nemati, S., Vest, A. N., Vaccarino, V., Lampert, R., Shah, A. J., & Clifford, G. D. (2017). Heart rate-based window segmentation improves accuracy of classifying posttraumatic stress disorder using heart rate variability measures. *Physiological measurement*, *38*(6), 1061.
- ²³ Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, *54*(1), 547-577.
- ²⁴ Dagum, P. (2018). Digital biomarkers of cognitive function. *npj Digital Medicine*, *1*(1), 10.
- ²⁵ Winnick, M. (2016, June 16). Putting a Finger on Our Phone Obsession. Retrieved September 16, 2018, from <https://blog.dscout.com/mobile-touches>
- ²⁶ Haller Jr, J. S. (1985). Medical thermometry--a short history. *Western journal of medicine*, *142*(1), 108-116.
- ²⁷ Martinez-Martine, N., Insel, T.R., Dagum, P., Greely, H. T., & Cho, M.K. (*in press*). Data mining for health: staking out the ethical territory of digital phenotyping. *npj Digital Medicine*.
- ²⁸ Chang, A. (2018). The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram. *Vox*, *March*, 23.
- ²⁹ Frenkel, S., Confessore, N., Kang, C., Rosenberg, M & Nicas, J. (2018). Delay, Deny and Deflect: How Facebook's Leaders Fought Through Crisis. *The New York Times*, *November, 14, 2018*.

³⁰ Martinez-Martin, N., & Kreitmair, K. (2018). Ethical Issues for Direct-to-Consumer Digital Psychotherapy Apps: Addressing Accountability, Data Protection, and Consent. *JMIR mental health*, 5(2).

³¹ Aitken, M., Clancy, B., & Nass, D. (2017, November 7). The Growing Value of Digital Health. Retrieved September 16, 2018, from <https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health>

³² American Medical Association (2016, September 26). Digital Health study physicians' motivations and requirements for adopting digital clinical tools. [Survey]. Retrieved November 17, 2018: <https://www.ama-assn.org/sites/default/files/media-browser/specialty%20group/washington/ama-digital-health-report923.pdf>

³³ *Ibid.*

³⁴ Prior, C. E. (1868). A few practical notes on the use of the thermometer in disease. *The British Medical Journal*, 1(384), 1-451.