The digital phenotype

Sachin H Jain, Brian W Powers, Jared B Hawkins & John S Brownstein

In the coming years, patient phenotypes captured to enhance health and wellness will extend to human interactions with digital technology.

In 1982, the evolutionary biologist Richard Dawkins introduced the concept of the "extended phenotype"[1], the idea that phenotypes should not be limited just to biological processes, such as protein biosynthesis or tissue growth, but extended to include all effects that a gene has on its environment inside or outside of the body of the individual organism. Dawkins stressed that many delineations of phenotypes are arbitrary. Animals and humans can modify their environments, and these modifications and associated behaviors are expressions of one's genome and, thus, part of their extended phenotype. In the animal kingdom, he cites damn building by beavers as an example of the beaver's extended phenotype[1].

As personal technology becomes increasingly embedded in human lives, we think there is an important extension of Dawkins's theory—the notion of a 'digital phenotype'. Can aspects of our interface with technology be somehow diagnostic and/or prognostic for certain conditions? Can one's clinical data be linked and analyzed together with online activity and behavior data to create a unified, nuanced view of human disease? Here, we describe the concept of the digital phenotype. Although several disparate studies have touched on this notion, the framework for how digital technologies will be integrated into the patient journey and play a role in precision medicine has yet to be described. We attempt to define digital phenotype and further describe the opportunities and challenges in incorporating these data into healthcare.

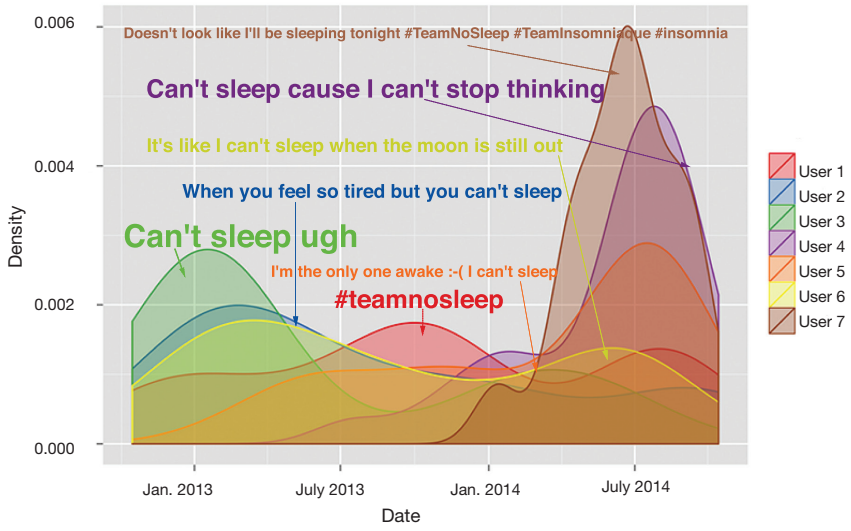

Figure 1 Timeline of insomnia-related tweets from representative individuals. Density distributions (probability density functions) are shown for seven individual users over a two-year period. Density on the *y* axis highlights periods of relative activity for each user. A representative tweet from each user is shown as an example.

Defining the digital phenotype

The growth and evolution of digital products and their application to health supports this interpretation of the extended phenotype. Through social media, forums and online communities, wearable technologies and mobile devices, there is a growing body of health-related data that can shape our assessment of human illness. Such data have substantial value above and beyond the physical exam, laboratory values and clinical imaging data—our traditional approaches to characterizing a disease phenotype. When gathered and analyzed appropriately, these data have the potential to fundamentally alter our notion of the manifestations of disease by providing a more comprehensive and nuanced view of the experience of illness. Through the lens of the digital phenotype, an individual's interaction with digital technologies affects the full spectrum of human disease from diagnosis, to treatment, to chronic disease management. Early examples of digital tracking include the use of cell phone activity to measure's one's activity levels and the association with depression by the Boston-based startup company Ginger.io. There are, of course, limitations to what can be measured and by whom when considered in the context of personal privacy.

Exploiting the digital phenotype

As a corollary to traditional forms of disease expression, digital phenotypes can expand our ability to identify and diagnose health conditions. Some of the earliest and most

Sachin H. Jain is at Caremore Health System, Cerritos, California, USA; Brian W. Powers, Jared B. Hawkins and John S. Brownstein are at Harvard Medical School, Boston, Massachusetts, USA; Sachin H. Jain is at Boston-Virginia Medical Center, Boston, Massachusetts, USA, and Merck and Co., Inc., Boston, Massachusetts, USA; and John S. Brownstein and Jared B. Hawkins is at Children's Hospital Boston, Boston, Massachusetts, USA.
e-mail: shjain@post.harvard.edu

well-developed uses of online health data have involved disease surveillance, such as the use of internet search data to identify outbreaks of communicable diseases like flu and salmonella[2]. Similar approaches can be applied to glean insights at the individual level. Information from digital products and social media provide new venues to identify and track disease symptoms. For example, researchers have demonstrated that Google search data can be used to identify suicidal ideation[3]. Thus, appreciating an individual's digital phenotype could aid in early disease detection, identifying symptoms before traditional phenotypic expression—and potentially building tools for early intervention.

The role of digital phenotypes in diagnosis extends beyond surveillance and early detection. Digital phenotypes redefine disease expression in terms of the lived experience of individuals, which expands our ability to classify and understand disease. For a patient with insomnia, data regarding the timing and hours of one's digital footprint can be considered part of the disease's expression. To illustrate this point, we queried our own database of geo-tagged tweets (i.e., tweets for which we have precise latitude and longitude information) for various insomnia-related terms. In **Figures 1** and **2**, we depict the insomnia-related tweets of several Twitter users, demonstrating how these data could be used to track symptoms over time and assess insomnia from a population health perspective.

Similarly, for a bipolar patient whose mania is manifested in rapid, uninterruptible speech or hypergraphia, their disease could be characterized by the frequency, length and content of participation in social media. Through these varied applications, digital phenotypes can help ensure that early manifestations of disease do not go unnoticed and allow the healthcare system to develop more nimble, targeted and prompt interventions.

As continuously measured manifestations of biologic disease, digital phenotypes can also be a useful adjunct for traditional approaches to disease treatment and management. By redefining the manifestation of illness, they provide new ways to measure disease and therapeutic response in ways that matter most to patients. This approach is especially important for the wide spectrum of diseases characterized by functional limitations. Several online platforms, such as PatientsLikeMe, CureTogether and Inspire, allow individuals to continuously rate their personal assessment of their disease status and track this information over time. Coupled to online communities, these venues also allow experimentation with new treatment strategies and the active sharing of results (http://www.nature.com/nbt/journal/v27/n10/full/nbt1009-888.html).

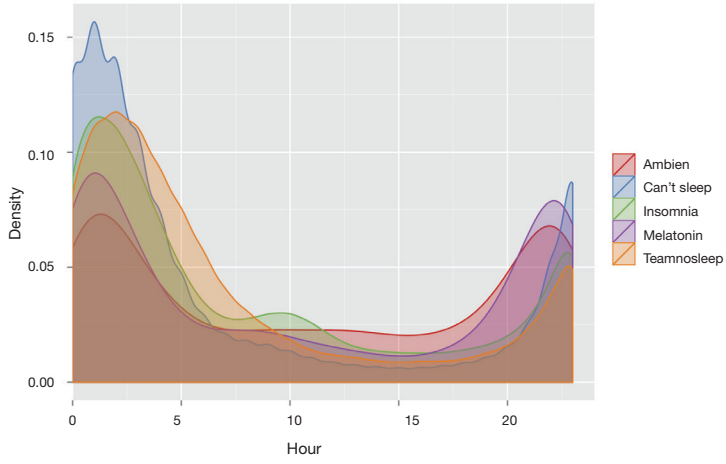

Figure 2 Time distribution of insomnia-related tweets. Density curves (probability density functions) for multiple insomnia terms are shown, illustrating what time of day (using a 24-hour clock) tweets were posted. Density on the *y* axis highlights times of relative activity for each keyword. Tweets were from our own database of geo-tagged tweets, with 1,315,236 tweets shown here across a two-year period. Note that the time of each tweet was converted into the user's local time.

The information on treatment efficacy gleaned from these data is an important addition to traditional evaluations of therapeutics. For example, by tracking digital phenotypes among members of an online disease community, researchers were able to demonstrate the lack of efficacy of lithium in slowing disease progression in individuals with amyotrophic lateral sclerosis. These findings were later replicated in several slower and more expensive randomized controlled trials[4].

In terms of strategies for disease management, a recent systematic review highlighted several successful efforts to use wearable accelerometers to track functional outcomes in patients with neurological disorders and optimize the design of rehabilitation programs[5]. Relatedly, data from personal activity trackers, such as Fitbit and the forthcoming Apple Watch, could be valuable in patients with cardiometabolic diseases where activity levels are an important variable for prognosis and management. Whereas these diseases are presently characterized by the outcome of their activity levels—weight and BMI (body mass index)—they could easily be understood differently in terms of activity levels themselves. This would create greater opportunities for primary prevention and lifestyle modification. Although healthcare has repeated the need for preventive and predictive approaches to managing disease for decades, the digital phenotype and associated interventions could be a new approach to identifying and incentivizing healthy individual behaviors.

Into the digital era

As new products enter the market and individuals interact with digital communities in new ways, our knowledge of digital phenotypes will continue to expand and evolve. Building our understanding of digital phenotypes and their application to improving care will require refined methods for capturing and analyzing various streams of digital health data in a rigorous fashion. A key potential limitation of the success of these approaches will be the physician's and healthcare system's ability to integrate these data into clinical practice in a way that is ethically sound, legally permissible and respectful of patient privacy. Clinicians already suffer from an overload of data—ensuring that data are incorporated in a value-added way will be challenging. New tools to help interpret and simplify data for use in everyday clinical decisions will be vitally important.

Whatever the pace and obstacles, the convergence of digital technologies and biology is an inevitability that is bound to change how we both understand and treat disease—and the concept of the digital phenotype is one that will only continue to grow in its significance.

ACKNOWLEDGMENTS

J.S.B and J.B.H are supported by the National Human Genome Research Institute BD2K grant 1U54HG007963-01.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Dawkins, R. *The Extended Phenotype: the Gene as the Unit of Selection*. (Oxford: Oxford University Press; 1982).
2. Brownstein, J.S., Freifeld, C.C. & Madoff, L.C. *N. Engl. J. Med.* **360**, 2153–2155 (2009).
3. Gunn, J.F. III & Lester, D. *J. Affect. Disord.* **148**, 411–412 (2013).
4. Wicks, P., Vaughan, T.E., Massagli, M.P. & Heywood, J. *Nat. Biotechnol.* **29**, 411–414 (2011).
5. Steins, D., Dawes, H., Esser, P. & Collett, J. *J. Neuroeng. Rehabil.* **11**, 36 (2014).