Chapter 15

Introduction to Digital Phenotyping for Global Health

Olivia Mae Waring and Maiamuna S. Majumder

Abstract Background: The advent of personal mobile devices and the popularity of the Internet are generating unprecedented amounts of data, which can be harnessed to shed light on—and potentially mitigate—public health concerns. **Objective:** To provide a theoretical overview of digital phenotyping (DP), as well as examples of DP in practice. **Results:** Digital phenotyping has been successfully employed to diagnose PTSD in trauma-stricken populations and to localize the source of infectious disease outbreaks. These are only two of the many potential applications of this technology. **Conclusion:** Digital phenotyping is a relatively low-cost, yet powerful tool for both assessing the health status of individuals as well as analyzing broader public health trends.

Keywords Digital phenotyping · Active and passive data streams · Machine learning · Deep phenotyping · Precision medicine · mHealth · Electronic medical records

Learning Objectives

- (1) Understand how digital phenotyping is similar to and distinct from mobile health tools and electronic medical records.
- (2) Explore potential applications of digital phenotyping towards public health, including mental health diagnostics and infectious disease outbreak tracking.
- (3) Discuss the limitations and ethical ramifications of this emergent technology.

O. M. Waring (✉)

Health Sciences and Technology Division of Harvard University and the Massachusetts Institute of Technology, Cambridge, MA, USA

e-mail: omwaring@mit.edu

M. S. Majumder

Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

© The Author(s) 2020

L. A. Celi et al. (eds.), *Leveraging Data Science for Global Health*,
https://doi.org/10.1007/978-3-030-47994-7_15

15.1 Introduction

Nestled in our pockets and purses at any given moment is a tiny yet startlingly powerful device capable of tracking the most intimate details of our everyday lives: our locations, our physical movements, our social habits, our very heartbeats. The proliferation of smart phones in both the developed and developing worlds has brought with it a commensurate rise in the amount of highly personalized data available for mining. It is by now well-known that corporate behemoths such as Google and Facebook collect and leverage this data to generate advertisements targeted at niche demographics (Maréchal 2018). In recent years, medical practitioners and computer scientists alike have been exploring the possibility that information harvested from digital devices could be harnessed to improve healthcare throughout the world, with applications as varied as predicting infectious disease outbreaks to diagnosing mental illness.

15.2 What Is Digital Phenotyping?

For many readers, the term “phenotype” likely conjures memories of Mendel’s peas and Punnett Squares, a mainstay of high school biology textbooks. A phenotype simply refers to the collection of empirically observable traits that define an individual (such as the color, shape, and size of each plant and its component parts in the case of Mendel’s peas). In a medical context, “phenotype” acquires a more specialized meaning, namely “some deviation from normal morphology, physiology, or behavior” (Robinson 2012). In general, **digital phenotyping** refers to the combined art and science of leveraging information from digital devices and data streams to characterize individual human beings. Put more poetically, it allows us to cultivate a database of “digital fingerprints” that “reflect the lived experiences of people in their natural environments, with granular temporal resolution”. Traits that comprise the **human phenotype** include “behavioral patterns; sleep patterns; social interactions; physical mobility; gross motor activity; cognitive functioning; speech production” and more (Onnela 2016).

For a formal definition of **digital phenotyping** (DP), we take our cue from one of the leading innovators in the field, Jukka-Pekka Onnela of the T.H. Chan School of Public Health Harvard University. He characterizes the process of digital phenotyping as “the moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices” (Onnela 2016). **Deep phenotyping** refers to phenotypic characterization with extremely high accuracy, thus enabling the pursuit of **precision medicine** (i.e. medicine that is tailored to the patient’s particular clinical profile) (Robinson 2012). Akin to genomic medicine, which leverages minute knowledge of an individual’s genetic sequence to devise highly personalized treatment regimens and targeted therapies (Roberts 2017), deep phenotyping could help usher in an era of bespoke clinical interventions.

Digital phenotyping is merely the latest in a long history of intersections between medicine and the digital world. At this point in our discussion, it behooves us to draw a distinction between several related, yet distinct concepts that lie at the intersection of healthcare and digital devices:

- **Mobile health** (also mHealth): the delivery of healthcare services via mobile communication devices (Fedele 2017). An example of mHealth would be an app that sends appointment reminders to patients and clinicians.
- **Electronic medical/health records** (EMR/EHR): any digital platform for collecting and collating patient and population-level health data, in order to streamline care across healthcare settings (Carter 2015). An example of EMR would be a computer system for cataloguing patients, their symptoms, and their treatment plans at a particular hospital.
- **Digital phenotyping**: the collection and analysis of moment-by-moment individual-level human phenotype data “in the wild” using data culled from personal digital devices (Insel 2017). An example of digital phenotyping in action would be an app that monitors the user’s heart rate during a run, thereby contributing to a longitudinal profile of cardiac health.

Mobile health tools and EHRs/EMRs can be used to inform digital phenotyping, and therefore DP is in some ways an extension of both these domains.

15.3 Tools of the Trade

Digital phenotyping relies on the ubiquity of digital data collection devices. Smartphone penetration is on the rise globally; as of 2014, the PEW Research Center reported that 64% of Americans owned a smartphone (American Trends Panel Survey, 2014). Recent years have witnessed a particularly precipitous increase in smartphone usage throughout the developing world, with some estimates placing global smartphone penetration as high as 70% by 2020 (Prieto 2016). Thus, the potential for harvesting data from internet-connected devices is difficult to overstate.

There are two primary categories of data that can be captured via digital phenotyping. The first are **active data streams**, which require the concerted input of the subjects being studied (Torous 2016). These include social media activity, video or audio recordings, and responses to surveys. **Passive data streams**, on the other hand, do not require the user’s active participation—or in some cases, even their permission (the ethics of which will be discussed later in this chapter) (Bourla 2018). Sources of passive data include GPS coordinates, WiFi and Bluetooth connectivity, accelerometer-derived data, screen state, and others. When it comes to medical data collection, digital phenotyping practitioners tend to prefer passive data streams; in Onnela’s words, “our overall philosophy is to do as much as possible using passively collected data, because this is the only way to run long-term studies without significant problems with subject adherence.” An experimentalist runs a protocol many times over to correct for human error and false positives, and the same philosophy

applies to digital phenotyping: the more data that can be collected on a given patient in a longitudinal, non-invasive way, the more robust the analytical algorithms operating on that data will be (Onnela 2019).

Let us apply this rationale to the domain of mental health diagnostics, which we will revisit later in this chapter. Traditional depression screenings involve a paper questionnaire that is administered during infrequent office visits. The latency period between appointments and cumbersome nature of the questionnaire make it difficult to collect sufficiently large quantities of information on a given patient. These questionnaires also rely on potentially faulty recollections of symptoms, and patients might be tempted to answer inaccurately so as not to disappoint the healthcare provider. As a more granular alternative, the Onnela group has pioneered the notion of micro-surveys, which are administered by one's smartphone three times per day and prompt the user to answer just a few questions about mood, sleep quality, suicidal ideations, etc. By reducing the volume and simultaneously increasing the frequency of data collection, adherence and accuracy is improved. Once again invoking our comparison to experimental science, we can refer to the paper questionnaire as an "in vitro" data collection modality, whereas the smartphone app implements "in vivo" data collection; one approach elicits symptoms in a simulated "laboratory" setting, the other in "real life" (Onnela 2016).

Oftentimes, the reams of data collected by digital devices are difficult for humans to parse and interpret, much less parlay into effective treatments. In order to facilitate the analysis of digital phenotyping data, the Onnela Group has developed the Biewe analytics pipeline (Onnela 2019). The Amazon Web Services-based platform (accessible via a web browser and also available as an Android and iOS app) provides a suite of tools for managing and making sense of phenotypic information.¹ Furthermore, machine learning (ML)—which is emerging as a powerful diagnostic tool in its own right and is already being deployed across a wide range of medical specialties to improve and personalize care—can also be combined with digital phenotyping to detect disease and devise new therapies.

15.4 Case Studies

In the sections that follow, we will discuss two promising application of digital phenotyping. The first employs machine learning to characterize the speech patterns of individuals afflicted with PTSD. The second analyzes social media trends to localize infectious outbreaks. While these two applications of DP are quite distinct—the former provides insights into the mental states of individual patients, while the latter aggregates numerous data points to arrive at broader public health conclusions—they both demonstrate the striking potential of this nascent technology.

¹Learn more at <https://www.hsph.harvard.edu/onnella-lab/beiwe-research-platform/>.

15.4.1 Case Study 1: Mental Health Diagnostics

One application of digital phenotyping that has shown a great deal of promise in recent years is mental health diagnostics, particularly in resource-limited contexts. The UN estimates that there were over 65 million refugees throughout the world as of 2015, displaced from their homelands by warfare or natural disasters and exposed to torture, genocide, forced isolation, sexual violence, and other harrowing traumas (Ruzek 2016). As many as 86% of refugees worldwide are thought to suffer from Post-Traumatic Stress Disorder (PTSD) (Owen 2015). Traditional clinical screenings for PTSD and other affective disorders such as depression require patients to sit with a professional healthcare provider and answer a series of diagnostic questions; however, this approach is not generalizable to low and middle-income countries for at least two primary reasons. One is the limited availability of professional psychiatric care within refugee camps and other resource-limited contexts (Springer 2018). The second is that, due to cultural taboos and deeply-entrenched stigmas associated with mental health issues, patients might not provide honest answers to their psychiatric evaluators (Doumit 2018). Thus, web and smartphone-based telemental health interventions are being touted as potential solutions to the high rates of PTSD in humanitarian settings.

Researchers and clinicians have noted that patients diagnosed with depression often speak more slowly and with more pauses than non-depressed individuals (Mundt 2012). These so-called “suprasegmental” features of speech (which include speed, pitch, inflection, etc.) are far more difficult to manipulate at will and therefore serve as reliable diagnostic biomarkers. Thus, **suprasegmental analysis** provides an alternative to traditional question and answer-based diagnostic protocols and has been successfully employed to diagnose generalized anxiety disorder, depression, and other affective ailments (Mundt 2007; Nilsonne 1987; Stassen 1991). A well-trained machine learning algorithm can detect depression based on the following suprasegmental features: voice quality, resonance, pitch, loudness, intonation, articulation, speed, respiration, percent pause time, pneumo-phono-articulatory coordination, standard deviation of the voice fundamental frequency distribution, standard deviation of the rate of change of the voice fundamental frequency, and average speed of voice change (Mundt 2007; Nilsonne 1987).

In recent years, researchers at the University of Vermont and the United States Department of Defense have proposed the development of an ML-driven, cell phone-based approach that uses suprasegmental linguistic features to diagnose PTSD (Xu 2012). An AI-driven chatbot could perform suprasegmental analysis on the incoming audio signal and then, using a pretrained classifier, categorize the caller according to the severity of his or her symptoms, as shown in Fig. 15.1. The most urgent cases (i.e. those that scored above a certain empirical threshold) would be escalated to local responders or remote psychotherapists. This low-cost mobile platform would allow more PTSD victims to obtain immediate symptomatic support and potentially life-saving diagnoses, even without access to the internet. This methodology has already been partially validated by several studies. A group at MIT

Fig. 15.1 A trained classifier analyzes the suprasegmental (blue), lexical (purple), and semantic (green) features of an utterance to assess the speaker’s psychiatric condition

has developed an ML-driven artificial intelligence agent that assigns a “probability of depression” index to audio content (Alhanai 2018). A study conducted by the University of Vermont in conjunction with the United States Department of Defense, in which acoustic biomarkers were used to screen for PTSD, determined that the standard deviation of fundamental frequency was found to be significantly smaller in PTSD sufferers than in non-sufferers. Analyses of this and other features were combined to achieve a detection accuracy of 94.4% using only 4 s of audio data from a given speaker (Xu 2012).

15.4.2 Case Study 2: Mapping the Spread of Infectious Disease Using Social Media and Search Engines

While Case Study 1 focuses on diagnosing individuals with a particular affliction, Case Study 2 invokes existing knowledge of disease prevalence to predict future epidemiological outcomes. Imagine that several hundred Twitter users independently complain of nausea, vomiting, diarrhea, and fever, while also referencing food they recently consumed at a particular fast-food joint (Harris 2018). A natural language processing engine that detects references to symptoms and maps them onto specific diseases, in conjunction with a geolocation module that aggregates tweets based on proximity, could conceivably be used to spot an *E. coli* outbreak originating from a single restaurant. This gives rise to a powerful platform for identifying infectious disease outbreaks—foodborne and otherwise—and pinpointing their sources far more rapidly and accurately than through formal reporting channels. Researchers have noted that engaging with disease outbreaks via tweets (as in the case of a

St. Louis food-poisoning incident) results in a larger number of reports filed with the relevant public health organizations (Harris 2017). The organization HealthMap streamlines and formalizes this sort of analysis, harnessing social media trends and aggregating news sources to predict and characterize the spread of disease throughout the world (Freifeld 2008) (Fig. 15.2).

Search engine trends can also provide insight into infectious disease outbreaks. The search engine query tool Google Search Trends can be used to determine interest in a particular search term, and it allows for an impressive degree of geographical and temporal granularity. Search interest can then be used to approximate the incidence of a given disease or condition or to analyze population-level sentiment associated with said disease or condition. In one recent study, Google Trends data was able to recapitulate the traditional surveillance data of the 2015–2016 Colombian Zika outbreak with remarkable accuracy, suggesting that non-traditional digital data sources could be used to track epidemiological transmission dynamics in near real-time (Majumder 2016). Another example can be found in the work of Mauricio Santillana, whose team was able to retroactively reconstruct influenza transmission dynamics in Argentina, Bolivia, Brazil, Chile, Mexico, Paraguay, Peru, and Uruguay over a four-year period by analyzing internet search terms. Their model substantially outperformed other methods of recapitulating historical flu observations, paving the way for search term-based predictive outbreak modeling (Clemente 2019).

15.5 Limitations and Ethical Considerations

There are of course ethical qualms associated with digital phenotyping, as with any methodological innovation. The nature of digital phenotyping—particularly when it comes to passive data streams—is such that technology users are constantly generating data for mining, oftentimes without their knowledge or explicit consent (Martinez-Martin 2018). Researchers anticipate that it won't be long before digital phenotyping costs as little as \$1 per subject per year—less than \$100 over the course of a lifetime. This data, while costing a pittance to acquire, is of enormous monetary value to advertisers, political operatives, and even entities with more sinister agendas (Kozłowska 2018). Indeed, a recent New York Times exposé described a mobile app targeted at children under thirteen called “Fun Kid Racing,” which allows users to race virtual cars against cartoon animals. Unbeknownst to most parents, the app was passively tracking the location of its underage users and selling this data to private companies. Concerned privacy advocates wonder how we can ensure that this highly sensitive data doesn't end up in the hands of predators (Valentino-deVries 2018).

It can also be argued that, simply by virtue of its dependence on smartphones and other connected devices, digital phenotyping perpetuates global disparities. Internet penetration is still relatively low in many parts of the developing world and almost nonexistent in vulnerable settings such as refugee camps or areas stricken by natural

Fig. 15.2 Anonymized tweet used to identify and localize the source of a foodborne illness

disasters (Ruggiero 2012). Thus, those who stand to benefit most from the healthcare applications of digital phenotyping may be those most likely to be excluded from the trend.

15.6 Conclusion

The scientists of yesteryear probed the boundaries of human knowledge with telescopes and microscopes, looking to the distant reaches of space and the intricate interplay of molecules for insights into the world and its workings. As we venture further into the twenty-first century, however, researchers are increasingly directing their attention closer to home, focusing on the everyday human experience. A more nuanced understanding of our own behavioral patterns, as gleaned through digital phenotyping, could usher humanity into a new era of highly personalized healthcare while illuminating the best and worst inclinations of our species. The data-generation capability is already in place; deciding how to utilize this data is an enterprise limited only by the collective imagination of digital phenotyping practitioners.

References

- Alhanai, T. G. (2018). Detecting Depression with Audio/Text Sequence Modeling of Interviews. (MIT, Interviewer).
- American Trends Panel Survey. (2014). Pew Research Center. (3–27 October).
- Bourla, A. F. (2018). Assessment of mood disorders by passive data gathering: The concept of digital phenotype versus psychiatrist's professional culture. *L'Encephale*, 168–175.
- Carter, J. (2015). Electronic medical records and quality improvement. *Neurosurgery Clinics of North America*, 245–251.
- Clemente, L. L. (2019). Improved real-time influenza surveillance: Using internet search data in eight Latin American Countries. *JMIR Public Health Surveillance*.
- Doumit, M. F. (2018). Focus groups investigating mental health attitudes and beliefs of parents and teachers in South Lebanon: Are they culturally determined? *Journal of Transcultural Nursing*, 240–248.
- Fedele, D. C. (2017). Mobile health interventions for improving health outcomes in youth: A meta-analysis. *JAMA Pediatrics*, 461–469.
- Freifeld, C. M. (2008). HealthMap: global infectious disease monitoring through automated classification and visualization of internet media reports. *Journal of the American Medical Informatics Association*, 150–157.
- Harris, J. H. (2017). Using twitter to identify and respond to food poisoning: The food safety STL project. *Journal of Public Health Management Practice*, 577–580.
- Harris, J. H. (2018). Evaluating the Implementation of a twitter-based foodborne illness reporting tool in the city of St. Louis Department of Health. *International Journal of Environmental Research and Public Health*.
- Insel, T. (2017). Digital phenotyping: technology for a new science of behavior. *Journal of the American Medical Association*, 1215–1216.

- Kozłowska, I. (2018). *Facebook and data privacy in the age of Cambridge analytica*. Retrieved from The Henry M. Jackson School of International Studies. <https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/>.
- Majumder, M. S. (2016). Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015-2016 Colombian Zika Virus disease outbreak. *JMIR Public Health and Surveillance*.
- Maréchal, N. (2018). *Targeted advertising is ruining the internet and breaking the world*. Retrieved from Motherboard. https://motherboard.vice.com/en_us/article/xwjden/targeted-advertising-is-ruining-the-internet-and-breaking-the-world.
- Martínez-Martin, N. I. (2018). Data mining for health: staking out the ethical territory of digital phenotyping. *npj Digital Medicine*.
- Mundt, J. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *J Neurolinguistics*, 50–64.
- Mundt, J. V. (2012). Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological Psychiatry*, 580–587.
- Nilsson, A. (1987). Acoustic analysis of speech variables during depression and after improvement. *Acta Psychiatr Scand*, 235–245.
- Onnela, J. R. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*, 1691–1696.
- Onnela, J. (2019). *Bieve Research Platform*. Retrieved from Onnela Lab. <https://www.hsph.harvard.edu/onnella-lab/beiwe-research-platform/>.
- Owen, J. J.-B. (2015). mHealth in the wild: Using novel data to examine the reach, use, and impact of PTSD coach. *JMIR Ment Health*.
- Prieto, R. N. (2016). *10th annual cisco visual networking index (VNI) mobile forecast projects 70 percent of global population will be mobile users*. Retrieved from The Network. <https://newsroom.cisco.com/press-release-content?articleId=1741352>.
- Roberts, M. K. (2017). The current state of implementation science in genomic medicine: Opportunities for improvement. *Genetics in Medicine*, 858–863.
- Robinson, P. (2012). Deep phenotyping for precision medicine. *Human mutation*, 777–780.
- Ruggiero, K. R. (2012). Randomized controlled trial of an internet-based intervention using random-digit-dial recruitment: The disaster recovery web project. *Contemporary Clinical Trials*, 237–246.
- Ruzek, J. K. (2016). Mobile mental health interventions following war and disaster. *Mhealth*.
- Springer, P. S. (2018). Global proofing a collaborative care telemental health intervention in Brazil. *Families, Systems, and Health*.
- Stassen, H. B. (1991). Speech characteristics in depression. *PSP*, 88–105.
- Torous, J. S. (2016). Bipolar disorder in the digital age: New tools for the same illness. *International Journal of Bipolar Disorder*, 25.
- Valentino-deVries, J. S. (2018). How game apps that captivate kids have been collecting their data. *New York Times*.
- Xu, R. M. (2012). A voice-based automated system for PTSD screening and monitoring. *Studies in Health Technology and Informatics*, 552–558.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.