

## Digital Health Data-Driven Approaches to Understand Human Behavior

**Cite this article as:** Lisa A. Marsch, Digital Health Data-Driven Approaches to Understand Human Behavior, *Neuropsychopharmacology* doi:10.1038/s41386-020-0761-5

This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

□ This article is licensed under a Creative Commons Attribution 4.0 International License, 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 article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article'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. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.nature.com/authors/policies/license.html#AAMtermsV1>

**Digital Health Data-Driven Approaches to Understand Human Behavior**

Lisa A. Marsch, PhD

Director, Center for Technology and Behavioral Health

Andrew G. Wallace Professor,

Geisel School of Medicine at Dartmouth College, USA

[lisa.a.marsch@dartmouth.edu](mailto:lisa.a.marsch@dartmouth.edu)

Author accepted manuscript

## Abstract

Advances in digital technologies and data analytics have created unparalleled opportunities to assess and modify health behavior and thus accelerate the ability of science to understand and contribute to improved health behavior and health outcomes. Digital health data capture the richness and granularity of individuals' behavior, the confluence of factors that impact behavior in the moment, and the within-individual evolution of behavior over time. These data may contribute to discovery science by revealing digital markers of health/risk behavior as well as translational science by informing personalized and timely models of intervention delivery. And they may help inform diagnostic classification of clinically problematic behavior and the clinical trajectories of diagnosable disorders over time. This manuscript provides a review of the state-of-the-science of digital health data-driven approaches to understanding human behavior. It reviews methods of digital health assessment and sources of digital health data. It provides a synthesis of the scientific literature evaluating how digitally-derived empirical data can inform our understanding of health behavior, with a particular focus on understanding the assessment, diagnosis and clinical trajectories of psychiatric disorders. And, it concludes with a discussion of future directions and timely opportunities in this line of research and its clinical application.

## Introduction

### **Overview and limitations of theoretical models of human behavior and diagnostic models of psychiatric disorders.**

Human behavior is one of the biggest drivers of health and wellness as well as mortality and morbidity. Indeed, health risk behavior, including poor diet, physical inactivity, tobacco, alcohol and other substance use, causes as much as 40% of the illness, suffering, and early death related to chronic diseases [1-3]. Health risk behavior is linked to obesity, Type 2 diabetes [4], heart disease, liver disease, kidney failure and neurological diseases. It is also linked to many mental health disorders including anxiety and depression [5,6]. And it greatly increases one's risk for a wide variety of cancers. For example, heavy alcohol use greatly increases risk of breast [7-9], esophageal and upper [10] digestive and liver cancers [11,12]. Smoking is strongly linked to lung cancer and is also a major contributor to esophageal cancer [13-16]. And, obesity increases risk of colorectal and esophageal cancer [17-19].

Research designed to explain and predict health behavior and events influencing health outcomes has heavily relied on theoretical models of health behavior and behavior change [20,21]. At the psychological level, the cognitive literature has focused on such performance-related processes as goal maintenance in working memory, impulsivity, and cognitive homeostasis. The affective science and social psychology literatures have focused on emotion regulation processes, social influences and resource models. In parallel, the health psychology and behavioral medicine literatures have focused on processes such as self-efficacy and outcome expectancies. At the behavioral level, focus has been largely placed on behavioral disinhibition and temporal discounting. At the neural level [22], health behavior can be conceptualized in terms of top-down control (implemented by fronto-parietal networks) over

impulsive drives or habits (implemented by subcortical and ventromedial prefrontal regions [23]). And an emerging framework from neuroeconomics has characterized decision processes in terms of goal-directed versus habitual or Pavlovian control over action [24-26].

Overall, these models afford a conceptual framework for illustrating causal processes of key constructs hypothesized to influence or change a target behavior. Theoretical models may be useful for developing, implementing, and evaluating behavior change interventions. And, interventions informed by theories of human behavior are generally more effective compared with those that are not [27]. Collectively, various theoretical models have articulated that an individual's beliefs and attitudes, behavior intentions, level of motivation for behavior change and social and cognitive processes impact health behavior [28].

Despite the promise of theoretical models of health behavior, their ability to explain and predict health behavior has been only modestly successful [28-30]. Many theoretical models have regarded human behavior as linear or static in nature and have not recognized that behavior is dynamic and responsive to diverse social, biological, and environmental contexts. And, theoretical models have heavily focused on between-person differences in behavior and have not embraced the study of important within-person differences in behavior. Further, many theoretical models of health behavior and behavior change have often been derived within siloed disciplines (e.g., health psychology, neuroscience) with little crosstalk [31,32].

Additionally, research examining factors that influence health behavior has tended to examine a small set of potential moderators or mediators of health behavior at a specific level of analysis (e.g., emotion regulation alone or impulsivity alone) and may lead to over-simplified accounts of behavior [33-36] change. Finally, little research has established the temporal precedence of a broad array of potential factors impacting health behavior [37]. More frequent and longer

assessment of moderators, mediators, and outcome(s) will be necessary to elucidate the temporal dynamics between changes in specific mechanisms [38] and behavior [39,40].

Similar limitations are evident in our current models for understanding and determining clinical diagnoses for psychological or psychiatric disorders. The current process for identifying diagnosable disorders heavily relies on measuring the number and type of symptoms a person may be experiencing as well as associated distress or impairment. Although this current diagnostic process provides a useful common language of mental disorders for clinicians, the process is largely based on consensus from expert panels and may oversimplify our understanding of human behavior [41]. And indeed, many mental health clinicians do not measure behavior, cognition and emotion when ascribing a psychiatric disorder to a patient. Further, mental health professionals usually interact with, and provide diagnoses to, patients at a specific moment in patients' lives, but recent evidence shows that people with psychological disorders may experience many different kinds of disorders across diagnostics families over their lifespan [42]. There is tremendous opportunity to understand psychological/biological systems that span the full range of human behavior from normal to abnormal and to empirically assess how they are situated in environmental and neurodevelopmental contexts [43]. Examining a broad array of factors impacting health behavior at multiple levels of empirical analysis and over time will enable a more comprehensive picture of health behavior and will increase our ability to develop more impactful interventions and better understand the conditions under which replications of effects do and do not occur.

### **The Promise of Digital Health in Understanding Human Behavior**

Advances in digital technologies and data analytics have created unparalleled opportunities to

assess and modify health behavior and thus accelerate the ability of science to understand and contribute to improved health behavior and health outcomes. Digital health refers to the use of data captured via digital technology to measure individuals' health behavior in daily life and to provide digital therapeutic tools accessible anytime and anywhere [44,45]. For example, smartphones have an array of native sensors including Bluetooth, GPS, light sensor, accelerometer, microphone, and proximity sensors as well as systems logs of calls, and short message service use. Smartphones, as well as some wearable devices (e.g., smartwatches), thus enable passive, ecological sensing of behavioral and physiological features, such as one's sleep, physical activity, social interactions, electrodermal activity and cardiac activity [46]. Individuals can also offer responses to questions they are prompted to answer on mobile devices (sometimes called "ecological momentary assessment" or EMA) to provide snapshots into, for example, their context, social interactions, stress, pain, mood, eating, physical activity, mental health symptoms and substance use. And, social media data, that many individuals produce in high volume, provide information about individuals' behavior, preferences and social networks. These "digital exhaust" [47] data or "digital footprints" [48] enable the continuous measurement of individuals' behavior and physiology in naturalistic settings.

These digital data may greatly complement and extend traditional sources of clinical data (which is typically captured on an episodic basis in a clinical context) with intensive, longitudinal ecologically-valid data. Digital health data capture the richness and granularity of individuals' behavior, the confluence of factors that impact behavior in the moment, and the within-individual evolution of behavior over time. As such, they may contribute to discovery science by revealing digital markers of health and risk behavior [49,50]. They may help us better develop empirically-based diagnostic classifications of aberrant/dysfunctional behavior and the clinical trajectories of diagnosable disorders over time [50]. And, they may help us in translational science by informing more personalized, biomarker-informed, and timely models of intervention delivery.

As the majority of the world has access to digital technology -- indeed, there are 8 billion mobile phone subscriptions worldwide [51]– digital health data-driven approaches can be used to understand human behavior across the population.

## **The State of the Science of Digital Health Data-Driven Approaches to Understand Human Behavior**

This manuscript provides a review of the state-of-the-science of digital health data-driven approaches to understanding human behavior. The manuscript first describes various methods of digital health assessment and sources of digital health data. It then provides a synthesis of the scientific literature evaluating how digitally-derived empirical data can inform our understanding of health and risk behavior. It then focuses on how digital health may help us develop a better empirically-based understanding in the assessment, diagnosis and measurement of clinical trajectories of aberrant/dysfunctional disorders in the field of psychiatry (a field that has led pioneering research in digital health [52]). Finally, it concludes with a discussion of future directions and timely opportunities in this line of research and its clinical application, including the development of personalized digital interventions (e.g., behavior change interventions) informed by digital health assessment.

### *Digital Health Assessment Methods*

Although digitally-derived data have been used to understand behavior and context in the field of Computer Science for over 15 years, a primary term currently used to capture digital health assessment is “digital phenotyping” [53] and is increasingly used by scientists, funders, as well



as the popular press. Digital phenotyping [54] primarily employs passively sensed data to allow for a moment-by-moment (*in situ*) quantification of behavior. These data can include data derived from smartphone or smartwatch sensors (e.g., an individual's activity, location), features of voice and speech data collected by mobile devices (e.g., prosody and sentiment), as well as data that captures a person's interaction with their mobile device (e.g., patterns of typing or scrolling). Digital phenotyping largely employs passive data (to reduce burden to participants in data collection), and some researchers confine their definition of digital phenotyping to passive data. However, digital measurement and analytics also encompass many other sources of data that are actively generated by individuals, including social media data, EMA data, and online search engine activity.

Overall, digital phenotyping focuses on the use of such digital data to understand and predict health outcomes of behaviors of interest. Sophisticated inferences from these data are increasingly possible due to the rapidly-advancing fields of big-data analytics and advanced Artificial Intelligence (including advanced machine learning approaches that focus on the creation of systems that learn from data instead of simply following programmed instructions).

Behavioral health systems that leverage passive sensing and machine learning to learn and adapt to a person's actual behavior and surroundings offer a promising foundation for predictive modeling of an individual's behavioral health trajectory and may support new breakthrough intervention technologies targeting health behavior. These developments enable behavioral monitoring to occur in the background as individuals go about their lives and build dynamic computational models tailored to the user that can lead to effective interventions.

And digital phenotyping may reveal new insights into how other data sources (such as genetic, molecular and neural circuitry data) interrelate with clinically observable psychopathology [55,56].

*Overview of the Scientific Literature on the Application of Digitally-Derived Empirical Data to Understand Health Behavior and Psychopathology.*

A robust and rapidly growing scientific literature is increasingly demonstrating the potential utility of digital assessment in revealing new insights into human behavior, including psychological and psychiatric disorders.

*Digital Health Biomarkers of Health and Risk Behavior Captured Via Mobile Technology*

Continuous smartphone sensing (e.g., of activity, mobility, sleep) has been shown to be significantly linked to mental well-being, academic performance (Grade Point Average), and behavioral trends of a college student body (such as increased stress, reduced sleep and reduced affect as the college term progresses and stress increases [57]. These patterns may help us understand, in close to real time, when individuals may be at risk of academic and/or mental health decline. Assessment of individual's interactions with mobile devices (e.g., swipes, taps and keystroke events) have been shown to capture neurocognitive function in the real world and may provide an ecological surrogate for laboratory-based neuropsychological assessment [58]. And, continuous smartphone monitoring can measure brain health and cognitive impairment in daily life [59]. And, digital data derived from mobile sensing (e.g., calling, texting, conversation and app use) have also been used to characterize behavioral sociability patterns and to map these behaviors onto personality traits [60]. Further, phenotypic data gathered via wearable sensors have shown that several metrics of sleep (total sleep time and sleep efficiency) are associated with cardiovascular disease risk markers such as waist

circumference and [61] body mass index and that insufficient sleep is linked to premature telomere attrition. Thus, these digitally derived health risk data can provide real time insights into biological aging.

*Digital Health Measurement of Aberrant/Dysfunctional Behavior and the Clinical trajectories of Diagnosable Disorders Over Time Captured via Mobile Technology*

Digital assessment has also illuminated novel insights into the nature and course of psychological and psychiatric disorders. High-frequency assessment of cognition and mood via wearable devices among persons with major depressive disorder has been shown to be feasible and valid over an extended period [62]. Behavioral indicators passively collected through a mobile sensing platform (e.g., the sum of outgoing calls, a count of unique numbers texted, the dynamic variation of voice, speaking rate) have been shown to predict symptoms of depression and PTSD [63]. Features derived from GPS data collected via phone sensors, including location variance, entropy and circadian movement, have been shown to predict severity of depressive symptoms and that these relationships can differ at different points in time (e.g., weekend vs. weekday [64]). And assessment of voice data has identified vocal acoustic biomarkers that have shown promise in predicting treatment response among persons with depression [65].

Movement data from actigraphs alone, a single measure of gross motor activity from a sensor worn on the wrist, were able to identify the diagnostic group status of individuals with major depression or bipolar vs. healthy controls 89% of the time. This level of accuracy in diagnostic classification is greater than published inter-reliability rates for second raters using the Structured Clinical Interview for the DSM (SCID). And results showed that actigraphy data

predicted the majority of variation in patients' depression severity over an approximately two-week period [66].

Emotion dynamics captured over time via digital technology have been shown to differentially predict bipolar and depressive symptoms concurrently and prospectively [67]. And EMA data captured on smartphones has been shown to predict future mood among persons with bipolar disorder [68]. Additionally, smartphone usage patterns have been shown to be linked to functional brain activity related to depression. For example, phone unlock duration has been shown to be positively linked to resting-state functional connectivity between the subgenual cingulate cortex (an area understood to be involved in depression) and the ventromedial/orbitofrontal cortex [69]. Results suggest that digital biomarker data may reflect readily-capturable data that relate to brain functioning.

Further, a small pilot study evaluated changes in mobility patterns and social behavior among persons diagnosed with schizophrenia using passively collected smartphone data. Results indicated that the rate of behavioral anomalies that were identified in the 2 weeks prior to a clinical relapse were markedly higher (71%) than rates of behavioral anomalies during other periods of time [70]. And other research has underscored the significant variability across individuals in digital indicators of a psychotic relapse [71] – thus underscoring the multi-dimensional nature of a diagnosis of a psychotic disorder.

Additionally, a small series of case studies demonstrated that self-reported psychotic symptoms are linked to various behaviors (cognition scores on games) and activity levels (step count) among persons with psychotic illness. Importantly, results revealed considerable variability in the patterns in these data streams across individuals, underscoring the utility of these approaches in understanding and monitoring within-individual clinical trajectories [72]. And other

research has demonstrated that decreased variability in physical activity and noisy conditions on an inpatient psychiatric unit, captured via multimodal measurement, is associated with violent ideation among inpatients with serious mental illness [73].

Assessment of geography via passive sensing of geolocation using GPS has demonstrated that drug craving, stress, and mood among persons with an opioid use disorder were predicted by exposure to visible signs of environmental disorder along a GPS-derived [74,75] track (such as visible signs of poverty, violence and drug activity). A recent digital health EMA study demonstrated a stronger link between drug craving and drug use than between stress and drug use -- a result that was not well-documented or understood from prior traditional clinical assessment [76]. And, among smokers trying to quit, lapses to smoking were shown to be associated with increases in negative mood for many days (and not just hours) before a smoking lapse [77]. These studies reveal new insights into the dynamic nature of drug use events and the confluence of factors that impact them.

Unfortunately, only a few studies have included randomized controlled evaluations of the clinical utility of digital phenotyping in the clinical treatment of psychological disorders. Among these studies, one recent, controlled study that investigated the effect of smartphone monitoring of persons with bipolar disorder did not show a statistically significant benefit on depressive or manic symptoms compared to a control group, although persons with smartphone monitoring reported higher quality of life and lower stress [78,79].

*Digital Health Measurement of Health Behavior Captured via Additional (Non-Mobile) Data*

*Sources*

In addition to data captured via mobile devices, other sources of digital data have been shown to reveal insights into human behavior. For example, social media data have provided new insights into mental health and substance use behavior. In one study, a deep-learning method was able to identify individuals' risk for substance use using content from their Instagram profiles [80]. And another evaluation demonstrated that community-generated Instagram data (post captions and comments from friends or followers), when evaluated along with user-generated content (individuals' post captions and comments), were able to identify depression among individuals. Other work has also demonstrated that Facebook status updates can predict postpartum depression [81] and that depression can be identified via daily variation in word sentiment analysis among Twitter and Facebook users [82,83]. Such methods offer promise for conducting population-level risk assessments and inform population-level interventions [84].

Data from online search engine activities are another source of consumer-generated digital data that can reveal individual-level as well as population-level behavioral patterns. For example, online health-seeking behavior has been shown to predict real-world healthcare utilization [85]. Online search activity has been shown to be related to changes in use of new substances [86], and substance use search data have been strongly correlated with overdose deaths [87]. And a recent study analyzed over 10 million Google search queries across the U.S. related to mental health during the COVID-19 global pandemic. Results revealed that mental health search queries increased rapidly prior to the issuance of stay-at-home order within states, and these searches markedly decreased after the announcement and implementation of these orders (presumably once a response/management plan was in place [88]).

## Future Research Directions/Opportunities

Overall, the existing scientific literature demonstrates a compelling “proof of concept” that digital health data can provide new insights into human behavior, including psychopathology. This line of research offers great promise for advancing our theoretical models of health behavior and informing behavior change interventions that are responsive to the dynamic nature of health behavior.

The promise of digital health is particularly compelling when applied to the field of psychiatry. Digital assessment allows for the continuous, empirical quantification of clinically useful digital biomarkers that can be useful in identifying and refining diagnostic processes over time. These data may also be useful as outcomes in measurement-based care. These data may help us generate predictive models that reflect the confluence of factors, and their relations over time, that may inform when an individual may be at risk for a clinically significant event (such as a relapse or psychotic event). These methods may help detect a problem before it occurs and inform in-the-moment preventative interventions. And, given that psychiatric conditions are often chronic and recurrent, digital data captured in an intensive longitudinal manner can inform strategies for optimizing responsive and adaptive models of clinical care over time.

Thus, digital health offers value along a full spectrum from measurement to intervention delivery -- by providing novel digital biomarkers, new insights into clinical diagnoses of psychiatric disorders, personalized intervention delivery on digital platforms, as well as digital outcome measurement over time. These multiple applications of digital health can complement one another by measuring behavior and informing interventions that are responsive to that measurement.

Despite the promise of digital health data-driven approaches to understanding human behavior, there remain many gaps and opportunities in the field. As noted above, most digital health research has not embraced rigorous experimental research designs. Indeed, only a paucity of trials has embraced well-powered, randomized, controlled research designs to allow for causal inference about the value of digital assessment and associated data analytics in informing clinical outcomes [89]. Additionally, tremendous variability exists in the specific digital metrics being employed in digital health research – ranging from smartphone sensing data, smartwatch sensing data, EMA data, social media data and online search engine data. And within each of these categories, there is also great variability in the types of features that are being extracted and applied to clinical inference. For example, in smartphone sensing alone, some research focuses heavily on GPS, other work focuses on actigraphy, while still other research focuses on movement. The specific features and sources of digital health data (including the potential combination of multiple sources of digital data) that provide maximal precision in characterizing human behavior and behavioral disorders remain understudied as do the psychometric characteristics (e.g., validity and reliability) of such metrics [90]. In order to realize the potential of digital health and provide the most robust and replicable results, a priority focus on experimental rigor and reproducibility is critically needed.

Additionally, digital health research to date has been conducted within our existing classification systems (e.g., patients with bipolar disorder or depression) which, as noted above, can be refined with digital health approaches. And most digital health research has been focused on disease-specific models (e.g., focusing on depression alone or substance use alone). The rich, granular data afforded by digital health approaches offer tremendous opportunity to transcend siloed disease-specific models of behavior and care to empirically embrace, understand, and treat the complexity and interrelatedness of behavioral patterns and clinical disorders. Indeed, scientific research has demonstrated that many disorders co-occur and interrelate in meaningful



ways and that these disorders evolve and change over the lifespan. Digital health offers great (but yet unrealized) promise to provide a data-informed understanding of this full spectrum of health and wellness. This may include the development of an ontology of behavior that is informed by digital health data which may enable a new understanding of co-occurring aberrant/dysfunctional behaviors and their evolution over time. And this may include digital therapeutic interventions that are responsive to the combination of needs and goals of each individual and their evolution over time.

Finally, much of the current research appears to be grounded in assumptions that digital health data will be of interest and of value to consumers, patients and clinicians. Although one could make the case that patients may value self-monitoring and feedback on their behavior and their clinical status and that clinicians may welcome actionable digital health data that can aid them in the care of patients, this may not always be the case. For example, if patients do not experience value in generating and sharing these data, they will not be inclined to do so (or to do so for any extended period of time). If providers receive large volumes of unsolicited data and/or data that do not directly inform their clinical work, they may perceive such a model to be burdensome and unhelpful. And if patients do not understand the privacy and security considerations of how their sensitive data will be handled and/or if healthcare systems do not understand data sharing/protection policies of industry vendors, this will undoubtedly impact adoption[91]. Indeed, it is possible that the current scientific literature largely reflects a subset of the population that are willing to share personal health data collected on digital devices, which may not be broadly generalizable. A broader dialogue is needed to establish fundamental principles of privacy and research ethics in the digital health space. This may include establishing best practices for ensuring protections of patient privacy and sensitive information while still allowing for data to be shared between parties (e.g., patients and clinicians) in accordance with patient and provider preferences. And, this may include informed

consent processes that are adaptive and dynamic in response to each individual's digital literacy and data sharing preferences [92]. Overall, as research and clinical application of digital measurement of behavior expands, there is an urgent need to ensure implementation science approaches are employed to systematically assess the preferences of all relevant digital health stakeholders and to inform models of development and deployment that have the greatest chance of scalability and sustainability. This will undoubtedly require an interdisciplinary effort across the scientific arena (including behavioral science, data science, computer science and neuroscience) as well as the digital health industry and experts in public policy.

### **Conclusion**

Digital health and data analytics are transforming our world. And the real-world precision assessment that digital health methods enable are providing unprecedented insights into human behavior and psychiatric disorders and can inform interventions that are personalizable and adaptive to individuals' changing needs and preferences over time. Now is the moment of opportunity to embrace a systematic, rigorous, and comprehensive research agenda to realize this vision.

## **Funding and Disclosure**

### **Funding**

Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health [Grant Number P30DA029926].

### **Disclosure**

The author is affiliated with Pear Therapeutics, Inc., HealthSim, LLC, and Square2 Systems, Inc. Conflicts of interest are extensively managed by her academic institution, Dartmouth College.

Author accepted manuscript

## References:

- 1 Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA*. 2004;291(10):1238-45.
- 2 Saint Onge JM, Krueger PM. Health lifestyle behaviors among U.S. adults. *SSM - Population Health*. 2017;3:89-98.
- 3 Murray CJ, Atkinson C, Bhalla K, Birbeck G, Burstein R, Chou D, et al. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA*. 2013;310(6):591-608.
- 4 Haque M, McKimm J, Sartelli M, Samad N, Haque SZ, Bakar MA. A narrative review of the effects of sugar-sweetened beverages on human health: A key global health issue. *J Popul Ther Clin Pharmacol*. 2020;27(1):e76-e103.
- 5 Huang Y, Li L, Gan Y, Wang C, Jiang H, Cao S, et al. Sedentary behaviors and risk of depression: a meta-analysis of prospective studies. *Transl Psychiatry*. 2020;10(1):26.
- 6 Allen MS, Walter EE, Swann C. Sedentary behaviour and risk of anxiety: A systematic review and meta-analysis. *J Affect Disord*. 2019;242:5-13.
- 7 Key J, Hodgson S, Omar RZ, Jensen TK, Thompson SG, Boobis AR, et al. Meta-analysis of studies of alcohol and breast cancer with consideration of the methodological issues. *Cancer Causes Control*. 2006;17(6):759-70.
- 8 Seitz HK, Pelucchi C, Bagnardi V, La Vecchia C. Epidemiology and pathophysiology of alcohol and breast cancer: Update 2012. *Alcohol Alcohol*. 2012;47(3):204-12.
- 9 Nelson DE, Jarman DW, Rehm J, Greenfield TK, Rey G, Kerr WC, et al. Alcohol-attributable cancer deaths and years of potential life lost in the United States. *Am J Public Health*. 2013;103(4):641-8.
- 10 Gronbaek M, Becker U, Johansen D, Tonnesen H, Jensen G, Sorensen TI. Population based cohort study of the association between alcohol intake and cancer of the upper digestive tract. *BMJ*. 1998;317(7162):844-7.
- 11 McKillop IH, Schrum LW. Alcohol and liver cancer. *Alcohol*. 2005;35(3):195-203.
- 12 Chuang SC, Lee YC, Wu GJ, Straif K, Hashibe M. Alcohol consumption and liver cancer risk: a meta-analysis. *Cancer Causes Control*. 2015;26(9):1205-31.
- 13 Zhang Y. Epidemiology of esophageal cancer. *World J Gastroenterol*. 2013;19(34):5598-606.
- 14 Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev*. 2010;19(8):1893-907.
- 15 Parsons A, Daley A, Begh R, Aveyard P. Influence of smoking cessation after diagnosis of early stage lung cancer on prognosis: systematic review of observational studies with meta-analysis. *BMJ*. 2010;340:b5569.
- 16 Pesch B, Kendzia B, Gustavsson P, Jockel KH, Johnen G, Pohlabein H, et al. Cigarette smoking and lung cancer--relative risk estimates for the major histological types from a pooled analysis of case-control studies. *Int J Cancer*. 2012;131(5):1210-9.
- 17 Moghaddam AA, Woodward M, Huxley R. Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. *Cancer Epidemiol Biomarkers Prev*. 2007;16(12):2533-47.
- 18 Dai Z, Xu YC, Niu L. Obesity and colorectal cancer risk: a meta-analysis of cohort studies. *World J Gastroenterol*. 2007;13(31):4199-206.
- 19 Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med*. 2003;348(17):1625-38.
- 20 Glanz K, Rimer BK, Lewis FM. Health behavior and health education: Theory, research, and practice. 3rd ed. Jossey-Bass: San Francisco, CA, US; 2002.

- 21 Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From Theory to Intervention:  
Mapping Theoretically Derived Behavioural Determinants to Behaviour Change  
Techniques. *Applied Psychology*. 2008;57(4):660-80.
- 22 Morgenstern J, Naqvi NH, Debellis R, Breiter HC. The contributions of cognitive  
neuroscience and neuroimaging to understanding mechanisms of behavior change in  
addiction. *Psychol Addict Behav*. 2013;27(2):336-50.
- 23 Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal  
choice. *Nat Neurosci*. 2007;10(12):1625-33.
- 24 Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of  
value-based decision making. *Nat Rev Neurosci*. 2008;9(7):545-56.
- 25 Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for  
exploratory decisions in humans. *Nature*. 2006;441(7095):876-9.
- 26 Lee SW, Shimojo S, O'Doherty JP. Neural computations underlying arbitration between  
model-based and model-free learning. *Neuron*. 2014;81(3):687-99.
- 27 Glanz K, Bishop DB. The role of behavioral science theory in development and  
implementation of public health interventions. *Annu Rev Public Health*. 2010;31:399-418.
- 28 Naslund JA, Aschbrenner KA, Kim SJ, McHugo GJ, Unutzer J, Bartels SJ, et al. Health  
behavior models for informing digital technology interventions for individuals with mental  
illness. *Psychiatr Rehabil J*. 2017;40(3):325-35.
- 29 Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior  
models in the age of mobile interventions: are our theories up to the task? *Transl Behav  
Med*. 2011;1(1):53-71.
- 30 Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing  
Models and Theories for Digital Behavior Change Interventions. *Am J Prev Med*.  
2016;51(5):825-32.
- 31 de Ridder D, Wit J. Self-Regulation in Health Behavior: Concepts, Theories, and Central  
Issues. 2008. p. 1-23.
- 32 Vohs KD, Baumeister RF. *Handbook of self-regulation: Research, theory, and  
applications*, 2nd ed. Guilford Press: New York, NY, US; 2011.
- 33 Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten  
strategies for increasing the timeliness and usefulness of eHealth research. *J Med  
Internet Res*. 2014;16(2):e36.
- 34 Moos RH. Theory-based processes that promote the remission of substance use  
disorders. *Clin Psychol Rev*. 2007;27(5):537-51.
- 35 Kazdin AE, Nock MK. Delineating mechanisms of change in child and adolescent  
therapy: methodological issues and research recommendations. *J Child Psychol  
Psychiatry*. 2003;44(8):1116-29.
- 36 Longabaugh R. The search for mechanisms of change in behavioral treatments for  
alcohol use disorders: a commentary. *Alcohol Clin Exp Res*. 2007;31(10 Suppl):21s-32s.
- 37 Baraldi AN, Wurpts IC, Mackinnon DP, Lockhart G. Evaluating mechanisms of behavior  
change to inform and evaluate technology-based interventions. *Behavioral healthcare  
and technology: Using science-based innovations to transform practice*. New York, NY,  
US: Oxford University Press; 2015. p. 187-99.
- 38 Collins LM, Graham JW. The effect of the timing and spacing of observations in  
longitudinal studies of tobacco and other drug use: temporal design considerations. *Drug  
Alcohol Depend*. 2002;68 Suppl 1:S85-96.
- 39 Roche MJ, Jacobson NC, Pincus AL. Using repeated daily assessments to uncover  
oscillating patterns and temporally-dynamic triggers in structures of psychopathology:  
Applications to the DSM-5 alternative model of personality disorders. *J Abnorm Psychol*.  
2016;125(8):1090-102.

- 40 Frank B, Jacobson NC, Hurley L, McKay D. A theoretical and empirical modeling of  
anxiety integrated with RDoC and temporal dynamics. *J Anxiety Disord.* 2017;51:39-46.
- 41 Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA.*  
2017;318(13):1215-16.
- 42 Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, et al. Longitudinal  
Assessment of Mental Health Disorders and Comorbidities Across 4 Decades Among  
Participants in the Dunedin Birth Cohort Study. *JAMA Netw Open.* 2020;3(4):e203221.
- 43 Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for  
psychiatry. *Am J Psychiatry.* 2014;171(4):395-7.
- 44 Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of  
healthcare. *Eur Heart J.* 2016;37(18):1428-38.
- 45 Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and  
recommendations for general practice. *Heredity.* 2020;124(4):525-34.
- 46 Trifan A, Oliveira M, Oliveira JL. Passive Sensing of Health Outcomes Through  
Smartphones: Systematic Review of Current Solutions and Possible Limitations. *JMIR*  
*Mhealth Uhealth.* 2019;7(8):e12649.
- 47 Boyd D, Crawford K. CRITICAL QUESTIONS FOR BIG DATA. *Information,*  
*Communication & Society.* 2012;15(5):662-79.
- 48 Digital Footprints: An Internet Society Reference Framework.  
[https://www.internetsociety.org/wp-content/uploads/2017/08/Digital20Footprints20-  
20An20Internet20Society20Reference20Framework.pdf](https://www.internetsociety.org/wp-content/uploads/2017/08/Digital20Footprints20-20An20Internet20Society20Reference20Framework.pdf).
- 49 Fagherazzi G. Deep Digital Phenotyping and Digital Twins for Precision Health: Time to  
Dig Deeper. *J Med Internet Res.* 2020;22(3):e16770.
- 50 Washington P, Park N, Srivastava P, Voss C, Kline A, Varma M, et al. Data-Driven  
Diagnostics and the Potential of Mobile Artificial Intelligence for Digital Therapeutic  
Phenotyping in Computational Psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging.*  
2019.
- 51 Jonsson P, Carson S, Blennerud G, Kyohun Shim J, Arendse B, Hussein A, et al.  
Ericsson Mobility Report. [https://www.ericsson.com/4acd7e/assets/local/mobility-  
report/documents/2019/emr-november-2019.pdf](https://www.ericsson.com/4acd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf).
- 52 Insel TR. Digital phenotyping: a global tool for psychiatry. *World Psychiatry.*  
2018;17(3):276-77.
- 53 Mohr DC, Shilton K, Hotopf M. Digital phenotyping, behavioral sensing, or personal  
sensing: names and transparency in the digital age. *NPJ Digit Med.* 2020;3:45.
- 54 Torous J, Onnela JP, Keshavan M. New dimensions and new tools to realize the  
potential of RDoC: digital phenotyping via smartphones and connected devices. *Transl*  
*Psychiatry.* 2017;7(3):e1053.
- 55 Torous J, Kiang MV, Lorme J, Onnela JP. New Tools for New Research in Psychiatry: A  
Scalable and Customizable Platform to Empower Data Driven Smartphone Research.  
*JMIR Ment Health.* 2016;3(2):e16.
- 56 Onnela JP, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance  
Behavioral and Mental Health. *Neuropsychopharmacology.* 2016;41(7):1691-6.
- 57 Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. in *Proceedings of the 2014*  
*ACM International Joint Conference on Pervasive and Ubiquitous Computing* 3–14  
(Association for Computing Machinery, Seattle, Washington, 2014).
- 58 Dagum P. Digital biomarkers of cognitive function. *NPJ Digit Med.* 2018;1:10.
- 59 Chen R, Jankovic F, Marinsek N, Foschini L, Kourtis L, Signorini A, et al. in *The 25th*  
*ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)* (ACM,  
Anchorage, AK, 2019).

- 60 Harari GM, Muller SR, Stachl C, Wang R, Wang W, Buhner M, et al. Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *J Pers Soc Psychol*. 2019.
- 61 Teo JX, Davila S, Yang C, Hii AA, Pua CJ, Yap J, et al. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. *Commun Biol*. 2019;2:361.
- 62 Cormack F, McCue M, Taptiklis N, Skirrow C, Glazer E, Panagopoulos E, et al. Wearable Technology for High-Frequency Cognitive and Mood Assessment in Major Depressive Disorder: Longitudinal Observational Study. *JMIR Ment Health*. 2019;6(11):e12814.
- 63 Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders. *J Med Internet Res*. 2017;19(3):e75.
- 64 Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*. 2016;4:e2537.
- 65 Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. *Biol Psychiatry*. 2012;72(7):580-7.
- 66 Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. *NPJ Digit Med*. 2019;2:3.
- 67 Sperry SH, Walsh MA, Kwapil TR. Emotion dynamics concurrently and prospectively predict mood psychopathology. *J Affect Disord*. 2020;261:67-75.
- 68 Busk J, Faurholt-Jepsen M, Frost M, Bardram JE, Vedel Kessing L, Winther O. Forecasting Mood in Bipolar Disorder From Smartphone Self-assessments: Hierarchical Bayesian Approach. *JMIR Mhealth Uhealth*. 2020;8(4):e15028.
- 69 Huckins JF, daSilva AW, Wang R, Wang W, Hedlund EL, Murphy EI, et al. Fusing Mobile Phone Sensing and Brain Imaging to Assess Depression in College Students. *Front Neurosci*. 2019;13:248.
- 70 Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology*. 2018;43(8):1660-66.
- 71 Ben-Zeev D, Brian R, Wang R, Wang W, Campbell AT, Aung MSH, et al. CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatr Rehabil J*. 2017;40(3):266-75.
- 72 Wisniewski H, Henson P, Torous J. Using a Smartphone App to Identify Clinically Relevant Behavior Trends via Symptom Report, Cognition Scores, and Exercise Levels: A Case Series. *Front Psychiatry*. 2019;10:652.
- 73 Ben-Zeev D, Scherer EA, Brian RM, Mistler LA, Campbell AT, Wang R. Use of Multimodal Technology to Identify Digital Correlates of Violence Among Inpatients With Serious Mental Illness: A Pilot Study. *Psychiatr Serv*. 2017;68(10):1088-92.
- 74 Epstein DH, Tyburski M, Craig IM, Phillips KA, Jobes ML, Vahabzadeh M, et al. Real-time tracking of neighborhood surroundings and mood in urban drug misusers: application of a new method to study behavior in its geographical context. *Drug Alcohol Depend*. 2014;134:22-29.
- 75 Epstein DH, Tyburski M, Kowalczyk WJ, Burgess-Hull AJ, Phillips KA, Curtis BL, et al. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. *NPJ Digit Med*. 2020;3:26.
- 76 Preston KL, Kowalczyk WJ, Phillips KA, Jobes ML, Vahabzadeh M, Lin JL, et al. Before and after: craving, mood, and background stress in the hours surrounding drug use and stressful events in patients with opioid-use disorder. *Psychopharmacology*. 2018;235(9):2713-23.

- 77 Shiffman S, Waters AJ. Negative affect and smoking lapses: a prospective analysis. *J Consult Clin Psychol*. 2004;72(2):192-201.
- 78 Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. *Psychol Med*. 2015;45(13):2691-704.
- 79 Faurholt-Jepsen M, Frost M, Christensen EM, Bardram JE, Vinberg M, Kessing LV. The effect of smartphone-based monitoring on illness activity in bipolar disorder: the MONARCA II randomized controlled single-blinded trial. *Psychol Med*. 2020;50(5):838-48.
- 80 Hassanpour S, Tomita N, DeLise T, Crosier B, Marsch LA. Identifying substance use risk based on deep neural networks and Instagram social media data. *Neuropsychopharmacology*. 2019;44(3):487-94.
- 81 Choudhury MD, Counts S, Horvitz E. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 3267–76 (Association for Computing Machinery, Paris, France, 2013).
- 82 Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting Depression From Language-Based Emotion Dynamics: Longitudinal Analysis of Facebook and Twitter Status Updates. *J Med Internet Res*. 2018;20(5):e168.
- 83 Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, et al. in *Proceedings of the 26th International Joint Conference on Artificial Intelligence* 3838–44 (AAAI Press, Melbourne, Australia, 2017).
- 84 Ricard BJ, Marsch LA, Crosier B, Hassanpour S. Exploring the Utility of Community-Generated Social Media Content for Detecting Depression: An Analytical Study on Instagram. *J Med Internet Res*. 2018;20(12):e11817.
- 85 White RW, Horvitz E. From health search to healthcare: explorations of intention and utilization via query logs and user surveys. *J Am Med Inform Assoc*. 2014;21(1):49-55.
- 86 Perdue RT, Hawdon J, Thames KM. Can Big Data Predict the Rise of Novel Drug Abuse? *Journal of Drug Issues*. 2018;48(4):508-18.
- 87 Niforatos JD, Zheutlin AR, Pescatore RM, Raja AS. Public interest in medication-assisted treatment for opioid used disorder in the United States. *Am J Emerg Med*. 2019;37(10):1983-85.
- 88 Jacobson NC, Lekkas D, Price G, Heinz MV, Song M, O'Malley AJ, et al. Flattening the Mental Health Curve: COVID-19 Stay-at-Home Orders Result in Alterations in Mental Health Search Behavior in the United States. *PsyArXiv*. 2020.
- 89 Ebner-Priemer U, Santangelo P. Digital phenotyping: hype or hope? *Lancet Psychiatry*. 2020;7(4):297-99.
- 90 Cohen AS, Schwartz E, Le T, Cowan T, Cox C, Tucker R, et al. Validating digital phenotyping technologies for clinical use: the critical importance of "resolution". *World Psychiatry*. 2020;19(1):114-15.
- 91 Hirschtritt ME, Insel TR. Digital Technologies in Psychiatry: Present and Future. *Focus (Am Psychiatr Publ)*. 2018;16(3):251-58.
- 92 Nebeker C, Leow AD, Moore RC. From return of information to return of value: Ethical considerations when sharing individual-level research data. *Journal of Alzheimer's Disease*. 2019;71(4):1081-88.