ANNUAL REVIEWS **Further**

Click here to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning

David C. Mohr,[1] Mi Zhang,[2] and Stephen M. Schueller[1]

[1]Center for Behavioral Intervention Technologies and Department of Preventive Medicine, Northwestern University, Chicago, Illinois 60611; email: d-mohr@northwestern.edu, schueller@northwestern.edu

[2]Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824; email: mizhang@egr.msu.edu

Annu. Rev. Clin. Psychol. 2017. 13:23–47

First published online as a Review in Advance on March 17, 2017

The *Annual Review of Clinical Psychology* is online at clinpsy.annualreviews.org

https://doi.org/10.1146/annurev-clinpsy-032816-044949

Copyright © 2017 by Annual Reviews. All rights reserved

Keywords

mental health, mHealth, machine learning, pervasive health, wearables, sensors

Abstract

Sensors in everyday devices, such as our phones, wearables, and computers, leave a stream of digital traces. Personal sensing refers to collecting and analyzing data from sensors embedded in the context of daily life with the aim of identifying human behaviors, thoughts, feelings, and traits. This article provides a critical review of personal sensing research related to mental health, focused principally on smartphones, but also including studies of wearables, social media, and computers. We provide a layered, hierarchical model for translating raw sensor data into markers of behaviors and states related to mental health. Also discussed are research methods as well as challenges, including privacy and problems of dimensionality. Although personal sensing is still in its infancy, it holds great promise as a method for conducting mental health research and as a clinical tool for monitoring at-risk populations and providing the foundation for the next generation of mobile health (or mHealth) interventions.

Contents

1. INTRODUCTION 24
2. FROM DATA TO KNOWLEDGE: A HIERARCHICAL MODEL 25
2.1. Raw Sensor Data 25
2.2. Feature Extraction: Data to Information 25
2.3. Behavioral Markers: Information to Knowledge 26
2.4. Clinical Targets 26
3. REVIEW OF PERSONAL SENSING RESEARCH 27
3.1. Mobile Phones 27
3.2. Other Devices and Platforms 31
4. METHODS 33
4.1. Supervised Learning 34
4.2. Unsupervised Learning 35
4.3. Semisupervised Learning 35
4.4. Active Learning 35
4.5. Deep Learning 36
5. CURRENT CHALLENGES IN PERSONAL SENSING 36
5.1. Study Quality and Reproducibility 36
5.2. The Curse of Variability 37
5.3. The Unknown Expiration Date 37
5.4. Balancing Accuracy and Invisibility 38
5.5. The Certainty of Uncertainty 38
5.6. Privacy, Ethics, and the Naked Truth 38
6. POTENTIAL APPLICATIONS OF PERSONAL SENSING 39
6.1. Integration into Existing Models of Care 39
6.2. Behavioral Intervention Technologies 40
6.3. Epidemiology 40
7. SUMMARY AND CONCLUSIONS 40

1. INTRODUCTION

Ubiquitous sensing: use of networked sensors weaved into everyday life to capture information about humans, environments, and their interactions anytime and everywhere

Sensor: a device that measures a physical property and produces a corresponding output

This article is the story of the collision of several innovations—ubiquitous sensing, big data, and mobile health (or mHealth)—and their potential to revolutionize mental health research and treatment. A sensor is any device that detects and measures a physical property. Sensors are as old as civilization itself. The Sumerians developed scales, which are essentially weight sensors, some 9,000 years ago, and we have continued to develop new sensors ever since. The use of sensors to measure physical properties for the purpose of understanding psychological states, or psychophysiology, has long been a core discipline within psychology. Advances in sensor technology have accelerated throughout the past decades, with sensors becoming smaller, lighter, and more accurate. Furthermore, they have become increasingly ubiquitous and embedded into networks such that they can provide vast amounts of data almost anywhere and nearly instantaneously.

Today, people are measured continuously by sensors. Many sensors are embedded in mobile phones, measuring location, movement, communication or social interaction, light, sound, digital devices in the area, and more. Smartwatches and wearable devices containing onboard sensors that track activity and physiological functions are increasingly popular. People leave digital traces when

they make credit card purchases, send a tweet, or visit a website. This digital exhaust produced by sensors has rich information about people's behavior, and, potentially, about their beliefs, emotions, and, ultimately, mental health.

Various terms have been used to describe the utilization of ubiquitous sensor data to estimate behaviors, such as reality mining (Eagle & Pentland 2006), personal informatics (Li et al. 2010), digital phenotyping (Jain et al. 2015, Torous et al. 2016), and personal sensing (Klasnja et al. 2009). We use the term personal sensing because it is easily understood and conveys the intimacy of the information. In this article, we provide an overarching model of personal sensing, review the literature on using sensors to detect mental health conditions and related behavioral markers, provide an overview of methods, and describe some of the grand challenges and opportunities in this emerging field.

Wearable: a computing technology designed to be worn; many contain embedded sensors for specific purposes, most commonly for monitoring health or fitness

Behavioral marker: behaviors, thoughts, feelings, traits, or states identified using personal sensing

Feature: a measureable property of a phenomenon, which is proximal to, and constructed from, sensor data

Machine learning: a subdiscipline of artificial intelligence that builds algorithms that have the ability to learn without explicitly programmed instructions

2. FROM DATA TO KNOWLEDGE: A HIERARCHICAL MODEL

The goal of applying personal sensing to mental health is to convert the potentially large amount of raw sensor data into meaningful information related to behaviors, thoughts, emotions (for simplicity, in this article we refer to these collectively as behavioral markers), and clinical states and disorders. Although there are many approaches to sensemaking, we present a layered, hierarchical sensemaking framework, as this illustrates a number of processes and issues. In this framework, raw sensor data are captured and converted into features that contain information. These features can then be used to define behavioral markers, often through machine learning. In the end, the entire set of features and behavioral markers can be used to identify clinical states, similar to diagnosing a disorder. Although some methods, such as deep learning (discussed in Section 4.4), do not necessarily require these steps, we believe this framework is useful both because it is more likely to be viable in most academic research contexts and because it illustrates several important issues in sensemaking. We use a simplified version of a mobile-phone sensing platform for detecting common mental health problems as an example (**Figure 1**); however, platforms could include data from any source.

2.1. Raw Sensor Data

The boxes at the bottom of the figure represent the inputs to the sensing platform in the form of raw phone sensor data. For the most part, unprocessed, raw sensor data do not contain sufficient information for the inferences we aim to make.

2.2. Feature Extraction: Data to Information

To add information, raw sensor data must be transformed into features. Features are constructs measured by, and proximal to, the sensor data. In **Figure 1**, features are depicted in the layer above the inputs. This is, arguably, the most important step in sensemaking (Bengio et al. 2013). There are a number of ways to construct and extract features. One common approach is to use domain expertise or brainstorming to inject human intelligence for feature construction. For example, raw data about phone usage may be of minimal value. If you are interested in in-phone communication, relevant features might be the number and duration of incoming calls or short message service (SMS) text messages, the number and duration of outgoing calls and SMS messages, the number of missed calls, and the ratios of these features. In addition, features can also be extracted statistically using algorithms, such as slow feature analysis and stacked autoencoders (Vincent et al. 2010, Wiskott & Sejnowski 2002), that can automatically discover new feature representations. Finally, some features estimate observable states using machine learning. For

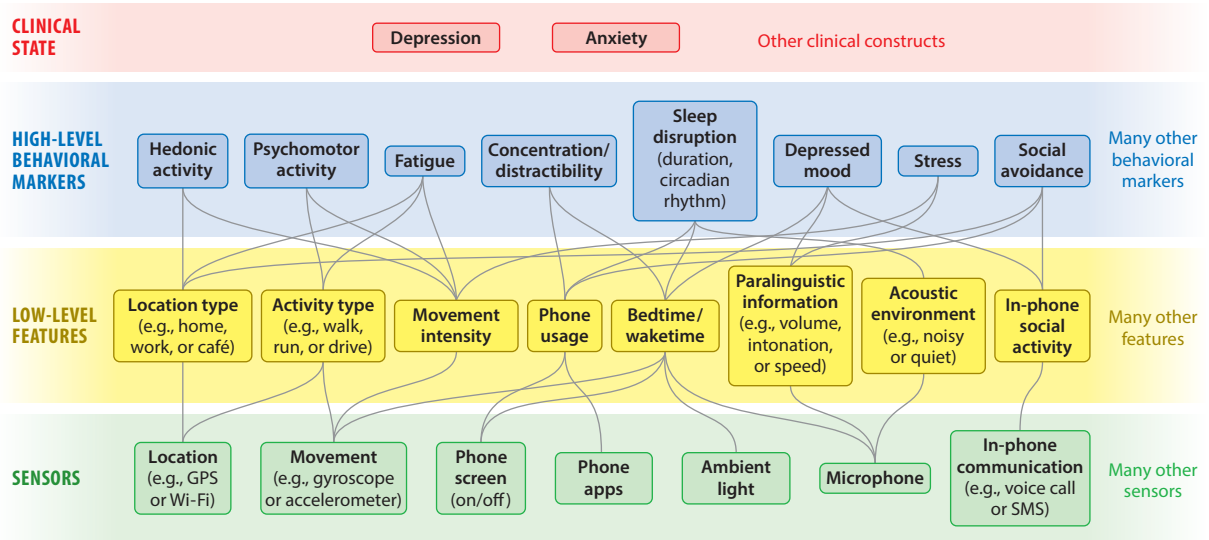

Figure 1

Example of a layered, hierarchical sensemaking framework. Green boxes at the bottom of the figure represent inputs to the sensing platform. Yellow boxes represent features. Blue boxes represent high-level behavioral markers. Abbreviations: GPS, global positioning system; SMS, short message service.

example, bedtime or waketime can be estimated using a number of sensors and features related to light, sound, and phone use (Zhenyu et al. 2013).

2.3. Behavioral Markers: Information to Knowledge

Behavioral markers are higher-level features, reflecting behaviors, cognitions, and emotions, that are measured using low-level features and sensor data. This is similar to the notion of latent constructs in psychological methodology. Some examples of potential high-level behavioral markers are represented in **Figure 1**. Behavioral markers are most commonly developed using machine-learning and data-mining methods to uncover which features and sensor data are useful in detecting the marker. For example, a behavioral marker for circadian sleep rhythm might include features such as bedtime and waketime, sleep duration, and phone usage. Markers of sleep quality might include ambient sound features, but may also include bedtime and wake time (Abdullah et al. 2014). Furthermore, the accuracy of such features may be enriched by including additional helper features, such as age (older people use phones differently from younger people) or whether it is a workday or non-workday.

2.4. Clinical Targets

One would not attempt to diagnose a mental health disorder on the basis of one or two questions about symptoms (although one might use them for screening purposes). Similarly, as discussed in Section 3, limited sets of features have been only modestly successful at predicting clinical targets. We expect that clinical targets will be better predicted by applying machine-learning methods to a larger number of behavioral markers and features. However, these may not have a one-to-one correspondence to symptoms used to diagnose disorders. Some symptoms may simply not be

detectable, and personal sensing may uncover other predictors that have not been considered to date.

Social media: technological tools that allow people, companies, and organizations to share user-generated information and connect with other users through networks

3. REVIEW OF PERSONAL SENSING RESEARCH

Most work on personal sensing for mental health has used mobile phone sensors. Therefore, we review this work before reviewing work from other areas, including wearables, social media, and computers.

3.1. Mobile Phones

Mobile phones are commonly used for research because they are widely owned: 72% of Americans own a smartphone, up from 35% in 2011 (Poushter 2016). Furthermore, people keep their phones on or near them and use them frequently. On average, people check their phones 46 times per day, and for younger people that figure is 85 times per day (Andrews et al. 2015, Eadicicco 2015). The phone also has an increasingly large number of embedded sensors. Here, we focus on behavioral markers in three areas related to mental health: sleep and social context, which are potentially more observable, and mood and stress, which are internal states.

3.1.1. Behavioral markers. Although a growing literature examines detection of an increasingly broad range of behavioral markers using mobile phone sensors, we describe the work on detection of sleep, social context, mood, and stress as examples because these markers have received the most attention.

3.1.1.1. Sleep. Sleep disturbance is a common symptom, occurring across many mental health conditions (Sivertsen et al. 2009, Taylor et al. 2005). Those disturbances can be reflected by patients' sleep periods (i.e., when and how long a person sleeps) and sleep quality (i.e., how well a person sleeps). By leveraging built-in sensors, a number of smartphone-based sensing systems have been developed to passively monitor sleep periods. Several groups have shown that sleep duration can be estimated with approximately 90% accuracy, without asking the user to do anything special with the phone, by using data from a number of sensors, such as the accelerometer, microphone, ambient light sensor, screen proximity sensor, running process, battery state, and display screen state (Chen et al. 2013, Min et al. 2014). Among heavy phone users, such as undergraduate students, sleep periods can be detected simply by observing phone screen lock and unlock events, which is less of a drain on the phone battery than other methods (Abdullah et al. 2014). Sleep period markers can then be used to create circadian-aware systems. For example, non-workday sleep duration can be used to estimate a person's chronotype (e.g., morning lark versus night owl), and changes in sleep patterns across workdays and non-workdays can identify social jet lag, which is the difference between a person's biological sleep rhythm and external requirements (Abdullah et al. 2014, Murnane et al. 2015). Such sleep period markers have also been correlated with the severity of depressive symptoms (Wang et al. 2014).

Sleep quality can also be effectively inferred by using smartphone sensors. For example, common events that interfere with sleep quality, such as body movement, coughing and snoring, and ambient noise, can be reliably detected using a smartphone's microphone when the phone is kept in the user's room. Acoustic features have been associated with both short-term (i.e., 1-night) and long-term sleep quality measured using actigraphy and self-report (Hao et al. 2013). A number of studies have used multimodal sensing schemes—including the accelerometer, microphone, light

sensor, screen proximity sensor, running process, battery state, and display screen state—to infer sleep stages and sleep quality (Gu et al. 2014, Min et al. 2014).

3.1.1.2. Social context. Here, we focus on phone-based research; we discuss work using social media networks in Section 3.2.2. Large population-based phone datasets can provide dynamic information about individual movement and proximity to others that can be used to calculate people's proximity in a social network. Patterns of movement and of colocation can be used to infer relationships and predict new social ties (Hsieh & Li 2014, Pham et al. 2013, Wang et al. 2011).

On a smaller scale, in a classic study, Eagle and colleagues (Eagle & Pentland 2006, Eagle et al. 2009) were able to identify friends and non-friends with a high degree of accuracy using Bluetooth sensors, which can detect other Bluetooth-enabled devices up to 15 meters away. People in proximity to one another only during work hours were more likely to be colleagues than friends, but proximity during the evening or weekends was an indicator of friendship. Relational status can then be used to identify other psychological targets. For example, calling friends during work hours was associated with lower job satisfaction (Eagle et al. 2009). Although such methods hold promise, their application is limited by the small percentage of people who leave their Bluetooth sensors in discoverable mode. Nevertheless, this work (Eagle & Pentland 2006, Eagle et al. 2009) underscores the utility of colocation and time as important features that indicate the nature of relationships.

Other forms of social sensing have used remote communication tools within the phone, including calls and SMS messages. Contact lists (address books) within a person's phone can contain information about relationships. For example, contact fields sometimes include family role (e.g., Aunt Julie), relationship context [e.g., Kaitlyn (Peg's friend)], phone type (e.g., Mom at home), or an honorific (e.g., Mrs., Mr., or Dr.), which can be mined to infer relationships (Wiese et al. 2014). However, contact lists are vulnerable to idiosyncratic labeling methods and tell us little about the frequency of contact.

Patterns of time and frequency and the regularity of incoming and outgoing calls and SMS messages have also been have been used to classify a person's contacts into a relationship domain (family, friend, or work colleague), with more than 90% accuracy (Min et al. 2013). For example, longer calls were associated with family; the work domain was characterized by fewer weekend calls and a lower likelihood of SMS messages; and friends and social contacts were characterized by more SMS messages sent during the week. The strength of social ties can also be estimated to some degree, with higher levels of in-phone communication frequency, call duration, and communication initiated by the phone owner being associated with a stronger relationship (Wiese et al. 2015). However, this signal is noisy because low levels of communication do not necessarily mean weak ties (we speak infrequently with some people who are very close to us), as much communication may occur outside of the phone, such as face to face or using other media, and, increasingly, other applications (or apps) such as Snapchat and WhatsApp.

3.1.1.3. Mood and stress. Mood and stress are internal states that are likely to be more distal from the sensors and features normally used in personal sensing. A number of studies have attempted to leverage a broad array of built-in mobile phone sensors to predict mood (LiKamWa et al. 2013, Ma et al. 2012, Madan et al. 2010). In the earliest study, Madan et al. (2010) found that decreases in calls, SMS messaging, Bluetooth-detected contacts, and location entropy (a measure of the temporal dispersion of locations) were strongly related to feeling sad and stressed among students, as measured by daily ecological momentary assessments (EMAs). Moodscope (LiKamWa et al. 2013) was used to infer mood, labeled by daily EMAs, from data from 32 participants over

2 months. The number and length of communications (calls, SMS text messages, and emails), the number of apps used and usage patterns, web browser history, and a person's location could be used to estimate a user's daily mood average with an initial accuracy of 66%, which gradually improved to 93% after a 2-month personalized training period. Similarly, using location, motion detectors, light, and ambient sound, Ma et al. (2012) achieved approximately 50% accuracy for daily moods during a 30-day period with 15 participants. A recent attempt to replicate the Moodscope findings with a cohort of 27 students failed to perform better than chance (Asselbergs et al. 2016). The varying results and failure to replicate suggest that although a number of small studies have demonstrated the technical feasibility of sensing mood, these findings do not appear to generalize.

At least one study has attempted to detect stress using the swipe, scroll, and text-input interactions with a phone (Ciman et al. 2015). This work, based in part on literature showing that stress can be detected through computer mouse and keyboard interactions (see Section 3.2.3), found that under laboratory conditions, features derived from a person's scroll, swipe, touch, and text-input interactions with a phone could differentiate a laboratory-induced stressful state from a normal state. It remains an open question whether real-world instances of these interactions provide a strong enough signal of stress.

A large body of literature has demonstrated that affect and mood can be detected through the paralinguistic features of speech (Calvo & D'Mello 2010). StressSense (Lu et al. 2012) is a smartphone sensing system that uses the phone's built-in microphone to capture human speech during social interactions to infer a user's level of perceived stressed by analyzing paralinguistic information, such as pitch and speaking rate. Under quasi-experimental conditions using a mock job interview and a marketing task, StressSense achieved 76–81% accuracy in identifying stress. These findings were then extended to a real-world evaluation. Following 7 participants over 10 days, the sensed stress marker correlated with self-reported stress at $r = 0.59$ (Adams et al. 2014). Another example, Emotion Sense (Rachuri et al. 2010) proposed that audio-based emotion recognition could identify up to 14 different emotions clustered into five broader emotion groups (happy, sad, fear, anger, and neutral). In an initial proof-of-principle 10-day study involving 18 participants, the distribution of the emotions detected through Emotion Sense generally reflected the self-reports of the participants.

Detecting mood or subjective stress is likely a challenge for many commonly available sensors in smartphones. Given the history of paralinguistic voice features predicting mood, using a built-in microphone would seem promising from an analytical perspective, but it may pose challenges from logistical and ethical perspectives for acquiring samples of sufficient quantity and quality. This illustrates a disconnect that sometimes occurs between technical and laboratory proof-of-concept and real-world feasibility.

3.1.2. Clinical disorders. Some studies have examined the possibility of using smartphone sensor data to detect the presence and severity of mental health disorders, including depression, bipolar disorder, and schizophrenia.

3.1.2.1. Depression. Early work using smartphones for personal sensing began by examining depression. Madan et al. (2010) followed 70 undergraduates living in a residence hall and found that decreases in total communication were associated with greater depression. Depression, however, was assessed using only a single item. A second study, StudentLife, used the Patient Health Questionnaire-9 (PHQ-9) to assess depression among 48 students over 10 weeks (Wang et al. 2014). Similar to the study by Madan et al. (2010), conversation frequency and duration, measured using the microphone, as well as colocation with other students, detected using a global positioning system (GPS) and Bluetooth, were significantly related to depression. In addition, depression was

Global positioning system (GPS): sometimes refers to location services that fuse GPS with other signals, such as Wi-Fi, to obtain greater accuracy with less battery drain

associated with a previously developed sleep-duration classifier (Chen et al. 2013). The relationship between sensed social contact and depression was also observed in a nonstudent population using elderly people living in a retirement community (Berke et al. 2011).

These relationships are perhaps unsurprising, given that sleep disruption is a symptom of depression and social withdrawal and avoidance are factors significantly related to common mental health problems, such as depression and anxiety (Hames et al. 2013). The fact that sleep, social withdrawal, and anxiety were inferred through smartphone sensors demonstrates the utility of the hierarchical model displayed in **Figure 1**.

Although sensing has the potential to automate the detection of behaviors we know are related to disease states, it also has the potential to uncover new information that may lead to new understanding. A case in point is the emerging work on the relationship of GPS data to depression. The first study in this area, using 2 weeks of data from 28 participants, found that a number of GPS-derived location features were associated with depression (Saeb et al. 2015). The number of places a person visited was not related to depression; however, location entropy (the variability in time spent in different locations) was, such that the more time clustered around a few locations, the more likely the person was to be depressed; more equal time distributions were related to lower depression scores. A feature measuring periodicity, or the circadian rhythm of movement through geographical space, was particularly strongly related to depression, suggesting that disruption in the regularity of movement was associated with a greater severity of depressive symptoms. These findings were then replicated in the StudentLife dataset described above (Saeb et al. 2016). A third study, using somewhat different methods, similarly found that similar GPS features could estimate depression (Canzian & Musolesi 2015).

This general relationship between mobility and depression has been explored in more detail. For example, the relationship between GPS features and depression is stronger on non-workdays than it is on workdays when much movement is driven by social expectations (Saeb et al. 2016). This suggests that distinguishing between times when behaviors are more under the individual's control versus when they are not may identify features that can be used to increase the accuracy of models. GPS features appear to predict depression many weeks in advance, although the relationship between depression and subsequent GPS features degrades quickly over time, suggesting that a lack of mobility may be an early warning signal of depression.

3.1.2.2. Bipolar disorder. The MONARCA project (MONitoring, treAtment and pRediCtion of bipolAr Disorder Episodes) pioneered the use of smartphone-based behavior monitoring technologies for mental health (Gravenhorst et al. 2015). MONARCA leveraged a variety of phone sensors to detect the mental states of patients, as well as changes in mental states. To validate the effectiveness of their smartphone-based sensing system, the MONARCA team conducted a series of real-world studies among bipolar patients from a rural psychiatric hospital in Austria. Based on 12 patients followed for 12 weeks, accelerometry, location, or fused accelerometry–location features produced clinical state (depression/mania) recognition accuracy of 72–81% and state-change detection with a precision and recall of, respectively, 96% and 94% (Grünerbl et al. 2014). By fusing phone call features with paralinguistic information, a state-recognition accuracy of 76%, as well as a precision and recall accuracy of, respectively, 97% and 97% for state-change detection, were achieved (Grünerbl et al. 2015). Another analysis of 18 patients over 5 months indicated that a smartphone app can be used to identify stress and mood (Alvarez-Lozano et al. 2014). For example, higher use of social and entertainment apps was associated with lower stress and irritability.

Work has also examined the potential for GPS features, originally developed for depression (Saeb et al. 2015), to detect depressive episodes among bipolar patients. The same features,

including entropy and circadian rhythm, remain strongly related to the severity of depression in this population (Palmius et al. 2016). Furthermore, when combined, these features could distinguish depressed from non-depressed states with 85% accuracy. This underscores the potential utility of features across diagnoses when examining similar states.

3.1.2.3. Schizophrenia. Work on sensing in schizophrenia has begun just recently. A survey of patients with schizophrenia suggested that most are comfortable using a smartphone with sensing, and are interested in potentially receiving feedback and suggestions from such a system, although a minority voiced concerns that it might upset them or were concerned about a loss of privacy (Ben-Zeev et al. 2016). In a first study, 34 patients with schizophrenia were provided with a smartphone for 2–8.5 months that collected a variety of sensor data. Personalized models used a number of features to predict EMA responses. For example, changes in physical activity, detected conversations, and later bedtimes were associated with self-reported worry that someone was intending to harm the participant, and with self-reported auditory and visual hallucinations (Wang et al. 2016).

3.2. Other Devices and Platforms

We have described personal sensing using mobile phones because this is the most ubiquitous personal sensing platform, harnessing data from people's lives with little to no ongoing effort or actions on the part of the user. However, many other sources of data exist, which have their own strengths and weaknesses. Below we review data from wearables, social media, and computers.

3.2.1. Wearables. Wearable devices (or wearables) are sensor-enabled technologies designed to be worn for specific purposes, most commonly health and fitness. These devices, such as Fitbit and Jawbone, track activities continuously, for example, how many steps people take, how many miles they run, and how long they sleep. Wearables, which use sensors designed for their specific targets, and that are intended to be worn in a specified and consistent manner (e.g., on the wrist or clipped to the belt), may provide data that is of significantly higher quality than that provided by smartphones, which are not designed specifically for health tracking. However, the increase in data quality may be offset by other drawbacks. Wearables are less prevalent than smartphones, with only 19% ownership among Americans (Ricker 2015). Their use is higher among those already motivated to keep a watchful eye over their health, and many people abandon using them soon after purchase (Piwek et al. 2016).

The most widely used sensor in wearables is the accelerometer. Accelerometer-based wearable devices have been developed for tracking physical exercise (Choudhury et al. 2008), detecting falls (Li et al. 2009), and monitoring activities of daily living (Spenkelink et al. 2002). In a large study of 2,862 participants, greater levels of accelerometer-based physical activity were strongly associated with decreased rates of depression (Vallance et al. 2011).

Increasingly, wearable devices are including a broader range of sensors that can measure variables that are useful for mental health researchers, such as skin conductance and heart rate. For example, investigators have noted that greater asymmetries in skin conductance amplitude on the left and right sides of the body are an indicator of emotional arousal (Picard et al. 2016). Many of these sensors are now available in smartwatches, which attempt to leverage a behavioral and cultural pattern to avoid the problem of abandonment seen with other wearables. It remains to be seen whether smartwatches will attain the ubiquity enjoyed by smartphones.

Wearables are being developed that are dedicated to behaviors that have been difficult to sense. Eating and appetite, for example, are often disrupted in mental health conditions, but are

difficult to detect through commonly available sensors. Two specific methods of detection, gestures and sound, may be particularly useful for assessing these behaviors. Because eating and drinking activities normally involve repetitive wrist movements and rotations, wrist-worn wearables that include an accelerometer and gyroscope have shown promise in capturing these activities (Edison et al. 2015, Sen et al. 2015). Eating and drinking may also produce idiosyncratic sounds through chewing and swallowing. A microphone attached at the neck can classify sounds produced by eating and drinking with reasonable accuracy (Kalantarian et al. 2015, Rahman et al. 2014, Yatani & Truong 2012).

3.2.2. Social media. With 65% of Americans using social media in 2015 (Perrin 2015), platforms such as Facebook and Twitter have become common places where people share their opinions, feelings, and daily experiences. The field of psycholinguistics has long demonstrated that the linguistic analysis of speech can be used for diagnostic classifications (Oxman et al. 1982, Rude et al. 2004). Thus, social media postings, which consist largely of language, are a potential source of information about mental health, as well as people's thoughts and feelings related to those conditions. Using a large dataset of more than 28,000 Facebook users who completed a personality survey, Schwartz et al. (2014) found that features generated from posts were modestly related to depression severity. Themes related to depression include depressed mood, hopelessness, hopelessness and helplessness, symptoms, relationships and loneliness, hostility, and suicidality. Similarly, De Choudhury (2013b) found that depressed Twitter users can be distinguished from non-depressed users based on later posting times, less frequent posting, greater use of first person pronouns, and greater disclosure about symptoms, treatment, and relationships. Furthermore, the development of a future depressive episode could be predicted with 70% accuracy. In a large sample of Twitter users, rates of depression were consistent with geographical, demographic, and seasonal patterns reported by the US Centers for Disease Control and Prevention (CDC) (De Choudhury et al. 2013a).

Social media likely will be helpful in identifying behavioral markers that are strongly related to cognitive and motivational factors, which are difficult to evaluate through nonverbal sensors. For example, language features from Facebook posts have shown modest but consistent correlations with Big 5 personality factors (Park et al. 2015). Twitter-derived features related to suicidal ideation have been shown to correlate strongly with rates of completed suicides from the CDC (Jashinsky et al. 2014). Suicidal ideation has also been associated with language used in social media that shows heightened self-attention focus, poor linguistic coherence and coordination with the community, reduced social engagement, and manifestations of hopelessness, anxiety, impulsiveness, and loneliness (De Choudhury et al. 2016). Thus, language generated naturalistically through social media may be a useful tool for sensing mental health conditions, and it may be particularly well suited for behavioral markers that involve cognitive or motivational states that are beyond the reach of nonverbal sensors.

3.2.3. Computers. Many people spend a considerable amount of their lives at computers and many interactions still take place through the mouse and keyboard. A number of studies have examined whether mouse movements and keyboard taps can provide information about a person's mental state. An early study explored a broad range of possible features derived from mouse movements to predict experimentally induced emotions (Maehr 2008). Although most features showed no relationships to emotions, motion breaks, or discontinuities in movement, were related to overall arousal and discrete emotions, such as disgust and anger. Motion breaks resemble the pause features observed in speech that have been related to stress, and it is possible that this is a

general behavioral pattern when one is stressed, observable across multiple channels. Another study explored whether muscle tension would change the dynamics of the movement (resonant frequency and damping ratio) and, thus, be an observable correlate of stress. Data collected in a laboratory setting demonstrated that simple models of arm–hand dynamics applied to mouse motions were strongly related to concurrently collected physiological measures of stress and arousal (Sun et al. 2014). This signal remained strong across a variety of mouse tasks, including clicking, dragging, and steering.

Similar affective inferences have been made from keyboard activity. One study tracking everyday computer use for 1 month among 12 participants found promising accuracy (70–88%) for self-reported emotion (Epp et al. 2011). Models were trained using a decision-tree classifier and features derived from short key sequences, such as duration of and latency between keystrokes, to predict common discrete emotional states. Although these results were encouraging, classification rates represented only a modest gain over baseline classification.

3.2.4. Additional sources of data. Phones, wearables, social media, and computers are far from the only technologies that produce digital traces. Other potential streams of data include purchasing behavior, browsing history, or productivity apps, such as calendars and email. Furthermore, social context could be better understood by making use of Google maps or other repositories of images of environments. Such images can be mined to determine the environmental factors that affect mental health and well-being (e.g., the amount of green space or number of trees in a neighborhood, or cleanliness or number of surfaces tagged with graffiti). We have not discussed these, primarily because they have not yet been investigated in relationship to mental health. Another rapidly expanding area is ambient intelligence, in which sensors are installed on everyday objects and in places where people live to sense people's movements, gestures, habits, and intentions, and respond to needs in a seamless and nonintrusive manner (Acampora et al. 2013). Indeed, such ambient systems have the potential to provide visibility into the most intimate spheres of a person's life.

4. METHODS

The field of personal sensing is young, with almost all of the research occurring during the past few years. Most of the studies have been conducted by computer science and engineering groups using research models that are very different from those commonly used in the clinical and behavioral sciences (Intille 2013). As we in the behavioral sciences begin work in this area, it is important to understand the fundamental differences between the methods used by computer scientists and those with which we are more familiar.

First, engineering and computer science research is typically exploratory in nature, focused on solving a problem. In the area of personal sensing, computer scientists tend to collect as much data as possible, using data-mining methods to develop classification algorithms. Although these analytical methods employ techniques such as cross-validation to avoid overfitting in the models, these methods are, nonetheless, quite different from commonly used clinical methods, which come from a positivist tradition and are hypothesis driven and confirmatory rather than exploratory. Said in a different way, clinical scientists tend to design a study to test an answer to a question. Engineers tend to design a study to find an answer to a question.

Second, engineers and computer scientists, in their quest for novel solutions, tend to have a higher tolerance for risk and change in their research than do clinical scientists. Clinical scientists place a much higher value on eliminating as many threats to internal validity as possible, and have

a low tolerance for methods that might limit confidence in the results. As such, clinical scientists aim to avoid type I errors or spurious positive findings, but engineers and computer scientists see type II errors as a greater threat because they may lead to overlooking a potentially novel and useful solution.

Supervised learning: a category of machine learning that uses labeled data provided by a set of training samples to construct an algorithm

Unsupervised learning: a category of machine learning that attempts to uncover underlying structure in data and does not require labeled data

Semisupervised learning: a category of machine learning that combines aspects of supervised and unsupervised methods by using samples with both labeled and unlabeled data

Label: in machine learning, a label refers to that which is being predicted, similar to a dependent variable in statistics

Finally, engineering and computer science are frequently focused on a proof of principle, but clinical scientists value generalizability. It is perhaps only a slight exaggeration to say that computer scientists in personal sensing are asking, "Does this work at all?"; clinical scientists want to ask, "Will this work for a population under all circumstances?" Thus, most of the studies in personal sensing have been small, generally with sample sizes of 7–30 participants, commonly using convenience samples consisting of college students. (The social media studies are the exception, using very large datasets.) It is not uncommon for comparatively large percentages of enrolled participants, sometimes on the order of half the sample, to be excluded from analyses due to any number of problems in data acquisition or data quality. Thus, a clinical scientist might see these as offering little assurance that the findings might extend outside of the research context. However, engineers and computer scientists have demonstrated that a novel solution may have value.

The main analytical method used for personal sensing is machine learning (Bishop 2006). The goal of machine learning is to identify potentially complex relationships among data, and to use the identified relationships to make predictions about new data. We review three commonly used machine-learning analytical methods: supervised learning, unsupervised learning, and semisupervised learning, as well as a new trend in machine learning called deep learning.

4.1. Supervised Learning

Supervised learning is a category of algorithms in machine learning that aims to learn a function that maps data to labels provided by a set of training samples. A label in machine learning refers to that which is being predicted, similar to a dependent variable in statistics. In personal sensing, labels are often users' self-reports. A training sample is a pair consisting of a data instance and its label. Like a teacher supervising learning in a classroom, the labels supervise the learning process, which occurs through training samples. The learned mapping function is then applied to data in the absence of labels to predict their labels. If labels are categorical values, the supervised learning algorithms are called classification algorithms, and the mapping function is referred to as a classifier. Learning algorithms for continuous values are called regression algorithms, and the mapping function is called a regression function.

Classification is the most commonly used supervised learning method. For example, activity recognition can be formulated as a classification problem in which sensor data are mapped to different activity labels, such as walking, running, or sleeping. There are two families of classification algorithms: generative algorithms and discriminative algorithms. Generative algorithms learn a classifier of the joint probability of the data instances and their labels, and then calculate the posterior probability by applying Bayes' theorem to predict labels of new data instances (Ng & Jordan 2002). Naive Bayes' models, hidden Markov models, and Gaussian mixture models are some of the most commonly used generative algorithms. In contrast, discriminative algorithms build a model to describe the boundaries that separate different labels. Examples of discriminative algorithms include logistic regression, support vector machine, and conditional random fields. Generative and discriminative algorithms have unique strengths and weaknesses. In practice, the classification performance of discriminative algorithms tends to be better than that of generative algorithms (Bishop & Lasserre 2007). However, generative algorithms can identify data that come from a new label that is not included in the training samples, for example, identifying the new activities of a user (e.g., yoga) that are not included in the training samples.

4.2. Unsupervised Learning

The goal of unsupervised learning is to find hidden structure within the data. In unsupervised learning, the training samples do not have labels and contain only data instances. There are three families of unsupervised learning algorithms: clustering, anomaly detection, and dimensionality reduction, each aiming to identify different structures within the data. Clustering algorithms (such as *K*-means and hierarchical clustering) aim to divide data instances into separate clusters such that data in the same cluster are similar but they are dissimilar from data in different clusters. Anomaly detection algorithms (such as one-class support vector machines) aim to identify the few instances that are very different from the majority of the data. Finally, dimensionality reduction algorithms (such as feature selection and principal component analysis) aim to remove multicollinearity and retain the most important information about the data to avoid the effects of the curse of dimensionality (Bishop 2006), thereby improving the generalization performance of machine-learning models.

Curse of dimensionality: as the number of dimensions expands, the data in the space become sparse, which prevents machine-learning methods from being efficient

Active learning: subset of semisupervised learning; an algorithm queries a user for additional labels when it is uncertain how to classify a set of data

In personal sensing, unsupervised learning algorithms are often used to preprocess sensor data before using supervised methods for further processing. For example, clustering algorithms have been applied to GPS data (i.e., pairs of latitudes and longitudes) to create a heat map and to find points of interests to the user (e.g., home, workplace) (Saeb et al. 2015). Anomaly detection algorithms have been used to detect changes in the mental states of bipolar patients so that just-in-time interventions can be delivered (Grünerbl et al. 2014). Finally, dimensionality reduction algorithms have been applied to identify the most important behavioral markers to best predict the mental states of depressive patients (Saeb et al. 2015).

4.3. Semisupervised Learning

As its name implies, semisupervised learning is in between supervised and unsupervised learning. It uses training samples that contain both labeled and unlabeled data to achieve better performance than could be achieved by simply using supervised learning trained on the labeled data (Zhu & Goldberg 2009). Semisupervised learning is practical for personal sensing, in which there is a large ratio of unlabeled to labeled data. For example, it would be burdensome, expensive, and time-consuming, if not impossible, to label every minute of GPS or accelerometry data collected throughout a day (Chapelle et al. 2006). Semisupervised learning addresses this problem by leveraging the intrinsic structure of the unlabeled data taken together with information provided by the labeled data.

4.4. Active Learning

Active learning is a special case of semisupervised learning. Active learning algorithms query a user to provide additional labels when the algorithm detects that a user's behavior or state deviates from what it has been trained to before and, thus, the algorithm is uncertain how to classify it. Therefore, in contrast to supervised learning models, which are static and cannot be updated after the training period, active learning algorithms are able to update users' models after getting additional labels. In this way, they can provide evolving models that adapt to users' changing behaviors and states.

Active learning is especially useful in generating personalized models from group models. Group models are algorithms that, once developed, are intended to run passively, with no input from the user (much like activity-tracking devices). Personalized models require user labeling to create a model that is specific to the individual. Personalized models tend to perform better than group models, but they incur labeling burden. However, this labeling burden may be somewhat

mitigated with hybrid models that are initiated with group models and optimized through user labeling via active learning. In particular, active learning can help derive personalized models with less labeling, as the algorithm requests labels only when they are needed (Settles 2010).

4.5. Deep Learning

During the past decade, deep learning, a new trend in machine learning, has emerged (Schmidhuber 2015). Methods developed based on deep learning have dramatically improved the state of the art, and have beaten other machine-learning methods in a wide range of applications, such as identifying objects in images (Krizhevsky et al. 2012), recognizing speech (Hinton et al. 2012), translating languages (Sutskever et al. 2014), understanding the genetic determinants of diseases (Xiong et al. 2015), and predicting health status using electronic health records (Miotto et al. 2016).

The success of deep learning is rooted in a revolutionary way of extracting features from data. It is well understood that the performance of machine-learning methods largely depends on the features chosen (Bengio et al. 2013), which traditionally has required considerable human effort and domain knowledge. Although these hand-engineered features exhibit great performance in small datasets, they do not generalize well to challenging problems involving large-scale datasets (LeCun et al. 2015). In contrast, deep learning adopts a data-driven approach in which a general-purpose procedure automatically learns features from data, with no prior domain knowledge needed. These self-learned features are organized in a multilevel hierarchy in which higher-level features are defined from lower-level ones, similar to the layered, hierarchical sensemaking framework illustrated in **Figure 1**.

Deep learning may be vulnerable to overfitting at smaller sample sizes, thus often making traditional machine-learning methods a better fit. However, once an adequate sample size has been obtained, deep learning exhibits superior capability at capturing the intricate characteristics of data that traditional machine-learning methods fail to capture. Therefore, deep learning achieves much better performance than other methods as the sample size increases. Furthermore, although self-learned, multilevel features generated purely by machines may not be well understood by humans, it is possible that these features may uncover new understanding about the constructs we are measuring, but may do so by increasing the complexity beyond human understanding.

5. CURRENT CHALLENGES IN PERSONAL SENSING

5.1. Study Quality and Reproducibility

A growing number of studies appears to replicate the findings of other studies (albeit using small, narrow samples) using machine-learning methods that estimate behavioral markers, such as mood, stress (LiKamWa et al. 2013, Ma et al. 2012, Madan et al. 2010), and sleep (Abdullah et al. 2014, Chen et al. 2013, Min et al. 2014), by using a combination of phone sensor data and features. However, because computer science and engineering tend to value technical novelty over generalizability, studies that appear to address the same behavioral marker usually use different sensors, different sets of features, different methods of measuring the behavioral markers, and varying research designs (e.g., giving people phones versus having them use their own, studying them for varying periods of time, or having varying numbers of participants excluded). The machine-learning methods used vary, and the results or weightings, particularly for group models, are not necessarily comparable across studies. In addition, it is unclear how many attempts have not been published due to failure. The one replication study we are aware of that used nearly identical methods to the original research was unable to reproduce the very strong findings in the

original paper on prediction of mood (Asselbergs et al. 2016). Thus, studies examining the use of machine-learning methods to estimate behavioral markers indicate that it is feasible under narrow conditions; however, the reliability needed for clinical use has not been demonstrated.

Furthermore, the availability of easy-to-use tools for machine learning is expanding faster than the expertise, resulting in a growing number of publications using questionable methods. A recent review of papers that used sensor data to detect disease states found that half used inappropriate cross-validation techniques, which greatly overestimated prediction accuracy (Saeb et al. 2016). Furthermore, papers that used these inappropriate techniques were cited just as often as papers using proper techniques, suggesting that poor-quality information is having the same impact as high-quality information. Although the papers cited in this review did not evidence these types of methodological problems, it would behoove interested scientists to explore the field with a healthy mix of excitement and skepticism.

5.2. The Curse of Variability

As we move from narrow proof-of-concept studies to testing in broader populations, the sources of variability expand enormously, emanating from a variety of sources, including data types, characteristics of people, and different environments. Sensors in smartphones vary from manufacturer to manufacturer, model to model, and version to version, affecting the data collected. People's characteristics might impact the relationship between constructs or how they use the measurement devices. For example, age may be related to the number of social contacts, with older people having fewer contacts and wanting less contact, but it may also be related to how social activity is measured with a phone (e.g., older people are more likely to call and less likely to send text messages than younger people). Where people carry their smartphone (pocket, handbag, backpack) can profoundly affect the sensor data. Environment and seasonality represent additional important dimensions. For example, GPS and accelerometer data in winter will look different in Minneapolis relative to Miami.

When dimensionality is high, individual small studies are unlikely to be adequately powered to create reliable and generalizable classifiers for use in larger populations. Efforts such as the Precision Medicine Initiative (Collins & Varmus 2015), which plans to enroll more than 1,000,000 people, may provide such opportunities, but it is unclear what data will be collected and how. An approach used in other fields that have similar problems, such as genomics, is to pool data across studies. A challenge in pooling data is to find a scientifically valid balance between identifying uniform variables, which makes data pooling straightforward (e.g., using the exact same questions) but can be hard to implement, and using statistical methods to manage heterogeneity by providing similar, if not identical, data points (Fortier et al. 2011). The field of personal sensing in mental health is still young and small enough that some agreement on a core set of clinical assessment methods (EMA or self-report) may be possible, thereby providing uniform anchors to which the broad range of sensor data could be tied as it evolves and changes over time and across research projects.

5.3. The Unknown Expiration Date

Personal sensing algorithms will likely have shelf lives, which may be relatively short. As devices and sensors are updated, the associated raw data will change over time. Additionally, the way we use these devices and platforms will change as well. Just in the past few years, smartphone use has changed dramatically. We spend more time reading and watching movies on our phones, and communications have shifted away from calls and SMS texts to messaging apps and social media. Social media are becoming increasingly more visual relative to being text based, and interfaces and notification methods are changing when, how, and what people write. As people change how

they use the devices that provide the data, machine-learning algorithms will become increasingly inaccurate.

Google Flu Trends offers a high-profile, cautionary tale. Launched in 2008, it mined flu-related search terms, producing results that closely matched the CDC's surveillance data and provided the information more rapidly (Butler 2013). The system was rolled out to 29 countries and extended to other diseases. It performed remarkably well until it stopped working. How people conducted searches changed over time, rendering the algorithms ineffective. The changes in people's search strategies were driven at least in part by Google's own efforts to optimize search algorithms, which also altered the search recommendations provided to users, thus changing people's search behaviors and, ultimately, undermining Google Flu Trends' models.

5.4. Balancing Accuracy and Invisibility

A common goal in personal sensing is to make acquiring data and predicting behavioral markers as unobtrusive or invisible to the user as possible. On the one hand, requiring user actions will likely result in abandonment of the tool by some percentage of users. On the other hand, personalized, active-learning models, which require user labeling, perform better than static group models. Active-learning models also allow for recalibration over time, potentially eliminating the shelf-life problem.

Thus, users who provide a little bit of data would enjoy substantially higher predictive accuracy from the models. Rather than thinking of a sensing platform as a technology that autonomously creates information, it may be more useful to think of the sensing platform as a social machine in which the quality of prediction reflects a shared endeavor. The ability to accurately predict a marker or phenotype depends upon using data, passively and actively collected, from many other individuals who have come before. Returning labeled data back into the system can improve the accuracy for that individual user, as well as for all subsequent users, thus harnessing the wisdom of the crowd while contributing to the crowd.

5.5. The Certainty of Uncertainty

The output of any personal sensing system, even under the best of circumstances, will always have some degree of error and uncertainty. This error is always user-facing, affecting the quality of the experience. This raises several questions that can and should be considered from the early stages of research. First, how much uncertainty is acceptable, and how much accuracy is good enough (Lim & Dey 2011)? For example, if a system were designed to detect likely depression among general internal medicine patients, how many false positives would be acceptable to clinic staff? What levels of false negatives would be acceptable to a care system or to patients, and how could the effect of false negatives be mitigated? In addition, error can be shifted between false positives and false negatives, depending on where it can best be managed and produce the least harm. Or perhaps a metric can be used in a way that minimizes the effect of inaccuracy. For example, step counts from activity trackers may be inaccurate; however, to the degree that they are consistent within user, they can be used for day-to-day comparisons. Early stage research can explore the understanding and acceptability of error and uncertainty and how best to mitigate it (Kay et al. 2015).

5.6. Privacy, Ethics, and the Naked Truth

The use of passively collected digital data raises many issues of privacy and security, about which there are disagreements within the community of researchers, and there is also a lack of guidelines

(Shilton & Sayles 2016). We review some general themes and topics that are most relevant to leveraging passively collected data for mental health purposes.

The principle of privacy refers to ensuring that people have choice and control over the use of their own data and, some would argue, that they understand those choices (Shilton 2009). Security refers to the protections put in place to ensure that people's choices are followed. People's agreement to share their data usually revolves around two key concepts: trust and value. Trust refers to the notion that the data will be used appropriately, given a person's wishes and expectations. Value refers to the benefit that is accrued to the user or society based on the use of data.

An important aspect of trust has traditionally been the severance of the identity of the individual from the data provided from that individual, or de-identification. This is challenging because even a few pieces of information, such as sex, zip code, and birthdate, can identify most of the US population (Sweeney 2000). The data collected from devices may pose even greater risks of identification. GPS traces are the most personally identifying type of data; with only 4 spatiotemporal points, 95% of individuals can be identified (de Montjoye et al. 2013). Various methods can help obfuscate location data; however, none of these successfully preserves privacy and retains the usefulness of the data (Brush et al. 2010).

Privacy management needs to give participants as much control over their data as possible (Shilton 2009). Participants should be informed what the data might reveal about them, for how long the data will be used, who will be using it, and why. Data management tools can be designed to help people manage their data, including the abilities to define acceptable use, limit data access, delete data, or revoke consent altogether.

Encouraging greater openness, more transparency, and the development of better methods to share data is desirable for several reasons, including improving the quality of the scientific literature, providing opportunities to replicate findings, and creating tools that are valid, reliable, and generalizable. As standards and best practices evolve to ensure participants' privacy, the field will be best served if those standards place the participant at the center, such that trust can be established by providing clear understanding, choice, and control.

6. POTENTIAL APPLICATIONS OF PERSONAL SENSING

6.1. Integration into Existing Models of Care

A personal mental-health sensing platform with sufficient accuracy could enhance mental health care by helping identify people in need of treatment, accelerating access to treatment, and monitoring functioning during or after treatment. The inability to identify patients in need of treatment is a major failure point in our health-care system. In any given year, nearly 60% of all people with a mental health condition receive no treatment. Our health-care system relies almost entirely on people with mental health conditions presenting themselves for treatment. Thus, accessing care in a timely manner relies primarily on the patient, whose condition may involve a loss of motivation, stigmatization, a sense of hopelessness and helplessness, and, in some cases, impaired judgment, all of which may interfere with seeking help.

Although personal mental-health sensing holds great promise for monitoring at-risk populations to deliver care more rapidly and effectively, developing accurate algorithms alone will not solve these problems. This will require user-centric approaches to privacy and control, as well as providing sufficient value to all end users (patients and providers) to promote use. Such systems will also likely present situations for which there are no care guidelines. For example, if a mental-health sensing system forecasts that a bipolar patient has a high likelihood of having a relapse during the next 2 weeks, would clinic staff know what to do (Mayora et al. 2013)? Thus, the ability

to detect potential mental health problems opens enormous potential to improve access to care, but the solution will require considerable clinical and design research beyond the personal sensing described in this article.

6.2. Behavioral Intervention Technologies

Behavioral intervention technologies for mental health, such as websites or mobile apps, consist largely of psychoeducational content (as text or video) and interactive tools. The development of sensing capacity for health behaviors, such as physical activity or sleep, has resulted in apps that are less reliant on patients to log activities, presumably making apps easier to use and, therefore, more effective. As attractive as that may sound, here too, previous efforts have demonstrated many unknowns beyond personal sensing (Burns et al. 2011). Although research has discovered a lot about behavioral and environmental factors that contribute to mental illness, it has produced little granular knowledge about the wishes, goals, challenges, and aspirations of people on a moment-to-moment, hour-to-hour basis. This information is critical to designing the next generation of intervention technologies to fit into the fabric of people's daily lives. Applications have tended to be designed using a top-down approach, trying to get people to do what we think will help them. But the technologies that are adopted and widely used are commonly those that make some aspect of people's lives easier, helping people do or achieve something they are motivated to do. Success will be more likely if what is sensed, and how sensed data are used, speak to the user's personal goals, thus integrating treatment aims with making their goals and tasks easier on a daily basis, and thereby fitting treatment activities into people's common patterns and actions.

6.3. Epidemiology

Databases with genomic, epigenetic and other biological data are being integrated into clinical databases to explore genetic influences on disease. Although there is growing recognition that behavior is a critical factor, behavioral data have traditionally been collected using self-report measures, which provide only a periodic subjective snapshot. Personal sensing platforms can provide a continuous stream of objective data that can be used to explore interactions among behavioral markers, genetic and biological factors, and disorders.

7. SUMMARY AND CONCLUSIONS

A growing cloud of digital exhaust is emitted from our daily activities and actions. Some of these data are produced intentionally, such as through the use of wearables. But much of the data are a by-product of our daily actions captured through our smartphones, computers, purchasing, and the increasingly sensor-enabled objects in our lives. The promise for research into mental health, as well as for clinical care, is enormous. But the challenges are also large and manifold. Although the feasibility of personal sensing for mental health has been demonstrated, enormous challenges remain to move from proof of concept to tools that are useful in broader populations. The ultimate success of personal sensing in mental health will likely depend on the continued engagement of users who supply both passively collected data and some measure of active labeling. This, we believe, will require an infrastructure that is a social machine, sufficiently engaging to users to prevent obsolescence. Creating trust in these systems will require a recognition of the primacy of the user, instantiated by enabling people to understand, control, and own their data. Although the tasks are considerable, the potential benefits are also game-changing. The ability to continuously identify behaviors related to mental health has the potential to transform the delivery of care,

speeding recognition of people who are at risk or in need of treatment, and ushering in a new generation of highly personalized, contextualized, dynamic mobile health (or mHealth) tools that can listen rather than ask, and that seamlessly interact, learn, and grow with users.

SUMMARY POINTS

1. Because we use sensors in our everyday lives, personal sensing offers the potential to measure human behavior continuously, objectively, and with minimal effort from the user.
2. Translating raw sensor data into knowledge can be achieved using a layered, hierarchical approach in which sensor data are converted into features, and features are combined to estimate behaviors, moods, and clinical states.
3. A growing number of studies have found that phone sensor data (e.g., from the GPS, accelerometer, light or microphone) can, using machine learning, provide markers of sleep (e.g., bedtime or waketime, duration), social context (e.g., who is in the vicinity, relationship to in-phone contacts), mood, and stress.
4. Depression and mood states in bipolar disorder have been estimated using a variety of phone sensor data. GPS features measuring entropy and the circadian rhythm of movement have been correlated with depression.
5. Posts on social media (e.g., on Facebook or Twitter) can identify people who are depressed or likely to become depressed.
6. Although the work on phone sensor data has been promising, most studies have been small, on the order of 7–30 participants, who frequently are college students; there is little evidence to support replicability.
7. Machine-learning methods vary, some relying on user-generated labels and others uncovering patterns in unlabeled data. Labeling often improves and helps algorithms adapt to new circumstances. Thus, rather than an autonomous prediction machine, it may be more useful to think of a mental-health sensing platform as a social machine in which the quality of prediction is ensured through a shared endeavor.
8. Research suggests that it is feasible to obtain data from personal sensing using everyday sensors. However, numerous challenges must be overcome before it is viable for clinical deployment.

FUTURE ISSUES

1. Because of the amount of variability coming from differences in hardware, device-usage patterns, lifestyle, and the environment, personal sensing platforms will likely require a large user base to have widespread applicability.
2. Some data, such as those generated from GPS tracking, are impossible to de-identify while retaining utility. Thus, creating trust in these systems among participants will require a recognition of the primacy of the user, instantiated by enabling people to understand, control, and own their data.

3. Personal sensing offers the potential to develop a new class of intervention technologies that can reduce users' burden while creating highly tailored and contextualized interactions.
4. No sensing system will be 100% accurate and, thus, researchers, developers, and users must come to consensus about how much error is acceptable and how to better explain and display error to relevant stakeholders.
5. Improving systems will likely require some action on the part of users, and discovering ways to ensure that actions are directly associated with benefits will likely create more engaging and empowering systems.
6. Personal sensing can help improve screening for disorders and access to treatment, but it will require making advances in infrastructure and ensuring that it is integrated into the workflow, as well as making improvements in the underlying technology and knowledge, and improving algorithmic accuracy.
7. The field of personal sensing will likely continue to experience a tension between what is possible and what is feasible, which is related to a trade-off that occurs between small proof-of-concept studies demonstrating novelty and large studies demonstrating robustness and generalizability.
8. Integrating personal sensing data with clinical and genomic databases will offer the opportunity to deepen our understanding of the relationship between behavior and gene × behavior interactions on health, wellness, and disease.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Mental Health with grants P20MH090318, R01MH100482, and R01MH095753 to D.C.M. and K08MH102336 to S.M.S.

LITERATURE CITED

Abdullah S, Matthews M, Murnane EL, Gay G, Choudhury T. 2014. Towards circadian computing: "Early to bed and early to rise" makes some of us unhealthy and sleep deprived. *Proc. UbiComp '14: 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Seattle, WA*, pp. 673–84. New York: Assoc. Comput. Mach.

Acampora G, Cook DJ, Rashidi P, Vasilakos AV. 2013. A survey on ambient intelligence in healthcare. *Proc. IEEE* 101:2470–94

Adams P, Rabbi M, Rahman T, Matthews M, Voida A, et al. 2014. Towards personal stress informatics: comparing minimally invasive techniques for measuring daily stress in the wild. *Pervasive Health '14: Proc. 8th Int. Conf. Pervasive Comput. Technol. Healthc., Oldenburg, Ger*. Brussels: Inst. Comput. Sci. Social-Inform. Telecom. Eng. **http://eudl.eu/doi/10.4108/icst.pervasivehealth.2014.254959**

Alvarez-Lozano J, Osmani V, Mayora O, Frost M. 2014. *Tell me your apps and I will tell you your mood: correlation of apps usage with Bipolar Disorder State*. Presented at Int. Conf. Pervasive Technol. Relat. Assist. Environ., 7th, Rhodes, Greece

Andrews S, Ellis DA, Shaw H, Piwek L. 2015. Beyond self-report: tools to compare estimated and real-world smartphone use. *PLOS ONE* 10:e0139004

Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. 2016. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. *J. Med. Internet Res.* 18:e72

Ben-Zeev D, Wang R, Abdullah S, Brian R, Scherer EA, et al. 2016. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. *Psychiatr. Serv.* 67:558–61

Bengio Y, Courville A, Vincent P. 2013. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35:1798–828

Berke EM, Choudhury T, Ali S, Rabbi M. 2011. Objective measurement of sociability and activity: mobile sensing in the community. *Ann. Fam. Med.* 9:344–50

Bishop CM. 2006. *Pattern Recognition and Machine Learning*. New York: Springer

Bishop CM, Lasserre J. 2007. Generative or discriminative? Getting the best of both worlds. *Bayesian Stat.* 8:3–23

Brush AJB, Krumm J, Scott C. 2010. Exploring end user preferences for location obfuscation, location-based services, and the value of location. *Proc. UbiComp '10: 2010 ACM Conf. Ubiquitous Comput., Copenhagen, Denmark*, pp. 95–104. New York: Assoc. Comput. Mach.

Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, et al. 2011. Harnessing context sensing to develop a mobile intervention for depression. *J. Med. Internet Res.* 13:e55

Butler D. 2013. When Google got flu wrong. *Nature* 494:155–56

Calvo RA, D'Mello S. 2010. Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affect. Comput.* 1:18–37

Canzian L, Musolesi M. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. *Proc. UbiComp '15: 2015 ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Osaka, Japan*, pp. 1293–304. New York: Assoc. Comput. Mach.

Chapelle O, Schölkopf B, Zien A. 2006. *Semi-Supervised Learning*. Cambridge, MA: MIT Press

Chen ZY, Lin M, Chen FL, Lane ND, Cardone G, et al. 2013. Unobtrusive sleep monitoring using smartphones. *Pervasive Health '13: Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthc., Venice, Italy*, pp. 145–52. Washington, DC: IEEE

Choudhury T, Consolvo S, Harrison B, LaMarca A, LeGrand L, et al. 2008. The mobile sensing platform: an embedded activity recognition system. *IEEE Pervasive Comput.* 7:32–41

Ciman M, Wac K, Gaggi O. 2015. Assessing stress through human-smartphone interaction analysis. *Pervasive Health '15: Proc. 9th Int. Conf. Pervasive Comput. Technol. Healthc., Istanbul.* Brussels: Inst. Comput. Sci. Social-Inform. Telecom. Eng. **http://ieeexplore.ieee.org/document/7349382/**

Collins FS, Varmus H. 2015. A new initiative on precision medicine. *N. Engl. J. Med.* 372:793–95

De Choudhury M, Counts S, Horvitz E. 2013a. Social media as a measurement tool of depression in populations. *WebSci '13: Proc. 5th Annu. ACM Web Sci. Conf., Paris*, pp. 47–56. New York: Assoc. Comput. Mach.

De Choudhury M, Gamon M, Counts S, Horvitz E. 2013b. Predicting depression via social media. *Proc. 7th. Int. AAAI Conf. Weblogs Social Media, Boston*, pp. 128–37. Palo Alto, CA: Assoc. Adv. Artif. Intell.

De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. 2016. Discovering shifts to suicidal ideation from mental health content in social media. *CHI'16: Proc. 2016 CHI Conf. Human Factors Comput. Systems, San Jose, CA*, pp. 2098–110. New York: Assoc. Comput. Mach.

de Montjoye Y-A, Hidalgo CA, Verleysen M, Blondel VD. 2013. Unique in the crowd: the privacy bounds of human mobility. *Sci. Rep.* 3:1376

Eadicicco L. 2015. Americans check their phones 8 billion times a day. *Time*, Dec. 15. **http://time.com/4147614/smartphone-usage-us-2015/**

Eagle N, Pentland A. 2006. Reality mining: sensing complex social systems. *Pers. Ubiquitous Comput.* 10:255–68

Eagle N, Pentland A, Lazer D. 2009. Inferring friendship network structure by using mobile phone data. *PNAS* 106:15274–78

Edison T, Essa I, Abowd G. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. *Proc. UbiComp '15: 2015 ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Osaka, Jpn.*, pp. 1029–40. New York: Assoc. Comput. Mach.

Epp C, Lippold M, Mandryk RL. 2011. Identifying emotional states using keystroke dynamics. *CHI'11: Proc. SIGCHI Conf. Hum. Factors Comput. Sys.*, *Vancouver*, *Can.*, pp. 715–24. New York: Assoc. Comput. Mach.

Fortier I, Doiron D, Burton P, Raina P. 2011. Consolidating data harmonization—how to obtain quality and applicability? *Am. J. Epidemiol.* 174:261–64; author reply 65–66

Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, et al. 2015. Mobile phones as medical devices in mental disorder treatment: an overview. *J. Pers. Ubiquitous Comput.* 19:335–53

Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, et al. 2015. Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE J. Biomed. Health Inform*. 19:140–48

Grünerbl A, Osmani V, Bahle G, Carrasco JC, Oehler S, et al. 2014. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. *AH'14: Proc. 5th Augment. Hum. Int. Conf.*, *Kobe*, *Jpn.*, *Mar. 7–9*, Artic. 38. New York: Assoc. Comput. Mach.

Gu W, Yang Z, Shangguan L, Sun W, Jin K, Liu Y. 2014. Intelligent sleep stage mining service with smartphones. *Proc. UbiComp '14: 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, *Seattle*, *WA*, pp. 649–60. New York: Assoc. Comput. Mach.

Hames JL, Hagan CR, Joiner TE. 2013. Interpersonal processes in depression. *Annu. Rev. Clin. Psychol.* 9:355–77

Hao T, Xing G, Zhou G. 2013. iSleep: unobtrusive sleep quality monitoring using smartphones. *SenSys'13: Proc. 11th ACM Conf. Embed. Netw. Sens. Syst.*, *Rome*, Artic. 4. New York: Assoc. Comput. Mach.

Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, et al. 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Proc. Mag.* 29:82–97

Hsieh H-P, Li C-T. 2014. Inferring social relationships from mobile sensor data. *WWW '14 Companion: Proc. 23rd Int. Conf. World Wide Web*, *Seoul*, *Korea*, pp. 293–94. New York: Assoc. Comput. Mach.

Intille SS. 2013. Closing the evaluation gap in UbiHealth Research. *IEEE Pervasive Comput*. 12:76–79

Jain SH, Powers BW, Hawkins JB, Brownstein JS. 2015. The digital phenotype. *Nat. Biotechnol.* 33:462–63

Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, et al. 2014. Tracking suicide risk factors through Twitter in the US. *Crisis* 35:51–59

Kalantarian H, Alshurafa N, Le T, Sarrafzadeh M. 2015. Monitoring eating habits using a piezoelectric sensor-based necklace. *Comput. Biol. Med.* 58:46–55

Kay M, Patel SN, Kientz JA. 2015. How good is 85%? A survey tool to connect classifier evaluation to acceptability of accuracy. *CHI'15: Proc. 2016 CHI Conf. Human Factors Comput. Syst.*, *Seoul*, *Korea*, pp. 347–56. New York: Assoc. Comput. Mach.

Klasnja P, Consolvo S, McDonald DW, Landay JA, Pratt W. 2009. Using mobile & personal sensing technologies to support health behavior change in everyday life: lessons learned. *AMIA Annu. Symp. Proc.* 2009:338–42

Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional neural networks. *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, *Lake Taboe*, *NV*, pp. 1097–105. Red Hook, NY: Curran Assoc.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521:436–44

Li I, Dey A, Forlizzi J. 2010. A stage-based model of personal informatics systems. *CHI'10: Proc. 2016 CHI Conf. Human Factors Comput. Syst.*, *Atlanta*, pp. 557–66. New York: Assoc. Comput. Mach.

Li Q, Stankovic JA, Hanson MA, Barth AT, Lach J, Zhou G. 2009. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. *Proc. Sixth Int. Workshop Wearable Implant. Body Sens. Netw.*, pp. 138–43. Washington, DC: IEEE

LiKamWa R, Liu Y, Lane ND, Zhong L. 2013. MoodScope: building a mood sensor from smartphone usage patterns. *MobiSys '13: Proc. 11th Annu. Int. Conf. Mob. Syst. Appl. Serv.*, *Taipei*, *Taiwan*, pp. 389–402. New York: Assoc. Comput. Mach.

Lim BY, Dey AK. 2011. Investigating intelligibility for uncertain context-aware applications. *Proc. UbiComp '11: 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, *Beijing*, pp. 415–24. New York: Assoc. Comput. Mach.

Lu H, Rabbi M, Chittaranjan GT, Frauendorfer D, Mast MS, et al. 2012. Detecting stress in unconstrained acoustic environments using smartphones. *Proc. UbiComp '12: 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, *Pittsburgh*, *PA*, pp. 351–60. New York: Assoc. Comput. Mach.

Ma Y, Xu B, Bai Y, Sun G, Zhu H. 2012. Daily mood assessment based on mobile phone sensing. *Proc. 2012 9th Int. Conference Wearable Implant. Body Sens. Netw.*, *London*, pp. 142–47. Washington, DC: IEEE

Madan A, Cebrian M, Lazer D, Pentland A. 2010. Social sensing for epidemiological behavior change. *Proc. UbiComp '10: 2010 ACM Conf. Ubiquitous Comput.*, *Copenhagen*, *Den.*, pp. 291–300. New York: Assoc. Comput. Mach

Maehr W. 2008. *eMotion: Estimation of User's Emotional State by Mouse Motions*. Saarbrücken, Ger.: VDM

Mayora O, Arnrich B, Bardram J, Drager C, Finke A, et al. 2013. Personal health systems for bipolar disorder: anecdotes, challenges and lessons learnt from MONARCA Project. *Pervasive Health '13: Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthc.*, *Venice*, *Italy*, pp. 424–29. Washington, DC: IEEE

Min J-M, Doryab A, Wiese J, Amini S, Zimmerman J, Hong JI. 2014. Toss 'n' turn: smartphone as sleep and sleep quality detector. *CHI'10: Proc. 2016 CHI Conf. Human Factors Comput. Syst.*, *Toronto*, *Can.*, pp. 477–86. New York: Assoc. Comput. Mach.

Min J-M, Wiese J, Hong J, Zimmerman J. 2013. Mining smartphone data to classify life-facets of social relationships. *CSCW'13: Proc. 2013 Conf. Comput.-Support. Coop. Work Soc. Comput.*, *San Antonio*, *TX*, pp. 285–94. New York: Assoc. Comput. Mach.

Miotto R, Li L, Kidd BA, Dudley JT. 2016. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.* 6:26094

Murnane EL, Abdullah S, Matthews M, Choudhury T, Gay G. 2015. Social (media) jet lab: how usage of social technology can modulate and reflect circadian rhythms. *Proc. UbiComp '15: 2015 ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, *Osaka*, *Jpn.*, pp. 843–54. New York: Assoc. Comput. Mach.

Ng AY, Jordan MI. 2002. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. *NIPS'01: Proc. 14th Int. Conf. Neural Inf. Process. Syst. Nat. Synth.*, *Vancouver*, *Can.*, pp. 841–48. Cambridge, MA: MIT Press

Oxman TE, Rosenberg SD, Tucker GJ. 1982. The language of paranoia. *Am. J. Psychiatry* 139:275–82

Palmius N, Tsanas A, Saunders K, Bilderbeck AC, Geddes JR, et al. 2016. Detecting bipolar depression from geographic location data. *IEEE Trans. Biomed. Eng.* PP(99). **https://doi.org/ 10.1109/TBME.2016.2611862**

Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, et al. 2015. Automatic personality assessment through social media language. *J. Pers. Soc. Psychol.* 108:934–52

Perrin A. 2015. *Social media usage: 2005–2015*. Rep., Pew Research Center, Washington, DC. **http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/**

Pham H, Shahabi C, Liu Y. 2013. EBM: an entropy-based model to infer social strength from spatiotemporal data. *SIGNMOD '13: Proc. 2013 ACM SIGMOD Int. Conf. Manag. Data*, *New York*, pp. 265–76. New York: Assoc. Comput. Mach.

Picard RW, Fedor S, Ayzenberg Y. 2016. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emot. Rev.* 8:62–75

Piwek L, Ellis DA, Andrews S, Joinson A. 2016. The rise of consumer health wearables: promises and barriers. *PLOS Med.* 13:e1001953

Poushter J. 2016. *Smartphone ownership and Internet usage continues to climb in emerging economies*. Rep., Pew Research Center, Washington, DC. **http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/**

Rachuri KK, Musolesi M, Mascolo C, Rentfrow PJ, Longworth C, Aucinas A. 2010. Emotion sense: a mobile phones based adaptive platform for experimental social psychology research. *Proc. UbiComp '10: 2010 ACM Conf. Ubiquitous Comput.*, *Copenhagen*, *Denmark*, pp. 281–90. New York: Assoc. Comput. Mach.

Rahman T, Adams AT, Zhang M, Cherry E, Zhou B, et al. 2014. BodyBeat: a mobile system for sensing non-speech body sounds. *MobiSys '14: Proc. 12th Annu. Int. Conf. Mob. Syst. Appl. Serv.*, *Bretton Woods*, *NH*, pp. 2–13. New York: Assoc. Comput. Mach.

Ricker T. 2015. Wearables are booming but is anyone still wearing them? *The Verge*, Dec. 4. **http://www.theverge.com/2015/12/4/9848394/wearable-abandonment**

Rude SS, Gortner EM, Pennebaker JW. 2004. Language use of depressed and depression-vulnerable college students. *Cogn. Emot.* 18:1121–33

Saeb S, Lattie E, Schueller SM, Kording K, Mohr DC. 2016. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ* 4:e2537

Saeb S, Lonini L, Kording K, Mohr DC. 2016. Voodoo machine learning for clinical predictions. bioRxiv 059774. **https://doi.org/10.1101/059774**

Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, et al. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J. Med. Internet Res.* 17:e175

Schmidhuber J. 2015. Deep learning in neural networks: an overview. *Neural Netw.* 61:85–117

Schwartz HA, Eichstaedt J, Kern M, Park G, Sap M, et al. 2014. Towards assessing changes in degree of depression through Facebook. *Proc. Workshop Comput. Linguist. Clin. Psychol. Linguist. Signal Clin. Real., Baltimore, MD*, pp. 118–25. Stroudsburg, PA: Assoc. Comput. Linguist.

Sen S, Subbaraju V, Misra A, Balan RK, Lee Y. 2015. The case for smartwatch-based diet monitoring. *Proc. 2015 IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops), St. Louis, MO.*, pp. 585–90. Washington, DC: IEEE

Settles B. 2010. *Active learning literature survey*. Comput. Sci. Tech. Rep. 1648, Univ. Wisconsin, Madison. **http://burrsettles.com/pub/settles.activelearning.pdf**

Shilton K. 2009. Four billion little brothers? Privacy, mobile phones, and ubiquitous data collection. *Commun. ACM* 52:48–53

Shilton K, Sayles S. 2016. "We aren't all going to be on the same page about ethics": ethical practices and challenges in research on digital and social media. *Proc. 2016 49th Hawaii Int. Conf. Syst. Sci., Honolulu, HI*, pp. 1909–18. Washington, DC: IEEE

Sivertsen B, Krokstad S, Overland S, Mykletun A. 2009. The epidemiology of insomnia: associations with physical and mental health. The HUNT-2 study. *J. Psychosom. Res.* 67:109–16

Spenkelink CD, Hutten MM, Hermens HJ, Greitemann BO. 2002. Assessment of activities of daily living with an ambulatory monitoring system: a comparative study in patients with chronic low back pain and nonsymptomatic controls. *Clin. Rehabil.* 16:16–26

Sun D, Paredes P, Canny J. 2014. MouStress: detecting stress from mouse motion. *CHI'14: Proc. SIGCHI Conf. Human Factors Comput. Syst., Toronto, Can.*, pp. 61–70. New York: Assoc. Comput. Mach.

Sutskever I, Vinyals O, Le QV. 2014. Sequence to sequence learning with neural networks. *NIPS'14: Proc. 27th Int. Conf. Neural Inf. Process. Syst., Montreal, Can.*, pp. 3104–12. Cambridge, MA: MIT Press

Sweeney L. 2000. *Simple demographics often identify people uniquely*. Data Priv. Work. Pap. 3, Carnegie Mellon Univ., Pittsburgh, PA

Taylor DJ, Lichstein KL, Durrence HH, Reidel BW, Bush AJ. 2005. Epidemiology of insomnia, depression, and anxiety. *Sleep* 28:1457–64

Torous J, Kiang MV, Lorme J, Onnela JP. 2016. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health* 3:e16

Vallance JK, Winkler EA, Gardiner PA, Healy GN, Lynch BM, Owen N. 2011. Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005–2006). *Prev. Med.* 53:284–88

Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. 2010. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* 11:3371–408

Wang D, Pedreschi D, Song C, Giannotti F, Barabási A-L. 2011. Human mobility, social ties and link prediction. *KDD '17: Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., San Diego, CA*, pp. 1100–8. New York: Assoc. Comput. Mach.

Wang R, Chen FL, Chen Z, Li TX, Farari G, et al. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proc. UbiComp '14: 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Seattle, WA*, pp. 3–14. New York: Assoc. Comput. Mach.

Wang R, Aung MSH, Abdullah S, Brian R, Campbell AT, et al. 2016. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. *Proc. UbiComp '16: 2016 ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Heidelberg, Ger.* pp. 886–97. New York: Assoc. Comput. Mach.

Wiese J, Hong JI, Zimmerman J. 2014. Challenges and opportunites in data mining contact lists for inferring relationships. *Proc. UbiComp '14: 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Seattle, WA, Sept. 13–17*, pp. 643–647. New York: Assoc. Comput. Mach.

Wiese J, Min J-M, Hong J, Zimmerman J. 2015. "You never call, you never write": call and SMS logs do not always indicate tie strength. *CSCW'15: Proc. 18th ACM Conf. Comput. Support. Coop. Work Soc. Comput., Vancouver, Can.*, pp. 765–74. New York: Assoc. Comput. Mach.

Wiskott L, Sejnowski TJ. 2002. Slow feature analysis: unsupervised learning of invariances. *Neural Comput.* 14:715–70

Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, et al. 2015. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347:1254806

Yatani K, Truong KN. 2012. BodyScope: a wearable acoustic sensor for activity recognition. *Proc. UbiComp '12: 2012 ACM Conf. Ubiquitous Comput., Pittsburgh, PA*, pp. 341–50. New York: Assoc. Comput. Mach.

Zhenyu C, Lane N, Cardone G, Lin M, Choudhury T, Campbell A. 2013. Unobtrusive sleep monitoring using smartphones. *Pervasive Health '13: Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthc., Venice, Italy*, pp. 145–52. Washington, DC: IEEE

Zhu X, Goldberg AB. 2009. *Introduction to Semi-Supervised Learning*. San Rafael, CA: Morgan & Claypool