Identification, Prediction, and Intervention Via Remote Digital Technology: Digital Phenotyping & Deployment of Clinical Interventions in Psychiatry

Lior Carmi[1], Anzar Abbas[2], Katharina Schultebraucks[3], Isaac R. Galatzer-Levy[4,5]

[1] Post Trauma Center, Chaim Sheba Medical Center, Ramat Gan, Israel
[2] AiCure, New York, NY
[3] Emergency Medicine, Columbia University, New York, NY
[4] Reality Labs, Facebook, New York, NY
[5] Psychiatry, New York University School of Medicine, New York, NY

Abstract

The extensive use of smart technology (smartphones & wearables) and the vast amount of information they contain, has positioned remote devices and technology as a massive database for behavioral, personal, and social day–to–day activities. Harnessing smart devices into the clinical field has introduced new, real-time, data sources that hold promise in characterizing clinical functioning and intervening remotely on a scale and timeframe that would have been unimaginable a few years ago. This promise is beginning to come to fruition as both digital technology and the underlying data models to use the massive amounts of data they collect rapidly advance.

Remote characterization of clinical populations (known as digital phenotyping) and subsequent digital methods of intervention are highly relevant in psychiatry where behavioral patterns and changes in these patterns often characterize prediction or deterioration in each disorder. Specifically, several clinically mental situations would be prevented and better understood, by employing a digital personalized model which is capable to predict when certain deviations from a patient's usual behavior may lead, with high probability, to his or her health deterioration. While such methods hold promise, significant work is needed to understand clinical risk based on digital signals and to develop coordinated logistical systems to deploy useful interventions.

1 – The importance of measurement

Reliable measurement of a patient's health is central to medical care and scientific research. Fortunately, clinicians have access to tools that allow them to diagnose disease and assess symptom severity with great accuracy—tools like blood tests, biopsies, MRI scans, ultra-sounds, and gene sequencing—all of which provide objective measures of physical health. However, in the context of mental and behavioral health, measurement is far from straightforward.

Psychiatrists rely on rater-based scales or patient-report questionnaires to diagnose disease and assess symptom severity. These tools have well-established shortcomings, both in their clinical validity and the logistics associated with their use (Cuthbert, 2015). Hence, in the context of medical care and scientific research in mental health, clinicians must not only focus on treatment, but also on the measurement tools used to assess treatment. In this chapter, we discuss measurement challenges in psychiatry and the emergence of digital phenotyping as a solution for objective measurement of disease severity.

2 – The challenge of measurement in mental health

Measurement has historically been a recognized challenge in psychiatry. Before efforts such as the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) by the American Psychological Association or the *International Classification of Diseases* (ICD) by the World Health Organization, the same patient could get two different diagnoses from two different clinicians on the same day with no overlap in recorded symptomatology: Diagnosing psychiatric patients was an entirely subjective exercise.

But with the DSM or the ICD, a standardization in symptom classification and measurement emerged (American Psychiatric Association, 2013; World Health Organization, 2020). This allowed for the establishment of standardized scales for the assessment of disease severity, typically involving semi-structured interviews for observation of classified symptomatology. Examples of this include the HAM-D for depression (Willams, 1988), PANSS for schizophrenia (Kay et al., 1987), or the CARS for autism spectrum disorder (Schopler et al., 1980). It also led to the creation of patient-report scales, where a patient responds to questions, answers to which are meant to be indicative of disease severity. Examples of self-report scales include the BPRS for schizophrenia (Overall and Gorham, 1962), ASRS for attention-deficit/hyperactivity disorder (Kessler et al., 2005), or the PHQ-9 for depression (Kroenke et al., 2001). These scales have become the gold standard for the measurement of neuropsychiatric illness.

Yet, the question of subjectivity remains. Despite the progress made in the standardization of measurement, both rater-based and patient-report scales are prone to varying degrees of observer and patient bias (Fuchs, 2010). Moreover, considering recent developments in neuroscience, the symptom classification efforts themselves are grounded in outdated disease nosology (Cuthbert and Insel, 2010). There is a need for increased objectivity in assessment of mental health, leading to measurement that more accurately reflects patient functioning (Pallagrosi et al., 2016). Indeed, a core motive behind this book is to highlight the clinical shortcomings of existing measurement tools in mental health.

In addition to clinical challenges, current assessment tools present practical and logistical barriers (van Eijk, 2020). Every time a rater-based assessment is needed, the patient most likely has to appear in the clinic in person. They may have to take time off work or set aside part of their weekend. In many cases, they must depend on public transportation or drive long distances to get to the clinic. Then, in what may be an unfamiliar environment, they are interviewed by a clinician they may or may not feel comfortable with, who probes into their health and psychiatric functioning as part of assessments that can at times take as long as 90 minutes to administer. The experience can be no less frustrating when the patient is asked to come into the clinic to participate in paper-based patient-report questionnaires. Not only do such experiences lead to measurements not reflective of a patient’s day-to-day functioning (Schmuckler, 2001), the burden associated—both for the patient and the clinician—renders assessment of mental health impractical unless conducted far and few between, with clinicians having little to no visibility into patient health and behavior the moment they step outside the clinic.

3 – Virtual care and electronic patient self-report

A natural solution to the logistical challenges associated with clinical assessment has been the use of technology to conduct them virtually. Indeed, remote patient self-report is the most intuitive consequence of this (Coons et al., 2015). Self-report scales can easily be digitized and participated in electronically. This has even led to the emergence of ecological momentary assessments, a departure from the rigidity of some traditional self-report scales (Burke et al., 2017). Initial concerns regarding the validity of remote self-report in comparison to it being conducted in person have alleviated over time, with a plethora of studies confirming their accuracy and reliability (Löwe et al., 2004; Cavelti et al., 2012; Areán et al., 2016), not to mention their widespread use in clinical research (Moskowitz et al., 2006; Shiffman et al., 2008). The chapter on ecological momentary assessments for digital phenotyping in Section I of this book describes these developments in detail.

More recently, with the broader adoption of telemedicine and virtual care, clinician-administered psychiatric assessments are also being conducted remotely (Barnett and Huskamp, 2020; LaFrance et al., 2020, Wright, 2020). Some traditional scales are even being adopted so they can be conducted over video calls (Dorsey et al., 2020). The adoption of virtual care and its impact on measurement is further discussed in the chapter on telepsychiatry earlier in Section II of this book. Through the digitization of self-report and virtual care, practical and logistical challenges associated with clinical assessment are being addressed. However, the clinical shortcomings stemming from the still-subjective nature of these assessments remain. And with technology becoming a core component of clinical assessment, its potential to enable objective measurement is becoming more apparent—and digital phenotyping is coming into play.

4 – Digital phenotyping of mental health

Mental illness manifests itself in observable ways. The aim of an assessment tool is to measure such observable behavior in a standardized manner. Clinicians do just that during clinical assessments, i.e. they observe aspects of the patient's behavior, including their facial expressivity, characteristics of their speech, acoustics of their voice, patterns of movement, and the manner in which they respond to questions and stimuli. They then use these behavioral characteristics to make judgments on the patient's health and clinical functioning. Advancements in machine learning, supported by computer vision, natural language processing, and digital signal processing tools have led to tools that allow for the quantification of the same behavioral characteristics (Goyal et al., 2017; Hardeniya et al., 2016; Boersma & Weenink, 2018). Similarly, these behavioral characteristics can be used to measure health, forming the foundation for the field of digital phenotyping.

Digital phenotyping of mental health proposes solutions to many of the challenges presented by traditional clinical assessment tools (Insel, 2017). Given a clinician needn't directly be involved in the observation of behavior i.e. the collection of data, the patient may participate in the assessment on their own time and in their natural environment, removing many of the logistical barriers associated with clinical assessment. Most importantly, digital assessment of behavior leads to objective and sensitive measurements, removing clinician and patient bias from the assessment.

Digital medicine is based on four major pillars:

1. Continuous Passive Monitoring (CPM) of behavioral parameters or Smart Active Monitoring (SAM)
2. Identification of behavioral patterns that will lead to an Individualized Digital Phenotype (IDP) of a disorder.
3. Accurate detection of clinically relevant changes and accordingly Timely Precise Intervention (TPI), e.g., secondary prevention.
4. Optimized communication for assessment and intervention.

CPM is efficiently measuring behavioral patterns manifested through mobile phone usage. The information to be measured may consist of: Communication patterns, Activity patterns, Diurnal variation, and thus sleep changes. Active data sources such as SAM may be consisted of voice prosody, facial & eye coding, linguistic analysis, remote survey etc.

The remainder of this chapter provides an overview of recent developments in digital phenotyping, focusing primarily on ones that use consumer devices and hence do not require specialized hardware. There have been several attempts to classify digital phenotyping tools and the biomarkers they measure (Corovas et al., 2019). Here, we categorize digital phenotyping methods into those that rely on passive monitoring and those that require active assessments. Digital phenotyping efforts that depend on non-consumer devices have been omitted, including application of machine learning to datasets from medical imaging, electronic medical records, and genetic sequencing data (Abbas et al., 2021). Patient self-report, though indeed a form of digital phenotyping, has been mostly omitted from this chapter given detailed discussion of its merits and shortcomings in other chapters of this book.

4.1 – Passive monitoring

Digital phenotyping through passive monitoring involves recording of an individual's behavior while they go about their days as they would regularly. It is based on the notion that certain passively observable behavioral characteristics can serve as proxies of health. The most popular form of passive monitoring is measurement of actigraphy through wearables (Piau et al., 2019). This section also discusses how monitoring of an individual's electronic behavior can provide useful behavioral measures. There has been use of specially designed in-home sensors to measure patient behavior and consequently health (e.g., in Adib et al., 2015). However, since these devices are not necessarily consumer-grade or meant for broad adoption, a discussion on in-home sensors has been omitted from this chapter. Passively collected measures through digital phenotyping are a relatively novel area of research, partially due to the recent popularity of health-focused consumer devices and accessible data on individuals' online behavior. Yet, they have demonstrated marketed success as effective measurement tools.

Significant efforts have been made both in industry and academia, to develop novel digital phenotypes for remote monitoring health status including mental health. Areas of research interest include, the use of GPS to characterize behavioral activation and avoidance (Glenn & Monteith, 2014; John Torous, Staples, & Onnela, 2015), monitoring of physiology and sleep through smartphones and wearables, e.g., Fitbit, Garmin, or Apple watches (Onnela, Keshavan, Staples, Barnett, & Torous, 2018), and passive measurement of cognitive functioning based on keystroke activity, taps, and swipes via smartphone apps (Dagum, 2018). Such approaches show promise but require significant infrastructure investment to capture data, to process high data volumes, to develop theoretical and machine learning models to map digital signals to behavioral phenotypes, and finally to communicate results in an efficient and comprehensible manner. While significant research and development in the area of digital phenotyping has occurred in academia, attempts to fully develop and maintain technologies has occurred primarily in industry with both startups and major technology and insurance companies (Google/Verily, Amazon, Kaiser Permanente) all making large investments in digital phenotyping technologies.

4.1.1 – Actigraphy and tremor

Actigraphy, a term generally used to refer to movement activity measured through a wearable device (e.g. smartwatches, accelerometers, pedometers) has gained popularity alongside the devices that offer the measurements (Wright et al., 2017). The use of this technology is perhaps most promising for direct measures of fine motor behavior such as tremor and gait that are difficult to assess simply through observation, whether it be through purpose-built sensors (Jeon et al., 2017) or commercially available devices (Lamont et al., 2018). However, movement abnormalities form part of the symptomatology of a range of neuropsychiatric disorders, such as in individuals with depression (Santomas, 2020) or schizophrenia (Shin et al., 2016). Hence, actigraphy has been utilized as a proxy measure of overall movement behavior in a wide range of patient populations (Wright et al., 2017; Depp et al., 2019).

4.1.2 – Human-Computer Interaction

Given the integration of technology into most aspects of daily living, human-computer interaction could refer to any aspect of how a patient interacts with their devices or online. Intuitively, the concept behind using human-computer interactions is that such activity reflects multiple aspects of functioning that are clinically relevant in psychiatry such as motor functioning and cognition reflected in typing behavior to social functioning reflected in social media activity. Keystroke activity-based biomarkers use passively collected keyboard activity, whether on mobile phones or computer keyboards as correlates of mental health (Epp et al., 2011; Zululeta et al., 2018). Efforts to use social media activity overlap with efforts to understand natural language (also discussed in Section 4.2.2), utilizing an individual's online behavior, including the posts they generate and the posts they interact with, as indicators of their health and functioning (Coppersmith et al., 2014; McClellan et al., 2017). Though measurements derived from an individual's electronic behavior have shown promise as measures of health—particularly in the context social and cognitive functioning—the question of how they can be integrated into patient care and clinical research remains unclear.

4.2 – Active assessments

In contrast to passively acquired data, active assessments ask individuals to engage in pre-designed tasks or interactions that collect short bursts of data on their behavior, which can then be used to derive measures of health. These are closest to traditional clinical assessments in that they are meant to elicit behavior for targeted measurement of health, rather than deriving inferences from passive monitoring. They are different from traditional clinical assessments in that the collected data is used to objectively quantify behavioral characteristics to derive measures of health.

4.2.1 – Facial expressivity

Facial expressivity is an important measure during assessments of psychiatric functioning (e.g. passive assessment of a depressed patient's emotional experience or active evaluation of blunted affect in an individual with schizophrenia). Recognizing the subjectivity of such observation, efforts to standardize facial measurements date back decades (Ekman & Friesen, 1978). The Facial Action Coding System, which catalogues all possible combinations of facial musculature arrangements, formed the foundation for objective labeling of facial activity and subsequently emotional expressivity (Ekman, 1997). Though facial coding in this manner showed strong direct relationships between facial expressivity and psychiatric functioning (such as in Cohn et al., 2009), manually coding facial activity is not scalable as a clinical measure given the effort required.

With advancements in computer vision, the same coding of facial activity can be conducted except using automated tools that are openly available to researchers (Baltrušaitis et al., 2016; Baltrušaitis et al., 2018). As a result, measurement of facial expressivity to assess psychiatric health and functioning has become simple to integrate into clinical research. Several studies have used these methods in the laboratory to demonstrate the relationship between facial expressivity as measured through computer vision with symptom severity across psychiatric disorders (Jiang et al., 2020; Haque et al., 2018; Corcoran & Cecchi, 2018). More recent efforts have built smartphone-based platforms to do so in the real world in patients receiving treatment (Galatzer-Levy et al., 2020a; Galatzer-Levy et al., 2020b). If the collection of video of patient video can be made scalable and secure, computer vision-based measurement of facial expressivity can be a valuable assessment of psychiatric functioning.

4.2.2 – Voice and speech

Similar to facial expressivity, assessing a patient's speech is a critical part of clinical assessment: The clinician makes observations not only on what the patient is saying, but also *how* they are saying it. They then use their judgement as clinicians to respond to items in the clinical assessment that may refer to the patient's speech, such as verbal fluency or social withdrawal (Kay et al., 1987). Yet, this measurement is subjective and efforts have long been under way in laboratory research to standardize analysis of voice (Oğuz et al., 2011). Consequently, a field of vocal acoustics has emerged which uses techniques in digital signal processing to identify features of a voice's waveform that are related to the speaker's health (Godino-Llorente et al., 2008; Jadoul et al., 2018). As a result, several acoustic properties of voice have been related to psychiatric functioning (Hashim et al., 2017; Parola et al., 2020), including measures as simple as the loudness or fundamental frequency of voice (Quatieri & Malyska, 2012) to properties such as the harmonics-to-noise ratio (Shama et al., 2006) and normalized amplitude quotient (Airas & Alku, 2006).

In addition to the acoustics of an individual's voice, recent advancements in machine learning tools have allowed for widespread adoption of natural language processing in clinical research (Chowdhary, 2020). With these tools, the analysis can focus on *what* the patient is saying by transcribing their speech or analyzing written text to automatically analyze characteristics of language that are indicative of psychiatric functioning (Althoff et al., 2016; Cook et al., 2016; Stewart & Velupillai, 2020). This includes simple measurements like the lengths of pauses between words to complex language characteristics such as emotional valence of speech, lexical diversity, and deriving cognitive measures based on speech (Patel et al., 2015; He et al., 2017). Several efforts have utilized vocal and speech measures to quantify disease severity in the context of medical care and clinical research and have built software platforms to collect such data from patients in a scalable manner (Komeili et al., 2019). If the collection of such data, which is still considered Protected Health information when collected to assess patient health, can be collected securely in a scalable manner, digital measurements of voice and speech can contribute significantly to efforts in digital phenotyping.

4.2.3 – Movement behavior

Measurement of movement through actigraphy is discussed in the earlier section on passive monitoring of behavior. However, measurements of movement through active assessments expands the data through which motor functioning can be assessed. Primarily, it refers to the use of computer vision to more directly measure an individual's movement activity during active assessments, similar to how motor functioning would be assessed as part of a traditional clinical assessment (Goetz et al., 2008).

As with actigraphy, one of the primary reasons to utilize such technology is for more sensitive quantification of motor abnormalities. Traditional clinical assessments such as the UPDRS (Goetz et al., 2008) or the TETRAS (Elble, 2016) classify tremor into discrete scores, given that is all that is possible with rater observation. In comparison, when computer vision is used on videos of patients performing similar assessments, the quantification of tremor can be made using a continuous measure and by definition be a more sensitive assessment in addition to not being reliant on subjective clinician observation (Williams et al., 2019; Pang et al., 2020; Nieto-Hidalgo et al., 2018).

Moving beyond tremor, computer vision based measurement of movement can identify other clinically meaningful aspects of an individual's behavior. Head movement has been shown to reflect motor retardation and in some cases overall disease severity in psychiatric populations (Abbas et al., 2020). If oculomotor activity is considered as an aspect of movement, then eye gaze directionality, eye blink behavior, pupil dilation, etc. can serve as important measures of psychiatric functioning. Examples of this include measuring saccades for assessment of schizophrenia (Huang et al., 2020), blink durations for assessment of fatigue (Wang et al., 2017), pupil dilations for assessments of attention and arousal (Miller et al., 2019), and eye gaze directionality in response to pre-designed stimuli designed to measure aspects of social and attentional functioning (Hashemi et al., 2012).

4.2.4 – Cognitive functioning

Digital measurement of cognitive functioning requires a special mention as it is perhaps farthest ahead in its adoption in clinical research and medical care, with several commercial efforts making it easily accessible (Kaser et al., 2017). This is partially due to the fact that traditional paper-based cognitive assessments, similar to patient report questionnaires, have been relatively simple to digitize and be performed over computers, tablets, and smartphones (Lancaster et al., 2019; Au et al., 2017; Hafiz et al., 2019). In fact, the experience of cognitive assessments has in some cases been enhanced significantly by the creativity possible with digitization that was not the case with traditional paper-based assessments.

In addition to performance-based cognitive testing, some of the digital measurements discussed above have also been either shown to be correlates of cognition or in themselves are directly reflective of cognitive functioning. The most common example of this has been the use of text or speech data to extract characteristics of language that are indicative of cognitive functioning

(Thapa et al., 2020; Clarke et al., 2020). In fact, studies using natural language processing to extract speech characteristics have shown that speech can be indicative of cognitive decline in individuals with Alzheimer's Disease years before diagnosis and noticeable cognitive impairment (Beltrami et al., 2018; Filiou et al., 2020; Lopez-de-Ipina et al., 2018; Shibata et al., 2018).

5 – Challenges faced by digital phenotyping

Each of the efforts in digital phenotyping discussed are associated with their own merits and drawbacks. However, novel digital measures of mental health share common obstacles before their potential can be fully realized. First, routes to validation of methods remain unclear. Any methods developed or code utilized must be made open for evaluation, which is often not the case with commercial efforts. A potential solution is a common repository of methods to be established for sharing of methods between academia, medicine, and industry. Without it, the fields remain disconnected, skepticism persists, and progress is slowed. Second, all efforts must take into consideration regulations pertaining to data privacy, particularly when working with Protected Health Information. Most digital phenotyping methods discussed above require collection of identifying data, which requires special handling (Cohen & Mello, 2018). Third, there is need for clear routes to approval of novel tools from regulatory authorities such as the Food and Drug Administration. Without it, a digital measurement may accumulate widespread scientific support yet still not be 'valid' as a clinical decision making tool in contexts of medical care and clinical research (Manta et al., 2020). Finally, novel tools must adapt the way in which data is collected and presented in order to integrate with the existing healthcare ecosystem. By doing so, they enable clinicians to make informed decisions without having to allocate extra time towards independent software tools or being inundated with additional of data streams with information that may be difficult to interpret (Abbas et al., 2021).

5.1 Privacy and anonymity

A major issue related to this kind of monitoring is the issue of anonymity and the need to keep personal privacy confidential (Insel, 2017). An authorized app must ensure that all the data that will be gathered in the app will be completely anonymous, i.e., with no personal information stored on the servers.

All data acquired during the day are encrypted directly inside the smartphone memory. Personal information, e.g., telephone numbers, person names or specific location, is whitened by coding it into hash values. By doing so, behavioral patterns such as communication patterns, diurnal variation, movement patterns etc. may be recognized without abrogating privacy. Using this strategy allows counting the number of calls or messages that were sent but not the content or the actual digits. Similarly, one could also measure the distance travelled in a certain time frame without knowing where a trauma victim or responder was during the use of the app. Thus, digital medicine aims to measure and compare the amount of activity done using the smartphone without saving specific personal data.

6 – Promise and future of digital measurement

It is important to distinguish between measurement of behavioral characteristics, measurement of clinically meaningful symptomatology, and measurement of overall disease severity. Computer vision-based quantification of facial activity is a measurement of a behavioral characteristic. The use of facial activity to derive blunted affect in a patient with schizophrenia, for example, is measurement of clinically meaningful symptomatology. However, to arrive at a composite measure of disease severity, one must successfully integrate several measures of clinically meaningful symptomatology to make a prediction. So far, this chapter has discussed individual clinical measures i.e. disparate digital markers of disease severity, which in their best form are proxies of overall disease severity. However, the promise of digital phenotyping lies beyond isolated application of individual measures. Ultimately, neuropsychiatric illness is defined by manifestation of a group of symptoms that characterize a disease. Traditional clinical assessments aim to quantify the severity of all such symptoms to arrive at composite measures of disease severity, However, this is not as simple in the context of digital measurement—and for good reason.

Traditional clinical assessments typically call for scoring of symptom severity on equally weighted discrete scales (e.g. on a scale of 0 to 4, how severe is the patient's tremor?; Elble, 2016). Each of the scores are then combined using simple arithmetic operations such as addition or averaging. In the case of digital measurement, each symptom may be quantified on a continuous scale with its own range of expected values and likely have separate units of measurement (Galatzer-Levy et al., 2020a). Hence, amalgamation of individual measures into a single composite score is less straightforward. The natural solution here is the application of machine learning to train models that make 'predictions' of overall disease severity. However, a significant amount of data must be collected to train such models and their applicability universally must be thoroughly examined before their use. Just as importantly, such a project would require significant technological effort to integrate data streams from individual measurement tools into a unified data infrastructure, as any single academic group or digital assessment tool is unlikely to collect all relevant measures. Past efforts to combine even two such data streams have demonstrated the challenge associated with amalgamation of multi-modal digital measurements (Schultebraucks et al., 2020). However, if such challenges can be overcome i.e., accurate, objective, and sensitive digital measurements can acquired using consumer devices in large patient populations, it could lead to the development of unprecedented models of disease severity, a significant departure from reliance on paper-pencil clinician assessments.

6.1 Data Models and Engineering Needs

Apps are not magic. Both their development and maintenance require large infrastructure investments in computing power as well as storage and movement of data. They also require large teams of software engineers to build front end, backend, plumbing, and delivery systems. These systems must be maintained and updated due to changes in user needs and the software platforms they depend on (i.e. iOS/Android). As such, both the development and maintenance of tools for the identification and intervention require significant initial and sustained investments and efforts to be of clinical value.

References

Abbas, A., Schultebraucks, K., Galatzer-Levy, I. R. (2021). Digital Measurement of Mental Health: Challenges, Promises, and Future Directions. *Psychiatr Ann.* 2021;51(1)

Abbas, A., Yadav, V., Smith, E., Ramjas, E., Rutter, S. B., Benavides, C., ... & Perez-Rodriguez, M. M. (2020). Computer vision-based assessment of motor functioning in schizophrenia: Use of smartphones for remote measurement of schizophrenia symptomatology. *medRxiv*.

Adib F, Mao H, Kabelac Z, Katabi D, Miller RC. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* 2015 (pp. 837-846).

Airas, M., & Alku, P. (2006). Emotions in vowel segments of continuous speech: analysis of the glottal flow using the normalised amplitude quotient. *Phonetica*, *63*(1), 26-46.

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, *4*, 463-476.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Areán, P. A., Ly, K. H., & Andersson, G. (2016). Mobile technology for mental health assessment. *Dialogues in clinical neuroscience*, *18*(2), 163.

Au, R., Piers, R. J., & Devine, S. (2017). How technology is reshaping cognitive assessment: Lessons from the Framingham Heart Study. *Neuropsychology*, *31*(8), 846.

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016, March). Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-10). IEEE.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018, May). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 59-66). IEEE.

Barnett, M. L., & Huskamp, H. A. (2020). Telemedicine for mental health in the United States: Making progress, still a long way to go. *Psychiatric Services*, *71*(2), 197-198.

Beltrami, D., Gagliardi, G., Rossini Favretti, R., Ghidoni, E., Tamburini, F., & Calzà, L. (2018). Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline?. *Frontiers in aging neuroscience*, *10*, 369.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0. 37. *URL http://www. praat. org/. Retrieved March*, *14*, 2018.

Burke, L. E., Shiffman, S., Music, E., Styn, M. A., Kriska, A., Smailagic, A., ... & Mancino, J. (2017). Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. *Journal of medical Internet research*, *19*(3), e77.

Cavelti, M., Kvrgic, S., Beck, E. M., Kossowsky, J., & Vauth, R. (2012). Assessing recovery from schizophrenia as an individual process. A review of self-report instruments. *European psychiatry*, *27*(1), 19-32.

Chowdhary, K. R. (2020). Natural language processing. In *Fundamentals of Artificial Intelligence* (pp. 603-649). Springer, New Delhi.

Clarke, N., Foltz, P., & Garrard, P. (2020). How to do things with (thousands of) words: Computational approaches to discourse analysis in Alzheimer's disease. *Cortex*.

Cohen, I. G., & Mello, M. M. (2018). HIPAA and protecting health information in the 21st century. *Jama*, *320*(3), 231-232.

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., ... & De la Torre, F. (2009, September). Detecting depression from facial actions and vocal prosody. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (pp. 1-7). IEEE.

Cook, B. L., Progovac, A. M., Chen, P., Mullin, B., Hou, S., & Baca-Garcia, E. (2016). Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Computational and mathematical methods in medicine*, *2016*.

Coons, S. J., Eremenco, S., Lundy, J. J., O'Donohoe, P., O'Gorman, H., & Malizia, W. (2015). Capturing patient-reported outcome (PRO) data electronically: the past, present, and promise of ePRO measurement in clinical trials. *The Patient-Patient-Centered Outcomes Research*, *8*(4), 301-309.

Coppersmith, G., Dredze, M., & Harman, C. (2014, June). Quantifying mental health signals in Twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 51-60).

Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ digital medicine*. 2019;2(1):1-5.

Coravos, A., Khozin, S., & Mandl, K. D. (2019). Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ digital medicine*, *2*(1), 1-5.

Corcoran, C. M., & Cecchi, G. A. (2018). Computational approaches to behavior analysis in psychiatry. *Neuropsychopharmacology*, *43*(1), 225.

Cuthbert, B. N. (2015). Research Domain Criteria: toward future psychiatric nosologies. *Dialogues in clinical neuroscience*, *17*(1), 89.

Cuthbert, B. N., & Insel, T. R. (2010). Toward new approaches to psychotic disorders: the NIMH Research Domain Criteria project.

Depp, C. A., Bashem, J., Moore, R. C., Holden, J. L., Mikhael, T., Swendsen, J., ... & Granholm, E. L. (2019). GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. *NPJ digital medicine*, *2*(1), 1-7.

Dorsey, E. R., Bloem, B. R., & Okun, M. S. (2020). A New Day: the role of telemedicine in reshaping care for persons with movement disorders. *Movement Disorders*, *35*(11), 1897-1902.

Dagum, P. (2018). Digital biomarkers of cognitive function. npj Digital Medicine, 1(1), 10. doi:10.1038/s41746-018-0018-4

Ekman, P., & Friesen, W. V. (1978). *Manual for the facial action coding system*. Consulting Psychologists Press.

Ekman, R. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

Elble, R. J. (2016). The essential tremor rating assessment scale. *Journal of Neurology & Neuromedicine*, *1*(4).

Epp, C., Lippold, M., & Mandryk, R. L. (2011, May). Identifying emotional states using keystroke dynamics. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 715-724).

Filiou, R. P., Bier, N., Slegers, A., Houzé, B., Belchior, P., & Brambati, S. M. (2020). Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: a scoping review. *Aphasiology*, *34*(6), 723-755.

Fuchs, T. (2010). Subjectivity and intersubjectivity in psychiatric diagnosis. *Psychopathology*, *43*(4), 268-274.

Galatzer-Levy, I., Abbas, A., Koesmahargyo, V., Yadav, V., Perez-Rodriguez, M. M., Rosenfield, P., ... & Hansen, B. J. (2020b). Facial and vocal markers of schizophrenia measured using remote smartphone assessments. *medRxiv*.

Galatzer-Levy, I., Abbas, A., Yadav, V., Koesmahargyo, V., Aghjayan, A., Marecki, S., ... & Sauder, C. (2020a). Remote digital measurement of visual and auditory markers of Major Depressive Disorder severity and treatment response. *medRxiv*.

Glenn, T., & Monteith, S. (2014). New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Current Psychiatry Reports, 16(12), 523.

Godino-Llorente, J. I., Osma-Ruiz, V., Sáenz-Lechón, N., Cobeta-Marco, I., González-Herranz, R., & Ramírez-Calvo, C. (2008). Acoustic analysis of voice using WPCVox: a comparative study with Multi Dimensional Voice Program. *European archives of oto-rhino-laryngology*, *265*(4), 465-476.

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., ... & Dubois, B. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, *23*(15), 2129-2170.

Goyal, K., Agarwal, K., & Kumar, R. (2017, April). Face detection and tracking: Using OpenCV. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)* (Vol. 1, pp. 474-478). IEEE.

Hafiz, P., Miskowiak, K. W., Kessing, L. V., Jespersen, A. E., Obenhausen, K., Gulyas, L., ... & Bardram, J. E. (2019). The internet-based cognitive assessment tool: System design and feasibility study. *JMIR formative research*, *3*(3), e13898.

Haque, A., Guo, M., Miner, A. S., & Fei-Fei, L. (2018). Measuring depression symptom severity from spoken language and 3D facial expressions. *arXiv preprint arXiv:1811.08592*.

Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). *Natural language processing: python and NLTK*. Packt Publishing Ltd.

Hashemi, J., Spina, T. V., Tepper, M., Esler, A., Morellas, V., Papanikolopoulos, N., & Sapiro, G. (2012, November). A computer vision approach for the assessment of autism-related behavioral markers. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1-7). IEEE.

Hashim, N. W., Wilkes, M., Salomon, R., Meggs, J., & France, D. J. (2017). Evaluation of voice acoustics as predictors of clinical depression scores. *Journal of Voice*, *31*(2), 256-e1.

He, Q., Veldkamp, B. P., Glas, C. A., & de Vries, T. (2017). Automated assessment of patients 'self-narratives for posttraumatic stress disorder screening using natural language processing and text mining. *Assessment*, *24*(2), 157-172.

Huang, L., Wei, W., Liu, Z., Zhang, T., Wang, J., Xu, L., ... & Le Meur, O. (2020). Effective schizophrenia recognition using discriminative eye movement features and model-metric based features. *Pattern Recognition Letters*, *138*, 608-616.

Insel, T. R. (2017). Digital phenotyping: technology for a new science of behavior. *Jama*, *318*(13), 1215-1216.

Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing parselmouth: A python interface to praat. *Journal of Phonetics*, *71*, 1-15.

Jeon, H., Lee, W., Park, H., Lee, H. J., Kim, S. K., Kim, H. B., ... & Park, K. S. (2017). Automatic classification of tremor severity in Parkinson's disease using a wearable device. *Sensors*, *17*(9), 2067.

Jiang, Z., Harati, S., Crowell, A., Mayberg, H., Nemati, S., & Clifford, G. (2020). Classifying Major Depressive Disorder and Response to Deep Brain Stimulation Over Time by Analyzing Facial Expressions. *IEEE Transactions on Biomedical Engineering*.

Kaser, M., Zaman, R., & Sahakian, B. J. (2017). Cognition as a treatment target in depression. *Psychological medicine*, *47*(6), 987-989.

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin*, *13*(2), 261-276.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., ... & Ustun, T. B. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychological medicine*, *35*(2), 245.

Komeili, M., Pou-Prom, C., Liaqat, D., Fraser, K. C., Yancheva, M., & Rudzicz, F. (2019). Talk2Me: Automated linguistic data collection for personal assessment. *Plos one*, *14*(3), e0212342.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606-613.

LaFrance Jr, W. C., Ho, W. L. N., Bhatla, A., Baird, G., Altalib, H. H., & Godleski, L. (2020). Examination of Potential Differences in Reporting of Sensitive Psychosocial Measures via Diagnostic Evaluation Using Computer Video Telehealth. *The Journal of Neuropsychiatry and Clinical Neurosciences*, appi-neuropsych.

Lamont, R. M., Daniel, H. L., Payne, C. L., & Brauer, S. G. (2018). Accuracy of wearable physical activity trackers in people with Parkinson's disease. *Gait & posture*, *63*, 104-108.

Lancaster, C., Koychev, I., Blane, J., Chinner, A., Wolters, L., & Hinds, C. (2019). The Mezurio smartphone application: Evaluating the feasibility of frequent digital cognitive assessment in the PREVENT dementia study. *medRxiv*, 19005124.

Lopez-de-Ipina, K., Martinez-de-Lizarduy, U., Calvo, P. M., Mekyska, J., Beitia, B., Barroso, N., ... & Ecay-Torres, M. (2018). Advances on automatic speech analysis for early detection of Alzheimer disease: a non-linear multi-task approach. *Current Alzheimer Research*, *15*(2), 139-148.

Löwe, B., Kroenke, K., Herzog, W., & Gräfe, K. (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of affective disorders*, *81*(1), 61-66.

Manta, C., Patrick-Lake, B., & Goldsack, J. C. (2020). Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. *Digital Biomarkers*, *4*(3), 69-77.

McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., & Landwehr, J. (2017). Using social media to monitor mental health discussions− evidence from Twitter. *Journal of the American Medical Informatics Association*, *24*(3), 496-502.

Miller, A. L., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language*, *104*, 25-42.

Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience*, *31*(1), 13.

Nieto-Hidalgo, M., Ferrández-Pastor, F. J., Valdivieso-Sarabia, R. J., Mora-Pascual, J., & García-Chamizo, J. M. (2018). Gait Analysis Using Computer Vision Based on Cloud Platform and Mobile Device. *Mobile Information Systems*, *2018*.

Oğuz, H., Kiliç, M. A., & ŞAFAK, M. A. (2011). Comparison of results in two acoustic analysis programs: Praat and MDVP. *Turkish Journal of Medical Sciences*, *41*(5), 835-841.

Overall, J. E., & Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychological reports*, *10*(3), 799-812.

Pallagrosi, M., Fonzi, L., Picardi, A., & Biondi, M. (2016). Association between clinician's subjective experience during patient evaluation and psychiatric diagnosis. *Psychopathology*, *49*(2), 83-94.

Pang, Y., Christenson, J., Jiang, F., Lei, T., Rhoades, R., Kern, D., ... & Liu, C. (2020). Automatic detection and quantification of hand movements toward development of an objective assessment of tremor and bradykinesia in Parkinson's disease. *Journal of Neuroscience Methods*, *333*, 108576.

Parola, A., Simonsen, A., Bliksted, V., & Fusaroli, R. (2020). Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis. *Schizophrenia Research*, *216*, 24-40.

Patel, R., Jayatilleke, N., Broadbent, M., Chang, C. K., Foskett, N., Gorrell, G., ... & Roberts, A. (2015). Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. *BMJ open*, *5*(9).

Piau, A., Wild, K., Mattek, N., & Kaye, J. (2019). Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. *Journal of medical Internet research*, *21*(8), e12785.

Onnela, J.-P., Keshavan, M., Staples, P., Barnett, I., & Torous, J. (2018). 150. Automated Longitudinal Latent Interval Estimation With Applications to Sleep. Biological Psychiatry, 83(9, Supplement), S61. doi:https://doi.org/10.1016/j.biopsych.2018.02.168

Quatieri, T. F., & Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity. In *Thirteenth Annual Conference of the International Speech Communication Association*.

Santomas, K. (2020). Passive Monitoring of Physical Activity Using a Fitbit Charge 3 in Individuals at Risk for Depression.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*(4), 419-436.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of autism and developmental disorders*.

Schultebraucks, K., Shalev, A. Y., Michopoulos, V., Grudzen, C. R., Shin, S. M., Stevens, J. S., ... & Marmar, C. R. (2020). A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nature medicine*, *26*(7), 1084-1088.

Shama, K., Krishna, A., & Cholayya, N. U. (2006). Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP Journal on Advances in Signal Processing*, *2007*, 1-9.

Shibata, D., Ito, K., Nagai, H., Okahisa, T., Kinoshita, A., & Aramaki, E. (2018). Idea density in Japanese for the early detection of dementia based on narrative speech. *PloS one*, *13*(12), e0208418.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1-32.

Shin, S., Yeom, C. W., Shin, C., Shin, J. H., Jeong, J. H., Shin, J. U., & Lee, Y. R. (2016). Activity monitoring using a mHealth device and correlations with psychopathology in patients with chronic schizophrenia. *Psychiatry Research*, *246*, 712-718.

Stewart, R., & Velupillai, S. (2020). Applied natural language processing in mental health big data. *Neuropsychopharmacology*, *46*(1), 252.

Thapa, S., Adhikari, S., Naseem, U., Singh, P., Bharathy, G., & Prasad, M. (2020, November). Detecting Alzheimer’s Disease by Exploiting Linguistic Information from Nepali Transcript. In *International Conference on Neural Information Processing* (pp. 176-184). Springer, Cham.

Torous, J., Onnela, J., & Keshavan, M. (2017). New dimensions and new tools to realize the potential of RDoC :digital phenotyping via smartphones and connected devices. Translational psychiatry, 7(3), e1053.

van Eijk, R. P. (2020). Frequent self-assessments in ALS Clinical Trials: worthwhile or an unnecessary burden for patients?. *Annals of Clinical and Translational Neurology*, *7*(10), 2074.

Wang, M., Guo, L., & Chen, W. Y. (2017). Blink detection using Adaboost and contour circle for fatigue recognition. *Computers & Electrical Engineering*, *58*, 502-512.

Williams, J. B. (1988). A structured interview guide for the Hamilton Depression Rating Scale. *Archives of general psychiatry*, *45*(8), 742-747.

Williams, S., Shepherd, S., Fang, H., Alty, J., O'Gorman, P., & Graham, C. D. (2019). Computer vision of smartphone video has potential to detect functional tremor. *Journal of the neurological sciences*, *401*, 27-28.

World Health Organization (2020). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/

Wright, A. J. (2020). Equivalence of remote, digital administration and traditional, in-person administration of the Wechsler Intelligence Scale for Children, (WISC-V). *Psychological assessment*.

Wright, S. P., Collier, S. R., Brown, T. S., & Sandberg, K. (2017). An analysis of how consumer physical activity monitors are used in biomedical research. *The FASEB Journal*, *31*(1_supplement), 1020-24.

Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S. A., ... & Leow, A. (2018). Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study. *Journal of medical Internet research*, *20*(7), e241.

Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S. A., ... & Leow, A. (2018). Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study. *Journal of medical Internet research*, *20*(7), e241.