

REVIEW ARTICLE OPEN

Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety

Kit Huckvale¹, Svetha Venkatesh² and Helen Christensen^{1,3}

The use of data generated passively by personal electronic devices, such as smartphones, to measure human function in health and disease has generated significant research interest. Particularly in psychiatry, objective, continuous quantitation using patients' own devices may result in clinically useful markers that can be used to refine diagnostic processes, tailor treatment choices, improve condition monitoring for actionable outcomes, such as early signs of relapse, and develop new intervention models. If a principal goal for digital phenotyping is clinical improvement, research needs to attend now to factors that will help or hinder future clinical adoption. We identify four opportunities for research directed toward this goal: exploring intermediate outcomes and underlying disease mechanisms; focusing on purposes that are likely to be used in clinical practice; anticipating quality and safety barriers to adoption; and exploring the potential for digital personalized medicine arising from the integration of digital phenotyping and digital interventions. Clinical relevance also means explicitly addressing consumer needs, preferences, and acceptability as the ultimate users of digital phenotyping interventions. There is a risk that, without such considerations, the potential benefits of digital phenotyping are delayed or not realized because approaches that are feasible for application in healthcare, and the evidence required to support clinical commissioning, are not developed. Practical steps to accelerate this research agenda include the further development of digital phenotyping technology platforms focusing on scalability and equity, establishing shared data repositories and common data standards, and fostering multidisciplinary collaborations between clinical stakeholders (including patients), computer scientists, and researchers.

npj Digital Medicine (2019)2:88; <https://doi.org/10.1038/s41746-019-0166-1>

INTRODUCTION

Digital phenotyping¹ (or personal sensing²) is the moment-by-moment, in situ quantification of the individual-level human phenotype using data from personal digital devices. It seeks to exploit the potential of data that are automatically generated and aggregated by smartphones, wearables and other connected devices to measure (or offer robust proxies for) human behavior and function in both health and disease. Today, these data streams include sensor measurements, activity logs and user-generated content.³

Data-driven, objective measurement of individual function is of specific interest in psychiatry, which has previously relied almost exclusively on self-reports of mental health symptoms, which has few biological markers, and where diagnostic categories remain unclear.^{4–6} Building on the widespread adoption of smartphones as the principal enabling technology, digital phenotyping has been enthusiastically adopted as a research theme in mental health. Our searches identified over 80 peer-reviewed publications since 2015 that focus on digital phenotyping for psychiatric conditions.

Many of these studies appear to anticipate that digital phenotyping should play a role in routine clinical practice, for example by enhancing aspects of clinical diagnosis and treatment through earlier detection of condition onset, relapse or treatment response. As a result, there is a timely opportunity to consider what this vision of clinical digital phenotyping might require in terms of scope, quality and safety in order to be used in practice.

Three factors motivate this question. The first is the historically slow pace of translation of health innovations into practice. Reported lag times of 17 years⁷ are at least partially accounted for by mismatches between the outputs of research and what—in terms of both design and supporting information⁸—is needed for adoption. The second, relatedly, is the formalization of approaches to health technology assessment which act to codify criteria for adoption, such as cost-effectiveness.⁹ The third acknowledges the risk posed by technology change: should it take 17 years to find practical uses for digital phenotyping, it may well be that the underlying technologies are obsolete. There is a risk that, without such consideration, the potential benefits of digital phenotyping are delayed or not realized because clinically-feasible approaches, and the evidence required to support clinical commissioning, are not generated through timely research.

The purpose of this review is to highlight how developments in digital phenotyping have created a broad range of potential clinical uses, to identify salient gaps, and to highlight opportunities for action intended to promote future clinical adoption, quality and safety (summarized in Table 1). The review argues that a range of developments are needed to accelerate progress, which include the development of scalable data collection infrastructure that addresses equity and privacy issues, the application of methods from machine learning to process and analyze signals, and the development of validation approaches for data quality that address challenges of bias and noise inherent in population-scale phenotypic data. We also highlight the potential

¹Black Dog Institute, UNSW Sydney, Sydney, NSW, Australia; ²A2I2, Deakin University, Geelong, VIC, Australia and ³Mindgardens Neuroscience Network, Sydney, NSW, Australia
Correspondence: Kit Huckvale (c.huckvale@unsw.edu.au)

Received: 23 April 2019 Accepted: 9 August 2019

Published online: 06 September 2019

Table 1. Seven priorities: opportunities and practical steps for progressing a vision of clinical digital phenotyping

#	Priority
1	Applying digital phenotyping to the mechanisms and behaviors underlying psychiatric disorders rather than outcomes alone.
2	Prioritizing research into digital phenotyping according to realistic clinical uses.
3	Anticipating clinical quality, safety and acceptability issues that will act as barriers to implementation and uptake.
4	Combining digital phenotyping with digital interventions.
5	Developing data collection platforms with a focus on issues of equity, trust and privacy.
6	Developing shared data resources to accelerate collaborative research, replication and scale-up studies.
7	Establishing strong collaborations with healthcare professionals, providers and computer science.

clinical value of integration between digital phenotyping and digital interventions in order to accelerate adoption.

Reflecting the topic of this special collection, we draw examples from mental health in general, and adolescents and young adults, specifically. This group represents a potentially important target group for mental health-focused digital phenotyping. Both the incidence and overall prevalence of serious mental disorders peaks in those aged 18–30. Unlike younger children who often do not yet have a personal smartphone, device ownership is ubiquitous, and—in Australia—higher in this group than in any other demographic. Nevertheless, the opportunities we identify are not restricted to this age group or condition area and are of potential relevance in any clinical domain where digital phenotyping is being considered.

Mapping digital phenotyping to potential clinical applications: examples from youth mental health

There are now a broad range of potential applications for digital phenotyping with clinical relevance in youth mental health and that are the subject of active evaluation. These span the breadth of care stages, from screening, diagnosis, monitoring, and treatment, including early intervention and relapse prevention. These are summarized in Table 2 and discussed below. (Because the primary youth literature is limited, we also include examples of digital phenotyping that are relevant in youth either because of the developmental significance of the condition or because the peak age of onset occurs in this age group.)

Mood disorder identification, tracking, and predicting subsequent treatment response. Within student and young adults cohorts, passive detection of activity changes using accelerometry, GPS, and phone utilization data has shown promise for identifying individuals at risk for self-reported depression or anxiety^{10–12} as well as potential upstream determinants of future mental ill health, such as self-reported stress.^{13–17} Although mood disorders are most commonly clinically diagnosed between the ages of 25 and 30,¹⁸ it is increasingly recognized that psychological distress may predate this by many years, either as subclinical symptoms or because of delayed help seeking relating to stigma and poor expectations of clinical support. Digital phenotyping strategies that can identify these at-risk and un-diagnosed individuals might offer a way to alleviate significant morbidity and future clinical demand, as well assist parental and self-monitoring in adolescents once diagnosis is confirmed.¹⁹ Further work is required, however, to elucidate the relationship between population scale measurement of constructs such as “stress” that are operationalized in different ways and the ultimate development of mood disorders.

Separately, an important potential opportunity in the clinical management of depression is predicting treatment response.²⁰ Only 50% of individuals respond to the first treatment they are offered,²¹ and lengthy trial-and-error approaches incur significant patient and health-service costs. Digital phenotyping using voice analysis²² provides a proof of concept for new methods to predict

treatment response but improvements in prediction accuracy are now needed to enable clinical uses.

Bipolar disorder and relapse prevention. In bipolar disorder (BPD), there has been substantial progress in the development of digital-phenotyping techniques for condition monitoring and relapse detection. Changes in location and activity patterns,^{12,23–25} keyboard interaction dynamics,^{26,27} social phone utilization metrics^{24,25,28} (such as calls placed and received) and voice^{26,29} (for example, captured from phone calls) have been used, alongside active measures, to predict both manic and depressive states. Relapse is common in BPD, with 70% of individuals experiencing deterioration or recurrence within 5 years of a manic episode.³⁰ Despite subtle symptoms routinely being present at up to 4 weeks³¹ before relapse, access to timely treatment remains a major issue, partly because symptoms can be highly patient-specific³¹ and partly because comorbid factors, such as drug use and co-existing psychiatric disease,³² affect the capacity of individuals to respond effectively. Early-warning sign interventions that rely on self-monitoring are desirable for young adult BPD patients³³ and effective in increasing time-to-recurrence while reducing hospitalisation.³⁴ The development of digital phenotyping-based methods promises early warning sign services that could be offered to individuals who would otherwise find it hard to sustain self-monitoring.^{28,35,36} The first randomized³⁷ and cohort³⁷ studies are now either underway or will start shortly.

Opioid overdose detection and harm reduction. Opioid overdose carries a high risk of respiratory failure and death. In 2016, 245 Australians aged 15–34 died of opioid overdose, of which 216 (88%) were accidental.³⁸ This represents a 31% increase in yearly deaths compared to a decade earlier (6.98 deaths per 100,000 in 2016 compared to 5.33 in 2006). Opioid toxicity is reliably treatable using the drug naloxone if caught in time, but users often struggle to identify signs. A smartphone-based, harm-reduction solution that uses digital phenotyping to detect signs of respiratory distress³⁹ raises the prospect of reducing accidental overdoses by, for example, contacting community first responders to administer naloxone⁴⁰ or prioritizing those with detected near-overdoses for methadone therapy in order to avoid future events.⁴¹

Detection of harmful alcohol drinking behaviors and exposures to alcohol-related messaging. Some of the only phenotyping literature that focusses explicitly on a youth population has explored whether alcohol-related exposures can be predicted using passive monitoring of location data^{42–47} and, separately, if alcohol consumption behaviors can be predicted using smartphone sensing and activity data.^{48,49} These uses highlight how digital phenotyping can also be used for public health purposes by generating information not only about individuals but concerning aggregate patterns of behavior that can then be used to inform structural interventions, for example, ensuring that retailers are complying with the law in relation to the supply of

Table 2. Spectrum of youth-relevant mental health applications being explored using digital phenotyping

Prevention	Screening and early diagnosis	Monitoring	Treatment
Fostering resilience and health promoting behaviors	Proactive identification of undiagnosed conditions and/or formal confirmation of a specific condition	Early detection of condition changes, adverse events, and relapse	Tailored intervention, engagement and treatment efficacy monitoring
<p>Stress identification^{13–17} Passive detection of changes in self-perceived stress in order to foster self-regulation and resilience or trigger proactive help-seeking before the onset of frank mental health symptoms.</p> <p>High-risk alcohol use detection^{49,50} Passive identification of high-risk drinking episodes using activity and phone utilization data in order to trigger prevention interventions.</p>	<p>Mood disorder detection^{10–12} Passive detection of activity changes using accelerometry, GPS, phone utilization data in order to identify individuals at risk for depression or anxiety.</p> <p>Suicidality detection^{51–53} Automatic natural language processing of social media posts to identify at-risk individuals.</p>	<p>Mood disorder self-monitoring¹⁹ Using activity and location data to discreetly monitor mood changes as part of combined parent-child self-monitoring intervention.</p> <p>BPD relapse prediction^{12,23,26–29,37,38} Passive monitoring for depressive (using keyboard signals) and manic (using voice signals) signs indicative of relapse, enabling “early warning sign” interventions.</p> <p>Opioid overdose detection⁴⁰ Active abnormal respiratory pattern detection post opioid use using smartphone “sonar” (combining speaker and microphone.)</p> <p>Schizophrenia relapse prediction¹²⁴ Passive monitoring for early-warning signs using accelerometry and heart variability in order to detect relapse early and enable medical intervention.</p>	<p>Depression therapy enhancement^{22,112} Sensor-derived signals (e.g. location information) used to tailor therapy in order to maximize user engagement and treatment effect or identify when treatment is not working.</p>
BPD bipolar disorder			

alcohol to minors in locations where problem drinking emerges as a pattern from digital data.

Identification of risk of suicide in the wild. Automatic natural language processing of social media posts has been used successfully to identify individuals with evidence of psychological distress^{50–52} that might place them at risk of self-harm or suicide. Suicide is a leading cause of death amongst children and young people,⁵³ and predicting rapid escalations in the risk of suicide is a policy priority, particularly given the development of new, effective interventions, such as ketamine for rapid reduction of depressive and suicidal symptoms.⁵⁴ Proactive screening in online environments raises important privacy questions but recognizes that many suicides occur out of the blue, prior to contact with health professions. Within at-risk populations, signals from smartphones⁵⁵ and clinical measurements (such as electrocardiography to detect heart rate variability⁵⁶) may offer a discreet way to provide a safety net.

Gaps and opportunities

Despite the potential described above, today only a few research and healthcare organizations are collecting digital signals, and these activities are largely exploratory. The data that result are small-scale (typically involving a few tens of people monitored for only a short period of time³⁶), partial, unstandardized and often not linkable, resulting in multiple, small data “silos”.⁵⁷ These are not suitable for robust identification of digital biomarkers concerning mental illness onset, treatment response or relapse. They are also often insufficient for effective analysis (for example, beyond simple correlations), because the data are small-scale and noisy. The development and application of appropriate methods for study execution and the analysis of digital phenotyping data has already been identified as a priority for the future clinical relevance of the field.⁵⁸ Sitting alongside this, we perceive several

additional opportunities that collectively stand to accelerate the clinical impacts of digital phenotyping.

Opportunity 1: applying digital phenotyping to the mechanisms and behaviors underlying psychiatric disorders rather than outcomes alone

An emerging template for contemporary studies in digital phenotyping is to explore through correlations,^{59,60} modelling^{61,62} or machine learning,^{17,63} the relationship between a set of smartphone-derived sensor and utilization features and the result of a self-completed outcomes instrument, such as the PHQ-9 for depression. This approach has yielded new, clinically-relevant phenomena, such as apparent changes in smartphone-measured sleep continuity⁶⁴ and location-based activity⁶⁵ that precede a major depressive episode by weeks and could therefore have potential uses for onset and relapse prediction. This kind of data-led, clinical endpoint-based approach recognizes that there are a large number of potential sensor and analytics data sources, each of which may be (at least in advance) of uncertain significance in relation to a given clinical outcome, as well as being amenable to any number of summary representations (for example, due to complex temporality and missingness.⁶⁶) Nevertheless, we want to highlight the potential value—both for clinical applications and research—of a complementary, mechanistic approach that considers not only clinical outcomes but also intermediate functional and behavioral states,² as well as disease-related processes, as potential targets for prediction using digital phenotyping.

Consider, for example, the relationship between cognitive dysfunction and depression, which typically presents first in early adulthood.¹⁸ Subjective impairment of thought and concentration forms part of the diagnostic criteria for major depressive disorder,⁶⁷ while objective testing consistently identifies a range of specific functional deficits in executive function (including processing speed), memory and attention.^{68–70} These deficits are

present at initial diagnosis^{71,72} and have been identified as potential trait markers for depression⁷³ in at least a subset of individuals.⁷² It is already known that cognitive dysfunction improves with therapy and may predict specific treatment response,^{74,75} such as the likely success of talking therapies, disease course⁷² and future neurodegenerative illness.⁷⁶

This example illustrates firstly how clinically useful opportunities can exist for measuring specific facets of a disorder rather than its overall state using digital phenotyping. Being able to quantitate cognitive change using a phenotyping-based approach is attractive because current psychometric and neuropsychiatric tests rarely reflect specific cognitive processes unambiguously, can be unwieldy, take time, and are poorly standardised.⁷³ Yet, because cognitive function may predict treatment response,²¹ it is attractive as a target for prediction, particularly given the resource and patient costs associated with the selection of ineffective therapy.⁷⁷ The research goal here should be to identify proxies for cognitive tests which are practical to apply quickly and routinely, and which offer timely and more precise signals of improvement.

Secondly, the causal and temporal relationships between cognitive dysfunction and affective symptoms is itself an open research question that is amenable to exploration using phenotyping.^{73,76} The potential feasibility of discreet, continuous digital phenotyping in young adults offers a route to address the specific call for longitudinal studies⁷³ that can assess if and how cognitive symptoms precede the peak onset of depression in the mid-late twenties.¹⁸ Because depression state accounts for a only small proportion of the observed variation in cognitive function between individuals,⁶⁸ the capacity of digital phenotyping to capture detailed within-individual data⁶⁶ is also important.

Thirdly, this focus can act as a rational guide as to what signals are captured from users' digital behavior in the first place. For example, the observation that specific sub-measures of executive function, such as semantic and phonemic fluency, are significantly reduced in first episode depression⁷² reasonably directs attention to device activities where these capabilities might be exercised, such as word-finding whilst typing. In BPD, similar metrics have yielded new potential cognitive markers.²⁶ Effort can then be directed to the feasibility of collecting these data, for example the technical ability to monitor a user's on-screen keyboard versus understandable potential privacy concerns.

This guided approach is important not only because there is otherwise a large space of things that could be measured, but because it is becoming clear that digital phenotyping is not always optimal for the detection of particular behavioral signals. For example, sleep detection using smartphones overestimates sleep duration and underestimates sleep disturbance compared to formal actigraphy,⁷⁸ while self-reported mood using experience sampling methods substantially improves model prediction compared to digital phenotyping alone in depression.⁶³ Having reference standards (such as validated functional measurement instruments) that are conceptually "closer" to the original data sources should make it easier to critically select, assess and refine what is used for modelling. Because each measurement comes with a concrete cost both in terms of implementation and user experience (e.g. in battery life impact, data transfer costs and/or perceived impacts on privacy or acceptability) a rational approach to selection may also help to avoid wasted effort.

The mechanistic approach we describe here emphasizes the value of the existing physiological measurement and psychometric literature in suggesting intermediate targets for proxy prediction by digital phenotyping. We are not new in promoting this type of strategy.^{2,79,80} For example, Mohr and colleagues advocate using digital phenotyping to build markers of behavioral traits and use these, in turn, to explore relationships with higher-level states.² Our approach is complementary in advocating a focus on signs and symptoms with established (or emerging) direct practical clinical use. Importantly, neither precludes using

these intermediate targets in turn to predict clinical endpoints. It is an empirical question as to whether models built in this way will have greater predictive power than those that attempt to link raw data directly to outcomes, and future work should critically assess this potential.

Opportunity 2: prioritizing research into digital phenotyping according to realistic clinical uses

Digital phenotyping strategies, as with any health technology intended for clinical use, will ultimately need to demonstrate both efficacy and cost-effectiveness. What is acceptable performance, however, is substantially contingent on the intended clinical application. An illustrative comparison can be made for BPD between screening, e.g. for detecting onset in early adulthood,^{12,81} and monitoring existing patients, e.g. for relapse detection.²⁸

One argument made in favor of digital phenotyping is that the ubiquity of consumer digital devices will enhance the reach of the services that result.^{5,58} From the point of view of the technical performance of digital phenotyping-based screening for new conditions, however, a population-focused strategy cannot escape the challenges that apply to any screening programme.⁸² For example, assuming a point prevalence for clinical BPD of 0.6% in adolescents,⁸³ any new digital phenotyping screening test would need to have a specificity of at least 99.4% (assuming perfect sensitivity, and no targeting other than by age) if the group of those who positive is not to be dominated by false positives. Even assuming an appetite for false positives, representative of existing clinical tests,⁸⁴ that allows for 10 false alerts for every true case, a specificity of 94.0% is required. (For context, the best specificity of the relevant BPD studies we reviewed was 87.2%.¹²) Specificity problems may be enough to discount a screening test, given the dual burden of unnecessary worry in a non-clinical population and actual healthcare costs associated with managing people who present incorrectly as screen positives. Parallel issues affect sensitivity, particularly when differences between healthy and diseased populations are small or where measured changes account for only a small proportion of inter-individual variance.⁷⁰ These issues are not simply theoretical. Commercial digital phenotyping platforms whose stated purpose is to support population-scale diagnostic screening are already being piloted in health systems despite unclear evidence concerning their psychometric properties.^{85,86}

By contrast, specificity may be less of concern for monitoring of those with an established condition population for signs of relapse. The absolute numbers involved are likely be smaller, limiting the scope for burden on service delivery. Individuals may find a false positive risk acceptable if this means that genuine episodes are not missed. And, importantly, these trade-offs can be explicitly stated in advance so that individuals can make an informed choice. Finally, established caring relationships might mean that false positives can be more efficiently triaged out (based on the known profile of each patient, particularly if digital phenotyping can be paired with continuous self-monitoring) without excessive cost or distress. In this scenario, rather than maximizing specificity, it may well be that sensitivity becomes the most important issue given the costs and burden associated with an unmanaged relapse.

Technical test performance is not the only relevant concern. Established principles for clinical screening programmes⁸² require, for example, that conditions have a prodromal phase in which early intervention yields clinical benefit. Despite interest in digital phenotyping for Parkinson's Disease (PD) that focuses on motor symptoms,⁸⁷ early therapy is not clearly beneficial for the control of these symptoms.⁸⁸ Similarly, for depression the clinical significance in terms of future progression to depression of subclinical low mood in otherwise healthy individuals—

particularly at a population scale—is unclear. Importantly, this does not imply that these approaches lack value. For example, a digital phenotyping test deployed prospectively to those who present with motor symptoms could offer a low-cost aid to diagnosis (or subtype elucidation) in PD, which is frequently misdiagnosed.^{89–91} Focus could also be directed towards targets that are modifiable. Unlike motor symptoms, cognitive impairment, which is prevalent in PD, appears to respond to targeted intervention, for example using exercise.⁹² Digital phenotyping used to either detect or monitor cognitive changes could usefully inform rational clinical management in early disease (when intensive clinical monitoring is otherwise not warranted.) Prevention strategies may also be more tolerant of false positive results if the follow-on intervention is simple, low-cost and generally acceptable. Many digital interventions could fall into this category of response.

The differences between the two scenarios outlined at the start of this section (both ostensibly using digital phenotyping as a diagnostic test for a defined change in health status) underline how the choice of application shapes not only the risks and potential costs involved but the standards to which tests should be held. While it is not unreasonable to assume that there will be improvements in classification performance as the field develops, and while calls for larger scale studies³⁶ (which promise better performance) are timely, a clear sense of the ultimate clinical goals remains important to gauge progress. In our view, there is no reason to delay this. For some applications, it will not be possible to achieve diagnostic accuracy statistics of 99% or higher. In order to avoid wasting effort and time, and to ensure that the evidence base needed to support clinical commissioning develops effectively, consideration of how clinical priorities (whether on grounds of burden of disease, potential resource saving or unmet need) intersect with technical feasibility should be a routine feature of research goal-setting now. Without this, there is a risk of outputs that have no realistic prospect of being used in clinical practice. At the very least, these test accuracy statistics should be fully reported; it is not uncommon to see sensitivity (recall) and the positive predictive value (precision) being reported, but not specificity. Study authors may reasonably contend that the balance of these issues will vary according to setting, provider risk appetite and patient attitudes. Decision frameworks, such as net benefit,⁸⁴ offer researchers a way to model tradeoffs between costs-harms under a range of clinically realistic scenarios without having to commit to a particular solution. These kinds of models should be routinely reported in digital phenotyping studies.

Opportunity 3: anticipating clinical quality, safety and acceptability issues that will act as barriers to implementation and uptake

Implementation-relevant concepts of quality and safety are well-operationalized in clinical medicine, for example as the six Institute of Medicine quality domains.⁹³ These span safety, effectiveness, patient-centredness, timeliness, efficiency and equity. While each is relevant from the point of future implementation of digital phenotyping, several dimensions are salient.

Person-centered care. Because patients and consumers are the ultimate gatekeepers of whether it is used, clinical digital phenotyping will rely on a person-centered approach. For complex long-term conditions, there are benefits in monitoring strategies that moderate treatment burden⁹⁴ by reducing explicit self-monitoring and the constant reminders of health status and functional limitations that this can entail. Conversely, some groups may prefer the active engagement that self- or professionally-supported care entails. For example, a qualitative study of potential young adult users of app-based behavior change interventions found that most were not receptive to the use of

contextual tailoring, of the kind that digital sensing could provide, to augment these tools.⁹⁵ Beyond individual preferences, the potential for consequential impacts on self-management and self-regulation skills of increasingly automated measurement remains an open question.

Digital phenotyping relies, both in development and subsequent application, on the ongoing willingness of users to grant access to the various data streams, from on-device sensors to third party social media, that provide insights into their daily behavior. There are multiple potential trade-offs that patients and the public might want to consider in making an informed choice about whether to consent to pervasive monitoring.⁹⁶ Privacy and data governance are understandable topical concerns, given the repeated identification of poor privacy practices by large internet companies and repeated failures observed in related consumer technologies, such health apps.⁹⁷ The development of digital phenotyping entails technical choices including where, for example, data will be processed and stored—particularly if machine learning models are cloud-based—with real potential implications on the acceptability of the ultimate solution to users. In addition, although more trusting of doctors than other groups, patients appear to be wary of sharing certain types of data, such as location, which is routinely used in digital phenotyping.⁹⁸ Strategies will therefore be needed to empower providers to negotiate appropriate access to these data.

How these trade-offs play out will inform the feasibility of different digital phenotyping approaches. For example, individuals with a serious mental health issue may be highly motivated to avoid the risk of relapse even if this requires extensive data about everyday life, including sources such as voice samples.²⁹ There is an opportunity for user-centered research that explores the detail of these compromises.^{98,99} This should seek to focus effort towards applications that are likely to be acceptable to both patients and clinicians (and therefore actually usable), to identify effective strategies for supporting individuals in making informed choices about digital phenotyping without the risk of coercion, and in identifying user “red lines” (for example, about how data will be handled) that have practical implications for the design and cost of the technology platforms that underpin digital phenotyping.

Equity. Equity is a relevant issue in digital phenotyping for at least three reasons. First, there is a risk of excluding groups of users if underlying technology development favors certain commercial platforms or is predicated on sensing or other technologies only available in latest generation devices. For example, a majority of studies to date collect data using the Android mobile device operating system, reflecting technical challenges in enabling reliable continuous sensing on Apple devices.³⁶ In Australia, however, for example, Apple devices account for over half of mobile market share.¹⁰⁰ Addressing this disparity should be a priority, therefore. There is a strategic opportunity for the research community to proactively engage with Apple and Google not only to address salient technical challenges but also to ensure that digital phenotyping is understood as a valid (and valued) use case. Without this kind of engagement, there is a risk that unforeseen changes to privacy rules or platform software will unexpectedly disrupt the function of digital phenotyping apps.

A second issue relates to the use of machine learning as a foundational technology for translating digital phenotyping signals into usable information. Machine learning models trained in limited populations (such as college students) can demonstrate unacceptable bias in real-world applications (a problem known as “distributional shift”¹⁰¹), such as image classifiers trained on majority white populations that consistently fail in other groups.¹⁰² Selection bias has already been identified as a potential risk in digital phenotyping studies of BPD.³⁶ Consequently, digital

phenotyping applications that incorporate machine learning need to attend to evolving standards and evaluation methods designed to assure fairness in clinical machine learning.¹⁰³ These include, for example, ensuring that test/training populations have the same distributional characteristics as the populations in which the digital phenotyping will be used, attending specifically to model performance in 'protected groups' who represent minorities and those historically subject to inequity, and considering the potential for digital phenotyping to reinforce existing subtle biases in the clinical management of patients.¹⁰³ One immediate consequence for digital phenotyping research is to challenge the assumption that convenience samples are routinely good enough to condition models.

Finally, and relatedly, assumptions about how users interact with their devices may not be valid for different user groups. For example, in our experience, many adolescents have limited mobile data allowances that limit the potential for bulk data collection. Because the activation of device sensors is associated with additional power demands, limited battery capacity—or the ability to charge devices on demand—may also be relevant, for example in homeless youth.¹⁰⁴ Similarly, the assumption that smartphones are "always carried and always on" may not be valid in older populations, limiting their ability to derive value from digital phenotyping strategies that rely on continuous signals. Recognizing that there may be constraints associated with specific populations does not mean that nothing can be done. For example, in youth and other populations who are sensitive to cellular data costs, it may be feasible to configure digital phenotyping apps that wait for the availability of a wireless (i.e. no cost) data connection before attempting to transmit data for analysis. Alternatively, where data latency needs to be controlled, it may be appropriate to offer resource to covers the cost of cellular connections. Where device energy concerns exist, it may be feasible to selectively activate sensors, for example in response to contextual triggers, to reduce the total impact on battery life. There is also an opportunity for future digital phenotyping analyses to routinely model the minimal data required to generate usable signal.

Efficiency and safety. Particularly in primary care, where practitioner time and resources are constrained, it is imperative that digital phenotyping strategies can be effectively integrated into clinical workflows. This means understanding early the value that healthcare professionals will attach to different forms of information arising from digital phenotyping and anticipating practical concerns such as clinical systems integration – and upstream requirements, such as certification and data standardization.

A related concern is the validation and safety assurance process for algorithms intended for clinical use, particularly where these are based on machine learning techniques that may have subtle, hard-to-anticipate failure modes.¹⁰¹ The development of standardized approaches for documenting digital phenotyping strategies, including machine learning feature and algorithm definitions, is an open opportunity, as are approaches to validation and testing that can reliably uncover safety-relevant issues.

Opportunity 4: combining digital phenotyping with digital interventions

Digital phenotyping is anticipated to create clinical value through "closing the loop" between detecting clinical phenomena and taking action by using signals to trigger, tailor and deliver personalized digital treatment or prevention interventions.⁵ This is particularly relevant to psychiatry, where the development and adoption of both personalized treatments and digital interventions is a priority.¹⁰⁵ Digital interventions can incorporate health promotion, lifestyle education, and psychological therapies, and have a proven record in the treatment and prevention of depression and anxiety,¹⁰⁶ smoking cessation¹⁰⁷ and for the

management of diabetes,¹⁰⁸ asthma¹⁰⁹ and cardiovascular disease.¹¹⁰ Contextually enhanced eHealth interventions that tailor advice and guidance to the setting and experiences of individuals^{111,112} offer a potential avenue to reduce treatment costs¹¹³ while addressing the challenge of poor adherence seen with current digital interventions.¹¹⁴

Many existing digital phenotyping applications appear to be already intervention-like, for example integrating experience sampling as a principal data source and summarizing longitudinal data used for modelling and prediction in ways that are intended to be accessible to users. The need to package phenotyping within an app wrapper for deployment to users' smartphones creates an obvious context to extend this with intervention content that is tailored and responsive to the signals generated through digital biomarkers in order to return value to users. There are multiple ways in which this could be achieved. For example, models resulting from digital phenotyping studies could simply be integrated into future interventions and used, for example, to trigger contextual intervention content (Fig. 1a). Or, alternatively, digital phenotyping data could be used to drive online optimization,¹¹⁵ where intervention tailoring models are continuously updated (Fig. 1b).

Enhancing self-management in this way is a potentially good fit with stepped care approaches, such as those now established in the management of mood disorders,¹¹⁶ where objective data-driven guidance can be used to enhance the capacity of individuals to effectively self-manage early-stage and less severe illness. It offers a route to maximize the potential value that can be extracted from the large amounts of data that are necessary to drive digital phenotyping by creating a context where data can be visualized and used for structured self-reflection with a defined therapeutic purpose. In addition, the use of digital sensing data for features such as tailoring and soft recommendations—rather than in formal diagnostic or therapeutic processes—may be more realistic in terms of safety and technical feasibility. Because within-individual variation appears to be an effective predictor of condition onset or change in mental health,⁶⁶ closed loop interventions may be particularly valuable here. For example, future approaches could leverage Bayesian optimization¹¹⁷ to build n-of-1 predictive models tailored to the specific user. These closed-loop systems offer a new context to explore mechanisms and trajectories of illness development and treatment response. In addition, integration with digital interventions may itself create entirely new opportunities for digital phenotyping, for example, using automatically collected data about interactions with the intervention itself to generate 'engagement phenotypes' that can be subsequently used for tailoring.

Next steps

In order to respond effectively to the opportunities identified above, practical and coordinated action stands to help accelerate both research and the ultimate development of real-world health applications for digital phenotyping.

Development of shared platforms for data collection. One of the catalysts for digital phenotyping has been the research-led development of sophisticated open technology platforms such as Beiwe,¹ Purple Robot⁶² and Monsenso.¹¹⁸ Reflecting the opportunities identified we have identified above, priorities for the future development of these platforms should include addressing equity concerns by supporting Apple devices, anticipating information and clinical data governance issues as platforms move from research to practical uses (for example, ensuring that cloud-based data processing is within compatible jurisdictions), incorporating features that can expedite data validation and quality assurance, and supporting future integration with digital interventions.

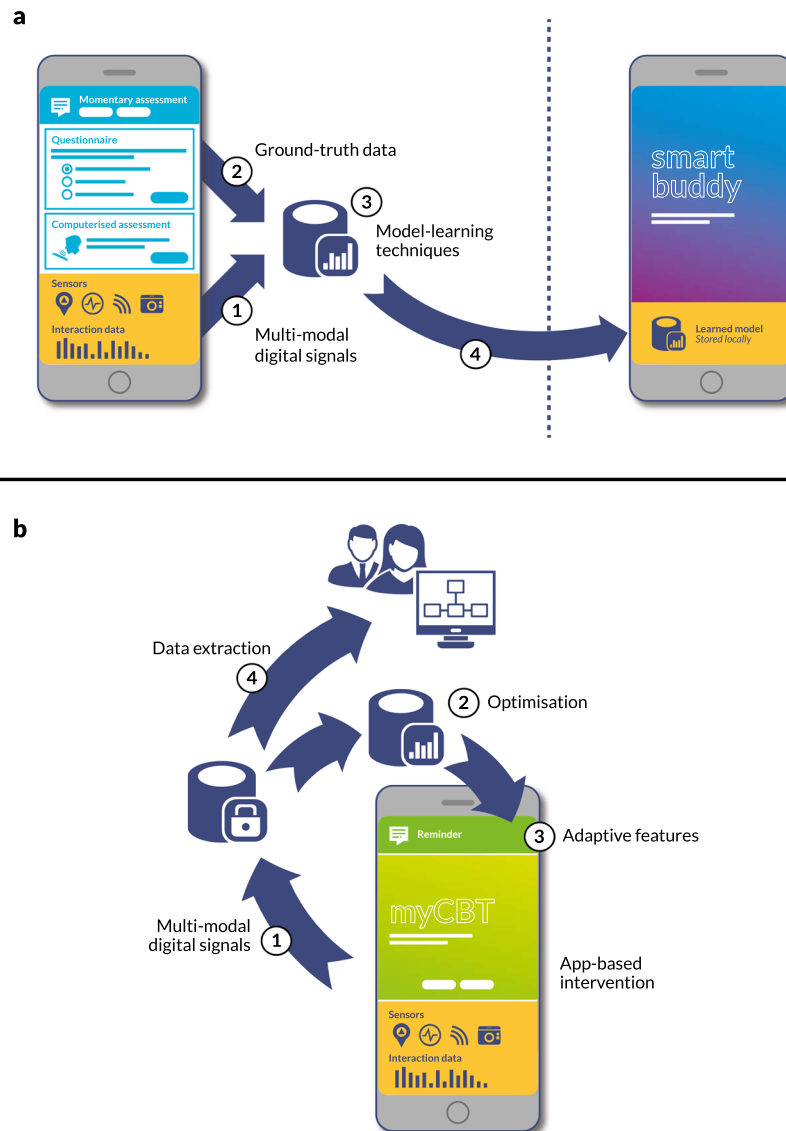


Fig. 1 Two models of integration between digital phenotyping and digital interventions. Figures and letters refer to those shown in the diagram. Model (A) describes a “learn-then-implement” approach where (1) multi-modal digital signals (e.g. sensor data) are combined with (2) ground-truth data (such as self-reported mental health) and used to learn a digital phenotyping predictive model, for example, predicting a change in mental health status from GPS and activity data. This model can then be deployed into future interventions (4) to trigger intervention components based on changes in mental health state predicted by digital signals alone. Model (B) describes a “continuous learning” approach, where (1) digital signals are automatically collected alongside intervention outcomes data. These are used to (2) continuously update and refine an intervention model conditioned on some goal, for example achieving a positive change in mental health status. This model is then used to trigger and tailor different aspects of the intervention (3). The resultant outcomes feed back into the learning process. Data collected via this approach can also be extracted for analysis (4)

Development of shared data repositories. Digital phenotyping studies typically generate rich datasets⁶⁴ which may be exploited for multiple analytical purposes. As a result, there is an opportunity to consider how these might be structured as reusable resources. For example, the development of biobanks has resulted in large numbers of research publications, clarity around researcher market needs and rapid technology development. UK Biobank,^{119,120} opened to research in 2012, provides a case study of how a single, well-designed, public resource can make a significant scientific contribution. By 2018, our bibliometric analysis suggests that research studies using UK Biobank represented nearly a tenth (9.2%, $n = 462/1727$) of annual global biobank-related publications. Of the biobank studies published in the fifty most important clinical and general science journals in

2018, well over half (57%, $n = 77/120$) used data from UK Biobank.

For digital phenotyping, potential benefits include avoiding duplication of effort, accelerating research, opening the field to a wider range of researchers beyond those already invested in digital health, and being able to pool datasets to tackle issues of statistical power and heterogeneity. Acknowledging topical interest in replicability in psychiatry, there may be specific value in the collaborative development of data standards. Coordination is relevant not only to how raw telemetry data are persisted, but also ensuring consistent acquisition of metadata that affect analysis (such as originating device types, measurement scale/precision and demographic details), socializing best practice around data cleaning and validation pipelines and, where supervised machine learning is used, documenting feature

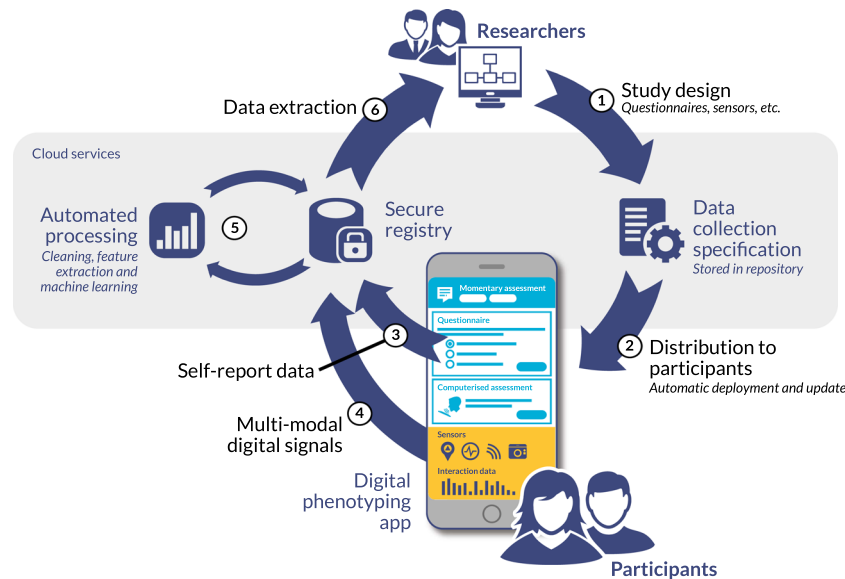


Fig. 2 Black Dog Institute/Deakin model for a scalable, integrated multi-user platform for digital phenotyping research. Figures and letters refer to those shown in the diagram. In this model, (1) researchers specify the study design, define which questionnaires and sensors are required to deliver a digital phenotyping study (and optionally how these are integrated with any intervention components, such as self-guided therapy.) This specification is then hosted alongside others in a secure online repository. When each study commences, the specification is automatically downloaded (2) to users' devices by a digital phenotyping app. This app can be a multi-study coordination tool that acts to coordinate data collection, a bespoke, study-specific data collection app, or a hybrid data collection intervention. Collected (3) self-report (e.g. questionnaires and momentary assessments) and (4) digital data (e.g. sensor measurements and device interaction data) is uploaded automatically to a secure online registry. Platform modules automatically manage potential barriers to data collection, such as user battery life and limited connectivity, through smart scheduling and caching. Automated processing pipeline (5) normalizes and converts raw data into standardized intermediate features and labelled outputs using machine learning. Researchers can start to extract registry data (6) as soon as it is received, accelerating analysis, permitting study designs that involve expert feedback, and allowing any data collection issues to be identified and addressed early in the research process. Rights management enables future researchers to request from users' access to previously-collected data

algorithms so that they can be replicated. Development of shared repositories for digital phenotyping will also require consideration of ethics issues, secure storage, linkage potential and logistical issues of exchange potentially large datasets securely across different jurisdictions. At the simplest level, collaboration could involve the development of working groups that work to harmonize data collection and reporting to expedite replication and scale-up studies. Nevertheless, we think that more significant value will be realized by being able to combine and pool data for reuse. Recognizing this opportunity, in Australia, a multi-center consortium is being assembled to develop a large-scale phenotyping databank. The Black Dog Institute and the Applied Artificial Intelligence Institute (A2I2) at Deakin University are building a hybrid data collection platform and data repository that will permit multiple primary and secondary analysis studies to be conducted on shared infrastructure (Fig. 2).

Development of multidisciplinary collaborations involving clinical disciplines, providers, patients, user-centered design, and computer science. The success of digital phenotyping is contingent on hospitals, clinicians and health companies wanting to participate in the development of useful products for their patients and health care organizations. There is an opportunity for researchers to engage with these stakeholders to better understand their priorities and needs in relation to digital phenotyping.

Equally, the willingness of patients and the public to adopt digital phenotyping technologies should not be assumed.¹²¹ Topical user-centered research questions include what expectations different user groups hold about when and how the clinical information that digital phenotyping generates should be returned to them; what ways of presenting, summarizing and guiding appropriate self-management exist that can create genuine value for users; and how different groups weight the

potential trade-offs between intrusiveness and personal health value. Given both the rapid evolution of privacy issues affecting consumer technologies and the litany of recent high profile commercial privacy breaches, finding ways to substantively represent the views of patients and the public on an ongoing basis should be seen as a strategic priority, not only to understand the boundaries of what kinds of information can be consumed by digital phenotyping but to assure that community consent exist to develop the field in the first place.

There is also a specific need to work alongside the computer science community to ensure that digital phenotyping research continues to benefit from the latest developments in machine learning, the sub-discipline of computer science concerned with the creation of algorithms and models without relying on explicit programming. Removing the need for human programming is also important for interventions, such as personalized digital therapies involving multiple treatment options, timings, individual preferences and capabilities, where the data space is too complex for humans to easily interpret (or interpret at all) and where the form of good solutions cannot be specified or predicted.¹²² Modelling techniques that can efficiently model intra-individual variation (even with sparse data) now exist and are a promising candidate for analyzing personal longitudinal tracking data. The development of explainer mechanisms and layered models may also offer new ways to interpret how individual signals are integrated into predicting clinical phenomena, with relevance for mechanistic insights into the development and evolution of clinical conditions, such as depression in young people. The opportunity to combine these refined or enhanced phenotype datasets with genetic and imaging data, along with personal, self-report and health information is likely to add value to multiple medical research disciplines and accelerate behavioral health.

CONCLUSIONS

For digital phenotyping to drive benefits in mental health and other clinical domains, serious consideration must be given to the practicalities of future clinical application. To be used, and to be useful, digital phenotyping must fit with established norms of quality and safety, be cost-effective and feasible. The research agenda that responds to these challenges will necessarily be multifaceted and multidisciplinary, spanning consumer and health stakeholder engagement, implementation science, technical development, intervention design and economic evaluation. Importantly, this call should not be interpreted as reducing the value of basic research into mechanistic or technical aspects of digital phenotyping that may not have immediate clinical applications. Nor should it discourage approaches that will necessitate changes to clinical workflows, training or patient experience.

Because many serious mental illnesses first present in youth, and because this group is an enthusiastic adopter of consumer technologies, the successful development of digital phenotyping is of specific relevance to the future, effective care of young people with psychological distress. Only a focused approach will ensure that today's young people—rather than some future generation—start to realize benefits of improved and better personalized diagnosis, monitoring and intervention.

Equally important is the development of global leadership and collaboration to tackle head on questions of trust and access to data, replicability of findings and capacity-building within clinical workforces for this new science of behavior. Because digital phenotyping stands to address genuine gaps in assessment and treatment of mental health issues, psychiatry is particularly well placed to show leadership in this newest of “big data” disciplines.

DATA AVAILABILITY

The literature search results that support this narrative review are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

This was an invited submission with a steer to explore digital phenotyping using technology. K.H. and H.C. jointly conceived the key arguments and basic structure of the paper. K.H. wrote the first draft. H.C. and K.H. reviewed and revised the draft. S.V. provided critical feedback. All authors contributed to the final version of the paper.

ADDITIONAL INFORMATION

Competing interests: H.C. is director of Black Dog Institute which develops apps and internet interventions for mental health but with no personal financial gain. H.C. stands to receive royalties as a creator of Moodgym, but to date no financial gain. H. C., S.V., and K.H. are involved in the development of the Black Dog Institute/Deakin University digital phenotyping platform described in this paper. K.H. and S.V. declare no other competing financial or non-financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Torous, J., Kiang, M. V., Lorme, J. & Onnela, J. P. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *Jmir Mental Health* **3**, e16 (2016).
2. Mohr, D. C., Zhang, M. & Schueller, S. M. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. in *Annual Review of Clinical Psychology*. Vol 13. (Widiger, T. & Cannon, T.D. eds.) 23–47 (Annual Reviews, 2017).
3. Onnela, J.-P. & Rauch, S. L. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* **41**, 1691–1696 (2016).
4. Raballo, A. Digital phenotyping: an overarching framework to capture our extended mental states. *Lancet Psychiatry* **5**, 194–195 (2018).

5. Insel, T. R. Digital phenotyping: technology for a new science of behavior. *JAMA* **318**, 1215–1216 (2017).
6. Bickman, L., Lyon, A. R. & Wolpert, M. Achieving precision mental health through effective assessment, monitoring, and feedback processes: introduction to the special issue. *Adm. Policy Ment. Health Ment. Health Serv. Res.* **43**, 271–276 (2016).
7. Morris, Z. S., Wooding, S. & Grant, J. The answer is 17 years, what is the question: understanding time lags in translational research. *J. R. Soc. Med.* **104**, 510–520 (2011).
8. Glasgow, R. E. & Emmons, K. M. How can we increase translation of research into practice? Types of evidence needed. *Annu. Rev. Public Health* **28**, 413–433 (2007).
9. Mathes, T., Jacobs, E., Morfeld, J.-C. & Pieper, D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. *BMC Health Serv. Res.* **13**, 371 (2013).
10. Place, S. et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. *J. Med. Internet Res.* **19**, e75 (2017).
11. Faherty, L. J. et al. Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy. *J. Am. Med. Inform. Assoc.* **24**, 746–753 (2017).
12. Palmius, N. et al. Detecting bipolar depression from geographic location data. *IEEE Trans. Biomed. Eng.* **64**, 1761–1771 (2017).
13. Ashok, C. K., Karunanidhi, S. & Narayanan, R. Validation of stress assessment using mobile phone. *J. Psychosoc. Res.* **11**, 479–488 (2016).
14. Gjoreski, M., Gjoreski, H., Lustrek, M. & Gams, M. Automatic detection of perceived stress in campus students using smartphones. in *Proc. 2015 International Conference on Intelligent Environments* (Weber, M. et al. eds.) 132–135 (IEEE, Prague, 2015).
15. Sano, A. Measuring college students' sleep, stress, mental health and wellbeing with wearable sensors and mobile phones. *Diss. Abstr. Intl.: Sect. B: Sci. Engi.* **78**, 117–120 (2017).
16. Sano, A. et al. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. in *Proc. 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks*. (IEEE, Cambridge, 2015).
17. Sano, A. et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. *J. Med Internet Res* **20**, e210 (2018).
18. Weissman, M. M. et al. Cross-national epidemiology of major depression and bipolar disorder. *JAMA* **276**, 293–299 (1996).
19. Truong, A. L. et al. Smartphone and online usage-based evaluation in teens (SOLVD-TEEN): can an app help teens and their parents with depression? *J. Am. Acad. Child Adolesc. Psychiatry* **56**, S216 (2017).
20. Godlewska, B. R. et al. Predicting treatment response in depression: the role of anterior cingulate cortex. *Int. J. Neuropsychopharmacol.* **21**, 988–996 (2018).
21. Rush, A. J. et al. STAR*D: revising conventional wisdom. *CNS drugs* **23**, 627–647 (2009).
22. Mundt, J. C., Vogel, A. P., Feltner, D. E. & Lenderking, W. R. Vocal acoustic biomarkers of depression severity and treatment response. *Biol. psychiatry* **72**, 580–587 (2012).
23. Gruenerbl, A. et al. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. in *Proc. 5th Augmented Human International Conference* 38 31–38:38 (ACM, Kobe, 2014).
24. Faurholt-Jepsen, M. et al. Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. *Int. J. Methods Psychiatr. Res.* **25**, 309–323 (2016).
25. Faurholt-Jepsen, M. et al. Smartphone data as objective measures of bipolar disorder symptoms. *Psychiatry Res.* **217**, 124–127 (2014).
26. Stange, J. et al. Convergence of active and passive assessments of affective instability in predicting the prospective course of bipolar disorder: the bi affect study. *Neuropsychopharmacology* **43**(Supplement 1), S164 (2017).
27. Zulueta, J. et al. Predicting mood disturbance severity in bipolar subjects with mobile phone keystroke dynamics and metadata. *Biol. Psychiatry* **81**, S195–S196 (2017).
28. Faurholt-Jepsen, M. et al. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. *Bipolar Disord.* **17**, 715–728 (2015).
29. Faurholt-Jepsen, M. et al. Voice analysis as an objective state marker in bipolar disorder. *Transl. Psychiatry* **6**, e856 (2016).
30. Tohen, M., Waternaux, C. M. & Tsuang, M. T. Outcome in mania. A 4-year prospective follow-up of 75 patients utilizing survival analysis. *Arch. Gen. psychiatry* **47**, 1106–1111 (1990).
31. Perry, A., Tarrier, N., Morriss, R., McCarthy, E. & Limb, K. Randomised controlled trial of efficacy of teaching patients with bipolar disorder to identify early

- symptoms of relapse and obtain treatment. *BMJ (Clin. Res. ed.)* **318**, 149–153 (1999).
32. Biskin, R. S. The lifetime course of borderline personality disorder. *Can. J. psychiatry Rev. Can. de. Psychiatr.* **60**, 303–308 (2015).
 33. Nicholas, J., Boydell, K. & Christensen, H. Beyond symptom monitoring: consumer needs for bipolar disorder self-management using smartphones. *Eur. Psychiatry* **44**, 210–216 (2017).
 34. Morriss, R. K. et al. Interventions for helping people recognise early signs of recurrence in bipolar disorder. *Cochrane Database Syst. Rev.* **2007**, CD004854 (2007).
 35. Faurholt-Jepsen, M., Frost, M., Bardram, J. E. & Kessing, L. V. Smartphone based treatment in bipolar disorder. *Eur. Psychiatry* **33**(Supplement), S32–S33 (2016).
 36. Faurholt-Jepsen, M., Bauer, M. & Kessing, L.V. Smartphone-based objective monitoring in bipolar disorder: status and considerations. *Intl. J. Bipolar Disord.* **6**, 6 (2018).
 37. Faurholt-Jepsen, M. et al. Reducing the rate and duration of Re-ADMISSions among patients with unipolar disorder and bipolar disorder using smartphone-based monitoring and treatment - the RADMIS trials: study protocol for two randomized controlled trials. *Trials* **18**, 277 (2017).
 38. Roxburgh, A., Dobbins, T., Degenhardt, L. & Peacock, A. *Opioid, amphetamine, and cocaine-induced deaths in Australia: August 2018*. (National Drug and Alcohol Research Centre, UNSW, Sydney, 2018).
 39. Nandakumar, R., Gollakota, S. & Sunshine, J. E. Opioid overdose detection using smartphones. *Sci. Transl. Med.* **11**, eaau8914 (2019).
 40. Giglio, R. E., Li, G. & DiMaggio, C. J. Effectiveness of bystander naloxone administration and overdose education programs: a meta-analysis. *Inj. Epidemiol.* **2**, 10 (2015).
 41. Russolillo, A., Moniruzzaman, A. & Somers, J. M. Methadone maintenance treatment and mortality in people with criminal convictions: a population-based retrospective cohort study from Canada. *PLOS Med.* **15**, e1002625 (2018).
 42. Byrnes, H. F. et al. Brief report: using global positioning system (GPS) enabled cell phones to examine adolescent travel patterns and time in proximity to alcohol outlets. *J. Adolesc.* **50**, 65–68 (2016).
 43. Byrnes, H. F. et al. Association of environmental indicators with teen alcohol use and problem behavior: Teens' observations vs. objectively-measured indicators. *Health Place* **43**, 151–157 (2017).
 44. Byrnes, H. F. et al. Tracking adolescents with global positioning system-enabled cell phones to study contextual exposures and alcohol and marijuana use: a pilot study. *J. Adolesc. Health* **57**, 245–247 (2015).
 45. Byrnes, H. F. et al. Using GPS-EMA techniques to examine contextual exposures in activity spaces vs residential areas: relations with teen AOD and problem behavior. *Alcohol: Clin. Exp. Res.* **41**, 171A (2017).
 46. Byrnes, H. F. et al. Presence and characteristics of alcohol outlets perceived during daily travels: relationships with teen alcohol use, attitudes, and access. *Alcohol: Clin. Exp. Res.* **42**, 54A (2018).
 47. Boyle, S. C. The social mindfeed project: using objective assessment methods to better understand the nature of social-media based peer alcohol influence. *Alcohol: Clin. Exp. Res.* **42**, 280A (2018).
 48. Bae, S., Chung, T., Ferreira, D., Dey, A. K. & Suffoletto, B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addict. Behav.* **83**, 42–47 (2018).
 49. Santani, D. et al. DrinkSense: characterizing youth drinking behavior using smartphones. *Ieee Trans. Mob. Comput.* **17**, 2279–2292 (2018).
 50. O'Dea, B., Larsen, M., Batterham, P., Calear, A. & Christensen, H. Talking suicide on Twitter: linguistic style and language processes of suicide-related posts. *Eur. Psychiatry* **33**, S274 (2016).
 51. O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L. & Christensen, H. A linguistic analysis of suicide-related Twitter posts. *Crisis* **38**, 319–329 (2017).
 52. O'Dea, B. et al. Detecting suicidality on Twitter. *Internet Interv.* **2**, 183–188 (2015).
 53. WHO. *Preventing suicide: a global imperative*, (Stylus Publishing, 2014).
 54. Canuso, C. M. et al. Efficacy and safety of intranasal esketamine for the rapid reduction of symptoms of depression and suicidality in patients at imminent risk for suicide: results of a double-blind, randomized, placebo-controlled study. *Am. J. psychiatry* **175**, 620–630 (2018).
 55. Vahabzadeh, A., Sahin, N. & Kalali, A. Digital suicide prevention: can technology become a game-changer? *Innov. Clin. Neurosci.* **13**, 16–20 (2016).
 56. Wilson, S. T. et al. Heart rate variability and suicidal behavior. *Psychiatry Res* **240**, 241–247 (2016).
 57. Wang, T., Azad, T. & Rajan, R. *The emerging influence of digital biomarkers on healthcare*. (Rock Health, San Francisco, 2016).
 58. Torous, J., Staples, P. & Onnela, J.-P. Realizing the potential of mobile mental health: new methods for new data in psychiatry. *Curr. Psychiatry Rep.* **17**, 61 (2015).
 59. Kamath, J. et al. Prediction of clinical depression using smartphone sensory data. *Neuropsychopharmacology* **41**, S536–S537 (2016).
 60. Saeb, S. et al. The relationship between clinical, momentary, and sensor-based assessment of depression. in *Proc. 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)* 229–232 (ICST Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2015).
 61. Frank, E. et al. Sensing depression: Using smartphone sensors to predict changes in depression severity. *Neuropsychopharmacology* **43**, S346 (2017).
 62. Saeb, S. et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J. Med. Internet Res.* **17**, e175 (2015).
 63. Farhan, et al. Behavior vs. Introspection: refining prediction of clinical depression via smartphone sensing data. in *Proc. 2016 IEEE Wireless Health (WH)* 1–8 (IEEE, Bethesda, 2016).
 64. Aung, H. et al. Continuous behavioral data as a depression biomarker. *Neuropsychopharmacology* **41**, S488–S489 (2016).
 65. Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P. & Mohr, D. C. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ* **4**, e2537 (2016).
 66. Barnett, I. et al. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology* **43**, 1660–1666 (2018).
 67. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association, Washington, DC, 2013).
 68. McDermott, L. M. & Ebmeier, K. P. A meta-analysis of depression severity and cognitive function. *J. Affect. Disord.* **119**, 1–8 (2009).
 69. Rock, P. L., Roiser, J. P., Riedel, W. J. & Blackwell, A. D. Cognitive impairment in depression: a systematic review and meta-analysis. *Psychol. Med.* **44**, 2029–2040 (2014).
 70. Lee, R. S. C., Hermens, D. F., Porter, M. A. & Redoblado-Hodge, M. A. A meta-analysis of cognitive deficits in first-episode major depressive disorder. *J. Affect. Disord.* **140**, 113–124 (2012).
 71. Cha, D. S. et al. Perceived sleep quality predicts cognitive function in adults with major depressive disorder independent of depression severity. *Ann. Clin. psychiatry : Off. J. Am. Acad. Clin. Psychiatr.* **31**, 17–26 (2019).
 72. Vicent-Gil, M. et al. Cognitive predictors of illness course at 12 months after first-episode of depression. *Eur. Neuropsychopharmacol.* **28**, 529–537 (2018).
 73. Bortolato, B., Carvalho, A. F. & McIntyre, R. S. Cognitive dysfunction in major depressive disorder: a state-of-the-art clinical review. *CNS Neurol. Disord. drug targets* **13**, 1804–1818 (2014).
 74. Roiser, J. P., Elliott, R. & Sahakian, B. J. Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology* **37**, 117–136 (2012).
 75. Dawson, E. L. et al. Executive functioning at baseline prospectively predicts depression treatment response. *Prim. Care Companion J. Clinical Psychiatry* **19**, 16m01949 (2017).
 76. John, A., Patel, U., Rusted, J., Richards, M. & Gaysina, D. Affective problems and decline in cognitive state in older adults: a systematic review and meta-analysis. *Psychol. Med.* **49**, 353–365 (2019).
 77. Simon, G. E. & Perlis, R. H. Personalized medicine for depression: can we match patients with treatments? *Am. J. psychiatry* **167**, 1445–1455 (2010).
 78. Kolla, B. P., Mansukhani, S. & Mansukhani, M. P. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert Rev. Med. Devices* **13**, 497–506 (2016).
 79. Williams, L. M. et al. The ENGAGE study: integrating neuroimaging, virtual reality and smartphone sensing to understand self-regulation for managing depression and obesity in a precision medicine model. *Behav. Res. Ther.* **101**, 58–70 (2018).
 80. Bagot, K. S. et al. Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. *Dev. Cogn. Neurosci.* **32**, 121–129 (2018).
 81. Faurholt-Jepsen, M. et al. Objective smartphone data as a potential diagnostic marker of bipolar disorder. *Aust. N.Z. J. Psychiatry* **53**, 119–128 (2019).
 82. Dobrow, M. J., Hagens, V., Chafe, R., Sullivan, T. & Rabeneck, L. Consolidated principles for screening based on a systematic review and consensus process. *CMAJ : Can. Med. Assoc. J.—J. de. l'Assoc. Med. Can.* **190**, E422–E429 (2018).
 83. Lewinsohn, P. M., Klein, D. N. & Seeley, J. R. Bipolar disorder during adolescence and young adulthood in a community sample. *Bipolar Disord.* **2**, 281–293 (2000).
 84. Vickers, A. J., Van Calster, B. & Steyerberg, E. W. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ (Clin. Res. ed.)* **352**, i6 (2016).
 85. Dagum, P. Digital biomarkers of cognitive function. *npj Digital Med.* **1**, 10 (2018).
 86. Sheridan, K. Mindstrong's mood-predicting app is shadowed by questions over evidence. *STAT* <https://www.statnews.com/2018/10/04/mindstrong-questions-over-evidence/> (2018).
 87. Adams, W. R. High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. *PLOS ONE* **12**, e0188226 (2017).

88. Stocchi, F., Vacca, L. & Radicati, F. G. How to optimize the treatment of early stage Parkinson's disease. *Transl. Neurodegener.* **4**, 4–4 (2015).
89. Hustad, E., Skogholt, A. H., Hveem, K. & Aasly, J. O. The accuracy of the clinical diagnosis of Parkinson disease. The HUNT study. *J. Neurol.* **265**, 2120–2124 (2018).
90. Schrag, A., Ben-Shlomo, Y. & Quinn, N. How valid is the clinical diagnosis of Parkinson's disease in the community? *J. Neurol., Neurosurg. Psychiatry* **73**, 529–534 (2002).
91. Pagan, F. L. Improving outcomes through early diagnosis of Parkinson's disease. *Am. J. Manag. Care* **18**, S176–182 (2012).
92. Goldman, J. G. et al. Cognitive impairment in Parkinson's disease: a report from a multidisciplinary symposium on unmet needs and future directions to maintain cognitive health. *npj Park. s. Dis.* **4**, 19 (2018).
93. Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. National Academy Press, Washington, DC, USA, 2001).
94. Demain, S. et al. Living with, managing and minimising treatment burden in long term conditions: a systematic review of qualitative research. *PLOS ONE* **10**, e0125457–e0125457 (2015).
95. Dennison, L., Morrison, L., Conway, G. & Yardley, L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *J. Med. Internet Res.* **15**, e86 (2013).
96. Bauer, M. et al. Ethical perspectives on recommending digital technology for patients with mental illness. *Int. J. bipolar Disord.* **5**, 6–6 (2017).
97. Huckvale, K., Torous, J. & Larsen, M. E. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. *JAMA Netw. Open* **2**, e192542–e192542 (2019).
98. Nicholas, J. et al. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: Cross-sectional questionnaire study. *JMIR Mhealth Uhealth* **7**, e12578 (2019).
99. Torous, J., Rodriguez, J. & Powell, A. The new digital divide for digital biomarkers. *Digit. Biomark.* **1**, 87–91 (2017).
100. StatCounter. Mobile operating system market share in Australia - January 2019. <http://gs.statcounter.com/os-market-share/mobile/australia> (2019).
101. Challen, R. et al. Artificial intelligence, bias and clinical safety. *BMJ Qual. Saf.* **28**, 231–237 (2019).
102. Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V. & Jain, A. K. Face recognition performance: role of demographic information. *IEEE Trans. Inf. Forensics Secur.* **7**, 1789–1801 (2012).
103. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. Ensuring fairness in machine learning to advance health equity. *Ann. Intern. Med.* **169**, 866–872 (2018).
104. Adkins, E. C. et al. Exploring the potential of technology-based mental health services for homeless youth: A qualitative study. *Psychol. Serv.* **14**, 238–245 (2017).
105. Holmes, E. A. et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry* **5**, 237–286 (2018).
106. Deady, M. et al. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. *BMC Psychiatry* **17**, 310–310 (2017).
107. Do, H. P. et al. Which eHealth interventions are most effective for smoking cessation? A systematic review. *Patient Prefer. adherence* **12**, 2065–2084 (2018).
108. Kitsiou, S., Paré, G., Jaana, M. & Gerber, B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. *PLOS ONE* **12**, e0173160 (2017).
109. Jeminiwa, R. et al. Impact of eHealth on medication adherence among patients with asthma: a systematic review and meta-analysis. *Respir. Med.* **149**, 59–68 (2019).
110. Carbo, A. et al. Mobile technologies for managing heart failure: a systematic review and meta-analysis. *Telemed. e-Health* **24**, 958–968 (2018).
111. Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M. & Weidt, S. Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR MHealth UHealth* **4**, e111 (2016).
112. Weidt, S., Wahle, F., Rufer, M., Horni, A. & Kowatsch, T. MOSS—mobile sensing and support detection of depressive moods with an app and help those affected. *Ther. Umsch.* **72**, 553–555 (2015).
113. McCrone, P. et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *Br. J. psychiatry : J. Ment. Sci.* **185**, 55–62 (2004).
114. Torous, J., Nicholas, J., Larsen, M. E., Firth, J. & Christensen, H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid. Based Ment. Health* **21**, 116–119 (2018).
115. Hoi, S.C., Sahoo, D., Lu, J. & Zhao, P. Online learning: a comprehensive survey. in *arXiv preprint arXiv:1802.02871*. (2018).
116. Ho, F. Y.-Y., Yeung, W.-F., Ng, T. H.-Y. & Chan, C. S. The efficacy and cost-effectiveness of stepped care prevention and treatment for depressive and/or anxiety disorders: a systematic review and meta-analysis. *Sci. Rep.* **6**, 29281–29281 (2016).
117. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & Freitas, Nd Taking the human out of the loop: a review of bayesian optimization. *Proc. IEEE* **104**, 148–175 (2016).
118. Faurholt-Jepsen, M. et al. Daily electronic monitoring of subjective and objective measures of illness activity in bipolar disorder using smartphones—the MON-ARCA II trial protocol: a randomized controlled single-blind parallel-group trial. *BMC Psychiatry* **14**, 309 (2014).
119. Allen, N. et al. UK Biobank: current status and what it means for epidemiology. *Health Policy Technol.* **1**, 123–126 (2012).
120. Collins, R. What makes UK Biobank special? *Lancet (Lond., Engl.)* **379**, 1173–1174 (2012).
121. Huckvale, K., Wang, C. J., Majeed, A. & Car, J. Digital health at fifteen: more human (more needed). *BMC Med.* **17**, 62 (2019).
122. Alanazi, H. O., Abdullah, A. H. & Qureshi, K. N. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. *J. Med. Syst.* **41**, 69 (2017).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019