



Citation for published version:

Davidson, B 2022, 'The Crossroads of Digital Phenotyping', *General Hospital Psychiatry*, vol. 74, pp. 126-132.
<https://doi.org/10.1016/j.genhosppsy.2020.11.009>

DOI:

[10.1016/j.genhosppsy.2020.11.009](https://doi.org/10.1016/j.genhosppsy.2020.11.009)

Publication date:

2022

Document Version

Peer reviewed version

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The Crossroads of Digital Phenotyping

Brittany I. Davidson, Ph.D.^{1,2*}

¹Information, Decisions, and Operations Division, School of Management,
University of Bath

²Department of Computer Science, University of Bristol

Abstract

The term ‘Digital Phenotyping’ has started to appear with increasing regularity in medical research, especially within psychiatry. This aims to bring together digital traces (e.g., from smartphones), medical data (e.g., electronic health records), and lived experiences (e.g., daily activity, location, social contact), to better monitor, intervene, and diagnose various psychiatric conditions. However, is this notion any different from digital traces or the quantified self? While digital phenotyping has the potential to transform and revolutionize medicine as we know it; there are a number of challenges that must be addressed if research is to blossom. At present, these issues include; (1) methodological issues, for example, the lack of clear theoretical links between digital markers (e.g., battery life, interactions with smartphones) and condition relapses, (2) the current tools being employed, where they typically have a number of security or privacy issues, and are invasive by nature, (3) analytical methods and approaches, where I question whether research should start in larger-scale epidemiological scale or in smaller (and potentially highly vulnerable) patient populations as is the current norm, (4) the current lack of security and privacy regulation adherence of apps used, and finally, (5) how do such technologies become integrated into various healthcare systems? This aims to provide deep insight into how the Digital Phenotyping could provide huge promise if we critically reflect now and gather clinical insights with a number of other disciplines such as epidemiology, computer- and the social sciences to move forward.

1 Introduction

The relatively recent rush of ‘*Digital Phenotyping*’ promises to capture a variety of objective data-streams, typically from smartphones, alongside patient’s lived experiences¹⁻³. In theory, this would reveal patterns of behaviour unique to the patient, which provides a more ‘complete’ picture, whereby helping clinicians to detect early warning signs of conditions. Currently, these methods and approaches are primarily being tested within psychiatry^{2,4-8}. Digital Phenotyping may allow us to see a patient’s pattern of behaviour, from which, we could infer when they are struggling with specific symptoms of interest¹, typically this relates to mental health outcome, including schizophrenia, bipolar, and various depressive disorders^{4,5,7-10} in the current literature. Hence, being able to capture fluctuations in digital behaviours provide clinicians additional insight into patients¹. For instance, a sudden drop in overall smartphone usage may be a sign of social withdrawal¹¹ or a shift towards or increase in usage of a smartphone at night may be a sign of erratic sleeping patterns, which could be useful contextual information for medical professionals. Having a deeper knowledge of patients and their daily patterns and routines could therefore provide previously unknown touchpoints for interventions, the detection of worsening symptoms, or an ability to accurately predict relapses⁸. However, we must remember, these digital traces can be highly revealing, where studies have shown we can predict a variety of demographics and personality traits from digital traces¹²⁻¹⁴, which requires great consideration regarding the ethics and standardisation as to how to appropriately utilize these data sources.

2 Digital what?

On the outset, digital phenotyping sounds revolutionary and therefore could change medicine as we know it. However, this concept has been seen previously. The use of digital technologies to better understand individuals, groups, or society remains a key topic across many disciplines. As a result, we see a number of terms generating hype, fear, or overpromising within siloed fields¹⁵⁻¹⁷. For example, the terms: ‘*digital trace*’^{18,19} or ‘*digital footprint*’²⁰, which are common in (digital) humanities, social and computer sciences, rose alongside the ‘*digitized*’²¹ or ‘*quantified*’ self²² (Table 1).

Table 1. Definition and examples of terms preceding ‘Digital Phenotyping’

Term	Definition	Examples
Digital trace	‘Digital records of activity and events that involve information technologies,’ ²³ .	Data from smartphones ^{24,25} , online
Digital footprint	‘The digital traces each one of us leaves behind as we conduct our lives,’ ²⁰	forums ^{26,27} , fitness trackers ^{28,29} or pedometers ^{30,31} , sociometric badges ^{32,33} ,
Digitized self	The self translated into a digital space ²¹	devices to check air quality (e.g.,
Quantified self	‘Any individual engaged in the self-tracking of any kind of biological, physical, behavioral, or environmental information,’ ²²	particulate matter) ²⁸

Generally speaking, all of these ideas are underpinned by the notion of passive sensing of individuals, whereby an app, sensor, or device is passively monitoring and capturing a variety of metrics (e.g., movement, interaction, etc.). This passive sensing or capturing of metrics are typically used to better understand individuals, for example, can we predict demographics, gender, (offline) behaviour, or personality?^{12–14}, which has a number of applications from business analytics, security, to health and medicine. More recently, medicine has coined a ‘new’ term under the guise of ‘*Digital Phenotyping*’, which appears to promise something similar: the ability to track biological, physical, behavioural, and external environmental data, which may be captured via smartphones or (passive) sensors (e.g. fitness trackers)^{22,34}. These data streams can then be linked to medical datasets (e.g., patient health records) or potentially ‘personal genomic’ datasets (e.g., 23andMe)²², which comes together as a form of ‘precision medicine’^{22,35,36}. Others within the technology—medical overlap have also suggested the use of the term ‘personal sensing’³.

3 The Promise

By utilising our moment-by-moment interactions with a device, typically smartphones in the current literature, we will be able to have deep understanding of people’s health-related behaviours³⁷. Equally, this will reveal interactions and behaviors that are uniquely human¹⁶. For example, via these digital traces, we will see nuanced patterns of behaviour that remain largely unknown and under-researched revealing habits and routines patients are likely unaware of. The idea would be to analyze digital traces from devices (e.g., usage logs, location, movement, app usage), engaging with ecological momentary assessment (EMA) (where

individuals are prompted to answer questions, e.g., ‘*how are you feeling today?*’ or record survey responses), and linking with a number of other datasets (e.g., census, health records, government data such as air pollution data, in-depth demographics like socioeconomic status). This will generate extremely rich datasets of individuals^{2,34,37}. In theory, the promise of digital phenotyping would be substantial for understanding of people (and society) at all levels: individual, group, and societal, especially as data would allow for time series analysis, whereby it would be possible to monitor public health phenomena among other measures of interest. Specifically, relating to health research, one of the key promises is to help those who are *most* vulnerable (e.g., those suffering with acute mental health conditions). Hence, digital phenotyping promises to deliver ‘just-in-time’ interventions alongside passive monitoring³⁷.

4 The Reality

4.1 Apps, Frameworks, and Tools

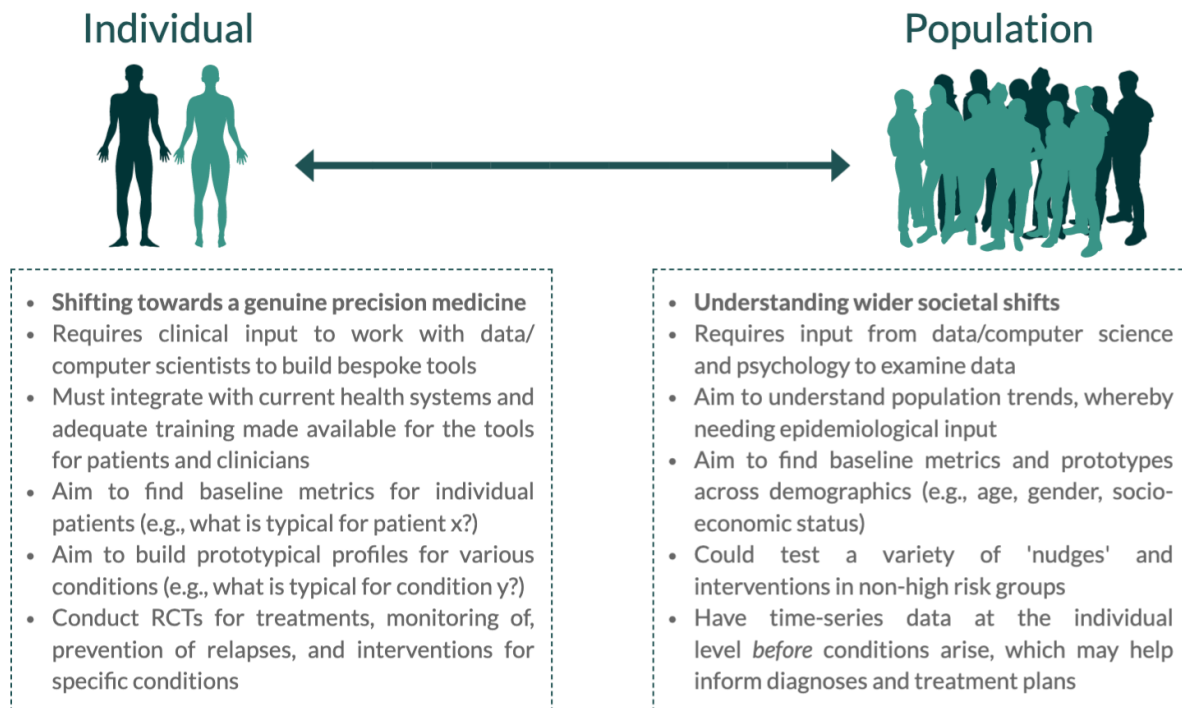
Digital phenotyping has a great promise to hold, yet a distinct definition or approach remains poorly defined. One of the key elements for such research is an app, tool, or framework in order to capture the data required¹. While there are many ‘medical’ apps^{42,43}, the uptake and adoption of these apps is often limited⁴⁴ and the success of such apps remains largely unknown. Using digital trace data from apps for applied medical purposes while not a strictly new concept; there remain a number of challenges. For example, digital trace data will greatly vary in quality, where errors can creep in at all stages of the research process, from the initial construct definitions for measurement, through to errors in data cleaning, analysis, and inferences, which relate more to the human-aspects of data analytics⁴⁵. However, these issues undoubtedly occur on the technical side of these apps, for example an app might fail *in situ* or there may be questions as to how an app collects such data (unless the researchers themselves built the app employed²³), additionally there may be measurement differences from device to device that are not necessarily easily explained⁴⁵. Hence, data quality is critical for modelling any medical outcomes, where it does not matter how sophisticated a model may be, if the data quality lacks due to the lack of understanding of the digital trace data collection or nonresponses from participants creating sparse data, our predictions or outcomes will be meaningless (‘*garbage in, garbage out,*’)^{23,45,46}.

¹ Examples of apps for capturing digital traces: PEG Log³⁸, Usage Logger²⁴, AWARE Framework³⁹, Funf in a Box⁴⁰, Contact Logger⁴¹.

4.2 Precision versus Population Perspective

Digital Phenotyping points towards the notion of precision medicine (defined as the ‘prevention and treatment strategies that take individual variability into account’⁴⁷, which would focus on N=1 samples), as well as the ability to encompass data from a number of databases (e.g., medical records), and devices (e.g., smartphones, fitness devices). This can, in theory, help diagnose conditions early, monitor conditions, and deploy ‘just-in-time’ interventions. From the outset, this is a vast set of goals to reach, all of which require different approaches to reach. For instance, there is an argument as to whether research would begin with a *general population* or within *clinical populations* (Figure 1). It is important to note that the current literature base (specifically using the term digital phenotyping) relies on data from extremely vulnerable patients (e.g., those with schizophrenia, bipolar, among other psychiatric disorders^{4,5,7,9,10}) rather than those from the general population. It is interesting as there is often little to no explanation as to why. It is plausible that the rationale assumes that these patients would have more obvious fluctuations in mood, and therefore, one might gather that changes in mood, behaviour, or wellbeing will be more easily detected and predicted¹⁰. However, this is problematic as if something falls awry as this may have devastating consequences that we are less likely to see within the general population. There are typically the high drop-out rates within patient samples, which not only causes issues in both the statistical analysis conducted, but more importantly, questions whether invasive studies such as these are suitable for those who are incredibly vulnerable. One might ask whether secondary data analyses would be more suitable for understanding patients with severe mental illness rather than active and invasive procedures. If digital phenotyping pursues clinical populations, it is important that new research also focuses on other kinds of patients in order to widen perspective and approaches to digital phenotyping and medical sensing or monitoring (e.g., diabetes monitoring via smartphones^{48,49}), which would likely cross-fertilize and aid development of research designs and protocols⁴⁴.

Figure 1. Precision Medicine Approach versus Population-Level Approach



Conversely, if the research starts with the general population, this would enable us to create a variety of 'baseline' measures across the population, noting demographics (e.g., age, socio-demographics, etc), health data (e.g., fitness tracking, clinical data, previous conditions, medications, etc.), and 'lived' experiences (e.g., EMA, however one should note, there is also a question as to what 'lived' experience data would be of interest). This would enable theory and metric development for general health and wellbeing (e.g., what metrics best predict young persons' fitness? How can we increase office-workers daily movement? Or, presenting a theory of digital health, whereby we consider *how* people use technologies for health monitoring, for example). Theory could focus on specific conditions (e.g., what do worsening symptoms look like digitally for *x* condition? How can we develop personalized sensors that track specific symptoms for conditions?). Over time, this would allow us to have datasets containing the history of people *before* a diagnosis. Hence, with a variety of metrics captured over time, in theory, this would enable a retrospective analysis to see what lead to a patient being diagnosed with *x*, which could be used as training data for predictive modelling for others. However, research is yet to take this angle and perspective.

However, similar to the current work, starting in clinical populations may also be useful. Here, it would be sensible to start a sample size of $N=1$ (however, research is yet to do this), where

each patient is a ‘case study’ in their own right. Here, potentially less data could be captured as it would be an idea to measure the symptoms relating to their illness and others of interest to capture. The ‘lived’ experience could capture highly insightful data to link to their captured digital data that aims to measure certain symptoms and behaviors associated with symptoms. With careful monitoring, patient-specific ‘baselines’ can be detected, whereby specific anomaly detection algorithms could be trained for them, which would aim to reduce problems with noisy data (as seen in Barnett et al’s⁸ work with schizophrenia patients). A key reason to be cautious with anomaly detection and wider machine learning classification problems relates to false positives or false negatives, which can have devastating consequences. The distinction here would be: can we reliably and accurately distinguish between a mood fluctuation and a relapse or change in symptoms⁵⁰? Hence, starting this research within the general population to formulate a number of reliable metrics for mood, behavior, health outcome prediction is undoubtedly a safer route, which will provide findings that other areas of health and medicine can draw and build upon.

4.3 *Metrics, Measurement, and Theory*

Delving deeper into the realities and practicalities of digital phenotyping, there is a question as to what data is needed and *how* this data will be gathered. For instance, in order to capture a digital trace, an app will be required on the smartphone to gather such data, alongside scripts to pre-process, and analyse the data. To have apps that capture the specific types of data desired ethically by ensuring privacy and security for users is not necessarily simple both technically in terms of ensuring apps are secure and private by design, but also when considering the personal and invasive data being collected that can reveal a lot about an individual (e.g., daily routines, when they are at home versus at work, etc.). Further, at this stage in research, there is no defined approach as to how to conduct digital phenotyping in terms of conceptualisation of research questions or hypotheses or measurement or how inferences will be made based on the data collected. Further, there remains no discussion of open science practices⁵¹ (e.g., open source apps, data sharing, preregistration), which is important for big data analytics particularly in applied settings to promote transparency, openness, and reproducibility.

Current apps for capturing digital traces can be augmented with EMA or lived experiences, where we can analyse links between online interactions and other factors of interest (e.g., mood, exercise, symptom tracking). However, many apps collect some form of digital trace data alongside offering medical advice or clinical input (e.g., therapy sessions). For example,

‘mindstrong’⁵², which measures ‘taps, scrolls, clicks and other activities on the smartphone touchscreen’⁵³ as their ‘digital biomarkers’ to correlate against mental health scales. While this approach has great potential, it is too early to make predictions regarding interventions without deeper understanding of how digital interactions link to offline behaviors or attitudes. This would need to come from both large-scale studies (data science and epidemiology approaches) as well as at the individual level (precision, within-subjects approaches) (Figure 1).

Specifically focusing on metrics, one could argue to take entire traces of phones, fitness devices, among others. However, first, a viable research question is needed to inform what data is required and clinical input would be critical to inform this. We must move away from the notion of collecting any and all data from devices to then only consider research questions and analysis plans after data are collected as it is unethical to collect data for the sake of it. A way forward would be to push for preregistration of digital phenotyping studies on the Open Science Framework⁵¹ (or similar), where all research questions, hypotheses, data collection protocols, and analysis plans would be documented *before* data collection took place in order to avoid ‘data mining’. Additionally, it is critically important for us to have a solid, reliable, and accurate understanding of these measures^{23,45}. For example, mobiles or smartwatches (worn to bed) could provide estimates of sleeping patterns, movement or activity levels, and location, which could reveal insight into mood or personality. However, there will be a question as to how accurate and reliable are these devices or apps at logging various everyday interactions or uses^{3,38,54?}

From a theoretical or top-down perspective, if the interest is looking at depressive mood symptoms in the general population, we might want to consider what these symptoms look like behaviourally (offline) and consider how this might map onto digital interactions. Teams should work with clinicians to sense check and seek advice as to how to map digital behaviors to offline symptoms of depression. For instance, some work has shown overall screen time to *reduce* during worsening symptoms of depression as it is a form of social withdrawal¹¹. We might also analyze location, where we may expect participants to spend more time at home and not going out to socialize at the weekend or in the evenings. Having some form of theoretical mechanism to examine is critically important to avoid ‘data mining’ in these circumstances due to working with vulnerable patients or simply, extremely sensitive data (even more so when this is linked to other sensitive data). It is not uncommon to see research with >100 variables^{4,55} being modelled to predict an outcome or mood severity, rather than focusing on

few, but specific measures/metrics. By focusing on specific digital interactions and outcomes (e.g., behaviors, clinical outcomes of interest), this encourages ‘behavioral analytics’^{26,50,56}, where we have underlying mechanisms or theories to understand *why* variable *x* relates to symptom *y*. For example, having a constant measure of someone’s smartphone charge *could* be of interest as a proxy for risk taking, however, as this is not validated nor tested. Hence, groundwork must be completed to verify potential associations. This type of analysis also relates to a number of issues regarding the inferences that may be made about users and the (likely) lack of control patients will have over these inferences that might impact them in terms of clinical decision-making or treatment⁵⁷. Hence, valid and reliable measures and understanding of measures is critical particularly in applied medical settings.

We must also note, this will be an iterative process of learning. Digital phenotyping will demonstrate that some metrics will be predictive of for some, but not others. Health and medicine is incredibly complex, which complicates modelling and prediction⁴⁶. However, prediction techniques may improve when combining a variety of metrics and data sources. This by nature, returns to the point of what makes a suitable metric (and baseline) for individuals, groups, or the population? Since the current research almost exclusively relies on smartphones, which contain a number of sensors, but is the data quality arising from these devices and sensors detailed and reliable enough for precision medicine? Do we need to build specific sets of sensors for subsets of the population or those with particular conditions? These questions are critically important to the future of digital phenotyping and how this field would want to proceed. It requires incremental research across the general population and single case studies (N=1) for granular approaches.

Finally, there are numerous questions as to whether the patients can see the day-to-day measurement from devices or whether any or all patients would receive feedback. While there are many examples of feedback being incredibly useful or necessary (e.g., with diabetics monitoring sugar levels), this may cause stress and worry if patients see unexpected changes (e.g., weight monitoring, number of steps taken/distance walked)⁵⁸. Therefore, we must also consider bespoke devices for the individual, where we return to the notion of personal and precision medicine (LHS Figure 1). Hence, there are a number of research avenues to follow in regard to what the patient sees, how this is monitored, who analyses this data, however there is a distinct lack of protocols and regulations in place for if, or perhaps when, these methods and apps go wrong.

4.4 *Reporting of Findings*

While early findings and pilot studies may look promising, the reporting of findings must remain balanced and limitations stated fully. For instance, it is relatively common to see very small sample sizes across the current literature using the term digital phenotyping (e.g., $N \sim 19$). This is not inherently problematic, as we note that this is perhaps due to these samples using specific groups of patients with rare and acute disorders. However, it is problematic that the statistics run across data collected are not built for such underpowered studies or violate a number of underlying assumptions of statistical inferences^{59,60}. When handling to extremely small sample sizes and underpowered studies, additional caution must be taken when reporting these findings if the researchers do not actively address these statistical inference issues within data analysis. For example, some studies mention a sample size of 15 in the abstract and method, but only it becomes clear in the findings that only *three* of the patients provided usable data for a number of reasons⁸, while others do not mention data cleaning or preparation processes¹⁰. This is not good practice in terms of being open and transparent but causes concern for how these inferences are made and whether it is based on high-enough quality data. This again would be easier if digital phenotyping studies (or studies using digital traces) were preregistered, where these preregistrations are read and validated before any data analysis can take place⁵¹. Preregistrations require some detail regarding hypotheses, statistical methods, including power calculations and stopping rules regarding sample size, which would help with a number of these issues seen across the current literature base⁶¹. Finally, it is crucial to not oversell or overstate the results across any research. It is especially important when this research (currently) focuses almost entirely on patients with acute psychiatric illnesses, where results come from work that is underpowered and potentially extremely unreliable. This could become dangerous if this causes shifts in research focus towards unfavourable directions^{62,63}.

5 **Ethics, Security & Privacy, and Regulation**

5.1 *Privacy and Data Sharing of (Commercial) Apps*

Digital phenotyping also raises a number of ethical, regulatory, and legal questions (relating to any population tested). Current ethics standards are not yet up to date with research utilizing digital traces (from scraping online communities to device traces) as this is still a ‘new’ type

of research and data². There are large numbers of apps commercially build to help manage symptoms of mental health (e.g., Personal Mood Journal, Daylio, Calm, Schizophrenia Storylines, etc.), however these are not necessarily helpful to those wanting to conduct digital phenotyping studies. This would vary due to the (usual) lack of open source materials, not knowing how data is collected, shared, and used, and how do get data from these apps in a format useful to researchers. Hence, there is a rise in researchers developing their own apps for research purposes (this is not exclusive to medicine, e.g., Contact Logger, PEG Log, or Usage Logger in psychological science). It is important to note that open source and materials will not solve all of these issues, but it will allow for innovation, openness, transparency, and scrutiny regarding how apps work, how secure these apps are, and what data privacy is afforded (especially those made for research)⁶⁵, which is critical for researchers conducting work utilizing digital traces. As stated previously, digital phenotyping or digital trace research is still fairly new and will remain an iterative process of learning as implications become clearer with more research.

When focusing on apps built for research purposes, questions of security and privacy remain just as important, for example, some research goes as far as using voice clips form smartphones as a metric^{55,66}, which could be a serious ethical and privacy-related issue⁴⁴. There are intrinsic security issues with many smartphone apps, where security is not built into the app from the outset, which is problematic^{11,38}. This means when using or building new apps for digital trace research, there needs to be more openness and transparency about the security and privacy of these apps, noting that some apps may already be clear on their current security protocols.

Huckvale et al.⁶⁷ examined a number of data sharing and privacy problems across smartphone apps for depression and smoking, which revealed data sharing practices that included identifiable information of individuals. This confirms that many health apps (commercial or not) do not abide by privacy-oriented statues (e.g., HIPAA, GDPR)⁴⁴ and that sharing data with subsidiaries or third parties is indeed the norm⁶⁸. This is highly problematic as this also means reidentification can become easier by the increased ability to triangulate data sources from a variety of data (e.g., activity trackers, smart watches, smartphones)⁶⁹. Similarly, there remain questions as to whether there are appropriate app security features in place to protect users

² There are some frameworks available that evaluate apps for mental health (e.g., American Psychiatric Association have the APP ADVISOR, which aims to assess apps⁶⁴ and the Division of Digital Psychiatry at BIDMC (Harvard Medical School) also have a new app rating platform for mental health⁴³

(e.g., apps requesting invasive accesses on smartphone ecosystems, data sharing with third parties). The more users of a particular app means the amount of potentially highly personal and intimate data continues to grow, and if this is also being shared with third parties, if there was a data breach, the implications could be detrimental^{70,71} (e.g., in September 2020 alone, 95 data breaches have been reported within HIPAA-covered entities consisting of an estimated 9.7 million records)⁷².

5.2 *Data Use and Misuse Implications*

From an academic research perspective, it is clear that research considering actual usage and measurement from devices such as smartphones or wearables (even simple wearables such as pedometers) is at its infancy. The unintended consequences of using wearable technologies such as smart watches that enable '*personal quantification*' remain largely unknown and actually, might cause the opposite desired effect with users losing interest in their devices or self-quantification^{34,58}. Hence, from a participant perspective, there is much to learn about how feedback from devices may impact wellbeing (which is a key set of future research in itself). This lack of knowledge, therefore, has an impact for research such as digital phenotyping, whether feedback is given, if so, how and at what time. From a research ethics perspective, we need specific guidelines from bodies such as the British Psychological Association (BPS), American Psychological Association (APA), (and other subject-specific bodies utilising such tools) regarding ethics and protocols for conducting this research as well as for those potentially developing their own tools to gather such data.

From a potential clinical perspective, as previously stated, there are many unanswered questions as to how these data are being monitored and whether outcomes might impact the person adversely both in terms of how feedback impacts individuals, but increasingly as health apps may be linked with medical insurance, etc., (e.g., if a patient does not wear a device for a day or two, will they be penalized? If they have one additional drink than their doctor's suggestion, how would this impact patients, if at all?). Adjacent to this, legally, there will be questions relating to data use and reuse, however, the more pressing question perhaps relates to when digital phenotyping will inevitably go wrong. If the field continues with highly vulnerable patients and one has a severe incident, where does the fault lie? With the developers of the applications or devices? Those analysing the data and missed an incident that was deemed a false negative? The clinicians that allowed the use of this application(s)? Or the patient themselves? This links back to the ethical questions as to whether these vulnerable

patients are truly able to consent to such invasive research being conducted on them, and this of course impacts both the academic research as well as any use of digital phenotyping in both healthcare and commercial settings.

6 Clinical Outcomes

While the potential remains great for clinical outcomes, where digital phenotyping promises to aid both the diagnosis of, but also the daily monitoring and employing interventions when needed, we are a long way from this being feasible. As stated previously, there are a number of concerns relating to *how* we capture symptoms as required. Similarly, the ethical issues regarding wearables and sensors remains unsettled. With increasing numbers of ‘black boxes’ in cars⁷³, which can quickly invalidate someone’s insurance for maintaining a few miles per hour over the speed limit and not noticing, the same has started with health insurance (e.g., Vitality creating incentives and rewards for using a wearable health tracker⁷⁴). It creates an issue as to whether patients being constantly measured means they are uncomfortable and feel restricted due to a lack of privacy⁷⁵⁻⁷⁸. Although, some patients appear to be happy to share various types of data for analytics such as this, however, we must never assume this is and will be the case¹¹. In addition, these questions of privacy and wellbeing are deserving of both quantitative and qualitative insights to understand how these devices and precision medicine can benefit the patient without being invasive and intrusive. A key example of extremely privacy evading technology-tracking would be the ‘*Screenome Project*’³, which monitors smartphone use and takes screenshots every five seconds. This data captures a vast amount of sensitive data from conversations, activities, and interactions, which raises ethical and privacy questions.

In contrast, there are concerns relating to how this data, the sensors, and the care with these become integrated across various medical systems, how this will align with national and international laws and regulations regarding data collection at the very least¹¹. If this was the case, decisions will need to be made to understand who has access to the data, similarly, will analyse and monitor this data: will it be GPs? Specific doctors within the wider healthcare system? Just an algorithm? Depending on the country of interest, there are nuanced questions as to what capacity the current medical systems have for this huge influx of additional data and

³ This research group has 30 million screenshots from 600 people, which they call a ‘screenome’⁷⁹.

computational power to clean, analyze, and report on the data (e.g., countries with a national health system versus private healthcare systems, centralised, or decentralised healthcare systems, or the amount of capital available for the digital transformation of services and mental health, etc.). We can consider whether this is the place of the hospitals to store this data if these devices were prescribed for the patients, or does it stay with the patients, or with the companies making the devices.

7 What is the Future for Digital Phenotyping?

There is great potential for the future of digital phenotyping—or whichever term is currently in fashion, however, we must proceed cautiously^{3,44} to ensure we protect participants. We also must remember that digital traces and sensors have been around for a long time, and digital phenotyping is merely an evolution rather than a revolution. Moving forward, within academia, there needs to be input from more than one discipline. For example, clinicians are critical in order to sense check measurement and inferences (e.g., how does digital behavior x relate to offline symptom y ?), analyses, and the potential integration into healthcare systems and clinical decision making. Technical input may be required for data capture (although this is perhaps decreasing with an array of new apps being released, e.g., Usage Logger, AWARE frameworks, Funf in a Box). Other researchers with statistical or methods knowledge are required to navigate and make sense of these data. Hence, bringing together a variety of backgrounds will enrich the work and it is critical to ensure digital phenotyping is done rigorously due to the nature of its medical applications. Should research priorities focus on better data capture, the focus should be to build ethical and secure apps. This must understand the legal, regulatory, and integration with current healthcare systems, alongside leading analytics from a theory-driven perspective, and to educate the clinicians who will use ideally co-design, test, and implement these apps or devices with their patients.

8 Declarations

Funding	Not applicable/no funding associated
Conflicts of interest	No competing interests/conflicts of interest
Availability of Data	Not applicable
Code Availability	Not applicable
Authors' Contributions	BID is the sole author
Ethics Approval	Not applicable
Consent to Participate	Not applicable
Consent for Publication	Not applicable

9 References

1. Insel TR. Digital phenotyping: a global tool for psychiatry. *World Psychiatry*. 2018;17(3):276-277. doi:10.1002/wps.20550
2. Torous J, Kiang M V, Lorme J, Onnala J-P. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Ment Health*. 2016;3(2):e16. doi:10.2196/mental.5165
3. Mohr DC, Shilton K, Hotopf M. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. *Npj Digit Med*. 2020;3. doi:10.1038/s41746-020-0251-5
4. Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: Prospective observational cohort study. *J Med Internet Res*. 2019;21(4). doi:10.2196/11029
5. Scott J, Hidalgo-Mazzei D, Strawbridge R, et al. Prospective cohort study of early biosignatures of response to lithium in bipolar-I-disorders: overview of the H2020-funded R-LiNK initiative. *Int J Bipolar Disord*. 2019;7(1). doi:10.1186/s40345-019-0156-x
6. Wisniewski H, Henson P, Torous J. Using a Smartphone App to Identify Clinically Relevant Behavior Trends via Symptom Report, Cognition Scores, and Exercise Levels: A Case Series. *Front Psychiatry*. 2019;10(September):1-9. doi:10.3389/fpsy.2019.00652
7. Chauvin JJ, Insel TR. Building the Thermometer for Mental Health. *Cerebrum Dana Forum Brain Sci*. 2018;2018(November):1-12.
8. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnala JP. Relapse prediction in schizophrenia through digital phenotyping: A pilot study. *Neuropsychopharmacology*. 2018;43(8):1660-1666. doi:10.1038/s41386-018-0030-z
9. Reilly T, Mechelli A, McGuire P, Fusar-Poli P, Uhlhaas PJ. E-clinical high risk for psychosis: viewpoint on potential of digital innovations for preventive psychiatry. *J Med Internet Res*. 2019;21(10). doi:10.2196/14581
10. Zulueta J, Piscitello A, Rasic M, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: A biaffect digital phenotyping study. *J Med Internet Res*. 2018;20(7):1-10. doi:10.2196/jmir.9775
11. Mou D. Battling severe mental illnesses with smartphones: how patients' smartphone data can help improve clinical care. *mHealth*. 2016;2(8). doi:10.21037/mhealth.2016.08.03
12. Hinds J, Joinson A. Human and Computer Personality Prediction From Digital Footprints. *Curr Dir Psychol Sci*. 2019;28(2):204-211. doi:10.1177/0963721419827849
13. Hinds J, Joinson AN. What demographic attributes do our digital footprints reveal? A systematic review. *PLoS ONE*. 2018;13(11):e0207112. doi:10.1371/journal.pone.0207112
14. Boyd RL, Pasca P, Lanning K. The Personality Panorama: Conceptualizing Personality Through Big Behavioural Data. *Eur J Personal*. n/a(n/a). doi:10.1002/per.2254

15. Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: Data science enabling personalized medicine. *BMC Med.* 2018;16(150). doi:10.1186/s12916-018-1122-7
16. Davidson BI, Ellis DA. Avoiding Irrelevance: The manifestation and impacts of technophobia in psychological science. *PsyArXiv*. Published online 2019.
17. Emanuel EJ, Wachter RM. Artificial Intelligence in Health Care: Will the Value Match the Hype? *JAMA - J Am Med Assoc.* 2019;321(23):2281-2282. doi:10.1001/jama.2019.4914
18. Howison J, Wiggins A, Crowston K. Validity issues in the use of social network analysis with digital trace data. *J Assoc Inf Syst.* 2011;12(12):767-797. doi:10.17705/1jais.00282
19. Freelon D. On the Interpretation of Digital Trace Data in Communication and Social Computing Research. *J Broadcast Electron Media.* 2014;58:59-75. doi:10.1080/08838151.2013.875018
20. Weaver SD, Gahegan M. Constructing, Visualizing and Analyzing a Digital Footprint. *Geogr Rev.* 2007;97(3):324-350.
21. Cupchik GC. The digitized self in the internet age. *Psychol Aesthet Creat Arts.* 2011;5(4):318-328. doi:10.1037/a0024820
22. Swan M. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data.* 2013;1(2). doi:10.1089/big.2012.0002
23. Vial G. Reflections on quality requirements for digital trace data in IS research | Elsevier Enhanced Reader. *Decis Support Syst.* 2019;126:1-5. doi:10.1016/j.dss.2019.113133
24. Geyer K, Ellis DA, Shaw H, Davidson BI. *Open Source Smartphone Apps and Tools for Measuring, Quantifying, and Visualizing Screen Time.* PsyArXiv; 2020. doi:10.31234/osf.io/eqhfa
25. Andrews S, Ellis DA, Shaw H, Piwek L. Beyond self-report: Tools to compare estimated and real-world smartphone use. *PLoS ONE.* 2015;10(10):1-9. doi:10.1371/journal.pone.0139004
26. Welser H, Cosley D, Kossinets G, et al. Finding social roles in Wikipedia. *Proc 2011 IConference - IConference 11.* Published online 2011:122-129. doi:10.1145/1940761.1940778
27. Gleave E, Welser HT, Lento TM, Smith MA. A Conceptual and Operational Definition of ‘ Social Role ’ in Online Community. *Proc 42nd Hawaii Int Conf Syst Sci.* Published online 2009:1-11. doi:10.1109/HICSS.2009.6
28. Sun Y, Mobasher A. Utilizing Crowdsourced Data for Studies of Cycling and Air Pollution Exposure: A Case Study Using Strava Data. *Int J Environ Res Public Health.* 2017;14(3):274. doi:10.3390/ijerph14030274
29. Jestico B, Nelson T, Winters M. Mapping ridership using crowdsourced cycling data. *J Transp Geogr.* 2016;52:90-97. doi:10.1016/j.jtrangeo.2016.03.006

30. Rowe DA, Mahar MT, Raedeke TD, Lore J. Measuring Physical Activity in Children With Pedometers: Reliability, Reactivity, and Replacement of Missing Data. *Pediatr Exerc Sci.* 2004;16:343-354.
31. Strycker LA, Duncan SC, Chaumeton NR, Duncan TE, Toobert DJ. Reliability of pedometer data in samples of youth and older women. *Int J Behav Nutr Phys Act.* 2007;4(1):4. doi:10.1186/1479-5868-4-4
32. Kim T, McFee E, Olguin DO, Waber B, Pentland A “Sandy.” Sociometric badges: Using sensor technology to capture new forms of collaboration. *J Organ Behav.* 2012;33(3):412-427. doi:10.1002/job.1776
33. Taylor P, Banks F, Jolley D, et al. Oral hygiene effects verbal and nonverbal displays of confidence. *J Soc Psychol.* 2020;0(0):1-15. doi:10.1080/00224545.2020.1784825
34. Piwek L, Ellis DA, Andrews S, Joinson A. The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Med.* 2016;13(2):1-9. doi:10.1371/journal.pmed.1001953
35. Barrett MA, Humblet O, Hiatt RA, Adler NE. Big Data and Disease Prevention: From Quantified Self to Quantified Communities. *Big Data.* 2013;1(3). doi:10.1089/big.2013.0027
36. Torous J, Onnela JP, Keshavan M. New dimensions and new tools to realize the potential of RDoC: Digital phenotyping via smartphones and connected devices. *Transl Psychiatry.* 2017;7(3):2-4. doi:10.1038/tp.2017.25
37. Marsch LA. Opportunities and needs in digital phenotyping. *Neuropsychopharmacology.* 2018;43(8):1637-1638. doi:10.1038/s41386-018-0051-7
38. Geyer K, Ellis DA, Piwek L. A simple location-tracking app for psychological research. *Behav Res Methods.* Published online 2018. doi:https://doi.org/10.3758/s13428-018-1164-y
39. Ferreira D, Kostakos V, Dey AK. AWARE: Mobile Context Instrumentation Framework. *Front ICT.* 2015;2. doi:10.3389/fict.2015.00006
40. Funf in a Box. funf | Open Sensing Framework. Accessed September 22, 2020. <http://www.funf.org/>
41. Keil TF, Koschate M, Levine M. Contact Logger: Measuring everyday intergroup contact experiences in near-time. *Behav Res Methods.* Published online January 21, 2020. doi:10.3758/s13428-019-01335-w
42. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. *Npj Digit Med.* 2020;3(1):1-8. doi:10.1038/s41746-020-00312-4
43. The Division of Digital Psychiatry at BIDMC. Digital Psychiatry Apps (BETA). Published 2020. Accessed September 22, 2020. <https://apps.digitalpsych.org/>
44. Torous J, Haim A. Dichotomies in the development and implementation of digital mental health tools. *Psychiatr Serv.* 2018;69(12):1204-1206. doi:10.1176/appi.ps.201800193

45. Sen I, Floeck F, Weller K, Weiss B, Wagner C. A Total Error Framework for Digital Traces of Humans. *ArXiv190708228 Cs*. Published online December 5, 2019. Accessed September 22, 2020. <http://arxiv.org/abs/1907.08228>
46. Wilkinson J, Arnold KF, Murray EJ, et al. Time to reality check the promises of machine learning-powered precision medicine. *Lancet Digit Health*. 2020;0(0). doi:10.1016/S2589-7500(20)30200-4
47. Collins FS, Varmus H. A New Initiative on Precision Medicine. *N Engl J Med*. 2015;372(9):793-795. doi:10.1056/NEJMp1500523
48. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: Randomized controlled trial. *J Med Internet Res*. 2013;15(11). doi:10.2196/jmir.2588
49. Katz D, Dalton N, Holland S, O’Kane A, Price BA. Questioning the reflection paradigm for diabetes mobile apps. In: *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*. ; 2017:315-326. doi:10.1007/978-3-319-49655-9_39
50. Davidson BI, Jones SL, Joinson AN, Hinds J. The evolution of online ideological communities. *PLOS ONE*. 2019;14(5):e0216932. doi:10.1371/journal.pone.0216932
51. Centre for Open Science. OSF | Center For Open Science. Accessed September 22, 2020. <https://www.cos.io/products/osf>
52. mindstrong. mindstrong: Mental health care made for you. Published 2020. Accessed April 27, 2020. <https://mindstrong.com/>
53. mindstrong. Using science to help us seek the truth. Published 2020. Accessed April 27, 2020. <https://mindstrong.com/science/>
54. Ellis DA. *Smartphones within Psychological Science*. Cambridge University Press; 2020.
55. Bourla A, Mouchabac S, El Hage W, Ferreri F. e-PTSD: an overview on how new technologies can improve prediction and assessment of Posttraumatic Stress Disorder (PTSD). *Eur J Psychotraumatology*. 2018;9(sup1). doi:10.1080/20008198.2018.1424448
56. Angeletou S, Rowe M, Alani H. Modelling and Analysis of User Behaviour in Online Communities. *Semantic Web – ISWC 2011*. 2011;7031(Chapter 3):35-50. doi:10.1007/978-3-642-25073-6
57. Wachter S, Mittelstadt B. A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI [forthcoming]. *Columbia Bus Law Rev*. Published online 2018:1-84.
58. Etkin J. The Hidden Cost of Personal Quantification. *J Consum Res*. 2016;42(6):967-984. doi:10.1093/jcr/ucv095
59. Chen A, Zhu W. Revisiting the assumptions for inferential statistical analyses: A conceptual guide. *Quest*. 2001;53:418-439. doi:10.1080/00336297.2001.10491756

60. Vadillo MA, Konstantinidis E, Shanks DR. Underpowered samples, false negatives, and unconscious learning. *Psychon Bull Rev.* 2016;23:87-102. doi:10.3758/s13423-015-0892-6
61. van't Veer A, Giner-Sorolla R. Public template for pre-registration. OSF. Published June 18, 2014. Accessed September 24, 2020. <https://osf.io/k5wns/>
62. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *J Am Med Assoc.* 2002;288(3):358-362. doi:10.1001/jama.288.3.358
63. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: Review. *BMJ Online.* 2009;338. doi:10.1136/bmj.b1732
64. APA. APA App Advisor. Accessed September 22, 2020. <https://www.psychiatry.org/psychiatrists/practice/mental-health-apps>
65. von Hippel E. Learning from open-source software. *MIT Sloan Manag Rev.* 42(4):82-86.
66. Onnela JP, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology.* 2016;41(7):1691-1696. doi:10.1038/npp.2016.7
67. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. *JAMA Netw Open.* 2019;2(4):e192542-e192542. doi:10.1001/jamanetworkopen.2019.2542
68. Razaghpanah A, Nithyanand R, Vallina-Rodriguez N, et al. Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem. In: *Proceedings 2018 Network and Distributed System Security Symposium.* Internet Society; 2018. doi:10.14722/ndss.2018.23353
69. McCoy TH, Hughes MC. Preserving Patient Confidentiality as Data Grow: Implications of the Ability to Reidentify Physical Activity Data. *JAMA Netw Open.* 2018;1(8):e186029-e186029. doi:10.1001/jamanetworkopen.2018.6029
70. McCoy TH, Perlis RH. Temporal Trends and Characteristics of Reportable Health Data Breaches, 2010-2017. *JAMA.* 2018;320(12):1282-1284. doi:10.1001/jama.2018.9222
71. Mylonas A, Kastania A, Gritzalis D. Delegate the smartphone user? Security awareness in smartphone platforms. *Comput Secur.* 2013;34:47-66. doi:10.1016/j.cose.2012.11.004
72. HIPAA. September 2020 Healthcare Data Breach Report: 9.7 Million Records Compromised. *HIPAA J.* Published online 2020. Accessed October 27, 2020. <https://www.hipaajournal.com/september-2020-healthcare-data-breach-report-9-7-million-records-compromised/>
73. Filipova-Neumann L, Welzel P. Reducing asymmetric information in insurance markets: Cars with black boxes. *Telemat Inform.* 2010;27(4):394-403. doi:10.1016/j.tele.2010.03.003

74. Fitness Tracker Offers | Activity Tracking. Vitality. Accessed September 24, 2020. <https://www.vitality.co.uk/rewards/partners/activity-tracking/>
75. Volkman R. Privacy and Authenticity. In: *Computer Ethics Philosophical Enquiry*. ; 2014.
76. Alan F. Westin. Impact of Surveillance on Privacy. *Columbia Law Rev.* 1966;66(6):1003-1050.
77. Westin AF. Social and political dimensions of privacy. *J Soc Issues.* 2003;59(2):431-453. doi:10.1111/1540-4560.00072
78. Westin AF. Privacy and Freedom. *Am Sociol Rev.* 1968;33(1):173. doi:10.2307/2092293
79. Reeves B, Robinson T, Ram N. Time for the Human Screenome Project. *Nature.* 2020;577:314-317.