

Peer Review File

Article information: <https://dx.doi.org/10.21037/atm-21-6016>

First External Peer Review

Reviewer A

Comment 1: Figure 1 of a screen shot is not common practice for showing the results of a search string. A PRISMA flow diagram would be more appropriate.

Reply 1: Noted. However, Figure 1, referred to in the introduction, was not intended to show the results of a search string but rather to demonstrate the exponentially increasing interest in this field. We have clarified this in the figure legend of Figure 1.

Changes in the text: The figure legend of Figure 1 now reads “Figure 1. Search hits for “wearable” and “psychology” on PubMed, National Library of Medicine, showing exponentially increasing interest in this field.”

Comment 2: When mentioning brand names of devices, it may be better to refer to the operating system versus the brand. For example, to say iOS and Android devices, as the exclusion of a brand name could cause the reader to falsely believe that you are excluding certain types of brands. You mention FitBit, Oura Ring and Garmin, but do not mention the Apple Watch which is arguably a leader in the use of wearable devices and their integration into clinical studies on health through their Research Kit program.

Reply 2: Thank you for pointing this out. Indeed, the Apple Watch is an important wearable device which warrants mentioning in this paper. Otherwise, we decided to keep referring to devices by their brand name as it is more recognizable. Additionally, we now clarify that we are not excluding certain types of brands.

Changes in the text: Lines 97-99 now read “Smartphones and other wearable devices (including but not limited to the Fitbit, Apple Watch, Oura Ring, and Garmin watch) may enhance the early identification and monitoring of mental disorders by providing data associated with a patient’s mental health.”

Comment 3: While I personally find Google Scholar to be incredibly helpful for literature searches, the issue with Google is that their search algorithm is subject to change and therefore the results of your search may not always be reproducible. Please comment on this limitation, as well as the use of only 3 databases.

Reply 3: This is a good point. We updated our literature search as of April and included

a fourth database in our search (Cochrane).

Changes in the text: In lines 136-137, we clarify “A literature review was performed using PubMed, ScienceDirect, the Cochrane Library, and Google Scholar, of articles published through April 4, 2022.” Consequently, we found some additional relevant studies, such as those spoken about in lines 256-265 regarding social jetlag.

Comment 4: More information is required in the Methods section. Did you follow a specific search string or did you just do a general search using the terms mentioned? How were articles selected? Did you use a title/abstract followed by full-text review? Was there specific inclusion and exclusion criteria applied? How many independent reviewers sorted through the articles? What data was extracted from the chosen manuscripts? Etc.

Reply 4: These are good points. Our search was not as rigorous as you describe and we have accordingly reframed our study as a commentary. Our methods section has been modified in clarification.

Changes in the text: Our Methods section from lines 136-143 now reads “A literature review was performed using PubMed, ScienceDirect, the Cochrane Library, and Google Scholar, of articles published through April 4, 2022. The Boolean search terms used included: “wearable devices”, “wearables”, “smartphone”, “mental health”, “psychology”, and “psychiatry”. A focus was placed on papers demonstrating the use of smartphones and other wearable sensors for the detection and monitoring of mental illness. We did not consider the use of these devices We did not consider articles not written in English, and not of the following manuscript types: encyclopedia, conference abstracts, correspondence, editorials, short communications, and corrections.”

Comment 5: The organization of the manuscript is a bit confusing. Definitions of things like inertial measurement units should be in the background or methods. The Results section should be clearly labelled.

Reply 5: Thank you for pointing this out. We have now explained inertial measurement units in the Introduction. In addition, given that we reframed our article to be a commentary, we decided to retain our original decision to integrate the results and discussion of our article under subheadings corresponding to the different wearable metrics explored. This is not a new approach, and is similarly demonstrated in other papers published by this journal (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8246192/>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8106090/>).

Changes in the text: Lines 118-119 read “Meanwhile, increased physical activity levels captured using inbuilt inertial measurement units (which document acceleration patterns)”.

Comment 6: There should also be a section in Results that provides the reader with a snapshot of the included articles (e.g. "The initial search yielded X articles. After removing duplicates, X articles were screened by title and abstract and Y articles were screened by full text. Overall, X articles were included in this review. Of the X articles, X focused on physical activity, X focused on GPS tracing...").

Reply 6: This is a good point and would certainly be warranted if our study used a systematic search as in a systematic or scoping review. We did not perform a systematic search and have reframed our study to be a commentary as per recommendations.

Changes in the text: Clarification in the Methods.

Comment 7: The discussion section should also be clearly labelled as it is hard to distinguish the results from the discussion and therefore provide real feedback on your overall synthesis and interpretation of the literature based on the results.

Reply 7: Given that we reframed our article to be a commentary, we decided to retain our original decision to integrate the results and discussion of our article under subheadings corresponding to the different wearable metrics explored. This is not a new approach, and is similarly demonstrated in other papers published by this journal (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8246192/>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8106090/>).

Changes in the text: None.

Comment 8: I would suggest either re-framing the paper as a commentary on the potential for the use of wearable devices for the detection, monitoring and study of mental illness OR providing the reader with enough information in the methods section for them to have confidence in your interpretations and conclusions of your data as a systematic or even scoping (which may be more appropriate given your research question) review.

Reply 8: Thank you, this was very helpful. In keeping with your recommendations, we have reframed our paper as a commentary as this better aligns with our methodology.

Changes in the text: Clarification in the methods.

Reviewer B

This manuscript presented a narrative review on the promise of using wearable devices to identify and treat mental illness. While there have been similar reviews on this topic, this review provides an updated primer given the continued rise of this work. I have several comments that hopefully can improve the quality of this manuscript.

Comment 1: There are a lot of recent papers that are not included in the authors' review, particularly within the field of schizophrenia. Is there a reason these papers were not included? The authors should consider broadening their keyword search to be more inclusive.

Reply 1: To make our search more inclusive, we added an additional database and performed a more recent search as of April 4, 2022. We found additional relevant studies, including a paper regarding schizophrenia by Buck et al. (<https://pubmed.ncbi.nlm.nih.gov/30940400/>). It should also be noted that, in accordance with recommendations from other reviewers, we have reframed our article as a commentary – hence, our paper does not have a systematic search as a systematic or scoping review would.

Changes in the text: We have clarified our updated search in our Methods as in lines 136-137 which now read “A literature review was performed using PubMed, ScienceDirect, the Cochrane Library, and Google Scholar, of articles published through April 4, 2022.” Our discussion of the paper by Buck et al. is found in lines 214-218: “Similarly, Buck et al. (2019) found that the number and duration of outgoing calls and the total frequency of text messages was significantly correlated with schizophrenia relapse ($p \leq 0.031$) (30). These findings suggest that data from smartphones could be used to predict relapse of mental disorders before they occur, enabling possibilities of prevention.”

Comment 2: The topic of risk factors or at-risk individuals is mentioned in the introduction but is not a topic of focus in the review. For example, how exactly can wearables be used to identify risk factors? Ideally, risk factors should be studied in people without mental illness or showing attenuated symptoms-- groups not included in the review. This needs to be addressed somehow.

Reply 2: This is a good point and constitutes another barrier to clinical implantation. Although wearable devices can theoretically be used to enable the early identification of mental illness in at-risk individuals, it is unclear how at-risk individuals could be identified as appropriate candidates for wearable-based monitoring. We have now flagged this consideration.

Changes in the text: Lines 317-325 now read “Despite the exciting possibilities of wearable devices in the detection and monitoring of mental illness, many barriers to

clinical implementation exist. For example, given that it is not feasible to screen for mental illness indiscriminately, it is unclear how individuals should be selectively screened to allow for the early detection of mental illness. One approach is to screen individuals who are at risk of developing mental illness. While this may be possible in individuals who have had prior mental illness with access to a mental health professional, it is unlikely that members of the community can be identified as appropriate candidates for wearable-based monitoring with their consent before they present to a mental health professional of their own accord.”

Comment 3: It is unclear what exactly wearables/digital phenotyping is capturing from this review. Is the focus on group differences between those with/without mental illness on specific metrics? That is not interesting as those with mental illness, particularly serious, such as bipolar disorder and PTSD, are bound to show differences in many metrics just given the severity of the illness. Instead, a focus should be on mechanisms of mental illness that can be elucidated using wearables. The authors briefly touch on this in the limitation section, but I think this paper could benefit from a more vigorous discussion of causal mechanisms. Discussing risk factors as mentioned above could also help with this.

Reply 3: Thank you for pointing these out. We have now explained the term “digital phenotype”. Although it is true that some diagnoses may be obvious and digital phenotyping may be redundant, we are of the stance that digital phenotyping may still be useful in a variety of other situations. For example, a subtle deterioration in metrics may indicate a need for assessment and allow for the early detection of mental illness, whilst tracking wearable metrics over time may glean insight into disease progression or recovery. Individuals who are known to be at risk of mental illness such as those with past mental illness may also be monitored allowing for the early detection of any relapse in disease.

Changes in the text: Lines 103-104 now read “Together, these metrics form the “digital phenotype” of the individual, which refers to the behaviour and characteristics of an individual as inferred by their interaction with digital devices.” Lines 107-113 now read: “Although currently not directly corresponding with diagnosable mental illnesses identified in the DSM-V and although digital phenotyping may be redundant in some cases due to the diagnosis being obvious, a deterioration in these metrics may enable the early identification of mental illness and the progression of these metrics may correspond with disease severity. Furthermore, individuals who are at risk of mental illness such as those with previous mental illness could be monitored allowing for the early identification of disease onset or relapse”.

Comment 4: The authors discuss that clinicians and healthcare providers could use this information in real-time to guide treatment decisions. While this may be the ultimate

goal, I feel as though we are long away from that. There are many challenges to implementing wearables for treatment in practice. For example, is it feasible for providers to give idiographic treatment for every patient? Further, there would need to be sufficient normative data for every metric for this to be reliable. These limitations and others should be discussed.

Reply 4: Yes, while the possibilities of wearable-based monitoring in psychology is exciting, there are many barriers to clinical implementation. We have now flagged this point by creating a new subheading “Barriers to clinical implementation”. Regarding your point about whether data from these sensors should be used to give idiographic data for each patient, our stance is that wearable-based data only acts as a clinical adjunct (and not a substitute for other assessments) – and we have now clarified this in the Introduction. Yes, there needs to be sufficient normative data for comparisons to be made and pathological findings to be identified. We have now raised this consideration in our “Barriers to clinical implementation” section.

Changes in the text: Also, we have clarified that wearable sensors should only be used as a clinical adjunct in lines 160-161 “Wearable sensors can act as a clinical adjunct and provide objective data capture that can provide insight into a person’s mental health”. In our “Barriers to clinical implementation” section, we clarify in lines 345-355 “Other barriers to clinical implementation include device validation and the lack of sufficient normative data for various health metrics. For example, in a meta-analysis by Haghayegh et al. it was found that, compared to the gold-standard polysomnography, Fitbit models had poor specificity (0.10-0.52) in correctly identifying sleep epochs (50). Advancements in wearable technology accuracy must take place before these devices can be recommended for clinical use. In addition, health metrics obtained from patients must be compared against population norms. While this is possible for some metrics such as daily step count where there exist large databases organised by sex and age (51), a normative range for other metrics such as social jetlag has not been defined. Additional work is required to gather normative values for a larger range of psychology-related metrics before wearable-based monitoring of mental illness can be applied in clinical settings.”

Comment 5: In the section starting on line 246, the authors discuss using summary scores to assess mental disorders. However, there are many issues with this approach. Namely, when using summary scores, it essentially assumes that all items are equally weighted and equally represented by the sum score, which is hardly ever the case. This would also be true in the authors' example related to text and call logs. For example, utilizing a summary score on text and call logs may not capture the nuance, such as the person the patient is interacting with, the duration of the interactions, and the purpose of the interaction.

Reply 5: Thanks for raising this issue. We have now acknowledged the limitations

associated with summary scores.

Changes in the text: Lines 305-308 now read “Although summary scores may oversimplify metrics and lack nuance associated with, in an example pertaining to call metrics, the purpose of each call and the identity of the caller, they may still be useful in broadly categorizing patients into varying levels of disease severity.”

Comment 6: A critical point that needs to be addressed is why we should use wearables and, more specifically, their utility above traditional assessments. This is something that I have struggled with when using digital phenotyping as well. We validate digital phenotyping metrics by correlating them with clinician-rated or self-reports that we deem sub-par, then state that wearables are better. This rationale is insufficient, because then why not just use traditional assessments that save time and resources. The authors should include studies showing that wearables capture dynamic changes (e.g., across time) that conventional assessments do not measure. This would provide a more robust rationale than simply saying traditional measures are limited by retrospective bias.

Reply 6: Thank you for the comment – indeed, this is a concept that we have struggled with as well. We have now emphasized the potential for wearable devices to capture dynamic changes over time in our Introduction.

Changes in the text: Lines 104-107 read “Moreover, contrary to traditional single timepoint based assessments such as patient-reported outcomes, wearable-based metrics can be collected continuously (for as long as the device is worn), allowing dynamic changes in disease status over time to be captured”.

Reviewer C

Using smartphones and wearables for mental illness is an interesting topic in the field of mHealth and eHealth. The authors summarized recent studies in this topic in this narrative review. I have the following major comments:

Comment 1: A flowchart showing the selective procedure of the literature including the number of papers in each step is needed. Now I cannot even see how many studies the authors have reviewed in their manuscript.

Reply 1: Yes, our search was not rigorous enough for this paper to qualify as a systematic review. As per the recommendations of another reviewer, we have reframed our article to be a commentary and have restructured our paper accordingly,

Changes in the text: The title of this paper is now “A commentary on the potential of smartphones and other wearable devices to be used in the identification and monitoring of mental illness”. Additional clarification in the Methods regarding our search methodology.

Comment 2: A table summarizing the key features of the final selected literature is suggested, such as the years of the studies, countries, the technologies used in these studies (e.g., apps, smartphones, or wearable devices).

Reply 2: Given that we did not perform a systematic search and as per the recommendations of another reviewer, we have reclassified our study as a commentary. Since we did not perform a systematic search, it would be misleading for us to present a table of selected studies.

Changes in the text: Clarification in the Methods.

Comment 3: The authors summarized the pros of smartphones and wearables in mental illness. But what are the cons? I think the reliability and validity of these new technologies still need to be further tested. For example, negative results of validating wearables in measuring sleep (see doi: 10.2196/16273, DOI: 10.1016/j.smr.2019.101227) and other physiological signals such as SpO₂ (see DOI: 10.2196/24171, doi: 10.1055/a-1337-2790) have been reported.

Reply 3: This is an important limitation to add. We have now added this consideration to our “Barriers to clinical implementation section”.

Changes in the text: Lines 356-362 now read “Moreover, some metrics may not yet be collected with sufficient accuracy with the current state of wearable technology. Haghayegh et al. (2019) performed a systematic review and found that the wrist-based Fitbit had poor specificity (0.58-0.69) when detecting sleep epochs (52). Similarly, Hermand, Coll, Richalet & Lhuissier (2021) found that the Garmin Forerunner (a wrist-based oximeter) had a >50% error rate when reporting oxygen saturation (p<0.001) (53). Before widespread clinical uptake can be expected, wearable devices need to

demonstrate improved accuracy.”

Comment 4: In the section 'other relevant metrics' the authors should also discuss blood pressure and SpO₂, because they are also biomarkers related to mental health and currently some smartwatches have added these functions.

Reply 4: Thanks for pointing this out. We have added another paragraph under this section to discuss blood pressure and SpO₂.

Changes in the text: Lines 279-290 now read: Blood pressure and oxygen saturation are additional metrics which can be captured by wearable sensors but which, to the authors' knowledge, have not yet been demonstrated to be useful in the identification and monitoring of mental illness (43). However, it is reasonable to speculate that these metrics may still be useful in mental illness detection and monitoring. For example, it is well established that psychological stress is associated with increased blood pressure. A meta-analysis by Gasperin et al. (2009) demonstrated that subjects who had stronger responses to psychologically stressful tasks were 21% more likely to develop an increase in blood pressure compared to those with weaker responses ($p < 0.001$) (44). Meanwhile, oxygen saturation is relevant in the monitoring of sleep apnoea, which itself is linked to mental health disease such as anxiety and suicide ideation – Kaufmann et al. (2017) found that past year sleep apnoea was associated with a 3.11 (95%CI = 2.77 – 3.50) times increase in the odds of reporting depression in the past year.

Comment 5: I am glad to see that you have a section of sleep patterns, and circadian rhythm is also mentioned. I suggest the authors to add the discussion of social jetlag in this section too, because social jetlag is associated with mental health (see DOI: 10.1080/07420528.2019.1636813) and it can be measured in large population in real life by smartwatches (see doi: 10.2196/13482.).

Reply 5: Thank you for your insight. We have added a section on social jetlag as you have advised.

Changes in the text: Lines 256-265 now read “Interestingly, social jetlag has also been proposed as another sleep-related metric associated with mental health. For example, Islam et al. found that greater social jetlag was associated with an increased likelihood of having depressive symptoms where Japanese employees with at least two hours of social jetlag (defined as the difference in sleep timing between work days and non-work days) were 2.14 times more likely to demonstrate depressive symptoms ($p = 0.01$) (37). Social jetlag can be measured on a population scale, as shown by Zhang, Cajochen & Khatami (2019) who monitored social jetlag in 71,176 participants (38). In this way, the monitoring of social jetlag on a large scale may contribute to population screening for mental illness, allowing for targeted early intervention. Clearly, paucity of research in this area represents a clear avenue for future research.”

Second External Peer Review

Reviewer Comments:

Comment 1: Thank you for clarifying your intended purpose of Figure 1. However, I do not believe that Figure 1 adds substantial value to the overall paper and reader. The same information could be conveyed in a simple sentence.

Reply 1: Thanks for clarifying your stance on Figure 1. As recommended, we will delete it and convey the same idea in text.

Changes in the text: Figure 1 has been removed. Lines 62-64 reads “Recently, research interest in wearable devices has expanded towards mental health outcomes – a field that has historically lacked quantifiable biological indicators of health”.

Comment 2: As you note in your response, you have revised your aims and title of the paper to be a commentary on the use of wearable devices. However, the manuscript still contains references to summarizing the literature, which is no longer your stated aim as per your response and title. Please clarify.

As a commentary, this paper is overly focused on a synthesis of the literature and is missing the larger component of discussion on the future, excitement, and promise of wearable technologies that make commentaries interesting to read.

Reply 2: Thanks for pointing this out. While many of the studies we discussed were performed to showcase future applications of wearable technologies, we acknowledge that a commentary should additionally have clear authorial input. We have revised our manuscript accordingly to clarify our commentary on the future prospects and excitement of wearable technologies in psychology while also commenting on the current barriers to clinical implementation.

Changes in the text: The section “Future considerations in the wearable-based monitoring of mental illness” which spans lines 284-373 comments on the future prospects and current barriers to the clinical implementation of wearable-based mental health identification and monitoring.