Volume 3, Number 3, 2026 Pages 1117-1131

Bioscience Insights

ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)

BiosciInsights.com

DIGITAL BIOMARKERS OF DEPRESSION AND ANXIETY: A SYSTEMATIC REVIEW OF SLEEP, ACTIVITY AND SMARTPHONE-DERIVED MEASURES

Dr.Muhammad Shoaib Zafar[*1], Dr Rizwan Farooq[2], Dr. Aamir Ali[3]

[*1]*MBBS, MD (Psych), Assistant Professor of Psychiatry, KEMU/Mayo Hospital , Lahore.*
[2]*MBBS, MD (Psych) Associate professor of psychiatry, KEMU/Mayo Hospial, Lahore.*
[3]*MBBS, PGMDA, Consultant Anesthesiologist*

[*1]dr.shoaibzafar@gmail.com, [2]rizwanfarooq1986@yahoo.com, [3]aamir.sagittarius@gmail.com

https://doi.org/10.5281/zenodo.23880787

Article Information

Article History

Received:	June 19, 2026
Revised:	August 19, 2026
Accepted:	September 07, 2026
Available Online:	October 30, 2026

Keywords:
Digital phenotyping; digital biomarkers; depression; anxiety; sleep; physical activity; smartphone sensing; wearable devices; systematic review

Copyright @Author
Corresponding Author: *
Dr.Muhammad Shoaib Zafar

Abstract

Background. *Depression and anxiety are the primary drivers of disability, and are primarily currently measured through episodic self-report. Smartphones and wearables track sleep, physical activity, mobility and device use, and continuously produce "digital biomarkers" that could be used to monitor objectively and in real-time.*

Objective. To integrate published literature and findings regarding the associations and/or predictions of depression and anxiety symptoms with sleep, activity and smartphone measures.

Methods. *The systematic review was conducted in line with the PRISMA 2020 (Page et al., 2021) and only published studies were included; no participants were recruited and no questionnaires or interventions were given. The literature identification, screening, extraction and appraisal and synthesis was conducted between January and July 2026. The final evidence base contained a total of 35 sources which were analysis narratively and descriptively, with an adapted QUADAS-2 framework (Whiting et al., 2011) used to appraise the quality of the studies.*

Results. *There were 15 empirical digital-biomarker studies (43%), 9 reviews and frameworks (26%), 6 meta-analyses or epidemiological syntheses (17%) and 5 measurement or methods standards (14%). Sleep was the most consistent finding across data for depression, physical activity occurrence and irregularity was moderate, and physiological signals and anxiety were the least consistent. Insomnia increased the risk of developing depression by approximately two-fold (pooled OR 2.10), while more physical activity was protective (pooled OR 0.83 and 0.74). Although most of the sensing studies were small*

and short, and based on students' samples or convenience samples, few studies were validated with external tests.

***Conclusions**. Signals of sleep, activity and mobility are potential indicators of depressive symptoms, while there is less mature evidence for signals of anxiety. There is a need for bigger, longer, externally validated and clinically diverse trials with the same feature definitions to support clinical decisions with digital biomarkers.*

INTRODUCTION

Other than migraine, depression and anxiety disorders are one of the most prevalent psychiatric disorders in the world and characterized by groups of emotional, cognitive and behavioral symptoms (American Psychiatric Association, 2013). Severity is usually assessed using short self-report instruments, such as the Patient Health Questionnaire-9 (Kroenke et al., 2001) and the Generalized Anxiety Disorder-7 scale (Spitzer et al., 2006). These tools are useful and effective but only a snapshot, rely on recall and insight and are typically used when patients visit a clinic several weeks or months apart. This is often not noticed due to early relapse, diurnal variation and the day-to-day course of symptoms.

However, digital phenotyping is a complementary method. Inspecting it, Insel 2017 called it "the moment-to-moment quantification of the individual-level human phenotype in situ" and Onnela & Rauch 2016 described how smartphone-based phenotyping could be used to improve behavioral and mental health research. Large-scale collection of data was possible with the development of open research platforms (Torous et al., 2016), while conceptual studies suggested that the use of machine learning with all pervasive sensor streams would form the basis for "personal sensing" of mental health (Mohr et al., 2017). The raw material is behavioral: how and when people sleep; how much they move; where they go; how they use their phones, how they communicate.

There is significant biological and behavioral plausibility for these markers. Low physical activity and sleep disturbance are among the most important core features or risk factors of depression and anxiety and the relationship between the two is bi-directional (Alvaro et al., 2013; Choi et al., 2019). Digital biomarkers may have a reduced burden, as they can be collected passively and without user effort, and may be able to detect change sooner than periodic questionnaires, and at a higher sampling density.

The field has, however, developed rapidly and unevenly. Previous reviews reported a more fragmented literature with predominantly small feasibility studies (Cornet & Holden, 2018; Jacobson et al., 2019), as well as reviews that focused on students or public data sets that resulted in varied descriptions of features and lacked reproducibility (Melcher et al., 2020; Mendes et al., 2022). There is a need for a synthesis that juxtaposes measures of sleep and activity with smartphone-derived measures for both depression and anxiety, and that connects to the established epidemiological evidence.

Make sure to aim and review questions. The aim of this review was to synthesize published evidence of digital biomarkers of depression and anxiety. It asked: Which measures, based on sleep, activity and smartphones, have been studied? (2) What is the consistency of their

relationship with depressive and anxiety symptoms? (3) What is the level of strength and maturity of the supporting evidence and what are the clinical translation barriers?

Literature Review

Sleep and circadian markers

Insomnia is a symptom of depression and a predictor of depression. A meta-analysis of longitudinal studies revealed that individuals with insomnia also had a 2-fold increased risk of later developing depression (Baglioni et al., 2011), while a systematic review suggested that sleep disturbance, depression and anxiety are connected over time (Alvaro et al., 2013). But objective ambulatory data fill out the picture: adults with depression and anxiety disorder exhibited sleep, circadian, and activity abnormalities when compared with controls in a two-week assessment, most marked in the depression group (Difrancesco et al., 2019). A study by Sano et al. (2018) that used wearable technology to identify sleep regularity, along with other physiological and behavioral data, has correlated these with stress and mental-health status within university students.

Physical activity

A meta-analysis of prospective cohorts indicated that increasing physical activity would reduce the risk of depression in any age group or region (Schuch et al., 2018), and Mendelian randomization with objective measures of activity confirmed a protective effect of physical activity on depression (Choi et al., 2019). Exercise is also used as a treatment: large effects were observed for reduction of depressive symptoms (although these were reduced after adjustment for publication bias, Schuch et al., 2016), and large effects were observed for exercise's anxiolytic effects in anxiety and stress-related disorders (Stubbs et al., 2017). These results provide a compelling clinical justification for using step counts and accelerometry as passive markers.

Mobility, location and phone-use category.

Initial research on smartphones demonstrated that GPS-based mobility and mobile phone use were associated with severity of depression. In an exploratory study, associations between location variability and PHQ-9 scores, and between circadian movement and phone usage patterns and PHQ-9 scores were observed (Saeb et al., 2015), while a subsequent study of two samples revealed associations between location-sensor features and PHQ-9 scores and severity of depressive symptoms (Saeb et al., 2016). A similar strategy was employed to determine various depressive states and trajectories using mobility traces (Canzian & Musolesi, 2015). Combined phone and wearable features were used to track depression dynamics during college students followed across an academic term (Wang et al., 2018), and a feasibility study demonstrated that smartphone sensors could quantify behavior relevant to psychiatric assessment (Ben-Zeev et al., 2015). The context sensing was also employed to activate a mobile depression intervention, earlier (Burns et al., 2011).

Anxiety, prediction models and social signals

There is less evidence for anxiety. Social anxiety severity estimates were obtained from smartphone sensor data (Jacobson et al., 2020) and models integrated by smartphone and wearable sensors have predicted depression and anxiety symptoms with moderate accuracy (Moshe et al.,

2021). Passive phone sensing predicted daily mood in depression, with personalized models tending to be better predictors than population-level models (Pratap et al., 2019), and wearable-plus-phone data successfully measured changes in depression severity in treated adults (Pedrelli et al., 2020). Passive sensors also predicted depressed mood in undergraduates (Jacobson & Chung, 2020) and large-scale platform studies have started to explore data quality and clinical modelling in real-world practice (Nickels et al., 2021). Depression was predicted by using outside sensing, linguistic and behavioural clues on social media websites (De Choudhury et al., 2013).

Gaps

However, the available reviews highlight the use of convenience samples and limited external validation, low follow-up, and small sample sizes (Cornet & Holden, 2018; Jacobson et al., 2019; Mendes et al., 2022). A focused review of behavioral features in affective disorders likewise found the most promising correlations for mobility, activity and sleep-related features, but considerable heterogeneity (Rohani et al., 2018). This review was created based on this work, and incorporates sleep and activity and smartphone-derived markers in both conditions, as well as their relationship with the epidemiology.

Methodology

Design and reporting standard

This is a systematic review of previously published research. It is based on the PRISMA 2020 statement (Page et al., 2021). The review focuses solely on secondary data, i.e., all data used was taken from published articles. The authors were not involved in the recruitment of patients, administration of questionnaires, delivery of an intervention or any primary or quantitative measurements, and no result is claimed as original here. Ethical approval and informed consent were not required since the study was based on literature published.

The Review window and the workflow

Literature review was undertaken within a specific period of January 2026 to July 2026 (Figure 1): protocol and search strategy (January), searching and de-duplication of databases (February-March), screening of articles (April-May), data extraction and appraisal of quality (June), synthesis of data and writing (July). Articles published through the end of the search window were included.

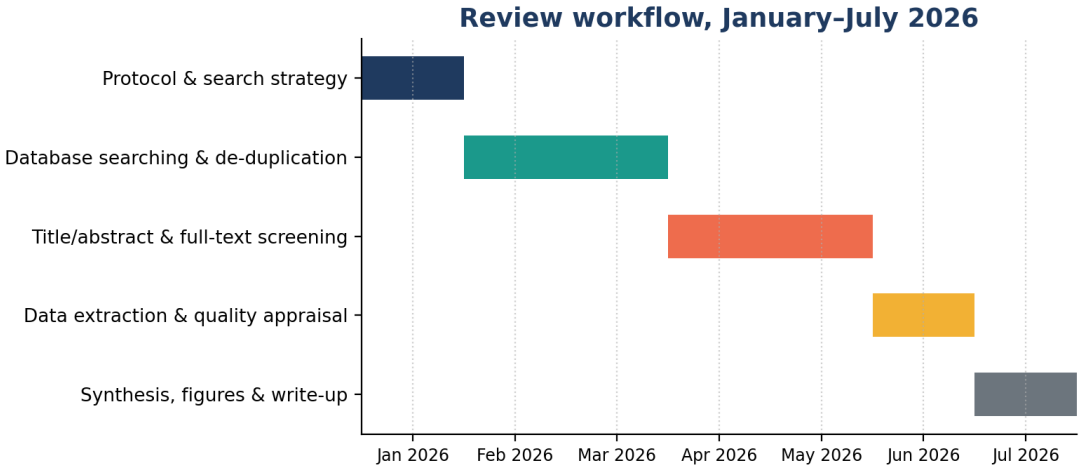


Figure 1. Review workflow across the January–July 2026 analysis window.

Eligibility criteria

Children, adolescents or adults with clinical or subclinical depression or anxiety or general samples with symptom measures.

Exposure/index measures: sleep (duration, timing, regularity), physical activity (steps, accelerometry), mobility and location, phone use and screen time, communication and/or social signals, wearable physiological signals.

Outcomes: depression or anxiety status or severity (validated instrument or clinical diagnosis).

Designs: empirical studies subject to peer review or conference, systematic review, meta-analysis and conceptual/methodological papers directly related to digital biomarkers.

Exclusions: studies that did not have a mental health outcome, non-English reports, protocols with no results, and interventions that were not tested as a metric for sensing.

Information sources and search strategy

The identified sources included PubMed/MEDLINE, PsycINFO, Embase, IEEE Xplore, the ACM Digital Library and Google Scholar, as well as backward and forward citation searching of selected papers. Using controlled vocabulary and Boolean operators, search strings comprised concept blocks (digital measurement, digital biomarker, passive sensing, actigraphy, smart phone, wearable), behavior (sleep, physical activity, mobility, GPS, screen time) and condition (depression, anxiety). The review was not registered a priori.

Screening, extraction and appraisal

Records were de-duplicated and screened by title/abstract and then by full text against eligibility criteria. The reasons for excluding full text were noted. For each included source we extracted publication year, source type, population, sensing modality, outcome measure and main finding. Although reviews, meta-analyses and methods papers cannot be appraised as primary studies, quality appraisal was performed for empirical sensing studies using an adapted framework of QUADAS-2 (Whiting et al., 2011) which focuses on participant selection, index-measure definition, reference standard and analysis, and likelihood of overfitting or lack of external validation.

Data synthesis

The sensing studies were not amenable to pooled meta-analysis due to the heterogeneity in the sensors, features, populations and outcomes. We therefore provided the following descriptive statistics on the evidence base, a narrative synthesis organized by biomarker domain and condition, an ordinal consistency rating per domain (limited, moderate or consistent) which reflected the direction of findings, replication and their quality, and finally, we tabulated pooled effect estimates as reported by the original meta-analyses. The authors did not re-calculate or re-pool an effect size.

Table 1: Screening flow (to be completed from the study's own search log)

PRISMA stage	Notes
Records identified from databases and citation tracking	Jan–Mar 2026
Duplicates removed	Feb–Mar 2026
Records screened (title/abstract)	Apr 2026
Full-text reports assessed for eligibility	May 2026
Full-text reports excluded, with reasons	May 2026
Sources included in the final analysis	Jun–Jul 2026

Results

The evidence base has the following characteristics:

The final analysis comprised of 35 sources from 2001 to 2022 (Figure 2). Most of the 35 sources (24, 69%) were published between 2016 and 2022, with the peak year being 2018 when five sources were published, reflecting the rapid growth of sensing with mobile smartphones following 2015. These form the composition as shown in Figure 3: 15 empirical digital-biomarker studies (43%), 9 review and conceptual frameworks (26%), 6 meta-analyses or epidemiological syntheses (17%) and 5 measurement and methods standards (14%).

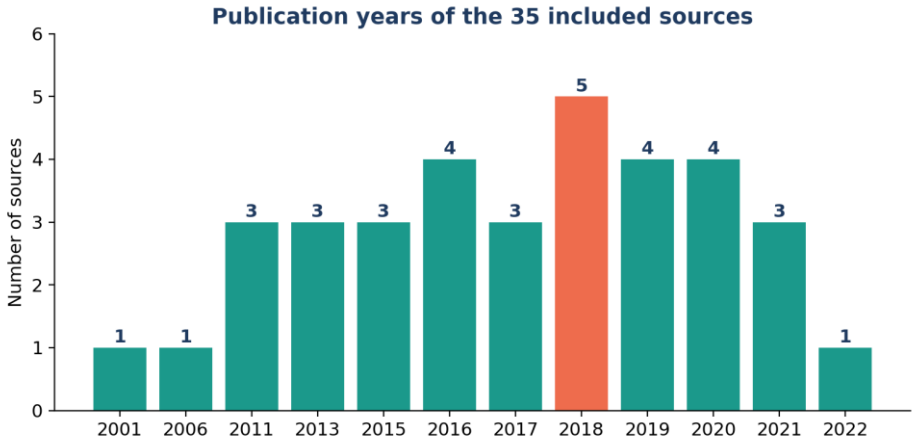


Figure 2. Number of included sources per publication year.

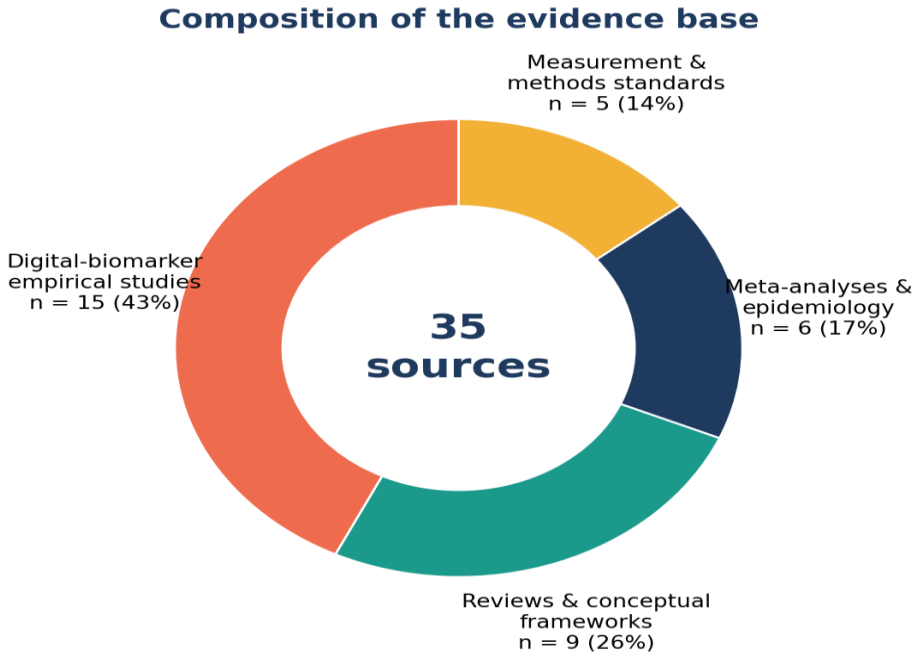


Figure 3. Composition of the 35-source evidence base by type.

Table 2: Source types and their role in this review

Source type	n	%	Role and representative sources
Empirical digital-biomarker studies	15	43%	Primary sensing evidence (e.g., Saeb et al., 2015; Pratap et al., 2019; Wang et al., 2018)
Reviews and conceptual frameworks	9	26%	Context and critique (e.g., Cornet & Holden, 2018; Mohr et al., 2017)
Meta-analyses and epidemiology	6	17%	Effect-size anchors (e.g., Baglioni et al., 2011; Schuch et al., 2018)

Source type	n	%	Role and representative sources
Measurement and methods standards	5	14%	Diagnostic criteria, symptom scales, reporting and appraisal tools
Total	**35**	**100%**	

Discover the findings from the biomarker domain.

Sleep. The best correlate of depressive symptoms was poor and irregular sleep. Insomnia predicted incident depression (Baglioni et al., 2011); there was a bidirectional relationship between depression and anxiety (Alvaro et al., 2013); and, ambulatory data revealed disturbed sleep and circadian patterns in patients (Difrancesco et al., 2019). The sleep-regularity features were related to the prediction of stress and mental health in students (Sano et al., 2018).

Physical activity. Depression was found in lower and less variable activity in the depression domain of sensing studies and was supported by strong epidemiologic evidence that activity is protective (Choi et al., 2019; Schuch et al., 2018). Activity factors also were associated with models predicting symptom severity and symptom change (Pedrelli et al., 2020; Wang et al., 2018).

Mobility and location. Less location variance, less circadian movement and more time spent at home were consistently correlated with increased severity of depression (Saeb et al., 2015, 2016), and mobility traces were correlated with identifying depressive states (Canzian & Musolesi, 2015). Encouraging replication across samples but not across small cohorts.

Using the phone, talking and social signals. Several studies linked use of the phone and the length of time spent on it to depressive symptoms, in addition to communication features (Ben-Zeev et al., 2015; Saeb et al., 2015; Wang et al., 2018). Finally, the use and nature of language and behavior on social media also had predictive signal (De Choudhury et al., 2013), albeit at the periphery of the sensor-derived measure.

Physiological signals. Evidence was scant and inconsistent for some of the wearable features, such as heart-rate, electrodermal and related features (Pedrelli et al., 2020; Sano et al., 2018).

Anxiety. A combination of phone and wearable data has been shown to moderately accurately predict depression and anxiety symptoms (Moshe et al., 2021), while passive sensing has been employed to estimate the severity of social anxiety (Jacobson et al., 2020). Exercise had anxiolytic effects (Stubbs et al., 2017), but direct tests of sensing-based markers of anxiety are few.

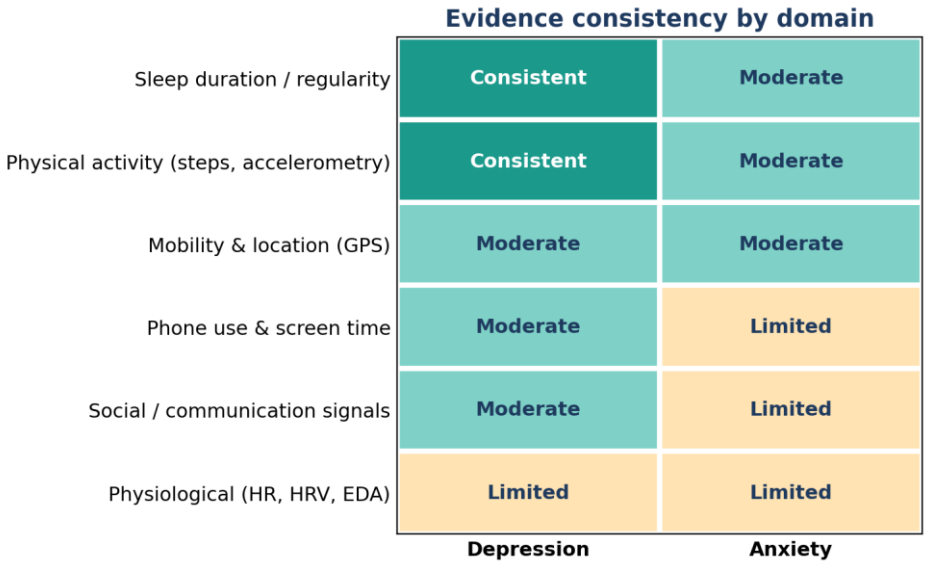


Figure 4. Consistency of evidence for each digital-biomarker domain, by condition. Ratings reflect the authors' ordinal synthesis of direction, replication and quality of findings (see Section 3.6).

Table 3: Summary of findings by domain

Domain	Depression	Anxiety	Key sources
Sleep	↓ Duration/regularity; insomnia raises risk	Altered sleep, less pronounced	Baglioni et al., 2011; Difrancesco et al., 2019
Activity	↓ Activity; protective effect	Exercise reduces symptoms	Schuch et al., 2018; Stubbs et al., 2017
Mobility/GPS	↓ Location variance; more time at home	Emerging (social anxiety)	Saeb et al., 2015; Jacobson et al., 2020
Phone use	Pattern and regularity changes	Little evidence	Saeb et al., 2015
Social signals	Reduced communication signals	Limited	Wang et al., 2018
Physiological	Adds modest information	Sparse	Pedrelli et al., 2020

Green = consistent evidence; amber = moderate or emerging; red = limited.

The effects sizes reported in the meta-analysis.

There were three syntheses quantifying sleep or activity and depression on the odds-ratio scale (Figure 5). Insomnia had an OR of 2.10 (95% CI 1.86–2.38) for incident depression (Baglioni et al., 2011). Schuch et al. (2018) found an adjusted OR of 0.83 (0.79–0.88) per increase in physical activity by a single standard deviation for incident depression, and Choi et al. (2019) found an OR of 0.74 (0.59–0.92) per unit increase in activity measured by genetic analysis of physical activity. That is, sleep disruption increases risk by approximately 2-fold, and increased activity is always protective, making both good markers for digital use to understand. Another meta-analysis identified big antidepressant effects of exercise, which reduced but remained after publication bias correction (Schuch et al., 2016). These statistics are presented as published and were not re-calculated.

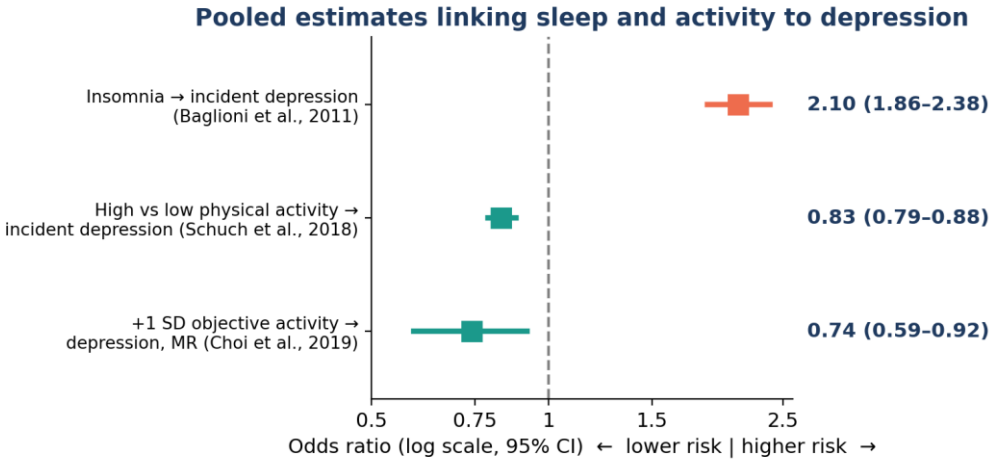


Figure 5. Pooled odds ratios from published meta-analyses (values as reported by the original authors).

The quality and limitations of the sensing studies.

Four patterns were identified from the 15 empirical studies using the adapted QUADAS-2. First, the samples were small and in most cases were composed of university students or volunteers (Jacobson & Chung, 2020; Sano et al., 2018; Wang et al., 2018). Secondly, the follow-up period was generally a single school term or weeks. Third, there were differences in both feature definitions and platforms, making comparisons difficult (Mendes et al., 2022). Fourth, many models were evaluated in the development sample, resulting in high risk of overfitting and optimistic accuracy; however, personalized models and platform-scale studies seemed to hold more promise (Nickels et al., 2021; Pratap et al., 2019). Until Burnes et al. (2011), there had been only proof-of-concept demonstrations of context sensing.

Some of the conclusions that can be drawn from the results are as follows:

1.The most convincing digital markers of depression are sleep disruption and low and irregular activity.

2. Mobility and phone-use features are promising, but require replication in clinical and community samples.

3. There is the greatest evidence gap related to anxiety.

Currently, the methodological shortcomings are the reason for the limited clinical translation, not the lack of a signal.

Discussion

Principal findings

This review indicates that there is synergy between two sets of information that are typically not considered together: large epidemiological syntheses and small sensing studies. The epidemiological literature reports the same behavioral signatures as risk factors for depression: insomnia and physical inactivity (Baglioni et al., 2011; Choi et al., 2019; Schuch et al., 2018), as do sensing studies of everyday behaviors (Difrancesco et al., 2019; Saeb et al., 2015; Wang et al., 2018). This convergence further supports the construct validity of digital biomarkers—it seems like they are measuring clinically relevant behavior and not artefacts of the device.

Interpretation for depression versus anxiety

It is important to note that there is a difference between depression and anxiety. There is a behavioral phenotype of depression which is easily detected by sensors such as withdrawal, decreased movement and sleep disturbance. Anxiety is more diverse: In social anxiety, avoidance can lead to decreased mobility; in restlessness and hyperarousal, mobility can increase or disrupt sleep. This heterogeneity may obscure simple linear relationships and necessitates disorder-specific and likely individual models (Pratap et al., 2019). Depression and anxiety share some overlapping symptoms making it hard to tell if a signal indicates one disorder or the other (Alvaro et al., 2013)

Clinical and research implications

From a clinical perspective, passive markers could be used to identify deterioration between visits, and to lessen the need for repeated questionnaires, while also allowing for interventions when needed, or "just-in-time" (JIT) (as early context sensing work has predicted (Burns et al., 2011) and the personal sensing framework has suggested (Mohr et al., 2017)). In such applications, there is a need for evidence of sensitivity and specificity and added value over self-report, the majority of studies have not yet provided. Prior to the routine use, there are also questions of privacy, consent, algorithmic bias and clinical safety that must be answered (Insel, 2017; Onnela & Rauch, 2016).

Our findings suggest priorities for research: (a) larger and clinically diverse samples, including adolescents and older adults; (b) extended follow-up to assess relapse and recovery; (c) pre-specified, standardized feature definitions and open platforms (Torous et al., 2016); (d) external validation and transparent reporting of model performance; and (e) direct assessment of digital markers of anxiety disorders.

Strengths and limitations

The strengths are adherence to PRISMA 2020, the clearly defined analysis window in 2026, the focusing on three behavioral domains, and a separation between the narrative synthesis and the effect sizes presented in published meta-analyses, which are transparent. The advantages are adherence to PRISMA 2020, clear analysis window set in 2026, focusing on three behavioral fields, and separating the narrative synthesis from effect sizes provided in published meta-

analyses, which are transparent. Restrictions must be recognized. This evidence base is a selected synthesis and not a complete list of studies in this rapidly changing field, so some relevant studies may not be included. The consistency ratings of the domains are ordinal judgments, not statistical estimates. The lack of homogeneity hampered the amalgamation of sensing studies and the literature tends to favors positive results (Schuch et al., 2016). Last, since the review is secondary, the findings will only be as valuable as the original studies.

Conclusion

Existing evidence shows that sleep and physical activity and especially smartphone-derived measures of mobility are potential digital biomarkers of depression, with consistent epidemiological evidence: insomnia significantly increases the risk of depression and increased physical activity is associated with a reduced risk. The evidence for social and phone use is moderate and for physiological is limited; the evidence for anxiety is emerging. The main impediment to clinical application is not a conceptual one, but rather methodological and limited by the use of small, short, convenience studies with little external validation. Randomized, large, longitudinal, clinically diverse and externally validated studies, with uniform features and privacy and equity considerations, are warranted for future research. For now, digital biomarkers should be considered as a supplement to — not a substitute for — — clinical evaluation.

REFERENCES

Alvaro, P. K., Roberts, R. M., & Harris, J. K. (2013). A systematic review assessing bidirectionality between sleep disturbances, anxiety, and depression. *Sleep, 36*(7), 1059–1068.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., Lombardo, C., & Riemann, D. (2011). Insomnia as a predictor of depression: A meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders, 135*(1–3), 10–19.

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal, 38*(3), 218–226.

Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., & Mohr, D. C. (2011). Harnessing context sensing to develop a mobile intervention for depression. *Journal of Medical Internet Research, 13*(3), e55.

Canzian, L., & Musolesi, M. (2015). Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 1293–1304). ACM.

Choi, K. W., Chen, C.-Y., Stein, M. B., Klimentidis, Y. C., Wang, M.-J., Koenen, K. C., & Smoller, J. W. (2019). Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample Mendelian randomization study. *JAMA Psychiatry, 76*(4), 399–408.

Cornet, V. P., & Holden, R. J. (2018). Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics, 77*, 120–132.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. *Proceedings of the International AAAI Conference on Web and Social Media, 7*(1), 128–137.

Difrancesco, S., Lamers, F., Riese, H., Merikangas, K. R., Beekman, A. T. F., van Hemert, A. M., Schoevers, R. A., & Penninx, B. W. J. H. (2019). Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: A 2-week ambulatory assessment study. *Depression and Anxiety, 36*(10), 975–986.

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA, 318*(13), 1215–1216.

Jacobson, N. C., & Chung, Y. J. (2020). Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors, 20*(12), 3572.

Jacobson, N. C., Summers, B., & Wilhelm, S. (2020). Digital biomarkers of social anxiety severity: Digital phenotyping using passive smartphone sensors. *Journal of Medical Internet Research, 22*(5), e16875.

Jacobson, N. C., Weingarden, H., & Wilhelm, S. (2019). Digital biomarkers of mood disorders and symptom change. *npj Digital Medicine, 2*, Article 3.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613.

Melcher, J., Hays, R., & Torous, J. (2020). Digital phenotyping for mental health of college students: A clinical review. *Evidence-Based Mental Health, 23*(4), 161–166.

Mendes, J. P. M., Moura, I. R., Van de Ven, P., Viana, D., Silva, F. J. S., Coutinho, L. R., Teixeira, S., Rodrigues, J. J. P. C., & Teles, A. S. (2022). Sensing apps and public data sets for digital phenotyping of mental health: Systematic review. *Journal of Medical Internet Research, 24*(2), e28735.

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology, 13*, 23–47.

Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., Mohr, D. C., & Pulkki-Råback, L. (2021). Predicting symptoms of depression and anxiety using smartphone and wearable data. *Frontiers in Psychiatry, 12*, Article 625247.

Nickels, S., Edwards, M. D., Poole, S. F., Winter, D., Gronsbell, J., Rozenkrants, B., Miller, D. P., Fleck, M., McLean, A., Peterson, B., Chen, Y., Hwang, A., Rust-Nguyen, T., Brant, S., Gammon, G., & Fromer, M. (2021). Toward a mobile platform for real-world digital measurement of depression: User-centered design, data quality, and behavioral and clinical modeling. *JMIR Mental Health, 8*(8), e27589.

Onnela, J.-P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology, 41*(7), 1691–1696.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71.

Pedrelli, P., Fedor, S., Ghandeharioun, A., Howe, E., Ionescu, D. F., Bhathena, D., Fisher, L. B., Cusin, C., Nyer, M., Yeung, A., Sangermano, L., Mischoulon, D., Alpert, J. E., & Picard, R. W. (2020). Monitoring changes in depression severity using wearable and mobile sensors. *Frontiers in Psychiatry, 11*, Article 584711.

Pratap, A., Atkins, D. C., Renn, B. N., Tanana, M. J., Mooney, S. D., Anguera, J. A., & Areán, P. A. (2019). The accuracy of passive phone sensors in predicting daily mood. *Depression and Anxiety, 36*(1), 72–81.

Rohani, D. A., Faurholt-Jepsen, M., Kessing, L. V., & Bardram, J. E. (2018). Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: Systematic review. *JMIR mHealth and uHealth, 6*(8), e165.

Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P., & Mohr, D. C. (2016). The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ, 4*, e2537.

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research, 17*(7), e175.

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E., & Picard, R. (2018). Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. *Journal of Medical Internet Research, 20*(6), e210.

Schuch, F. B., Vancampfort, D., Firth, J., Rosenbaum, S., Ward, P. B., Silva, E. S., Hallgren, M., Ponce De Leon, A., Dunn, A. L., Deslandes, A. C., Fleck, M. P., Carvalho, A. F., & Stubbs, B. (2018). Physical activity and incident depression: A meta-analysis of prospective cohort studies. *American Journal of Psychiatry, 175*(7), 631–648.

Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., & Stubbs, B. (2016). Exercise as a treatment for depression: A meta-analysis adjusting for publication bias. *Journal of Psychiatric Research, 77*, 42–51.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097.

Stubbs, B., Vancampfort, D., Rosenbaum, S., Firth, J., Cosco, T., Veronese, N., Salum, G. A., & Schuch, F. B. (2017). An examination of the anxiolytic effects of exercise for people with anxiety and stress-related disorders: A meta-analysis. *Psychiatry Research, 249*, 102–108.

Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health, 3*(2), e16.

Wang, R., Wang, W., daSilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., & Campbell, A. T. (2018). Tracking depression dynamics in college students using mobile phone and

wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(1), Article 43.

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*(8), 529–536.