The Relation between passively collected GPS mobility metrics and depressive symptoms: A systematic review and meta-analysis.

Yannik Terhorst, Johannes Knauer, Paula Philippi, Harald Baumeister

Submitted to: Journal of Medical Internet Research
on: September 11, 2023

Table of Contents

Original Manuscript **5**
Supplementary Files **28**
Multimedia Appendixes **29**
Multimedia Appendix 1 ..... **29**

The Relation between passively collected GPS mobility metrics and depressive symptoms: A systematic review and meta-analysis.

Yannik Terhorst[1] BA, MA; Johannes Knauer[1] BA, MA; Paula Philippi[2] BA, MA; Harald Baumeister[1] MD, PhD

[1]Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education University Ulm Ulm DE
[2]Department of Clinical Child and Adolescent Psychology and Psychotherapy, Institute of Psychology University of Wuppertal Wuppertal DE

Corresponding Author:
Yannik Terhorst BA, MA
Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education
University Ulm
Lise-Meitner-Str. 16, 89081 Ulm, Germany
Ulm
DE

Abstract

Background: Objective unobtrusively collected GPS features (e.g., homestay, distance) from everyday devices like smartphones may offer a promising augmentation to current assessment tools for depression. However, to date there is no systematic and meta-analytical evidence on the associations between GPS features and depression.

Objective: The present systematic review with meta-analysis investigated the between-person and within-person correlations between GPS features and depressive symptoms. Furthermore, it critically reviews the quality and potential publication bias in the field.

Methods: We searched MEDLINE, PsycInfo, Embase, CENTRAL, ACM, IEEE Xplore, PubMed, and Web of Science to identify eligible articles focusing on the correlations between GPS features and depression. In- and exclusion criteria were applied in a two-stage inclusion process conducted by two independent reviewers. Between and within-person correlations were analyzed using random effects models. Study quality was determined by comparing studies against the STROBE guidelines. Publication bias was investigated using Egger's test and funnel plots.

Results: A total of k=19 studies involving N=2,930 participants were included in the analysis. Mean age was M=28.42 (SD=18.96) with 59.64% participants being female. Significant between-person correlations between GPS features and depression were identified: Distance (r=-0.25, 95%-CI: -0.29 to -0.21), normalized entropy (r=-0.17, 95%-CI: -0.29 to -0.04), location variance (r=-0.17, 95%-CI: -0.26 to -0.04), entropy (r=-0.13, 95%-CI: -0.23 to -0.04), number of clusters (r=-0.11, 95%-CI: -0.18 to -0.03), and homestay (r=0.10, 95%-CI: 0.00 to 0.19). Studies reporting within-correlations (k=3) were too heterogenous to conduct meta-analysis. A deficiency in study quality and research standards was identified: All studies followed exploratory observational designs, but no study referenced or fully adhered to the international guidelines for reporting observational studies (STROBE). 79% of the studies were underpowered to detect a small correlation (r=.20). Results showed evidence for potential publication bias.

Conclusions: Our results provide meta-analytical evidence for between-person correlations of GPS features and depression. Hence, depression diagnostics may benefit from adding GPS features as an integral part in future assessment and expert tools. However, confirmatory studies for between-person correlations and further research on within-person correlations are needed. In addition, the methodological quality of the evidence needs to improve. Clinical Trial: https://osf.io/cwder

(JMIR Preprints 11/09/2023:51875)
DOI: https://doi.org/10.2196/preprints.51875

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

The Relation between passively collected GPS mobility metrics and depressive symptoms: A systematic review and meta-analysis.

Yannik Terhorst[1,2,3], Johannes Knauer[1], Paula Philippi[4], Harald Baumeister [1]

[1] Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, University Ulm, Lise-Meitner-Str. 16, 89081 Ulm, Germany

[2] Department of Psychology, LMU Munich, Munich, Germany

[3] DZPG (German Center for Mental Health), Partner Site München, Munich, Germany

[4] Department of Clinical Child and Adolescent Psychology and Psychotherapy, Institute of Psychology, University of Wuppertal, Gaußstr. 20, 42119, Wuppertal, Germany

*** Correspondence:**
Yannik Terhorst, Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, University Ulm, Lise-Meitner-Str. 16, 89081 Ulm, Germany, Phone: +49-(0) 731/50 32812, Email: yannik.terhorst@uni-ulm.de

Abstract

Background: Objective unobtrusively collected GPS features (e.g., homestay, distance) from everyday devices like smartphones may offer a promising augmentation to current assessment tools for depression. However, to date, there is no systematic and meta-analytical evidence on the associations between GPS features and depression.

Objective: The present systematic review with meta-analysis aimed to investigate the between-person and within-person correlations between GPS mobility and activity features and depressive symptoms, and to critically review the quality and potential publication bias in the field.

Methods: We searched MEDLINE, PsycInfo, Embase, CENTRAL, ACM, IEEE Xplore, PubMed, and Web of Science to identify eligible articles focusing on the correlations between GPS features and depression from 06.12.2022 to 24.03.2023. Inclusion and exclusion criteria were applied in a two-stage inclusion process conducted by two independent reviewers. To be eligible, studies needed to report correlations between wearable-based GPS variables (e.g., total distance) and depression symptoms measured with a validated questionnaire. Studies with underage persons and other mental health disorders were excluded. Between and within-person correlations were analyzed using random effects models. Study quality was determined by comparing studies against the STROBE guidelines. Publication bias was investigated using Egger's test and funnel plots.

Results: A total of k=19 studies involving N=2,930 participants were included in the analysis. The mean age was M=38.42 (SD=18.96) with 59.64% of participants being female. Significant between-person correlations between GPS features and depression were identified: Distance (r=-0.25, 95%-CI: -0.29 to -0.21), normalized entropy (r=-0.17, 95%-CI: -0.29 to -0.04), location variance (r=-0.17, 95%-CI: -0.26 to -0.04), entropy (r=-0.13, 95%-CI: -0.23 to -0.04), number of clusters (r=-0.11, 95%-CI: -0.18 to -0.03), and homestay (r=0.10, 95%-CI: 0.00 to 0.19). Studies reporting within-correlations (k=3) were too heterogeneous to conduct meta-analysis. A deficiency in study quality and research standards was identified: All studies followed exploratory observational designs, but no study referenced or fully adhered to the international guidelines for reporting observational studies (STROBE). 79% of the studies were underpowered to detect a small correlation (r=.20). Results showed evidence for potential publication bias.

Conclusions: Our results provide meta-analytical evidence for between-person correlations of GPS mobility and activity features and depression. Hence, depression diagnostics may benefit from adding GPS mobility and activity features as an integral part of future assessment and expert tools. However, confirmatory studies for between-person correlations and further research on within-person correlations are needed. In addition, the methodological quality of the evidence needs to improve.

Trial Registration: https://osf.io/cwder

Funding: Self-funded by the authors

Keywords: Smart Sensing, Digital Phenotyping, Depression, GPS, Meta-Analysis

Introduction

Depressive disorders are one of the most prevalent mental disorders worldwide. The global prevalence rate is estimated to be 4.4% [1,2]. Associated consequences of depression are severe not only for affected individuals (e.g., globally 43 million years lived with disability in 2017) but also for the economy and society [3–5]. Costs are even higher and long-term health consequences are more severe if the disorder is not properly diagnosed or treated [6,7]. To target the global burden caused by depression and to initiate effective treatment, timely diagnosis is a key bottleneck in healthcare [8–10].

Various diagnostic approaches like structured clinical interviews [11], clinician-rated screening scales [12,13], and self-report screening assessments [14] are available. However, their reliability and accuracy are often hindered by social desirability, recall, or confirmation biases [15–17]. Therefore, augmentations and extensions of existing approaches by objective data sources could potentially make an important contribution to the improvement in the diagnostic process of depression and other mental disorders.

Facing an ever-growing digitalization in daily living, the availability of various sensor data in smartphones, wearable devices, cars, and smart home devices may provide a way towards objective diagnosis in a timely manner and sensor-informed diagnostic support tools for clinical personnel [18–21]. In the context of medical applications, the approach of using sensor data to infer mental health is referred to as smart sensing or digital phenotyping [18,19]. In short, raw data from software (e.g., screen status of the smartphone [on/off]) and hardware-based (e.g., GPS coordinates) sensors are processed to derive higher-level features (e.g., total smartphone usage time, total distance, circadian rhythm), which are then linked to clinical symptoms or diagnoses [18,19].

Earlier studies highlight the potential of smart sensing [22]: For instance, Asare and colleagues found an Area under the Curve (AUC) of the Receiver Operating Curve (ROC) of 94.69%-99.06% in the prediction of depression status (depressed or not depressed) using a supervised machine learning model on behavioral markers from the smartphone (i.e., app usage, screen usage, and network usage features) [23]. Furthermore, studies have shown, that smart sensing data can improve prediction and increase explained variance over self-report ratings such as ecological momentary assessments [24,25]. In line with these findings, various descriptive reviews underline the potential shown in the present literature [26–33].

While there is a broad array of sensors, in the context of depression, in particular, GPS-based mobility and activity features (e.g., total distance, places visited, time spent at home, etc.) could be promising candidates for objective and unobtrusive measurement of core symptoms of depression such as reduced activity, fatigue, loss of functioning, and diminished interest [24,27,34,35]. Individual studies find medium to high correlations between depressive symptoms and GPS mobility and activity features (e.g., number of location clusters: r=-0.38, circadian movement: r=-0.34, home stay: r=0.22 [34]; see also e.g., [24,35–37]). However, the heterogeneity in the findings and in the methodology between studies is substantial [27,31]. In addition, a mean sample size of M=23.1 (SD=27.9) was reported in the systematic review of Holden and Cohen [26], clearly highlighting the issue of underpowered trials in the field. Meta-analytical evidence is highly needed to provide an estimate of the relationship between GPS mobility and activity features and depressive symptoms to determine their applicability for clinical settings and to transfer research from exploratory investigations to confirmatory studies. However, up to date no quantitative meta-analysis has been conducted on the associations between GPS mobility and activity features and depression.

Therefore, the present study systematically reviewed and meta-analyzed the current evidence on the associations between GPS mobility and activity features and depression. Besides, core aspects of study quality (e.g., adherence to international reporting guidelines, pre-registrations, presence of a-priori and post-hoc power analysis [31,38,39]) and the potential of publication bias in the field were investigated. Accordingly, the following research questions were investigated:

1. What is the pooled between-person correlation of GPS mobility and activity features and depression severity?
2. What is the pooled within-person correlation of GPS mobility and activity features and depression severity?
3. To what extent are international reporting guidelines (i.e., STROBE [38]) reflected in the studies?
4. Are small study effects as an indicator of potential publication bias observable in the literature?

Methods

Study design

The present study is a systematic review and meta-analysis. All procedures have been pre-registered in the open science framework under https://osf.io/cwder. Reporting is conducted in accordance with the updated PRISMA guidelines for the reporting of systematic reviews [40] (Multimedia Appendix 1 – Supplement 1).

Search and Eligibility Criteria

Given the interdisciplinary nature of the field we searched MEDLINE, PsycInfo, Embase, CENTRAL, ACM, IEEE Xplore, PubMed, and Web of Science. The database-specific search strings can be found in Multimedia Appendix 2 – Supplement 1. After the removal of duplicates by automatic tools of the databases, all identified articles were screened in a two-stage process: a) title and abstract screening and b) full-text screening, both performed by two independent reviewers (YT, JK). If duplicates were encountered in this screening process, duplicates were removed by hand. In addition to the database search, we searched all reference lists of the included studies for further eligible studies. This search and inclusion process was started on 06.12.2022 (database searches from 06.12. until 13.06.) and completed by 24.03.2023 (last reference search). Any disagreements between the reviewers were resolved in discussions.

To be eligible, the following inclusion criteria had to be met: (1) any kind of GPS sensor data collection; (2) the data was collected by a wearable device such as smartphone or smartwatch; (3) an assessment of depressive symptoms was conducted either by self-report or by clinical diagnostic scales, and (4) reported outcomes included correlations between the collected GPS sensor data and depressive symptoms. Studies that (1) comprised participants younger than 18 years; or (2) included participants with disorders other than depression (e.g., bipolar disorder) were excluded.

Measured variables and coding

All data was extracted by two independent reviewers (JK, YT). The following study characteristics and empirical data points were obtained:
Study characteristics: To describe the included studies, we extracted the authors' names, publication year, measures of depression (e.g., PHQ-9), GPS-sensor features (i.e., mobility and activity features such as total distance), study design, study setting, sensing framework (app name) and sample characteristics (age, gender, population, country).
Quantitative and empirical data: For the statistical analysis, correlation coefficients (both between-person and within-person) and sample size were extracted (see further analysis details below). Between-person correlations were defined as the interindividual associations between two variables (e.g., do individuals with higher time spent at home tend to show higher depression). Within-person correlations were defined as intraindividual associations of two variables across time (e.g., does an

individual tend to show higher depression in weeks with more time spent at home). Corresponding authors of eligible studies with insufficiently reported information for meta-analysis (e.g., reporting of p-values without correlation coefficients) were contacted for the missing information (i.e., repeated e-mail containing a study description, and extraction template for the needed information). Furthermore, all included studies were compared against international guidelines for reporting observational studies (STROBE [38]). In addition, the presence of a pre-registration and a-priori or post-hoc power-analysis were rated as additional criteria (see assessment of research standards and small study effects below). The rating was performed independently by two researchers (JK, YT). Disagreements were resolved in discussion. Study designs other than exploratory observational intensive longitudinal designs were planned to be compared against corresponding guidelines of the equator network (https://www.equator-network.org/) but were not present in the included studies.

Statistical Analysis

Meta-analysis of correlation

We conducted a random-effects meta-analysis of correlations. Pooling was based on inverse variance weighting [41]. Maximum likelihood was used as the estimator. Following the Cochrane handbook for meta-analysis, the minimum required number of studies reporting on the correlations between a feature (e.g., home stay) and depression to run meta-analysis was two [42]. For all significance tests and range of confidence intervals, alpha was set to 5%. Heterogeneity of effect sizes was evaluated by I^2 [43]. For heterogeneity, we defined an I^2 of 25% as the threshold for low, 50% for moderate, and 75% for high heterogeneity in the present meta-analysis. 95% confidence and prediction intervals were calculated along the pooled correlation estimates.

Assessment of research standards and small study effects

As a general criterion for the adherence to international reporting guidelines, we assessed the reference and adherence to STROBE [38] of the included observational studies. The STROBE checklist was rated for each study by two independent researchers. Disagreements were resolved in discussion.

Facing the replication crisis in research [39,44], we additionally investigated the percentage of pre-registrations and the presence of a-priori and post-hoc power-calculation. Potential publication bias was investigated using (1) funnel plots, and (2) Egger's test [45,46]. Egger's test was only conducted in analyses with at least ten studies [47]. Funnel plots display the effect sizes reported in studies (x-axis) in respect to the standard error (y-axis). The core assumption is that in case of no publication bias, all points in the funnel plot should be distributed equally around an average effect with the high precision studies (low standard error) at the top of the figure and closely around the average, and studies with low precision (high standard error) broadly distributed around the average effect. In contrast to this, a scenario with publication bias (e.g., small studies are only published if they show high effects, while large studies with lower effects are published either way due to the number of participants) would lead to an asymmetric funnel plot, where an association between standard error and reported effect is visible. Capitalizing on this idea, Eggert's test is a regression model investigating whether the standard error is significantly influencing the average reported effect. Both asymmetries in funnel plots and significant Eggert tests indicate potential publication bias, and hence, potentially biased meta-analytically results (e.g., systematic overestimation of the true effect due to unpublished high standard error studies with no effects). For a more in-depth introduction see (e.g., [41]).

Software

All analyses were conducted in R. Meta was used as the core package in the analysis [48]. For an

overview of all loaded packages and versions see Multimedia Appendix 3 – Supplement 1. Both the analysis code and the dataset containing the correlations are available under CC-BY 4.0 license at https://osf.io/ce45a/.

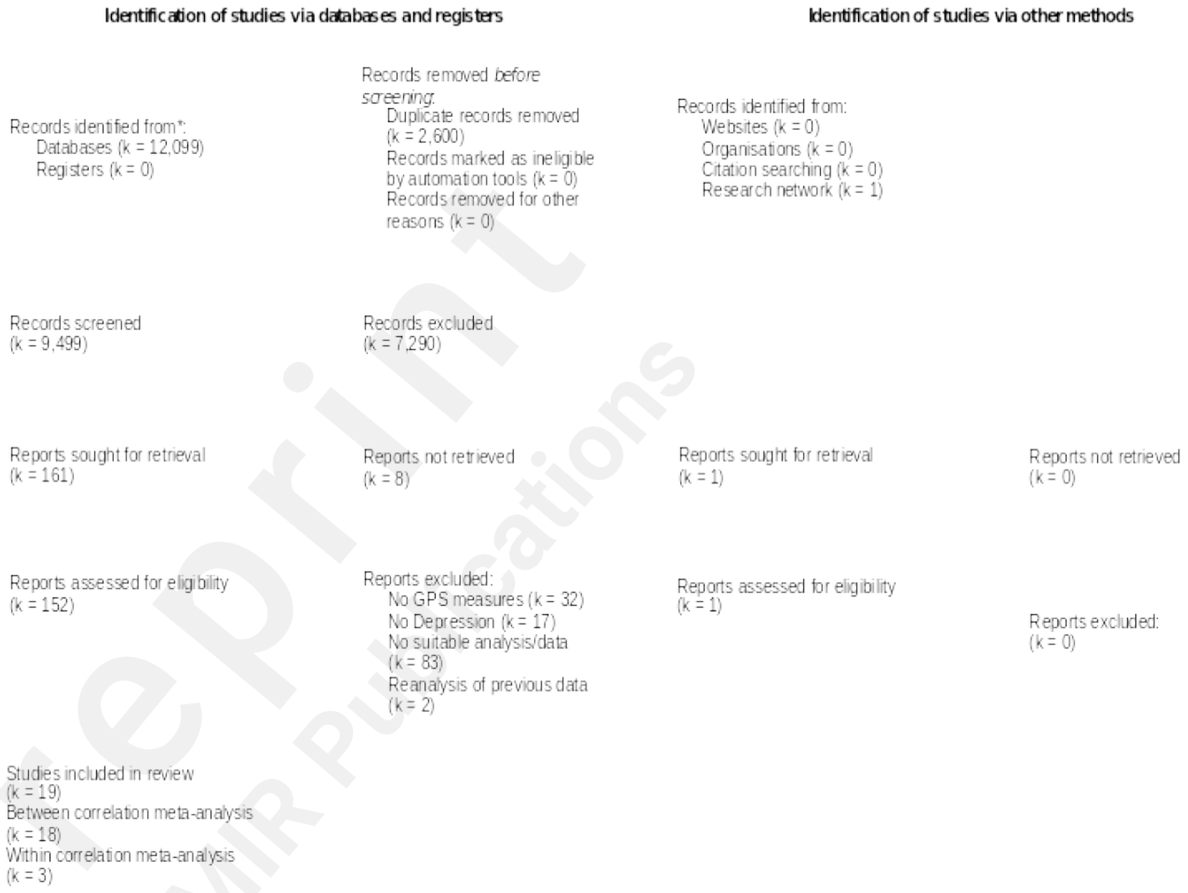
Results

We identified a total of k= 9,499 unique records in the systematic literature search. One additional study was screened and included after being forwarded by researchers contacted due to insufficient data reported in an identified study. In the following screening and inclusion process conducted by two independent reviewers, k=19 studies were finally included [24,34–37,49–60]. Two of the studies were eligible for between- and within-person correlation meta-analysis [34,53]. For further details please see the PRIMSA flow-chart in Figure 1.

Combined, the included studies comprised a total of N = 2,930 participants. The sample sizes ranged from a minimum of n = 18 to n = 1046 (M=154.21, median=72, SD=235.54). The mean age across all included studies was M=38.42 (SD=18.96). The average percentage of female participants was 59.64% (SD=22.99%). Eight (42.11%) of the included studies explicitly targeted students and young adults (e.g., recruitment at universities), three studies (15.79%) recruited from the general public, four studies (21.05%) used non-tailored online and social media recruitment, two studies (10.53%) targeted older adults, and two studies (10.53%) provided no further information on the recruitment strategies and targeted population. Please see Table 1 for further details on the descriptive characteristics of the studies.

A total of k=13 unique sensing frameworks were applied among the included studies. PurpleRobot was the most frequently used framework (k=3, 15.79%). Based on the sensed GPS data, the studies reported n=12 distinct activity and mobility features. The number of unique studies per activity and mobility feature ranged from k=1 to k=14 and a combined sample size of n=69 to n=2287 per feature. A summary of the features can be found in Table 2 for between-person features and Table 3 for within-person features.

Figure 1. Study Flow.



Note: * MEDLINE, PsycInfo, Embase, CENTRAL, ACM, IEEE Xplore, PubMed, and Web of Science

Table 1 Main Characteristics of Included Studies.

Reference	Country	N	Depression scale	Mean age (SD)	Software	Period	Target population	Severity
Boukhechba et al. (2018)	USA	72	DASS	19.8 (2.4)	Sensus	2 weeks	Students/Youth	3.5
Canzian and Musolesi (2015)	UK	28	PHQ-8	31	Mood- Traces	Avg. 71 days	General public	/
Currey and Torous (2022)	USA	147	PHQ-9		mind-LAMP	4 weeks	Students/Youth	/
DeMasi and Recht (2017)	USA	33	BDI		/	8 weeks	Students/Youth	12.7
Di Matteo (2021)	CAN	71	PHQ-8	30.6 (9.4)	Logger	2 weeks	Online recruitment	9.1
Farhan et al. (2016)	USA	79	PHQ-9		Life-Rhythm	/	Students/Youth	/
Giannouli et al. (2019)	GER	69	GDS	69.5 (4.9)	Ufall & custom app	1 week	Older adults	1.39
Lu et al. (2018)	USA	103	QIDS		Life-Rhythm & Fitbit	/	Students/Youth	/
MacLeod et al. (2021)	CAN	121	CES-DC	18 (2.76)	Custom app	2 weeks	Students/Youth	32.6
Moshe et al. (2021)	/	55	DASS-21	42.8 (11.6)	AWARE	30 days	Online recruitment	3.8
Nickels et al. (2021)	USA	379	PHQ-9		/	12 weeks	Depressed & non-depressed	
Saeb et al. (2015)	USA	28	PHQ-9	28.9 (10.1)	Purple-Robot	2 weeks	General public	5.6
Saeb et al. (2015a)	/	18	PHQ-9		Purple-Robot	2 weeks	Online recruitment	5.8
Saeb et al. (2016)	/	48	PHQ-9		Student-Life	10 weeks	Students/Youth	/
Saeb et al. (2017)	USA	206	PHQ-9	39.3 (10.3)	Purple-Robot	6 weeks	General public	9.7
Tung et al. (2014)	CAN	54	GDS	72.6	VALMA	3 days	Older adults	/
Wang et al. (2018)	USA	83	PHQ-8 & PHQ-4	20.1	Student-Life	9 weeks	Students/Youth	6.1
Zhang et al. (2022)	NL, ESP, UK	290	PHQ-8	/	/	2 weeks	/	/

Table 2 Between person features.

details_cor	gps_var	Definition	studies	freq	n_total	n_mean	n_sd
between	homestay	Percentage of time spent home.	Boukhechba_2018, Currey_2022, DiMatteo_2021, Farhan_2016_android, Farhan_2016_iOS, Lu_2018_android, Lu_2018_iOS, Mohr_2023, Moshe_2021, Nickels_2021, Saeb_2015a, Saeb_2015b, Saeb_2016, Saeb_2017	14	2216	158.29	266.40
between	location variance	Variance of latitude and longitude values.	Currey_2022, DeMasi_2017, DiMatteo_2021, Farhan_2016_android, Farhan_2016_iOS, Lu_2018_android, Lu_2018_iOS, Mohr_2023, Moshe_2021, Nickels_2021, Saeb_2015a, Saeb_2015b, Saeb_2016, Zhang_2022	14	2287	163.36	276.10
between	entropy	Distribution of time spent at different location clusters.	Currey_2022, DiMatteo_2021, Farhan_2016_android, Farhan_2016_iOS, Lu_2018_android, Lu_2018_iOS, Mohr_2023, Moshe_2021, Nickels_2021, Saeb_2015a, Saeb_2015b, Saeb_2016, Zhang_2022	13	2254	173.38	284.71
between	n clusters	Number of unique location clusters.	Currey_2022, DiMatteo_2021, Farhan_2016_android, Farhan_2016_iOS, Lu_2018_android, Lu_2018_iOS, Mohr_2023, Nickels_2021, Saeb_2015a, Saeb_2015b, Saeb_2016, Wang_2018, Zhang_2022	13	2282	175.54	283.85
between	norm entropy	Entropy normalized by the number of location clusters.	DiMatteo_2021, Farhan_2016_android, Farhan_2016_iOS, Lu_2018_android, Lu_2018_iOS, MacLeod_2021, Mohr_2023, Moshe_2021, Saeb_2015a, Saeb_2015b, Saeb_2016, Zhang_2022	12	1849	154.08	290.46
between	distance	Total distance between coordinates.	DiMatteo_2021, Lu_2018_android, Lu_2018_iOS, MacLeod_2021, Mohr_2023, Moshe_2021, Saeb_2015a, Saeb_2015b, Saeb_2016, Tung_2014, Zhang_2022	11	1824	165.82	301.64
between	speed moving	Speed at GPS data point collection.	Farhan_2016_android, Farhan_2016_iOS, Lu_2018_android, Lu_2018_iOS, Mohr_2023, Saeb_2016	6	1266	211.00	409.36
between	time moving	Time spent in moving states in percentage.	Farhan_2016_android, Farhan_2016_iOS, Lu_2018_android, Lu_2018_iOS, MacLeod_2021, Zhang_2022, DiMatteo_2021, Saeb_2015a, Saeb_2015b, Saeb_2016	10	748	74.80	81.54
between	circadian movement	Amount of energy in frequency periods (e.g., 30 min) based on least-squares spectral analysis.	DiMatteo_2021, Mohr_2023, Saeb_2015a, Saeb_2015b, Saeb_2016	5	1201	240.20	450.90
between	life space area	Convex hull of GPS coordinates	Giannouli_2019	1	69	69.00	NA
between	maximum action range	Longest (straight-line) distance away from home	Giannouli_2019	1	69	69.00	NA

Table 3 Within person features.

details_cor	gps_var	Definition	studies	freq	n_total	n_mean	n_sd
within	distance	Total distance between coordinates.	Canzian_2015, Saeb_2016, Zhang_2022	3	356	118.67	148.46
within	entropy	Distribution of time spent at different location clusters.	Saeb_2016, Zhang_2022	2	328	164.00	178.19
within	location_variance	Variance of latitude and longitude values.	Saeb_2016, Zhang_2022	2	328	164.00	178.19
within	n_clusters	Number of unique location clusters.	Saeb_2016, Zhang_2022	2	328	164.00	178.19
within	norm_entropy	Entropy normalized by the number of location clusters.	Saeb_2016, Zhang_2022	2	328	164.00	178.19
within	circadian_movement	Amount of energy in frequency periods (e.g., 30 min) based on least-squares spectral analysis.	Saeb_2016	1	38	38.00	NA
within	homestay	Percentage of time spent home.	Saeb_2016	1	38	38.00	NA
within	speed_moving	Speed at GPS data point collection.	Saeb_2016	1	38	38.00	NA
within	time_moving	Time spent in moving states in percentage.	Saeb_2016, Zhang_2022	2	328	164.00	178.19

Meta-analysis: Between person correlations

Distance was the mobility and activity feature most strongly associated with depressive symptoms, with a meta-analytically pooled between-person correlation of r = -0.25 (95%-CI: -0.29 to -0.21), followed by normalized entropy (r = -0.17, 95%-CI: -0.29 to -0.04), location variance (r = -0.17, 95%-CI: -0.26 to -0.06), entropy (r = -0.13, 95%-CI: -0.23 to -0.04), number of clusters (r = -0.11, 95%-CI: -0.18 to 0.03) and home stay (r = 0.10, 95%-CI: 0.00 to 0.19). In contrast, the features circadian movement, transition time, speed moving and time moving indicate no significant correlations with depression. Please see Table 4 for a summary of all pooled between-person correlations. Feature-specific forest plots can be found in Multimedia Appendix 4 – Supplement 1.

Table 4. Pooled between-person correlations.

Feature	N	Cor	CI	Prediction Interval	I^2
circadian movement	1201	-0.36	-0.71 to 0.12	-0.91 to 0.64	86%
distance	1824	-0.25	-0.29 to -0.21	-0.30 to -0.20	0%
norm entropy	1849	-0.17	-0.29 to -0.04	-0.44 to 0.13	69%
location variance	2287	-0.17	-0.26 to -0.06	-0.36 to 0.04	58%
entropy	2254	-0.13	-0.23 to -0.04	-0.35 to 0.10	57%
n_cluster	2282	-0.11	-0.18 to -0.03	-0.25 to 0.04	36%
homestay	2216	0.10	0.00 to 0.19	-0.09 to 0.27	50%
speed moving	1266	-0.01	-0.07 to 0.06	-0.09 to 0.07	0%
time moving	748	-0.05	-0.21 to 0.11	-0.45 to 0.36	68%

Meta-Analysis: Within-Person Correlations

The three identified studies reporting on within-person correlations (see Table 3) differed widely in the applied methodology of analysis. Saeb and colleagues [34] correlated GPS mobility and activity features with the change in PHQ-9 scores, while Zhang and colleagues [53] used autoregressive models to estimate the correlations over time, and Canzian and colleagues [57] applied time series analysis to investigate the correlations within each participant across multiple days. Due to this heterogeneity in these studies to derive the reported correlations, we did not perform a meta-analysis for within-person correlations.

Assessment of Research Standards and Small Study Effects

All included studies followed an observational study design. Assessment of the international reporting standards for observational studies revealed that not a single study referenced the international reporting guidelines STROBE. Comparing the studies against the STROBE checklist showed an overall agreement across all items of 73.84%. Only 50% of the studies reported how potential sources of bias were addressed. Regarding missing data, 33.33% reported how missing data was handled and 16.67% reported rates

of missingness for the variables of interest. A total of 38.89% of the studies listed reasons for the exclusion of participants at each stage of the study. For all STROBE ratings please see Multimedia Appendix 5 – Supplement 1.

In addition, neither pre-registrations nor study protocols were found for the included studies. A-priori power analyses for sample size planning or post-hoc for discussion were also not reported in any of the included studies. Power analyses of the included studies indicate that k=16 studies were sufficiently powered with a power of 80% to detect a correlation of r=.50 (84.21%), k=7 (36.84%) for r=.30, k=4 (21.05%) for r=.20, and k=1 (5.26%) for r=.10.

For all mobility and activity features except for time moving Egger's test showed a significant asymmetry in the funnel plot (Table 6). Funnel plots for features with less than k=10 studies additionally indicate asymmetry. Please see Multimedia Appendix 6 (Supplement 1) for the funnel plots of all between-person features.

Table 6. Eggert's test for funnel plot asymmetry

Feature	Intercept	95%-CI	p
homestay	1.55	0.56 to 2.54	.010
location variance	-1.44	-2.58 to -0.31	.029
entropy	-2.11	-2.87 to -1.34	<.001
n clusters	-1.29	-2.22 to -0.35	.021
norm entropy	-2.30	-3.71 to -1.33	<.001
distance	1.01	0.46 to 1.56	.006
time moving	-0.42	-2.74 to 2.66	.976

Discussion

This systematic review with meta-analysis provides pooled correlation estimates for the between-person correlation of GPS-based smart sensing mobility and activity features and depression. We identified robust small to medium between-person correlations for multiple mobility and activity features (i.e., distance, normalized entropy, entropy, homestay, number of clusters, and homestay). In contrast to the findings on between-person correlations, the current literature for within-person correlations did not allow for meaningful meta-analyses. Furthermore, the present study clearly highlights a lack of quality in the literature. While all studies followed an observational study design, not a single study referenced the international reporting guidelines for observational studies (STROBE [38]). Although the reporting in studies addressed STROBE items in most cases (73.84%), deficits were found in addressing sources of bias and missing data. Moreover, we found strong evidence for small study effects potentially indicating publication bias. Most studies were underpowered to detect correlations of the magnitude indicated by the meta-analysis.

Overall, the present analysis highlights the potential of GPS features to infer depression, which is in line with previous published descriptive reviews in the field [26,27,30,32,33]. For instance, with a total sample size of N=1824 and homogenous findings across studies (I^2=0%), distance (r=-0.25) might be a promising objective and unobtrusively collectible GPS feature in expert systems for (assisted) diagnoses, just-in-time interventions and personalization of treatment of depression in future [20,61]. Analogously, normalized entropy, entropy, homestay, number of clusters, and homestay were significant markers (|r|: 0.10 to 0.17). However, it is important to note that all studies followed an exploratory design. Both the correlations and the clinical feasibility of such smart sensing augmented expert systems need to be proven in confirmatory trials before clinical applications. The identified pooled correlations and their confidence intervals offer a strong foundation for a-priori power analysis to guide confirmatory studies in their study and sample planning. [62,63]

However, while some of the investigated correlations are statistically significant and pooled estimates can serve as a basis for future study planning from a statistical point, this does not release from the question of what a clinically relevant correlation is. The process involved in obtaining GPS features (e.g., collection of GPS-coordinates) as well as the features itself (e.g., how much time was spent at home or at work) are highly sensitive. Hence, a discussion taking the clinical significance as well as ethical and privacy issues into account is of utmost importance [20,21,61,64–66]. Furthermore, it needs to be discussed to which extent GPS features as a standalone metric can be used in the diagnostic process or rather as an add-on to existing procedures and measurements (e.g., patient-reported outcome, ecological momentary assessment, medical record data) [20,24,67]. This discussion would strongly benefit from future studies illuminating these questions from patients' and health care providers' (e.g., psychotherapists) perspectives.

Another question in the field is, whether GPS features cannot only serve to determine depression from a between-person perspective but also to inform researchers and

clinicians about changes within persons (e.g., is the increase of time spent at home a reliable indicator for rising depression severity?). The present systematic review showed too sparse evidence and too high heterogeneity in applied analyses to run a meta-analysis on the so far conducted studies. Although the present study cannot provide answers regarding the robustness across studies and hence evidence for the significance of within-person correlations, in particular, the included study by Zhang and colleagues on the longitudinal relationship of GPS mobility and activity features and depression [53] found higher within-person correlations than between person correlations underlining the potential benefit of within-person features in clinical applications and the importance of further studies on within-person correlations.

In line with previous reviews on the methods used in smart sensing studies [26,27,31], this study critically highlights gaps in the adherence to international reporting guidelines, potential publication bias, and high heterogeneity in assessment and analysis methodology. Guidelines like STROBE [38] for observational studies and other study design specific guidelines (e.g., https://www.equator-network.org/) offer a strong starting point to increase the study and especially reporting quality in the field. However, research on smart sensing and objective sensor data collection comes with its own challenges like a) high heterogeneity in hardware (e.g. used devices) and GPS sampling (e.g., many different manufacturers of sensors and variety in the precision of sensing and data quality), b) a plethora of data preprocessing decisions (e.g., sampling rate, outliner detection or feature selection and definition), c) missing data handling in large fine-grained data sets, and d) analyses methodology (e.g., how to deal with dependencies resulting from multiple measurements from the same individual [24]) [31]. Pre-registration and open-access scripts for data preparation and analyses and extensions to existing international guidelines (e.g., standards for feature calculation, reporting guidelines, missing data handling, etc.) are highly needed to move towards a standardized and reliable research field.

The current lack of standardization and heterogeneity between studies in assessment frameworks, samples (e.g., age range), sensor sampling rates, devices, feature calculation, methods of handling missing data or accounting for dependencies in the data structure should also be considered when interpreting the present findings. Due to the already low number of studies, we were unable to conduct meaningful sensitivity analysis on more homogenous groups of studies (e.g., based on data quality, devices) or meta-regression models to control for other variables (e.g., for systematic mobility and activity differences across age groups). As the number of studies and their quality are likely to increase over time, analyses on homogenous groups and meta-regression analyses might become feasible in future. Nevertheless, we want to point out that the heterogeneity for some features (e.g., distance) was very low, underlining the robustness of the pooled correlation estimates for these features. Moreover, the weighted pooling in the meta-analysis gave stronger influence on larger studies with more precise estimations of correlations, counteracting the influence of small studies to some extent [68].

Besides analyses on GPS mobility and activity features in more homogenous study groups, other GPS features should be focused on in the future. For instance, GPS sensor

data allows to derive environmental features such as green space, blue space, air pollution, or even regional data like social deprivation scores (e.g., ZIP-code based). First studies, using smartphone-based GPS collection to infer depression based on such environmental GPS features are promising, and it needs to be investigated to what extent such features show meta-analytically robust correlations [69–71]. Analogously, meta-analytical evidence for other sensors and features is lacking (e.g., smartphone screen usage, app usage, and communication and sociability features) and of high importance to inform researchers and practitioners on which objective markers might be suited for depression diagnosis, early-warning, just-in-time interventions, and other clinical applications [21,30–32,61,65,72]. Furthermore, extensions to other prevalent mental disorders (e.g., anxiety) and conditions of leading disease burden (e.g., pain) are highly needed to investigate the potential of unobtrusive and objective smart sensing data in health care.

Conclusions

This present systematic review and meta-analysis provides evidence for small to medium between-person correlations of GPS mobility and activity features collected via smartphones and depression. In the future, GPS mobility and activity features such as distance may become an important augmentation in the assessment of depression and clinical applications (e.g., decision support systems). However, replications in confirmatory studies and improvements in the study quality and standards in the field are needed to draw robust conclusions. Besides, more research on the within-person correlations is necessary to determine the potential of GPS features in not only between-person applications but also from a longitudinal perspective. To fully exploit the potential of smart sensing, further research on associations of depression and other mental disorders with GPS mobility and activity features as well as environmental GPS features (e.g., green space) and other sensor modalities (e.g., smartphone usage features), and how smart sensing features can be integrated in existing information systems and complex prediction models along patient-reported outcomes, clinician ratings, ecological momentary assessments, and medical record data is of high importance.

Acknowledgments

We would like to thank all the student assistants involved in the project. Generative AI tools (e.g., ChatGPT) were not used in any portion of the manuscript writing.

Conflicts of Interest

None declared.

Data Availability

Both the analysis code and the dataset containing the correlations are available under CC-BY 4.0 license at https://osf.io/ce45a/.

Author Contributions

Conceptualization: YT; Data curation: YT, PP, JK; Formal Analysis: YT; Funding acquisition: HB; Investigation: YT, PP, JK; Methodology: YT; Project administration: YT; Resources: YT, PP, JK, HB; Software: YT, JK ; Supervision: YT, HB; Validation: YT, JK; Visualization: YT; Writing – original draft: YT; Writing – review & editing: YT, PP, JK, HB.

Funding

This study was self-funded by the authors. YT is supported by the initial phase of the German Center for Mental Health [Deutsches Zentrum für Psychische Gesundheit (DZPG), grant: 01EE2303A].

References

[1] World Health Organization. Depression and other common mental disorders: global health estimates 2017.

[2] Walker ER, McGee RE, Druss BG. Mortality in Mental Disorders and Global Disease Burden Implications. JAMA Psychiatry 2015;72:334. https://doi.org/10.1001/jamapsychiatry.2014.2502.

[3] James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789–858. https://doi.org/10.1016/S0140-6736(18)32279-7.

[4] Greenberg PE, Fournier A-A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015;76:155–62. https://doi.org/10.4088/JCP.14m09298.

[5] Greenberg PE, Fournier A-A, Sisitsky T, Simes M, Berman R, Koenigsberg SH, et al. The Economic Burden of Adults with Major Depressive Disorder in the United States (2010 and 2018). Pharmacoeconomics 2021;39:653–65. https://doi.org/10.1007/s40273-021-01019-4.

[6] Trautmann S, Rehm J, Wittchen H-U. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016;17:1245–9. https://doi.org/10.15252/embr.201642951.

[7] Saarni SI, Suvisaari J, Sintonen H, Pirkola S, Koskinen S, Aromaa A, et al. Impact of psychiatric disorders on health-related quality of life: general population survey. Br J Psychiatry 2007;190:326–32. https://doi.org/10.1192/bjp.bp.106.025106.

[8] Kramer T, Als L, Garralda ME. Challenges to primary care in diagnosing and managing depression in children and young people. BMJ 2015;350:h2512. https://doi.org/10.1136/BMJ.H2512.

[9] Wurcel V, Cicchetti A, Garrison L, Kip MMA, Koffijberg H, Kolbe A, et al. The Value of Diagnostic Information in Personalised Healthcare: A Comprehensive

Concept to Facilitate Bringing This Technology into Healthcare Systems. Public Health Genomics 2019;22:8–15. https://doi.org/10.1159/000501832.
[10] Kroenke K. Depression screening and management in primary care. Fam Pract 2018;35:1–3. https://doi.org/10.1093/FAMPRA/CMX129.
[11] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. 2013. https://doi.org/10.1176/APPI.BOOKS.9780890425596.
[12] Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003;54:573–83. https://doi.org/10.1016/S0006-3223(02)01866-8.
[13] Fried EI, Flake JK, Robinaugh DJ. Revisiting the theoretical and methodological foundations of depression measurement. Nat Rev Psychol 2022;1:358–68. https://doi.org/10.1038/s44159-022-00050-2.
[14] Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, et al. Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression. JAMA 2020;323:2290. https://doi.org/10.1001/jama.2020.6504.
[15] Jacobucci R, Grimm KJ. Machine Learning and Psychological Research: The Unexplored Effect of Measurement. Perspect Psychol Sci 2020;15:809–16. https://doi.org/10.1177/1745691620902467.
[16] McNamara ME, Zisser M, Beevers CG, Shumake J. Not just “big” data: Importance of sample size, measurement error, and uninformative predictors for developing prognostic models for digital interventions. Behav Res Ther 2022;153:104086. https://doi.org/10.1016/J.BRAT.2022.104086.
[17] Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211. https://doi.org/10.2147/JMDH.S104807.
[18] Onnela J-P, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. Neuropsychopharmacology 2016;41:1691–6. https://doi.org/10.1038/npp.2016.7.
[19] Garatva P, Terhorst Y, Messner E-M, Karlen W, Pryss R, Baumeister H. Smart Sensors for Health Research and Improvement. In: Montag C, Baumeister H, editors. Digit. Phenotyping Mob. Sens. 2nd ed., Berlin: Springer; 2023, p. 395–411. https://doi.org/10.1007/978-3-030-98546-2_23.
[20] Terhorst Y, Knauer J, Baumeister H. Smart Sensing Enhanced Diagnostic Expert Systems. In: Montag C, Baumeister H, editors. Digit. Phenotyping Mob. Sens. 2nd ed., Berlin: Springer; 2023, p. 413–25. https://doi.org/10.1007/978-3-030-98546-2_24.
[21] Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017;13:23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949.
[22] Abd-Alrazaq A, AlSaad R, Shuweihdi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in

detecting and predicting depression. Npj Digit Med 2023;6:84. https://doi.org/10.1038/s41746-023-00828-5.

[23] Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting Depression From Smartphone Behavioral Markers Using Machine Learning Methods, Hyperparameter Optimization, and Feature Importance Analysis: Exploratory Study. JMIR MHealth UHealth 2021;9:e26540. https://doi.org/10.2196/26540.

[24] Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. Front Psychiatry 2021;12:625247. https://doi.org/10.3389/fpsyt.2021.625247.

[25] Terhorst Y, Messner E-M, Opoku Asare K, Montag C, Kannen C, Baumeister H. Which Smartphone-Based Sensing Features Matter in Depression Prediction? Results from an observation study. [Manuscript submitted for publication]. Department of Clinical Psychology and Psychotherapy, Ulm University 2024. https://doi.org/10.2196/preprints.55308.

[26] Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018;77:120–32. https://doi.org/10.1016/j.jbi.2017.12.008.

[27] Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review. JMIR MHealth UHealth 2018;6:e165. https://doi.org/10.2196/mhealth.9691.

[28] Benoit J, Onyeaka H, Keshavan M, Torous J. Systematic Review of Digital Phenotyping and Machine Learning in Psychosis Spectrum Illnesses. Harv Rev Psychiatry 2020;28:296–304. https://doi.org/10.1097/HRP.0000000000000268.

[29] Dlima SD, Shevade S, Menezes SR, Ganju A. Digital Phenotyping in Health Using Machine Learning Approaches: Scoping Review. JMIR Bioinforma Biotechnol 2022;3:e39618. https://doi.org/10.2196/39618.

[30] Nouman M, Khoo SY, Mahmud MAP, Kouzani AZ. Recent Advances in Contactless Sensing Technologies for Mental Health Monitoring. IEEE Internet Things J 2022;9:274–97. https://doi.org/10.1109/JIOT.2021.3097801.

[31] De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. Npj Digit Med 2022;5:3. https://doi.org/10.1038/s41746-021-00548-8.

[32] Zarate D, Stavropoulos V, Ball M, de Sena Collier G, Jacobson NC. Correction: Exploring the digital footprint of depression: a PRISMA systematic literature review of the empirical evidence. BMC Psychiatry 2022;22:530. https://doi.org/10.1186/s12888-022-04153-1.

[33] Peterson B, Gonzalez D, Perez-Haddock Y, Frias J, Tourgeman I. PO110 / #755 THE UTILITY OF PASSIVE AND PORTABLE SENSOR DATA FOR MONITORING THE SYMPTOMATOLOGY OF DEPRESSION AND ANXIETY. Neuromodulation Technol Neural Interface 2022;25:S247.

https://doi.org/10.1016/j.neurom.2022.08.283.
[34] Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537. https://doi.org/10.7717/peerj.2537.
[35] Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015;17:e175. https://doi.org/10.2196/jmir.4273.
[36] Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proc ACM Interactive, Mobile, Wearable Ubiquitous Technol 2018;2:1–26. https://doi.org/10.1145/3191775.
[37] Currey D, Torous J. Digital phenotyping correlations in larger mental health samples: analysis and replication. BJPsych Open 2022;8:1–7. https://doi.org/10.1192/bjo.2022.507.
[38] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. Ann Intern Med 2007;147:573. https://doi.org/10.7326/0003-4819-147-8-200710160-00010.
[39] Maxwell SE, Lau MY, Howard GS. Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? Am Psychol 2015;70:487–98. https://doi.org/10.1037/a0039400.
[40] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://doi.org/10.1136/bmj.n71.
[41] Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis With R: A Hands-On Guide. 1st ed. Boca Raton, FL and London: Chapman & Hall/CRC Press; 2021.
[42] Cochrane Handbook for Systematic Reviews of Interventions | Cochrane Training n.d. https://training.cochrane.org/handbook/current (accessed July 16, 2023).
[43] Borenstein M, Higgins J, Hedges L V, Rothstein HR. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. Res Synth Methods 2017;8:5–18.
[44] Aarts AA, Anderson JE, Anderson CJ, Attridge PR, Attwood A, Axt J, et al. Estimating the reproducibility of psychological science. Science (80-) 2015;349. https://doi.org/10.1126/SCIENCE.AAC4716/SUPPL_FILE/AARTS-SM.PDF.
[45] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629–34. https://doi.org/10.1136/BMJ.315.7109.629.
[46] Mathur MB, VanderWeele TJ. Sensitivity analysis for publication bias in meta-analyses. J R Stat Soc Ser C Appl Stat 2020;69:1091–119. https://doi.org/10.1111/rssc.12440.
[47] Sterne JAC, Sutton AJ, Ioannidis J, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in

meta-analyses of randomised controlled trials. Br Med J 2011;343:302–7.
[48] Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. BMJ Ment Heal 2019;22:153–60. https://doi.org/10.1136/ebmental-2019-300117.
[49] MacLeod L, Suruliraj B, Gall D, Bessenyei K, Hamm S, Romkey I, et al. A mobile sensing app to monitor youth mental health: Observational pilot study. JMIR MHealth UHealth 2021;9. https://doi.org/10.2196/20638.
[50] Lu J, Shang C, Yue C, Morillo R, Ware S, Kamath J, et al. Joint Modeling of Heterogeneous Sensing Data for Depression Assessment via Multi-task Learning. Proc ACM Interactive, Mobile, Wearable Ubiquitous Technol 2018;2:1–21. https://doi.org/10.1145/3191753.
[51] Nickels S, Edwards MD, Poole SF, Winter D, Gronsbell J, Rozenkrants B, et al. Toward a mobile platform for real-world digital measurement of depression: User-centered design, data quality, and behavioral and clinical modeling. JMIR Ment Heal 2021;8. https://doi.org/10.2196/27589.
[52] Tung JY, Rose RV, Gammada E, Lam I, Roy EA, Black SE, et al. Measuring life space in older adults with mild-to-moderate Alzheimer's disease using mobile phone GPS. Gerontology 2014;60:154–62. https://doi.org/10.1159/000355669.
[53] Zhang Y, Folarin AA, Sun S, Cummins N, Vairavan S, Bendayan R, et al. Longitudinal Relationships Between Depressive Symptom Severity and Phone-Measured Mobility: Dynamic Structural Equation Modeling Study. JMIR Ment Heal 2022;9. https://doi.org/10.2196/34898.
[54] Saeb S, Zhang M, Kwasny M, Karr C, Kording K, Mohr D. The Relationship between Clinical, Momentary, and Sensor-based Assessment of Depression. Proc. 9th Int. Conf. Pervasive Comput. Technol. Healthc., vol. 3, ICST; 2015, p. 103–11. https://doi.org/10.4108/icst.pervasivehealth.2015.259034.
[55] Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety. JMIR MHealth UHealth 2017;5:e112. https://doi.org/10.2196/mhealth.7297.
[56] Boukhechba M, Daros AR, Fua K, Chow PI, Teachman BA, Barnes LE. DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. Smart Heal 2018;9–10:192–203. https://doi.org/10.1016/j.smhl.2018.07.005.
[57] Canzian L, Musolesi M. Trajectories of depression: Unobtrusive Monitoring of Depressive States by means of Smartphone Mobility Traces Analysis. Proc 2015 ACM Int Jt Conf Pervasive Ubiquitous Comput - UbiComp '15 2015:1293–304. https://doi.org/10.1145/2750858.2805845.
[58] DeMasi O, Recht B. A step towards quantifying when an algorithm can and cannot predict an individual's wellbeing. UbiComp/ISWC 2017 - Adjun Proc 2017 ACM Int Jt Conf Pervasive Ubiquitous Comput Proc 2017 ACM Int Symp Wearable Comput 2017:763–71. https://doi.org/10.1145/3123024.3125609.
[59] Di Matteo D. Inference of anxiety and depression from smartphone-collected data. Diss Abstr Int Sect B Sci Eng 2022;83.
[60] Giannouli E, Fillekes MP, Mellone S, Weibel R, Bock O, Zijlstra W. Predictors of real-life mobility in community-dwelling older adults: An exploration based

on a comprehensive framework for analyzing mobility. Eur Rev Aging Phys Act 2019;16:1–13. https://doi.org/10.1186/s11556-019-0225-2.

[61] Steele R, Hillsgrove T, Khoshavi N, Jaimes LG. A survey of cyber-physical system implementations of real-time personalized interventions. J Ambient Intell Humaniz Comput 2022;13:2325–42. https://doi.org/10.1007/s12652-021-03263-0.

[62] Schuster R, Kaiser T, Terhorst Y, Messner EM, Strohmeier L-M, Laireiter A-R. Sample size, sample size planning, and the impact of study context: systematic review and recommendations by the example of psychological depression treatment. Psychol Med 2021;51:902–8. https://doi.org/10.1017/S003329172100129X.

[63] Barnett I, Torous J, Reeder HT, Baker J, Onnela J-P. Determining sample size and length of follow-up for smartphone-based digital phenotyping studies. J Am Med Inform Assoc 2020;27:1844–9. https://doi.org/10.1093/jamia/ocaa201.

[64] Montag C, Baumeister H, editors. Digital Phenotyping and Mobile Sensing. 2nd ed. Cham: Springer International Publishing; 2023. https://doi.org/10.1007/978-3-030-98546-2.

[65] Terhorst Y, Weilbacher N, Suda C, Simon L, Messner E-M, Sander LB, et al. Acceptance of smart sensing: a barrier to implementation—results from a randomized controlled trial. Front Digit Heal 2023;5. https://doi.org/10.3389/fdgth.2023.1075266.

[66] Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The Role of Data Type and Recipient in Individuals' Perspectives on Sharing Passively Collected Smartphone Data for Mental Health: Cross-Sectional Questionnaire Study. JMIR MHealth UHealth 2019;7:e12578. https://doi.org/10.2196/12578.

[67] Terhorst Y, Sander LB, Ebert DD, Baumeister H. Optimizing the predictive power of depression screenings using machine learning. Digit Heal 2023;9. https://doi.org/10.1177/20552076231194939.

[68] Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis With R: A Hands-On Guide. 1st ed. Boca Raton, FL and London: Chapman & Hall/CRC Press; 2021.

[69] Roberts H, Helbich M. Multiple environmental exposures along daily mobility paths and depressive symptoms: A smartphone-based tracking study. Environ Int 2021;156:106635. https://doi.org/10.1016/j.envint.2021.106635.

[70] Mennis J, Mason M, Ambrus A. Urban greenspace is associated with reduced psychological stress among adolescents: A Geographic Ecological Momentary Assessment (GEMA) analysis of activity space. Landsc Urban Plan 2018;174:1–9. https://doi.org/10.1016/j.landurbplan.2018.02.008.

[71] Liu Z, Chen X, Cui H, Ma Y, Gao N, Li X, et al. Green space exposure on depression and anxiety outcomes: A meta-analysis. Environ Res 2023;231:116303. https://doi.org/10.1016/j.envres.2023.116303.

[72] Rottstädt F, Becker E, Wilz G, Croy I, Baumeister H, Terhorst Y. Enhancing the Acceptance of Smart Sensing in Psychotherapy Patients: Findings from a Randomized Controlled Trial. Front Digit Heal 2024;6.

https://doi.org/10.3389/fdgth.2024.1335776.

Supplementary Files

Multimedia Appendixes

Supplement 1.
URL: http://asset.jmir.pub/assets/754856cdcecd4d5742246f5857ced3b1.docx