

SmartSense-D: A safety, feasibility, and acceptability pilot study of digital phenotyping in young people with major depressive disorder

Andres Camargo^{1,2} , Scott D Tagliaferri^{1,2} , Simon D'Alfonso³ , Tianyi Zhang³ , Zamantha Munoz^{1,2} , Pemma Davies^{1,2}, Mario Alvarez-Jimenez^{1,2}, Niels van Berkel⁴ , Vassilis Kostakos³  and Lianne Schmaal^{1,2}

Abstract

Background: Digital assessment of behaviours, including physical activity, sleep, and social interactions could be associated with changes in mood and other mental health symptoms. This study assessed the safety, feasibility, acceptability, and potential predictive value of passive and active sensing in young people with major depressive disorder (MDD).

Methods: Over eight weeks, passive (smartphone sensing, actigraphy) and active (ecological momentary assessment; EMA) data were collected from 40 young participants with MDD (aged 16–25 years). We assessed the safety, feasibility, and acceptability of daily active and passive sensing in this population. Additionally, linear mixed models and correlation analysis explored associations between passive and active sensing measures.

Results: Of the 48 young participants, 83% (n = 40) completed the full protocol. No adverse events were reported. Over eight weeks, participants averaged 35.9 days (65.3%) with EMAs and 37.9 days (69%) with actigraphy data. Smartphone sensors recorded communication for 21.1 days (38.4%), location for 43.1 days (78.4%), maximum unlock duration for 43.4 days (79%), social media use for 34.8 days (63.3%), and inter-key delay for 32.8 days (59.6%). Regarding acceptability, 83.1% found the application usable and comfortable. Secondary measures showed significant correlations between sleep and physical activity, and between location and phone use sensors. There was a significant negative association between daily positive mood ratings and QIDS total scores (Beta coefficient [95% CI]: 2.66 [−3.98, −1.34]; p = 0.002).

Conclusion: Passive and active sensing methods were safe, and acceptable among young people with MDD.

Keywords

Digital phenotyping, passive sensing, actigraphy, ecological momentary assessment, depressive symptomatology, wearables

Received: 10 July 2024; accepted: 11 March 2025

Introduction

Depression is the leading cause of disability worldwide, due to early age of onset, high population prevalence, chronicity and recurrence, and comorbidity with physical illness.¹ Depression onset peaks during adolescence and early adulthood, with 12-month prevalence rates of 4–5%.^{2,3} The depressive symptoms experienced by young people (16–25 years) cause greater impairment than any other mental or physical disorder.⁴ Investigating how depression progresses over time in young people is crucial to better

¹Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

²Orygen, Parkville, Australia

³School of Computing and Information Systems, The University of Melbourne, Parkville, Australia

⁴Department of Computer Science, Aalborg University, Aalborg, Denmark

Corresponding author:

Lianne Schmaal, Orygen, 35 Poplar Rd, Parkville, VIC, 3052 Australia.
Email: lianne.schmaal@unimelb.edu.au



understand the underlying causes and to identify potential risk factors and early warning signs of relapse.⁵

The use of digital tools in mental health has the potential to improve symptom tracking, risk factor identification, and treatment support.⁶ Digital phenotyping is the real-time, data-driven quantification of individual phenotypes, gathered directly from personal digital devices (e.g., smartphones).^{7,8} More recent definitions of digital phenotyping consider categories of active and passive sensing.⁹ Active sensing refers to data collected through individual interactions, for example, self-report questions administered via ecological momentary assessment (EMA).⁹ Passive sensing data is generated without active user input via sensors on a personal device, such as data on application usage, communication, and physical location.¹⁰ These methods of active and passive data collection could be used in conjunction to monitor changes in mood and identify relevant risk factors.

EMA is a form of active sensing that can be used to track mood states daily via notifications to electronic devices,¹¹ and has been successfully implemented to assess psychological states, behaviours, and contextual factors.⁸ This use of EMA can maximise ecological validity, reduce recall biases, and capture more dynamic information on depressive symptomatology.¹¹ To date, depressive symptoms have been successfully collected daily using active EMA methods in adults with mild-to-moderate depression.¹² The limited number of studies exploring this approach in young people with major depressive disorder (MDD) highlights the need for further research.¹³ Evaluating the use of EMA in young individuals with MDD is critical, given this is the peak time for the burden of the condition.

Digital devices can collect real-time data using embedded sensors.¹⁴ Studies suggest passive assessment of communication, location, and phone lock/unlock events could be useful in predicting depressive symptoms among college students.^{15,16} Furthermore, actigraphy could identify the relationship between sleep disturbances and physical activity with depressive symptoms reported by young people.^{17,18} However, the feasibility of implementing a combination of such approaches in clinical populations remains unclear.

Previous research suggests that both active and passive sensing are feasible and acceptable, however, most findings are derived from samples of the general population.¹⁹ Growing evidence indicates that digital phenotyping could potentially facilitate the detection of mental disorders.²⁰ However, there is limited research on these data collection methodologies in young people with MDD.²¹ From a methodological standpoint, it is crucial to further investigate the safety, feasibility, and acceptability of digital phenotyping in a clinical sample of young people with major depressive disorder (MDD). This will ensure that the data collection process addresses individual concerns and mitigates technical challenges effectively. Moreover, evaluating the safety, feasibility and acceptability of a combined protocol

involving daily EMA surveys, actigraphy and passive sensing in a clinical sample of help-seeking young individuals with moderate to severe MDD is important, as many previous studies examining associations between digital phenotypes and depressive symptoms have been conducted in general or student populations^{22,23} and have mostly not evaluated collection of EMA, passive sensing and actigraphy data simultaneously. Such findings could provide evidence-based recommendations for improving implementation of digital phenotyping in research and clinical practice. Therefore, the primary aim of this study was to determine the safety, feasibility, and acceptability of comprehensive digital phenotyping among young people with MDD. Young people with MDD filled out twice daily EMA surveys to track their mood in real-time. The AWARE-Light smartphone sensing app was used to passively collect real-time data from mobile phone sensors, providing insights into communication, location (e.g., time spent at home), phone usage.²⁴ Sleep and physical activity was tracked by GENEActiv devices (Activinsights, Kimbolton, England), which is a wrist-worn device that has been widely utilised in research to objectively measure and monitor sleep patterns and physical activity in young people.²⁵ The secondary aim was to explore the prospective association between digital phenotyping measures, including active and passive sensing, and changes in depressive symptoms over time.

Methods

Design

This pilot longitudinal study included a baseline clinical assessment, eight weeks of EMA, actigraphy and passive sensing, and a follow-up clinical assessment at the end of the data collection period. This research project was approved by the Human Research Ethics Committee of The University of Melbourne (Ethics ID Number: 1955691.4). This study was conducted in line with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cohort studies.²⁶

Sample

A total of 40 young people were recruited from four *head-space* youth mental health clinics in Melbourne, Australia. Inclusion criteria included: (1) aged 16 to 25 years inclusive; (2) diagnosis of MDD based on DSM-V and MINI interview;²⁷ (3) familiarity with and willingness to use available crisis support in case of emergency and; (4) possession and regular use of an Android smartphone as the used application only collects data from Android devices. Exclusion criteria were: (1) presence of an intellectual disability, as determined by the treating clinician; (2) inability to sufficiently converse in or read English; (3) imminent

risk of suicide, determined in consultation with the treating clinician; (4) actively manic or actively psychotic, based on MINI interview; (5) insufficient familiarity with smartphones to complete the daily EMA surveys and; (6) not owning an Android phone. Written informed consent was obtained from all participants prior to enrolment. For participants aged 16 to 18, additional consent was required from a legally nominated guardian. Participants with cognitive impairment, intellectual disability, or mental illness who were unable to provide informed consent were excluded. The consent process and the Participant Information and Consent Form (PICF) were approved by the Human Research Ethics Committee (HREC) before implementation.

Study procedure

Screening/baseline assessment. Participants with MDD were enrolled as per inclusion criteria. This included meeting the criteria for a current depressive episode of MDD, as determined by the MINI interview during screening.²⁷ While direct psychometric validation with an Australian sample is limited, the MINI's reliability and validity across international studies suggest it is a suitable tool for evaluating depressive symptoms among younger populations.^{27,28} Additionally, the MINI has been employed in research examining diagnostic tools for mental health, including studies comparing the classification of major depression across different diagnostic interviews.²⁹ Eligible participants were asked to complete a survey, including demographics, education background, history of medical conditions, previous pharmaceutical treatments, and psychosocial support. Participants further completed the Quick Inventory of Depressive Symptomatology (QIDS16) at baseline and follow-up. The QIDS is a 16-item self-report questionnaire with a total score ranging 0–27.³⁰ The QIDS measures low mood, concentration, self-criticism, suicidal ideation, interest, energy/fatigue, irregular sleep patterns, changes in appetite and weight, and psychomotor agitation/retardation.³¹ The baseline average scores for the QIDS-16 revealed that participants had an average score of 15.65 (scale range 0–27), placing them within the moderate-to-severe range of depressive symptomatology. Furthermore, the QIDS has been used in various clinical trials registered in Australia, focusing on interventions for depression and related conditions. These trials often include young adult participants, further supporting the instrument's use in this demographic within the Australian context.^{32,33} The participants were then given a wrist-based actigraphy device, and the setup of a smartphone application for collection of active and passive sensing data. Fourteen participants experienced compatibility issues with the AWARE-Light app on specific Android smartphone models. To facilitate their participation and ensure appropriate collection of EMA surveys and sensor

data, participants used loan smartphones, either sourced from personal spare devices or provided by the study team.

Active sensing measures

Ecological momentary assessment (EMA). The AWARE-Light app¹ delivered EMA prompts twice daily, at midday and 8pm, and collected passive sensing data over an eight-week period (Figure 1). Although passive data collection spanned the full eight weeks, EMA surveys were required for at least six weeks, as participants had the option to take a break from EMA surveys during week 4 and/or week 8. Six participants took these breaks, while the other 30 participants kept providing EMA surveys for the entire assessment period. The EMA questionnaire was developed by integrating items from established psychometric tools, including The Pittsburgh Sleep Quality Index (PSQI), Depression, Anxiety, and Stress Scale (DASS), Penn State Worry Questionnaire (PSWQ), Ruminative Response Scale (RRS), and UCLA Loneliness Scale, to capture participants' emotional states, sleep quality, stress, anxiety, worry, rumination, and social interactions.^{34–38} The EMA schedule was implemented to capture distinct temporal dynamics, with morning questionnaires assessing post-sleep states and evening questionnaires summarising daily experiences. The morning survey included 12 items on positive and negative affect, sleep quality, and substance use since last survey. An additional question was included to explore the impact of any significant events on mood, behaviour, or phone usage. The evening survey included the same items as the morning survey except for the sleep item, with eight additional items on perceived stress, anxiety, worry, rumination, loneliness, interaction with other people, how pleasant the most positive event of the day was, and how pleasant the most negative event of the day was. For this study, mood average scores and mood variability, were derived and included in the statistical analyses. Positive and negative mood scores from all morning and evening surveys over the eight-week period were averaged to calculate one overall average score for each variable. Positive and negative mood variability, referring to the magnitude of the mood shift, was calculated as the standard deviation (SD) of the average mood scores.³⁹

Passive sensing measures

Communication, location, and phone usage. The AWARE-Light smartphone sensing Android app²⁴ was also used to passively collect data from participants' mobile phone sensors.²⁴ Passive sensing data was gathered from multiple smartphone sensors, with detailed features listed in Supplementary Table 1. Data from the location sensor (GPS and Networks surrounded), configured to record every 180 s, included latitude, longitude, speed, altitude,

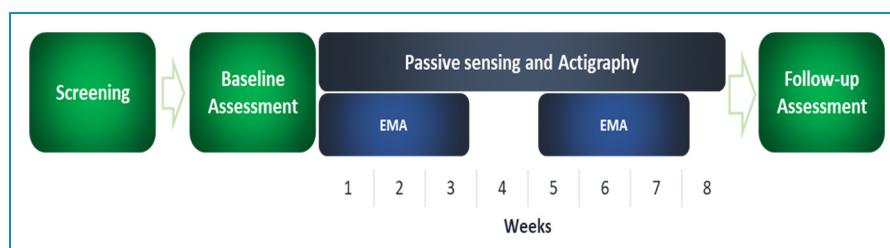


Figure 1. Assessment timeline.

and estimated accuracy, which were used to calculate features such as normalised location entropy, number of location transitions, and location variance. These features were calculated solely based on the recorded data points, meaning that no assumptions or inferred values are introduced, ensuring the derived features reflect only the available data and remain free from imputation bias. The communication sensor logged the time and duration of incoming, outgoing, and missed calls, as well as sent and received messages, capturing metrics like the average call duration and the number of distinct contacts. The screen usage sensor recorded screen states (e.g., locked, unlocked, on, or off) and the corresponding times, while the keyboard sensor logged keystroke timing and application use, with participants given the option to anonymise text content. The application usage sensor logged app usage events (recording whenever an app is used). Features were extracted from these raw data using the Reproducible Analysis Pipeline for Data Streams (RAPIDS), enabling the calculation of behavioural measures across key domains, including location (e.g., normalised location entropy), communication (e.g., distinct contacts), and phone usage (e.g., social media duration). In terms of sensor-level validity, we assessed the quality of each passive sensor independently, marking data as valid only if the sensor provided complete information for that day. Subsequently, for day-level validity, we defined data as valid if at least one sensor, specifically the location sensor, collected accurate information.

Sleep and physical activity. Participants were provided with a GENEActiv sleep and physical activity tracker (Activinsights, Kimbolton, England) to monitor changes in sleep patterns and physical activity. The actigraphy device was delivered to the participants home and replaced after three weeks to prevent the battery from running out. Participants were instructed to wear the actigraphy device on the wrist of their choosing throughout the day and to remove it only when it needed to be swapped for a new one. The GENEActiv devices were set up to sample at a measurement frequency of 10.0 Hz and data sleep and physical activity features were extracted from the raw actigraphy device. The accelerometer data processing R-package GGIRR, provided an estimation of the average time of sleep (sleep efficiency, duration, regularity) and

physical activity features (coded into light, moderate or vigorous levels) recorded over the study period.⁴⁰

Follow-up assessment. After the eight-week data collection period, participants were requested to complete the QIDS again. Additionally, participants were provided with an online debriefing questionnaire. The debrief questionnaire included two sections. The first section comprised 18 questions that evaluated usability and comfort on a scale from one (strongly disagree) to seven (strongly agree). These items assessed the user experience in navigating the AWARE-light application, changes in mobile phone usage, comfort levels with the application, battery drainage, frequency of EMAs, and willingness to participate in future studies. Additionally, two questions (eight and 11) with reversed scales were included to identify potential privacy concerns and the extent of discomfort associated with using the AWARE light app. The second section included eight questions assessing potential concerns related to collecting passive sensing data from specific smartphone sensors (applications, communication, location, light, keyboard, network, and screen usage). Similarly, participants were asked about their level of comfort using the actigraphy device over the study period. Participants rate these questions on a scale from one (extremely uncomfortable) to seven (extremely comfortable). The specific questions are reported in (Supplementary Table 2).

Outcome measures

Primary outcomes. For safety, the evaluation was based on reports from participants and treating clinicians, who were asked to report any significant adverse events associated with the questionnaires, interviews, ecological momentary assessments (EMA), and passive sensing methods. Additionally, the medical officer of the study assessed any potential adverse effects associated with the study.

For feasibility, we assessed the following: (1) attrition rates; measured by the number of participants who withdrew, and (2) the number and percentage of data collected from both active and passive sensing relative to the number of potential assessments.

Feasibility was defined as (1) completion of full assessment protocol in >70% of participants, (2) 80% of

participants completed 65–70% EMA surveys in the first 3-week EMA period, and (3) 80% of participants with missing data <30% of days were set for actigraphy for 6 weeks (minimum data collection period considering the optional breaks) and all passive sensors over the 8-week assessment period, with missing data being defined as no data available on a certain day. This threshold was determined based on prior research into the feasibility of passive smartphone data collection,⁴¹ complemented by consensus among experts and developers of the AWARE-Light app. This criterion was considered to balance practical implementation and the reliability of data required for robust analysis. To ensure data reliability, we first analysed location data, which is typically recorded at short, regular intervals, to identify inconsistencies across all sensors. The presence of irregularities in location data, indicative of potential sensor malfunctions, resulted in the exclusion of 13 out of 38 participants who exhibited inconsistent sensor patterns.

Acceptability was determined based on the scores and feedback provided through the debriefing questionnaire administered at the follow-up assessment, (Supplementary Table 6).

Secondary outcomes. Although feasibility remains the primary focus, we incorporated exploratory analyses to investigate preliminary associations between sensor-derived measures and psychological constructs. These findings offer early insights into the potential applicability of this approach, laying a foundation for evidence-based, in-depth analyses in future research. By addressing feasibility and adherence, this study contributes to the design of robust longitudinal research that can prioritise examining relationships between passive sensing features and mood changes over time. Using aggregated scores (average scores across the 8-week study period), associations between the different smartphone sensors were explored. Additionally, correlations between passive and active sensing data and changes in depressive symptoms (QIDS scores) were analysed over an 8-week follow-up period.

Data analysis. For the primary outcomes, feasibility ($n = 48$), safety ($n = 48$), and acceptability ($n = 36$) were summarised via descriptive measures (Figure 2). For secondary analyses, data from 40 participants were included in analyses including sleep and physical activity (actigraphy), and mood (EMA) data (Figure 2). Data from 27 participants were included in models incorporating smartphone sensor variables, because of more missing data due to technical issues (Figure 2).

First, Pearson correlation coefficients were examined to assess associations between sleep, physical activity, mood, and smartphone sensors (*rcorr function in Hmisc R package*).⁴² Second, linear mixed models⁴³ (*lmer function in lme4 R package*)⁴⁴ with a restricted maximum likelihood estimation approach were used to examine the association

between each active and passive sensing measure that were averaged across the follow-up period and the total QIDS scores (baseline, follow-up). This statistical approach is well-suited to our study as it accounts for the hierarchical structure of the data, including repeated measures within individuals, while accommodating both fixed and random effects, and can appropriately deal with missing data.⁴⁵ In each linear mixed model, participants were included as random factors, and age, sex assigned at birth, predictor (active or passive sensing measure), time (baseline, follow-up), and a predictor*time interaction were included as fixed factors. To avoid introducing bias or distorting temporal patterns, when extended periods of missing data occurred at specific locations (e.g., absence of recorded activity or sensor input), these gaps were retained without any attempts to estimate or impute the missing information.

For all analyses, significance levels were set at an alpha of 0.05, corrected for multiple comparisons using the Benjamini-Hochberg False Discovery Rate (FDR).⁴⁶ Given the exploratory nature of our secondary analyses examining associations between each digital phenotyping measure and the QIDS score and the likely non-independency between the different linear mixed models, we chose FDR as correction method.⁴⁷

Results

Participant characteristics

The demographic data of participants are reported in Table 1. Among the 40 participants, for sex assigned at birth, 28 (70%) were female and 12 (30%) were male. For gender, 22 (55%) identified as female, 4 (10%) as gender queer, 10 (25%) as male, and 4 (10%) as transgender. Fourteen out of 40 participants (35%) had completed secondary school, ten participants (25%) completed vocational education and training, two had attained undergraduate degrees (5%) and one (2.5%) a postgraduate degree, thirteen participants (32.5%) were enrolled in education when they joined the study. Similarly, 5 (12.5%) participants were engaged in full-time paid employment, 4 (10%) in part-time paid employment, 9 (22.5%) in casual paid employment, 19 (47.5%) were unemployed, and 3 (7.5%) were involved in other unpaid roles, such as volunteering.

Primary findings

Safety. There were no reports of any adverse events.

Feasibility

Full protocol completion (criterion 1). Of the 48 young people who consented to the study, 83% ($n = 40$) completed the full assessment protocol, which exceeded our predefined

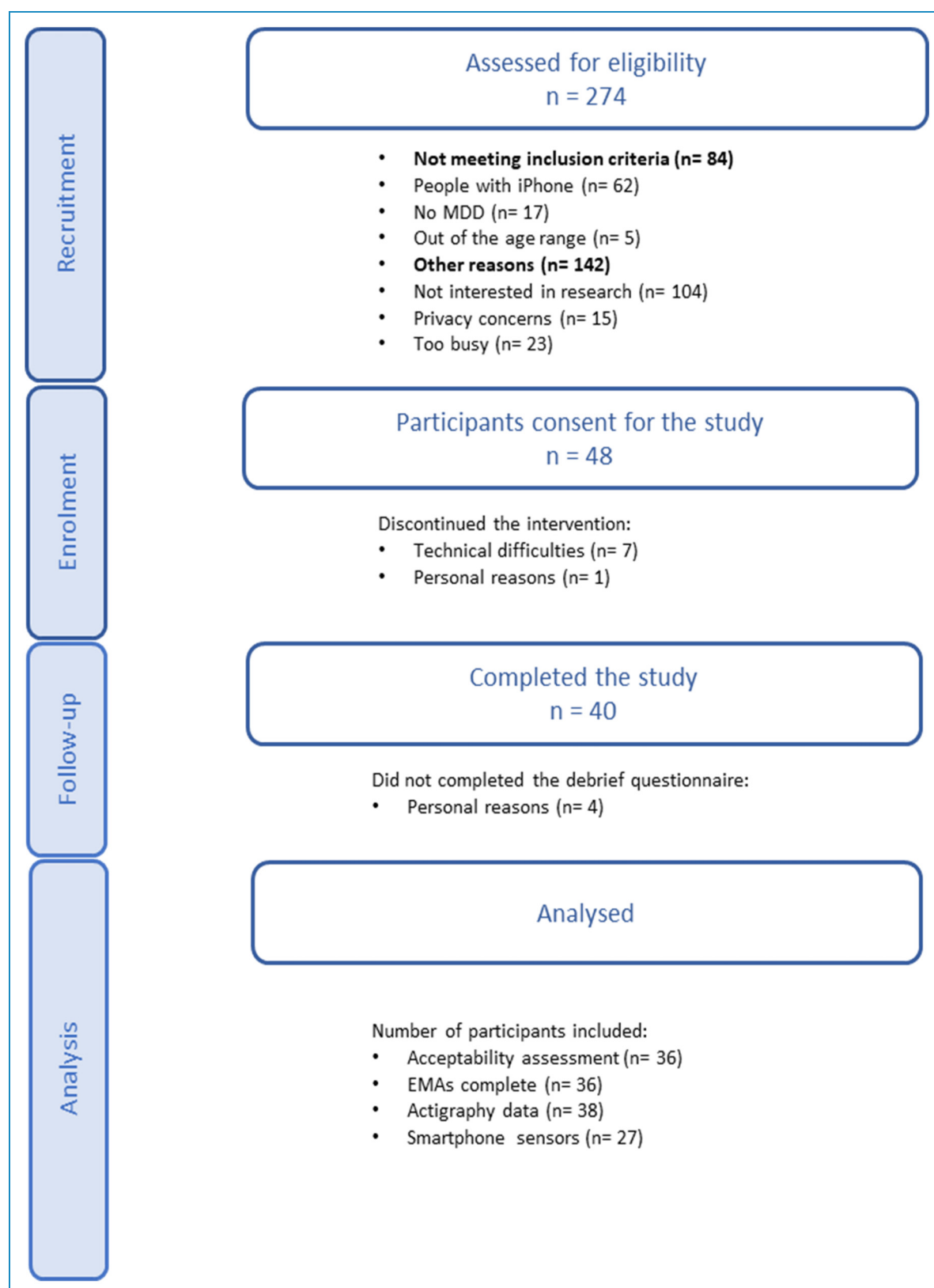


Figure 2. Flow diagram of participant selection at the eight-week study period.

Table 1. Participant characteristics.

Characteristics n = 40	Number (%)
Sex assigned at birth	
Female	28 (70)
Male	12 (30)
Gender	
Female	22 (55)
Gender queer	4 (10)
Male	10 (25)
Transgender	4 (10)
Age group (years), n(%)	
16–18	10 (25)
19–22	17 (42)
23–25	13 (32)
Current education, n(%)	
Full-time student	9 (22.5)
Part-time student	4 (10)
Not in education	27 (67.5)
Education level achieved, n(%)	
Secondary school	14 (35)
Vocational education and training	10 (25)
University undergraduate	2 (5)
University postgraduate	1 (2.5)
Current education	13 (32.5)
Employment, n(%)	
Full-time paid employment	5 (12.5)
Part-time paid employment	4 (10)
Casual paid employment	9 (22.5)
Unemployed	19 (47.5)
Other (volunteer unpaid)	3 (7.5)

(continued)

Table 1. Continued.

Characteristics n = 40	Number (%)
People who live with, n(%)	
Both parents	18 (45)
Mother only	8 (20)
Partner	6 (15)
Others	8 (20)

feasibility criterion (1). Seven participants withdrew due to technical issues with the AWARE-Light application, and one participant dropped out for personal reasons.

EMA completion (criterion 2). For EMA, during the first three weeks, 26 out of 36 (72%) of participants completed at least 65% of surveys, which is below our predefined feasibility criterion of 80% for the first 3 weeks (2). Over the minimum of 6-weeks of EMA surveys, 22 out of 36 (61%) of participants completed at least 65% of surveys. Individual EMA completion rates are detailed in Supplementary Table 3.

Missing days of actigraphy and passive sensing data (criterion 3). A total of 38 participants (95%) provided actigraphy data, with one actigraphy watch was defective and another failed to store data. Of these, 20 out of 38 (56%) of participants had valid actigraphy data, defined as data collected for more than 39 days (70% of the total 56 days). Non-compliance, such as failing to consistently wear the devices, contributed to missing data. Additionally, limited availability of wristband devices caused delays in distribution, leading to varied monitoring start and end times. Individual actigraphy completion rates are detailed in Supplementary Table 4.

Among all participants who collected data during the study period, only those who did not encounter significant technical difficulties (27 out of 38 participants) were included in the passive sensing analysis. Of these 27 participants, 22 (81%) provided passive sensing data for at least 39 days, meeting the threshold of 70% of the 56-day (8-week) collection period. Data completeness varied across passive sensing features. Location and max unlock duration had the highest levels of completeness, with 70.37% of participants (19 out of 27) meeting the criterion. Social media use followed with 37.04% (10 out of 27 participants). In contrast, communication and inter-key delay exhibited the lowest completeness rates, with only 25.93% (7 participants) providing more than 39 days of data. The average number

of days with available data reflected these trends: communication data was recorded for a mean of 21.1 days ($SD = 17.2$), location for 43.1 days ($SD = 11.9$), max unlock duration for 43.4 days ($SD = 11.9$), social media use for 34.8 days ($SD = 12.8$), and inter-key delay for 32.8 days ($SD = 12.4$). Detailed information is provided in Supplementary Table 5.

Acceptability. The user feedback data for the AWARE app revealed a positive overall experience. Overall, 83.1% of participants indicated agreement with usability and comfort questions (strongly agree: 39.5%; agree: 32.3%; slightly agree: 11.3%) (Supplementary Table 6; Supplementary Figure 1). For the reverse scaled questions, 61.1% of participants disagreed with privacy concerns, and 83.3% disagreed with the application making them feel upset (Supplementary Table 7; Supplementary Figure 2). Regarding passive sensing, 79.8% of the participants indicated being comfortable with passive sensing (extremely comfortable: 36.8%; very comfortable: 29.5%; slightly comfortable: 13.5%) (Supplementary Table 8; Supplementary Figure 3).

Secondary outcomes

Correlation analysis. Results from the correlation analyses are reported in Supplementary Tables 9 and 10. Correlation matrices are reported in Supplementary Figures 4 and 5.

For the first correlation analysis ($n = 38$; excluding smartphone sensing measures), results showed there was a significant negative correlation between sleep efficiency and light physical activity ($r = -0.53$, $p = 0.006$), sleep regularity and light physical activity ($r = -0.54$, $p = 0.006$), and sleep regularity and moderate physical activity ($r = -0.57$, $p = 0.003$). There was also a significant positive correlation between light and moderate physical activity ($r = 0.82$, $p < 0.001$). There were significant negative correlations between sleep efficiency and moderate physical activity and positive mood and light physical activity, however, these became non-significant following FDR adjustment.

For the second correlation analyses with smartphone sensors ($n = 27$), additional positive correlations were found between outgoing and incoming calls ($r = 0.90$, $p < 0.001$), location entropy and location transitions ($r = 0.65$, $p = 0.009$), and max unlock duration and social media use ($r = 0.58$, $p = 0.039$).

Linear mixed models. Associations of sleep, physical activity, daily mood ratings, smartphone sensors with depressive symptoms scores were assessed separately in linear mixed models (Table 2). The findings revealed a significant main effect of positive mood ratings on depressive symptoms severity (QIDS) over the 8-week data collection period (β [95%CI]: -2.66 [-3.98 , -1.34]; $p = 0.003$). This suggests that lower positive mood ratings were consistently linked

with higher levels of depressive symptoms across the duration of the study. There were also significant associations between light and moderate physical activity and depressive symptoms, however, these became non-significant after FDR adjustment. No other significant associations were observed.

Discussion

Our study evaluated the safety, feasibility, and acceptability of digital phenotyping methods in measuring depressive symptoms among young people with MDD. Our results showed that active and passive sensing methods were perceived as safe and acceptable by young people with MDD. EMA assessments were feasible, however technical issues caused significant data missingness for actigraphy and passive sensing. Correlations between sleep regularity, sleep efficiency and physical activity (light, moderate and vigorous) measures were found. Finally, we observed a negative association between daily positive mood ratings and depressive symptoms across time.

No reports of serious adverse events throughout the eight weeks were reported in our study. In terms of feasibility, the attrition rates for the study were 20%, with seven participants withdrawing due to constant difficulties impeding data collection through the AWARE-Light application. This exceeded our pre-defined feasibility criteria of 70% of participant completion.

The completion of EMAs within the first three weeks of data collection (78%) was around our predefined feasibility criteria of 80%. Additionally, 56% of the participants collected actigraphy data over 39 days of data collection. Over eight weeks, the data availability for smartphone sensing varied across features, with communication data recorded for an average of 21 days, while location data was available for 43 days. Maximum unlock duration was recorded for 43 days, social media use for 35 days, and inter-key delay for 33 days. These findings indicate the need for increased attention in addressing the persistent challenge of collecting passive sensing data. These results are aligned with previous studies as lower feasibility of passive sensing data, particularly missing data, has been seen in studies with lengthy duration.³³ Potential strategies to enhance the feasibility and efficiency of collecting passive sensing data can be implemented. Automated data quality alerts could flag issues such as missing data or sensor malfunctions in real time, enabling the capability to intervene without the need for constant monitoring. Streamlined, user-friendly dashboard systems could further support data management by providing an overview of key metrics like missing data or sensor uptime, reducing the cognitive and logistical demands on researchers / clinicians.

In terms of acceptability, a debriefing questionnaire filled out by 36 participants revealed positive perceptions among participants, specifically regarding the usability, comfort, and privacy concerns of passive and active sensing data.

Overall, 83.3% percent of participants agreed with the usability of passive and active data collection, suggesting that the instruments used (i.e., app and wearables) were easy to use and to interact with. To our knowledge, this is the first feasibility study assessing the AWARE-Light application as a data collection method in a clinical population, however, related work indicates that this smartphone application was originally designed to provide a user-friendly configuration for smartphone research studies using passive and active sensing, including EMA.²⁴ Although our study shows that using passive and active sensing data is acceptable for young people, it may also be useful to examine individual responses in future studies to identify whether the usability perception

is linked to specific application features.²⁴ Therefore, these findings need to be confirmed in larger samples.

Although 86.1% of participants expressed comfort with the EMA surveys, 11.2% disagreed with the app's level of comfort. Limitations most frequently reported were battery life while using the app, and the frequency of EMAs (i.e., twice a day). Previous research has highlighted that there is a trade-off between amount of data collected and battery life.⁹ Similarly, participants indicated that completing EMA can be burdensome, which may impact participants' engagement and comfort.⁴⁸ From that perspective, passive sensing may be a more acceptable approach to collecting daily information from participants, but the value of these

Table 2. Linear mixed models for the relationships between sleep, physical activity, mood and QIDS total score across time.^a

Factors	Variable	N ^b	Beta Coefficient	L95%CI	U95%CI	p-value	FDR p-value ^c
Sleep	Sleep duration (hours)	40	-0.18	-0.57	0.22	0.365	0.598
	Sleep efficiency (%)	40	-7.81	-28.22	12.60	0.445	0.667
	Sleep regularity (%)	40	-0.09	-0.20	0.01	0.073	0.328
Physical activity	Light activity (min)	40	0.83	0.09	1.58	0.029	0.244
	Moderate activity (min)	40	1.63	0.07	3.18	0.041	0.244
	Vigorous activity (min)	40	3.00	-15.95	21.95	0.751	0.811
Mood	Positive mood (mood scale)	40	-2.66	-3.98	-1.34	<0.001	0.003
	Negative mood (mood scale)	40	1.24	-0.44	2.92	0.145	0.372
	Positive mood variability (mood scale)	40	3.79	-11.32	18.90	0.616	0.759
	Negative mood variability (mood scale)	40	1.72	-5.48	8.92	0.633	0.759
Communication	Outgoing calls (min)	27	0.01	-0.01	0.02	0.359	0.598
	Incoming calls (min)	27	0.01	-0.01	0.03	0.545	0.754
Location	Location entropy (nats) ^d	27	1.16	-8.13	10.44	0.804	0.811
	Number location transitions (nats)	27	0.20	-0.13	0.53	0.237	0.474
	Location variance (transitions)	27	0.24	-1.80	2.28	0.811	0.811
Phone use	Max duration unlock (min)	27	0.06	-0.01	0.12	0.092	0.333
	Social media use (min)	27	0.09	-0.04	0.22	0.166	0.374
	Inter key delay (ms)	27	0.03	-0.01	0.07	0.119	0.357

^aSeparate linear mixed-effects models encompassing each predictor of depressive symptoms from (derived from QIDS) T1 baseline and T2 follow-up. All analyses included participant as a random factor and age, sex assigned at birth, predictor, time and a predictor*time interaction as fixed factors.

^bOnly 27 individuals were included in the analysis on these sociability factors (communication, location, and phone use) due to issues with smartphone sensors.

^cP-values were adjusted for the FDR.³³

^dRefers to the natural units of information used to measure uncertainty in the context of geographic locations.

passive sensors in tracking or predicting depressive outcomes needs to be further explored.

Only seven out of 36 participants expressed privacy concerns about passive sensing data collection. In keeping with the findings of other studies, data privacy concerns do not appear to concern young people, this may be because younger individuals tend to be more at ease with providing data via a smartphone.^{49,50} Participants also did not express concerns about the management of their personal information, possibly due to the involvement of healthcare providers from *headspace* and Orygen, considered as reliable mental health organisation in Australia.⁵¹

Secondary findings

Negative correlations between sleep regularity and physical activity (light and moderate) were found. These results may require more exploration as changes in physical activity throughout adolescence are observed.⁵² However, in adolescence, decreased physical activity and increased sedentary behaviour may significantly impact sleep habits.⁵³ In non-clinical youth populations, a study showed consistent results indicating that higher physical activity levels were associated with reduced mean value, night-to-night variation, and shifts in awakenings, suggesting a potential promotion of more sleep regularity.⁵⁴

Interestingly, we found no associations between passive sensing variables and depressive symptoms or daily mood ratings, diverging from prior research that linked prolonged screen unlock times and typing delays with elevated anxiety and depression scores in healthy college students. This contrast shows the complexity of these relationships and illustrates the need for further investigation into the unique interplay between phone use and mood in specific populations.^{55,56} Although previous studies have reported associations between social media use and daily mood ratings,^{57,58} our findings did not reveal any correlation between these variables. This discrepancy might be attributed to our method of averaging data over the 8-week period, which may not have adequately accounted for the significant daily fluctuations and individual differences in these variables.

Similarly, previous studies examining the link between social media and depressive symptoms revealed that increased time spent on social platforms may influence mental health by providing more opportunities for users to engage in online activities like appearance comparisons.^{59,60} Our results showed that higher daily positive mood ratings (averaged across the 8-week data collection period) were significantly associated with lower depressive symptoms across the eight weeks. This finding is not surprising and is consistent with previous studies showing that young adults with depression display significantly lower positive moods compared to healthy people.^{61,62}

Strengths and limitations

Our study had several strengths. Our study reduced the potential for biases and enhanced the accuracy of the data collected using objective methods recording sleep patterns, physical activity levels, and social interactions, and included a clinical sample with moderate to severe depression compared to general or student populations included in many previous studies. Our study also had several limitations. Although we collected eight weekdays of passive and active sensing data, deriving general conclusions from only 40 participants limits the extent to which these findings can be applied to a broader population. Additionally, 23% of young people in the recruitment process used iPhones, which affected the feasibility of our study. In addition, we cannot generalise our findings to iPhone users. A potential limitation can be that The EMA survey was specifically designed for this study and had not been previously validated. Yet despite the survey items being adapted from validated questionnaires, including the PSQI, DASS, PSWQ, RRS, and the UCLA Loneliness Scale, this was the first time the survey was implemented.

Furthermore, analyses included aggregated data from passive sensing and EMA measures (i.e., averaged across the eight-week data collection period), by which we did not make optimal use of the rich information on dynamic changes in passive and active sensing measures and how they relate to depressive symptoms trajectories over time, which we plan to explore in future work.








Conclusions

Our research showed that the use of active and passive sensing techniques is safe and acceptable for young people with MDD. While the study demonstrated that active and passive sensing methods were feasible, challenges related to data completeness, mostly caused by technical issues with the passive sensing app and non-compliance for actigraphy, were observed among young participants, indicating areas for improvement in future implementations. Furthermore, significant correlations were identified among sleep, physical activity, and communication measures across 8 weeks. Further investigations should focus on the relationship between dynamic changes in passive and active sensing methods and trajectories of depressive symptoms over time.

Acknowledgments

We thank Headspace National Youth Mental Health Foundation for supporting the recruitment process. Also, thank you to those who volunteered to participate in the study. M.A-J. was supported by an Investigator Grant (APP1177235) from the National Health and Medical Research Council and a Dame Kate Campbell Fellowship from The University of Melbourne.

ORCID iDs

Andres Camargo  <https://orcid.org/0000-0002-3644-4864>
 Scott D Tagliaferri  <https://orcid.org/0000-0003-3669-4131>
 Simon D'Alfonso  <https://orcid.org/0000-0001-7407-8730>
 Tianyi Zhang  <https://orcid.org/0000-0002-0778-8844>
 Zamantha Munoz  <https://orcid.org/0009-0003-9873-348X>
 Niels van Berkel  <https://orcid.org/0000-0001-5106-7692>
 Vassilis Kostakos  <https://orcid.org/0000-0003-2804-6038>

Statements and declarations

Ethical Considerations

ID Number: 1955691.4

Author Contributions/CRedit

Conceptualization: AC, SDT, SDA, LS

Data curation: AC, ZM, PD, TZ

Formal Analysis: AC

Funding acquisition: SDA, LS

Investigation: AC, PD, ZM, SDA

Methodology: SDA, MA, NvB, VK, LS

Project administration: PD, AC

Resources: LS, SD

Software: SDA, VK

Supervision: SDT, SDA, LS

Validation: SDT

Visualization: AC, SDT

Writing – original draft: AC, SDT, SDA, LS

Writing – review & editing: All

Approved final manuscript: All

Funding

This study was funded by The University of Melbourne (MDHS establishment grant awarded to Prof Schmaal) and by a grant from the Cisco University Research Program Fund and the Silicon Valley Community Foundation (ID: 2021-232746 3696).

Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

References

1. Organization WH. *Global diffusion of EHealth: making universal health coverage achievable: report of the third global survey on EHealth*. Geneva, Switzerland: World Health Organization, 2017.
2. Kessler RC, Angermeyer M, Anthony JC, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's world mental health survey initiative. *World Psychiatry* 2007; 6: 168–176.
3. Petito A, Pop TL, Namazova-Baranova L, et al. The burden of depression in adolescents and the importance of early recognition. *J Pediatr* 2020; 218: 265–267.e1.
4. Digital health tools for the passive monitoring of depression: a systematic review of methods | npj Digital Medicine, <https://www.nature.com/articles/s41746-021-00548-8> (accessed 3 June 2023).
5. Kwong ASF, Manley D, Timpson NJ, et al. Identifying critical points of trajectories of depressive symptoms from childhood to young adulthood. *J Youth Adolesc* 2019; 48: 815–827.
6. Areán PA, Hoa Ly K and Andersson G. Mobile technology for mental health assessment. *Dialogues Clin Neurosci* 2016; 18: 163–169.
7. Onnela JP and Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016; 41: 1691–1696.
8. Batra S, Baker RA, Wang T, et al. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. *Med Devices Evid Res* 2017; 10: 237–251.
9. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology* 2021; 46: 45–54.
10. Moura I, Teles A, Viana D, et al. Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: a systematic literature review. *J Biomed Inform* 2023; 138: 104278.
11. Arney MF, Schatten HT, Haradhvala N, et al. Ecological momentary assessment (EMA) of depression-related phenomena. *Curr Opin Psychol* 2015; 4: 21–25.
12. Birney AJ, Gunn R, Russell JK, et al. Moodhacker Mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. *JMIR MHealth UHealth* 2016; 4: e4231.
13. Beames JR, Kikas K and Werner-Seidler A. Prevention and early intervention of depression in young people: an integrated narrative review of affective awareness and ecological momentary assessment. *BMC Psychol* 2021; 9: 13.
14. Maharjan SM, Poudyal A, van Heerden A, et al. Passive sensing on mobile devices to improve mental health services with adolescent and young mothers in low-resource settings: the role of families in feasibility and acceptability. *BMC Med Inform Decis Mak* 2021; 21: 17.
15. Wang R, Wang W, daSilva A, et al. Tracking depression dynamics in college students using Mobile phone and wearable sensing. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2018; 2: 43:1–43:26.
16. Crouse JJ, Carpenter JS, Song YJC, et al. Circadian rhythm sleep–wake disturbances and depression in young people: implications for prevention and early intervention. *Lancet Psychiatry* 2021; 8: 813–823.
17. Blake MJ, Trinder JA and Allen NB. Mechanisms underlying the association between insomnia, anxiety, and depression in adolescence: implications for behavioral sleep interventions. *Clin Psychol Rev* 2018; 63: 25–40.
18. Zarate D, Stavropoulos V, Ball M, et al. Exploring the digital footprint of depression: a PRISMA systematic literature review of the empirical evidence. *BMC Psychiatry* 2022; 22: 21.

19. Cornet VP and Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. *J Biomed Inform* 2018; 77: 120–132.
20. Tekin Ş. Is big data the new stethoscope? Perils of digital phenotyping to address mental illness. *Philos Technol* 2021; 34: 447–461.
21. Vaid SS and Harari GM. Smartphones in personal informatics: a framework for self-tracking research with Mobile sensing. In: Baumeister H and Montag C (eds) *Digital phenotyping and Mobile sensing: new developments in psychoinformatics*. Studies in Neuroscience, Psychology and Behavioral Economics. Cham, Switzerland: Springer International Publishing, 2019, pp.65–92. doi:10.1007/978-3-030-31620-4_5.
22. Beames JR, Han J, Shvetcov A, et al. Use of smartphone sensor data in detecting and predicting depression and anxiety in young people (12–25 years): a scoping review. *Heliyon* 2024; 10: e35472.
23. Marciano L, Vocaj E, Bekalu MA, et al. The use of Mobile assessments for monitoring mental health in youth: umbrella review. *J Med Internet Res* 2023; 25: e45540.
24. van Berkel N, D'Alfonso S, Kurnia Susanto R, et al. AWARE-Light: a smartphone tool for experience sampling and digital phenotyping. *Pers Ubiquitous Comput*. Published online 2022. doi:10.1007/s00779-022-01697-7
25. Antczak D, Lonsdale C, del P o, et al. Reliability of GENEActiv accelerometers to estimate sleep, physical activity, and sedentary time in children. *Int J Behav Nutr Phys Act* 2021; 18: 73.
26. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth* 2019; 13: S31–S34.
27. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; 59: 22–33. quiz 34–57.
28. Pettersson A, Modin S, Wahlström R, et al. The Mini-International Neuropsychiatric Interview is useful and well accepted as part of the clinical assessment for depression and anxiety in primary care: a mixed-methods study. *BMC Fam Pract* 2018; 19: 19.
29. Wu Y, Levis B, Sun Y, et al. Probability of major depression diagnostic classification based on the SCID, CIDI and MINI diagnostic interviews controlling for hospital anxiety and depression scale – depression subscale scores: an individual participant data meta-analysis of 73 primary studies. *J Psychosom Res* 2020; 129: 109892.
30. Yeung A, Feldman G, Pedrelli P, et al. The quick inventory of depressive symptomatology, clinician rated and self-report. *J Nerv Ment Dis* 2012; 200: 712–715.
31. Rush AJ, Trivedi MH, Ibrahim HM, 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: 573–583.
32. Zhang WY, Zhao YJ, Zhang Y, et al. Psychometric properties of the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR) in depressed adolescents. *Front Psychiatry* 2020; 11. doi:10.3389/fpsy.2020.598609
33. ANZCTR - Registration. <https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=ACTRN12621001358831> (accessed 15 November 2024).
34. de la Vega R, Tomé-Pires C, Solé E, et al. The Pittsburgh sleep quality Index: validity and factor structure in young people. *Psychol Assess* 2015; 27: e22–e27.
35. Szabo M and Lovibond PF. Development and psychometric properties of the DASS-youth (DASS-Y): an extension of the depression anxiety stress scales (DASS) to adolescents and children. *Front Psychol* 2022; 13. doi:10.3389/fpsyg.2022.766890
36. Young CC and Dietrich MS. Stressful life events, worry, and rumination predict depressive and anxiety symptoms in young adolescents. *J Child Adolesc Psychiatr Nurs* 2015; 28: 35–42.
37. Young CC and Dietrich MS. Screening for rumination and brooding may be a feasible method of identifying adolescents at high risk for depression. *J Pediatr Nurs* 2014; 29: 688–695.
38. Cole A, Bond C, Qualter P, et al. A systematic review of the development and psychometric properties of loneliness measures for children and adolescents. *Int J Environ Res Public Health* 2021; 18: 3285.
39. van Genugten CR, Schuurmans J, Hoogendoorn AW, et al. A data-driven clustering method for discovering profiles in the dynamics of Major depressive disorder using a smartphone-based ecological momentary assessment of mood. *Front Psychiatry* 2022; 13: 755809.
40. Migueles JH, Rowlands AV, Huber F, et al. GGIR: a research community-driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. *J Meas Phys Behav* 2019; 2: 188–196.
41. Hossain T, Ahad MAR and Inoue S. A method for sensor-based activity recognition in missing data scenario. *Sensors* 2020; 20: 3811.
42. Harrell FE Jr. Hmisc: Harrell Miscellaneous. Published online July 17, 2003;5.2-1. doi:10.32614/CRAN.package.Hmisc
43. Lindstrom ML and Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990; 46: 673–687.
44. Kuznetsova A, Brockhoff PB and Christensen RHB. lmerTest Package: tests in Linear Mixed Effects Models. *J Stat Softw* 2017; 82. doi:10.18637/jss.v082.i13
45. Cnaan A, Laird NM and Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med* 1997; 16: 2349–2380.
46. Kwong KS, Holland B and Cheung SH. A modified Benjamini-hochberg multiple comparisons procedure for controlling the false discovery rate. *J Stat Plan Inference* 2002; 104: 351–362.
47. Gelman A, Hill J and Yajima M. Why we (usually) don't have to worry about multiple comparisons. *J Res Educ Eff* 2012; 5: 189–211.
48. Dao KP, Cocker KD, Tong HL, et al. Smartphone-Delivered ecological momentary interventions based on ecological

- momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. *JMIR MHealth UHealth* 2021; 9: e22890.
49. Roberts C, Herzing JME, Sobrino Piazza J, et al. Data privacy concerns as a source of resistance to complete Mobile data collection tasks via a smartphone app. *J Surv Stat Methodol* 2022; 10: 518–548.
 50. Hill CA, Biemer PP, Buskirk TD, et al. *Big data meets survey science: a collection of innovative methods*. Hoboken, NJ: John Wiley & Sons, 2020.
 51. Cherif E, Bezaz N and Mzoughi M. Do personal health concerns and trust in healthcare providers mitigate privacy concerns? Effects on patients' intention to share personal health data on electronic health records. *Soc Sci Med* 2021; 283: 114146.
 52. Dumith SC, Gigante DP, Domingues MR, et al. III. Physical activity change during adolescence: a systematic review and a pooled analysis. *Int J Epidemiol* 2011; 40: 685–698.
 53. Duraccio KM and Jensen CD. Associations between physical and sedentary activity regularity and sleep in preschoolers and kindergartners. *Sleep Health* 2017; 3: 263–268.
 54. Hrafnkelsdottir SM, Brychta RJ, Rognvaldsdottir V, et al. Less screen time and more physical activity is associated with more stable sleep patterns among Icelandic adolescents. *Sleep Health* 2020; 6: 609–617.
 55. Yang X, Knights J, Bangieva V, et al. Association between the severity of depressive symptoms and human-smartphone interactions: longitudinal study. *JMIR Form Res* 2023; 7: e42935.
 56. Rozgonjuk D, Levine JC, Hall BJ, et al. The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. *Comput Hum Behav* 2018; 87: 10–17.
 57. Vidal C, Lhaksampa T, Miller L, et al. Social media use and depression in adolescents: a scoping review. *Int Rev Psychiatry Abingdon Engl* 2020; 32: 235–253.
 58. Kelly Y, Zilanawala A, Booker C, et al. Social Media use and adolescent mental health: findings from the UK millennium cohort study. *eClinicalMedicine* 2018; 6: 59–68.
 59. Fardouly J, Magson NR, Rapee RM, et al. The use of social media by Australian preadolescents and its links with mental health. *J Clin Psychol* 2020; 76: 1304–1326.
 60. Cohen R, Newton-John T and Slater A. The relationship between Facebook and Instagram appearance-focused activities and body image concerns in young women. *Body Image* 2017; 23: 183–187.
 61. Peeters F, Berkhof J, Delespaul P, et al. Diurnal mood variation in major depressive disorder. *Emotion* 2006; 6: 383–391.
 62. Schacter HL and Margolin G. When it feels good to give: depressive symptoms, daily prosocial behavior, and adolescent mood. *Emotion* 2019; 19: 923–927.