

Predictors of engagement with remote sensing technologies for symptom measurement in Major Depressive Disorder

Matcham F^{1*}, Carr E^{2*}, White KM¹, Leightley D¹, Lamers F³, Siddi S⁴, Annas P⁵, de Girolamo G⁶, Haro JM⁴, Horsfall M³, Ivan A¹, Lavelle G¹, Li Q⁷, Lombardini F⁴, Mohr DC⁸, Narayan VA⁷, Penninx BWJH³, Oetzmann C¹, Coromina M⁹, Simblett SK¹⁰, Weyer J¹¹, Wykes T^{10,12}, Zorbas S¹¹, Brasen JC⁵, Myin-Germeys I¹³, Conde P², Dobson RJB², Folarin AA², Ranjan Y², Rashid Z², Cummins N², Dineley J^{2,14}, Vairavan S⁷, Hotopf M^{1,12}, on behalf of the RADAR-CNS consortium¹⁵.

1. Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK.
2. Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK.
3. Department of Psychiatry and Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands.
4. Parc Sanitari Sant Joan de Déu, Fundació San Joan de Déu, CIBERSAM, Universitat de Barcelona, Barcelona, Spain.
5. H. Lundbeck A/S, Valby, Denmark.
6. IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy.
7. Janssen Research and Development, LLC, Titusville, NJ, USA.
8. Center for Behavioral Intervention Technologies, Department of Preventative Medicine, Northwestern University, Chicago, IL, USA.
9. Parc Sanitari Joan de Déu, Barcelona, Spain.
10. Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK.
11. RADAR-CNS Patient Advisory Board.
12. South London and Maudsley NHS Foundation Trust, London, UK.
13. Department for Neurosciences, Center for Contextual Psychiatry, KU Leuven, Leuven, Belgium.
14. EIHW – Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany.
15. www.radar-cns.org

*Joint first authors

Corresponding author: Faith Matcham, faith.matcham@kcl.ac.uk.

Abstract

Background

Remote sensing for the measurement and management of long-term conditions such as Major Depressive Disorder (MDD) is becoming more prevalent. User-engagement is essential to yield any benefits. We tested three hypotheses examining associations between clinical characteristics, perceptions of remote sensing, and objective user engagement metrics.

Methods

The Remote Assessment of Disease and Relapse – Major Depressive Disorder (RADAR-MDD) study is a multicentre longitudinal observational cohort study in people with recurrent MDD. Participants wore a Fitbit and completed app-based assessments every two weeks for a median of 18 months. Multivariable random effects regression models pooling data across timepoints were used to examine associations between variables.

Results

A total of 547 participants (87.8% of the total sample) were included in the current analysis. Higher levels of anxiety were associated with lower levels of perceived technology ease of use; increased functional disability was associated with small differences in perceptions of technology usefulness and usability. Participants who reported higher system ease of use, usefulness, and acceptability subsequently completed more app-based questionnaires and tended to wear their FitBit activity tracker for longer. All effect sizes were small and unlikely to be of practical significance.

Limitations

Symptoms of depression, anxiety, functional disability, and perceptions of system usability are measured at the same time. These therefore represent cross-sectional associations rather than predictions of future perceptions.

Conclusions

These findings suggest that perceived usability and actual use of remote measurement technologies in people with MDD are robust across differences in severity of depression, anxiety, and functional impairment.

Keywords: Major Depressive Disorder, Remote Sensing, Cohort study, Engagement, Predictors

Introduction

Depressive disorders are the third leading cause of Years Lived with Disability (YLD), contributing to 14.3% of all-age YLD (James et al., 2018). Major depressive disorder (MDD) is a severe form of depression, widely experienced as a chronic condition characterised by periods of recovery and relapse (Verduijn et al., 2017). Symptoms include anhedonia and low mood, often combined with a range of other cognitive, physical, behavioural and emotional symptoms such as a loss of concentration, changes in appetite, reduced levels of activity and feelings of guilt or worthlessness. The chronic nature of MDD means that those affected should ideally have regular symptom monitoring, which is conventionally conducted via self-reported questionnaires or clinical interviews. These forms of symptom measurement are necessarily infrequent and subject to recall bias (Schmier and Halpern, 2004).

Harnessing mobile technologies (such as smartphones and/or wearable devices) to assess, manage and support chronic illness is becoming increasingly popular for capture of high-frequency multimodal behavioural, psychological, and physiological data in naturalistic settings (Dobkin, 2013). This form of data collection via remote sensing overcomes many of the limitations posed by conventional outcome assessment at minimal additional burden to patients (Mohr et al., 2020). There is also a growing literature reporting associations between depression severity and data collected via inbuilt smartphone sensors and wearables such as homestay, sleep duration and activity levels (Rohani et al., 2018; Laiou et al., 2022; Zhang et al., 2021). This growing research and clinical interest have been matched by commensurate industry investment. Over \$400 million in venture capital investment have been allocated to the development of mental health applications (apps) throughout 2019 (Connolly et al., 2020).

Yielding benefits from any form of mobile technology relies on user-engagement (Bakker and Rickard, 2019). Engagement, defined as not only initial uptake of a new system, but sustained usage and interaction over time (O'Brien and Toms, 2008), is dependent on a range of interconnected processes, defined usefully in the Technology Acceptance Model (TAM, supplementary figure 1, (Lee et al., 2003)). The TAM, originally adapted from the Theory of Reasoned Action (Ajzen and Fishbein, 1980), proposes that adoption and use of a novel system is dependent on intention to use the system, which in turn is influenced by inter-related perceptions of usefulness and ease of use. External variables, which are independent from the technology being assessed, may include social, cognitive, or experiential variables, hypothesised to impact perceptions of system usefulness and ease of use. The TAM has been widely used to inform the development and implementation of health informatics; from the introduction of telemedicine and electronic health records to the development of mobile applications (apps) and digital interventions (Rahimi et al., 2018).

Remote sensing is a new application of the TAM, requiring individuals to engage with novel technologies for long periods of time, without necessarily receiving any intervention or immediate reward for doing so (McGrady et al., 2010). Recent systematic review evidence has emphasised the importance of this distinction: whereas engagement with digital behaviour change interventions (DBCI) tends to be linked with the target outcome, sustained commitment to remote sensing depends on less tangible outcomes to motivate users (White et al., 2022).

Remote sensing presents novel challenges, particularly in clinical populations. Use of a new technology may not solely depend on an individual's intentions and perceptions, but also presentation of symptoms which will influence their ability to adapt to the system requirements

and maintain use throughout fluctuations in symptom severity (Simblett et al., 2018). We therefore hypothesise an additional pathway between clinical characteristics and actual system use (hypothesis 1).

Within the context of MDD, there are several elements of living with chronic mental illness which could be linked with both perceptions of system usability and actual use. Cognitive models of depression highlight the importance of the negative schemas individuals with depression have about themselves, the world and the future (Beck, 1967). These dysfunctional attitudes contribute to a bias in information processing, whereby negative experiences and information are prioritised to fit with pre-existing negative schemas (Huntley and Fisher, 2016). Individuals with MDD experiencing high levels of depression or anxiety, faced with the challenge of mastering a new technological system, may therefore have worsened subjective perceptions of the usability or acceptability of a novel system (Patel et al., 2020).

The experience of living with MDD may also have a direct impact on engagement with technology. For example, symptoms of MDD include a loss of interest in daily activities and forgetfulness, which can affect intrinsic motivation to behave proactively (Franzen and Brinkmann, 2016). Using wearables or smartphone apps to monitor symptoms can be seen as a self-management behaviour, requiring individuals to sustain enthusiasm to continue to engage with the technology (Connolly et al., 2020). Furthermore, high levels of functional disability are prevalent in MDD, with depression frequently impacting the extent to which an individual can participate in and enjoy their normal daily activities (Luty et al., 2002). An increase in depression severity might be associated with reduced engagement with mobile technologies for some, whereas it may trigger more diligent health monitoring in others (Simblett et al., 2019b).

An analysis of students' smartphone use over one week revealed a significant association between reduced depression severity and increased general smartphone use (Elhai et al., 2018). Better emotional health has also been associated with increased engagement with both smartphone apps and wearable technologies in people with bipolar disorder (Van Til et al., 2020). However, there is no evidence that the severity of depression and anxiety symptoms impact health app use in people with pre-existing depression or anxiety (Mohr et al., 2021).

Whilst this is a burgeoning field of research, multiple systematic reviews have highlighted the lack of adequate reporting with regards to the acceptability, usability and indicators of long-term engagement with mobile technology (Batra et al., 2017; De Angel et al., 2022; Lee et al., 2021). One review examining usability of wearable devices in chronic conditions could not provide any substantial conclusions due to the poor quality of reporting in the field (Keogh et al., 2021). Similarly, in a systematic review examining the passive monitoring of depression via smartphone sensors, de Angel et al., (2022) were unable to conduct a meta-analysis for the same reason. There is a clear need for higher quality datasets in clinical populations.

Here, we report findings from Remote Assessment of Disease and Relapse—Major Depressive Disorder (RADAR-MDD; (Matcham et al., 2019)), a longitudinal, observational cohort study being conducted as part of the Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS; www.radar-cns.org) programme. RADAR-MDD aims to use data collected passively via wearable devices and smartphone sensors, combined with app-delivered assessments, to predict relapse in people with MDD. With over 600 individuals participating in the study for a median of 541 days (Matcham et al., 2022), RADAR-MDD provides insights into relationships between clinical presentation, perceptions of technology

usability, and long-term engagement in a major depression population. Accordingly, we aim to test the following hypotheses:

- H₁ More severe clinical symptoms (depression and anxiety) and functional impairment are associated with lower levels of FitBit wear-time and app questionnaire responses.
- H₂ More severe clinical symptoms and functional impairment are associated with lower levels of self-reported system (Fitbit and app) usability and acceptability.
- H₃ Low levels of self-reported system usability and acceptability are associated with lower levels of FitBit wear-time and app questionnaire responses.

Methods

Study Design

RADAR-MDD is a multicentre, longitudinal, observational cohort study in people with recurrent MDD. Participants were enrolled from three recruitment sites: King's College London (KCL; London, United Kingdom); Amsterdam University Medical Centre, location VUmc (Amsterdam, The Netherlands); and Centro de Investigación Biomédica en Red (CIBER; Barcelona, Spain). A detailed description of the recruitment process at each site is provided elsewhere (Matcham et al., 2019).

Study Population

A total of 623 individuals were enrolled in RADAR-MDD and followed-up for a median of 541 days (interquartile range (IQR): 401-730; (Matcham et al., 2022)). To be eligible for inclusion in RADAR-MDD individuals needed to: have recurrent MDD (a lifetime history of ≥ 2 episodes, the most recent of which must have been within the past 2 years); able and willing to complete self-reported assessments via smartphone; able to give informed consent; be fluent in English, Dutch, Catalan or Spanish; have an existing Android smartphone or be willing to swap to an Android for the purposes of the study; and be aged over 18 years old. Individuals with a lifetime history of bipolar disorder, schizophrenia, MDD with psychotic features or schizoaffective disorder, dementia, recent history of drug or alcohol misuse or major medical diagnosis were excluded.

Data Collection

RADAR-MDD collected data from a range of sources, detailed in other publications (Matcham et al., 2022, 2019). Here we describe the measures used in the current analysis (Figure 1).

Clinical characteristics

Participants completed assessments of depression, anxiety, and functional disability severity every 3 months via REDCap automated surveys (Harris et al., 2009). Only data collected at

3- and 12-month follow-up time points were used in the current analysis to align with when self-reported usability and acceptability questionnaires were delivered.

Current depression severity was measured using the Inventory of Depressive Symptomatology – Self Report (IDS-SR; (Rush et al., 2000)). The IDS-SR contains 30 items, covering typical MDD symptoms as well as subtypes including melancholic and atypical depression. Response options ranged from 0-3 with 0 indicating the lowest level of severity, frequency, or duration of a symptom, and 3 indicating the highest level of severity, frequency or duration of a symptom. Scores are summed to create an overall score ranging between 0-84, with higher scores indicating increasing depression severity.

Symptoms of anxiety were measured using the Generalised Anxiety Disorder (GAD-7; (Spitzer et al., 2006)) questionnaire that consists of 7 items assessing how often an anxiety symptom was experienced within the last two weeks. Response options ranged between 0 (“Not at all”) and 3 (“Nearly every day”). Responses to each item are summed to create a total score ranging from 0-21 with higher scores indicating greater anxiety severity.

Functional disability was assessed using the Work and Social Adjustment Scale (WSAS; (Mundt et al., 2002)). This questionnaire has 5 items relating to different aspects of function (ability to work, home management, social leisure activities, private leisure activities, and maintenance of relationships), each scored according to how severely their depression affects each domain. Scores for each item range between 0 (not at all) and 8 (very severely), with a total score ranging from 0-40 indicating increased severity of functional disability.

Usability and Acceptability

Self-reported perceptions of system usability and acceptability were measured via REDCap surveys emailed to participants at 3, 12, and 24 months after enrolment. The present analysis excludes usability and acceptability data collected at 24 months since, as this was the end of the study, information on subsequent app and Fitbit usage beyond 24 months were unavailable.

Perceptions of system usability were measured with the Post-Study System Usability Questionnaire (PSSUQ; (Lewis, 1992)). The PSSUQ contains 19-items with responses ranging from 1 (strongly agree) to 7 (strongly disagree) for a range of system usability descriptors such as efficiency, ease of manipulation and flexibility. A total score was derived by summing the individual items. Higher scores indicate lower levels of perceived system usability, however for ease of interpretation, scores were reversed in our analyses, with higher scores representing increased perceived usability.

Perceptions of system acceptability were measured using the Technology Acceptance Model – Fast Form (TAM-FF; (Chin et al., 2008)). The TAM-FF has 16-items which reflect three subscales: perceived usefulness, ease of use, and predicted future use. Each item is scored from 0 (strongly agree) to 7 (strongly disagree). We derive a total score by summing the 16 items as well as scores for subscales (perceived usefulness and ease of use) calculated as the mean of each subscale. Higher scores indicate lower levels of perceived system acceptability, however for ease of interpretation, scores were reversed in our analyses, with higher scores representing increased perceived acceptability.

Smartphone app and wearable device usage

We considered two measures of usage. First, engagement with the RADAR-base active remote measurement (aRMT) app (Matcham et al., 2022, 2019; Ranjan et al., 2019) was

assessed based on the number of PHQ-8 questionnaires completed in the three months after each follow-up assessment. Participants were sent fortnightly self-report PHQ-8 questionnaires via the mobile app. We counted the number of questionnaires (between 0 and 9) at 3-6 and 12-15 months separately (Figure 1).

Secondly, the proportion of the time that participants were wearing their FitBit device was assessed also in the three months after each follow-up assessment. This was measured based on the availability of heart rate data for each 15-minute period of each day. We calculated the proportion of 15-minute periods with heart rate data per day and then the mean of all days per month. We then summed and rescaled the means for the three months after each follow-up assessment (3-6 and 12-15 months, separately). This gave an outcome ranging 0 to 1 where '0' indicated participants never wore their FitBit; '1' indicated they always wore their FitBit.

For both usage measures, we excluded observations where a participant withdrew midway through a 3-month period (e.g., withdrew at 3-6 or 12-15 months). This was to ensure that usage measures were calculated for the same time period (3 months) for all participants.

Covariates

At enrolment, participants completed detailed socio-demographic questionnaires detailing their age, gender, educational attainment as well as indicators of socioeconomic status, social environment, previous experience of using mobile technologies, and lifetime traumatic events (measured via the List of Threatening Experiences – Questionnaire (LTE-Q; Brugha & Cragg, 1990). A detailed summary of the baseline demographic variables collected can be found in our protocol paper (Matcham et al., 2019). At follow-up assessments participants were asked about lifetime traumatic events in the past three months. All covariates were collected via the Research Electronic Data Capture (REDCap) software (Harris et al., 2009).

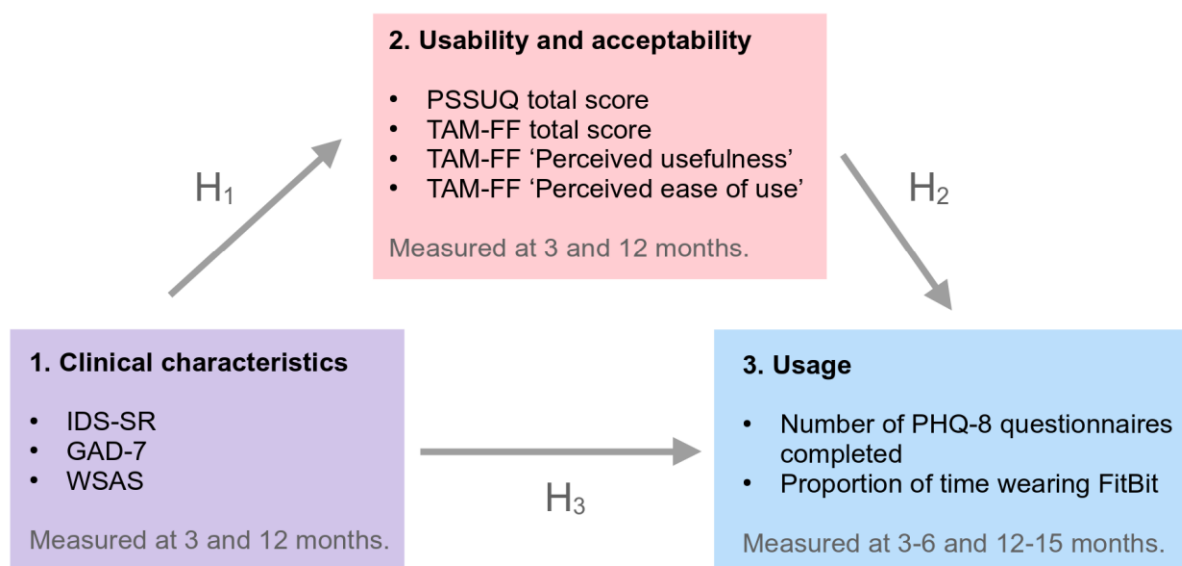


Figure 1. Summary of measures and follow-up periods.

Patient and Public Involvement

The study was co-developed with service users in our Patient Advisory Board. They were involved in the choice of measures, the timing and issues of engagement and have also been involved in developing the analysis plan and representatives are authors of this paper and critically reviewed it.

Statistical Analysis

The analysis was conducted in three stages. First, we described socio-demographic characteristics of the sample at enrolment (0 months) and clinical variables, usability and usage at follow-up (3 and 12 months) using appropriate summary statistics (median and inter quartile range (IQR) for continuous variables; frequency and percentage for categorical variables).

Second, we calculated Pearson correlations between clinical variables (in the current time-point), usability (in the current time-point) and usage (over the next three months). A sandwich estimator (Lumley, 2004) was used to account for repeated measures per participant (at 3 and 12 months). 95% confidence intervals were estimated with a percentile bootstrap (1000 repetitions).

Third, we used multivariable random effects regression models to consider three sets of relationships, shown in Figure 1, between: clinical variables and perceived usability, both measured in the current time-point (H_1); perceived usability in the current time-point and usage over the subsequent three months (H_2); and clinical variables in the current time-point and usage over the subsequent three months (H_3). We pooled information collected at 3 and 12 months, such that participants had one or two assessments of each outcome. All models included a participant-level random intercept.

Continuous outcomes (PSSUQ total score, TAM-FF total, and sub-scores) were modelled using linear regression models. The number of PHQ-8 questionnaires completed was modelled using a sequential ordinal model (Bürkner and Vuorre, 2019; Min and Agresti, 2005; Valle et al., 2019). This models the outcome as a counting process where achieving a count of k requires that all lower counts of 1 to $k-1$ have already been achieved. Lastly, the proportion of time wearing a FitBit was modelled using zero-inflated beta regression (Ospina and Ferrari, 2012). This model assumes the outcome has a mixed continuous-discrete distribution, where a beta distribution is used to model the continuous part of the outcome (the proportion of time wearing a FitBit, among those who ever wear a FitBit) and a logistic distribution is used to model the discrete part (the log odds of never wearing a FitBit).

We chose appropriate adjustment sets for each model by specifying directed acyclic graphs (DAGs) that specified relationships between potential confounders and the exposure and outcome of interest (see supplementary materials for details). Models considering clinical variables and usage (H_1) were adjusted for age, comorbid conditions, years of education, employment status and previous wearable use. Models for clinical variables and perceived usability (H_2) were adjusted for age and years of education. Models considering the effect of perceived usability on usage (H_3) were adjusted for comorbid conditions, employment status, previous wearable use, and physical impairment (WSAS).

All models were estimated in a Bayesian framework using the brms package (Bürkner and Vuorre, 2019) for R (R Core Team, 2021). We used noninformative priors (flat priors for fixed

effects and half student-t priors for random effects). Estimates were summarised based on the posterior distribution of expected values from each model. We calculated the difference in the expected values of the outcome for a 1 standard deviation (SD) difference in the exposure. All continuous variables were standardised (rescaled to have a mean of 0 and standard deviation of 1) before including in regression models. All code used in these analyses is available at <https://osf.io/e9by5>.

Results

The analytical sample included 547 individuals, having excluded participants without any information on clinical variables ($n=7$), usability and acceptability ($n=65$), and covariates ($n=4$). Excluded participants were similar to the analytical sample in terms of age (median of 51 and 48, respectively), male gender (25% and 24%), years of education (15 and 16) and the percentage reporting 2+ comorbid conditions (39% vs. 31%).

Table 1 describes the analytical sample. Most participants in the analytical sample were female (76%). The median age was 48 (interquartile range (IQR) = 32-58), 27% reported two or more comorbid conditions and 41% were in paid employment at the time of enrolment. Further description of sociodemographic covariates in relation to the outcomes is presented in Supplementary Table 1.

Figure 2 presents the distribution of usability and usage measures at 3- and 12-month follow-up assessments. Both usage outcomes (Fitbit wear time and PHQ8 completions) exhibited a large number of zeros. For FitBit wear time, 19% of assessments (202 of 1082) were zero, indicating that the participant did not wear their FitBit device during that 3-month period. The remaining 81% of responses were distributed around a median of 0.89 (IQR = 0.60, 0.96). For PHQ-8 completions, 19% (186 of 975) of assessments participants returned no questionnaires; the remaining responses were distributed around a median of 5 questionnaires returned (IQR = 3, 6).

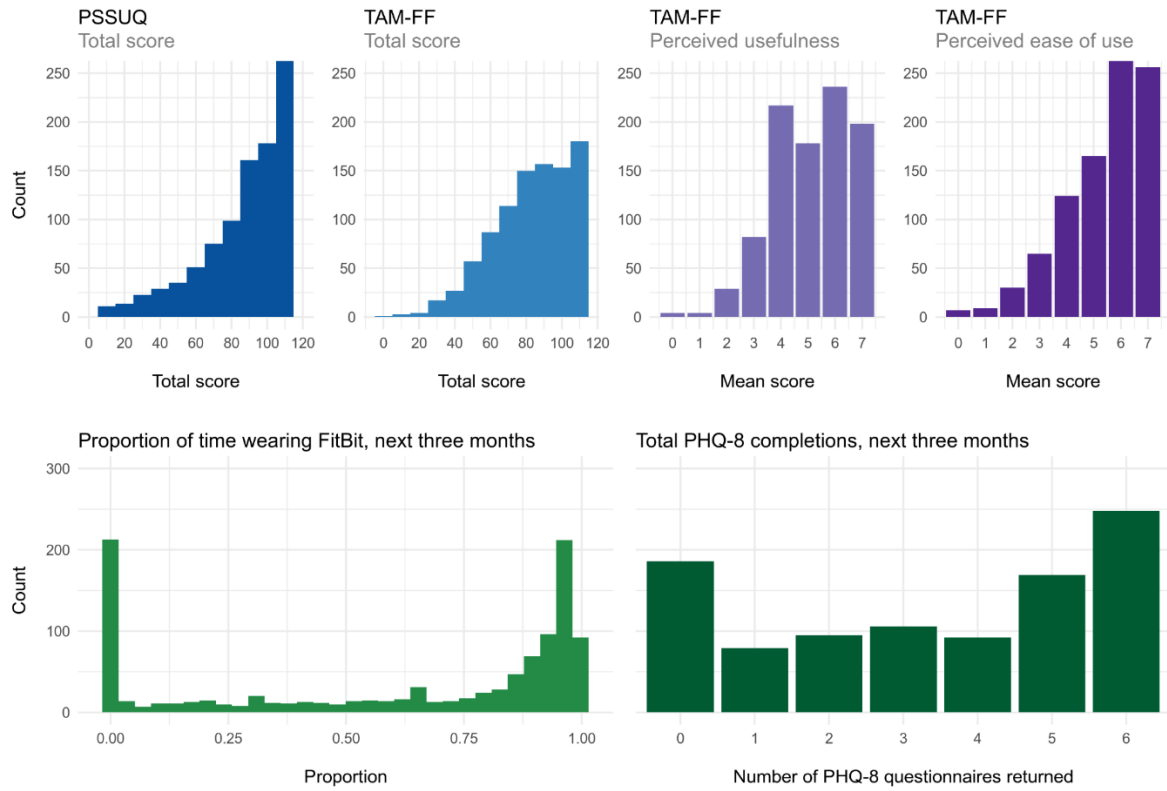


Figure 2. Distributions of (i) usability and acceptability measures; (ii) outcome measures. N = 547 participants.

Table 1. Description of the analytical sample.

	0 months N = 547	3 months N = 547	12 months N = 535
Demographic characteristics			
Age ^a	48 (32, 58) [18, 80]	–	–
Male gender ^b	133 (24%)	–	–
Number of comorbid conditions ^b			
0	279 (51%)	–	–
1	96 (18%)	–	–
2+	172 (32%)	–	–
Years of education ^a	16.0 (13.0, 19.0) [0.0, 55.0]	–	–
Currently working ^b	226 (41%)	–	–
Recent life events, last 3 months			
0	229 (42%)	225 (42%)	189 (45%)
1	164 (30%)	140 (26%)	127 (30%)
2	78 (14%)	79 (15%)	54 (13%)
3+	73 (13%)	97 (18%)	51 (12%)
Clinical variables			
Inventory of Depressive Symptomatology-Self Report (IDS-SR)			
Median (IQR) [Range]	30 (20, 41) [0, 70]	31 (22, 43) [0, 71]	30 (19, 42) [0, 69]
N missing	11	10	104
Generalised Anxiety Disorder-7			
Median (IQR) [Range]	8.0 (5.0, 13.0) [0.0, 21.0]	8.0 (5.0, 14.0) [0.0, 21.0]	7.0 (4.0, 13.0) [0.0, 21.0]
N missing	3	4	112
Work and Social Adjustment Scale (WSAS)			
Median (IQR) [Range]	20 (11, 28) [0, 40]	21 (11, 29) [0, 40]	19 (10, 29) [0, 40]
N missing	4	4	113

Usability and acceptability			
Post-Study System Usability Questionnaire (PSSUQ; reversed)			
Median (IQR) [Range]	–	94 (75, 106) [0, 114]	93 (73, 107) [0, 114]
N missing	–	5	120
Technology Acceptance Model (TAM; reversed)			
Median (IQR) [Range]	–	86 (71, 101) [7, 112]	87 (68, 103) [0, 112]
N missing	–	6	126
Perceived usefulness			
Median (IQR) [Range]	–	5.17 (4.17, 6.33) [0.50, 7.00]	5.33 (4.00, 6.50) [0.00, 7.00]
N missing	–	8	126
Perceived ease of use			
Median (IQR) [Range]	–	5.83 (4.67, 6.67) [0.17, 7.00]	5.67 (4.50, 6.83) [0.00, 7.00]
N missing	–	8	126
Usage			
Proportion of time wearing FitBit, next 3 months			
Median (IQR) [Range]	–	0.90 (0.56, 0.97) [0.00, 1.00]	0.74 (0.17, 0.94) [0.00, 1.00]
N missing	–	0	0
Number of PHQ-8 questionnaire completions, next 3 months			
0	–	86 (16%)	100 (23%)
1	–	39 (7.2%)	40 (9.3%)
2	–	53 (9.8%)	42 (9.7%)
3	–	63 (12%)	43 (10.0%)
4	–	56 (10%)	36 (8.3%)
5	–	95 (17%)	74 (17%)
6	–	151 (28%)	97 (22%)

^a Median (IQR) [Range]

^b n (%)

^c Recent life events was measured at enrolment and follow-up. At enrolment, this measure captured events occurring “in the last year”. At follow-up, it captured events occurring in the past three months.

Figure 3 presents bivariate correlations between clinical variables, usability and usage. We found small negative correlations between clinical variables (GAD-7, IDS-SR, WSAS) and both usability and usage. Conversely, we saw small positive correlations between usability and the two usage outcomes.

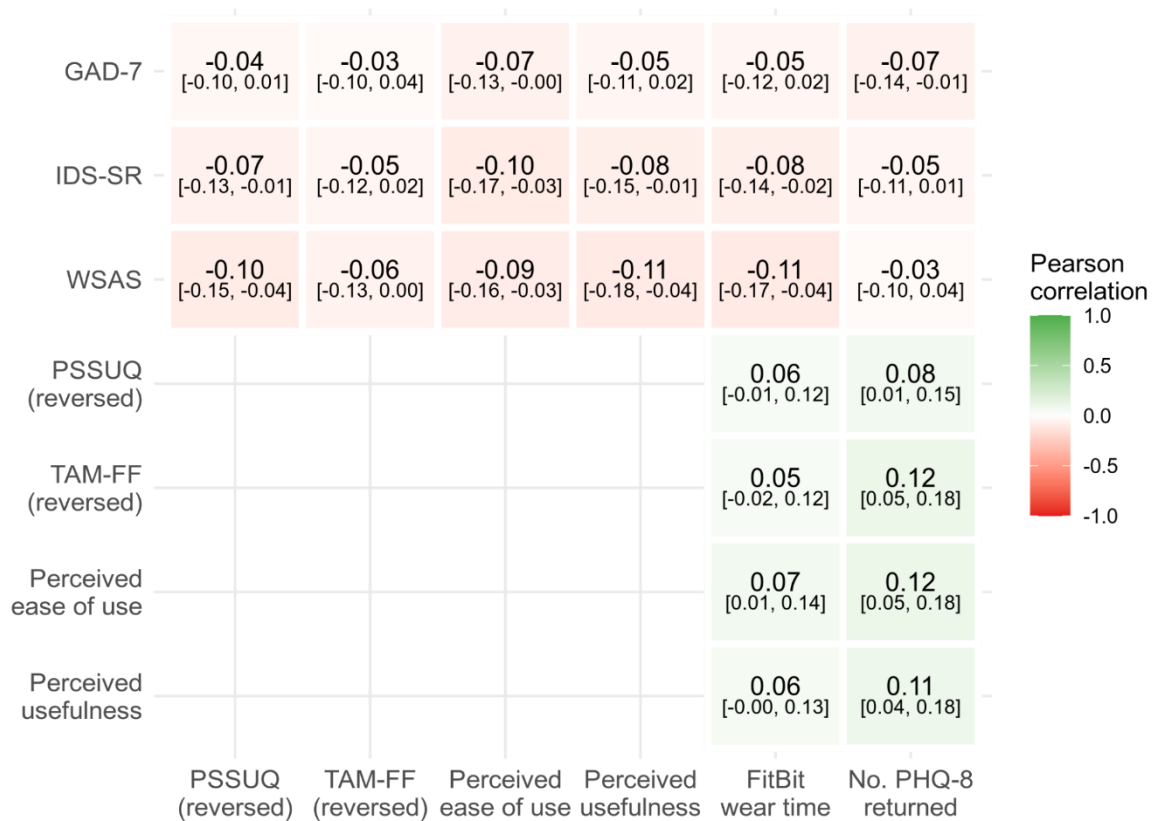


Figure 3. Pearson correlations between clinical variables, usability, and usage (547 participants) [95% bootstrap percentile confidence interval]

Tables 2-3 (and supplementary Figures 2-4) summarise a series of multivariable regression models that tested our three hypotheses in turn (H_1 , H_2 , and H_3). These tables and figures present the expected difference in the outcome for a one standard deviation (1 SD) difference in the exposure, holding other covariates to their average values.

Overall, we find results consistent with the correlation matrix above: there were very small negative associations for clinical variables with perceived usability (Table 2) and subsequent usage (Table 3); and very small positive associations for perceived usability and usage with subsequent usage (Table 3). In all cases, the observed effects sizes were consistent with no overall effect.

For example, regarding the effect of clinical variables on perceived usability (H_1), a 1 SD increase in GAD-7 was associated with reductions in perceived ease of use and usefulness of only 0.11 and 0.05, respectively. Both scales range from 0-7, making this a negligible change. Similarly, a 1 SD change in WSAS was associated with a 2-unit reduction in PSSUQ, a scale ranging from 0-114.

Table 2. Influence of clinical variables on perceived usability (H₁; 547 participants). Estimates show estimated difference in each outcome per 1 standard deviation difference in each predictor [95% Credible Intervals].

	Outcome			
	PSSUQ	TAM-FF	TAM-FF	TAM-FF
	Total score	Total score	Ease of use	Usefulness
	Estimated change in score	Estimated change in score	Estimated change in score	Estimated change in score
GAD-7	-0.76 [-2.48, 0.95]	-0.62 [-2.04, 0.79]	-0.11 [-0.21, -0.01]	-0.05 [-0.14, 0.04]
IDS-SR	-1.27 [-3.04, 0.45]	-0.54 [-2.00, 0.92]	-0.10 [-0.20, 0.01]	-0.08 [-0.18, 0.01]
WSAS	-1.95 [-3.72, -0.19]	-0.87 [-2.33, 0.59]	-0.10 [-0.20, 0.00]	-0.12 [-0.21, -0.02]

Adjusted for age and years of education.

95% CrI = 95% credible intervals.

We found similarly weak effects of perceived usability on the number of PHQ-8 questionnaires completed and FitBit wear time (H₂; Table 3). A 1 SD increase in the TAM-FF questionnaire was associated with a 0.29 increase in the number of questionnaires completed over the next 3 months. Perceived usability did not appear to influence FitBit wear time. The observed effects (e.g., increases of 2% or 3%) were in the expected direction but of no practical significance.

Finally, we found very small negative effects of clinical variables on the number of PHQ-8 questionnaires completed, but again these were not of practical significance. We found no evidence of effects of clinical variables on FitBit wear time.

Table 3. Influence of usability (H₂) and clinical variables (H₃) on the number of PHQ-8 completions and FitBit wear time (547 participants). Estimates show predicted difference in each outcome per standard deviation difference in each predictor [95% Credible Intervals].

	Outcomes measured over next 3 months	
	Number of PHQ-8 questionnaires completed	Proportion of time wearing FitBit
	Estimated change in mean count	Estimated change in proportion
H₂ Influence of usability and acceptability on subsequent usage^a		
PSSUQ	0.12 [-0.04, 0.28]	0.01 [-0.00, 0.03]
TAM-FF (Total score)	0.29 [0.12, 0.46]	0.01 [-0.00, 0.03]
TAM-FF (Ease of Use)	0.33 [0.15, 0.50]	0.02 [0.00, 0.04]
TAM-FF (Usefulness)	0.24 [0.07, 0.41]	0.02 [-0.00, 0.04]
H₃ Influence of clinical variables^b		
GAD-7	-0.13 [-0.25, -0.00]	-0.00 [-0.02, 0.02]
IDS-SR	-0.08 [-0.21, 0.05]	-0.02 [-0.04, 0.00]
WSAS	-0.06 [-0.18, 0.07]	-0.03 [-0.05, -0.01]

^a Adjusted for comorbid conditions, employment status, previous wearable use, and WSAS.

^b Adjusted for age, comorbid conditions, years of education, employment status, and previous wearable use, recent life events.

Discussion

We tested three hypotheses: 1) more severe clinical symptoms (depression, anxiety and functional disability) would be associated with lower levels of FitBit wear-time and fewer app questionnaire responses; 2) more severe clinical symptoms would be associated with reduced perceptions of system usability and acceptability; and 3) lower levels of self-reported system usability and acceptability would be associated with lower levels of FitBit wear-time and fewer app questionnaire responses. We found mixed evidence to support these hypotheses. Although the identified associations were in expected directions the effect sizes were extremely small. Large differences in each predictor were associated with very small differences in each outcome.

We found some evidence of an association between other clinical symptoms (anxiety and functional disability) and engagement with the RADAR-base system. Higher levels of anxiety were associated with a lower number of completed app-based questionnaires, but not the proportion of time spent wearing a FitBit. Higher levels of functional disability were associated with a reduction of FitBit wear-time, but not with the number of app-based questionnaires completed. Depression severity was not associated with any measure of acceptability or usability, or with Fitbit wear time or questionnaire completions.

These findings highlight a potential difference between elements of remote sensing requiring active engagement (such as responding to an app notification to complete a questionnaire) and passive engagement (such as wearing a fitness tracker). Active engagement requires more effort, time-management, cognitive capacity, and motivation; all of which may be compromised in people who are experiencing high levels of depression or anxiety (Lang et al., 1998; Rosa-Alcázar et al., 2021). Passive engagement may be robust to fluctuations in depression and anxiety, but more susceptible to changes in physical function, particularly as the technology in question measures physical activity. Although wearables require no active involvement from the wearer, they do provide feedback about the amount of activity an individual has undertaken. Individuals whose depression is having a significant impact on their ability to participate in their usual work, social or leisure activities may find this feedback de-motivating and increase the likelihood of them removing the device (Kerner and Goodyear, 2017; Kononova et al., 2019). On the other hand, providing a clear report with the results of the monitoring, with appropriate explanations, to each patient wearing a biosensor may provide the rationale to promote a change in individual behaviour. For instance, a patient receiving a report of a 7-day sensor monitoring of physical activity may realise that he/she is spending more time being sedentary than they thought, and this may trigger a behavioural change.

We reported small associations between increased anxiety and reduced levels of perceived ease of use; and increased functional disability and reduced perceived usefulness and overall usability scores. Tsai, Lin, Chang, Chang and Lee (2020) reported a similar result in their investigation of perceptions of cardiac warning systems. People with high levels of anxiety were found to rate the technology as having a lower perceived ease of use, with the authors concluding that a lack of familiarity with novel technology and anxiety about mastering a new skill were key drivers of this association (Tsai et al., 2020). Although this study examined a different population, similar results have been reported elsewhere, with anxiety often highlighted as having a moderating influence on perceived ease of use across a range of digital technologies (Baki et al., 2018; Saadé and Kira, 2007).

We identified an association between increased self-reported ease of use and usefulness, and an increase in the number of questionnaire completions, and overall Fitbit wear time. Higher overall acceptability scores were also associated with an increase in the number of questionnaires completed via the app. These findings are consistent with previous findings linking perceptions of system utility and usability with increased likelihood of long-term engagement (Or et al., 2011).

Recent research has highlighted potential psychological and cognitive antecedents of intentions to interact with mobile technologies. Demographic characteristics such as higher educational attainment and lower age have been associated with reporting higher levels of satisfaction and improved usage statistics of health e-systems (Harrati et al., 2016). A large-scale cross-sectional analysis highlighted several precursors of retention in digital health studies, with Pratab et al. (2020) emphasising that clinician referral, relevant clinical condition, financial incentive and older age are associated with increased participation. However, very little work has examined a wider range of clinical predictors of engagement with health technologies; systematic reviews in this field have largely identified a lack of adequate reporting in this field (Keogh et al., 2021; Lee et al., 2021) and have highlighted the lack of longitudinal data in clinical populations (De Angel et al., 2022). Our research goes some way to address these limitations in the existing literature.

Analytic complexities

There are several explanations for the small effect sizes identified in our analyses. Firstly, whilst we endeavoured to account for all relevant confounders in our analyses, there are unmeasured variables which may be critical in this context. Use of a novel health management system represents a complex and intricate relationship between personal, social, and technological factors, many of which we did not measure. Perceptions of self-efficacy, for example, have been found to be an important predictor of ratings of ease of use and motivation to use wearable devices (Rupp et al., 2018). Technological trust is a domain becoming increasingly important as new technologies are developed and integrated into daily life (Hancock et al., 2011). In this context, having faith that a wearable device accurately transforms biological data into accurate and understandable summaries is critical. There is also some evidence to suggest that different personality traits, such as extraversion and conscientiousness are associated with increased perceptions of device usability and motivation to engage with wearable technologies (Rupp et al., 2018). Future research in clinical populations would benefit from measuring these individual characteristics which represent diagnosis-agnostic individual differences which may be critical for understanding these relationships.

We also need to be mindful of the context in which these data have been collected. Participants enrolled in the RADAR-MDD study consented to take part in a longitudinal observational study with no health-related reward or benefit for doing so (Matcham et al., 2019). Participants were paid for enrolling in the study, participating in qualitative interviews, and for providing outcome data every three months, but were not financially incentivised for completing app-based questionnaires or for wearing the Fitbit. Participants received the same amount of money for participating regardless of how much they engaged with the technology, making financial reimbursement unlikely to influence our results. RADAR-MDD represents use of technology in a unique environment; highly motivated participants were

in close contact with the research team throughout the course of follow-up, with all concerns and complaints (technology- or health-related) managed and supported (Oetzmann et al., 2022). Close communication between researchers and participants is essential to maintain trust and collaboration throughout the course of follow-up (Garland, Plemmons & Koontz, 2006). This ongoing relationship with the research team may have mitigated the associations between perceptions of the system usability and actual system use; participants knew that the system was in development and that their participation would improve the functionality of the technology. If the system were implemented in the real-world, these associations may have been much more prominent; whilst our findings may be generalisable to other research protocols, they may not generalise to real-world remote data collection.

Study limitations

We also need to acknowledge the limitations in our measurement of engagement. Typical methods of measuring app-engagement involve mining data logs to determine the number of times an app has been opened, notifications received, notifications responded to, amount of time spent using an app, for example (Cao & Lin., 2017). Whilst we had this data available via Google analytics, we identified large discrepancies between the amount of number of notifications reported as having been received by the participants, and the number of questionnaires completed, with many more questionnaires completed than notifications sent. As a questionnaire was only actionable upon receipt of a notification, we questioned the accuracy of Google metrics and instead chose to focus on completion rates as an indicator of engagement. Therefore in lieu of “gold-standard” methods of quantifying engagement with remote sensing technologies, our measures of system use (Fitbit wear-time and questionnaire completions) should be viewed as proxy indicators of system engagement.

Furthermore, whilst the TAM-FF and PSSUQ are useful indicators of perceived usability and acceptability of the technology they pose limitations. Firstly, they ask about the whole system, rather than differentiating between the Fitbit and the questionnaires. Whilst participants could make this distinction themselves while they were answering the questions, most did not. This means that if there are discrepancies between how people feel about the Fitbit and how they feel about the app-based questionnaires, we are not able to determine these differences.

Finally, although we report associations between clinical characteristics and perceptions of system usability and engagement in the following 3-months, symptoms of depression, anxiety, functional disability, and perceptions of system usability are measured at the same time. These therefore represent cross-sectional associations rather than predictions of future perceptions.

Conclusions

We found some evidence to suggest that anxiety and functional disability may influence perceptions of system usability and system use in people with MDD, albeit with extremely small effect sizes. The RADAR-MDD reported high levels of engagement throughout the course of follow-up (Matcham et al., 2022) which appears to be largely robust across different levels of depression severity. These findings highlight several directions for future

research. More comprehensive assessments into personality traits, technology anxiety and normative beliefs may provide further insight into the driving factors of engagement. The standardisation of engagement metrics would allow future studies to implement outcome measurements which are more direct indicators of system engagement. Our next steps will be to examine the ability of the data collected via the remote sensing devices to predict outcomes such as relapse and deterioration in depressive state. Finally, our findings should provide some reassurance for researchers aiming to test the utility of remote sensing in people with depression – despite all the challenges the condition imposes, people with more severe depressive symptoms are still able to participate effectively.

References

- Ajzen, I., Fishbein, M., 1980. *Understanding Attitudes and Predicting Social Behavior*. Prentice Hall, New Jersey.
- Baki, R., Birgoren, B., Aktepe, A., 2018. A meta-analysis of factors affecting perceived usefulness and perceived ease of use in the adoption of e-learning systems. *Turkish Online Journal of Distance Education* 19, 4–42. <https://doi.org/10.17718/tojde.471649>
- Bakker, D., Rickard, N., 2019. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. *Australian Psychologist* 54, 245–260. <https://doi.org/10.1111/ap.12383>
- Batra, S., Baker, R.A., Wang, T., Forma, F., DiBiasi, F., Peters-Strickland, T., 2017. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. *Med Devices (Auckl)* 10, 237–251. <https://doi.org/10.2147/MDER.S144158>
- Beck, A.T., 1967. *Depression; causes and treatment*. Harper & Row, New York, USA.
- Brugha, T.S., Cragg, D. 1990. The List of Threatening Experiences: The reliability and validity of a brief life events questionnaire. *Acta Psychiatrica Scandinavica* 82, 77-81. <https://doi.org/10.1111/j.1600-0447.1990.tb01360.x>.
- Bürkner, P.-C., Vuorre, M., 2019. Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science* 2, 77–101. <https://doi.org/10.1177/2515245918823199>
- Cao, H., Lin M., 2017. Mining smartphone data for app usage prediction and recommendations: A survey. *Pervasive and Mobile Computing* 37, 1-22. <https://doi.org/10.1016/j.pmcj.2017.01.007>.
- Chin, W.W., Johnson, N., Schwarz, A., 2008. A fast form approach to measuring technology acceptance and other constructs. *MIS Quarterly* 32, 687–703. <https://doi.org/10.2307/25148867>
- Connolly, S.L., Hogan, T.P., Shimada, S.L., Miller, C.J., 2020. Leveraging implementation science to understand factors influencing sustained use of mental health apps: A narrative review. *Journal of Technology in Behavioral Science* 2020 6:2 6, 184–196. <https://doi.org/10.1007/S41347-020-00165-4>
- De Angel, V., Lewis, S., White, K., Oetzmann, C., Leightley, D., Oprea, E., Lavelle, G., Matcham, F., Pace, A., Mohr, D.C., Dobson, R., Hotopf, M., 2022. Digital health

- tools for the passive monitoring of depression: A systematic review of methods. *npj Digit. Med.* 5, 1–14. <https://doi.org/10.1038/s41746-021-00548-8>
- Dobkin, B.H., 2013. Wearable motion sensors to continuously measure real-world physical activities. *Current opinion in neurology* 26, 602. <https://doi.org/10.1097/WCO.0000000000000026>
- Elhai, J.D., Tiamiyu, M.F., Weeks, J.W., Levine, J.C., Picard, K.J., Hall, B.J., 2018. Depression and emotion regulation predict objective smartphone use measured over one week. *Personality and Individual Differences, Examining Personality and Individual Differences in Cyberspace* 133, 21–28. <https://doi.org/10.1016/j.paid.2017.04.051>
- Franzen, J., Brinkmann, K., 2016. Anhedonic symptoms of depression are linked to reduced motivation to obtain a reward. *Motiv Emot* 40, 300–308. <https://doi.org/10.1007/s11031-015-9529-3>
- Garland, A.F., Plemmons, D., Koontz, L., 2006. Research-practice partnership in mental health: Lessons from participants. *Adm Policy Ment Health* 33, 517–528. <https://doi.org/10.1007/s10488-006-0062-2>
- Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., Parasuraman, R., 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Hum Factors* 53, 517–527. <https://doi.org/10.1177/0018720811417254>
- Harrati, N., Bouchrika, I., Tari, A., Ladjailia, A., 2016. Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Computers in Human Behavior* 61, 463–471. <https://doi.org/10.1016/j.chb.2016.03.051>
- Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G., 2009. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 42,377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Huntley, C.D., Fisher, P.L., 2016. Examining the role of positive and negative metacognitive beliefs in depression. *Scandinavian Journal of Psychology* 57, 446–452. <https://doi.org/10.1111/sjop.12306>
- James, S.L., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., et al., 2018. 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. *The Lancet* 392, 1789– 1858. [https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)
- Keogh, A., Argent, R., Anderson, A., Caulfield, B., Johnston, W., 2021. Assessing the usability of wearable devices to measure gait and physical activity in chronic conditions: a systematic review. *J NeuroEngineering Rehabil* 18, 138. <https://doi.org/10.1186/s12984-021-00931-2>
- Kerner, C., Goodyear, V.A., 2017. The motivational impact of wearable healthy lifestyle technologies: A self-determination perspective on Fitbits with adolescents. *American Journal of Health Education* 48, 287–297. <https://doi.org/10.1080/19325037.2017.1343161>

- Kononova, A., Li, L., Kamp, K., Bowen, M., Rikard, R., Cotten, S., Peng, W., 2019. The use of wearable activity trackers among older adults: Focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. *JMIR Mhealth Uhealth* 7, e9832. <https://doi.org/10.2196/mhealth.9832>
- Laiou P, Kaliukhovich DA, Folarin AA, Ranjan Y, Rashid Z, Conde P, et al., 2022. The association between home stay and symptom severity in Major Depressive Disorder: Preliminary findings from a multicentre observational study using geolocation from smartphones. *JMIR Mhealth Uhealth* 10, e28095. <https://doi.org/10.2196/28095>.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1998. Emotion, motivation, and anxiety: brain mechanisms and psychophysiology. *Biological Psychiatry* 44, 1248–1263. [https://doi.org/10.1016/S0006-3223\(98\)00275-3](https://doi.org/10.1016/S0006-3223(98)00275-3)
- Lee, S., Kim, H., Park, M.J., Jeon, H.J., 2021. Current advances in wearable devices and their sensors in patients with depression. *Frontiers in Psychiatry* 12, 962. <https://doi.org/10.3389/fpsy.2021.672347>
- Lee, Y., Kozar, K.A., Larsen, K.R.T., 2003. The technology acceptance model: past, present, and future. *CAIS* 12. <https://doi.org/10.17705/1CAIS.01250>
- Lewis, J.R., 1992. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. *Proceedings of the Human Factors Society Annual Meeting* 36, 1259–1260. <https://doi.org/10.1177/154193129203601617>
- Lumley, T. 2004. Analysis of complex survey samples. *Journal of Statistical Software*, 9(8), 1-19.
- Luty, S.E., Joyce, P.R., Mulder, R.T., McKenzie, J.M., 2002. Social adjustment in depression: the impact of depression severity, personality, and clinic versus community sampling. *Journal of Affective Disorders* 70, 143–154. [https://doi.org/10.1016/S0165-0327\(01\)00331-7](https://doi.org/10.1016/S0165-0327(01)00331-7)
- Matcham, F., Barattieri di San Pietro, C., Bulgari, V., de Girolamo, G., Dobson, R., Eriksson, H., et al., 2019. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): A multi-centre prospective cohort study protocol. *BMC Psychiatry* 19, 72. <https://doi.org/10.1186/s12888-019-2049-z>
- Matcham, F., Leightley, D., Siddi, S., Lamers, F., White, K.M., Annas, P., et al., 2022. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): Recruitment, retention, and data availability in a longitudinal remote measurement study. *BMC Psychiatry*. <https://doi.org/10.21203/rs.3.rs-612374/v1>
- McGrady, E., Conger, S., Blanke, S., Landry, B.J.L., 2010. Emerging technologies in healthcare: Navigating risks, evaluating rewards. *Journal of Healthcare Management* 55, 353–364. <https://doi.org/10.1097/00115514-201009000-00011>
- Mohr, D.C., Kwasny, M.J., Meyerhoff, J., Graham, A.K., Lattie, E.G., 2021. The effect of depression and anxiety symptom severity on clinical outcomes and app use in digital mental health treatments: Meta-regression of three trials. *Behaviour Research and Therapy* 147, 103972. <https://doi.org/10.1016/j.brat.2021.103972>

- Mohr, D.C., Shilton, K., Hotopf, M., 2020. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. *npj Digital Medicine*. <https://doi.org/10.1038/s41746-020-0251-5>
- Mundt, J.C., Marks, I.M., Shear, M.K., Greist, J.H., 2002. The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *Br J Psychiatry* 180, 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- O'Brien, H.L., Toms, E.G., 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology* 59, 938–955. <https://doi.org/10.1002/asi.20801>
- Oetzmann, C., White, K.M., Ivan, A., Julie, J., Leightley, D., Lavelle, G., Hotopf, M., Matcham, F., n.d. A framework for recruiting into a remote measurement technologies (RMTs) study: Experiences from a major depressive disorder cohort. OSF Preprints. <https://doi.org/10.31219/osf.io/ns7dc>.
- Or, C.K.L., Karsh, B.-T., Severtson, D.J., Burke, L.J., Brown, R.L., Brennan, P.F., 2011. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. *Journal of the American Medical Informatics Association* 18, 51–59. <https://doi.org/10.1136/jamia.2010.007336>.
- Ospina, R., Ferrari, S.L.P., 2012. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis* 56, 1609–1623. <https://doi.org/10.1016/j.csda.2011.10.005>
- Patel, S., Akhtar, A., Malins, S., Wright, N., Rowley, E., Young, E., Sampson, S., Morriss, R., 2020. The acceptability and usability of digital health interventions for adults with depression, anxiety, and somatoform disorders: Qualitative systematic review and meta-synthesis. *Journal of Medical Internet Research* 22, e16228. <https://doi.org/10.2196/16228>.
- Pratab, A., Neto, E. C., Snyder, P., Stepnowsky, C., Elhadad, N., Grant D., Mohebbi, M. H., Mooney, S., Suver, C., Wilbanks, J., Mangravite, L., Heagerty, P. J., Arian, P., & Omberg, L. 2020. Indicators of retention in remote digital health studies: A cross-study evaluation of 100,000 participants. *Npj Digital Medicine* 3, 21. Doi: <https://doi.org/10.1038/s41746-020-0224-8>.
- R Core Team., 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rahimi, B., Nadri, H., Afshar, H.L., Timpka, T., 2018. A systematic review of the technology acceptance model in health informatics. *Appl Clin Inform* 09, 604–634. <https://doi.org/10.1055/s-0038-1668091>
- Ranjan, Y., Rashid, Z., Stewart, C., Conde, P., Begale, M., Verbeeck, D., Boettcher, S., Hyve, T., Dobson, R., Folarin, A., Consortium, T.R.-C., 2019. RADAR-Base: Open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. *JMIR mHealth and uHealth* 7, e11734. <https://doi.org/10.2196/11734>

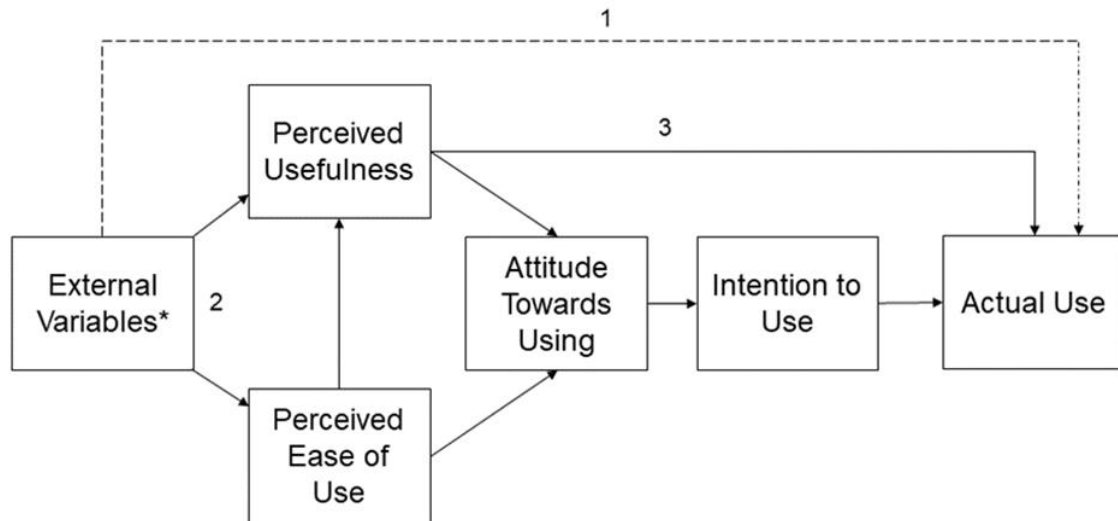
- Rohani, D.A., Faurholt-Jepsen, M., Kessing, L.V., Bardram, J.E., 2018. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. *JMIR Mhealth Uhealth* 6, e9691. <https://doi.org/10.2196/MHEALTH.9691>
- Rosa-Alcázar, A.I., Rosa-Alcázar, Á., Martínez-Esparza, I.C., Storch, E.A., Olivares-Olivares, P.J., 2021. Response inhibition, cognitive flexibility and working memory in obsessive-compulsive disorder, generalized anxiety disorder and social anxiety disorder. *International Journal of Environmental Research and Public Health* 18, 3642. <https://doi.org/10.3390/ijerph18073642>
- Rupp, M.A., Michaelis, J.R., McConnell, D.S., Smither, J.A., 2018. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. *Appl Ergon* 70, 77–87. <https://doi.org/10.1016/j.apergo.2018.02.005>
- Rush, A.J., Carmody, T., Reimitz, P.-E., 2000. The Inventory of Depressive Symptomatology (IDS): Clinician (IDS-C) and Self-Report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research* 9,45-59. <https://doi.org/10.1002/mpr.79>
- Saadé, R.G., Kira, D., 2007. Mediating the impact of technology usage on perceived ease of use by anxiety. *Computers & Education* 49, 1189–1204. <https://doi.org/10.1016/j.compedu.2006.01.009>
- Schmier, J., Halpern, M.T., 2004. Patient recall and recall bias of health state and health status. *Expert Review of Pharmacoeconomics and Outcomes Research* 4, 159–163. <https://doi.org/10.1586/14737167.4.2.159>
- Simblett, S., Evans, J., Greer, B., Curtis, H., Matcham, F., Radaelli, M., Mulero, P., Arévalo, M.J., Polhemus, A., Ferrao, J., Gamble, P., Comi, G., Wykes, T., 2019. Engaging across dimensions of diversity: A cross-national perspective on mHealth tools for managing relapsing remitting and progressive multiple sclerosis. *Multiple Sclerosis and Related Disorders* 32, 123–132. <https://doi.org/10.1016/j.msard.2019.04.020>
- Simblett, S., Greer, B., Matcham, F., Curtis, H., Polhemus, A., Ferrão, J., Gamble, P., Wykes, T., 2018. Barriers to and facilitators of engagement with remote measurement technology for managing health: Systematic review and content analysis of findings. *Journal of Medical Internet Research* 20, e10480. <https://doi.org/10.2196/10480>
- Simblett, S., Matcham, F., Siddi, S., Bulgari, V., Pietro, C.B. di S., López, J.H., Ferrão, J., Polhemus, A., Haro, J.M., Girolamo, G. de, Gamble, P., Eriksson, H., Hotopf, M., Wykes, T., Consortium, R.-C., 2019. Barriers to and facilitators of engagement with mhealth technology for remote measurement and management of depression: Qualitative analysis. *JMIR mHealth and uHealth* 7, e11325. <https://doi.org/10.2196/11325>
- Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med* 166, 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>

- Tsai, T.-H., Lin, W.-Y., Chang, Y.-S., Chang, P.-C., Lee, M.-Y., 2020. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLOS ONE 15, e0227270. <https://doi.org/10.1371/journal.pone.0227270>
- Van Til, K., McInnis, M.G., Cochran, A., 2020. A comparative study of engagement in mobile and wearable health monitoring for bipolar disorder. Bipolar Disord 22, 182–190. <https://doi.org/10.1111/bdi.12849>
- Verduijn, J., Verhoeven, J.E., Milaneschi, Y., Schoevers, R.A., van Hemert, A.M., Beekman, A.T.F., Penninx, B.W.J.H., 2017. Reconsidering the prognosis of major depressive disorder across diagnostic boundaries: full recovery is the exception rather than the rule. BMC Medicine 15, 215. <https://doi.org/10.1186/s12916-017-0972-8>
- White, K.M., Williamson, C., Bergou, N., Oetzmann, C., De Angel, V., Matcham, F., Henderson, C., Hotopf, M., n.d. Exploring the definition, measurement and reporting of engagement in studies of remote measurement technologies for physical and mental health symptom tracking: A systematic review. manuscript in preparation.
- Zhang, Y., Folarin, A. A., Sun, S., Cummins, N., Ranjan, Y., Rashid, Z., et al., 2021. Predicting depression symptom severity through individuals' nearby Bluetooth Device Count Data collected by mobile phones: Preliminary longitudinal study. JMIR Mhealth Uhealth 9, e29840. <https://doi.org/10.2196/29840>.

Predictors of long-term engagement with remote sensing technologies for symptom measurement in Major Depressive Disorder

Supplementary Materials

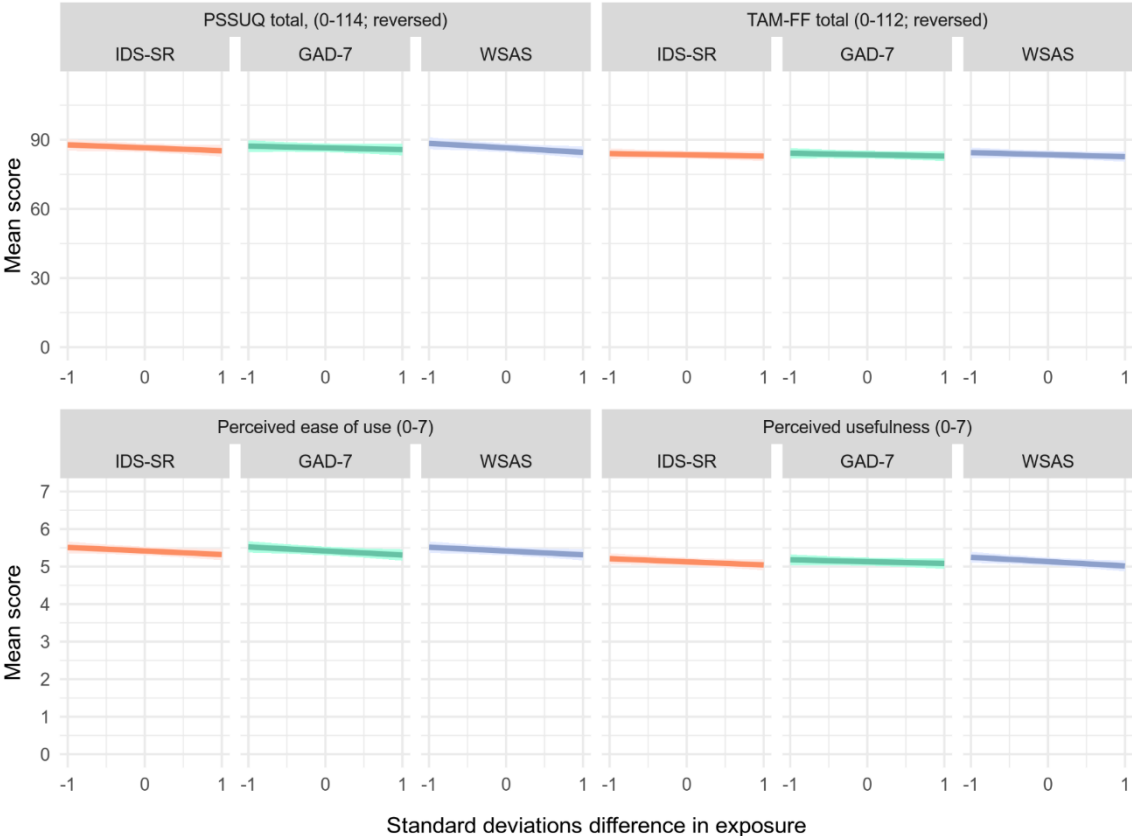
Supplementary Figure 1: Technology Acceptance Model and tested hypotheses.



Technology Acceptance Model, adapted from Rahimi et al., 2018. Hypothesis 1: Association between clinical presentation and actual system use. Hypothesis 2: Association between external variables (clinical presentation) and perceptions of system usability and acceptability. Hypothesis 3: Association between perception of usability/acceptability and actual system use. *Specifically, clinical symptoms such as depression, anxiety, and functional disability.

Supplementary Figure 2: Associations of clinical variables with usability and acceptability (n=547 participants)

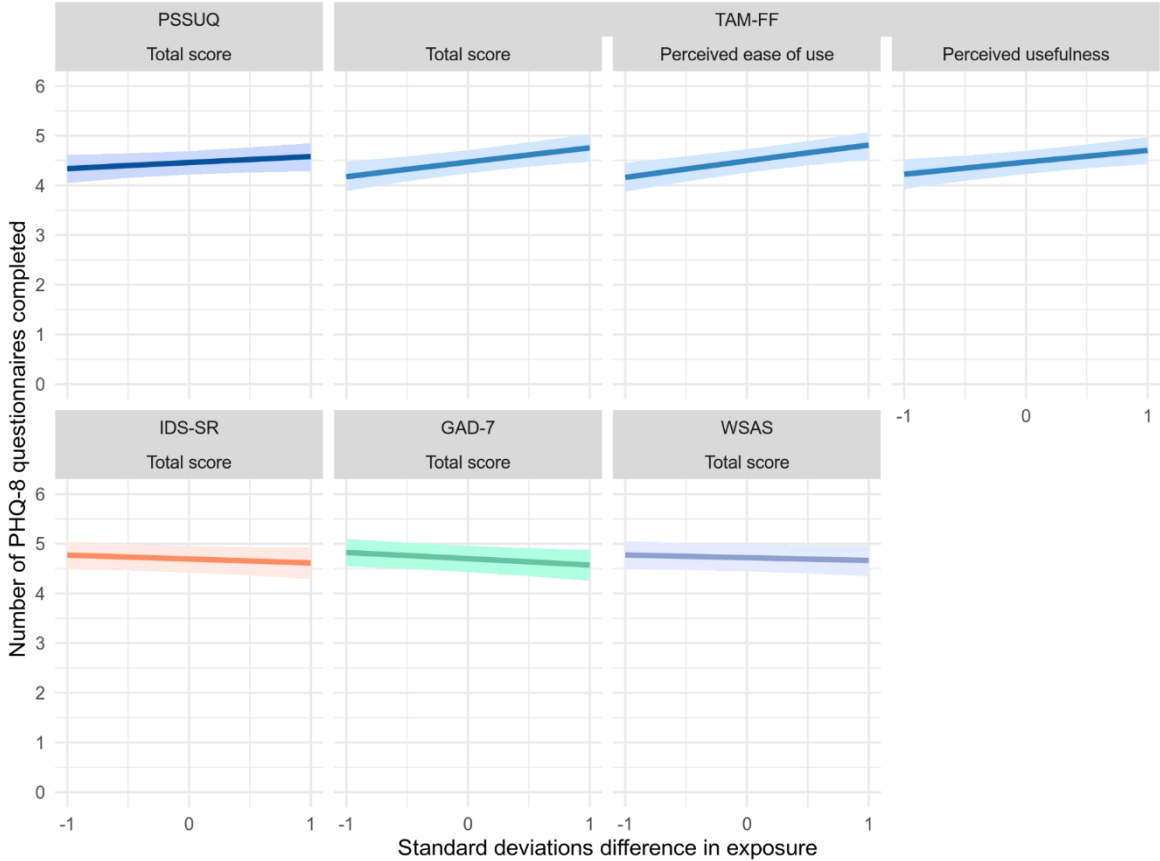
The estimated marginal means¹ for each questionnaire (y-axis) plotted against a \pm one standard deviation (SD) difference for each exposure (y-axis), holding other covariates to their median values.



¹ Estimated marginal means are computed from posterior draws of the expected value of the posterior predictive distributions. See [this page](#) for details.

Supplementary Figure 3: Effect of perceived usability and acceptability on the number of PHQ-8 questionnaires completed over the following three months (n=547 participants)

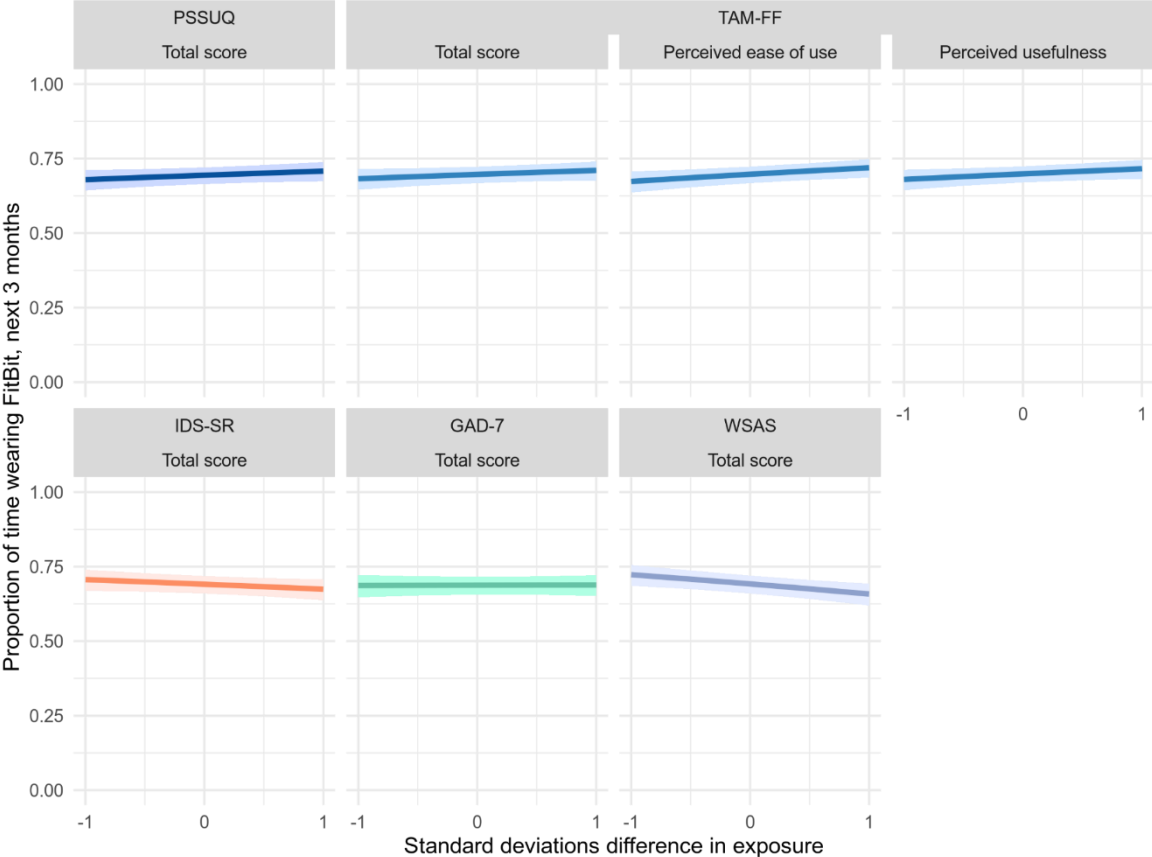
The estimated mean number of PHQ-8 questionnaires completed over the following three months (y-axis) plotted against a \pm one SD range of each exposure (y-axis),



holding other covariates at their median values.

Supplementary Figure 4: Effect of perceived usability and acceptability on the proportion of time wearing a FitBit device over the following three months (n=547 participants)

The estimated wear time proportion (y-axis) plotted against a \pm one SD range of each exposure (x-axis), holding other covariates at their median values.

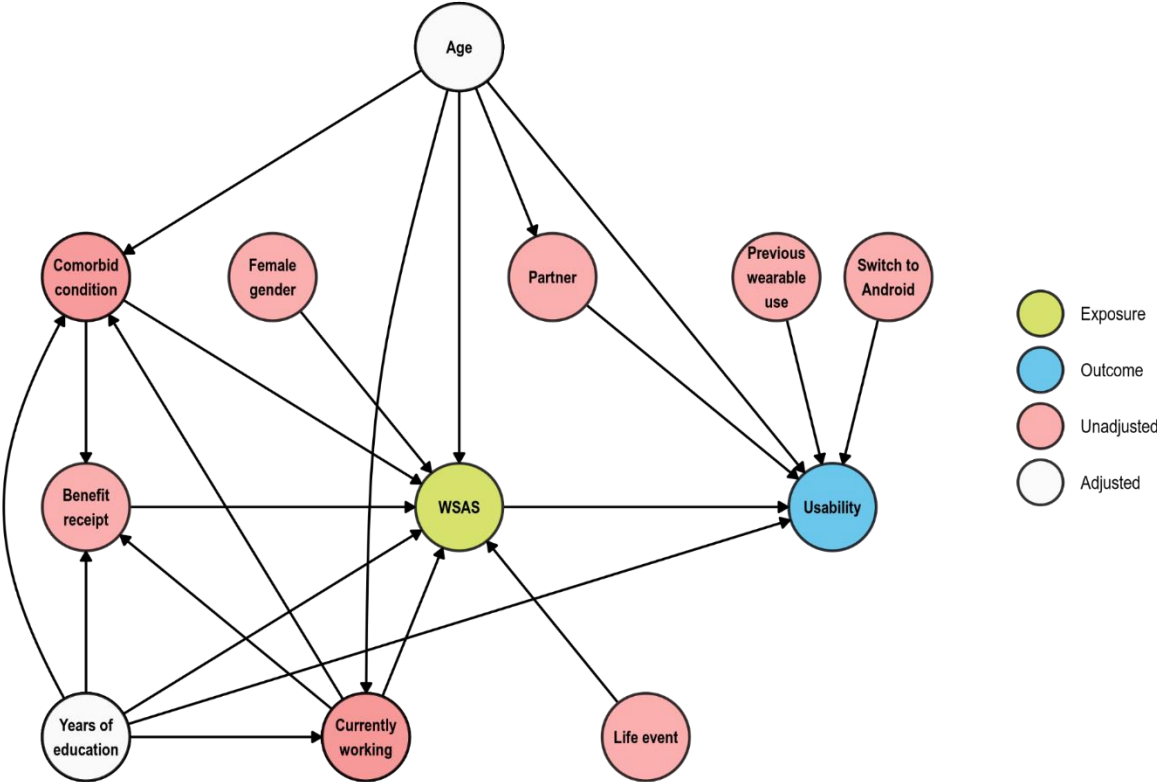


Directed acyclic graphs (DAGs)

The figures below present the DAGs used to inform the modelling of each pair of exposures (shown in yellow) and outcomes (shown in blue). Variables that were adjusted for in our models are shown in white.

H₁ Clinical variables and perceived usability and acceptability

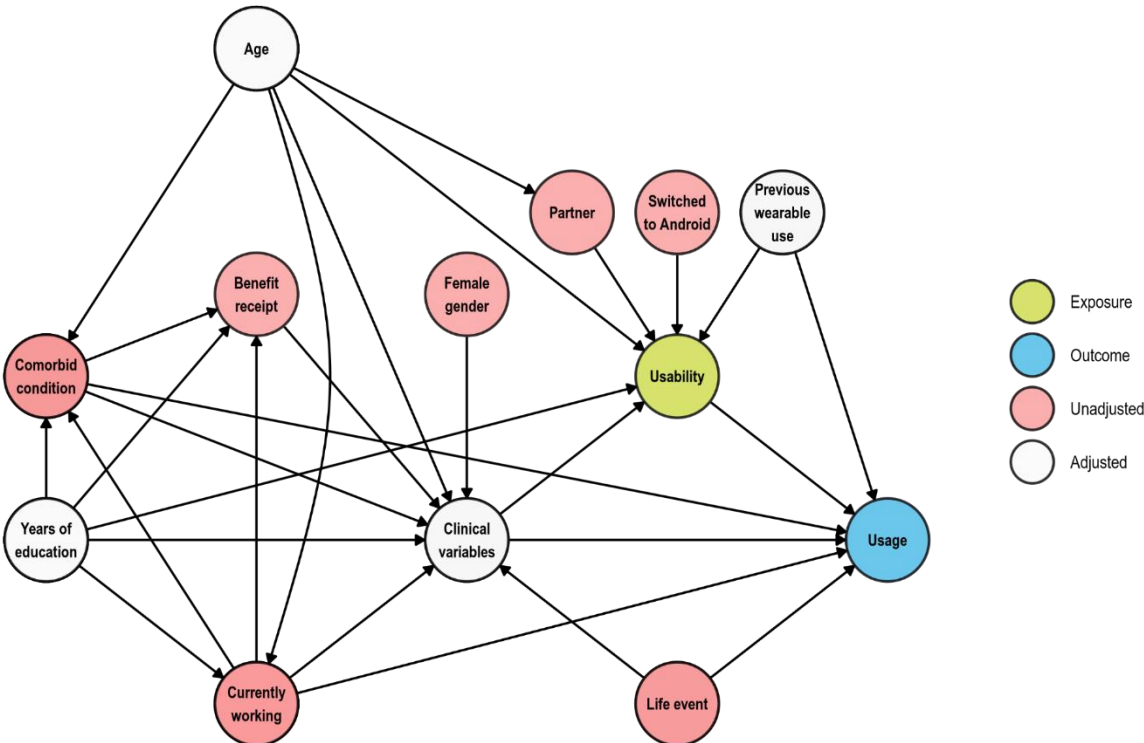
The adjustment set for this model was: **Age, years of education.**



H₂ Perceived usability and usage over the following three months

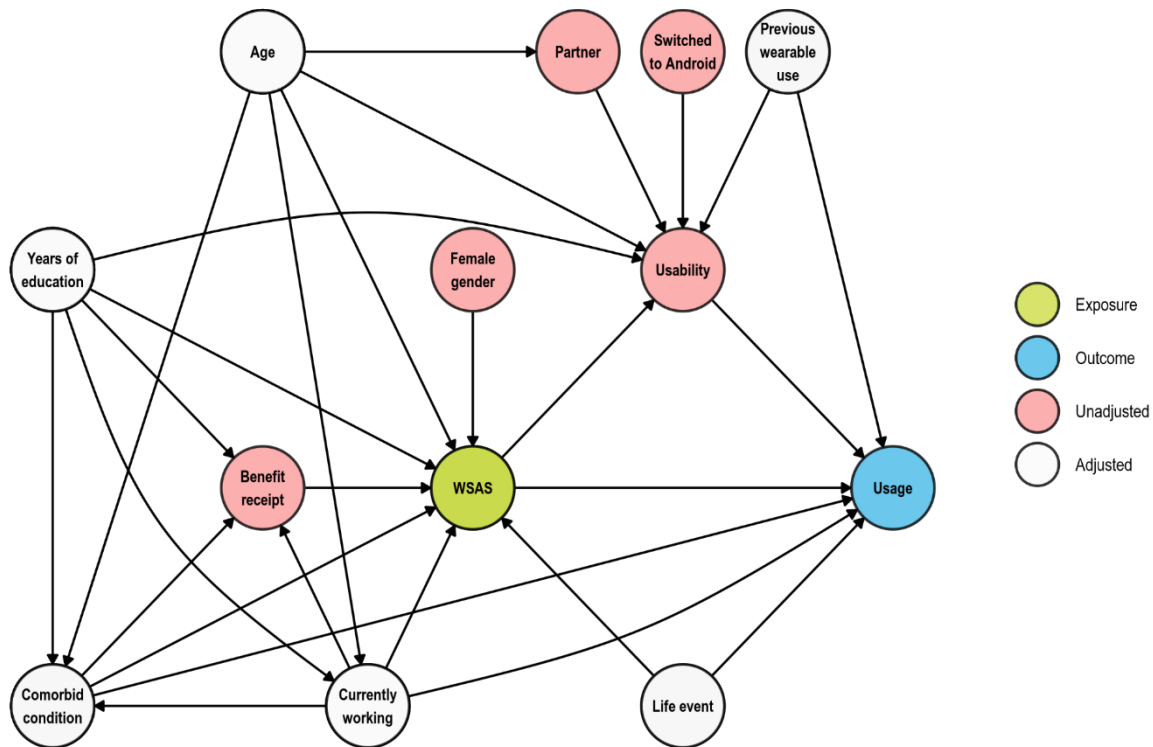
The adjustment set for this model was: **Age, the clinical variable*, previous wearable use, years of education.**

*either WSAS, IDS-SR, or GAD-7.



H₃ Clinical variables and usage over the following three months

The adjustment set for this model was: **Age, comorbid conditions, employment status, life event in the past three months, previous wearable use, years of education.**



Supplementary Table 1: Descriptive statistics for usage outcomes by sociodemographic variables

This table summarises the two usage outcomes (PHQ-8 questionnaires completed and FitBit wear time) by sociodemographic variables. The numbers represent the median outcome for each group from all responses collected during follow-up. The number in brackets represents the number of responses; each individual responded multiple times handled via random effect models.

	Number of PHQ-8 questionnaires completed, next 3 months	Proportion of time wearing FitBit, next 3 months
	Median [No. avail.]	Median [No. avail.]
Age group		

18-30	4 [218]	0.65 [249]
31-45	4 [209]	0.78 [238]
46-65	5 [464]	0.86 [505]
66+	6 [84]	0.94 [90]
Gender		
Male	4 [245]	0.79 [263]
Female	5 [730]	0.82 [819]
Number of comorbid conditions		
0	5 [494]	0.84 [550]
1	4 [168]	0.76 [190]
2	5 [313]	0.80 [342]
Years of education		
<10	6 [77]	0.62 [86]
10-20	5 [754]	0.82 [836]
>20	5 [144]	0.82 [160]
Employment status		
Working	5 [574]	0.82 [633]
Not working	4 [401]	0.80 [449]