

Characterising longitudinal patterns in cognition, mood and activity in depression with 6-week high-frequency wearable assessment: an observational study

Francesca Cormack, Margaret McCue, Caroline Skirrow, Nathan Cashdollar, Nick Taptiklis, Tempest van Schaik, Ben Fehnert, James King, L Chrones, Sara Sarkey, Jasmin Kroll, Jennifer Barnett

Submitted to: JMIR Mental Health
on: March 02, 2023

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
---------------------------------	----------

Preprint
JMIR Publications

Characterising longitudinal patterns in cognition, mood and activity in depression with 6-week high-frequency wearable assessment: an observational study

Francesca Cormack^{1, 2, 3} PhD; Margaret McCue⁴ MSc; Caroline Skirrow^{1, 5} PhD; Nathan Cashdollar¹ PhD; Nick Taptiklis¹ BA; Tempest van Schaik⁶ PhD; Ben Fehnert^{3, 6} BSc; James King^{3, 6} BSc; L Chrones⁴ BSc; Sara Sarkey⁴ MSc; Jasmin Kroll¹ PhD; Jennifer Barnett^{1, 2} PhD

¹Cambridge Cognition Cambridge GB

²Department of Psychiatry University of Cambridge Cambridge GB

³Cognition Kit Cambridge GB

⁴Takeda Pharmaceuticals USA Inc. Lexington US

⁵Department of Psychological Science University of Bristol Bristol GB

⁶Ctrl Group London GB

Corresponding Author:

Jasmin Kroll PhD

Cambridge Cognition

Tunbridge Court

Bottisham

Cambridge

GB

Abstract

Background: Cognitive symptoms are an underrecognized aspect of depression that are often untreated. High-frequency cognitive assessment holds promise for improving disease and treatment monitoring. Although we have previously found it feasible to remotely assess cognition and mood in this capacity further work is needed to ascertain the optimal methodology to implement and synthesize these techniques.

Objective: The objective of this study was to examine (1) longitudinal changes in mood, cognition, activity levels, and heart rate over 6 weeks; (2) diurnal and weekday-related changes; and (3) co-occurrence of fluctuations between mood, cognitive function, and activity.

Methods: Thirty adults with current mild-moderate depression, stabilized on antidepressant monotherapy, responded to testing delivered via an Apple Watch for six weeks. Outcome measures included cognitive function using the Cognition Kit N-Back, assessing three brief n-back tasks daily; self-reported depressed mood assessed once daily; daily total step count and average heart rate. Change over a six-week duration, diurnal and day-of-week variations, and covariation between outcome measures were examined using nonlinear and multilevel models.

Results: Participants showed initial improvement in the Cognition Kit N-Back performance followed by a learning plateau. Performance reached 90% of individual learning levels on average 10 days after study onset. N-back performance was typically better earlier and later in the day, and step counts were lower at the beginning and end of each week. Higher step counts overall were associated with faster n-back learning and increased daily step count was associated with better mood on the same ($p < 0.001$) and following day ($p = .02$). Daily n-back performance covaried with self-reported mood after participants reached their learning plateau ($p = 0.01$).

Conclusions: The current results support the feasibility and sensitivity of high-frequency cognitive assessments for disease and treatment monitoring in patients with depression. Methods to model the individual plateau in task learning can be used as a sensitive approach to better characterize changes in behaviour and improve clinical relevance of cognitive data. Wearable technology allows assessment of activity levels which may influence both cognition and mood.

(JMIR Preprints 02/03/2023:46895)

DOI: <https://doi.org/10.2196/preprints.46895>

Preprint Settings

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

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

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

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

Original Manuscript

Full title: Characterising longitudinal patterns in cognition, mood and activity in depression with 6-week high-frequency wearable assessment: an observational study

Short title:

Authors:

F Cormack^{1,2,3}, PhD; M McCue⁴, MSc; C Skirrow^{1,5}, PhD; N Cashdollar^{1,6}, PhD; N Taptiklis^{1,3}, BSc; T A van Schaik⁷, PhD; B Fehnert^{3,7}, J King^{3,7}, L Chrones⁴, S Sarkey⁴, J Kroll¹ PhD; JH Barnett^{1,2} PhD

Authors' names and affiliations:

¹Cambridge Cognition, Cambridge, United Kingdom; ²Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom; ³Cognition Kit, Cambridge, United Kingdom; ⁴Takeda Pharmaceuticals USA, Deerfield, IL, USA; ⁵Department of Psychological Science, University of Bristol; Bristol, United Kingdom; ⁶Cambridge Cognition, Cambridge, MA, USA; ⁷ Ctrl Group, London, United Kingdom

Corresponding Author's contact information:

Name: Francesca Cormack

Cambridge Cognition, Tunbridge Court, Bottisham, Cambridge, CB25 9TU, UK

email: Francesca.cormack@camcog.com

Telephone: +44 (0)1223 810700

Word counts:

Abstract: 325

Manuscript: 4686

Abstract

Background

Cognitive symptoms are an underrecognized aspect of depression that are often untreated. High-frequency cognitive assessment holds promise for improving disease and treatment monitoring. Although we have previously found it feasible to remotely assess cognition and mood in this capacity further work is needed to ascertain the optimal methodology to implement and synthesize these techniques.

Objective

The objective of this study was to examine (1) longitudinal changes in mood, cognition, activity levels, and heart rate over 6 weeks; (2) diurnal and weekday-related changes; and (3) co-occurrence of fluctuations between mood, cognitive function, and activity.

Methods

Thirty adults with current mild-moderate depression, stabilized on antidepressant monotherapy, responded to testing delivered via an Apple Watch for six weeks. Outcome measures included cognitive function, assessed with three brief n-back tasks daily; self-reported depressed mood assessed once daily; daily total step count and average heart rate. Change over a six-week duration, diurnal and day-of-week variations, and covariation between outcome measures were examined using nonlinear and multilevel models.

Results

Participants showed initial improvement in the Cognition Kit N-Back performance followed by a learning plateau. Performance reached 90% of individual learning levels on average 10 days after study onset. N-back performance was typically better earlier and later in the day, and step counts were lower at the beginning and end of each week. Higher step counts overall were associated with faster n-back learning and increased daily step count was associated with better mood on the same ($p < 0.001$) and following day ($p = .02$). Daily n-back performance covaried with self-reported mood after participants reached their learning plateau ($p = 0.01$).

Conclusions

The current results support the feasibility and sensitivity of high-frequency cognitive assessments for disease and treatment monitoring in patients with depression. Methods to model the individual plateau in task learning can be used as a sensitive approach to better characterize changes in

behaviour and improve clinical relevance of cognitive data. Wearable technology allows assessment of activity levels which may influence both cognition and mood.

Keywords (3-10): cognition, depression, digital biomarkers, mobile health, remote testing, ecological momentary assessment

Introduction

Background:

Major depressive disorder (MDD) is a debilitating condition and the leading cause of disease burden worldwide [1, 2]. MDD is characterized primarily by low mood, and/or reduced interest and pleasure in daily activities [3]. Cognitive deficits are a substantial problem in MDD patients with impairments in a range of domains reported including processing speed, attention, executive function, learning, and memory [4-6]. Despite impaired cognition being widely reported further studies using objective neuropsychological measures are required [7].

Self-Assessment of Major Depressive Disorder

MDD is typically assessed using retrospective self-report, where patients reflect on their experiences over a period of days or weeks. However, this method of reporting is subject to a variety of memory distortions [8, 9], and depression itself is linked to impaired recollection, memory bias and overgeneralization [10]. Direct comparison between retrospective recall and repeated real time assessments have shown negative emotional biases in depressed patients [11, 12] which persists beyond the depressive episode in a subset of patients [13]. Various explanatory factors have been proposed, such as the number of previous episodes and demographic factors, but overall results have been inconsistent [14].

If patients' recollections correspond poorly with their actual experiences this is likely to distort our understanding of disease course and treatment response [15]. Discrepancies between objectively measured cognitive function and patients' self-report have been demonstrated, with the latter being influenced by depressed mood [16-18]. This may be true even in cases of significant cognitive impairment where self-reported cognitive function is associated with subjective complaints of depressive symptoms but not objective cognitive outcomes [19]. Hence, it is plausible that self-reported and objective assessments are measuring different cognitive capabilities with objective measures needed to reveal underlying cognitive function.

Results from cognitive tests may also vary in relation to within-individual differences including dietary effects and sleep-wake cycle [20, 21]. These variances can make it difficult to differentiate clinical meaningful change from measurement error [22]. Higher frequency sampling is thought to generate more stable and reliable estimates of constructs of interest [22], by reducing state effects on punctual relationships which can obscure the signal in study interventions, and by narrowing the margin of error [23].

Real-Time Measurement of Cognition, Mood, and Activity

Advances in portable technology have enabled the precise, unobtrusive recording of real-time psychological, behavioural, and physiological measures. Higher frequency assessments in the context of everyday life can identify real life changes that are associated with clinical improvement [24], can improve sensitivity for detecting change [25], and help to identify shifts in depressive symptoms [26]. Additionally, this approach allows for the characterization of the temporal relationship of symptoms over time in relation to changes in an individual's behaviour and environmental influences [27, 28] creating a profile of the dynamic relationships between cognitive function, psychological and biological processes of an individual [29].

However, this approach also has important implications for sampling strategy and data analysis. Diurnal changes in mood and affect have been reported in individuals with depression where negative symptoms are more prevalent in the morning [30] and increased positive affect is seen later in the day [31, 32]. Further, research indicates cognitive function follows the same pattern as daily changes in mood, with worse performance in the morning and better performance in the evening on a range of cognitive measures, including memory, attention and psychomotor speed [33]. In the general population more positive and less negative moods are typically seen on weekends, with improvements starting on Fridays [34], whilst more activity, measured by higher step counts, is noted during weekdays in comparison with weekends [35, 36]. To account for the complexity of these fluctuations and how patients themselves may experience depression, novel data collection methods, such as ecological momentary assessments (EMA), are a promising opportunity to address this gap. Current scientific literature is mostly laboratory-based although EMA studies have shed light on various aspects of the disease such as rumination and emotion reactivity [37] however, less is known about the relationship between cognition function and mood in MDD.

Goal of This Study

Here we examine data from a six-week feasibility study of daily cognitive, mood, and activity assessment in adults with mild-to-moderate depression, stabilized on antidepressant treatment (<https://www.clinicaltrials.gov/ct2/show/NCT03067506>). Previously in this same sample, we demonstrated excellent compliance with daily cognitive and mood assessment on a smart watch, and good agreement with validated full length cognitive and self-report measures [38]. **High adherence rates were found (95%) with no deterioration over the course of the study. Adherence was not associated with depressive symptoms nor cognitive functioning.**

In the current study, we aim to optimize methods for analyzing high-frequency longitudinal data and characterize relationships between cognition, mood and activity data in order to facilitate future intervention studies. The current study aims to examine (1) longitudinal changes in mood, cognition, activity levels and heart rate over 6 weeks; (2) diurnal and weekday-related changes, and (3) co-occurrence of fluctuations between mood, cognitive function and activity.

Methods

This was an observational study aimed at characterising longitudinal patterns in cognition, mood, and activity in patients with major depressive disorder over six weeks with high-frequency assessment enabled by a wearable device.

Participants

Full recruitment details are provided in Cormack et al (2019) **which demonstrated the feasibility of high-frequency testing for a prolonged period of six weeks in participants with depression.** In brief, participants were recruited for a primary psychiatric diagnosis of MDD, with mild to moderate depression (as defined by PHQ-9 scores between 5 and 15, inclusive). They were recruited through a patient recruitment company with links to primary care providers and depression patient groups. Participants aged 18-65 years inclusive who were able to read and understand English were eligible for participation. **Individuals were excluded if they had a personal history of another psychiatric disorder (except nonprimary anxiety), mental, neurological or neurodegenerative disorders, or substance abuse or dependence. In total, thirty participants, with MDD, and receiving antidepressant monotherapy treatment, were enrolled.**

The study was reviewed and approved by the Proportionate Review Sub-Committee of the Wales Research Ethics Committee (REC) 6 at Swansea University and performed in accordance with the current version of the Declaration of Helsinki. All participants provided written informed consent

before being enrolled.

Measures

Cognitive function and mood were assessed with the Cognition Kit app, loaded onto the Apple Watch (high-resolution touch screen watch), and paired with an iPhone. Participants were asked to wear study equipment between 8 AM-10 PM for 6 weeks. **This duration was chosen as it is the time it typically takes for the improvement in mood to be seen following the start of administration of antidepressant medication.**

Prompts for cognitive assessment were given three times daily (morning, afternoon, and evening). Cognitive function was assessed with the Cognition Kit N-Back test, which has shown sensitivity to impairments in MDD [39]. Research suggests that the n-back task performance is a marker of cognitive function including aspects of working memory, task switching and attention [40]. During test administration, nine symbols randomly selected from a pool of 227 were presented briefly one at a time over 30 trials. Participants were asked to respond when any symbol was the same as the one presented 2 trials previously. The primary outcome measure was d-prime (the ratio of hits [correct detection of an n-back match] to false alarms [response during no match]). **Under the current implementation of the paradigm, values of d' range from -3.33 to 3.33. Test-retest reliability of high frequency testing using n-back d' is 0.8 in a mixed sample of participants with neurodegenerative disease, immune mediated inflammatory disorders and healthy controls [41]. Each assessment took 30 seconds to complete, after which participants were shown their test score.**

Prompts for mood assessment were given up to twice daily (afternoon and evening), with no prompt delivered in the evening if participants had completed the assessment in the afternoon. Mood was assessed with the following three questions adapted from the PHQ-2 and PDQ: How much have the following problems bothered you over the past day? (1) Lack of interest or pleasure in doing things; (2) Feeling down, depressed, or hopeless; and (3) Trouble concentrating on things (e.g., newspaper, TV). Responses were coded on a 4-point scale of severity of symptoms (1=no problem to 4=greatly). Responses were summated to provide a summary mood score. **The PHQ-2, a short validated form of the PHQ-9, has high accuracy and sensitivity to depression [42].**

Total daily step count and average heart rate, measures associated with depressive symptom severity, [43, 44] were acquired passively via the Apple Watch. **The Apple Watch has been found to be a reliable and valid tool for assessing heart rate variability and extremely accurate as a daily step count [45, 46]. Similarly the N-Back test has been previously used in wearable technology studies that**

incorporated high frequency testing demonstrating is feasibility [38, 47].

Procedure

Participants attended the study site where researchers introduced the devices and tasks to participants, who were given the opportunity to practice using the software and hardware and ask questions. Testing was completed in the following 6 weeks (42 days), with data uploading to secure study servers when transfer through Wi-Fi or roaming was possible. The study was completed with a home visit at 6 weeks, during which study hardware was returned.

Daily assessments were completed as a part of a larger test battery, including full-length rating scales and cognitive tests and semi-structured qualitative interview, described previously [38].

Statistical analysis

1) Data cleaning:

Data was harmonized by defining the first study day as the first day in which participants responded to assessment prompts for cognitive and mood assessment. Earlier days with only passive data were excluded from analyses. For activity and heart rate measures, non-wearing days (defined as days with <100 recorded steps [48, 49], or where no heart rate was recorded) were excluded from the analyses (25 observations out of 1,160).

Response frequency within each one-hour period collapsed across days (e.g. 6am-6.59am, 7am-7.59am, 8am-8.59am) was examined separately for mood and d-prime, to identify periods with sparse data. One-hour periods with $n < 25$ assessments overall were excluded. This left n-back data from 6am-11.59pm ($n = 29$ observations dropped), and mood data from 12 noon until 5.59pm ($n = 98$ observations dropped). For all included one-hour periods there were between 41-287 data points available (mean 182 mood, 180 n-back assessments per one-hour period).

2) Change over time

Since outcome measures are sampled repeatedly, hierarchical models are required to account for observations nested within individuals [50]. Random effects of participant with random intercept and random slope were applied to all models to control for between-participant variability, allowing the intercept and regression coefficient to vary between participants [51]. Model parameters were estimated using Maximum Likelihood, and model fits were compared with model likelihood ratios.

This allowed the direct comparison of models with different fixed-effect structures using chi-square analysis, and the selection of the best fitting model. Diagnostic tests included examination of normality of residuals, and their spread for each predictor variable.

Raw data were transformed using log and square root transformations into normally distributed data as appropriate. A series of longitudinal mixed-effects models examined change in each outcome measure over the six-week duration of the study to identify time-related trends in outcome measures. Each outcome measure was designated as the response variable and a fixed effect of time-on-task was specified. Intercept only models and linear, quadratic and cubic trends of time-on-task were examined. For unequally spaced time-points (cognitive assessment, mood), a continuous autoregressive correlation structure was applied. For summated daily indices (step-count, heart rate), a first-order autoregressive covariance structure was used.

3) Diurnal and weekday effects

Fixed effects of diurnal variation and/or weekday were examined by appending these as additional predictors (fixed effects) to the best fitting models. Time-of-day was treated as a continuous variable. Weekday was dummy coded from 1 to 7 (1=Monday to 7=Sunday). As in previous work, quadratic and linear fixed effects of time-of-day and day-of-week were examined [32]. For measures with variation both in time-of-day and day-of-week (d-prime, mood), models first examined the effects of time-of-day, and then incorporated weekday effects.

4) Characterising individual n-back learning curves

N-back learning curves were characterised following previously described methods [52]. This includes: (1) the “starting point”: the level at which performance begins; (2) the “asymptote”: the theoretically best score achievable towards which task performance tends with unlimited assessments; (3) the “slope”: the rate at which learning occurs; and (4) the “learning rate”: how quickly a pre-specified level of performance is reached.

Intercepts extracted for each participant from the best fitting mixed model for change over time, using the analysis steps described above, were implemented as individual starting points. These were subtracted from d-prime at each assessment occasion to provide a baseline adjusted d-prime. As in previous work, data were smoothed [53] by applying a three-assessment moving average to the data series, corresponding to the typical number of assessments per day.

For each individual participant, nonlinear regression was used to fit an inverse curve ($Y = a - (b/X)$), yielding personalised estimates for a (asymptote, the theoretically best score achievable: as $X \rightarrow \infty$, $Y \rightarrow a$), and b (slope) for baseline adjusted d-prime (Y) over each consecutive n-back assessment (X). 90% learning rate was defined as the number of trials required by each individual participant to reach 90% of potential beyond their starting point ($Y=0.9a$ when $X=10*b/a$) [52]. The stable maximum d-prime was calculated for participants by summing their asymptote and intercept.

Exploratory correlations of individual asymptotes and slopes were completed with summary measures from mood and activity assessments. Data were first examined for normality and parametric and non-parametric correlations were completed as appropriate.

5) Covariation of fluctuations with mood

Fluctuations over time were quantified from random effects residuals derived from the best fitting hierarchical models described above, regressing out significant effects of time-on-task (to control for practice effects), diurnal effects or weekday effects, as relevant to each outcome measure. Residuals reflect the deviation of each individual from their own slope at each moment in time (better or worse, higher or lower), quantifying the difference of each observation from what would be statistically expected. Since all measures with the exception of d-prime reflected overall daily measures, mean daily residuals for d-prime were computed by averaging d-prime residuals for each day, providing a consistent time-scale for covariation analyses. Hierarchical models were used to examine covariation between measures. All models specified a first-order autoregressive covariance structure, and random intercepts. Diagnostic tests included examination of normality of residuals, and their spread for each predictor variable.

To examine covariation of mood with d-prime and other assessment domains, a stepped approach was taken. First, a mixed model was used to predict fluctuations in mood by fluctuations within all other domains (fixed effects: d-prime, activity, heart rate). Next, direction of causality was examined using a time-lagged approach. Analyses examined whether the fluctuations in mood were predicted by fluctuations in the predictor variable on the previous or following day. Only outcome measures which were significant predictors in concurrent covariation analyses were taken forward and included in the lagged models.

Analyses were repeated after excluding the period during which 90% of n-back learning took place (computed from individual participant learning curves), with the exception of participants who showed no significant slope where all data was included. This helped to identify co-relation between

measures after the influences of task learning were minimised. Where significant associations with mean daily d-prime fluctuations were seen, analyses were repeated when constrained only to days in which three n-back assessments were available. This aimed to test whether covariation between d-prime and mood was affected by regression to the mean (where days with fewer n-back assessments are likely to show greater variability).

Results

Sample Characteristics:

The final sample (N=30) included 19 women and 11 men, aged between 19 and 63 (M=37.2, SD=10.4). All participants received antidepressant monotherapy and had been on their current medication for an average of 9.9 months (range 0.4-94.3, SD=9.5). Current medications included selective serotonin reuptake inhibitors (n=20), serotonin and norepinephrine reuptake inhibitors (n=5), tricyclic antidepressants (n=4) and serotonin antagonist and reuptake inhibitor (n=1). Mean depression symptom severity as measured by the PHQ-9 was 9.1 (range 5-15, SD=3.1).

Modelling changes in mood, cognition, activity levels and heart rate over six weeks

Model selection statistics and resultant model parameters are presented in Table 1. A cubic trend provided the best fit for change in cognitive performance over six weeks. Similarly, a linear trend provided the best fit for within-subjects change in mood over 6 weeks, in which depressive symptoms showed a subtle reduction over the course of the study. Heart rate and activity levels showed no overall change over time.

Data for d-prime and mood are presented in Figure 1. Here, individual daily scores (light grey lines) are shown alongside fitted random effects (bold grey) and the fitted fixed effect (orange). This illustrates the variability across participants in the direction and magnitude of change over the six weeks, and the remaining variability from these model fits.

Diurnal and weekday-related changes

D-prime showed a significant quadratic effect of time-of-day, with better performance seen first

thing in the morning and later at night (Figure 2A; time-of-day: Estimate=-0.02 (standard error (SE)=0.01), t-value=-2.33, p=0.02; time-of-day²: Estimate=0.001 (SE=0.0006), t-value=2.48, p=0.01). The model including time-of-day improved model fit (fit statistics: AIC=5952.15, BIC=6019.10, likelihood-ratio chi-squared test=6.16, p=0.01). No significant linear or quadratic effect of weekday was observed (p≥0.15).

Mood showed a subtle improvement over the course of the day, which did not reach significance thresholds and did not improve model fit (Figure 2B; time-of-day: Estimate=-0.01 (SE=0.006), p=0.06; Model fit statistics: AIC=436.22, BIC=476.20, Likelihood ratio=3.55, p=0.06). Effects of day-of-week were also not significant (p≥0.14).

Significant changes in activity and heart rate were also found over weekdays (Figure 2C-D). Step counts were lower at the week onset and end (weekday: Estimate=3.52 (SE=1.50), t-value=3.52, p=0.004; weekday²: Estimate=-0.64 (SE=0.17), t-value=-3.66, p=0.003). As expected, including these quadratic weekday effects improved model fit (fit statistics: AIC=10131.19, BIC=10166.41, Likelihood ratio=-5058.60, p=0.002). Similarly, heart rate showed a subtle quadratic effect of weekday with lower heart rates recorded on week onset and end (weekday: Estimate=0.01 (SE=0.006), t-value 1.96, p=0.05, weekday²: Estimate=-0.002 (SE=0.0008), t-value=-1.98, p=0.05), with improved model fit (AIC=-2312.81, BIC=-2272.56, Likelihood ratio=3.91, p=0.05).

Characterising the n-back learning curve

Adequate fit of the inverse learning curve model was seen for 27 out of 30 participants, in whom estimates for both slope and asymptote were significant (minimum t-value=2.18, p<0.04). In participants with adequate fit, 90% learning rates were reached after an average of 22.4 assessments (range 13-31 assessments), occurring after a mean of 10 days (range 6-24 days). Participants with data showing poor fit displayed a flatter trajectory and lower slope over the period of assessment, than those with a better fit (figure 3). Stable maximal d-prime was seen at a mean of 1.94 (range 0.74-2.93), with four individuals having maximums within the top quintile of positive n-back scores (at 2.59 and above), and no participants with 95% confidence intervals incorporating the highest possible score.

Three participants did not show a learning effect over the period of assessment, characterized by a non-significant slope. These participants performed overall above chance in n-back assessments

(mean scores between 0.9-1.7), indicating adequate understanding of test objectives but no clear learning pattern.

Learning slope and asymptote correlated moderately with one another ($\rho=0.70$ [95% confidence intervals: 0.39-0.84], $p<0.001$), but neither slope nor asymptote correlated with intercept (minimum $p=0.40$). Correlational analysis examining the relationship between learning parameters and mean mood and activity during the monitoring period revealed no significant association between learning slope or asymptote with mean mood ($\rho=-0.07$ -0.05, min $p=0.74$), but a significant association emerged with mean activity (slope: $\rho=0.44$ [0.05-0.75], $p=0.02$; asymptote: $\rho=0.45$ [0.08-0.68], $p=0.02$).

Co-occurrence of fluctuations between mood, cognitive function and activity

When examining the full assessment period, mood fluctuations did not significantly covary with fluctuations in daily heart rate ($t=1.2$, $p=0.23$), or d-prime ($t=-1.8$, $p=0.07$). However, daily mood fluctuations were seen with concurrent fluctuations in step count (Estimate=-0.001 (SE=0.0004), $t=-3.93$, $p=0.0001$), with relatively higher levels of activity associated with relatively better mood. Examining the direction of these effects using lagged models for the entire assessment period revealed that increased step count the day before was associated with better mood on the following day (prior day activity residual: Estimate=-0.001 (SE=0.0004), $t=-2.40$, $p=0.02$). However relatively more positive mood did not predict increased activity on the following day ($t=-0.49$, $p=0.62$).

Fluctuations between outcome variables and mood were re-examined after excluding the period during which 90% n-back learning rates were reached. Daily fluctuations in d-prime (Estimate=-0.06 (SE=0.02), $t=-2.51$, $p=0.01$) and activity (Estimate=-0.002 (SE=0.0005), $t=-3.50$, $p=0.0005$) were associated with fluctuations in mood (see figure 4 for individual examples of covariation of mood with d-prime). Covariation of mood with heart rate was non-significant ($t=1.74$, $p=0.08$). Estimates and significance levels remained similar after restricting analyses to days where all three n-back assessments were available (d-prime: Estimate=-0.07 (SE=0.03), $t=-2.57$, $p=0.01$; activity: Estimate=-0.002 (SE=0.0005), $t=-4.42$, $p<0.001$). Taking a lagged approach, but after excluding the learning period, increased step count was again associated with better mood on the next day (Estimate=-0.001 (SE=0.0005), $t=-2.00$, $p=0.05$). However relatively more positive mood did not predict increased step count on the following day ($t=-0.34$, $p=0.73$). This approach also indicated that relatively higher d-prime scores were not associated with better mood on the prior or following day (t range=0.61 to -0.77, p -range=0.44-0.72).

Discussion

Principal results and comparison with prior work

The current study characterises high-frequency assessments of mood, cognition and activity levels over a six-week period in patients with major depressive disorder (MDD). By modelling individual learning curves from cognitive testing, we show that after excluding an initial learning period, Cognition Kit N-Back test performance shows sensitivity to daily fluctuations in mood. The n-back test showed rapid early performance improvement followed by more incremental learning as participants near a stable performance level, a trend that is replicated from previous findings for mean daily data in this sample [38]. This pattern is similar to that seen in lab-based cognitive assessments [54, 55] and in practice effects identified during higher-frequency mobile cognitive assessments completed over one week of testing [22, 56]. Participants completed 90% of task learning after a mean of 22 assessments (~10 days after study onset). All participants achieved this level of improvement after 31 assessments (23 days). As such, a 10-day run-in before the introduction of any intervention may help to reduce the majority of learning effects in this n-back task and subsequent performance can be referenced relative to an individual's learning plateau as a proxy for baseline. Careful consideration of learning effects may be particularly important in intervention studies examining temporal associations between symptoms and accurate digital phenotyping [57]. As such, after removal of learning effects, the association between mood and n-back performance can be objectively measured with a brief three-times daily assessment. After excluding the initial learning period from the analysis, we found that better mood was associated with higher n-back scores on measurements taken on the same day. This approach demonstrates a potential method to 'baseline' individual performance in high-frequency assessments after their learning plateau is reached in order to disentangle clinically meaningful relationships between mood and cognition. The absence of a significant relationship between mood and n-back performance during the learning period could be attributable to the development of specific learning strategies applied by participants during the learning phase [58, 59], which obscure the relationship between mood and cognition.

Indeed, experimental studies have shown better cognitive performance when positive mood is induced [60-62], although findings are not always consistent [63]. This may be due to Nusbaum et al, (2018) examining cognitive flexibility as opposed to working memory which may have distinct underlying mechanisms contributing towards overall cognitive performance [64]. Working memory

impairment in patients with depression, have been widely observed and are associated with the number of hospitalisations and with overall prognosis [65, 66]. Performance in this domain is associated with negative symptom severity including lack of motivation and apathy [67], in addition to positive valence [68]. In the current study,. Changes may represent longer-term improvements in response to existing treatment regimens, or may reflect the natural history of remission and fluctuating symptom levels commonly reported in depression [69]. It has been suggested that enhanced patient understanding of mood and mood changes over time may help to improve depressive symptoms and their management [70, 71]. Whether these methods themselves have an impact on mood over time requires further clarification [72].

When examining diurnal and weekday effects, participants showed modestly better cognitive test performance early in the morning and late at night with slight decrease in function throughout the middle parts of the day. These findings are in agreement with the reported ‘afternoon slump’ in cognition in healthy adults [73]. Neither diurnal nor day-of-week effects on mood were seen in contrast to previous reports [31, 34]. However, in broad agreement with previous research [35, 36], we identified day-of-week effects for activity levels as measured by step count and similar patterns in mean heart rate. Step counts showed a quadratic effect of day of week, with lower counts registered at the beginning and end of each week. These findings highlight the importance of carefully considering the timing of assessments to ensure consistency of data sampling and controlling for trends in data over time. Longer term trends in mood data independent of any treatment effects highlight the importance of adequately controlling for subtle overall trends in time in future interventional studies, ideally using a randomised design and a placebo-controlled comparison group. Using a multilevel modelling approach, we were able to examine whether mood fluctuated synchronously with other measures of interest over time. We identified covariations in mood and activity levels, in keeping with a body of research showing improvements in depression symptoms with exercise [74, 75], or simple walking interventions [76]. Our results show that relatively increased mood was associated with relatively increased step count on the previous day (but not the following day), which indicated that the beneficial effects of exercise on wellbeing may well be protracted [77].

Limitations

Whilst associations between activity levels (step count), mood and cognition were identified, step counts are likely to have been influenced by wearing patterns [38], thereby reducing reliability of the

data. The summarised daily step counts also fail to elucidate whether activity was acute or prolonged, vigorous or light, and the exact timing of activity changes in relation to mood and cognitive assessments. Accelerometers can help to continuously monitor activity, and concurrent GPS and travel diary information has been found to help to classify and identify duration of walking [78]. Previous studies have examined associations between physical activity on mood in assessments triggered through GPS distance tracking [79, 80], a method which could be used to refine timing and accuracy of activity assessment in this context. Additionally, since this current study focused on patients with mild-to-moderate MDD who volunteered for participation, it is unclear whether results would generalize to patients with different clinical severity.

Conclusions

The current study indicates the importance of incorporating objective measures of cognitive testing and provides insight into fluctuations in mood and cognition in patients with MDD. The feasibility of remote high-frequency testing in MDD are promising for future research in this field and have important implications for clinical interventions. Whilst these methods can be used to monitor nuanced fluctuations between mood and cognition in a real-life setting, they may also be useful as a treatment tool. Understanding objective cognitive function in depression may also be used to target patients' acceptance of their objective difficulties in cognition and may be particularly relevant for interventions such as CBT and mindfulness [81]. Overall, while assessments of activity need to be further refined and improved, effects of step count on mood and cognition support the concurrent capture of data on activity, which may be an important contributor to variations in mood and changes in cognition in everyday life in patients with depression.

1. Ferrari, A.J., et al., *Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010*. PLoS Med, 2013. **10**(11): p. e1001547.
2. Moussavi, S., et al., *Depression, chronic diseases, and decrements in health: results from the World Health Surveys*. Lancet, 2007. **370**(9590): p. 851-8.
3. *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed.* Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed. 2013, Arlington, VA, US: American Psychiatric Publishing, Inc. xlv, 947-xlv, 947.

4. Hammar, Å. and G. Årdal, *Cognitive functioning in major depression - a summary*. Frontiers in Human Neuroscience, 2009. **3**.
5. McClintock, S.M., et al., *Association between depression severity and neurocognitive function in major depressive disorder: a review and synthesis*. Neuropsychology, 2010. **24**(1): p. 9-34.
6. Ahern, E. and M. Semkovska, *Cognitive functioning in the first-episode of major depressive disorder: A systematic review and meta-analysis*. Neuropsychology, 2017. **31**(1): p. 52-72.
7. Kaser, M., R. Zaman, and B.J. Sahakian, *Cognition as a treatment target in depression*. Psychol Med, 2017. **47**(6): p. 987-989.
8. Robinson, M.D. and G.L. Clore, *Belief and feeling: evidence for an accessibility model of emotional self-report*. Psychol Bull, 2002. **128**(6): p. 934-60.
9. Ebner-Priemer, U.W. and T.J. Trull, *Ambulatory assessment: An innovative and promising approach for clinical psychology*. European Psychologist, 2009. **14**(2): p. 109-119.
10. Dillon, D.G. and D.A. Pizzagalli, *Mechanisms of Memory Disruption in Depression*. Trends Neurosci, 2018. **41**(3): p. 137-149.
11. Ben-Zeev, D. and M.A. Young, *Accuracy of hospitalized depressed patients' and healthy controls' retrospective symptom reports: an experience sampling study*. J Nerv Ment Dis, 2010. **198**(4): p. 280-5.
12. Ben-Zeev, D., M.A. Young, and J.W. Madsen, *Retrospective recall of affect in clinically depressed individuals and controls*. Cognition and Emotion, 2009. **23**(5): p. 1021-1040.
13. Everaert, J., et al., *A meta-analytic review of the relationship between explicit memory bias and depression: Depression features an explicit memory bias that persists beyond a depressive episode*. Psychological Bulletin, 2022. **148**: p. 435-463.
14. Semkovska, M., et al., *Cognitive function following a major depressive episode: a systematic review and meta-analysis*. The Lancet Psychiatry, 2019. **6**(10): p. 851-861.
15. Bos, F.M., R.A. Schoevers, and M. aan het Rot, *Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review*. Eur Neuropsychopharmacol, 2015. **25**(11): p. 1853-64.
16. Popkin, S.J., et al., *Memory complaint and performance in normal and depressed older adults*. Exp Aging Res, 1982. **8**(3-4): p. 141-5.
17. Antikainen, R., et al., *Mood improvement reduces memory complaints in depressed patients*. Eur Arch Psychiatry Clin Neurosci, 2001. **251**(1): p. 6-11.
18. Naismith, S.L., et al., *Disability in major depression related to self-rated and objectively-measured cognitive deficits: a preliminary study*. BMC Psychiatry, 2007. **7**: p. 32.
19. Zlatar, Z.Z., et al., *Subjective Cognitive Decline Correlates With Depression Symptoms and Not With Concurrent Objective Cognition in a Clinic-Based Sample of Older Adults*. The Journals of Gerontology: Series B, 2017. **73**(7): p. 1198-1202.
20. Smith, A.P., R. Clark, and J. Gallagher, *Breakfast cereal and caffeinated coffee: effects on working memory, attention, mood, and cardiovascular function*. Physiol Behav, 1999. **67**(1): p. 9-17.
21. Wright, K., C. Lowry, and M. LeBourgeois, *Circadian and wakefulness-sleep modulation of cognition in humans*. Frontiers in Molecular Neuroscience, 2012. **5**.
22. Schweitzer, P., et al., *Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline*. Int J Methods Psychiatr Res, 2017. **26**(3).
23. Moore, R.C., et al., *Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults*. J Psychiatr Res, 2016. **75**: p. 116-23.

24. Barge-Schaapveld, D.Q., et al., *Changes in daily life experience associated with clinical improvement in depression*. J Affect Disord, 1995. **34**(2): p. 139-54.
25. Lenderking, W.R., et al., *Daily process methodology for measuring earlier antidepressant response*. Contemp Clin Trials, 2008. **29**(6): p. 867-77.
26. van de Leemput, I.A., et al., *Critical slowing down as early warning for the onset and termination of depression*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(1): p. 87-92.
27. Trull, T.J. and U. Ebner-Priemer, *Ambulatory assessment*. Annu Rev Clin Psychol, 2013. **9**: p. 151-76.
28. Marzano, L., et al., *The application of mHealth to mental health: opportunities and challenges*. Lancet Psychiatry, 2015. **2**(10): p. 942-8.
29. Sliwinski, M.J., et al., *Reliability and Validity of Ambulatory Cognitive Assessments*. Assessment, 2018. **25**(1): p. 14-30.
30. Leibenluft, E., B.M. Noonan, and T.A. Wehr, *Diurnal variation: reliability of measurement and relationship to typical and atypical symptoms of depression*. J Affect Disord, 1992. **26**(3): p. 199-204.
31. Peeters, F., et al., *Diurnal mood variation in major depressive disorder*. Emotion, 2006. **6**(3): p. 383-91.
32. Crowe, E., et al., *The intra-day dynamics of affect, self-esteem, tiredness, and suicidality in Major Depression*. Psychiatry Res, 2019. **279**: p. 98-108.
33. Porterfield, T., et al., *Neuropsychological function and diurnal variation in depression*. J Clin Exp Neuropsychol, 1997. **19**(6): p. 906-13.
34. Stone, A.A., S. Schneider, and J.K. Harter, *Day-of-week mood patterns in the United States: On the existence of 'Blue Monday', 'Thank God it's Friday' and weekend effects*. The Journal of Positive Psychology, 2012. **7**(4): p. 306-314.
35. Hirvensalo, M., et al., *Daily steps among Finnish adults: variation by age, sex, and socioeconomic position*. Scand J Public Health, 2011. **39**(7): p. 669-77.
36. Ludwig, V.M., et al., *Association between depressive symptoms and objectively measured daily step count in individuals at high risk of cardiovascular disease in South London, UK: a cross-sectional study*. BMJ Open, 2018. **8**(4): p. e020942.
37. Colombo, D., et al., *Current State and Future Directions of Technology-Based Ecological Momentary Assessment and Intervention for Major Depressive Disorder: A Systematic Review*. J Clin Med, 2019. **8**(4).
38. Cormack, F., et al., *Wearable Technology for High-Frequency Cognitive and Mood Assessment in Major Depressive Disorder: Longitudinal Observational Study*. JMIR Ment Health, 2019. **6**(11): p. e12814.
39. Snyder, H.R., *Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review*. Psychol Bull, 2013. **139**(1): p. 81-132.
40. Espeland, M.A., et al., *Performance of a computer-based assessment of cognitive function measures in two cohorts of seniors*. Int J Geriatr Psychiatry, 2013. **28**(12): p. 1239-50.
41. Cormack F, T.V., Taptiklis N, Kudelka J, Emmert K, Maetzler W., *App-based cognitive assessment and monitoring: a feasibility study in patients with immune-mediated inflammatory and neurodegenerative disorders*. Neurosci Appl, 2022. **1**.
42. Levis, B., et al., *Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-analysis*. JAMA, 2020. **323**(22): p. 2290-2300.

43. Schiweck, C., et al., *Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review.* Psychological Medicine, 2019. **49**(2): p. 200-211.
44. Ramsey, C.M., et al., *Daily steps and depressive symptoms: A longitudinal evaluation of patients with major depressive disorder in the precision medicine in mental health care study.* Journal of Affective Disorders, 2022. **300**: p. 334-340.
45. Veerabhadrapa, P., et al., *Tracking Steps on Apple Watch at Different Walking Speeds.* J Gen Intern Med, 2018. **33**(6): p. 795-796.
46. Fuller, D., et al., *Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review.* JMIR Mhealth Uhealth, 2020. **8**(9): p. e18694.
47. Brose, A., et al., *Daily variability in working memory is coupled with negative affect: The role of attention and motivation.* Emotion, 2012. **12**(3): p. 605-617.
48. Richardson, C.R., et al., *An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial.* J Med Internet Res, 2010. **12**(4): p. e71.
49. Moy, M.L., et al., *A pilot study of an Internet walking program and pedometer in COPD.* Respir Med, 2012. **106**(9): p. 1342-50.
50. Bringmann, L.F., et al., *A network approach to psychopathology: new insights into clinical longitudinal data.* PLoS One, 2013. **8**(4): p. e60188.
51. Verhagen, S.J., et al., *Use of the experience sampling method in the context of clinical trials.* Evid Based Ment Health, 2016. **19**(3): p. 86-9.
52. Feldman, L.S., et al., *A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task: defining "learning plateau" and "learning rate".* Surgery, 2009. **146**(2): p. 381-6.
53. Yamashita, M., M. Kawato, and H. Imamizu, *Predicting learning plateau of working memory from whole-brain intrinsic network connectivity patterns.* Sci Rep, 2015. **5**: p. 7622.
54. Collie, A., et al., *The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals.* J Int Neuropsychol Soc, 2003. **9**(3): p. 419-28.
55. Yamashita, M., M. Kawato, and H. Imamizu, *Predicting learning plateau of working memory from whole-brain intrinsic network connectivity patterns.* Scientific Reports, 2015. **5**(1): p. 7622.
56. Dupuy, M., et al., *Mobile cognitive testing in patients with schizophrenia: A controlled study of feasibility and validity.* Journal de Thérapie Comportementale et Cognitive, 2018. **28**(4): p. 204-213.
57. Griffin, B. and K.E.A. Saunders, *Smartphones and Wearables as a Method for Understanding Symptom Mechanisms.* Front Psychiatry, 2019. **10**: p. 949.
58. Gathercole, S.E., et al., *Working memory training involves learning new skills.* J Mem Lang, 2016. **105**: p. 19-42.
59. Tenison, C. and J.R. Anderson, *Modeling the distinct phases of skill acquisition.* J Exp Psychol Learn Mem Cogn, 2016. **42**(5): p. 749-67.
60. Nadler, R.T., R. Rabi, and J.P. Minda, *Better mood and better performance. Learning rule-described categories is enhanced by positive mood.* Psychol Sci, 2010. **21**(12): p. 1770-6.
61. Storbeck, J. and R. Maswood, *Happiness increases verbal and spatial working memory capacity where sadness does not: Emotion, working memory and executive control.* Cogn Emot, 2016. **30**(5): p. 925-38.

62. Yang, H., S. Yang, and A.M. Isen, *Positive affect improves working memory: implications for controlled cognitive processing*. Cogn Emot, 2013. **27**(3): p. 474-82.
63. Nusbaum, A.T., et al., *Induced Positive Mood and Cognitive Flexibility: Evidence from Task Switching and Reversal Learning*. Collabra: Psychology, 2018. **4**(1).
64. Periañez, J.A., et al., *Construct Validity of the Stroop Color-Word Test: Influence of Speed of Visual Search, Verbal Fluency, Working Memory, Cognitive Flexibility, and Conflict Monitoring*. Archives of Clinical Neuropsychology, 2020. **36**(1): p. 99-111.
65. Harvey, P.O., et al., *Executive functions and updating of the contents of working memory in unipolar depression*. J Psychiatr Res, 2004. **38**(6): p. 567-76.
66. Nikolin, S., et al., *An investigation of working memory deficits in depression using the n-back task: A systematic review and meta-analysis*. Journal of Affective Disorders, 2021. **284**: p. 1-8.
67. Zhu, Y., et al., *The Relationship Between Cognitive Dysfunction and Symptom Dimensions Across Schizophrenia, Bipolar Disorder, and Major Depressive Disorder*. Front Psychiatry, 2019. **10**: p. 253.
68. Storbeck, J. and R. Maswood, *Happiness increases verbal and spatial working memory capacity where sadness does not: Emotion, working memory and executive control*. Cogn Emot, 2016. **30**(5): p. 925-38.
69. Andrews, G., *Placebo response in depression: bane of research, boon to therapy*. Br J Psychiatry, 2001. **178**: p. 192-4.
70. Kauer, S.D., et al., *Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial*. J Med Internet Res, 2012. **14**(3): p. e67.
71. Groot, P.C., *Patients can diagnose too: How continuous self-assessment aids diagnosis of, and recovery from, depression*. J Ment Health, 2010. **19**(4): p. 352-62.
72. Hartmann, J.A., et al., *Experience sampling-based personalized feedback and positive affect: a randomized controlled trial in depressed patients*. PLoS One, 2015. **10**(6): p. e0128095.
73. Wertz, A.T., et al., *Effects of sleep inertia on cognition*. JAMA, 2006. **295**(2): p. 163-4.
74. Schuch, F.B., et al., *Exercise as a treatment for depression: A meta-analysis adjusting for publication bias*. J Psychiatr Res, 2016. **77**: p. 42-51.
75. Schuch, F.B., et al., *Neurobiological effects of exercise on major depressive disorder: A systematic review*. Neurosci Biobehav Rev, 2016. **61**: p. 1-11.
76. Robertson, R., et al., *Walking for depression or depressive symptoms: A systematic review and meta-analysis*. Mental Health and Physical Activity, 2012. **5**(1): p. 66-75.
77. Daley, A. and A. Welch, *The effects of 15 min and 30 min of exercise on affective responses both during and after exercise*. Journal of sports sciences, 2004. **22**: p. 621-8.
78. Kang, B., et al., *Walking objectively measured: classifying accelerometer data with GPS and travel diaries*. Med Sci Sports Exerc, 2013. **45**(7): p. 1419-28.
79. Reichert, M., et al., *Exercise versus Nonexercise Activity: E-diaries Unravel Distinct Effects on Mood*. Med Sci Sports Exerc, 2017. **49**(4): p. 763-773.
80. Reichert, M., et al., *Within-Subject Associations between Mood Dimensions and Non-exercise Activity: An Ambulatory Assessment Approach Using Repeated Real-Time and Objective Data*. Front Psychol, 2016. **7**: p. 918.
81. Serra-Blasco, M., et al., *Discrepancy between objective and subjective cognition in major depressive disorder*. European Neuropsychopharmacology, 2019. **29**(1): p. 46-56.