Predicting changes in mental health with temporal Ecological Momentary Assessment features: a longitudinal study on the predictive accuracy and optimal time windows of (measuring) daily affective states

Lea Berkemeier, Wim Kamphuis, Hilbrand Oldenhuis, Rudolf Verdaasdonk, Lisette van Gemert-Pijnen

Submitted to: JMIR Mental Health
on: September 21, 2025

Table of Contents

Predicting changes in mental health with temporal Ecological Momentary Assessment features: a longitudinal study on the predictive accuracy and optimal time windows of (measuring) daily affective states

Lea Berkemeier[1, 2, 3*] MSc; Wim Kamphuis[2*] PhD; Hilbrand Oldenhuis[3*] PhD; Rudolf Verdaasdonk[4*] Prof Dr, Ir; Lisette van Gemert-Pijnen[1*] Prof Dr

[1]Department of Psychology, Health and Technology Behavioral, Management and Social Sciences (BMS) University of Twente Enschede NL
[2]Department of Learning and Workforce Development The Netherlands Organization for Applied Scientific Research (TNO) Soesterberg NL
[3]Research Group Digital Transformation Hanze University of Applied Sciences Groningen NL
[4]TechMed Center Faculty of Science and Technology University of Twente Enschede NL
[*]these authors contributed equally

Corresponding Author:
Lea Berkemeier MSc
Department of Psychology, Health and Technology
Behavioral, Management and Social Sciences (BMS)
University of Twente
Drienerlolaan 5
Enschede
NL

Abstract

Background: With rising mental health complaints, the preventive monitoring of mental health becomes increasingly important. Different facets of mental health, such as perceived stress and mental wellbeing, have often been assessed with self-report questionnaires, administered at sporadic points in time. Ecological Momentary Assessment (EMA) may provide more continuous insights into these constructs. Research suggests that momentary affective states correlate with long-term mental health outcomes and that temporal EMA features of affective states (eg, average, variability, inertia) may be indicative of mental health complaints. However, findings have been mixed due to variability across settings and individuals. Long-term prediction performance has also not been thoroughly explored yet, partly due to the scarcity of longitudinal daily assessment data.

Objective: This paper examined the predictive accuracy of temporal EMA features for changes in mental health and investigated how the number of EMA data points (time windows) influences prediction performance.

Methods: Fifteen PhD candidates in the Netherlands, recruited via e-mail and other media, completed daily EMA of affect (valence, arousal; N=3953) and retrospective, validated monthly questionnaires (N=153) over 9 months. Retrospective outcomes included distress, perceived stress, personal accomplishment, exhaustion, and mental well-being. Temporal EMA features (average, variability, inertia, coefficient of variation) were aggregated to align with the monthly questionnaires (N=153) and used to predict changes in retrospective outcomes (R2pred) with leave-one-subject-out (LOSO) linear prediction models. In addition, within-subjects linear regressions were conducted to examine the explained in-sample variance. Prediction windows ranged from 3 to 30 days before questionnaire completion.

Results: The linear LOSO models yielded the highest median R2pred with a 3-day window for Distress (R2pred=0.44), Mental wellbeing (R2pred=0.25), and Personal accomplishment (R2pred=0.20), versus a 14-day window for Perceived stress (R2pred=0.16) and a 16-day window for Exhaustion (R2pred=0.19). Results did not show monotonic trends across time in predictive accuracy nor a consistently optimal window length. Within-subjects linear regressions indicated similar findings, with variation in performance within and across prediction windows.

Conclusions: Preliminary evidence suggests that temporal EMA features can predict changes in retrospective outcomes of mental health at a limited scale, with performance varying by window length. This discrepancy between momentary and retrospective assessments aligns with prior research on recall and may reflect individual differences in recall strategies. Findings suggest that shorter time windows between EMA and retrospective assessments or EMA alone, ideally combined with continuous objective measures, may offer a more detailed picture of changes in a person's mental health.

(JMIR Preprints 21/09/2025:84554)
DOI: https://doi.org/10.2196/preprints.84554

Preprint Settings

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

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

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

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Original paper

Predicting changes in mental health with temporal Ecological Momentary Assessment features: a longitudinal study on the predictive accuracy and optimal time windows of (measuring) daily affective states

Lea Berkemeier[1,2,3], Wim Kamphuis[2], Hilbrand Oldenhuis[3], Rudolf Verdaasdonk[4], and Lisette van Gemert-Pijnen[1]
Corresponding author: Lea Berkemeier, l.berkemeier@utwente.nl, +316 39 40 91 33
[1] Department of Psychology, Health and Technology, Centre for eHealth Research and Wellbeing, University of Twente, 7522NB Enschede, The Netherlands
[2] Department of Learning and Workforce Development, The Netherlands Organization for Applied Scientific Research (TNO), 3769 DE, Soesterberg, The Netherlands
[3] Research Group Digital Transformation, Hanze University of Applied Sciences, 9747 AS Groningen, The Netherlands
[4] TechMed Center, Faculty of Science and Technology, University of Twente, 7522 NB Enschede, The Netherlands

ORCID: 0000-0001-7065-5158 (L. Berkemeier), 0000-0002-8363-9277 (W. Kamphuis), 0000-0001-6493-5030 (H. Oldenhuis), 0000-0002-3115-637X (R. Verdaasdonk), 0000-0001-6511-7240 (L. van Gemert-Pijnen)

Abstract

Background: With rising mental health complaints, the preventive monitoring of mental health becomes increasingly important. Different facets of mental health, such as perceived stress and mental wellbeing, have often been assessed with self-report questionnaires, administered at sporadic points in time. Ecological Momentary Assessment (EMA) may provide more continuous insights into these constructs. Research suggests that momentary affective states correlate with long-term mental health outcomes and that temporal EMA features of affective states (eg, average, variability, inertia) may be indicative of mental health complaints. However, findings have been mixed due to variability across settings and individuals. Long-term prediction performance has also not been thoroughly explored yet, partly due to the scarcity of longitudinal daily assessment data.

Objective: This paper examined the predictive accuracy of temporal EMA features for changes in mental health and investigated how the number of EMA data points (time windows) influences prediction performance.

Methods: Fifteen PhD candidates in the Netherlands, recruited via e-mail and other media, completed daily EMA of affect (valence, arousal; N=3953) and retrospective, validated monthly questionnaires (N=153) over 9 months. Retrospective outcomes included distress, perceived stress, personal accomplishment, exhaustion, and mental well-being. Temporal EMA features (average, variability, inertia, coefficient of variation) were aggregated to align with the monthly questionnaires (N=153) and used to predict changes in retrospective outcomes (R^2_{pred}) with leave-one-subject-out (LOSO) linear prediction models. In addition, within-subjects linear regressions were conducted to examine the explained in-sample variance. Prediction windows ranged from 3 to 30 days before questionnaire completion.

Results: The linear LOSO models yielded the highest median R^2_{pred} with a 3-day window for Distress (R^2_{pred}=0.44), Mental wellbeing (R^2_{pred}=0.25), and Personal accomplishment (R^2_{pred}=0.20), versus a 14-day window for Perceived stress (R^2_{pred}=0.16) and a 16-day window for Exhaustion (R^2_{pred}=0.19). Results did not show monotonic trends across time in predictive accuracy nor a consistently optimal window length. Within-subjects linear regressions indicated similar findings, with variation in performance within and across prediction windows.

Conclusions: Preliminary evidence suggests that temporal EMA features can predict changes in retrospective outcomes of mental health at a limited scale, with performance varying by window length. This discrepancy between momentary and retrospective assessments aligns with prior research on recall and may reflect individual differences in recall strategies. Findings suggest that shorter time windows between EMA and retrospective assessments or EMA alone, ideally combined with continuous objective measures, may offer a more detailed picture of changes in a person's mental health.

Keywords: ecological momentary assessment; valence; arousal; affect; mental health; modeling; prediction; repeated measurement

Introduction

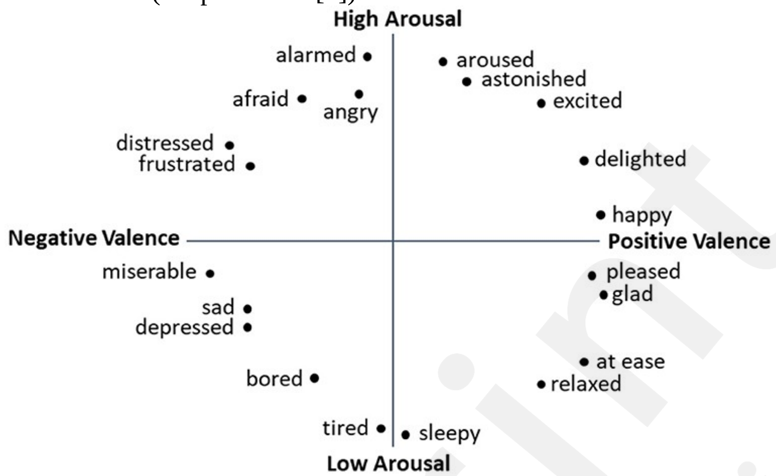
Mental health complaints such as burnout, anxiety, and depression have been on the rise globally, particularly among individuals in high-risk occupations like healthcare, military, police, and academia [1].The early detection and prevention of mental health decline are crucial to support individuals and reduce societal and economic costs [1]. Existing literature suggests associations between affective states and wellbeing or psychopathology [2], but it remains unclear how daily fluctuations in affect can predict long-term (changes in) mental health. This paper examines whether temporal dynamics of daily affective states can be used to predict changes in mental health over time.

Affective states

Affective states are defined as momentary feelings like mood and emotions [3], which tend to fluctuate in response to external and internal stimuli. According to the Circumplex Model of Affect [4], affective states can be categorized across two dimensions, namely valence (ranging from displeasure to pleasure) and arousal (ranging from deactivation to activation). By assessing and combining the valence and arousal levels, different types of affective states can be classified, as shown in Figure 1.

Affective states shape the perception and behavior of individuals in daily life and were found to be key indicators of psychological well-being and mental health outcomes [5]. Approximately 40-75% of individuals with mental disorders show symptoms that relate to unhealthy affective functioning [6,7]. For instance, symptoms such as mood swings or persistent sadness tend to be central diagnostic criteria for depression, anxiety disorders, or borderline personality disorder. In turn, wellbeing is often characterized by higher levels of positive affect [8], including feelings of happiness or pleasantness, see Figure 1.

Figure 1. Circumplex Model of Affect (adopted from [9]).



Temporal Ecological Momentary Assessment features

With the increasing availability of smartphones, it has become increasingly convenient for researchers to obtain frequent, repeated assessments of affective states through Ecological Momentary Assessments (EMA). EMA enables fine-grained measurements of a wide range of behaviors, contexts, and feelings in individuals' natural environment can be frequently and repeatedly measured, see [Error: Reference source not found,11]. In this context, EMA could be used to monitor changes in affective states, which may in turn predict mental health outcomes, see [12]. Repeated EMA of affective states enable the computation of temporal features, including 1) mean affect (average affective experience), 2) variability (degree of emotional fluctuation captured by standard deviation), or 3) inertia (autocorrelation of affective states across time, reflecting persistence of states), see [13]. These measures were shown to be important predictors of mental health constructs [13-15]. For example, high inertia and high emotional variability were linked to an increased risk for depression and other internalizing disorders [13,14,16].

Monitoring temporal EMA features may therefore help predict changes in wellbeing as mental health status and psychopathology, see [2,17,18]. To compute meaningful temporal EMA features and gain insights into different patterns and dynamics of a person's mental health trajectory, longitudinal data is needed [Error: Reference source not found,19,20]. Such trajectories may provide early warning signals for emerging mental health concerns (eg, [21-24]) and can inform Ecological Momentary Intervention (EMI), which intervenes on people's behavior in the moment to provide real-time support in their natural settings, see [8,25].

Current challenges of mental health predictions

Despite this promising approach, research using temporal EMA features to predict longitudinal mental health remains limited, see [2] and faces several challenges regarding individual differences and study design [19,20]. Regarding individual differences, individuals tend to differ in affect ratings, behaviors, and context. According to [26], p.7, "analyzing such individual, person-specific networks requires methodologies to capture the relevant variables to gather an individual's thoughts, experiences, and behaviors in situations with specific triggers". Ignoring such variances across individuals reduces ecological validity [27] and may help explain limited predictive success, see [28], highlighting the need for subject-dependent approaches.

Regarding study design, retrospective, self-report questionnaires remain the gold standard for the assessment of mental health and wellbeing, but tend to be administered infrequently [29], limiting insights into temporal dynamics of affective states. Conversely, Ecological Momentary Assessments (EMA) offer detailed insights into these real-world variations of affective states through repeated assessments, enhancing ecological validity and minimizing memory or recall biases [30-32]. Many existing studies tend to be of short duration, focus on within-day fluctuations at high sampling rates, or rely on clinical samples [24,33], leaving gaps in knowledge about long-term (between-days) patterns per person and predictions in general populations, see [8].

Objectives

This paper aims to improve the prediction of long-term mental health changes by examining longitudinal temporal EMA features of affective states. In this paper, "mental health" refers to retrospective self-report measures such as perceived stress, distress, exhaustion and mental wellbeing, capturing diverse facets and symptoms of mental health, see [34]. Participants of the study contributed nine months of data, consisting of daily EMA of affect, which were used to compute temporal features (average, variability and inertia) and to predict subsequent changes in multiple retrospective outcomes. By focusing on a non-clinical sample, this study also explores whether temporal EMA features generalize to the broader population, providing insights for preventive strategies and mental health monitoring. The research will focus on the following research questions:

1. To what extent can changes in retrospective outcomes of mental health be accurately predicted by temporal EMA features of affective states?
2. How many EMA datapoints are needed for accurate predictions of changes in retrospective outcomes of mental health?

Method

The study "CAMSTAM – Continuous Automatic Mental State Monitoring" was reviewed by the TNO Institutional Review Board (number: 2021-078, see [28]). The study design aims to explore the predictive accuracy of temporal EMA features on changes in retrospective measures of mental health by (1) creating temporal features from daily EMA assessments as predictor variables; (2) comparing prediction performances across time windows, ranging from 3 to 30 days prior to the monthly questionnaire completion. Through the longitudinal setup of the data collection, including repeated EMA measures and diverse retrospective outcomes of mental health, first insights into the design of future mental health monitoring protocols within non-clinical populations can be given.

Participants

The sample consisted of 16 PhD candidates who worked in the Netherlands and participated in a nine-month-long study that started at the end of March 2022. Participants were recruited via e-mail and other media and were invited for an in-person intake interview with the main researcher if interested. PhD candidates are considered a risk group due to high cognitive demands and workload. At least one in two PhD candidates experiences psychological distress, and at least one in three runs a higher than usual risk of developing a common psychiatric disorder [35]. Participants were included if they had completed at least 8 of the 9 retrospective questionnaires. One participant had to be excluded due to an insufficient number of monthly measurements (less than 7 out of 9 months), leaving a total of 15 participants for the data analysis. To be included in the study, participants had to have a sufficient Dutch proficiency (B2 level or higher) since the surveys and questionnaires were administered in Dutch. Participant's age ranged from 24 to 38 years, with a mean age of 29 years (SD = 4.36). Of the 15 participants, 9 (60%) were female and 6 (40%) were male. At the beginning of the data collection, 2 (13.3%) were first-year PhD candidates, 6 (40%) second-year, 2 (13.3%) third-year, and 5 (33.4%) final-year. Participants were compensated with 10€ per month, individual monthly feedback sessions with the main researcher. To be included in this current study, participants needed to provide 85% valid data points, which equals approximately 230 filled-in subjective questionnaires or 7.6 months of data collection. If participants had less than 85% valid data points after nine months, the data collection was extended to a maximum of eleven months. Exemptions were made in four cases ([1=64%; 2=66%; 3=71%; 4=74% respectively) where participants could not reach the 85% due to technical difficulties with the app used for the Ecological Momentary Assessments (EMA). These participants were still included in the analyses.

The data collection also involved objective measures, including behavioral data via smartphones and physiological data from wearing an OURA ring, see [28]. Consequently, participants signed an informed consent for all types of collected data and also had to fulfill the following criteria for participation: no heart rhythm disorder, interest in the use of wearables, and self-monitoring. For the current study, the objective measurements were excluded due to the scope of the study.

Measures

For an overview of the included subjective measurements in the data collection, see Table 1.

Table 1. Overview of the predictor and outcome measures.

Measure	Reference	Type	Frequency	Scale
Valence	See [36]	Predictor	Daily	Continuous (VAS 0-100)
Arousal	See [36]	Predictor	Daily	Continuous (VAS 0-100)
Four-Dimensional Symptom Questionnaire (4DSQ)	[37]	Outcome	4-weekly	Continuous (subscale scores questionnaire)
Perceived Stress Scale (PSS)	[38]	Outcome	4-weekly	Continuous (total score questionnaire)
Utrecht Burnout Scale (UBOS)	[Error: Reference source not found]	Outcome	4-weekly	Continuous (subscale scores questionnaire)
General Health Questionnaire (GHQ)	[40]	Outcome	4-weekly	Continuous (total score questionnaire)
Mental Health	[41]	Outcome	4-weekly	Continuous (total + subscale scores

Continuum Short-Form (MHC-SF)				questionnaire)	

Predictors
For the daily EMA questions, the ‘How am I’ app of [42] has been used. Participants filled in two questions at the end of each day, asking about their affective state (valence and arousal levels of that day). The EMA questions were available daily from 5 p.m. to 5 a.m., and participants could choose their preferred time for receiving reminders to complete the questions, based on their convenience or evening routine. For the valence question, participants rated on a scale from 0 (very unpleasant) to 100 (very pleasant) how pleasant their day was. For the arousal question, participants rated on a scale from 0 (very passive) to 100 (very active) how mentally active or passive they felt during the day. Similar EMA measures of valence and arousal were used in previous studies of [36, 43].

Based on prior literature on temporal EMA features [13,14], the following temporal features of affective states were computed as predictors preceding each monthly questionnaire date: average, variability (standard deviation), and inertia of valence and arousal. Additionally, the coefficient of variation of valence and arousal was included as an exploratory predictor to account for possible effects over time. All predictors were computed per person (within-subjects) and standardized within each training fold so that the scaling was determined from the training data only.

Outcome measures
Several questionnaires were filled in during the intake and every four weeks thereafter in Survalyzer [44] via personal links. The intake questionnaires were only used to create change scores for the first month, but were not used as outcome measures. The study included several retrospective, validated questionnaires measuring different facets of mental health, which were administered every 4 weeks, including the General Health Questionnaire (GHQ), Perceived Stress Scale (PSS), Mental Health Continuum Short Form (MHC-SF), Four-Dimensional Symptom Questionnaire (4DSQ), and Utrecht Burnout Scale (UBOS).

To measure the changes in the different questionnaires, retrospective outcomes were defined as month-to-month differences per person (eg, $Q_{change} = Q_t – Q_{t-1}$) and were kept in raw units for modeling. For example, a positive score for PSS would indicate an increase, a score of zero no change, and a negative score a decrease of perceived stress in comparison to the previous month. Several sub-scales were excluded from the analysis due to overall low variance (4DSQ Somatization, UBOS Depersonalization, GHQ) or computational issues (4DSQ Anxiety, 4DSQ Depression), leading to invalid cross-validation (CV) metrics. Consequently, the sub-scale Distress from 4DSQ, Personal Accomplishment and Exhaustion from UBOS, the MHC, and the PSS were included in the analyses.

The sub-scale Distress from the Four-Dimensional Symptom Questionnaire (4DSQ) contained 16 items to measure distress complaints of the previous month. Participants evaluated their distress complaints on a 5-point Likert scale ranging from no (0) to very often (5). Next, the last three categories (regularly, often, very often, or constantly) are merged into 2, see [37]. The Perceived Stress Scale (PSS) contained 10 items to measure experienced stressful moments. Participants evaluated their stressful moments of the last month with a 5-point Likert scale ranging from never (0) to very often (4). Items 4, 5, 7, and 8 were positively stated and thus reverse scored for comparison (4=0, 3=1, 1=3, 0=4). The sub-scale Personal accomplishment and Exhaustion from the Utrecht Burnout Scale (UBOS) consisted of 6 and 5 items, respectively, measuring burnout in terms of emotional exhaustion and personal accomplishment. Participants rated statements about how they felt at work the last month on a 7-point Likert Scale ranging from 0 (never) to 6 (always). The Mental Health Continuum Short-Form (MHC-SF; [41]) comprised 14 items measuring various feelings of mental wellbeing of the past month, which were rated on a 6-point Likert scale ranging from never (1) to every day (6). All five included questionnaires showed good internal consistency (Alpha > .70; [45]), see Table 2.

Table 2. Internal consistency of the included 4-weekly outcome measures.

Questionnaire	Scale	Sub-scale	Number of items	Cronbach’s α other studies	Cronbach’s α current study
Four-Dimensional Symptom Questionnaire (4DSQ)	5-point Likert from no (0) to very often (4)	Distress	16	0.93 [37]	0.90
Perceived Stress Scale (PSS)	5-point Likert scale ranging from never (0) to very often (4)	-	10	0.63 [46]	0.79
Utrecht Burnout Scale (UBOS)	7-point Likert Scale ranging from never (0) to always (6)	Personal Accomplishment	6	0.79 [Error: Reference source not found]	0.87
		Exhaustion	5	0.89 [Error: Reference source not found]	0.91

Mental Health Continuum Short-Form (MHC-SF)	6-point Likert scale ranging from never (0) to every day (5)	-	14	0.89 [41]	0.94

Data preprocessing and analysis

All preprocessing and analyses were done using RStudio [47], including the tidyverse library and supporting packages such as caret, glmnet, psych, and zoo. The EMA dataset contained 4782 observations, of which 562 (11.75%) had missing valence or arousal values. Missingness was slightly associated with weekday (missing rates: Monday [7.1%]; Tuesday [8.8%]; Wednesday [8.0%]; Thursday [8.6%]; Friday [10.5%]; Saturday [8.4%]; Sunday [10.9%]). Two preprocessing branches were evaluated, namely (1) excluding rows with missing EMA and (2) imputing missing EMA. Since imputation did not yield systematic improvements in out-of-sample R^2 (R^2_{pred}) or the Root Mean Squared Error (RMSE) across outcomes and time windows, it was decided to exclude or skip those missing EMA rows. Rows with insufficient data for calculating the dynamic measures predictors (fewer than two data points for averages or fewer than three for the standard deviation and inertia) were also excluded. The monthly surveys (N=161) had 3 missing entries (1.9%), which were reduced to 1 after excluding the participant noted in *Participants*.

Since participants started at different dates, monthly questionnaire timepoints were aligned per person. The variable "month" was coded from each individual's start date (T0) and the subsequent 4-weekly questionnaires (T1-T9). According to the guidelines of the different retrospective questionnaires, 4-weekly scores of each questionnaire were calculated, and the target outcome was defined as month-to-month change per person ($Q_{change} = Q_t – Q_{t-1}$), retained in raw units. For the interpretability of modeling error, the prediction model's RMSE was evaluated against the sample standard deviation (SD) of the change scores for each questionnaire.

For every person and monthly questionnaire date, we computed subject-dependent (within-subjects) features from rolling windows of 3-30 days immediately before the questionnaire date. Windows with fewer than 3 EMA days were skipped, due to the calculation requirements of the standard deviation and inertia predictors, leaving 28 time windows (3-30 days). Group-level features were explored but omitted from the final models due to the small correlations and to avoid collinearity. The final predictive models included eight temporal EMA features, all computed within person and window:

$$\left[Outcome\, measure \right] = valence\ \acute{x}_{within} + valence\ s_{within} + valence\ inertia_{within} + valence\ coefficient\ of\ variation_{within} + arou$$

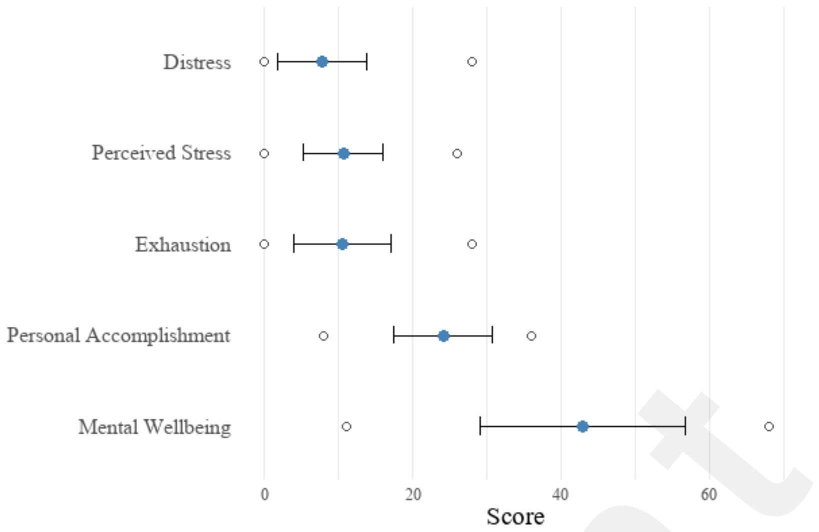
The model was executed for each of the following outcome measures: Distress (4DSQ_Distress), Perceived stress (PSS), Exhaustion (UBOS_Exhaustion), Personal accomplishment (UBOS_Personal_Accomplishment), and Mental wellbeing (MHC). Given the small per-person series (9 time points), validation used leave-one-subject-out (LOSO) instead of leave-one-month-out (per person) cross-validation. In each fold, one participant was held out, predictors were centered and scaled within the training fold, the model was trained on the remaining participants, and predictions were generated for all observations of the held-out participant. To ensure a stable estimation and viable folds, at least 10 rows were required for a given questionnaire-window combination and at least two eligible participants; otherwise, the combination was skipped. Two EMA-free baselines were computed in every LOSO fold to quantify the added value of EMA features: "No change" ($Q_{change} = 0$) and a "Train-Mean" (predicting training-set mean change). In addition to LOSO, we fit person-specific (within-subject) linear regressions using the same predictors to quantify in-sample explained variance (adjusted R^2) of changes per person and window. These results and summaries are provided in Appendix C.

Following [48], the prediction performance was evaluated on multiple measures, namely the cross-validated R^2_{pred} (variance explained in held-out subject) and RMSE (accuracy in terms of error). To address the first research question, "To what extent can changes in retrospective outcomes of mental health be accurately predicted by temporal EMA features of affective states?", LOSO R^2_{pred} distributions for included questionnaires at the 30-day window were summarized using boxplots, including median and interquartile range (IQR). To address the second research question, "How many EMA datapoints are needed for accurate predictions of changes in retrospective outcomes of mental health?", LOSO R^2_{pred} for included questionnaires were visualized with boxplots (median, IQR) across all window lengths (3-30 days) to assess how the predictive accuracy varies with the number of EMA days.

Results

At first, descriptive statistics were calculated for each retrospective questionnaire, see Figure 2. On average, participants' levels of Distress, Exhaustion, and Perceived stress were relatively low (lowest or second quartile) when compared to the range of the questionnaire (see Table 3). Indicators such as Personal accomplishment and Mental wellbeing showed higher levels (third quartile). Correlations between predictor-outcome combinations were considered low [49], see Appendix B.

Figure 2. Descriptive statistics per outcome variable.



Note. Steelblue dot = mean; Black Whiskers = mean +/- 1 SD; Grey circles: min and max

Table 3. Descriptive statistics of retrospective questionnaires.

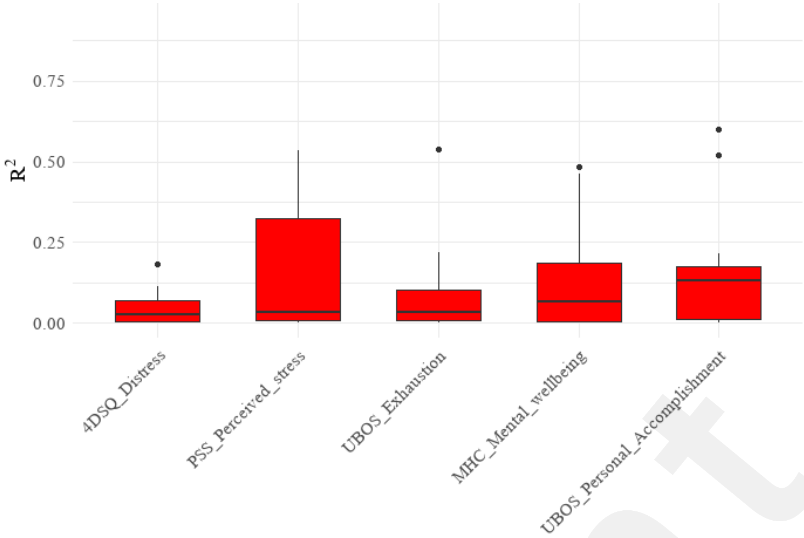
Questionnaire		Current study			
Name	Total score range	Mean	SD	Min	Max
4DSQ Distress	0-32	7.75	6.03	0	28
PSS Perceived stress	0-40	10.63	5.40	0	26
UBOS Exhaustion	0-30	10.44	6.61	0	28
UBOS Personal accomplishment	0-36	24.09	6.63	0	36
MHC Mental wellbeing	0-70	42.96	13.83	11	68

Predictive accuracy

To address the first research question, “To what extent can changes in retrospective outcomes of mental health be accurately predicted by temporal EMA features of affective states?”, the out-of-sample R^2 (R^2_{pred}) was calculated for LOSO linear prediction models at the 30-day time window, see Figure 3. The median R^2_{pred} for the included questionnaires ranged from 0.02 (Distress), 0.03 (Perceived stress), 0.03 (Exhaustion), 0.07 (Mental wellbeing), to .13 (Personal accomplishment) for the LOSO linear prediction models. At the same time, the lowest variability in performance was found for Distress (IQR = 0.07), followed by Exhaustion (IQR = 0.09), Personal accomplishment (IQR = 0.16), Mental wellbeing (IQR = 0.18), and Perceived stress (IQR = 0.32). Thus, the best model explained 13% of the held-out variance (Personal accomplishment), with notable between-subjects heterogeneity (especially for Perceived Stress). Additionally, within-subjects linear regression models showed similar patterns in terms of the explained in-sample variance (see Appendix C), highlighting between-subjects heterogeneity.

For the error (RMSE), the percentage of LOSO folds with RMSE below the outcome’s SD (higher = better) was compared with two simple baselines (“no-change” and “train-mean”), see Appendix A. The LOSO linear model of Distress (73%) outperformed the baseline models (61-65%), indicating an improvement with the LOSO model. The error rate of Mental wellbeing LOSO linear (73%) is slightly lower than the baseline models (75%), similarly to Exhaustion (59% versus 60-61%), indicating a comparable, thus no advantage, of the LOSO model. Lastly, the error rate of Perceived stress LOSO linear (47%) underperformed the baseline models (54%), similarly to Personal accomplishment (52% versus 60%). Overall, daily temporal EMA features of affective states offered limited out-of-sample explanatory power for month-to-month changes in these five outcomes, and only Distress showed a consistent RMSE benefit of the LOSO linear model.

Figure 3. Predictive accuracy (R2) per questionnaire, including all data (time window = 30).



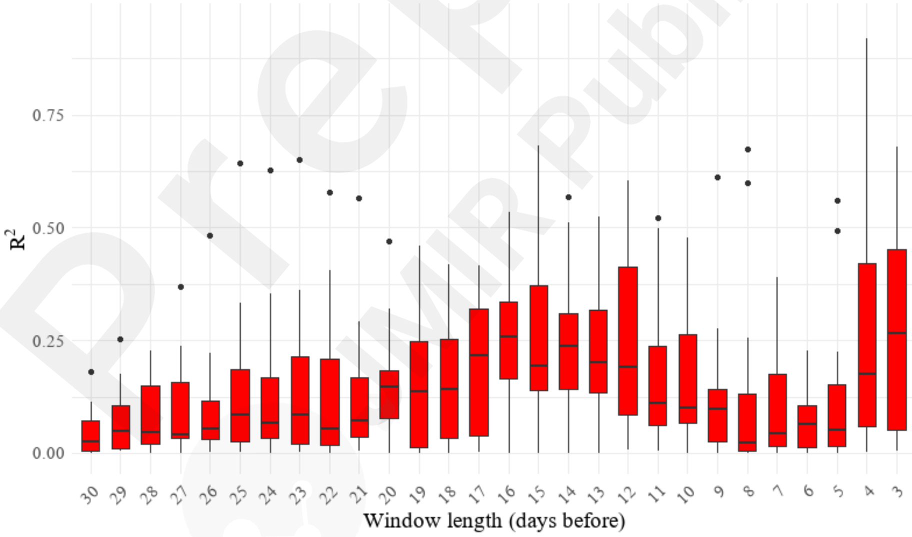
Number of EMA datapoints

To address the second research question, “How many EMA datapoints are needed for accurate predictions of changes in retrospective outcomes of mental health?”, LOSO linear models were calculated for time windows spanning 2-30 days before each questionnaire (ie, using the previous k days of EMA). In the following sections, the performance is reported as out-of-sample predictive R2 (R^2_{pred}) per outcome and an answer given to the second research question.

Distress

For Distress, the highest median R^2_{pred} occurred with a 3-day window (median R^2_{pred} = 0.44; IQR = 0.68), whereas the 30-day window yielded the most stable but lowest performance (median R^2_{pred} = 0.02; IQR = 0.07), see Figure 4. Overall, shorter windows tended to increase the predictive performance for Distress, but with larger variability across folds. The performance increase was not fully monotonic across window lengths, with a performance dip at 5-9 day windows (low median and narrow IQR), while some short windows (eg, 3-4 days) yielded high medians but wide IQR.

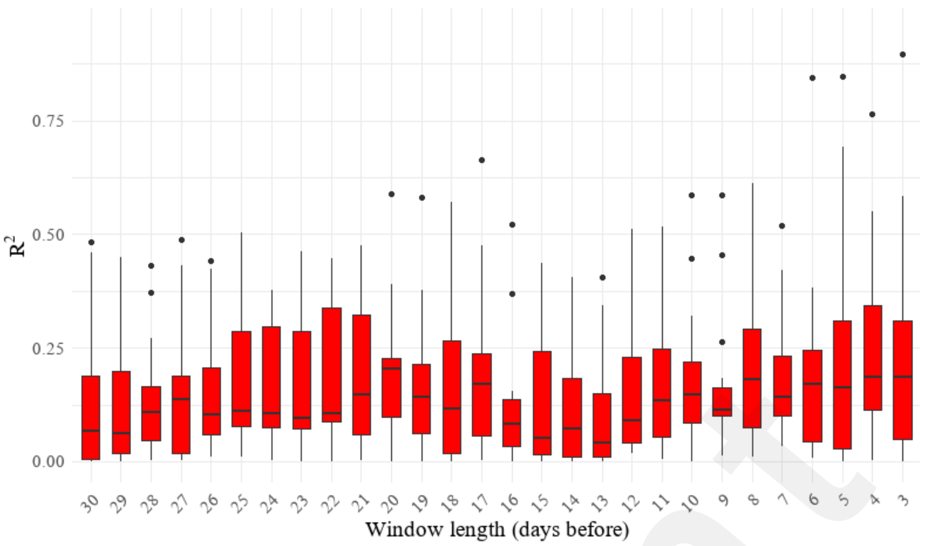
Figure 4. Predictive accuracy (R2) across time windows for 4DSQ Distress.



Mental wellbeing

For Mental wellbeing, the highest median R^2_{pred} occurred with a 3-day window (median R^2_{pred} = 0.25; IQR = 0.41). The lowest variability across folds was found at a 9-day window (median R^2_{pred} = 0.11; IQR = 0.06), see Figure 5. Overall, the performance tended to vary by window length and did not show monotonic trends: very short windows (eg, 3-6 days) yielded higher median R^2_{pred} but wider IQR, while several longer time windows (eg, 16, 18) yielded low medians with narrow IQR.

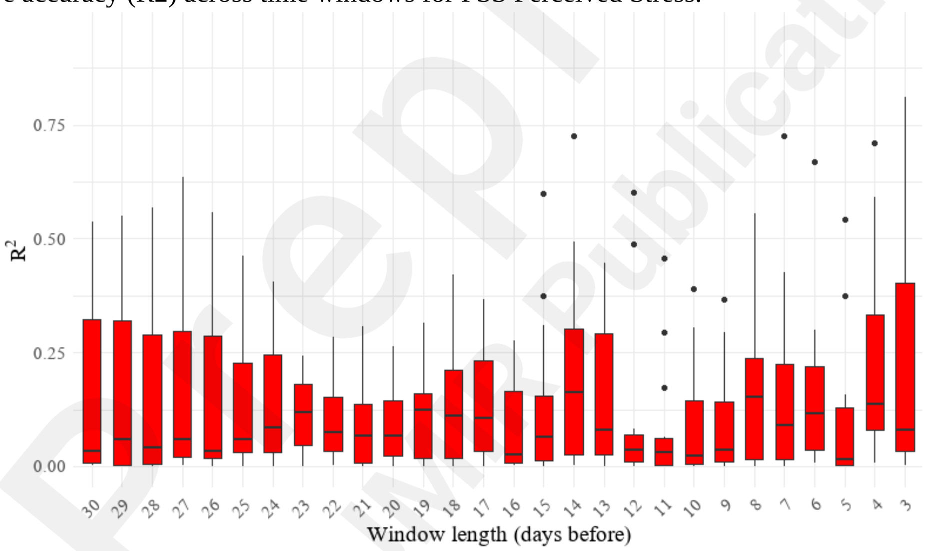
Figure 5. Predictive accuracy (R2) across time windows for MHC Mental wellbeing.



Perceived stress

For Perceived stress, the highest median R^2_{pred} was found for a 14-day window (median R^2_{pred} = 0.16; IQR = 0.28) while the lowest variability occurred at an 11-day window (median R^2_{pred} = 0.03; IQR = 0.06), see Figure 6. Overall, the performance tended to vary by window length and did not show monotonic trends: longer windows (eg, 24-30 days) generally showed lower medians with wider IQR, whereas shorter windows either showed low median and narrow IQR (eg, 5, 11, 12 day window) or wide IQR (eg, 3 day window).

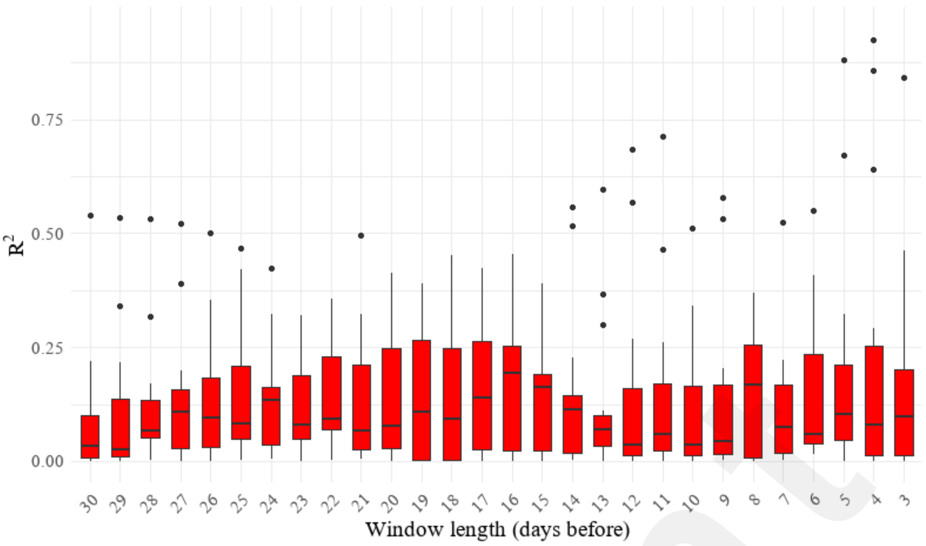
Figure 6. Predictive accuracy (R2) across time windows for PSS Perceived Stress.



Exhaustion

For Exhaustion, the highest median R^2_{pred} occurred at a 16-day window (median R^2_{pred} = 0.19; IQR = 0.23) while the lowest variability occurred at a 13-day window (median R^2_{pred} = 0.07; IQR = 0.07), see Figure 7. Across windows, the performance varied without a clear monotonic trend: several time windows showed lower medians (eg, 9-12, 29-30 day window), others displayed wide IQR (eg 4, 6, 18, 19 day window) and a few combined higher medians with wide IQR (eg, 8, 16 day window).

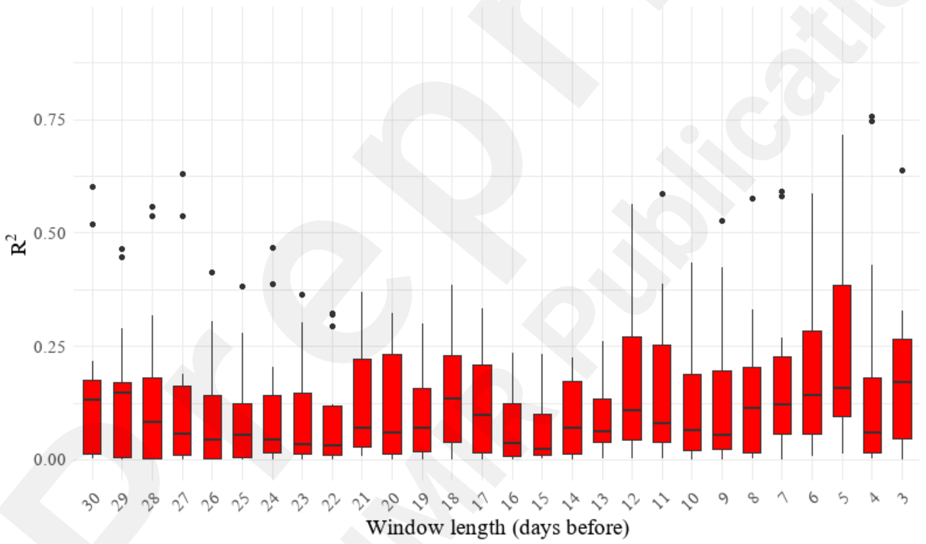
Figure 7. Predictive accuracy (R2) across time windows for UBOS Exhaustion.



Personal accomplishment

For Personal accomplishment, the highest median R^2_{pred} occurred at a 3-day window (median R^2_{pred} = 0.20; IQR = 0.26) while the lowest variability occurred at a 15-day window (median R^2_{pred} = 0.02; IQR = 0.09), see Figure 8. Across windows, the performance varied without a clear monotonic trend: several time windows showed lower medians and narrower IQR (eg, 12, 22 day window), others displayed higher medians with a wide IQR (eg, 3, 5, 6 day window).

Figure 8. Predictive accuracy (R2) across time windows for UBOS Personal Accomplishment.



In summary, there was no single time window length that consistently improved predictive performance across outcomes, but performance generally improved when optimizing the window relative to the 30-day window. For example, for Distress the median R^2_{pred} increased from 0.02 (30-day window) to 0.44 (3-day window). Peak median R^2_{pred} values occurred at short to mid-range windows (~3-16 days): Mental wellbeing peaked at a 3-day window (0.25), Perceived stress at 14 days (0.16), Exhaustion at 16 days (0.19), and Personal accomplishment at a 3-day (0.20). Importantly, these peak performances often coincide with wider variability across folds (eg, IQRs ~ 0.23-0.68), whereas the most stable windows tended to show lower medians (eg, Distress 30 days: IQR 0.07; Mental wellbeing 9 days: IQR 0.06; Perceived stress 11 days: IQR 0.06; Exhaustion 13 days: IQR 0.07; Personal accomplishment 15 days: IQR 0.09), thus showing an accuracy-stability trade-off. Similar trends indicating no preferred window and a within-subjects heterogeneity were visible in terms of the in-sampled explained variance (see Appendix C).

Discussion

Temporal EMA features of affective states are considered important predictors because they capture within-person fluctuations linked to psychopathology and mental wellbeing, yet most prior studies were of short duration. In the current study, participants provided daily EMA of valence and arousal and monthly retrospective questionnaires of mental health over nine months. Temporal features of affective states were derived from EMA and used to predict changes in retrospective outcomes.

Discussion of main findings

Starting with the first research question, “To what extent can changes in retrospective outcomes of mental health be accurately predicted by temporal EMA features of affective states?”, the results showed that temporal EMA features of affective states predicted changes in retrospective outcomes over time to a limited extent. With all observations included, median out-of-sample performance (R^2_{pred}) ranged from 0.02 (Distress) to 0.13 (Personal accomplishment). These findings align with prior work showing partial overlap between momentary and retrospective assessments, for instance, by discriminating ‘good days’ from ‘bad days’ [50], but limited agreement overall [51]. One explanation could be that momentary assessments reflect present, experiential knowledge, whereas retrospective assessments involve episodic memory processes [50,52]. Retrospective reports may overweigh the most intense or most recent experiences (episodic memory), reducing alignment with the preceding month’s EMA-derived temporal features and predictive accuracy. Another explanation could be limited variability in retrospective outcomes, which may result in the baseline models performing as well as the LOSO model.

Regarding the second research question, “How many EMA datapoints are needed for accurate predictions of changes in retrospective outcomes of mental health?”, median predictive accuracy improved for each retrospective outcome through window optimization compared to a 30-day window, indicating the latter may be less preferable. However, there was no monotonic trend or universal optimum, consistent with previous mixed results and recommendations on recall periods for subjective measures [53,54]. For instance, [54] suggested four weeks for the prediction of somatic symptom burden, whereas [53] argued for one week to minimize recall bias. The current results also support the conclusions of [55], namely that there is not a ‘one size fits all’ recall period because content and temporal variability differ across persons. Several time windows showed wide variability in predictive accuracy for both LOSO linear models and linear regression models (see Appendix C), likely reflecting individual differences in how experiences are perceived and reconstructed [26,51]. For instance, one individual may recall recent aspects of perceived stress (favoring shorter windows). In contrast, another recalls the most intense moments of a month (favoring longer windows), which may result in different predictive accuracies for shorter or longer time windows.

Strengths and limitations

Strengths of the current study include the extensive within-person longitudinal data, which yielded many observations for person-specific temporal EMA features of affective states. Furthermore, the repeated-measures nature of the data collection gives new insights into long-term prediction, and the use of multiple retrospective outcomes of mental health allowed a broad, multi-dimensional analysis. To the best of the authors’ knowledge, no prior longitudinal study has examined the predictive accuracy of temporal EMA features of affective states for changes in retrospective outcomes across 9 months while varying prediction windows.

Limitations of the current study include a small sample size (N = 15), although each participant contributed to a large number of longitudinal observations (~274 per person) which may be reused for future analyses. Furthermore, within-subjects cross-validation (leave-one-month-out) could not be performed and some retrospective outcomes showed limited variability or no extreme values. This led to the exclusion of certain questionnaires and possibly limited predictive accuracies. Some prediction models showed low R2 values or high variability, likely reflecting person-level differences in symptom severity over time, the multifaceted nature of the constructs, or temporal lags in temporal EMA features of affective states not captured by daily time windows. Finally, the current analyses rely on linear assumptions and between-subjects cross-validation, given the limited number of monthly observations per person. While this approach provides preliminary insights into predictive accuracies, it does not incorporate within-subjects cross-validation or more complex modeling approaches.

Implications

[50] recommended testing time-window effects on recall; however, the current findings suggest there is no single window that enhances prediction across retrospective outcomes. Future work should increase content overlap between EMA and retrospective assessment, see [50,56] to clarify convergence. At the same time, retrospective assessments contain ambiguities and confounders, including recall strategies, and prior work often favors EMA for capturing fine-grained predictors and their dynamics over time [50,51]. To avoid limited agreement and predictive accuracy between the momentary and retrospective assessments, future research should either investigate the predictive accuracy with a smaller time window between the retrospective outcome and EMA, keeping it as short as possible, such as one or two weeks or investigate the predictive accuracy of only EMA outcomes and trends throughout time, potentially in combination with other more objective or continuous data sources. This aligns with the conclusion of [11] that EMA can give insights into more fine-grained patterns of mental health and affective states in comparison to single-administered retrospective assessments.

Predicting mental health from temporal EMA features of affective states can be complex and likely requires high-temporal, multimodal data to grasp the context of subjective measures [57,58]. By extending intensive longitudinal design with passive sensing from wearable and smartphones (eg, [19]), one can integrate sensor data and self-report streams to track individual mental health trajectories [59-61]. Future research should test whether temporal EMA features of affective states align with passive sensor data and whether within-person changes are visible across both modalities. If passive sensing improves the predictive utility of EMA or temporal EMA features, one could consider replacing EMA measures with continuous measures to reduce participant burden while maintaining (or enhancing) data streams. For example, [62] provided preliminary evidence that depression in older adults living alone could be classified with a machine learning program using EMA and wearable data. Similarly, [27] highlights that the prediction of affective states with EMA is possible to a certain extent with objective data from wearables and smartphones. Furthermore, these continuous data streams of EMA and objective data can be integrated into mental health apps such as MoodPrism (see [12]) to promote self-awareness of one's mental health and, importantly, to automatically predict and warn individuals at risk of mental health decline.

Conclusion

Temporal EMA features can provide valuable insights into a person's mental health trajectory. In this longitudinal study, temporal EMA features of affective states could predict changes in retrospective mental health outcomes over time to a limited extent, and no single window length consistently optimized accuracy. Discrepancies between EMA-derived temporal features and retrospective outcomes may stem from individual differences in recall strategies or from differences between experiential and episodic memory. Given the advantages of EMA and the potential of passive sensing for contextual insight, future work could pursue longitudinal prediction with shorter time intervals between EMA and retrospective assessments. However, consistent with previous work, the most promising direction may lie in combining temporal EMA features of affective states with continuous objective measures (eg, wearables, smartphones) to detect long-term changes and fluctuations in mental health trajectories.

Acknowledgements

The authors would like to thank all participants who provided extensive data for this manuscript and the associated analyses.

Data Availability

The dataset supporting the findings, which also contains passive sensing data, will be pseudo-anonymized and made publicly available in a separate publication.

Conflict of interest

The authors declare no competing interests.

Appendix

The online version contains supplementary material available at:

References

1. Salvagioni DAJ, Melanda FN, Mesas AE, González AD, Gabani FL, Andrade SM. Physical, psychological and occupational consequences of job burnout: a systematic review of prospective studies. PLoS One. 2017;12(10):e0185781. doi:10.1371/journal.pone.0185781
2. Panaite V, Rottenberg J, Bylsma LM. Daily affective dynamics predict depression symptom trajectories among adults with major and minor depression. Affect Sci. 2020;1(3):186-198. doi:10.1007/s42761-020-00014-w
3. Niven K. Affect. In: Gellman MD, Turner JR, eds. Encyclopedia of Behavioral Medicine. New York, NY: Springer; 2013. doi:10.1007/978-1-4419-1005-9_1088
4. Russell JA. A circumplex model of affect. J Pers Soc Psychol. 1980;39(6):1161-1178. doi:10.1037/h0077714
5. Gross JJ, Uusberg H, Uusberg A. Mental illness and well-being: an affect regulation perspective. World Psychiatry. 2019;18(2):130-139. doi:10.1002/wps.20618
6. Gross JJ, Jazaieri H. Emotion, emotion regulation, and psychopathology: an affective science perspective. Clin Psychol Sci. 2014;2(4):387-401. doi:10.1177/2167702614536164
7. Jazaieri H, Urry HL, Gross JJ. Affective disturbance and psychopathology: an emotion regulation perspective. J Exp Psychopathol. 2013;4(5):584-599. doi:10.5127/jep.030312
8. de Vries LP, Baselmans BM, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud. 2021;22(5):2361-2408. doi:10.1007/s10902-020-00324-7
9. Schmitz-Hübsch A, Stasch SM, Becker R, Fuchs S, Wirzberger M. Affective response categories—toward personalized reactions in affect-adaptive tutoring systems. Front Artif Intell. 2022;5:873056. doi:10.3389/frai.2022.873056
10. Winstone L, Heron J, John A, et al. Ecological momentary assessment of self-harm thoughts and behaviors: systematic review of constructs from the Integrated Motivational-Volitional Model. JMIR Ment Health. 2024;11:e63132. doi:10.2196/63132
11. Wenze SJ, Miller IW. Use of ecological momentary assessment in mood disorders research. Clin Psychol Rev. 2010;30(6):794-804. doi:10.1016/j.cpr.2010.06.006
12. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health. 2016;3(4):e6202. doi:10.2196/mental.6202
13. Dejonckheere E, Mestdagh M, Houben M, et al. Complex affect dynamics add limited information to the prediction of psychological well-being. Nat Hum Behav. 2019;3(5):478-491. doi:10.1038/s41562-019-0555-0

14. Koval P, Ogrinz B, Kuppens P, et al. Affective instability in daily life is predicted by resting heart rate variability. PLoS One. 2013;8(11):e81536. doi:10.1371/journal.pone.0081536
15. Jenkins BN, Ong LQ, Ong AD, Lee HY, Boehm JK. Mean affect moderates the association between affect variability and mental health. Affect Sci. 2024;5(2):99-114. doi:10.1007/s42761-024-00238-0
16. Brose A, Schmiedek F, Koval P, Kuppens P. Emotional inertia contributes to depressive symptoms beyond perseverative thinking. Cogn Emot. 2015;29(3):527-538. doi:10.1080/02699931.2014.916252
17. Koval P, Brose A, Pe ML, et al. Emotional inertia and external events: the roles of exposure, reactivity, and recovery. Emotion. 2015;15(5):625-636. doi:10.1037/emo0000059
18. Suls J, Green P, Hillis S. Emotional reactivity to everyday problems, affective inertia, and neuroticism. Pers Soc Psychol Bull. 1998;24(2):127-136. doi:10.1177/0146167298242002
19. Hamaker EL, Wichers M. No time like the present: discovering the hidden dynamics in intensive longitudinal data. Curr Dir Psychol Sci. 2017;26(1):10-15. doi:10.1177/0963721416666518
20. Wright AGC, Zimmermann J. Applied ambulatory assessment: integrating idiographic and nomothetic principles of measurement. Psychol Assess. 2019;31(12):1467-1480. doi:10.1037/pas0000685
21. Czyz EK, Glenn CR, Arango A, Koo HJ, King CA. Short-term associations between nonsuicidal and suicidal thoughts and behaviors: a daily diary study with high-risk adolescents. J Affect Disord. 2021;292:337-344. doi:10.1016/j.jad.2021.05.104
22. Schoevers RA, van Borkulo CD, Lamers F, et al. Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. Psychol Med. 2021;51(11):1906-1915. doi:10.1017/S0033291720000689
23. Ammerman BA, Law KC. Using intensive time sampling methods to capture daily suicidal ideation: a systematic review. J Affect Disord. 2022;299:108-117. doi:10.1016/j.jad.2021.10.121
24. Rabasco A, Sheehan K. The use of intensive longitudinal methods in research on suicidal thoughts and behaviors: a systematic review. Arch Suicide Res. 2021;26(3):1007-1021. doi:10.1080/13811118.2021.1903635
25. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol. 2010;15(1):1-39. doi:10.1348/135910709X466063
26. Hofmann SG. Toward a cognitive-behavioral classification system for mental disorders. Behav Ther. 2014;45(4):576-587. doi:10.1016/j.beth.2014.03.001.
27. Ram N, Brinberg M, Pincus AL, Conroy DE. The questionable ecological validity of ecological momentary assessment: considerations for design and analysis. Res Hum Dev. 2017;14(3):253-270. doi:10.1080/15427609.2017.1340052
28. Berkemeier L, Kamphuis W, Brouwer AM, et al. Measuring affective state: subject-dependent and -independent prediction based on longitudinal multimodal sensing. IEEE Trans Affect Comput. 2025;16(2):827-843. doi:10.1109/TAFFC.2024.3474098
29. Trull TJ, Ebner-Priemer U. The role of ambulatory assessment in psychological science. Curr Dir Psychol Sci. 2014;23(6):466-470. doi:10.1177/0963721414550706
30. Schwarz N. Retrospective and concurrent self-reports: the rationale for real-time data capture. In: Stone A, Shiffman S, Atienza AA, Nebeling L, eds. The Science of Real-Time Data Capture: Self-Reports in Health Research. Oxford, UK: Oxford University Press; 2007:11-26. doi:10.1093/oso/9780195178715.003.0002
31. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol. 2008;4:1-32. doi:10.1146/annurev.clinpsy.3.022806.091415
32. Gratch I, Choo TH, Galfalvy H, et al. Detecting suicidal thoughts: the power of ecological momentary assessment. Depress Anxiety. 2021;38(1):8-16. doi:10.1002/da.23043
33. Niemeijer K, Mestdagh M, Verdonck S, Meers K, Kuppens P. Combining experience sampling and mobile sensing for digital phenotyping with m-Path Sense: performance study. JMIR Form Res. 2023;7:e43296. doi:10.2196/43296
34. Currey D, Hays R, Torous J. Digital phenotyping models of symptom improvement in college mental health: generalizability across two cohorts. J Technol Behav Sci. 2023;8(4):368-381. doi:10.1007/s41347-023-00310-9
35. Levecque K, Anseel F, De Beuckelaer A, Van der Heyden J, Gisle L. Work organization and mental health problems in PhD students. Res Policy. 2017;46(4):868-879. doi:10.1016/j.respol.2017.02.008
36. Kang S, Choi W, Park CY, et al. K-emophone: a mobile and wearable dataset with in-situ emotion, stress, and attention labels. Sci Data. 2023;10(1):351. doi:10.1038/s41597-023-02248-2
37. Terluin B, Smits N, Brouwers EP, de Vet HC. The Four-Dimensional Symptom Questionnaire (4DSQ) in the general population: scale structure, reliability, measurement invariance and normative data: a cross-sectional survey. Health Qual Life Outcomes. 2016;14(1):130. doi:10.1186/s12955-016-0533-4
38. van Eck M, Berkhof H, Nicolson N, Sulon J. The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol. Psychosom Med. 1996;58(5):447-458. doi:10.1097/00006842-199609000-00007

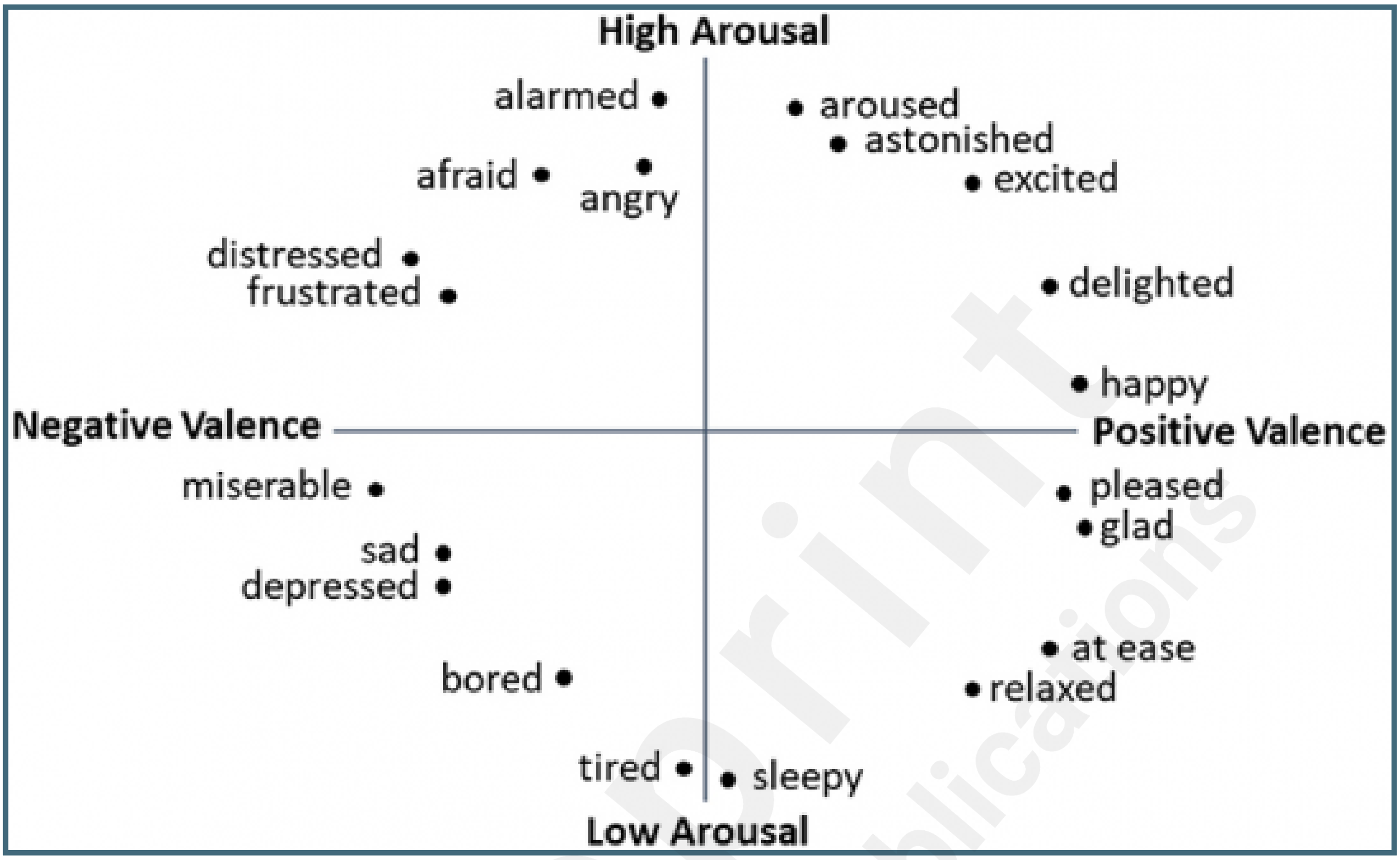
39. Schaufeli WB, van Dierendonck D. UBOS Utrechtse Burnout Schaal: Handleiding. Lisse, the Netherlands: Swets Test Publishers; 2000.
40. Koeter MWJ, Ormel J. General Health Questionnaire: Nederlandse Bewerking. Lisse, the Netherlands: Swets Test Services; 1991.
41. Lamers SM, Westerhof GJ, Bohlmeijer ET, ten Klooster PM, Keyes CL. Evaluating the psychometric properties of the Mental Health Continuum–Short Form (MHC–SF). J Clin Psychol. 2011;67(1):99-110. doi:10.1002/jclp.20741
42. TNO. How am I onderzoek, design-kit. https://www.monitorarbeid.tno.nl/nl-nl/coronacrisis/howami-onderzoek/. Accessed October 9, 2023.
43. Braunsmann L, Beermann F, Strüder HK, Abeln V. Self-selected versus imposed running intensity and the acute effects on mood, cognition, and (a)periodic brain activity. Cogn Neurodyn. 2024;18(5):2221-2241. doi:10.1007/s11571-024-10084-2
44. Survalyzer Survey Software. https://survalyzer.com/.
45. Nunnally JC, Bernstein IH. Psychometric Theory. 3rd ed. New York, NY: McGraw-Hill; 1994.
46. Korten NC, Comijs HC, Penninx BW, Deeg DJ. Perceived stress and cognitive function in older adults: which aspect of perceived stress is important? Int J Geriatr Psychiatry. 2017;32(4):439-445. doi:10.1002/gps.4486
47. Allaire JJ. RStudio: Integrated Development for R. RStudio. http://www.rstudio.com. Accessed April 4, 2024.
48. Langener AM, Bringmann LF, Kas MJ, Stulp G. Predicting mood based on the social context measured through the experience sampling method, digital phenotyping, and social networks. Adm Policy Ment Health. 2024;51(4):455-475.
49. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
50. Neubauer AB, Scott SB, Sliwinski MJ, Smyth JM. How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. J Pers Soc Psychol. 2020;119(1):185. doi:10.1037/pspp0000248
51. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol Assess. 2009;21(3):425-436. doi:10.1037/a0016869
52. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull. 2002;128(6):934-960. doi:10.1037/0033-2909.128.6.934
53. Cella D, Riley W, Stone A, et al; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179-1194. doi:10.1016/j.jclinepi.2010.04.011
54. Joustra ML, Janssens KAM, Schenk HM, Rosmalen JGM. The four week time frame for somatic symptom questionnaires reflects subjective symptom burden best. J Psychosom Res. 2018;104:16-21. doi:10.1016/j.jpsychores.2017.11.006
55. Norquist JM, Girman C, Fehnel S, DeMuro-Mercon C, Santanello N. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. Qual Life Res. 2012;21(6):1013-1020. doi:10.1007/s11136-011-0003-8
56. Mengelkoch S, Moriarity DP, Novak AM, Snyder MP, Slavich GM, Lev-Ari S. Using ecological momentary assessments to study how daily fluctuations in psychological states impact stress, well-being, and health. J Clin Med. 2023;13(1):24. doi:10.3390/jcm13010024
57. Gjoreski M, Gjoreski H, Luštrek M, Gams M. Continuous stress detection using a wrist device: in laboratory and real life. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. New York, NY: ACM; 2016:1185-1193. doi:10.1145/2968219.2968306
58. Mishra V, Hao T, Sun S, et al. Investigating the role of context in perceived stress detection in the wild. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. New York, NY: ACM; 2018:1708-1716. doi:10.1145/3267305.3267537
59. Sano A, Taylor S, McHill AW, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res. 2018;20(6):e210. doi:10.2196/jmir.9410
60. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. Sensors. 2020;20(12):3572. doi:10.3390/s20123572
61. Stamatis CA, Meyerhoff J, Meng Y, et al. Differential temporal utility of passively sensed smartphone features for depression and anxiety symptom prediction: a longitudinal cohort study. NPJ Ment Health Res. 2024;3(1):1. doi:10.1038/s44184-023-00041-y
62. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR

Mhealth Uhealth. 2019;7(10):e14149. doi:10.2196/14149

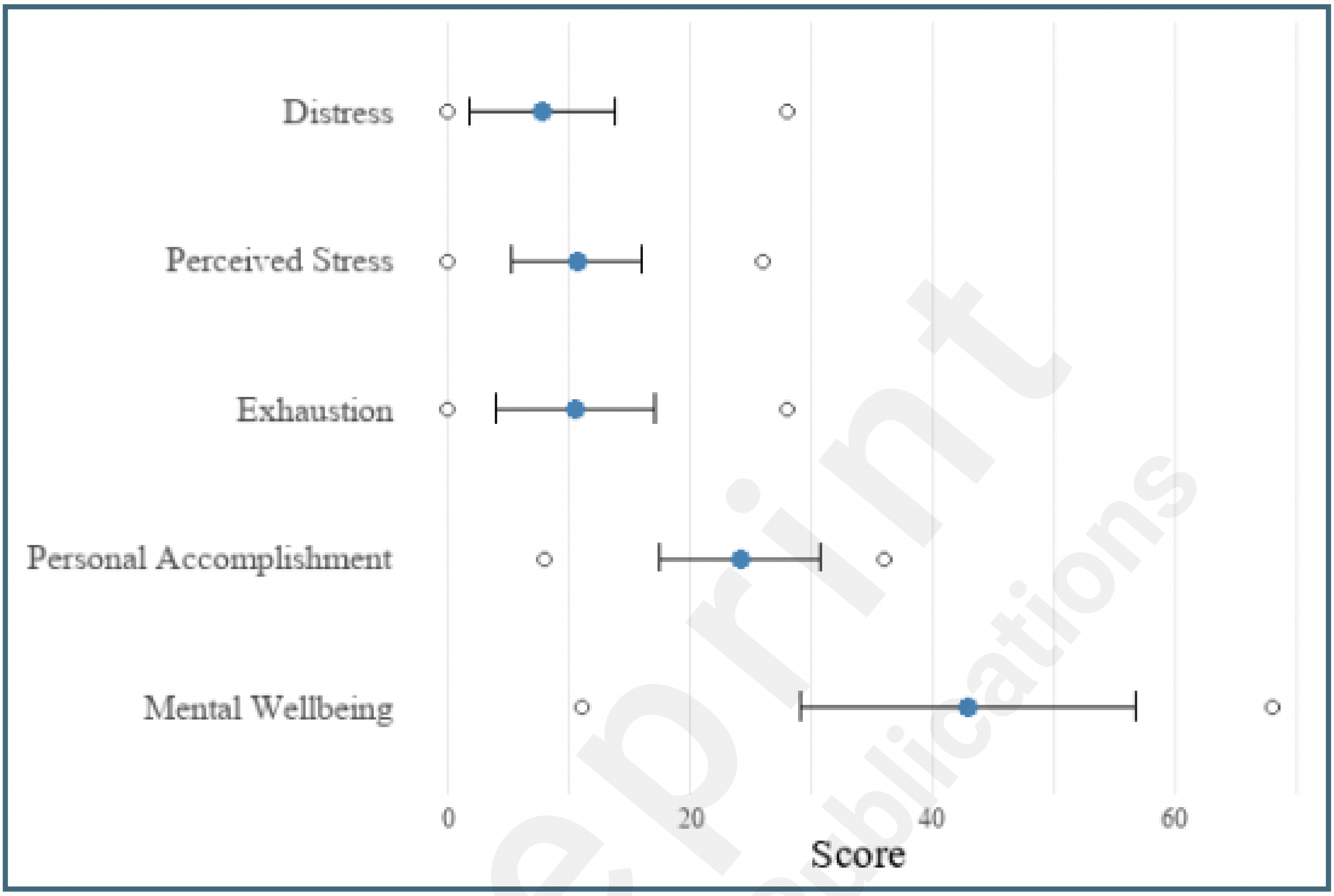
Supplementary Files

Figures

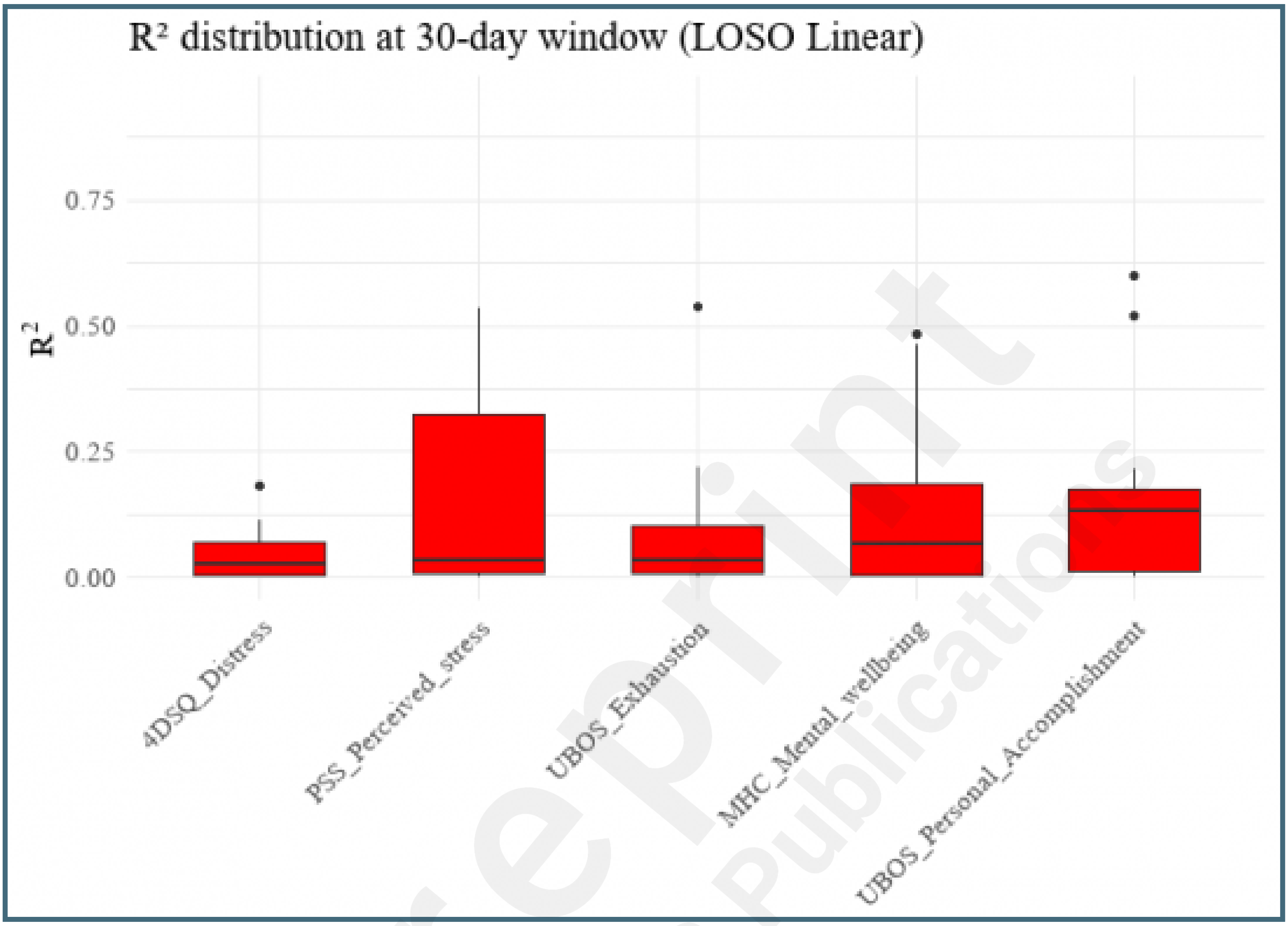
Circumplex Model of Affect (adopted from [9]).



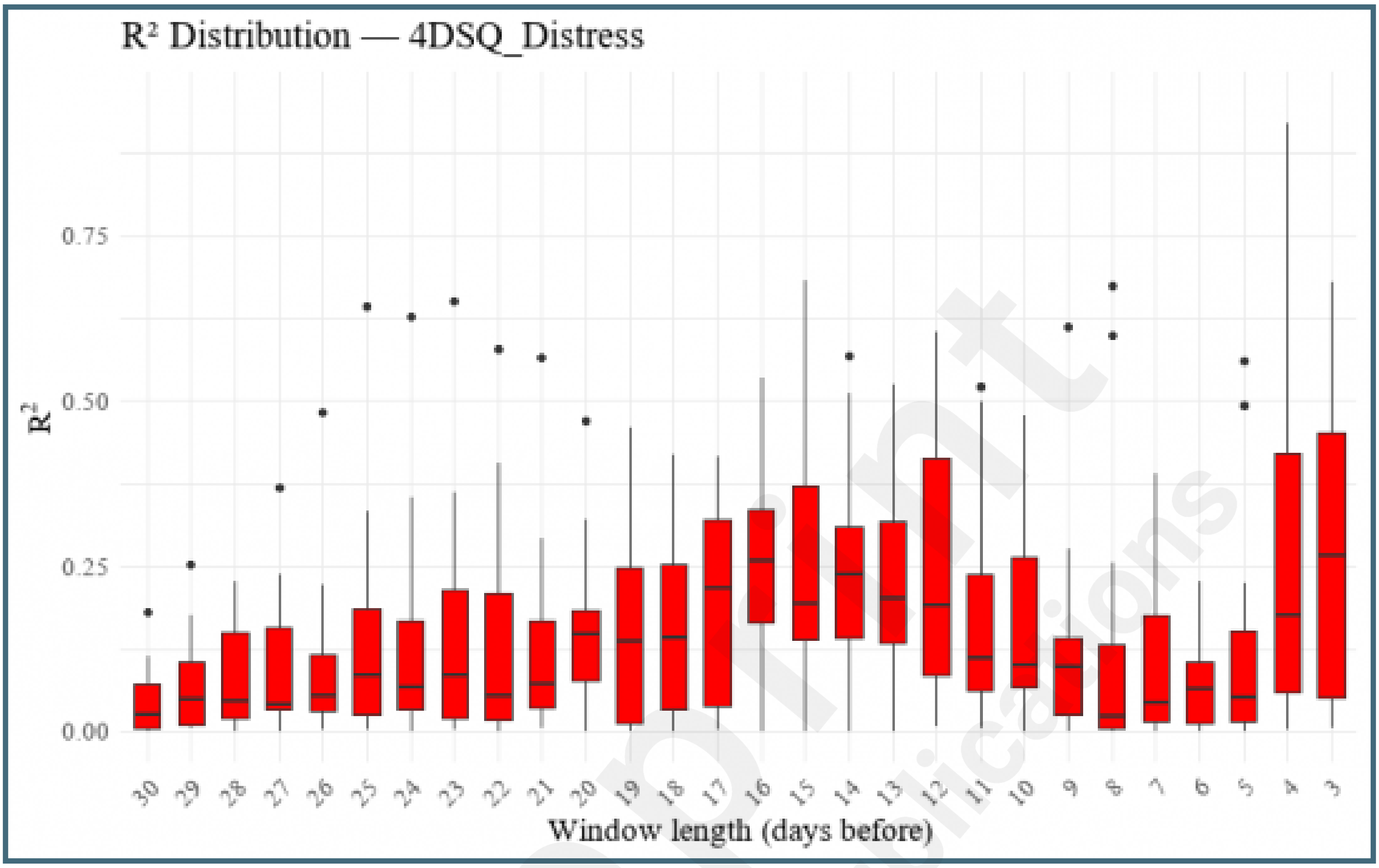
Descriptive statistics per outcome variable.



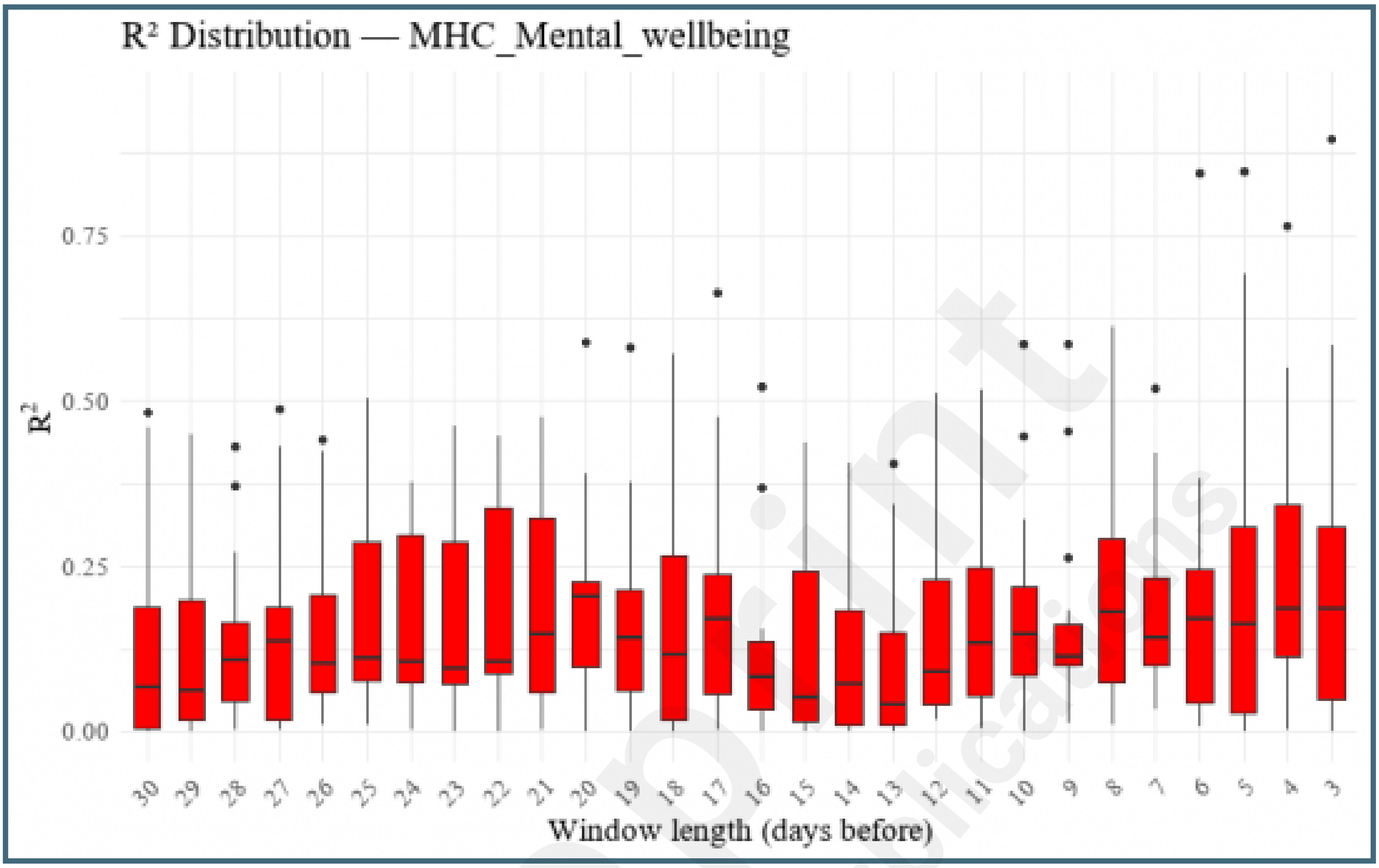
Predictive accuracy (R2) per questionnaire, including all data (time window = 30).



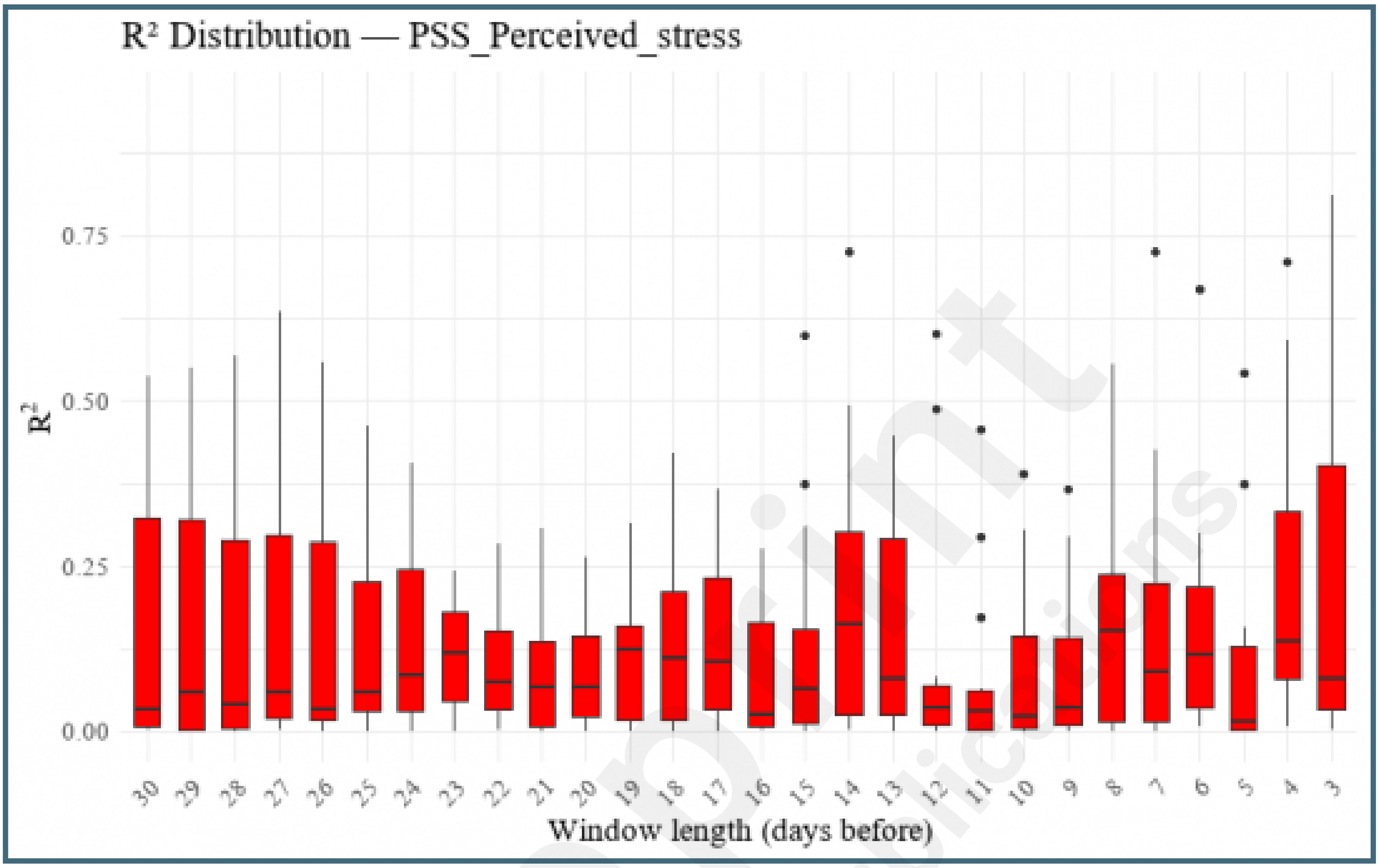
Predictive accuracy (R2) across time windows for 4DSQ Distress.



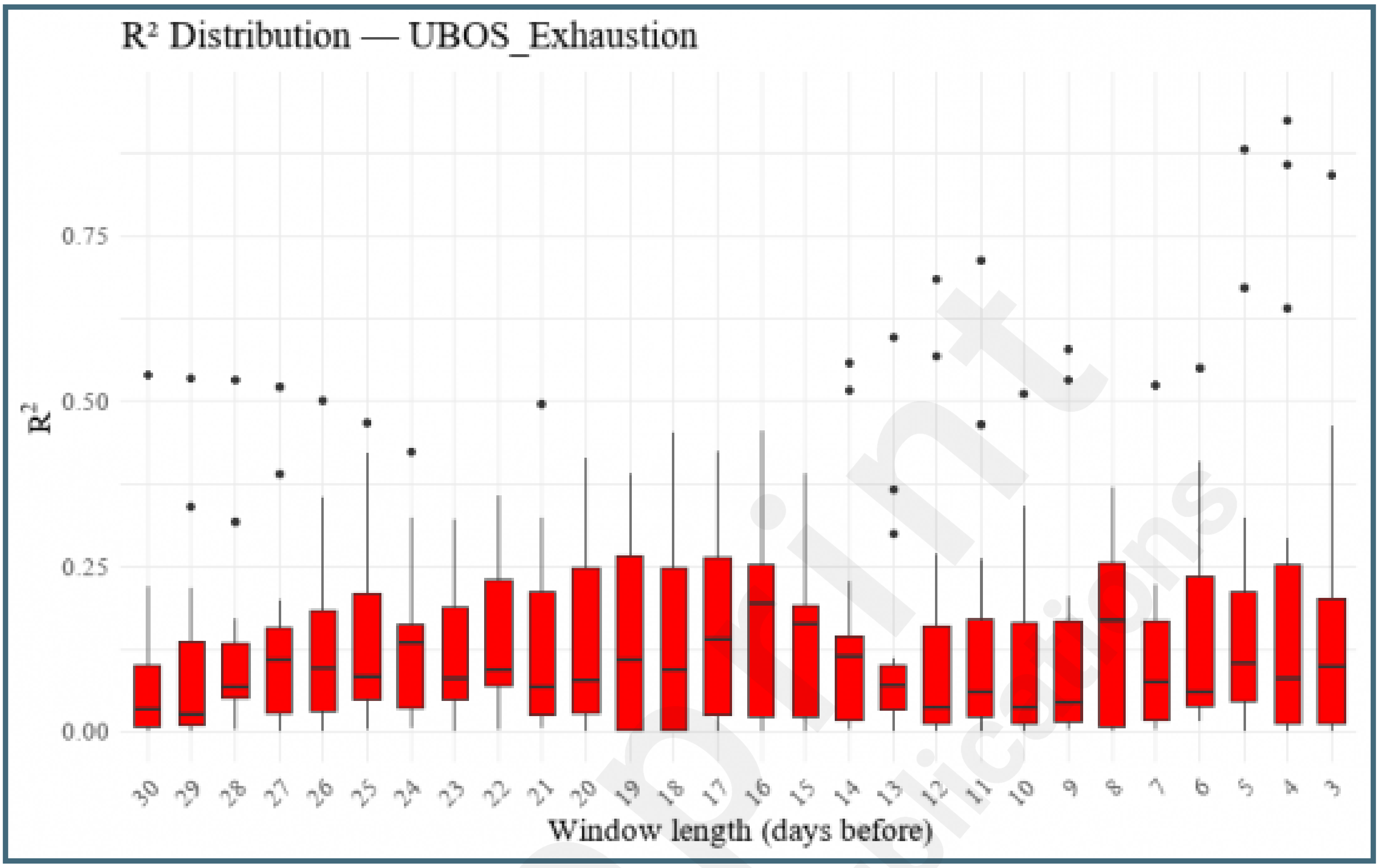
Predictive accuracy (R2) across time windows for MHC Mental wellbeing.



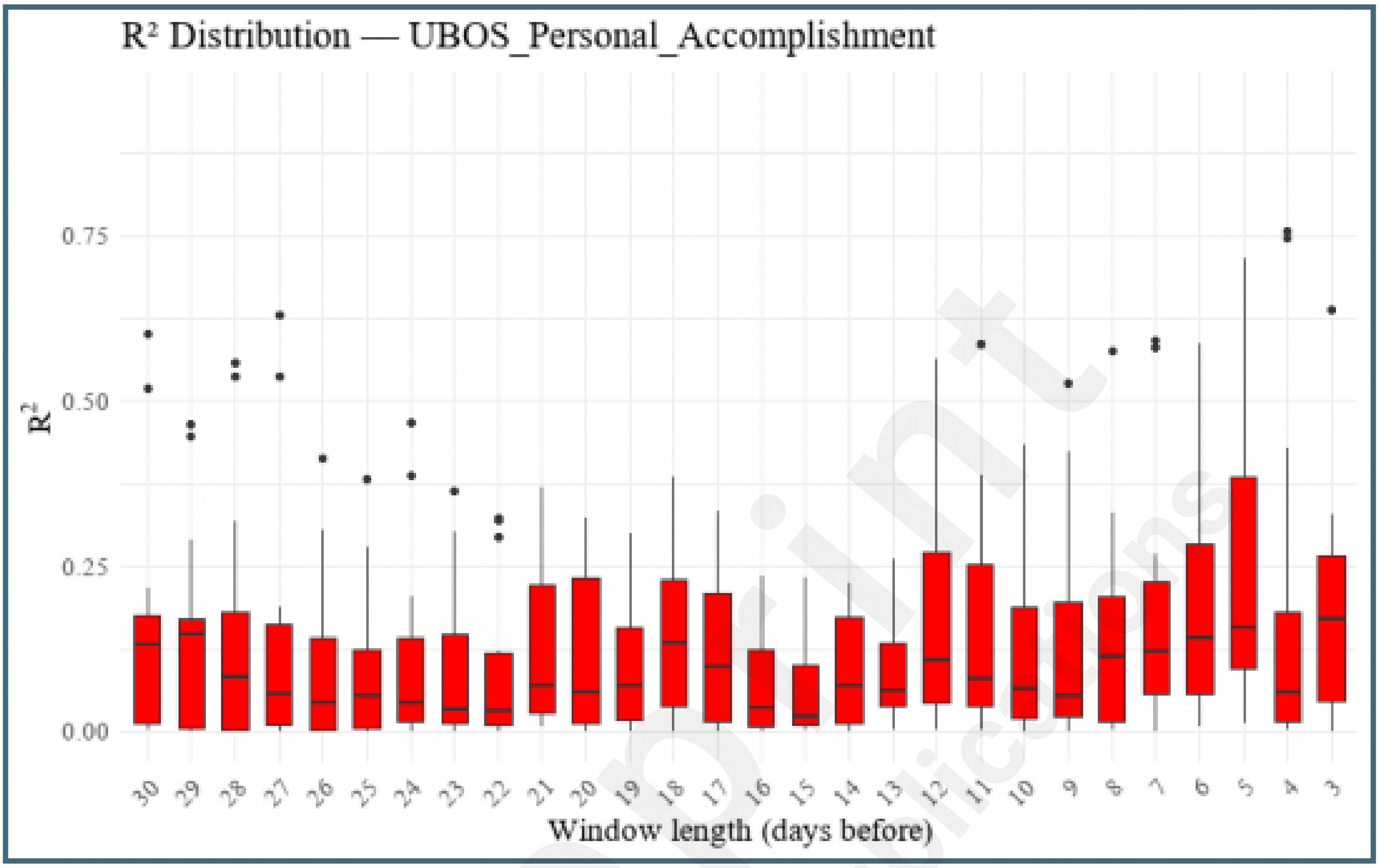
Predictive accuracy (R2) across time windows for PSS Perceived Stress.



Predictive accuracy (R2) across time windows for UBOS Exhaustion.



Predictive accuracy (R2) across time windows for UBOS Personal Accomplishment.



Multimedia Appendixes

Supplemental materials.
URL: http://asset.jmir.pub/assets/cfa8f23298a998ec8f0c723331aa2471.docx