

Original Paper

# Response Time Dynamics From Noncognitive Ordinal Ecological Momentary Assessment as a Proxy for Symptom Change in Geriatric Depression: Longitudinal Observational Study

Joohe Lee<sup>1</sup>, MA; Jeehang Lee<sup>1,2</sup>, PhD; Sehwan Park<sup>1</sup>, MA; Gangho Do<sup>1</sup>, BS; Jihye Noh<sup>3</sup>, MDM; Sangjoon Moon<sup>3</sup>, MS; Kyungmi Chung<sup>3,4,5</sup>, PhD; Sang Joon Son<sup>6\*</sup>, MD, PhD; Jin Young Park<sup>3,4,5\*</sup>, MD, PhD

<sup>1</sup>Medical Research Team, Digital Medic Co., Ltd., Seoul, Republic of Korea

<sup>2</sup>Department of Human-Centered Artificial Intelligence, College of Intelligence Information Engineering, Sangmyung University, Seoul, Republic of Korea

<sup>3</sup>Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, Republic of Korea

<sup>4</sup>Department of Psychiatry, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Republic of Korea

<sup>5</sup>Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Yongin, Republic of Korea

<sup>6</sup>Department of Psychiatry, Ajou University School of Medicine, Suwon, Republic of Korea

\*these authors contributed equally

## Corresponding Author:

Jin Young Park, MD, PhD  
Department of Psychiatry  
Yongin Severance Hospital, Yonsei University College of Medicine  
363, Dongbaekjukjeon-daero, Giheung-gu  
Yongin 16995  
Republic of Korea  
Phone: 82 31-5189-8148  
Fax: 82 31-5189-8565  
Email: [empathy@yuhs.ac](mailto:empathy@yuhs.ac)

## Abstract

**Background:** Depressive symptoms in older adults are amplified by social isolation and limited access to clinic-based mental health care. Ecological momentary assessment (EMA) enables remote self-monitoring and unobtrusively captures response times (RTs), which may serve as indicators of psychomotor and cognitive functioning.

**Objective:** This study investigated the use of EMA-based RT dynamics for predicting symptom change and profiling potential responders for repeated self-monitoring in late-life depression.

**Methods:** Forty-nine community-dwelling adults aged 65 years or older (mean age 70.7, SD 5.8 years; female: 35; male: 14) with a history of major depressive disorder received case management incorporating daily EMA. Participants provided self-reports of mood, appetite, sleep quality, and general well-being. Preassessment and postassessment included the 15-item Short Geriatric Depression Scale (GDS-15), the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R), the 9-item Patient Health Questionnaire, and the Beck Anxiety Inventory. RTs were cleaned with an asymmetric IQR rule,  $z$  standardized within-person  $\times$  response level, and modeled with exponential decay curves over successive EMA trials. The efficacy of EMA-adjunctive care was evaluated using pre-post comparisons of symptom scales. We then examined associations between RT-derived features and symptom change using correlational analyses. Finally, Bayesian multilevel modeling was applied to assess the clinical relevance of RT dynamics, including group differences in adaptation patterns.

**Results:** Older adults at risk for depression showed significant symptom reductions over the 4-week EMA-adjunctive care period across all 4 psychological scales (CESD-R: mean  $\Delta$  11.5; rank-biserial  $r=0.78$ ; GDS-15: mean  $\Delta$  2.14, Cohen  $d=0.76$ ), alongside high EMA adherence (>90%). In correlational analyses, descriptive EMA score metrics and raw RTs showed modest, symptom-specific associations with symptom change ( $\Delta$ CESD-R:  $|r|\approx 0.29$ ;  $\Delta$ 9-item Patient Health Questionnaire:  $|r|\approx 0.32$ ;  $\Delta$ Beck Anxiety Inventory:  $|r|\approx 0.35$ ) but were not significantly related to change in geriatric depression ( $\Delta$ GDS-15:  $|r|\approx 0.24$ ). In contrast, exponential-decay model parameters derived from standardized RT were significantly associated with

geriatric depressive symptom change ( $\Delta$  GDS-15), with the strongest effects observed for the feeling item (eg, decay rate  $\theta_b$ :  $r=-0.398$ , asymptote  $\theta_c$ :  $r=-0.321$ ). Bayesian multilevel modeling further indicated that EMA-adjunctive care responders showed faster RT adaptation than nonresponders (median decay-rate ratio $\approx$ 4.9, 95% credible interval 1.44-14.31), whereas differences in postadaptation RT levels were smaller and uncertain (median postadaptation RT ratio $\approx$ 1.25, 95% credible interval 0.95-1.58). Sensitivity analyses showed consistent decay-rate effects across alternative specifications.

**Conclusions:** Dynamic characteristics of EMA-based RTs emerged as a sensitive proxy for monitoring changes in depressive symptoms among older adults at risk. These findings highlight the potential use of RTs as digital biomarkers derived from brief, low-burden EMA self-monitoring, supporting the development of scalable and personalized mental health interventions for geriatric populations.

*JMIR Aging* 2026;9:e83891; doi: [10.2196/83891](https://doi.org/10.2196/83891)

**Keywords:** ecological momentary assessment; response time; digital biomarkers; geriatric depression; Bayesian multilevel modeling; self-monitoring; mobile health

## Introduction

### *Modern Hazards of Late-Life Depression*

Population aging has brought renewed attention to late-life depression. Growth in single-person households, shrinking family networks, and weaker community ties can intensify the risk of depression in older adults, particularly in highly urbanized settings, such as South Korea [1,2]. These shifts also strain clinic-centered care models, which are often inaccessible to older adults living alone and facing mobility barriers, underscoring the need for remote, daily-life-responsive geriatric mental health strategies.

### *Ecological Momentary Assessment: From Monitoring to Care*

Ecological momentary assessment (EMA) involves the real-time, in situ measurement of behaviors, emotions, and thoughts in everyday contexts [3,4]. By capturing experiences as they occur, EMA reduces recall bias and is well suited to symptoms that fluctuate with context [4]. Advances in mobile and sensor technologies have enabled unobtrusive, scalable EMA deployment, positioning it as a potential infrastructure for practical routine mental health support methods [5,6]. Repeated self-monitoring via EMA may also confer therapeutic benefits, particularly when combined with standard treatments [7-11]; however, effects appear heterogeneous, with reduced responsiveness among individuals with more severe baseline symptoms, highlighting the need to identify who benefits most from EMA-based approaches [12].

### *EMA Response Times: Linking Cognition and Depression in Older Adults*

Recent EMA research has moved beyond self-reported content to “side data” such as response time (RT), which may index underlying cognitive processes. Chung et al [13] reported a nonlinear (inverted U-shaped) association between EMA RT and depression severity, suggesting that both unusually fast and slow responses may reflect dysfunction (eg, impulsivity or cognitive slowing). Hernandez [14] further showed that EMA RTs correlate with Symbol Search

performance, indicating that RTs capture general processing capacity even in noncognitive EMA tasks.

EMA-derived RTs may therefore help identify individuals vulnerable to depressive symptoms. Consistent with *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)* descriptions of major depressive disorder, which include observable psychomotor retardation [15], Hernandez et al [16] applied drift diffusion modeling to binarized EMA responses and found that RT-derived parameters (eg, drift rate and boundary separation) systematically tracked traits such as neuroticism and depressive symptoms [16]. Together, these findings position EMA RT as a candidate marker linking cognitive efficiency with affective symptomatology.

This approach may be particularly relevant in geriatric populations, where aging and depression jointly exacerbate psychomotor slowing [17-19]. Older adults with major depressive disorder (MDD) show prolonged initiation and movement times relative to controls, implicating combined executive and motor delays [17], and reviews report especially large effects for psychomotor slowing in late-life samples [20]. Depressed older adults also appear less adaptable under repeated cognitive demands: greater symptom severity has been associated with slower reaction-time learning curves and reduced improvement across training blocks, with sluggish adaptation predicting later functional decline [21,22]. Some interventions may even backfire when effort exceeds coping capacity; for instance, one randomized controlled trial reported worsening depressive symptoms following speed-of-processing training [23]. Collectively, these results motivate EMA-based RT metrics as scalable markers for monitoring cognitive-motor slowing and adaptation in late-life depression.

### *Challenges in Utilizing Ordinal-Scale, Noncognitive EMA RTs*

Despite its potential usage, analyzing RTs from ordinal-scale EMA items presents distinct hurdles. While ordinal items, such as 5- or 7-point Likert scales, offer richer granularity than binary formats, their integration with RT data

poses additional analytical considerations and requires further preprocessing steps.

One complication arises from emotional clarity. As shown in Hernandez et al [16], individuals with clearer emotional awareness due to intense emotions tend to respond more quickly to EMA questionnaires. Thus, trait-level factors, such as baseline emotional reactivity, as well as situational factors, including bereavement, acute illness, and exposure to daily stressors, may affect the RT of EMA reports. This can mask or confound the expected slowing associated with cognitive impairment, making RT a less straightforward marker when emotional clarity varies across individuals and between occasions.

A further complication lies in interface variability. Mobile EMA apps often use different visual formats (eg, sliders, buttons, or scales), which can unintentionally influence how quickly users respond. For example, a response option that is most proximal to a frequently used finger can lead to faster responses. These app-specific differences, such as user interface composition, introduce another layer of noise that must also be accounted for when analyzing ordinal-scaled EMA-based RT.

## Research Purpose

In light of these issues, this study pursues the following objectives. First, we tested whether short-term EMA use ( $\leq 4$  wk) alongside usual case management is associated with symptom improvement among older adults at elevated risk for depression, evaluating EMA as a potential adjunctive intervention rather than a monitoring-only tool. Second, we examined how multiple EMA-derived features (eg, symptom ratings, response speed, and RT adaptation) relate to symptom change over time, with a focus on RT dynamics as a behavioral signature of improvement. We also introduce a scalable analytic approach that reduces key ecological confounds in RT (eg, momentary affect intensity and interface-related heterogeneity) and is broadly applicable across EMA apps and Likert-type scales. Third, we evaluated the clinical relevance of RT adaptation by using a Bayesian multilevel model to compare adaptive patterns between participants who benefited from EMA-adjunctive care and those who did not. With a modest sample size ( $n=50$ ), partial pooling in a multilevel framework improves estimate stability while accommodating missing and unbalanced EMA observations [24]. Additionally, Bayesian posterior inference provides interpretable uncertainty for group differences in adaptation parameters, thereby quantifying the evidence for between-group differences in adaptive patterns [24-26]. This research ultimately supports the objective, remote stratification of individuals who are most likely to benefit from EMA-enabled care.

## Methods

### Eligibility Criteria

Participants eligible for this study were aged 65 years or older, regardless of gender, and had expressed their willingness to participate through recruitment notices posted at the community-based Suwon Geriatric Mental Health and Welfare Center in Gyeonggi-do, Republic of Korea. Individuals who met the following eligibility criteria were included: (1) participants must have had a prior diagnosis of depressive disorder by a psychiatrist, with no change in medication for the last 3 months and without symptom aggravation; (2) absence of medical or neurological disorders such as dementia, stroke, or Parkinson disease that may affect the research; and (3) current recipients of case management services at the Suwon Geriatric Mental Health and Welfare Center.

### Ethical Considerations

All participants who met these criteria provided written informed consent prior to enrollment and completed demographic information and preassessment questionnaires. Those who completed these surveys were then given detailed guidance and support for installing and using the app, with additional assistance provided to those less familiar with digital tools. Given the specific population's limited experience with digital devices, only individuals who demonstrated the ability to follow the mobile app (BIG4+) usage guidelines independently were enrolled. Participation was entirely voluntary, and participants were free to withdraw at any time without any disadvantage. All procedures were approved by the Public Institutional Bioethics Committee designated by the Ministry of Health and Welfare of the Korean Government (MOHW; P01-202405-01-019) and were supported by Digital Medic Inc. Written informed consent was obtained from all participants. Participant privacy was protected by analyzing all data in deidentified form and restricting access to authorized research staff only. Electronic data were stored in password-protected encrypted files, and physical records were kept in locked storage. Participants received KRW 100,000 (approximately US \$67.50) for completing the full study protocol. For early withdrawal, compensation was provided on a prorated basis according to the week of completion.

### Research Design

The study used an observational, single-arm design. Participants were community-dwelling older adults enrolled through the Suwon Geriatric Mental Health and Welfare Center.

Case management was provided throughout the study, consisting of 1 monthly home visit and 1 to 2 interim telephone contacts, during which case managers monitored participants' daily living conditions and mental health status. In addition to this standard case management, participants were asked to use the Big4+ mobile app as an adjunctive care tool over a 4-week period. The Big4+ app, developed

specifically for research purposes, was available on both Android (AOS) and iOS platforms and collected daily self-reports on mood, appetite, sleep quality, and overall well-being using a 7-point Likert scale (1=minimum to 7=maximum).

Each participant visited the center twice: once prior to initiating app usage (baseline assessment, V1) and once at the end of the 4-week period (postassessment, V2). During both visits, participants completed a battery of self-report questionnaires administered in paper format by case managers, with each session lasting approximately 30 minutes.

Standardized psychological scales administered to participants included the Korean version of the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R) [27-29], the Korean version of the 15-item Short Geriatric Depression Scale (GDS-15) [30,31], the 9-item Patient Health Questionnaire-9 (PHQ-9) [32,33], and the Beck Anxiety Inventory (BAI) [34,35].

## Data Preprocessing

### Handling RT Outliers

To the best of our knowledge, there is no established gold standard for identifying outliers in RTs associated with noncognitive and ordinal EMA items. Furthermore, the population of this study (older adults at high risk for depression) was highly specific, making it necessary to define acceptable RTs empirically based on the distribution within the dataset. We adopted the IQR rule, where the IQR is calculated as the difference between the third quartile and the first quartile. Outlier boundaries were defined as specific multiples of the IQR from Q1 and Q3. Given that RT distributions were heavily right-skewed, we applied asymmetric boundaries across the dataset: a lower bound of  $Q1 - 0.5 \times IQR$  and an upper bound of  $Q3 + 1.5 \times IQR$ . These boundaries were calculated separately for each EMA item (feeling, appetite, feeling of sleep, and general evaluation). Trials with RTs falling outside these bounds were excluded from the analysis. The details of the predefined outlier thresholds and the decision-making process are provided in Section I in [Multimedia Appendix 1](#).

### Standardization of RTs

EMA RTs are noisy and can be confounded by momentary affect (eg, faster responses during more intense or clearly defined states) and device or app characteristics [16]. To improve comparability, we standardized RTs within individuals and response options, expressing each latency relative to that participant's typical speed at a given Likert level.

For individual  $i$ , response option  $j$  (1-7), and trial  $k$ , standardized RT (Z-RT) was as follows:

$$Z_{ij}^{(k)} = \frac{RT_{ij}^{(k)} - \bar{RT}_{ij}}{\sigma_{ij}},$$

where  $RT_{ij}$  and  $\sigma_{ij}$  are the mean and SD of RT for participant  $i$  at option  $j$ . This z-transformation reduces between-context and between-person variability, supporting more reliable downstream analyses.

## Extracting EMA-Based Digital Features

Descriptive features were extracted from 4 weeks of EMA data collected from each participant. For each EMA item, we computed the mean, SD, minimum, maximum, and interquartile values (25th, 50th, and 75th percentiles) of the reported scores to characterize overall mood patterns throughout the study period. In addition, identical descriptive statistics of EMA-based RTs were derived as indicators of cognitive processing speed and consistency.

Beyond overall response latency, we quantified how participants adapted to repeated EMA self-reflection over time. Such "practice effects" (ie, progressive reductions in RT with repeated exposure) are widely documented in cognitive tasks and are relevant to EMA given evidence that even noncognitive RTs reflect cognitive traits [16]. Although practice effects are often modeled with power-law decay [36], exponential decay may better capture adaptation in older adults, who typically show slower motor and cognitive adjustment [37,38]. Accordingly, we fit an exponential decay curve to each participant's longitudinal RT series (both raw RT and standardized RT):

$$f(t) = \theta_a \cdot \exp(-\theta_b \cdot t) + \theta_c,$$

where  $t$  indexes time or trials. Parameters were estimated via nonlinear least squares (Levenberg-Marquardt; `scipy.optimize.curve_fit`; max 10,000 iterations). The parameter  $\theta_a$  (amplitude) captures the initial elevation in RT above the asymptote (ie,  $f(0) - \theta_c$ ), reflecting the potential magnitude of improvement across repeated trials. The parameter  $\theta_b$  (decay rate) indexes the speed of adaptation, indicating how rapidly an individual becomes accustomed to repeated self-monitoring. The parameter  $\theta_c$  (asymptote or postadaptation RT) represents the asymptotic minimum RT, interpreted as the stabilized latency after maximal adaptation within the observation window. The model was fit to both raw and standardized RTs, with parameter bounds set to match the scale and range of each outcome.

In this stage, exponential decay parameters were estimated separately for each participant (ie, without partial pooling) to serve as person-level features that can be computed in naturalistic settings where no external reference data are available. This individual-fit approach also facilitates a fair comparison with other EMA-derived features (eg, self-report scores and RT descriptives), which were likewise calculated at the individual level without borrowing strength from group-level information. Technical details regarding the fitting of the exponential-decay model are provided in Section II in [Multimedia Appendix 1](#).

## Defining Symptom Improvement

A key step prior to the analyses was to select an appropriate metric for symptom change. Although treatment response is often operationalized using simple change scores (post-pre), this approach has been criticized for its dependence on baseline severity and the resulting negative correlation with baseline values [39-41]. Percentage change scores adjust for baseline differences and are widely used in other clinical domains [42,43] but can be difficult to interpret and may reduce statistical power due to skewed distributions [40,41]. Regression-based alternatives (eg, analysis of covariance or residualized change) can control for baseline levels yet may yield biased estimates when baseline severity is related to treatment effects, which is common in observational settings [40,41]. Given these considerations and prior methodological work supporting the robustness of simple change scores in nonrandomized contexts [40,44,45], we used simple change score  $\Delta$ (post-pre) as the primary symptom improvement outcome.

## Responder Classification (Minimally Detectable Change Criteria)

Participants were classified as “responders” or “nonresponders” to EMA-adjunctive care based on minimally detectable change (MDC) thresholds derived from the GDS-15. A recent meta-analysis by González-Domínguez et al [46], which included 5876 older adults across 21 studies (mean age 76 y), reported that a score reduction of 3.81 points on the GDS-15 corresponds to the MDC threshold, defined as the level of change with only a 5% probability of being attributable to measurement error (95% CI 3.59-4.04) [47]. Based on these findings, an MDC threshold of -3.81 points was applied, reducing the arbitrariness of classification and accounting for measurement error.

## Analysis

### Explorative Effects of EMA-Based Self-Monitoring

Explorative analysis was conducted to assess the benefits of EMA-adjunctive care in older populations. Given the single-arm design, the effectiveness of EMA-based self-monitoring in geriatric depression was assessed by comparing symptom scores collected at baseline (V1) and postintervention (V2) using paired statistical tests. The choice of test was determined based on the normality of the difference scores, as assessed by the Shapiro-Wilk test. When the assumption of normality was satisfied, a paired-samples  $t$  test was applied. In cases where the normality assumption was violated, the nonparametric Wilcoxon signed-rank test was used instead. All paired tests in this analysis were 1-tailed.

### Association Between EMA-Derived Features and Symptom Improvement

We examined whether EMA-derived features, particularly RT dynamics, track symptom improvement using correlational analyses. We computed Spearman correlations between

EMA-derived features and  $\Delta$  on the GDS-15 and assessed specificity using parallel analyses for the CESD-R, PHQ-9, and BAI.

## Clinical Relevance of EMA Trajectories for Symptom Improvement

To examine the clinical relevance of EMA-based RT dynamics, we fitted a moderated Bayesian multilevel exponential-decay model to test whether adaptive patterns of EMA-based RTs differed by the MDC group. The model equation is demonstrated as a fully randomized model. For detailed model specifications and rationales, see Section III in [Multimedia Appendix 1](#).

### Observation Model

For each observation  $n=1, \dots, N_{\text{obs}}$ ,

$$RT_n \sim \text{LogNormal}(\eta_n, \sigma_{\log}) \Leftrightarrow \log(RT_n) \sim \mathcal{N}(\eta_n, \sigma_{\log}^2),$$

we assumed each observation follows a log-normal distribution, ensuring positive predictions and accommodating the right-skew typical in RT data.

### Mean Structure (RT Adaptation)

Expected log-RT followed a subject-specific exponential decay over normalized time  $t \in [0, 1]$ :

$$\eta_n = \theta_a^{(i)} \cdot \exp(-\theta_b^{(i)} \cdot t) + \theta_c^{(i,r)},$$

where  $i$  indexes subjects and  $r \in 1, \dots, R$  indexes response options (Likert-7 scale).

### Multilevel Effects and MDC Moderation

Amplitude and decay rate were modeled on the log scale (log-linked) for positive constraints and allowed to vary by subject, with MDC group moderation:

$$\begin{aligned} \log \theta_a^{(i)} &= \beta_{0a} + \beta_{1a} \cdot g^{(i)} + \sigma_a \cdot z_a^{(i)}, & z_a^{(i)} &\sim \mathcal{N}(0, 1), \\ \log \theta_b^{(i)} &= \beta_{0b} + \beta_{1b} \cdot g^{(i)} + \sigma_b \cdot z_b^{(i)}, & z_b^{(i)} &\sim \mathcal{N}(0, 1). \end{aligned}$$

Baseline log-RT or postadaptation RT ( $\theta_c^{(i,r)}$ ) was decomposed into group, subject, and response-option components:

$$\theta_c^{(i,r)} = \underbrace{(\beta_{0c} + \beta_{1c} \cdot g^{(i)})}_{\text{Group-Moderated Baseline}} + \underbrace{u_c^{(i)}}_{\text{Subject Intercept}} + \underbrace{\alpha_c^{(r)}}_{\text{Option Main Effect}} + \underbrace{\epsilon_c^{(i,r)}}_{\text{Cell Residual}}$$

with  $u_c^{(i)} \sim \mathcal{N}(0, \sigma_{\text{subj}}^2)$ ,  $\epsilon_c^{(i,r)} \sim \mathcal{N}(0, \sigma_{\text{cell}}^2)$ , and  $\sum_{r=1}^R \alpha_c^{(r)} = 0$  for identifiability. The decomposition of baseline log-RT or postadaptation RT aligns with the framework advocating for the within-person and within-response option standardization of RTs. By explicitly modeling these components, the model accounts for response-option biases in observed latency while simultaneously partitioning subject-specific baseline variance.

## Priors and Inference

We used weakly informative priors  $\beta_{0*} \sim \mathcal{N}(\mu_{\text{emp}}, 3^2)$ ,  $\beta_{1*} \sim \mathcal{N}(0, 3^2)$ ,  $\alpha_c^{(r)} \sim \mathcal{N}(0, 1^2)$ ,  $\sigma_a$ ,  $\sigma_b$ ,  $\sigma_{\text{subj}}$ ,  $\sigma_{\text{cell}} \sim \text{HalfNormal}(1.5)$ , and  $\sigma_{\log} \sim \text{HalfNormal}(1)$ . Models were fit in PyMC using NUTS (4 chains; 1000 warm-up + 1000 draws per chain; target acceptance = 0.95) [48].

To determine the appropriate degree of hierarchical structure in EMA-based RT trajectories, we compared a set of nested Bayesian exponential-decay models that differed only in their random-effects specification. Specifically, we evaluated (1) a nonhierarchical model with fixed parameters across participants, (2) partially hierarchical variants allowing subject-level variation in selected parameters (eg, random  $\theta_b$  only; random  $\theta_a$  and  $\theta_b$ ; random  $\theta_b$  and  $\theta_c^{(l,r)}$ ), and (3) a full hierarchical model with subject-level variation in key components. Models were compared using Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO), summarizing out-of-sample predictive performance via the expected log pointwise predictive density leave-one-out cross-validation (ELPD-LOO). To isolate the contribution of random effects, MDC-group moderation terms were omitted during this model-structure comparison. Based on the best-performing random-effects structure (selected model; random  $\theta_b$  and  $\theta_c$ , with  $\theta_a$  fixed across subjects), we then tested clinical relevance by extending the selected model to include MDC group moderation, allowing these parameters to vary systematically between responders and nonresponders.

Group differences were quantified using Bayesian posterior summaries rather than frequentist hypothesis tests. For log-linked parameters (eg,  $\theta_a$ ,  $\theta_b$ ), effects were expressed as responder to nonresponder multiplicative ratios  $\exp(\beta_1)$ ; for baseline parameters on the log-RT scale (eg,  $\theta_c$ ), group effects were additive on the log scale and additionally converted to RT ratios via exponentiation. We report the posterior median of the responder-to-nonresponder ratio, computed draw-by-draw from the posterior samples, along with the 95% credible interval (CrI) for this ratio. A ratio of 1 indicates no difference between the groups on the original scale; ratios greater than 1 indicate larger parameter values in responders (eg, a ratio of 1.30 corresponds to ~30% higher values), whereas ratios less than 1 indicate smaller values in

responders relative to nonresponders. The 95% CrI represents the range containing 95% of the posterior probability for the quantity (given the specified model and priors). We also report the posterior probability of an increase,  $P(\text{Ratio}_{\text{Responder/Nonresponder}} > 1 | \text{data})$ , which quantifies the evidence that responders have larger parameter values than nonresponders.

Models were fit separately for each EMA item and additionally for an averaged RT outcome across items; for the averaged outcome, continuous scores were discretized into 7 bins to align with the response-option structure used in item-level models. Finally, we conducted sensitivity analyses by refitting the model under alternative weakly informative priors (Student  $t$ ) and a more complex plausible random-effects specification and verified that key group effect conclusions were robust across these reasonable modeling choices.

## Results

### Participants

A total of 50 older adults aged 65 years or older were recruited for the study (mean age 70.6, SD 5.8 y), of whom 72% (n=35) were female. Among the participants, 49 had a documented history of MDD, and 1 had a history of bipolar disorder. To ensure consistency and avoid confounding effects arising from the natural mood fluctuations associated with bipolar disorder, only participants with MDD were included in all statistical analyses (n=49). Baseline psychological assessments indicated clinically elevated symptoms across multiple domains. Preassessment scores were as follows: GDS-15 mean=9.37 (SD 4.42), CESD-R mean=33.94 (SD 18.24), PHQ-9 mean=13.04 (SD 7.40), and BAI mean=22.76 (SD 16.46). Following the study period, scores were as follows: GDS-15 mean=7.22 (SD 4.44), CESD-R mean=22.65 (SD 16.15), PHQ-9 mean=8.59 (SD 6.30), and BAI mean=13.31 (SD 11.52). All participants successfully installed and used the BIG4+ app over the 4-week period, with a mean compliance rate above 93%. All descriptive details are summarized in Table 1.

**Table 1.** Descriptives of demographics, psychological assessments, and mobile app (BIG4+) compliance (N=49).

Descriptives	Values
Gender, n (%)	
Women	35 (71.4)
Men	14 (28.6)
Age (y), mean (SD)	70.7 (5.8)
Operating system, n (%)	
Android	49 (100)
iOS	0 (0)
Income level, n (%)	
Lowest	0 (0)
Low	6 (12.2)

Descriptives	Values
Middle	15 (30.6)
High	8 (16.3)
Highest	18 (36.7)
No response	2 (4.1)
Residence type, n (%)	
Living with family	20 (40.8)
Living alone	29 (59.2)
Relationship with family, n (%)	
Very bad	5 (10.2)
Bad	8 (16.3)
Moderate	16 (32.7)
Good	15 (30.6)
Very good	4 (8.2)
No response	1 (2.0)
Intimate person present, n (%)	
Yes	28 (57.1)
No	19 (38.8)
No response	2 (4.1)
Belonging group present, n (%)	
None	15 (30.6)
1 group	19 (38.8)
2 groups	8 (16.3)
3 groups	5 (10.2)
Above 4 groups	2 (4.1)
Preassessment scale (V1), mean (SD)	
GDS-15 <sup>a</sup>	9.4 (4.4)
CESD-R <sup>b</sup>	33.9 (18.2)
PHQ-9 <sup>c</sup>	13 (7.4)
BAI <sup>d</sup>	22.8 (16.5)
Postassessment scale (V2), mean (SD)	
GDS-15	7.2 (4.4)
CESD-R	22.7 (16.1)
PHQ-9	8.6 (6.3)
BAI	13.3 (11.5)
BIG4+ adherence (%), mean (SD)	93.8 (9.1)

<sup>a</sup>GDS: 15-item Geriatric Depression Scale.

<sup>b</sup>CESD-R: Center for Epidemiologic Studies Depression Scale-Revised.

<sup>c</sup>PHQ-9: 9-item Patient Health Questionnaire.

<sup>d</sup>BAI: Beck Anxiety Inventory.

## EMA-Adjunctive Care and Short-Term Symptom Improvement in at-Risk Older Adults

To evaluate the benefit of EMA-based self-monitoring, preassessment and postassessment scores were compared across 4 psychological scales (GDS-15, CESD-R, PHQ-9, BAI). The results indicated that CESD-R, PHQ-9, and BAI scores violated the assumption of normality ( $P$  values = .01, .01, and .01, respectively), while the GDS-15 difference scores did not ( $P=.33$ ). Consequently, Wilcoxon signed-rank tests were applied to the CESD-R, PHQ-9, and BAI, and a paired-samples  $t$  test was used for the GDS-15 (Table 2).

All 4 measures showed statistically significant improvements following the intervention. CESD-R scores decreased by a mean of 11.50 points (SD 15.8; SE 2.26;  $W=1045.00$ ;  $P<.001$ ), with a rank biserial correlation of 0.78. PHQ-9 scores decreased by a mean of 4.50 points (SD 6.9; SE 1.00;  $W=947.00$ ;  $P<.001$ ), with an effect size of 0.75. BAI scores were reduced by 9.00 points (SD 11.7; SE 1.67;  $W=940.00$ ;  $P<.001$ ), with a rank biserial correlation of 0.82. GDS-15 scores showed a significant reduction of 2.14 points (SD 2.8; SE 0.40;  $t_{48}= 5.30$ ;  $P<.001$ ), corresponding to a Cohen  $d$  of 0.76.

**Table 2.** Paired tests of psychological assessment scores between preintervention (V1) and postintervention (V2).

Scales	Test	Statistics	<i>P</i> value	Rank biserial correlation	Cohen <i>d</i>
GDS-15 <sup>a</sup>	Student <i>t</i>	5.30	<.001	— <sup>b</sup>	0.76
CESD-R <sup>c</sup>	Wilcoxon <i>W</i>	1045.0	<.001	0.78	—
PHQ-9 <sup>d</sup>	Wilcoxon <i>W</i>	947.0	<.001	0.75	—
BAI <sup>e</sup>	Wilcoxon <i>W</i>	940.0	<.001	0.82	—

<sup>a</sup>GDS-15: 15-item Geriatric Depression Scale.

<sup>b</sup>Not applicable.

<sup>c</sup>CESD-R: Center for Epidemiologic Studies Depression Scale-Revised.

<sup>d</sup>PHQ-9: 9-item Patient Health Questionnaire.

<sup>e</sup>BAI: Beck Anxiety Inventory.

## Associations Between EMA-Derived Features and Symptom Improvement

Linear relationships between EMA-derived features and changes in depressive and anxiety symptoms were examined using correlational analyses. A full table of correlational analyses is presented in [Multimedia Appendix 2](#).

Descriptive statistics of EMA scores showed generally small-to-modest associations with mental health change scores ([Table 3](#)), with no significant correlations for

$\Delta$ GDS-15 across items. Significant associations emerged for other outcomes, suggesting item-specific links with symptom change:  $\Delta$ CESD-R was correlated with the minimum of appetite ( $r=0.294$ ;  $P=.04$ ) and general evaluation ( $r=0.29$ ;  $P=.04$ ).  $\Delta$ PHQ-9 was correlated with the median of feeling ( $r=0.291$ ;  $P=.04$ ), mean ( $r=0.325$ ;  $P=.02$ ), and minimum of appetite rating ( $r=0.291$ ;  $P=.04$ ), and the minimum of general evaluation ( $r=0.286$ ;  $P=.046$ ). For anxiety,  $\Delta$ BAI was associated only with the minimum sleep quality rating ( $r=0.335$ ;  $P=.02$ ).

**Table 3.** Associations between mental health change scores ( $\Delta$ ), EMA score, and EMA response time descriptives.<sup>a</sup>

Type, EMA <sup>b</sup> item, and feature	$\Delta$ GDS-15 <sup>c</sup> ( $r^p$ )	$\Delta$ CESD-R <sup>d</sup> ( $r^p$ )	$\Delta$ PHQ-9 <sup>e</sup> ( $r^p$ )	$\Delta$ BAI <sup>f</sup> ( $r^p$ )
EMA score				
Feeling				
Mean	0.094	0.239	0.251	0.204
SD	0.074	-0.246	-0.193	-0.255
Minimum	-0.01	0.267	0.234	0.26
Median	0.039	0.252	0.291*	0.234
Maximum	0.233	0.001	0.062	0.025
Appetite				
Mean	0.241	0.273	0.325*	0.152
SD	0.145	-0.221	-0.136	-0.163
Minimum	0.013	0.294*	0.291*	0.14
Median	0.193	0.245	0.264	0.107
Maximum	0.263	0.006	0.086	-0.07
Sleep quality				
Mean	0.119	0.074	0.197	0.16
SD	-0.004	-0.231	-0.123	-0.192
Minimum	-0.06	0.245	0.277	0.335*
Median	0.116	0.06	0.141	0.085
Maximum	0.028	0.00	0.038	0.042
General evaluation				
Mean	0.121	0.168	0.198	0.149
SD	0.016	-0.26	-0.196	-0.203
Minimum	0.029	0.29*	0.286*	0.261
Median	0.106	0.156	0.175	0.162
Maximum	0.024	0.072	0.04	0.101
EMA response time				
Feeling				
Mean	0.091	-0.073	0.2	-0.108

Type, EMA <sup>b</sup> item, and feature	$\Delta$ GDS-15 <sup>c</sup> ( $r^p$ )	$\Delta$ CESD-R <sup>d</sup> ( $r^p$ )	$\Delta$ PHQ-9 <sup>e</sup> ( $r^p$ )	$\Delta$ BAI <sup>f</sup> ( $r^p$ )
SD	0.092	0.074	0.091	-0.146
Minimum	0.099	-0.028	0.17	-0.039
Median	0.101	-0.129	0.175	-0.164
Maximum	0.107	0.103	0.181	-0.05
Appetite				
Mean	0.195	-0.122	0.037	-0.301*
SD	0.008	-0.141	-0.077	-0.282
Minimum	0.105	-0.114	0.04	-0.158
Median	0.229	-0.053	0.077	-0.261
Maximum	0.044	-0.159	-0.049	-0.35*
Sleep quality				
Mean	0.251	-0.067	0.011	-0.224
SD	0.189	-0.003	-0.106	-0.023
Minimum	0.17	-0.104	0.053	-0.139
Median	0.257	-0.015	0.078	-0.189
Maximum	0.143	-0.079	-0.13	-0.13
General evaluation				
Mean	0.084	-0.071	0.129	-0.034
SD	0.201	-0.148	-0.13	-0.159
Minimum	-0.004	-0.09	0.097	-0.018
Median	0.083	-0.024	0.146	-0.034
Maximum	0.138	-0.193	-0.068	-0.182

<sup>a</sup>Significance thresholds: \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

<sup>b</sup>EMA: ecological momentary assessment.

<sup>c</sup>GDS-15: 15-item Geriatric Depression Scale.

<sup>d</sup>CESD-R: Center for Epidemiologic Studies Depression Scale-Revised.

<sup>e</sup>PHQ-9: 9-item Patient Health Questionnaire.

<sup>f</sup>BAI: Beck Anxiety Inventory.

Descriptive statistics of raw EMA RTs showed limited associations with symptom change (Table 3). No RT features were significantly correlated with  $\Delta$ GDS-15,  $\Delta$ CESD-R, or  $\Delta$ PHQ-9. The only significant RT associations were observed for anxiety change ( $\Delta$ BAI), where mean ( $r = -0.301$ ;  $P = .04$ ) and maximum ( $r = -0.35$ ;  $P = .01$ ) of appetite RT were negatively correlated with  $\Delta$ BAI, underscoring the challenge of using noisy, naturalistic raw RT descriptives to track depressive symptom change in at-risk older adults.

In contrast, exponential-decay parameters derived from Z-RT (within subject  $\times$  response-option standardization) showed clearer associative patterns (Table 4). For the feeling item, both the decay rate and asymptote were significantly

associated with  $\Delta$ GDS-15 ( $\theta_b$ :  $r = -0.398$ ;  $P = .005$ ,  $\theta_c$ :  $r = -0.321$ ;  $P = .03$ ), representing the strongest  $\Delta$ GDS-15 correlations among the RT-derived features and exceeding those observed for EMA score descriptives. Feeling-item parameters also showed significant associations with other outcomes:  $\Delta$ CESD-R was related to model fit and dynamics ( $R^2$ :  $r = 0.358$ ;  $P = .01$ ,  $\theta_b$ :  $r = -0.302$ ;  $P = .04$ ,  $\theta_c$ :  $r = -0.305$ ;  $P = .04$ ), while  $\Delta$ BAI was associated only with goodness-of-fit ( $R^2$ :  $r = 0.33$ ;  $P = .02$ ). Notably, no significant associations were observed for decay parameters fitted to appetite, sleep quality, or general evaluation across change scores (all  $P > .05$ ), and no Z-RT decay features were significantly related to  $\Delta$ PHQ-9.

**Table 4.** Associations between mental health change scores and EMA Z-RT exponential decay parameters.<sup>a</sup>

EMA <sup>b</sup> item and RT <sup>c</sup> parameter	$\Delta$ GDS-15 <sup>d</sup> ( $r^p$ )	$\Delta$ CESD-R <sup>e</sup> ( $r^p$ )	$\Delta$ PHQ-9 <sup>f</sup> ( $r^p$ )	$\Delta$ BAI <sup>g</sup> ( $r^p$ )
Feeling				
$R^2$	0.19	0.358*	0.271	0.330*
$\theta_a$	0.209	0.144	0.017	0.244
$\theta_b$	-0.398**	-0.302*	-0.263	-0.186
$\theta_c$	-0.321*	-0.305*	-0.266	-0.166
Appetite				
$R^2$	0.034	0.198	0.206	0.266

EMA <sup>b</sup> item and RT <sup>c</sup> parameter	$\Delta$ GDS-15 <sup>d</sup> ( $r^2$ )	$\Delta$ CESD-R <sup>e</sup> ( $r^2$ )	$\Delta$ PHQ-9 <sup>f</sup> ( $r^2$ )	$\Delta$ BAI <sup>g</sup> ( $r^2$ )
$\theta_a$	0.047	0.17	0.158	0.188
$\theta_b$	0.23	0.072	-0.038	-0.12
$\theta_c$	0.172	0.004	-0.084	-0.153
Sleep quality				
$R^2$	-0.194	-0.196	-0.28	-0.021
$\theta_a$	0.107	0.165	0.042	0.206
$\theta_b$	0.182	0.137	0.213	-0.059
$\theta_c$	0.099	0.145	0.229	0.041
General evaluation				
$R^2$	0.206	-0.045	-0.05	-0.15
$\theta_a$	0.242	0.032	-0.042	-0.054
$\theta_b$	-0.104	0.227	0.181	0.127
$\theta_c$	-0.057	0.119	0.155	0.213

<sup>a</sup>Significance thresholds: \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

<sup>b</sup>EMA: ecological momentary assessment.

<sup>c</sup>RT: response time.

<sup>d</sup>GDS-15: 15-item Geriatric Depression Scale.

<sup>e</sup>CESD-R: Center for Epidemiologic Studies Depression Scale-Revised.

<sup>f</sup>PHQ-9: 9-item Patient Health Questionnaire.

<sup>g</sup>BAI: Beck Anxiety Inventory.

## Group Differences in RT Trajectories by EMA-Adjunctive Care Response

### Bayesian Multilevel Modeling

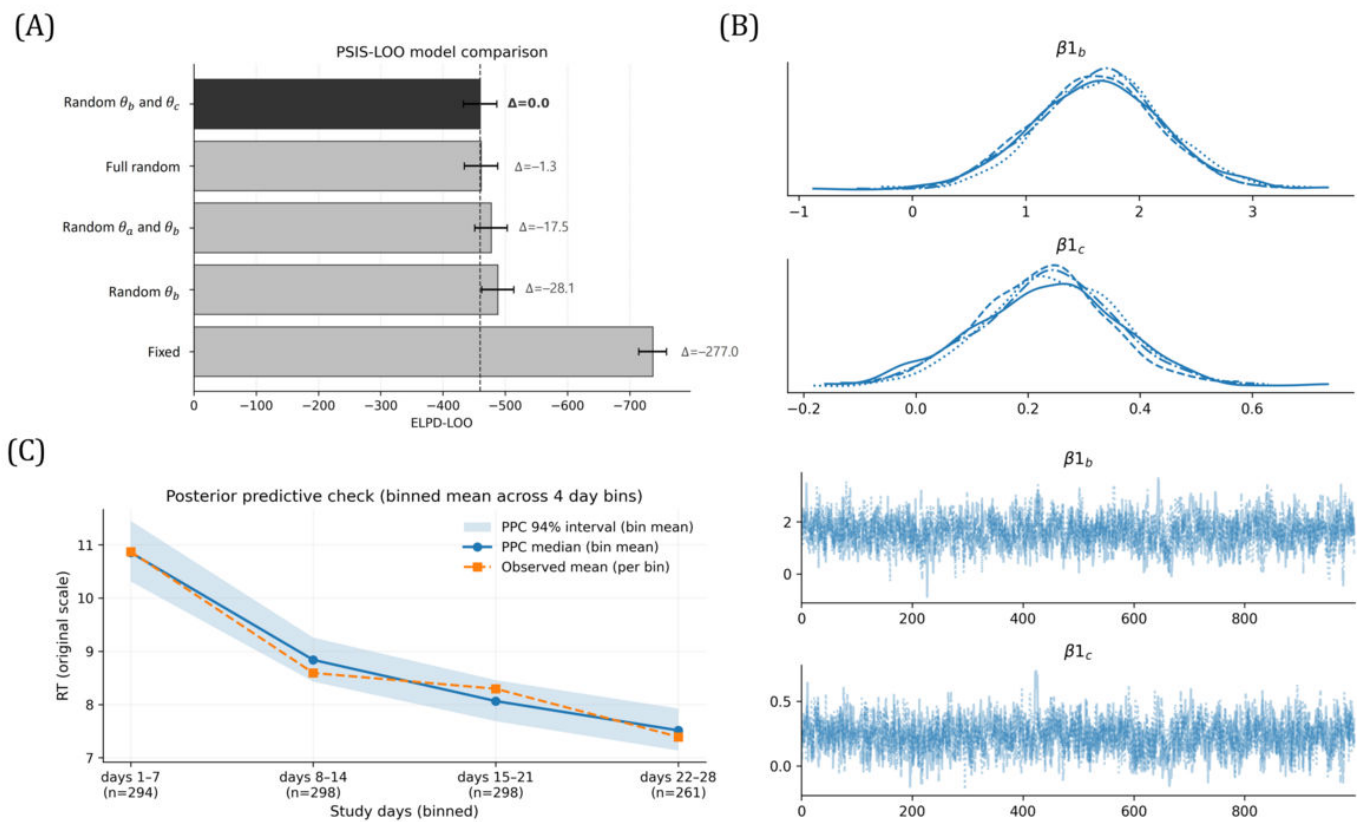
We next evaluated the clinical relevance of EMA-based RT trajectories in relation to depressive symptom improvement during EMA-adjunctive care. Responder-status moderation was tested via the posterior distribution of the group term ( $\beta_1$ ) in a Bayesian multilevel exponential-decay model, where responders were defined as participants whose GDS-15 change exceeded the MDC threshold ( $\Delta$ GDS-15  $\leq$  -3.81). Because prior correlational analyses indicated that only RT dynamic parameters from the feeling item were reliably associated with  $\Delta$ GDS-15, the following results focus on models fit to feeling-item RT (including model selection, adequacy checks, and group effects). Group differences are summarized using posterior estimates,

responder-to-nonresponder parameter ratios, and posterior probabilities of increase.

### Model Selection (Random-Effects Structure)

PSIS-LOO model comparison favored a partial hierarchical exponential-decay model with subject-level random effects on the decay rate ( $\theta_b$ ) and postadaptation RT ( $\theta_c$ ) (Figure 1A; ELPD-LOO = -459.7). A full random-effects model performed similarly ( $\Delta$ ELPD-LOO = 1.3) but showed higher effective complexity ( $p_{100}$ : 112.2 > 109.5), indicating limited predictive benefit from additionally allowing subject-specific variability in amplitude ( $\theta_a$ ). Models with reduced or no random effects showed substantially poorer predictive performance, supporting the need to model between-person heterogeneity in RT adaptation. Accordingly, the model with random decay rate ( $\theta_b$ ) and postadaptation RT ( $\theta_c$ ) was selected for subsequent analyses.

**Figure 1.** Model selection and diagnostic summary. (A) Pareto-smoothed importance sampling leave-one-out (PSIS-LOO) expected log predictive density (ELPD) by random-effects structure (higher is better). (B) Markov chain Monte Carlo (MCMC) trace plots for responder-status moderation terms ( $\beta_1$ ) on key parameters ( $\theta_b$  and  $\theta_c$ ). (C) Binned posterior predictive check comparing observed response time (RT) summaries with posterior predictive medians and 94% intervals. PPC: posterior predictive check.



## Model Adequacy and Diagnostics

Building on the selected random-effects structure (random  $\theta_b$  and  $\theta_c$ ), adding responder-status moderation produced a small but consistent gain in out-of-sample predictive performance relative to the nonmoderated model (ELPD-LOO:  $-457.2$  vs  $-459.7$ ;  $SE \approx 26.6$ - $26.7$ ). Although the absolute improvement was modest, it suggests that responder status explains some systematic variation in RT trajectories beyond subject-to-subject heterogeneity. LOO influence diagnostics were stable (99.7% of observations with Pareto  $k \leq 0.7$ ), indicating that the model's predictive evaluation was not driven by a small set of highly influential observations and that PSIS-LOO approximations were reliable. Markov chain Monte Carlo diagnostics further supported reliable posterior inference. Trace plots showed good mixing with no visible chain separation (Figure 1B), and convergence metrics ( $R \approx 1.00$ - $1.01$  with generally high ESS) indicated that posterior summaries are numerically stable and unlikely to reflect sampling pathologies. Finally, posterior predictive checks demonstrated that the model captures the key empirical pattern of interest, simulated trajectories reproducing the overall decay trend, and binned observed means closely tracked posterior predictive check medians while remaining largely within the 94% predictive intervals on the original RT scale (Figure 1C). Together,

these checks suggest that the moderated model provides an adequate and generalizable description of RT dynamics in this sample, supporting the downstream interpretation of responder-nonresponder differences in adaptation parameters.

## Responder Versus Nonresponder Differences in RT Adaptation

Responders exhibited markedly faster RT adaptation for the feeling item than nonresponders (Table 5). The responder-to-nonresponder ratio for the decay rate was substantially greater than 1 (median  $\theta_b$  ratio=4.86, 95% CrI 1.44-14.31;  $P$  (increase)=.99), indicating a steeper decline (4.86 times faster adaptation) in RT over repeated EMA administrations. Responders also exhibited higher postadaptation RT level than nonresponders (median exp ( $\theta_c$ ) ratio=1.25), indicating 1.25 times slower response after repeated administrations. However, group differences in the postadaptation RT level were more uncertain (95% CrI 0.95-1.58;  $P$  (increase)=.95), suggesting at most modest separation in late-phase baseline latency.

Group-averaged trajectories were consistent with the posterior estimates, showing a steeper RT decline among responders (Figure 2A). Differences in the postadaptation RT level became more apparent when model-implied trajectories

were extrapolated to 100 days beyond the observed study window (shaded region in Figure 2A). Subject-specific estimates supported the group-level findings (Figure 2B). Although posterior estimates varied within each group, such that some responders exhibited slower adaptation or lower baseline RT (and vice versa), individual trajectories generally

followed the overall group trend. Consistent with this, rank plots of participant-level parameters showed responders tending to appear toward higher values of the decay rate and postadaptation RT level, indicating generally larger estimated values.

**Table 5.** Group moderation effects across EMA domains<sup>a</sup>.

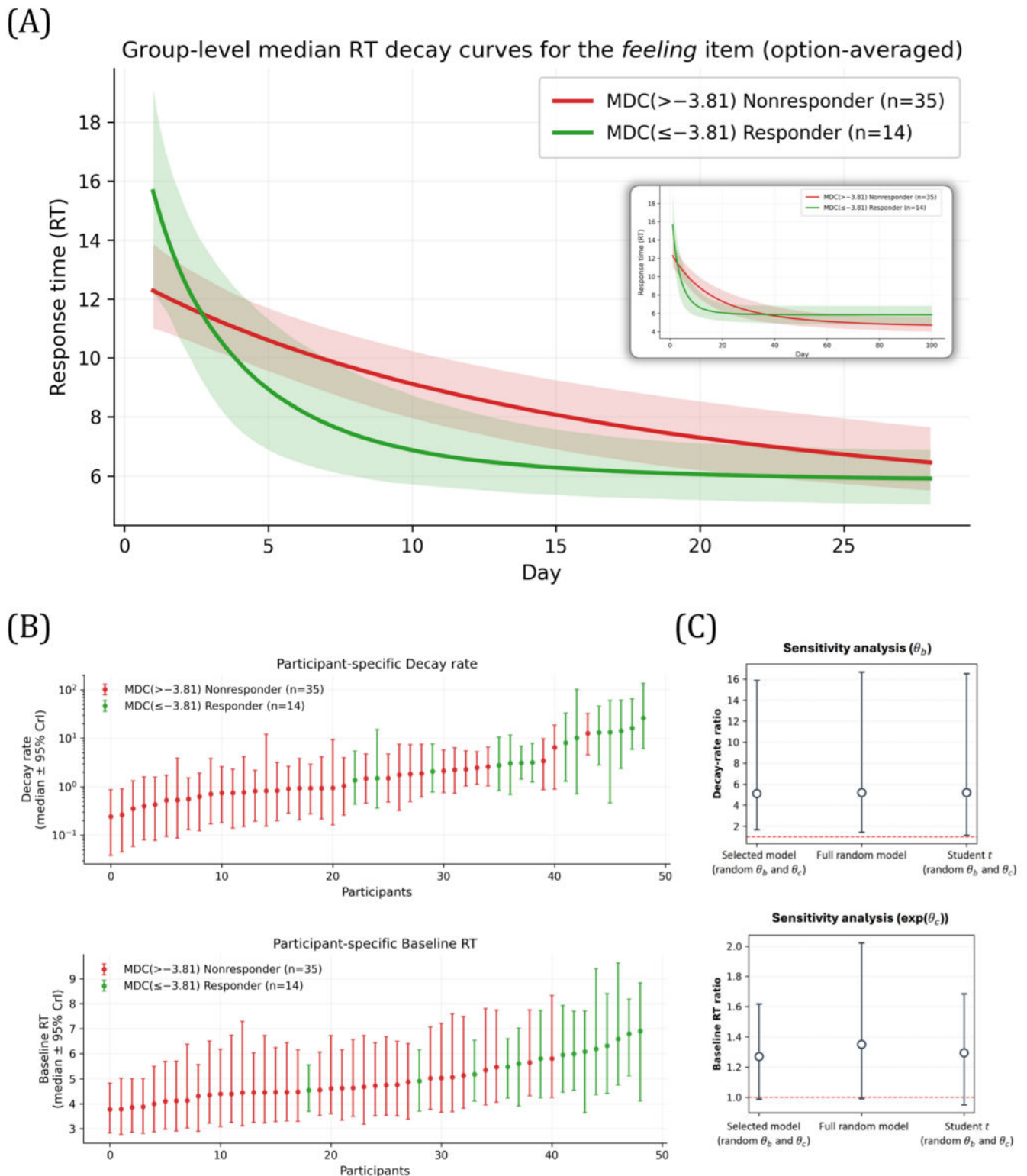
Quantity and EMA <sup>b</sup> item	Posterior median (IQR; width; 95% CrI <sup>c</sup> )	Posterior <i>P</i> (increase)
Decay rate ( $\theta_b$ ) ratio		
	4.86 (3.524-7.609; 4.085; 1.438-14.31)	0.99
Feeling		
Appetite	2.22 (1.565-3.169; 1.604; 0.793-6.561)	0.93
Sleep quality	1.69 (1.186-2.373; 1.187; 0.608-4.707)	0.84
General evaluation	3.32 (1.828-6.229; 4.4; 0.593-19.26)	0.92
Average	3.11 (2.117-4.592; 2.475; 0.991-10.86)	0.97
Baseline RT (exp ( $\theta_c$ )) ratio		
Feeling	1.25 (1.164-1.38; 0.2164; 0.953-1.581)	0.95
Appetite	1.04 (0.975-1.102; 0.1267; 0.862-1.247)	0.66
Sleep quality	0.96 (0.9008-1.025; 0.1244; 0.792-1.161)	0.33
General evaluation	1.09 (0.9891-1.2; 0.2108; 0.813-1.433)	0.73
Average	1.08 (1.015-1.154; 0.1392; 0.886-1.315)	0.80

<sup>a</sup>Ratios >1 indicate larger values in responder than nonresponder.

<sup>b</sup>EMA: ecological momentary assessment.

<sup>c</sup>CrI: credible interval.

**Figure 2.** Responder-status (minimally detectable change [MDC]) moderation results for the feeling item. (A) Model-implied response time (RT) trajectories by group (posterior median with 95% credible interval [CrI]). The shaded region shows an illustrative extrapolation of the fitted trajectories up to 100 days. (B) Participant-level estimates for key parameters ( $\theta_b$  and  $\theta_c$ ) by group. (C) Sensitivity analysis of group-effect ratios across alternative model specifications.



## Item-Specific Moderation Across EMA Domains

Additionally, moderation effects were examined across individual EMA items and an averaged RT outcome. Evidence for group separation was strongest for the feeling item, which showed the largest and most certain decay-rate difference ( $\theta_b$  ratio=4.86, 95% CrI 1.44-14.31;  $P=.99$ ). Other items and the averaged outcome showed smaller, less certain decay-rate ratios with CrIs generally overlapping 1. A similar pattern was shown for the postadaptation RT level, the feeling item again showed the largest (but still uncertain) group ratio ( $\exp(\theta_c)$  ratio=1.25, 95% CrI 0.95-1.58), whereas the remaining items showed minimal differences with intervals spanning 1.

## Sensitivity Analyses

Group-effect conclusions for the feeling item were robust across plausible alternative specifications, including a full random-effects model and a Student  $t$  likelihood or alternative priors (Figure 2C). The responder-to-nonresponder decay-rate ratio remained consistently around 5 (~4.8), with 95% CrIs entirely above 1 across specifications, indicating a stable group difference in RT adaptation speed. By contrast, the postadaptation RT ratio remained modest (~1.3), and its CrIs overlapped 1 in all specifications, suggesting that late-phase RT differences are small and uncertain in this dataset.

## Discussion

### Summary of Results

This study provides preliminary evidence for the feasibility and potential efficacy of EMA-based self-monitoring in reducing psychological symptoms among older adults at elevated risk. Across 4 validated scales (CESD-R, PHQ-9, BAI, and GDS-15), participants showed statistically significant symptom reductions over the 4-week period, with medium-to-large effect sizes. These findings align with prior work suggesting that EMA-related benefits may be enhanced when used as an adjunct to treatment as usual, and they extend that evidence to a geriatric context. The high adherence to the EMA mobile app (>90%) further indicates that EMA-based self-monitoring is feasible and acceptable in older adults, including those with limited familiarity with mobile technologies, supporting its promise as a scalable adjunct to depression care in late life.

Correlational analyses also indicated that EMA-derived features can predict variation in symptom change, but with important limitations. Descriptive statistics of EMA scores and raw EMA RTs showed symptom-specific associations across measures, yet they were not significantly related to changes in the geriatric depression scale ( $\Delta$ GDS-15), suggesting reduced sensitivity for tracking geriatric depressive symptom change using simple descriptives alone. In contrast, features derived from exponential-decay parameters fitted to standardized response times (Z-RT) showed consistent associations with depressive symptom change ( $\Delta$ GDS-15 and  $\Delta$ CESD-R). In particular, faster adaptation

(higher decay rate), higher postadaptation RT level (asymptote/baseline), and better fit quality for the emotionally salient feeling item are associated with greater symptom improvement, suggesting that RT dynamics during EMA engagement may reflect clinically meaningful change during EMA-adjunctive care.

These patterns were further corroborated by the Bayesian multilevel modeling results. Compared with nonresponders, responders to EMA-adjunctive care exhibited markedly faster adaptation (higher decay rate) and modestly higher postadaptation RT levels. This suggests that potential responders show a distinct trajectory characterized by more rapid accommodation to repeated self-reflection, although the underlying mechanism cannot be determined from the current design (ie, whether EMA-adjunctive care enhances self-monitoring efficiency, or whether individuals with greater baseline capacity for such adaptation are more likely to benefit). Evidence for between-group differences in the postadaptation RT level was comparatively uncertain, and the direction (ie, slightly higher baseline RT in responders) may appear counterintuitive if faster responding is assumed to reflect greater efficiency. One plausible interpretation is that faster late-phase responding among nonresponders could reflect a tendency toward more superficial or less attentive responding rather than more efficient self-reflection, but this remains speculative and warrants direct validation. Additionally, sensitivity analyses across plausible alternative specifications yielded the same qualitative conclusions, supporting the robustness of the central group-difference inference. Collectively, these results highlight the potential clinical use of behavioral signatures embedded in EMA responding, particularly RT adaptation dynamics, for tracking depressive symptom improvement in older adults.

Interestingly, consistent with the correlational findings, group contrasts were strongest when the model was fitted to the feeling item among the 4 EMA domains examined. Moreover, although prior work suggests that averaging RT across multiple EMA items can increase measurement reliability [14,16], group differences in adaptive patterns were larger when using the feeling item alone. This finding suggests that EMA response dynamics may be item-specific rather than purely reflecting a global response-speed trait, raising the possibility that different symptom domains could exhibit distinct behavioral signatures in EMA trajectories.

### Clinical Implications

These findings suggest that EMA-adjunctive care may be both feasible and clinically beneficial for older adults at risk of depression. Digital mental health interventions (DMHIs) are often considered difficult to implement in late-life populations because limited familiarity with digital tools can reduce sustained engagement. In contrast, adherence in this study remained high (on average >90%), indicating strong real-world acceptability. The EMA protocol used here, brief daily reporting on four items, appeared manageable even for participants who might be expected to show reduced willingness or capacity for sustained participation due to depressive burden. To support sustained engagement

in real-world deployment, the intended implementation can incorporate an explicit feedback loop, such as clinician-facing dashboards, periodic review during EMA engagements, or automated triggers that prompt timely outreach when a missing response is detected. Importantly, symptom reductions observed across validated scales indicate that EMA use in this context may confer benefits beyond feasibility, supporting EMA-adjunctive care as a promising component of geriatric depression management.

A second implication concerns the objective monitoring of symptom change. Remote psychiatric care still relies heavily on self-reported symptoms, ranging from retrospective standardized questionnaires to daily EMA ratings, both of which remain vulnerable to recall bias and other reporting distortions (eg, mood-congruent evaluation, self-reflective bias). EMA-derived RT metrics are not fully “passive” because they require active responding; however, they provide a comparatively objective behavioral signal that reflects underlying cognitive and psychomotor processes and is less directly shaped by deliberative self-presentation or response framing. The observed association between RT adaptation dynamics and clinically meaningful symptom improvement, therefore, supports RT trajectories as a complementary marker for tracking depressive symptom variation in naturalistic settings, potentially improving reliability when subjective reports are noisy or inconsistent.

Moreover, EMA-RT adaptation features may help make digital mental health interventions more actionable and personalized. As noted in the *Introduction* section, the effects of ecological momentary interventions are heterogeneous and depend on individual characteristics. Consequently, a practical barrier to scaling digital care is identifying early on who will benefit from a given protocol and who will require additional support. Our results suggest that RT adaptation patterns within an initial monitoring window (approximately 1 month) may provide an early indicator of EMA responsiveness. Such early stratification could guide stepped-care decisions (eg, prompting intensified clinician contact, alternative interventions, or safety monitoring for likely non-responders) while allowing responders to continue with lower-intensity, self-guided support, thereby supporting more efficient allocation of constrained clinical resources and optimizing just-in-time adaptive intervention approaches. We also note that this early stratification window may be shortened through systematic investigation of the optimal decision period. Specifically, building on the observed faster adaptation among responders, future work could quantify within-person day-to-day EMA-RT gradients (or slopes over rolling windows) and evaluate candidate “decision days” (eg, 7, 10, 14, or 21 days) to determine an earlier, reliable point for stepped-care decisions.

Finally, the item-specific links observed between EMA domains and symptom changes suggest a pathway to broaden EMA monitoring beyond depression. Because psychiatric symptoms frequently co-occur, identifying domain-specific behavioral signatures (eg, the feeling-depression linkage observed here) may enable EMA to track multiple symp-

tom dimensions simultaneously, supporting more nuanced, individualized monitoring and intervention planning.

## Limitations

First, regarding the observed effectiveness of the EMA-based self-monitoring, although significant reductions in depressive and anxiety symptoms were detected across multiple validated scales, the absence of a control group precludes definitive causal attribution. Participants were not randomized, and potential concurrent treatments, such as case management, pharmacotherapy, or psychotherapy, were not systematically recorded or controlled, introducing possible confounding influences. Additionally, postintervention assessments were conducted only once at the 4-week mark without follow-up measurements to evaluate the durability of treatment effects. As such, the long-term efficacy and sustainability of EMA-based self-monitoring in older adults with depression remain undetermined.

While several EMA-derived features were shown to be associated with symptom variation, some unexpected patterns emerged in the relationship between self-reported EMA scores and symptom change. Specifically, positive correlations were observed between mood-related EMA scores and symptom change in both depression and anxiety domains. These findings suggest that higher EMA ratings during the intervention period were paradoxically linked to less clinical improvement. One possible interpretation is that individuals with persistent depressive symptoms may have reduced self-reflective precision, leading to inflated or undifferentiated positive reports. Alternatively, healthier participants may have demonstrated more nuanced and accurate self-monitoring. However, this hypothesis requires further empirical validation.

Additionally, given the modest sample size ( $n < 50$ ), the correlational analyses were likely underpowered to detect small associations between symptom-change scores and EMA-derived features; thus, null or weak correlations should be interpreted cautiously. Larger cohorts will be needed to more precisely estimate small effects and confirm the robustness and generalizability of the observed associations, particularly those involving the exponential-decay parameters.

Finally, in examining dynamic EMA response trajectories, the most distinctive group-level differences between responders and non-responders emerged in the fitted parameters of the feeling item, whereas appetite, sleep quality, and general evaluation items yielded less pronounced contrasts. This discrepancy may reflect the unique psychological salience of the feeling item, which directly targets the immediate emotional state, in contrast to the more somatically anchored or abstract content of the other items. Alternatively, the fixed order of EMA item presentation, wherein the feeling item was always administered first, may have contributed to its stronger predictive utility. It is plausible that the first item elicits more cognitive effort or reflects the initiation speed of self-reflective responses, thus capturing more variance in related cognitive functioning. This potential order effect should be systematically examined in future studies to better

delineate what EMA-based response latencies truly index in ecological settings.

## Conclusions

This study demonstrates the feasibility of using EMA-based self-monitoring as both an adjunctive care and a behavioral phenotyping tool in older adults with a high risk of depression.

The implications for digital health care are manifold. First, the integration of response-time dynamics into mobile health platforms offers a nonobtrusive means of tracking cognitive and emotional engagement, even in populations with limited digital literacy. The method's compatibility with brief, low-frequency EMA schedules (eg, daily reports, about a month) makes it especially suitable for aging individuals, who

may be less tolerant of intensive digital protocols. Second, the use of within-subject standardization and curve-fitting models supports individualized tracking, allowing for the detection of meaningful intra-individual changes. This has significant implications for the early identification of symptom relapse, monitoring of treatment responsiveness, and tailored interventions based on dynamic behavioral signatures.

In sum, the introduced EMA-based modeling offers a novel, sensitive, and pragmatic approach for advancing mental health care in aging populations. Future work should aim to validate these findings in randomized controlled settings, investigate the mechanistic underpinnings of RT adaptations, and explore broader applications in digital psychiatry and geriatric care.

## Acknowledgments

We extend our sincere gratitude to the Suwon Geriatric Mental Health and Welfare Center for its invaluable role as an organizational partner and for its dedicated support in facilitating comprehensive data collection for this research. The authors used ChatGPT (OpenAI) to assist with language refinement. Sang Joon Son is the co-corresponding author (email: sjsonpsy@ajou.ac.kr; phone: 82 31-219-5180).

## Funding

This research was supported by the ASTRA Project through the National Research Foundation funded by the Ministry of Science and Information and Communications Technology (number RS-2024-00470056).

## Data Availability

The datasets generated or analyzed during this study are not publicly available due to ethical restrictions imposed by the institutional review board and the sensitive nature of clinical and behavioral data but may be obtained from the corresponding author upon reasonable request.

## Authors' Contributions

Conceptualization: Jeehang Lee (equal), Jooho Lee (lead)

Data curation: Jooho Lee

Formal analysis: Jeehang Lee (equal), Jooho Lee (lead)

Funding acquisition: GD (lead), JYP (equal)

Investigation: JN (supporting), KC (lead), SM (supporting), SP (equal)

Methodology: Jeehang Lee (equal), Jooho Lee

Project administration: KC (lead), JN (supporting), SM (supporting), SP (equal)

Resources: JN (supporting), KC (lead), SM (supporting), SP (equal)

Supervision: JYP (equal), SJS (lead)

Validation: JYP (equal), SJS (lead)

Visualization: JN (supporting), Jooho Lee (lead)

Writing – original draft: Jooho Lee

Writing – review & editing: Jeehang Lee (equal), Jooho Lee (lead), JYP (equal), SJS (equal)

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Methodological details.

[\[DOCX File \(Microsoft Word File\), 133 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Correlational analysis table.

[\[XLSX File \(Microsoft Excel File\), 28 KB-Multimedia Appendix 2\]](#)

## References

1. Future household projections 2022–2052 [Web Page in Korean]. Statistics Korea. 2024. URL: [https://mods.go.kr/board.es?mid=a10301010000&bid=207&list\\_no=432696&act=view&mainXml=Y](https://mods.go.kr/board.es?mid=a10301010000&bid=207&list_no=432696&act=view&mainXml=Y) [Accessed 2025-04-07]
2. Future population projections: 2022–2072 [Web Page in Korean]. Statistics Korea. 2024. URL: [https://mods.go.kr/board.es?mid=a10301020600&bid=207&act=view&list\\_no=428476](https://mods.go.kr/board.es?mid=a10301020600&bid=207&act=view&list_no=428476) [Accessed 2025-04-07]
3. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Ann Behav Med*. Sep 1, 1994;16(3):199–202. [doi: [10.1093/abm/16.3.199](https://doi.org/10.1093/abm/16.3.199)]
4. 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](https://doi.org/10.1146/annurev.clinpsy.3.022806.091415)] [Medline: [18509902](https://pubmed.ncbi.nlm.nih.gov/18509902/)]
5. Saeb S, Zhang M, Karr CJ, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J Med Internet Res*. Jul 15, 2015;17(7):e175. [doi: [10.2196/jmir.4273](https://doi.org/10.2196/jmir.4273)] [Medline: [26180009](https://pubmed.ncbi.nlm.nih.gov/26180009/)]
6. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Ment Health*. May 5, 2016;3(2):e16. [doi: [10.2196/mental.5165](https://doi.org/10.2196/mental.5165)] [Medline: [27150677](https://pubmed.ncbi.nlm.nih.gov/27150677/)]
7. Burns MN, Begale M, Duffecy J, et al. Harnessing context sensing to develop a mobile intervention for depression. *J Med Internet Res*. Aug 12, 2011;13(3):e55. [doi: [10.2196/jmir.1838](https://doi.org/10.2196/jmir.1838)] [Medline: [21840837](https://pubmed.ncbi.nlm.nih.gov/21840837/)]
8. Kramer I, Simons CJP, Hartmann JA, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry*. Feb 2014;13(1):68–77. [doi: [10.1002/wps.20090](https://doi.org/10.1002/wps.20090)] [Medline: [24497255](https://pubmed.ncbi.nlm.nih.gov/24497255/)]
9. Hur JW, Kim B, Park D, Choi SW. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: a randomized controlled trial. *Telemed J E Health*. Sep 2018;24(9):710–716. [doi: [10.1089/tmj.2017.0214](https://doi.org/10.1089/tmj.2017.0214)] [Medline: [29323626](https://pubmed.ncbi.nlm.nih.gov/29323626/)]
10. Marciniak MA, Shanahan L, Rohde J, et al. Standalone smartphone cognitive behavioral therapy-based ecological momentary interventions to increase mental health: narrative review. *JMIR mHealth uHealth*. Nov 12, 2020;8(11):e19836. [doi: [10.2196/19836](https://doi.org/10.2196/19836)] [Medline: [33180027](https://pubmed.ncbi.nlm.nih.gov/33180027/)]
11. Myin-Germeys I, Kasanova Z, Vaessen T, et al. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry*. Jun 2018;17(2):123–132. [doi: [10.1002/wps.20513](https://doi.org/10.1002/wps.20513)] [Medline: [29856567](https://pubmed.ncbi.nlm.nih.gov/29856567/)]
12. Bastiaansen JA, Ornée DA, Meurs M, Oldehinkel AJ. An evaluation of the efficacy of two add-on ecological momentary intervention modules for depression in a pragmatic randomized controlled trial (ZELF-i). *Psychol Med*. Dec 14, 2020;52(13):1–10. [doi: [10.1017/S0033291720004845](https://doi.org/10.1017/S0033291720004845)] [Medline: [33315003](https://pubmed.ncbi.nlm.nih.gov/33315003/)]
13. Chung K, Park JY, Joung D, Jhung K. Response time as an implicit self-schema indicator for depression among undergraduate students: preliminary findings from a mobile app-based depression assessment. *JMIR mHealth uHealth*. Sep 13, 2019;7(9):e14657. [doi: [10.2196/14657](https://doi.org/10.2196/14657)] [Medline: [31586362](https://pubmed.ncbi.nlm.nih.gov/31586362/)]
14. Hernandez R, Hoogendoorn C, Gonzalez JS, et al. Reliability and validity of noncognitive ecological momentary assessment survey response times as an indicator of cognitive processing speed in people's natural environment: intensive longitudinal study. *JMIR mHealth uHealth*. May 30, 2023;11:e45203. [doi: [10.2196/45203](https://doi.org/10.2196/45203)] [Medline: [37252787](https://pubmed.ncbi.nlm.nih.gov/37252787/)]
15. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. American Psychiatric Association; 1980. ISBN: 0890420332
16. Hernandez R, Hoogendoorn C, Gonzalez JS, Pyatak EA, Crespo-Ramos G, Schneider S. Reliability and validity of ecological momentary assessment response time-based measures of emotional clarity: secondary data analysis. *JMIR Ment Health*. Jul 18, 2024;11:e58352. [doi: [10.2196/58352](https://doi.org/10.2196/58352)] [Medline: [39024004](https://pubmed.ncbi.nlm.nih.gov/39024004/)]
17. Pier M, Hulstijn W, Sabbe BGC. Psychomotor retardation in elderly depressed patients. *J Affect Disord*. Jul 2004;81(1):73–77. [doi: [10.1016/j.jad.2003.08.002](https://doi.org/10.1016/j.jad.2003.08.002)] [Medline: [15183603](https://pubmed.ncbi.nlm.nih.gov/15183603/)]
18. Bonin P, Barry C, Méot A, Chalard M. The influence of age of acquisition in word reading and other tasks: a never ending story? *J Mem Lang*. May 2004;50(4):456–476. [doi: [10.1016/j.jml.2004.02.001](https://doi.org/10.1016/j.jml.2004.02.001)]
19. Beheydt LL, Schrijvers D, Docx L, Bouckaert F, Hulstijn W, Sabbe B. Cognitive and psychomotor effects of three months of escitalopram treatment in elderly patients with major depressive disorder. *J Affect Disord*. Dec 1, 2015;188:47–52. [doi: [10.1016/j.jad.2015.08.041](https://doi.org/10.1016/j.jad.2015.08.041)] [Medline: [26342888](https://pubmed.ncbi.nlm.nih.gov/26342888/)]
20. Bennabi D, Vandel P, Papaxanthis C, Pozzo T, Haffen E. Psychomotor retardation in depression: a systematic review of diagnostic, pathophysiologic, and therapeutic implications. *Biomed Res Int*. 2013;2013:158746. [doi: [10.1155/2013/158746](https://doi.org/10.1155/2013/158746)] [Medline: [24286073](https://pubmed.ncbi.nlm.nih.gov/24286073/)]
21. Wolinsky FD, Vander Weg MW, Martin R, et al. The effect of speed-of-processing training on depressive symptoms in ACTIVE. *J Gerontol A Biol Sci Med Sci*. Apr 2009;64(4):468–472. [doi: [10.1093/gerona/gln044](https://doi.org/10.1093/gerona/gln044)] [Medline: [19181719](https://pubmed.ncbi.nlm.nih.gov/19181719/)]

22. Brown PJ, Liu X, Sneed JR, Pimontel MA, Devanand DP, Roose SP. Speed of processing and depression affect function in older adults with mild cognitive impairment. *Am J Geriatr Psychiatry*. Jul 2013;21(7):675-684. [doi: [10.1016/j.jagp.2013.01.005](https://doi.org/10.1016/j.jagp.2013.01.005)] [Medline: [23567401](https://pubmed.ncbi.nlm.nih.gov/23567401/)]
23. Smith M, Jones MP, Dotson MM, Wolinsky FD. Speed of processing training and depression in assisted and independent living: a randomized controlled trial. *PLoS One*. 2019;14(10):e0223841. [doi: [10.1371/journal.pone.0223841](https://doi.org/10.1371/journal.pone.0223841)] [Medline: [31622386](https://pubmed.ncbi.nlm.nih.gov/31622386/)]
24. Veenman M, Stefan AM, Haaf JM. Bayesian hierarchical modeling: an introduction and reassessment. *Behav Res Methods*. Aug 2024;56(5):4600-4631. [doi: [10.3758/s13428-023-02204-3](https://doi.org/10.3758/s13428-023-02204-3)] [Medline: [37749423](https://pubmed.ncbi.nlm.nih.gov/37749423/)]
25. McGlothlin AE, Viele K. Bayesian hierarchical models. *JAMA*. Dec 11, 2018;320(22):2365-2366. [doi: [10.1001/jama.2018.17977](https://doi.org/10.1001/jama.2018.17977)] [Medline: [30535206](https://pubmed.ncbi.nlm.nih.gov/30535206/)]
26. Ghahramani Z. Bayesian non-parametrics and the probabilistic approach to modelling. *Philos Trans A Math Phys Eng Sci*. 2012;371(1984):20110553. [doi: [10.1098/rsta.2011.0553](https://doi.org/10.1098/rsta.2011.0553)] [Medline: [23277609](https://pubmed.ncbi.nlm.nih.gov/23277609/)]
27. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1(3):385-401. [doi: [10.1177/014662167700100306](https://doi.org/10.1177/014662167700100306)]
28. Eaton WW, Muntaner C, Smith C, Tien A, Ybarra M. Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-r). In: Maruish ME, editor. *Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults*. 3rd ed. Lawrence Erlbaum Associates; 2004:363-377. ISBN: 9781410610614
29. Lee S, Oh ST, Ryu SY, Jun JY, Lee K, Lee E, et al. Validation of the Korean version of Center for Epidemiologic Studies Depression Scale-Revised. *Int. J. Neuropsychopharmacol*. 2016;19(1):81-82. [doi: [10.1093/ijnp/pyw043.223](https://doi.org/10.1093/ijnp/pyw043.223)]
30. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clin Gerontol*. 1986;5:165-173. [doi: [10.1300/J018v05n01\\_09](https://doi.org/10.1300/J018v05n01_09)]
31. Bae JN, Cho MJ. Development of the Korean version of the Geriatric Depression Scale and its short form among elderly psychiatric patients. *J Psychosom Res*. Sep 2004;57(3):297-305. [doi: [10.1016/j.jpsychores.2004.01.004](https://doi.org/10.1016/j.jpsychores.2004.01.004)] [Medline: [15507257](https://pubmed.ncbi.nlm.nih.gov/15507257/)]
32. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. Sep 2001;16(9):606-613. [doi: [10.1046/j.1525-1497.2001.016009606.x](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
33. Han C, Jo SA, Kwak JH, et al. Validation of the Patient Health Questionnaire-9 Korean version in the elderly population: the Ansan Geriatric study. *Compr Psychiatry*. 2008;49(2):218-223. [doi: [10.1016/j.comppsy.2007.08.006](https://doi.org/10.1016/j.comppsy.2007.08.006)] [Medline: [18243897](https://pubmed.ncbi.nlm.nih.gov/18243897/)]
34. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol*. Dec 1988;56(6):893-897. [doi: [10.1037//0022-006x.56.6.893](https://doi.org/10.1037//0022-006x.56.6.893)] [Medline: [3204199](https://pubmed.ncbi.nlm.nih.gov/3204199/)]
35. Lee HK, Lee EH, Hwang ST, et al. Psychometric properties of the Beck Anxiety Inventory in the community-dwelling sample of Korean adults. *Korean J Clin Psychol*. Nov 2016;35(4):822-830. [doi: [10.15842/kjep.2016.35.4.010](https://doi.org/10.15842/kjep.2016.35.4.010)]
36. Newell A, Rosenbloom PS. Mechanisms of skill acquisition and the law of practice. In: Anderson JR, editor. *Cognitive Skills and Their Acquisition*. Lawrence Erlbaum Associates; 1981:1-55. ISBN: 9780898590937
37. Heathcote A, Brown S, Mewhort DJ. The power law repealed: the case for an exponential law of practice. *Psychon Bull Rev*. Jun 2000;7(2):185-207. [doi: [10.3758/bf03212979](https://doi.org/10.3758/bf03212979)] [Medline: [10909131](https://pubmed.ncbi.nlm.nih.gov/10909131/)]
38. Olivier GN, Walter CS, Paul SS, Dibble LE, Schaefer SY. How common is the exponential decay pattern of motor skill acquisition? A brief investigation. *Motor Control*. May 13, 2021;25(3):451-461. [doi: [10.1123/mc.2020-0043](https://doi.org/10.1123/mc.2020-0043)] [Medline: [33992025](https://pubmed.ncbi.nlm.nih.gov/33992025/)]
39. Kaiser L. Adjusting for baseline: change or percentage change? *Stat Med*. Oct 1989;8(10):1183-1190. [doi: [10.1002/sim.4780081002](https://doi.org/10.1002/sim.4780081002)] [Medline: [2682909](https://pubmed.ncbi.nlm.nih.gov/2682909/)]
40. Jennings MA, Cribbie RA. Comparing pre-post change across groups: guidelines for choosing between difference scores, ANCOVA, and residual change scores. *J Data Sci*. 2016;14(2):205-230. [doi: [10.6339/JDS.201604\\_14\(2\).0002](https://doi.org/10.6339/JDS.201604_14(2).0002)]
41. Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making. U.S. Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER); 2023. URL: <https://www.fda.gov/media/166830/download> [Accessed 2025-07-07]
42. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. Nov 2001;94(2):149-158. [doi: [10.1016/S0304-3959\(01\)00349-9](https://doi.org/10.1016/S0304-3959(01)00349-9)] [Medline: [11690728](https://pubmed.ncbi.nlm.nih.gov/11690728/)]
43. Warkentin LM, Das D, Majumdar SR, Johnson JA, Padwal RS. The effect of weight loss on health-related quality of life: systematic review and meta-analysis of randomized trials. *Obes Rev*. Mar 2014;15(3):169-182. [doi: [10.1111/obr.12113](https://doi.org/10.1111/obr.12113)] [Medline: [24118750](https://pubmed.ncbi.nlm.nih.gov/24118750/)]

44. Rogosa DR, Willett JB. Demonstrating the reliability the difference score in the measurement of change. *J Educ Meas.* Dec 1983;20(4):335-343. [doi: [10.1111/j.1745-3984.1983.tb00211.x](https://doi.org/10.1111/j.1745-3984.1983.tb00211.x)]
45. Castro-Schilo L, Grimm KJ. Using residualized change versus difference scores for longitudinal research. *J Soc Pers Relat.* Jan 2018;35(1):32-58. [doi: [10.1177/0265407517718387](https://doi.org/10.1177/0265407517718387)]
46. González-Domínguez NP, Wu Y, Fan S, et al. Individual patient data meta-analysis estimates the minimal detectable change of the Geriatric Depression Scale-15. *J Clin Epidemiol.* Sep 2024;173:111443. [doi: [10.1016/j.jclinepi.2024.111443](https://doi.org/10.1016/j.jclinepi.2024.111443)] [Medline: [38942179](https://pubmed.ncbi.nlm.nih.gov/38942179/)]
47. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* Oct 2006;59(10):1033-1039. [doi: [10.1016/j.jclinepi.2005.10.015](https://doi.org/10.1016/j.jclinepi.2005.10.015)] [Medline: [16980142](https://pubmed.ncbi.nlm.nih.gov/16980142/)]
48. Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Comput Sci.* 2016;2:e55. [doi: [10.7717/peerj-cs.55](https://doi.org/10.7717/peerj-cs.55)]

## Abbreviations

**BAI:** Beck Anxiety Inventory  
**CESD-R:** Center for Epidemiologic Studies Depression Scale–Revised  
**CrI:** credible interval  
**DSM-5:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*  
**ELPD-LOO:** expected log pointwise predictive density leave-one-out cross-validation  
**EMA:** ecological momentary assessment  
**GDS-15:** 15-item Geriatric Depression Scale  
**MDC:** minimal detectable change  
**MDD:** major depressive disorder  
**PHQ-9:** 9-item Patient Health Questionnaire  
**PSIS-LOO:** Pareto-smoothed importance sampling leave-one-out cross-validation  
**RT:** response time

*Edited by Jinjiao Wang; peer-reviewed by Joao Pedro Nunes, Raymond Hernandez; submitted 10.Sep.2025; final revised version received 18.Feb.2026; accepted 17.Mar.2026; published 08.May.2026*

*Please cite as:*

Lee J, Lee J, Park S, Do G, Noh J, Moon S, Chung K, Son SJ, Park JY  
*Response Time Dynamics From Noncognitive Ordinal Ecological Momentary Assessment as a Proxy for Symptom Change in Geriatric Depression: Longitudinal Observational Study*  
*JMIR Aging* 2026;9:e83891  
URL: <https://aging.jmir.org/2026/1/e83891>  
doi: [10.2196/83891](https://doi.org/10.2196/83891)

© Jooho Lee, Jeehang Lee, Sehwan Park, Gangho Do, Jihye Noh, Sangjoon Moon, Kyungmi Chung, Sang Joon Son, Jin Young Park. Originally published in *JMIR Aging* (<https://aging.jmir.org>), 08.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Aging*, is properly cited. The complete bibliographic information, a link to the original publication on <https://aging.jmir.org>, as well as this copyright and license information must be included.