

# UC San Diego

## UC San Diego Previously Published Works

### Title

Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults.

### Permalink

<https://escholarship.org/uc/item/7bq538mw>

### Authors

Moore, Raeanne C  
Depp, Colin A  
Wetherell, Julie Loebach  
[et al.](#)

### Publication Date

2016-04-01

### DOI

10.1016/j.jpsychires.2016.01.011

Peer reviewed



# Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults



Raeanne C. Moore<sup>a, b, c, \*</sup>, Colin A. Depp<sup>a, b, c</sup>, Julie Loebach Wetherell<sup>c, a</sup>, Eric J. Lenze<sup>d</sup>

<sup>a</sup> Department of Psychiatry, University of California San Diego, 9500 Gilman Drive 0993, La Jolla, CA 92093-0993, USA

<sup>b</sup> The Sam and Rose Stein Institute for Research on Aging, University of California San Diego, 9500 Gilman Drive 0664, La Jolla, CA 92093-0664, USA

<sup>c</sup> VA San Diego Healthcare System, 3550 La Jolla Village Drive, San Diego, CA 92161, USA

<sup>d</sup> Washington University School of Medicine, Department of Psychiatry, 660 So. Euclid Ave, Campus Box 8134, St. Louis, MO 63110, USA

## ARTICLE INFO

### Article history:

Received 2 July 2015

Received in revised form

9 December 2015

Accepted 15 January 2016

### Keywords:

Ecological momentary assessment

Mobile assessment

Randomized clinical trial

Psychometrics

Patient reported outcomes

Mindfulness-based stress reduction

## ABSTRACT

As mobile data capture tools for patient-reported outcomes proliferate in clinical research, a key dimension of measure performance is sensitivity to change. This study compared performance of patient-reported measures of mindfulness, depression, and anxiety symptoms using traditional paper-and-pencil forms versus real-time, ambulatory measurement of symptoms via ecological momentary assessment (EMA). Sixty-seven emotionally distressed older adults completed paper-and-pencil measures of mindfulness, depression, and anxiety along with two weeks of identical items reported during ambulatory monitoring via EMA before and after participation in a randomized trial of Mindfulness-Based Stress Reduction (MBSR) or a health education intervention. We calculated effect sizes for these measures across both measurement approaches and estimated the Number-Needed-to-Treat (NNT) in both measurement conditions. Study outcomes greatly differed depending on which measurement method was used. When EMA was used to measure clinical symptoms, older adults who participated in the MBSR intervention had significantly higher mindfulness and significantly lower depression and anxiety than participants in the health education intervention at post-treatment. However, these significant changes in symptoms were not found when outcomes were measured with paper-and-pencil measures. The NNT for mindfulness and depression measures administered through EMA were approximately 25–50% lower than NNTs derived from paper-and-pencil administration. Sensitivity to change in anxiety was similar across administration modes. In conclusion, EMA measures of depression and mindfulness substantially outperformed paper-and-pencil measures with the same items. The additional resources associated with EMA in clinical trials would seem to be offset by its greater sensitivity to detect change in key outcome variables.

Published by Elsevier Ltd.

## 1. Introduction

Ecological momentary assessment (EMA) is a data capture technique that involves repeated sampling of thoughts, feelings, or behaviors as close in time to the experience as possible in the naturalistic environment (Shiffman et al., 2008). Among the purported advantages of EMA is the mitigation of biases inherent in

retrospective self-reports, such as the concern that the participant's reporting of subjective experiences in the past may be influenced by their current state (Axelson et al., 2003; Ebner-Priemer and Trull, 2009; Granholm et al., 2008; Johnson et al., 2009; Moskowitz and Young, 2006; Shiffman et al., 2008; Trull and Ebner-Priemer, 2009). Among older adults, memory impairment and unfamiliarity with questionnaire formats may further limit the validity of assessment tools that require the participant to recall their experience over the past week or month (Lenze and Wetherell, 2009). Assessing symptoms such as depressed mood or anxiety, or psychological constructs such as mindfulness, with retrospective self-report measures is particularly problematic given their variability within and between days (Baer et al., 2009; Bishop et al., 2004; Lau et al.,

\* Corresponding author. Department of Psychiatry, University of California, San Diego 9500 Gilman Drive (0993), La Jolla, CA 92093-0993, USA.

E-mail addresses: [r6moore@ucsd.edu](mailto:r6moore@ucsd.edu) (R.C. Moore), [cdepp@ucsd.edu](mailto:cdepp@ucsd.edu) (C.A. Depp), [jlwetherell@ucsd.edu](mailto:jlwetherell@ucsd.edu) (J.L. Wetherell), [lenzee@psychiatry.wustl.edu](mailto:lenzee@psychiatry.wustl.edu) (E.J. Lenze).

2006; Orsillo, 2005; Starr and Davila, 2012). EMA queries about present moment experiences in *real time* multiple times throughout the day, which could create more stable estimates of phenomena that fluctuate over time compared to single time-point measurement. For some internal experiences, such as mindfulness, in-the-moment questions may better enable sampling of experiences without the retrospective judgments that are inherent in global self-reports.

With the emergence of smartphones, there is unprecedented capacity to obtain EMA data in naturalistic environments. Even with the ‘digital divide’ in older adults’ comfort and experience with technology, on average, relative to younger adults, a number of studies support the feasibility and acceptability of EMA techniques assessing multiple patient-reported outcomes with older adults (Cain et al., 2009). However, although much cross-sectional data support the feasibility and construct validity of EMA relative to traditional paper-and-pencil patient-reported outcomes, little is known about the sensitivity of EMA-based measures to change in clinical trials. The great majority of prior studies employing EMA have been observational studies and have not employed EMA in the context of detecting the effect of interventions. A number of authors have suggested that EMA could provide a useful approach to gathering patient-reported outcome measures and better representing the patient’s experience over time during treatment (Gwaltney et al., 2008). Measurement error known to be associated with traditional paper-and-pencil measures can result in low assay sensitivity and potentially smaller intervention effect sizes of clinical trials (Cain et al., 2009; Collins et al., 2003; Slater and Bick, 1994). However “head-to-head” comparisons addressing sensitivity to change with identical point-in-time paper-and-pencil measures have, to our knowledge, not been performed. There is non-trivial participant training, burden, and expense in implementing EMA, and so its use as an outcome measurement tool would need to be justified by evidence of increased reliability, validity, and sensitivity to change over traditional self-reports. The added challenges posed by EMA implementation may be more substantial in older adults, who may require more training and support in using EMA.

In this study, we examined the psychometric properties and sensitivity of EMA in contrast to paper-and-pencil measures among older adults who participated in a randomized controlled trial examining Mindfulness-Based Stress Reduction (MBSR) vs. a health education control group. Identical EMA and paper-and-pencil measures of depression, anxiety (derived from Patient Reported Outcome Management System [PROMIS] Short-Form), and mindfulness (derived from the CAMS-R; Feldman et al., 2007) were administered at baseline and post-treatment, affording us the opportunity to contrast the reliability, concordance, and ability to detect changes over the study period. This is the first study, to our knowledge, to examine sensitivity to change of EMA methods in contrast to paper-and-pencil measures, and among the first to measure sensitivity to change in mindfulness as assessed via EMA. Comparing these two assessment methods is important because ultimately mindfulness-based interventions needs to show efficacy for clinical outcomes if it is to be a treatment for late-life mental disorders; this requires reliable measurement of clinical outcomes (Bierman et al., 2005). We hypothesized that 1) EMA would be associated with greater internal consistency and item-total correlations than paper-and-pencil measures, 2) changes in EMA would be associated with larger effect sizes than paper-and-pencil measures.

## 2. Material and methods

### 2.1. Participants and design

This multisite study was conducted at Washington University in St. Louis and the University of California, San Diego, and was approved by both sites’ institutional review boards. This study represents a secondary aim of a randomized clinical trial in which participants with anxiety or depressive disorders and subjective cognitive complaints were randomized to either participate in MBSR or health education. Therefore, the study was statistically powered to detect change in anxious and depressive symptoms and to compare these two assessment methods and cross-validate these data with the same outcome measures collected by in-person raters. The primary aim of the clinical trial was to assess change in memory and executive functions. Expanded details of the two treatment conditions and the primary aim outcomes are described in a separate paper (Wetherell et al., 2016, under revision). Details about the patient-reported measures or EMA protocol have not been previously published.

All participants volunteered and provided written, informed consent. One hundred and three adults aged 65 years or older with clinically significantly anxiety-related distress and self-reported cognitive dysfunction were enrolled in the trial (Washington University:  $n = 52$ ; UCSD:  $n = 51$ ). The EMA program was still under development at the start of the trial, and this led to us being unable to capture EMA data on the first 10 participants. Given the focus on sensitivity to change, 21 participants were dropped because they completed less than 10 EMA surveys at baseline and an additional 5 were dropped due to insufficient EMA data a follow-up, resulting in a total of 67 participants included in this study.

Participants were excluded for: screening score  $< 22$  on the Penn State Worry Questionnaire-Abbreviated (PSWQ-A; Hopko et al., 2003); no self-reported cognitive dysfunction on screening question: “Have you noticed that you have any trouble with your memory or concentration?”; diagnosis of dementia based on known diagnosis or meeting criteria during screening exam (Katzman et al., 1983); lifetime diagnosis of psychotic or bipolar disorder; alcohol or substance use disorder within past six months; corticoid steroid use; current participation in psychotherapy, meditation practice, or yoga; unstable medical condition (e.g., congestive heart failure); or any condition or impairment likely to interfere with the ability to participate in MBSR.

### 2.2. Measures

#### 2.2.1. Demographic characteristics

These included age, sex, years of formal education, race/ethnicity, and marital status.

#### 2.2.2. EMA and paper-and-pencil clinical assessments

For depressive and anxiety symptoms, we used the National Institutes of Health (NIH) Patient Reported Outcomes Measurement Information System (PROMIS) adult depression and anxiety short form instruments (Bjorner et al., 2013). PROMIS derives from large item banks to measure patient-reported outcomes, and the psychometric properties of these item repositories have been rigorously tested (Cella et al., 2007; Reeve et al., 2007). The PROMIS short-form anxiety items focus on anxious apprehension (i.e., worry) and hyperarousal (i.e., tension, nervousness, and anxiousness). For the paper-and-pencil administration, we used the 7-item PROMIS anxiety scale. The PROMIS short-form depression items focus on negative mood (e.g., depressed, hopeless) and negative views of self (i.e., worthlessness, helplessness). For the paper-and-pencil administration, we used the 8-item PROMIS depression

scale. For EMA administration, we used the 4 anxiety and 4 depression items with the highest item-total correlations with their respective parent scales. For each instrument, participants rate the frequency of their symptoms on a 5-point scale ranging from 1 = *not at all* to 5 = *very much*. Total raw scores ranging from 4 to 20 for depression and anxiety symptoms at each EMA time point were calculated. Comparable raw scores based only on the corresponding anxiety and depression items were derived from the paper-and-pencil questionnaires and used in the analyses reported here.

To evaluate symptoms of mindfulness, we used the Cognitive Affective Mindfulness Scale-Revised (CAMS-R; Feldman et al., 2007). The full scale was administered in paper-and-pencil format. For EMA administration, the following four items from the CAMS-R were included: *I am preoccupied by the future*; *I am focused on the present moment*; *I am preoccupied by the past*; and *I am able to accept the thoughts and feelings I have*. All items are rated on a scale from 1 = *not at all* to 4 = *very much*. The items *I am preoccupied by the future* and *I am preoccupied by the past* were reverse coded, and all 4 items were summed to create a CAMS-R Total Score. The 4 items were chosen by the investigative team because they best represented the study aims and hypotheses about the construct of mindfulness, i.e., present moment orientation and nonjudgmental acceptance. Scores from the same 4 items were used for the paper-and-pencil measure to conduct the analyses reported here.

We note that the items for the PROMIS and CAMS-R measures were identical across EMA and paper-and-pencil measures but the frame of reference for the EMA version was in reference to the current state and for paper-and-pencil items to the past week. For example, for the PROMIS item “I felt anxious” the wording of the paper-and-pencil items were not changed: “In the past 7 day I felt anxious” from 1 = *Never* to 5 = *Always [Several times a day]*. The wording of the EMA item was converted to reflect present state: “At the moment I feel anxious” from 1 = *Not at all* to 5 = *Very Much*. Only the frame of reference for CAMS-R items were changed (i.e., paper-and-pencil version asked participant to reflect on experience over past 7 days, EMA version asked participant to reflect on present experience).

### 2.2.3. Procedure

After providing informed consent, participants meeting enrollment criteria completed an in-person pre-treatment assessment, including completion of the paper-and-pencil PROMIS measures and CAMS-R. Blind raters at Washington University and UCSD performed all assessments. Participants were then provided with a smartphone and sampled at-home with EMA surveys three times per day for typically 10 days at pre-treatment. The EMA assessments began the day immediately following the in-person visit. In some cases, participants had EMA assessments that lasted longer ( $n = 20$ ), as the EMA program did not stop assessments until the device was returned to research staff. All 12 items of interest (4 items from PROMIS Depression, 4 items from PROMIS Anxiety, and 4 items from CAMS-R) were sampled at all EMA time points. After participants completed their pre-treatment in-person and EMA assessments, they were randomized in groups of 5–8 people to either MBSR or to health education. Both MBSR and health education programs consisted of 8 once weekly, group-delivered sessions of approximately 90 minutes each. MBSR was conducted according to the protocol developed by Jon Kabat-Zinn, Ph.D. and colleagues at the University of Massachusetts, Boston (Stahl and Goldstein, 2010). We previously modified the MBSR meditation and light yoga sessions to reduce risk of injury to older patients (Lenze et al., 2014), and these modified sessions were administered in this study. Participation also included a half-day meditation

retreat.

The health education program was based on the health care self-management book written by Kate Lorig and colleagues (Lorig et al., 2012), and covered topics such as: finding resources, understanding and managing common conditions and symptoms, exercising for fitness, healthy eating, managing medications, expressing feelings, and communicating with health care providers. The original program included topics on relaxation and meditation strategies, which were removed for this study.

After completion of the treatment programs, participants returned to the laboratory and completed a post-treatment visit, including completion of the paper-and-pencil PROMIS measures and CAMS-R. They also completed another 10 days of at-home EMA assessments post-treatment. After completion of the EMA assessment period, participants returned the smartphone to the laboratory and were compensated for their participation. On average, each participant completed as many as 30 momentary assessments of depression, anxiety, and mindfulness at each time point (pre- and post-treatment). A summary of study procedures is presented in Fig. 1.

### 2.3. Statistical analyses

Data were analyzed using IBM SPSS version 22 (SPSS, 2010). Participants who had completed at least 10 EMA data points at baseline were included in the analyses. Group differences (MBSR vs. health education) were examined using t-tests for continuous and chi-squared for categorical variables. Next, Cronbach's alpha was

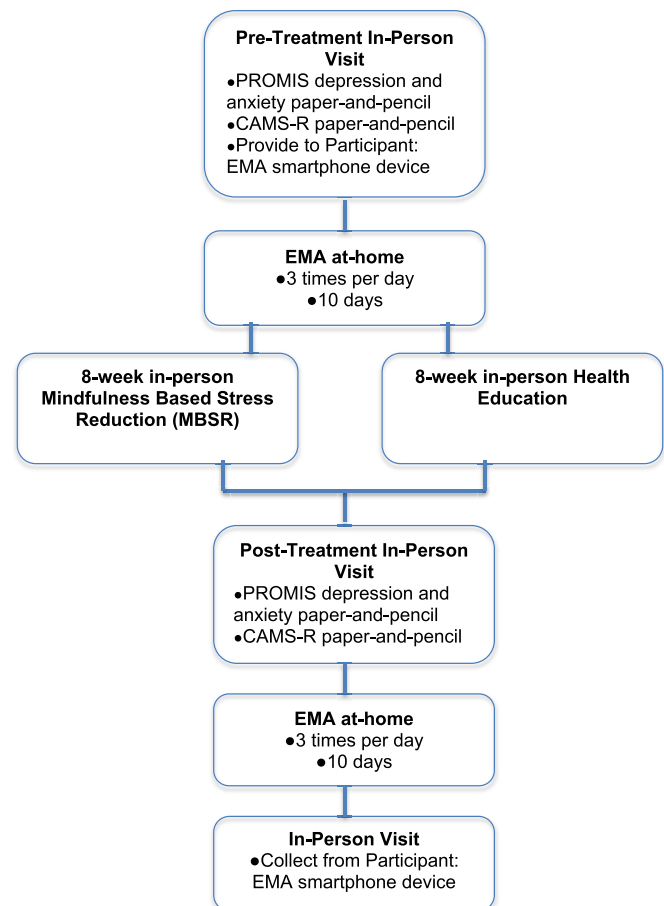


Fig. 1. Timing and frequency of EMA and paper-and-pencil assessments.

calculated for both EMA and paper-and-pencil data of the primary outcome variables: Mindfulness Total Score (4-item total from CAMS-R), Depression Total Score (4-item total from PROMIS depression scale), and Anxiety Total Score (4-item total from PROMIS anxiety scale). Baseline intercorrelations between study variables were examined using Pearson correlations, and Pearson correlations between EMA individual items on all outcome variables with the paper-and-pencil 4-item total score unstandardized predicted values were examined.

We examined group differences using data from all 67 randomized participants who provided sufficient baseline and follow-up data via both paper-and-pencil questionnaires and EMA. The primary outcomes were change in mindfulness, change in depressive symptoms, and change in anxiety symptoms; change in all three outcomes using EMA data was compared to data from the paper-and-pencil versions of these items (completed pre- and post-treatment). Change in these outcomes was analyzed using mixed-models, repeated measures analysis of variance with restricted maximum likelihood (REML) estimation. Treatment group (MBSR vs. health education), assessment point (baseline and post-treatment), and their interaction were fixed effects, and participants were treated as the random effect. The group-by-time interaction was the fixed effect of interest. The same procedure was applied separately for EMA data and paper-and-pencil data.

We calculated effect sizes for all outcomes. The first effect size was Cohen's *d* (Feingold, 2009). Next, as a marker of clinical significance, we calculated the Number-Needed-to-Treat (NNT) using the formula described by Furukawa and Leucht (Furukawa and Leucht, 2011):  $1/[NNT = 1/[\Phi(\delta - \Psi(\text{CER})) - \text{CER}]]$ . In this formula,  $\Phi$  = cumulative distribution function of the standard normal distribution;  $\Psi$  = inverse of  $\Phi$ ; CER = HE group's event rate; and  $\delta$  = population Cohen's *d*. NNT values were calculated based on the assumption that 20% of the health education group would have favorable outcomes.

### 3. Results

#### 3.1. Participant characteristics

As seen in Table 1, participants in both treatment conditions did not significantly differ on any demographic variables; therefore, none of these variables were included as covariates in the analyses. Moreover, no significant differences were observed between the two treatment conditions on mindfulness measured with either the paper-and-pencil measure or EMA. However, baseline group differences were observed for depression and anxiety as measured

with EMA, with more severe symptomatology reported by the participants randomly assigned to health education (Table 1).

#### 3.2. Internal consistency of EMA compared to paper-and-pencil assessment data

Cronbach's alphas for each total score (Depression, Anxiety and Mindfulness) were calculated for both EMA and paper-and-pencil data. For EMA, CAMS-R  $\alpha = 0.61$ ; Depression  $\alpha = 0.90$ ; Anxiety  $\alpha = 0.93$ . For the paper-and-pencil baseline data, CAMS-R  $\alpha = 0.53$ ; Depression  $\alpha = 0.84$ ; Anxiety  $\alpha = 0.85$ . To statistically compare differences in Cronbach's alpha coefficients between the two methods of administration, we used the conron calculation from the R programming language (Feldt et al., 1987). No significant differences were observed between the EMA and paper-and-pencil measures in terms of internal consistency (CAMS-R:  $\text{Chi}^2 = 0.17$ ,  $p = 0.68$ ; depression:  $\text{Chi}^2 = 1.03$ ,  $p = 0.31$ ; anxiety:  $\text{Chi}^2 = 2.66$ ,  $p = 0.10$ ).

#### 3.3. Correlations among baseline EMA mindfulness, depression, and anxiety questions

Baseline intercorrelations of individual items with EMA and paper-and-pencil total scores are provided in Table 2. To compare the relationship between individual EMA items with total scores from paper-and-pencil measures, we examined the correlations between EMA individual items with paper-and-pencil 4-item total score unstandardized predicted values (Table 3). For the paper-and-pencil variables, the mean predicted value for CAMS-R was 11.01 (SD = 2.28; range = 6–15); for depression, the mean predicted value was 9.45 (SD = 3.66; range = 4–17); and for anxiety, the mean predicted value was 12.64 (SD = 3.36; range = 4–19). As seen in Table 3, correlations between EMA individual items and paper-and-pencil predicted values ranged from small to medium.

#### 3.4. Sensitivity to change in EMA and paper-and-pencil approaches

In this subsample of participants who completed the randomized clinical trial, a significant difference between the MBSR and health education conditions were observed at the end of treatment based on the EMA measure of mindfulness but not on the abbreviated (i.e., 4 item) paper-and-pencil measure of mindfulness (Table 4). When translating these between-group effect sizes into Number-Needed-to-Treat (NNT) for clinical significance, the NNT for EMA = 7.5, whereas the NNT for paper-and-pencil = 13.6. Similarly, for depressive symptoms, within-group contrasts

**Table 1**  
Descriptive statistics for demographic variables and test scores for MBSR and health education groups.

	MBSR (N = 32)	Health education (N = 35)	t or $\text{Chi}^2$	df	p
Age (years), M (SD)	69.8(4.1) range: 65-79	72.0(5.5) range: 65-82	1.9	65	0.07
Female gender, n (%)	72%	81%	0.6	1	0.4
Education (years), M (SD)	15.7(3.0) range: 6-20	15.8(2.6) range: 9-20	0.2	65	0.9
Ethnicity (% White)	72%	89%	8.1	4	0.1
Marital status (% Married)	47%	63%	3.9	4	0.4
Site (% San Diego)	44%	46%	0.0	1	0.9
CAMS-R, paper-and-pencil, M (SD)	10.7(2.3)	11.0(2.2)	0.5	57	0.6
CAMS-R, EMA, M (SD)	13.3(2.7)	13.3(2.7)	0.3	1426	0.7
PROMIS depression, paper-and-pencil, M (SD)	9.7(3.6)	9.52(3.6)	-0.2	57	0.8
PROMIS depression, EMA, M (SD)	6.7(3.4)	7.7 (4.2)	5.4	1600	<0.001
PROMIS anxiety, paper-and-pencil, M (SD)	12.7(2.9)	13.2(3.5)	0.6	57	0.6
PROMIS anxiety, EMA, M (SD)	8.5(4.1)	10.0(4.5)	6.8	1574	<0.001

Note. MBSR = Mindfulness-Based Stress Reduction; CAMS-R = Cognitive and Affective Mindfulness Scale – Revised; PROMIS = Patient-Reported Outcomes Measurement Information System.

**Table 2**  
Baseline individual EMA items are highly correlated with EMA and paper-and-pencil total scores.

	Hopelessness (EMA)	Helplessness (EMA)	Depressed (EMA)	Worthless (EMA)
Depression total score (EMA)	0.92**	0.91**	0.85**	0.83**
Depression total score (paper-and-pencil)	0.86**	0.85**	0.80**	0.82**
	Tense (EMA)	Worried (EMA)	Anxious (EMA)	Nervous (EMA)
Anxiety total score (EMA)	0.90**	0.90**	0.91**	0.91**
Anxiety total score (paper-and-pencil)	0.86**	0.83**	0.85**	0.83**
	Preoccupied with past (EMA)	Focused on the present (EMA)	Accepts thoughts (EMA)	Preoccupied with future (EMA)
Mindfulness total score (EMA)	−0.65**	0.69**	0.68**	−0.70**
Mindfulness total (paper-and-pencil)	−0.61**	0.75**	0.71**	−0.55**

\*\* $p < 0.001$ .

**Table 3**  
EMA individual items are associated with paper-and-pencil total score unstandardized predicted values.

Mindfulness EMA individual items	Mindfulness total score paper-and-pencil unstandardized predicted values
I am preoccupied by the past	−0.40**
I am able to focus on the present moment	0.37**
I am able to accept the thoughts and feelings I have	0.21**
I am preoccupied with the future	−0.31**
Depression EMA individual items	Depression total score paper-and-pencil unstandardized predicted values
I feel hopeless	0.47**
I feel helpless	0.47**
I feel depressed	0.46**
I feel worthless	0.55**
Anxiety EMA individual items	Anxiety total score paper-and-pencil unstandardized predicted values
I feel tense	0.36**
I feel worried	0.35**
I feel anxious	0.38**
I feel nervous	0.37**

**Table 4**  
Effect sizes and number needed to treat for EMA and paper-and-pencil mindfulness, depression, and anxiety total score.

	Baseline mean (SD)	Post-mean (SD)	Condition (C) Time (T) Condition x time (TxC)	Cohen's d	NNT
<b>Mindfulness</b>					
<i>EMA assessment</i>					
MBSR	12.9(2.7)	14.6(3.0)	C = F(1,66.8) = 0.5(0.5)	0.4	7.5
Health education	13.2(2.7)	13.5(2.8)	T = F(1,2614.3) = 200.3(<0.001) TxC = F(1,2614,25) = 92.52(<0.001)		
<i>Paper-and-pencil assessment</i>					
MBSR	10.8(2.4)	12.7(2.8)	C = F(1,126) = 0.0(0.9)	0.2	13.6
Health education	11.2(2.2)	12.1(2.4)	T = F(1,126) = 10.7(0.0) TxC = F(1,126) = 1.2(0.3)		
<b>Depression</b>					
<i>EMA assessment</i>					
MBSR	6.9(3.4)	5.7(2.5)	C = F(1,67.2) = 1.9(0.2)	0.4	8.2
Health education	7.4(4.2)	7.2(3.9)	T = F(1,2934.2) = 97.8(<0.001) TxC = F(1,2934.2) = 37.6(<0.001)		
<i>Paper-and-pencil assessment</i>					
MBSR	9.4(3.8)	7.9(3.8)	C = F(1,126) = 0.2(0.7)	0.1	31.1
Health education	9.5(3.6)	8.3(3.8)	T = F(1,126) = 4.3(0.0) TxC = F(1,126) = 0.0(0.9)		
<b>Anxiety</b>					
<i>EMA assessment</i>					
MBSR	8.5(4.1)	7.1(3.9)	C = F(1,67.2) = 2.7(0.1)	0.4	7.7
Health education	9.4(4.5)	8.8(4.3)	T = F(1,2904.8) = 99.2(<0.001) TxC = F(1,2904.8) = 19.9(<0.001)		
<i>Paper-and-pencil assessment</i>					
MBSR	12.4(3.2)	10.9(3.9)	C = F(1,126) = 2.2(0.1)	0.4	7.3
Health education	12.9(3.5)	12.3(3.7)	T = F(1,126) = 2.7(0.1) TxC = F(1,126) = 0.6(0.4)		

Note. NNT = Number Needed to Treat.



indicated a significant difference between the MBSR and health education conditions when depression was measured with EMA, but not when depression was measured via paper-and-pencil (Table 4). The NNT when using EMA to assess depression pre- and post-treatment is 8.2, whereas the NNT was 31.1 when using paper-and-pencil surveys as the assessment method. Lastly, within-group contrasts for anxiety also indicated a significant difference between treatment conditions using EMA methods but not using paper-and-pencil methods (Table 4). However, NNTs were similar across these two methods of assessing anxiety (NNT = 7.7 with EMA; NNT = 7.3 with paper-and-pencil).

#### 4. Discussion

This study compared the sensitivity to change of clinical symptoms among psychologically distressed older adults across two different assessment methods: ecological momentary assessment (EMA) versus traditional paper-and-pencil measures. Results indicated greater improvement in mindfulness, depression, and anxiety in the MBSR intervention than the control intervention when symptoms were measured via EMA, but these effects were not seen for depression and mindfulness on the corresponding paper-and-pencil measures. Use of the NNT statistic indicated that out of every eight distressed older adults treated with MBSR, one older adult would demonstrate a clinically meaningful improvement in mindfulness compared with the health education condition, when change in mindfulness is measured via EMA. Comparatively, the NNT for change in mindfulness when mindfulness was measured via paper-and-pencil measures increased to 13.6, indicating that fourteen older adults would have to be treated with MBSR for one older adult to demonstrate a clinically meaningful improvement in mindfulness. The difference in NNT between EMA and paper-and-pencil methods were even more pronounced for clinically meaningful change in depression, with a NNT of 8.2 when depression was measured via EMA compared to a NNT of 31.1 when depression was measured via paper-and-pencil measures. These findings are particularly striking when considering that the *same items*, only differing in frame of reference, were administered across both assessment methods. While overall pre- to post-treatment differences were found for anxiety based on assessment method, differences in the NNTs were not observed. Our study indicates EMA measures of depression and mindfulness may be more sensitive to change in patient-reported outcomes following a mindfulness training intervention. In clinical trials, EMA could increase the precision of detecting and quantifying clinically significant effects, which may offset its additional subject and investigator burden via increased power and corresponding smaller sample sizes required.

The NNT is one of the most common ways for researchers and clinicians, as well as policy makers and grant funders, to understand the impact an intervention has on patients. Our results indicate that the NNT appears to be quite dependent on the manner in which the study outcome (in this case, mindfulness and depression) is measured. In a recent systematic review and meta-analysis of the literature on mediation programs for psychological stress, Goyal et al. (2014) found small to moderate effects of mindfulness interventions in improving stress or other mood or stress-related symptoms. The NNTs for paper-and-pencil measures in our study were high for depression and mindfulness (31 and 14, respectively), and, if only this modality has been used for assessing symptoms, may have reinforced the Goyal et al. conclusion regarding the modest effect size of these therapies.

There are several ways in which this difference in results based on administration modality may have arisen, including the capacity for EMA to indicate a greater mean difference and/or to tighten the

variability by virtue of repeated measurement. Inspection of the mean predicted values and dispersion estimates indicates it is most likely that EMA outperformed paper-and-pencil instruments by reducing the variability of estimates. This makes intuitive sense in that repeated measurement diminishes the possibility of state effects on point-in-time measures that can override the 'signal' produced by the study interventions. This narrowing of the estimation variability and its potential impact has been previously described in the estimation of cognitive ability with older adults with paper-and-pencil and EMA-administered cognitive assessments; mobile cognitive assessments were able to detect change with a smaller number of older adults than were standard point-in-time cognitive assessment instruments (Allard et al., 2014). The advantages of measurement methods with of greater sensitivity to change include more precise estimates of effect sizes, smaller sample size requirements to detect main intervention effects, and greater opportunity to detect subgroup effects (or moderator effects), which are critical to moving interventions towards personalized medicine. However, additional research on the equivalence and construct validity of EMA measures would be necessary, such as by comparison of different forms of administration to external criteria (e.g., clinician ratings). Some research has begun to indicate that EMA measures are indeed more convergent with clinician reports than are paper and pencil reports (for e.g., Depp et al., 2012). Moreover, the differential sensitivity of EMA may differ by outcome – in our study sensitivity to change in anxiety was comparable across EMA and paper-and-pencil outcomes. It is likely that recall biases or intra-subject variability may be more or less prominent depending upon the construct being assessed.

There are several limitations to this study. First, 26 (28%) of our participants did not complete ten or more EMA surveys at baseline or follow-up and were excluded from study analysis. Although these adherence rates are consistent with those reported in other EMA studies (e.g., Depp et al., 2012; Granholm et al., 2008), there is a possibility that this restricted sample biased the study results. Completion of brief paper-and-pencil measures produces far less subject burden than does EMA, and so investigators weighing advantages of EMA and paper-and-pencil administrations must consider the risk of subject loss and poor protocol adherence. Although EMA tools have evolved to become easier to respond to via touch screens and simple interfaces, further studies may benefit from incorporating real-time motivational incentives for completing surveys (such as micro-payments for each completed assessment, or a compliance-dependent financial bonus at the end of the study). Additionally, having research staff call participants on the first couple of days of EMA survey collection can help identify any problems participants may be having with the technology and promote adherence. Second, the 12 mindfulness, depression, and anxiety items were administered by themselves via EMA but embedded in longer scales when administered in paper-and-pencil format. The PROMIS measures (depression and anxiety) were developed to be unidimensional (i.e., that each item has good psychometric characteristics and investigators can mix and match the items), so the four-item subset should not have greatly influenced study outcomes. For the CAMS-R, it is possible that differences in results between EMA and paper-and-pencil administration were due to the effects that the additional paper-and-pencil items may have had on response patterns to the 4 items of interest. Additionally, we converted the wording of the EMA items to reflect present state, whereas the established paper-and-pencil administrations reflect past state. Our thesis is that frequent repeated assessments aggregated over time improve accuracy of measurement over retrospective point-in-time estimates, but we acknowledge that the difference in wording between modalities could have impacted results. Encouragingly, good reliability was found for both

CAMS-R EMA and paper-and-pencil administrations, as evidenced by the Cronbach's alpha values. Another limitation to address is that the clinical trial was not specifically designed to assess differences in the responsiveness to change in EMA versus retrospective paper-and-pencil reports. As such, we did not directly compare effect sizes or responsiveness as we would have done in a psychometrically focused experiment. To compare NNT effect sizes, a study would need to be designed and powered to detect differences in the psychometric properties of the instrument. However, because this is one of the only clinical trials, to our knowledge, to have available virtually parallel forms of commonly used measures of distress and mindfulness in EMA and retrospective formats, we hope that this paper can stimulate future research on the use of EMA in assessing clinical end-points in studies specifically designed for this purpose. Finally, we do not know from this study that the findings generalize beyond trials of MBSR or the older population employed here; it seems likely that EMA's benefits go beyond a specific outcome measure, population, or intervention type but this should be explicitly tested. We envision that a transition period of sorts should occur over the next 5–10 years, in which both EMA or other ambulatory assessment methods and traditional retrospective methods of measurement are used and compared head-to-head as part of clinical trials research.

In conclusion, our study provides initial evidence that ambulatory mobile assessment could enhance the ability to detect change in patient-reported outcomes in clinical trials when compared to standard paper-and-pencil administration. We encourage future interventions research to use more sensitive measures to assess treatment outcomes, in order to more directly examine whether an intervention demonstrates a clinically meaningful change in time-varying mindfulness and mood symptoms. Clinical trials are the critical step for determining whether scientific discoveries translate into public benefits, and one of the most important components of clinical trial methodology is getting a precise measurement of the outcomes, a necessary step for determining benefits of interventions. The implications of using EMA to improve outcome precision cannot be overstated, in terms of potential benefits for patients, care providers, and the public.

### Submission declaration

The authors assert that this work has not been published previously, this it is not under consideration for publication elsewhere, that its publication is approved by all authors, and that, if accepted, it will not be published elsewhere including electronically in the same form, in English or in any other language, without the written consent of the copyright-holder.

### Contributors

Raeanne Moore carried out the statistical analyses and wrote the paper. Colin Depp supervised statistical analyses and assisted with writing the paper. Julie Wetherell and Eric Lenze conceptualized the design of the study, supervised data collection, and assisted with writing the paper. All authors contributed to and have approved the final manuscript.

### Role of funding source

This work was supported primarily by National Institutes of Health (EJL, grant number R34AT007064 and AG049369, JLW, grant number R34AT007070, RCM, grant number K23MH107260, and CAD, grant number MH100417). Additional funding came from the Taylor Family Institute for Innovative Psychiatric Research (EJL).

### Conflicts of interest

EJL receives (past or present) non-governmental research support from the McKnight Brain Research Foundation, Taylor Family Institute for Innovative Psychiatric Research, Barnes-Jewish Foundation, Roche, Lundbeck, and Sidney R Baer Foundation.

### References

- Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., et al., 2014. Mobile technologies in the elderly detection of cognitive decline. *PLoS One* 9, e112197.
- Axelson, D., Bertocci, M., Lewin, D., Trubnick, L., Birmaher, B., Williamson, D., et al., 2003. Measuring mood and complex behavior in natural environments: use of ecological momentary assessment in pediatric affective disorders. *J. Child Adolesc. Psychopharmacol.* 13, 253–266.
- Baer, R.A., Walsh, E., Lykins, E.L.B., 2009. Assessment of Mindfulness. In: *Clinical Handbook of Mindfulness*, pp. 153–168.
- Bierman, E., Comijs, H., Jonker, C., Beekman, A., 2005. Effects of anxiety versus depression on cognition in later life. *Am. J. Geriatr. Psychiatry* 13, 686–693.
- Bishop, S., Lau, M., Shapiro, S., Carlson, L., Anderson, N., Carmody, J., et al., 2004. Mindfulness: a proposed operational definition. *Clin. Psychol. Sci. Pract.* 11, 230–241.
- Bjorner, J., Rose, M., Gandek, B., Stone, A., Junghaenel, D., Ware, J.J., 2013. Difference in method of administration did not significantly impact item response: an IRT-based analysis from the patient-reported outcomes measurement information system (PROMIS) initiative. *Qual. Life Res.* 23 (1), 217–227.
- Cain, A.E., Depp, C.A., Jeste, D.V., 2009. Ecological momentary assessment in aging research: a critical review. *J. Psychiatric Res.* 43, 987–996.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al., 2007. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. *Med. Care* 45, S3e11.
- Collins, R., Kashdan, T., Gollnisch, G., 2003. The feasibility of using cellular phones to collect ecological momentary assessment data: application to alcohol consumption. *Exp. Clin. Psychopharmacol.* 11, 73.
- Depp, C.A., Kim, D.H., de Dios, L.V., Wang, V., Ceglowski, J., 2012. A Pilot study of mood ratings captured by mobile phone versus paper-and-pencil mood charts in bipolar disorder. *J. Dual Diagn.* 8, 326–332.
- Ebner-Priemer, U., Trull, T., 2009. Ambulatory assessment: an innovative and promising approach for clinical psychology. *Eur. Psychol.* 14, 109.
- Feingold, A., 2009. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychol. Methods* 14, 43e53.
- Feldman, G., Hayes, A., Kumar, S., Greeson, J., Laurenceau, J.P., 2007. Mindfulness and emotion regulation: the development and initial validation of the cognitive and affective mindfulness scale revised (CAMS-R). *J. Psychopathol. Behav. Assess.* 29, 177–190.
- Feldt, L.S., Woodruff, D.J., Salih, F.A., 1987. Statistical inference for coefficient alpha. *Appl. Psychol. Meas.* 11.
- Furukawa, T.A., Leucht, S., 2011. How to obtain NNT from Cohen's d: comparison of two methods. *PLoS One* 6, e19070.
- Goyal, M., Singh, S., Sibinga, E.M., Gould, N.F., Rowland-Seymour, A., Sharma, R., et al., 2014. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. *JAMA Intern. Med.* 174, 357–368.
- Granholt, E., Loh, C., Swendsen, J., 2008. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. *Schizophr. Bull.* 34, 507–514.
- Gwaltney, C., Shields, A., Shiffman, S., 2008. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 11, 322–333.
- Hopko, D.R., Stanley, M.A., Reas, D.L., Wetherell, J.L., Beck, J.G., Novv, D.M., et al., 2003. Assessing worry in older adults: confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychol. Assess.* 15, 173–183.
- Johnson, E., Grondin, O., Barrault, M., Faytout, M., Helbig, S., Husky, M., et al., 2009. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. *Int. J. Methods Psychiatry Res.* 18, 48–57.
- Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., Schimmel, H., 1983. Validation of a short orientation-memory concentration test of cognitive impairment. *Am. J. Psychiatry* 140, 734–739.
- Lau, M., Bishop, S., Segal, Z., Buis, T., Anderson, N., Carlson, L., et al., 2006. The Toronto mindfulness scale: development and validation. *J. Clin. Psychol.* 62, 1445–1467.
- Lenze, E., Hickman, S., Hershey, T., Wendleton, L., Ly, K., Dixon, D., et al., 2014. Mindfulness-based stress reduction for older adults with worry symptoms and co-occurring cognitive dysfunction. *Int. J. Geriatr. Psychiatry* 29, 991–1000.
- Lenze, E., Wetherell, J.L., 2009. Bringing the bedside to the bench, and then to the community: a prospectus for intervention research in late-life anxiety disorders. *Int. J. Geriatr. Psychiatry* 24, 1–14.
- Lorig, K., Holman, H., Sobel, D., Laurent, D., Gonzales, V., Miner, M., 2012. Living a Healthy Life with Chronic Conditions: Self-management of Heart Disease,



- Arthritis, Diabetes, Depression, Asthma, Bronchitis, Emphysema and Other Physical and Mental Health Conditions, fourth ed. Bull Publishing Company.
- Moskowitz, D., Young, S., 2006. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *J. Psychiatry Neurosci.* 31, 13–20.
- Orsillo, S., 2005. Acceptance and Mindfulness-based Approaches to Anxiety: Conceptualization and Treatment.
- Reeve, B.B., Hays, R.D., Bjorner, J.B., Cook, K.F., Crane, P.K., Teresi, J.A., et al., 2007. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med. Care* 45, S22e31.
- Shiffman, S., Stone, A.A., Hufford, M.R., 2008. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4, 1–32.
- Slater, C., Bick, D., 1994. Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation. *JAMA* 271, 1377.
- SPSS, 2010. SPSS Statistics Base 22.0 ed. IBM Corporation, Chicago, IL.
- Stahl, B., Goldstein, E., 2010. *A Mindfulness-based Stress Reduction Workbook*. New Harbinger Press.
- Starr, L., Davila, J., 2012. Cognitive and interpersonal moderators of daily co-occurrence of anxious and depressed mood in generalized anxiety disorder. *Cognit. Ther. Res.* 26, 655–669.
- Trull, T., Ebner-Priemer, U., 2009. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. *Psychol. Assess.* 21, 457–462.
- Wetherell, J.L., Hershey, T., Hickman, S., Tate, S.R., Dixon, D., Bower, E.S., Lenze, E.J., 2016. Mindfulness-based stress reduction for older adults with stress disorders and neurocognitive difficulties: a randomized controlled trial. Under Revision.