Article

When Hindsight Is Not 20/20: Ecological Momentary Assessment of PTSD Symptoms Versus Retrospective Report

Assessment
2021, Vol. 28(1) 238–247

DOI: 10.1177/1073191119869826
journals.sagepub.com/home/asm
SAGE

Keke Schuler[1,2], Camilo J. Ruggero[1], Brittain Mahaffey[3], Adam Gonzalez[3], Jennifer L. Callahan[1], Adriel Boals[1], Monika A. Waszczuk[3], Benjamin J. Luft[3], and Roman Kotov[3]

Abstract

Assessment of posttraumatic stress disorder (PTSD) has relied almost exclusively on retrospective memory of symptoms, sometimes over long intervals. This approach creates potential for recall bias and obscures the extent to which symptoms fluctuate. The aim of the present study was to examine the discrepancy between retrospective self-reporting of PTSD symptoms and ecological momentary assessment (EMA), which captures symptoms closer to when they occur. The study also sought to estimate the degree to which PTSD symptoms vary or are stable in the short-term. World Trade Center responders ($N = 202$) oversampled for current PTSD (19.3% met criteria in past month) were assessed three times a day for 7 consecutive days. Retrospective assessment of past week symptoms at the end of the reporting period were compared with daily EMA reports. There was correspondence between two approaches, but retrospective reports most closely reflected symptom severity on the worst day of the reporting period rather than average severity across the week. Symptoms varied significantly, even within the span of hours. Findings support intervention research efforts focused on exploiting significant, short-term variability of PTSD symptoms, and suggest that traditional assessments most reflect the worst day of symptoms over a given period of recall.

Keywords

PTSD, ecological momentary assessment, retrospective reports, recall bias, assessments

Assessment of posttraumatic stress disorder (PTSD; American Psychiatric Association, 2013) symptoms in both clinical and research settings typically relies on retrospective reporting. This approach is standard and involves asking participants to report on symptoms in a global manner and that may have occurred days or weeks earlier (Schneider & Stone, 2016; J. P. Wilson & Keane, 2004). Despite its ubiquity, little is known about how well this standard manner of assessment reflects symptoms of PTSD as they occur in daily life.

Four limitations inherent to retrospective assessments can impact the validity and reliability of responses and obscure the nature of symptom burden. First, retrospective recall is vulnerable to cognitive biases (Shields, Shiffman, & Stone, 2016; Shiffman, Stone, & Hufford, 2008). For example, heuristic processes, such as peak-end rules, whereby the most intense or most recent experiences disproportionately influence recall, may affect reporting of symptoms (Schneider & Stone, 2016; Shiffman et al., 2008). Studies of pain patients showed higher average pain intensity reporting from recall than rated in daily reports (Stone et al., 2003; Stone, Schwartz, Broderick, & Shiffman, 2005). Second, affect at the time of a retrospective assessment may lead to greater recall of affect–congruent symptoms, with negative mood leading to reporting of more negative than actual or positive information (Clark & Teasdale, 1982; Shields et al., 2016; Shiffman et al., 2008). This bias may be more evident in recalling symptoms of PTSD, which is characterized with high negative affectivity (J. P. Wilson & Keane, 2004). Third, assessments typically occur in clinic or research settings, making them susceptible to contextual factors of test administration (Schwarz, 2012; Shields et al., 2016). Fourth, retrospective reporting of PTSD symptoms forces patients to

[1]University of North Texas, Denton, TX, USA
[2]Uniformed Services University of the Health Sciences, Bethesda, MD, USA
[3]Stony Brook University, Stony Brook, NY, USA

Corresponding Author:
Camilo J. Ruggero, Department of Psychology, University of North Texas, 1155 Union Circle #311280, Denton, TX 76203, USA.
Email: camilo.ruggero@unt.edu

Table 1. Six Possible PTSD Symptom Patterns.

PTSD symptom score	Description of calculation	Calculation formula
Average symptoms	The average of all momentary PTSD symptom scores across the recall period	$\Sigma(PTSD_1 + PTSD_2 + \ldots + PTSD_{20})/20$
Peak symptoms	The single maximum PTSD symptom score reported across all momentary assessments from the recall period	$\max(PTSD_1, PTSD_2, \ldots PTSD_{20})$
Primacy symptoms	The average momentary PTSD symptom scores in the first 2 days of the recall period	$\Sigma(PTSD_1 + PTSD_2 + \ldots + PTSD_5)/5$
Recency symptoms	The average momentary PTSD symptom scores in the past 2 days of the recall period	$\Sigma(PTSD_{15} + PTSD_{16} + \ldots + PTSD_{20})/6$
Average of daily peak symptoms	Each day's maximum PTSD symptom score was recorded, then these maximum scores were averaged across all days of the recall period	Step 1: $max_1 = \max(PTSD_1, PTSD_2) \ldots max_7 = \max(PTSD_{18}, PTSD_{19}, PTSD_{20})$. Step 2: $\Sigma(max_1 + max_2 + \ldots + max_7)/7$
Worst day symptoms	Each day's mean PTSD symptom score was recorded, then a single maximum PTSD symptom score was calculated across these mean scores within the recall period	Step 1: $mean_1 = \Sigma(PTSD_1 + PTSD_2)/2 \ldots mean7 = \Sigma(PTSD_{18} + PTSD_{19} + PTSD_{20})/3$ Step 2: $\max(mean_1, mean_2, \ldots mean_7)$

Note. PTSD = posttraumatic stress disorder. $PTSD_1$ to $PTSD_{20}$ denotes to PTSD symptom severity assessed at each time point during the EMA period.

aggregate reporting over long intervals (Shiffman et al., 2008), making it unclear whether symptom burden is chronic or episodic.

Ecological momentary assessment (EMA; Stone & Shiffman, 1994) methods overcome the aforementioned limitations to some extent by repeatedly assessing symptoms closer in time to when they occur in individuals' natural environment (Shiffman et al., 2008). Specifically, EMA has advantages of reducing certain cognitive biases and increasing ecological validity of self-report (Schneider & Stone, 2016; Shiffman et al., 2008). Such an approach can gauge the extent to which a symptom profile portrayed from clinic-based (or research-based) assessments of PTSD symptoms diverges from daily experience of them, and the extent to which symptoms fluctuate across time. Despite the advantages, EMA has limitations. The assessment is time consuming and therefore, could introduce burden for study participants (Schneider & Stone, 2016). Self-report EMA, in particular, can also be biased by various factors, such as characteristics of study participants, assessment time, or emotion state at the time of assessment (Schneider & Stone, 2016; Shiffman et al., 2008). However, in comparison with other single-point assessment methods or other types of ambulatory assessment (e.g., end-of-day reports), EMA is characterized by the least amount of recall bias (Schneider & Stone, 2016).

To date, only a few studies of PTSD have contrasted retrospective and EMA approaches, focusing broadly on congruence and reporting mixed results. Three studies, two using samples of trauma-exposed individuals (Carlson et al., 2016; Naragon-Gainey, Simpson, Moore, Varra, & Kaysen, 2012), and the other using a small sample of veterans (Westermeyer et al., 2015) revealed good consistency between the two approaches. Specifically, previous work has reported that retrospective reporting of PTSD symptoms explained approximately 55% to 70% of the variance in daily PTSD symptoms (Naragon-Gainey et al., 2012). Other studies have focused only on intrusive memories of a traumatic event and reported mixed results, with one finding traditional assessment overreporting daily symptoms (Kleim, Graham, Bryant, & Ehlers, 2013), and the other underreporting symptoms (Priebe et al., 2013). Furthermore, Westermeyer et al. (2015) attempted to determine certain daily PTSD symptoms (i.e., average daily symptoms, the most severe daily symptoms, and recent daily symptoms) best captured by retrospective reporting. They found that it closely matched the most severe daily symptoms; however, they only considered limited types of PTSD symptoms and only in a small sample of 17 participants.

Hence, existing literature is mixed and often based on small samples, leaving it unclear what exactly is being captured when clinicians or researchers use retrospective techniques to assess PTSD symptoms. Table 1 outlines six possible PTSD symptom patterns that a retrospective assessment may be capturing, operationalized for the present study. They include the overall average symptom severity score ("average symptoms"), the worst symptoms of the entire interval ("peak symptoms"), the average of the worst daily symptoms ("average of daily peak symptoms"), symptoms reflecting either the initial part of the reporting interval ("primacy symptoms"), or the most recent ones ("recency symptoms"), and the average symptoms experienced on the worst day of the entire interval ("worst day symptoms"). To date, no studies have tested which of these PTSD symptom patterns is best captured by traditional, retrospective methods.

Beyond congruence, few studies have examined the extent to which retrospective assessments obscure short-term fluctuations in symptoms (Black et al., 2016; Naragon-Gainey et al., 2012). Short-term symptom fluctuations are common in a range of medical and psychiatric conditions, including PTSD. The extent of such fluctuations is important to understand both for research in guiding emerging EMA interventions (Shields et al., 2016; Shiffman et al., 2008) and evaluating treatment efficacy (Shields et al., 2016). Black et al. (2016) examined the fluctuation of PTSD symptoms over a 28-day period and found substantial within-person, day-to-day variability. Naragon-Gainey et al. (2012) examined the fluctuation of each PTSD symptom dimension separately over a 1-month period, and found substantial overall daily variability, with the greatest fluctuations occurring for avoidance symptoms. Although informative, both studies had limitations, notably small sample size ($N = 9$; Black et al., 2016) and the inclusion of a nonclinical sample (i.e., undergraduates with a history of trauma exposure; Naragon-Gainey et al., 2012).

The present study extends existing research with data from a sample of responders to the World Trade Center (WTC) disaster. WTC responders were exposed to significant physical and emotional trauma and a substantial portion (9.7%) had WTC-related *Diagnostic and Statistical Manual of Mental Disorders–Fourth edition* (*DSM-IV*) PTSD more than a decade later (Bromet et al., 2016). The present study had two aims. The first aim of the study was to investigate the correspondence between the retrospective assessment and the aforementioned six types of PTSD symptom patterns obtained from the EMA reports. Based on prior research, we hypothesized retrospective assessments would be broadly correspondence with EMA assessments, but most closely match either momentary peak symptoms or the worst symptom severity level experienced on a given day. The second aim was to understand the PTSD symptom fluctuations over the EMA period and how such fluctuations were associated with the symptom severity. In the present study, the PTSD symptom fluctuations were characterized by three indices: variability, instability, inertia (see Analytic Plan for detail), each representing a different aspect of symptom fluctuations. We hypothesized that participants who had a PTSD diagnosis or presented with more severe PTSD symptoms at baseline would show greater symptom fluctuations, as indexed in higher levels in variability and instability.

Method

Participants and Procedure

The sample consisted of 202 participants ($M_{age} = 54.28$, $SD = 9.69$) oversampled for current PTSD ($n = 39$, 19.3%) and recruited from the Long Island site of WTC Health Program (WTC-HP) for an EMA study between October 2014 and February 2016. This is an unselected sample recruited from the WTC-HP, which monitors thousands of responders who worked or volunteered as a part of rescue, recovery, or cleanup of the WTC sites (Dasaro et al., 2015). Several recruitment methods were used, including flyers in the WTC clinic and study information presented directly to patients by nurses. The sample was primarily male (82.7%, $n = 167$), White (88.1%, $n = 178$), and non-Hispanic (80.7%, $n = 163$), with an average of 14.82 ($SD = 2.26$) years of education. Most of the participants were current or former police (62.9%, $n = 127$) and almost half (48.5%, $n = 98$) were retired, although many continued to work part-time. All participants provided written informed consent. The study was approved by the Stony Brook University Committees on Research Involving Human Subjects.

At baseline, participants completed a battery of self-report questionnaires and subsequently completed diagnostic interviews. Following the baseline assessment, participants completed 7 days of EMA on an iPod provided by the research site. All participants completed EMA surveys three times a day (i.e., mid-morning, afternoon, and evening before bed) over 7 consecutive days. Participants were prompted prior to the designated assessment times and prompts were personalized individually. Specifically, prior to start the EMA, participants were asked to provide information on their availabilities for the next 7 days. The assessment times and prompts were then tailored based on this information. Although a few variations of assessment time occurred, the assessment time tended to be at the same fixed time for all the participants. In the present study, the average EMA compliance rate was 93.8% (ranged from 55% to 100%). At the end of the EMA, participants completed retrospective assessment of their 7-day symptoms electronically.

Measures

PTSD Diagnosis. The structured clinical interview for *DSM-IV* (SCID; First, Spitzer, Gibbon, & Williams, 1997) was used for PTSD diagnosis at baseline. The SCID interviews were administrated by trained master-level interviewers, who were closely supervised by two clinical psychologists (C.R. and R.K.). Previous assessments of interrater reliability in this population demonstrated very good interrater agreement ($\kappa = 0.82$; Bromet et al., 2016).

EMA PTSD Symptoms. Eight items drawn from the PTSD checklist for *DSM-5* (PCL-5; Weathers et al., 2013) were used to assess PTSD symptoms during the EMA reporting period. Participants were instructed to "Tell us if you have had any of the following experiences in the past 5 hours" on a 5-point Likert-type scale from 1 (*not at all*) to 5 (*extremely*). Items were selected based on the King, Leskin, King, and

Weathers (1998) four-factor model: intrusion, avoidance, numbing, and hyperarousal (King et al., 1998), which has been received numerous empirical support over the past decades (e.g., Yufik & Simms, 2010). To minimize respondent burden, two items were chosen for each PTSD dimension. The eight items used in the EMA were as follows: "I had repeated, disturbing, and unwanted memories of the stressful experience" and "I felt very upset because something reminded you of the stressful experience" (intrusion); "I avoided memories, thoughts, or feelings related to the stressful experience" and "I avoided external reminders of the stressful experience" (avoidance); "I felt distant or cut-off from other people" and "I had strong negative feelings" (numbing); and "I felt jumpy or easily startled" and "I was 'super alert' or watchful or on guard" (hyperarousal). Using the equations provided by Mehl and Conner (2012), reliability for this abbreviated version of the PCL-5 was calculated within a multilevel framework. Specifically, for the eight PCL items used in the present study, the within-person reliability (i.e., reliability of change) was $R_C = .78$ and the between-person reliability was $R_{KF} = .99$.

Retrospective PTSD Symptoms. The PCL-5 (Weathers et al., 2013) was used at the retrospective assessment, tailored to inquire about only the period of EMA (i.e., past 7 days only). The PCL-5 is a self-report questionnaire consisting of 20 items corresponding to the 20 PTSD symptoms in *DSM-5* (American Psychiatric Association, 2013). Participants were instructed to rate "How much you have been bothered by that problem in the past 7 days" on a 5-point Likert-type scale from 1 (*not at all*) to 5 (*extremely*). Consistent with EMA, analyses included only the eight items administrated in the EMA. In the present study, the Cronbach's α of the scores from the eight EMA items was .95.

Analytic Plan

Prior to the main analyses, EMA responses were inspected. Responses were considered invalid if it was more than 2 hours from the designated assessment time participants were to complete it. For the invalid responses, only responses for the single time point were deleted, and the remaining responses for that person were retained. Due to the large amount of invalid responses in the very first assessment (i.e., midmorning survey of Day 1), data collected from this assessment were deleted for all participants, and therefore, were not included in the present analyses.

To examine the first hypothesis, the three analytic steps were followed. First, we calculated six variables to represent possible symptom patterns (see Table 1). Second, we examined correspondence between retrospective reporting and the six possible symptom patterns using correlation coefficients. Specifically, Pearson correlation was used to evaluate the association between momentary and retrospective assessments. Intraclass correlation coefficient (ICC) with absolute agreement was also used to capture the agreement between the two approaches. ICC is a widely used index to reflect the agreement between measures, with higher value indicating a better agreement between measures (Koo & Li, 2016). Based on Cohen's (1988) recommendations, in the present study, values of correlations or ICCs less than .30 were indicative of noncorrespondence between the two approaches. Differences in Pearson correlation coefficients and ICC coefficients were further examined using a Williams' (1959) test in *R* "psych" package. Finally, to investigate mean differences between retrospective and each of the six possible symptom patterns directly, a series of paired-sample *t* tests was performed. A Cohen's *d* was then calculated for each paired-sample *t* test to reflect standardized mean difference between retrospective assessment and each symptom pattern calculated in the present study.

To test the second hypothesis, the following analytic approaches were used. First, following the recommendations by Trull, Lane, Koval, and Ebner-Priemer (2015), we calculated three indices to represent variability, instability, and inertia, respectively. For variability, a within-person variability was computed for each participant (i.e., standard deviation based on each person's average momentary assessment; Wichers et al., 2010). This value represents the extent to which PTSD symptoms deviates from its mean for each participant (Trull et al., 2015; Wichers et al., 2010). For instability, we calculated root mean squared successive difference (RMSSD; Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009; Koval, Pe, Meers, & Kuppens, 2013). Although there are other indices of instability proposed in the literature (e.g., Jahng, Wood, & Trull, 2008), the RMSSD was chosen for the present study because it captures amplitude, frequency, and temporal dependency (Ebner-Priemer et al., 2009) and has been particularly used to characterize PTSD symptom instability (Naragon-Gainey et al., 2012). For inertia, a within-person autocorrelation value was calculated for each participant (Trull et al., 2015). Multilevel modeling was used to estimate the within-person autocorrelation of PTSD symptom severity assessed during the EMA. Specifically, following the approach described by Kuppens, Allen, and Sheeber (2010), we estimated a multilevel model with PTSD symptoms at time *t* was predicted by itself at the time *t*-1 at Level 1. Random effects in the intercept and slope were estimated at Level 2.

Second, Pearson correlations were used to examine associations between current PTSD diagnosis, retrospective PTSD symptom severity, and the three indices calculated above (i.e., within-person SD, RMSSD, and within-person autocorrelation). Given the binary nature of current PTSD diagnosis variable, independent-samples *t* test was then conducted to specifically examine whether variability (within-person SD), instability (RMSSD), and inertia differed between the two groups. Finally, to compare the

Table 2. Descriptive Statistics of PTSD Symptoms, Correlations, and Comparisons Between Retrospective and Indices of Momentary Assessments.

Assessments	*M* (*SD*)	*t*	Cohen's *d*	*r*	ICCs
Retrospective	1.93 (0.89)	—	—	—	—
EMA					
Average symptoms	1.60 (0.68)	11.57***	.42	.91***	.81***
Peak symptoms	2.18 (1.03)	−7.14***	−.26	.88***	.84***
Primacy symptoms	1.59 (0.67)	10.69***	.43	.86***	.76***
Recency symptoms	1.62 (0.72)	10.54***	.38	.89***	.81***
Average of daily peak symptoms	1.76 (0.78)	6.49***	.20	.91***	.89***
Worst day symptoms	1.92 (0.89)	0.424	.01	.89***	.89***

Note. $N = 202$. PTSD = posttraumatic stress disorder; EMA = ecological momentary assessment; r = Pearson correlation coefficients between retrospective assessment and momentary assessments; ICCs = intraclass correlation coefficients between retrospective and momentary assessments. Statistical analyses were paired-samples t tests comparing scores from retrospective with EMA assessments.
***$p < .001$.

within-person variability and between-person variability of PTSD symptoms, between-person variability across the sample was calculated using PTSD symptoms reported at the retrospective assessment. The comparison was operationalized by creating a ratio of the two variability values.

Results

Descriptive statistics of study variables, a summary of paired-samples t test results, and correlations between each possible symptom pattern and the retrospective assessment are presented in Table 2. Correlation results showed that all symptom patterns, except primacy symptoms, were highly correlated with retrospective reporting (rs = .89 to .91, ps $< .001$). The average of daily peak symptoms in the sample evidenced the highest correlation ($r = .91$) and was significantly higher than correlations with other EMA symptoms (i.e., peak symptoms, primacy symptoms, recency symptoms, and worst day symptoms; $t = -5.33$ to -2.42, ps $< .05$). However, ICCs revealed a much sharper difference, with the worst day symptoms showing the highest agreement with the retrospective reporting (ICC = .89). Importantly, this agreement was significantly higher than the agreements with the four other EMA symptom patterns (i.e., average, primacy, recency, and peak symptoms; $t = -6.44$ to -4.12, ps $< .001$) but did not differ from the agreement with the average of daily peak symptoms.

With respect to the mean-level differences between retrospective and each of the six possible symptom patterns, results showed that the average retrospective PTSD symptom score was lower than the average score of peak symptoms but higher than average scores of the rest of the five possible scores. Importantly, results from paired-sample t tests further revealed that the average retrospective score was significantly different from all the momentary scores, except the worst day symptom score. In other words, retrospective assessment most closely aligned with (i.e., did not significantly differ from) the worst day of symptoms.

Descriptive statistics and Pearson correlations for associations between current PTSD diagnosis, retrospective PTSD symptom severity, and the three indices representing variability, instability, and inertia, are summarized in Table 3. It is important to point out that these correlation coefficients represented bivariate associations between PTSD symptom severity and each index of symptom fluctuations. Both retrospective PTSD symptom severity and current PTSD diagnosis were significantly associated with variability (rs = .75 and .40, respectively; $p < .01$) and instability (rs = .73 and .42, respectively; $p < .01$), with more severe PTSD symptoms being associated with higher levels of variability and instability. Given the binary nature of current PTSD diagnosis variable, further independent-samples t test was conducted to compare the variability (within-person SD) and instability (RMSSD) between current PTSD group and non-PTSD group. Results showed that compared with the participants in non-PTSD group, participants who had a current PTSD diagnosis had significantly higher levels of variability ($M_{\text{variability}} = .44$, $p < .001$) and instability ($M_{\text{instability}} = .53$, $p < .001$) in PTSD symptoms over the course of EMA.

Retrospective PTSD symptom severity also showed a positive association with inertia, although the magnitude of this correlation was small ($r = .12$, $p < .05$). The association between inertia and current PTSD diagnosis was not significant. This result was further confirmed using an independent t test, which revealed that inertia did not differ between current PTSD group and non-PTSD group ($t = -.15$, $p = .882$). Past work has identified several limitations associated the two-step approach (i.e., inertia value is calculated for each participant and then used in a statistical test to examine group differences or associations with other variables) when analyzing inertia (e.g., Jahng et al.,

Table 3. Descriptive Statistics and Zero-Order Correlations for Indices of Fluctuations, Current PTSD Diagnosis and Retrospective Assessments.

Variables	*M* (*SD*) or % (*n*)	Correlations				
		1	2	3	4	5
1. Variability (within-person SD)	0.26 (0.22)	1.00				
2. Instability (RMSSD)	0.31 (0.25)	.96**	1.00			
3. Inertia (within-person autocorrelation)	0.31 (0.09)	.02	−.22**	1.00		
4. Retrospective PTSD symptoms	1.93 (0.89)	.75**	.73**	.15*	1.00	
5. Current PTSD diagnosis[a]	19.3% (39)	.40**	.42**	.02	.58**	1.00

Note. PTSD = posttraumatic stress disorder; RMSSD = root mean square successive difference.
[a]Given the binary nature of current PTSD diagnosis variable, the correlation reported here is the point biserial correlation coefficient.
*$p < .05$. **$p < .01$.

2008). To overcome the limitations, multilevel modeling was recommended (Jahng et al., 2008; Kuppens et al., 2010). Following this recommendation, inertia was then modeled within a multilevel framework. Results from the models showed positive significant autocorrelations, suggesting that PTSD symptoms at time *t* were significantly predicted by the symptoms at time *t*-1. At Level 2, both current PTSD diagnosis (estimate = .87, $p < .001$) and retrospective PTSD symptom severity (estimate = .33, $p < .001$) were significant positive predictors of individual differences in the intercept (i.e., initial symptom severity) for PTSD symptoms. The cross-level interaction represented how current PTSD diagnosis or retrospective PTSD symptoms was associated with individual differences in inertia for PTSD symptoms. Consistent with the Pearson correlation results, only retrospective PTSD symptoms (estimate = .06, $p = .004$) was significantly associated with the inertia for PTSD symptoms and the current PTSD diagnosis was not associated with the inertia for PTSD symptoms. However, it is worth pointing out that the effect of retrospective PTSD symptom severity on inertia was small, and therefore, this particular finding needs to be interpreted with caution.

To explicitly compare within-person variability and between-person variability of PTSD symptoms, a ratio of the two was calculated. Specifically, in the present sample, within-person variability ranged from 0 to 1.08, with an average of .26. In participants with current PTSD diagnosis, the average within-person variability of PTSD symptoms was .44. The between-person variability in the sample was .89. Therefore, the average within-person variability in PTSD symptoms across 7 days was approximately a third (29.2%) of the variability in symptoms observed between participants. For those diagnosed with current PTSD, the average within-person variability in PTSD symptoms across the 7 days was almost half (49.4%) of the between-person variability within the entire sample. Given that participants did not start on the same exact day of week, Figure 1 illustrates day of week (Figure 1A) and time of day (Figure 1B) trends in the present sample, stratified by presence/absence of a current PTSD diagnosis at baseline.

Discussion

The present study used EMA, a method that incorporates repeated measurement of symptoms closer in time to their occurrence, to investigate which pattern of PTSD symptoms is most congruent with traditional retrospective assessment. Two important findings emerged. First, consistent with prior studies and with our hypothesis, a traditional retrospective approach had the best congruence with symptoms experienced on the worst day of the reporting period, not with the overall average of symptoms experienced during that time. This finding suggests that traditional retrospective reports reflect more severe periods of PTSD symptoms rather than an aggregate of their day-to-day occurrence. Second, PTSD symptoms showed substantial fluctuations across time, and this was particularly true for individuals with current PTSD and who had more severe PTSD symptoms.

Consistent with prior work (Naragon-Gainey et al., 2012; Westermeyer et al., 2015), retrospective reports show an acceptable association with average EMA assessments. However, the present study revealed for the first time in the study of PTSD that retrospective reports more closely reflect PTSD symptoms experienced on the worst day of a reporting period. This close correspondence suggests that when recalling symptoms retrospectively, participants weigh more extreme experiences over others. Such a phenomenon would not be unique to PTSD. For example, EMA studies have found that patients tend to report their worst pain (Schneider, Stone, Schwartz, & Broderick, 2011) and that smokers tend to overestimate negative affect and their smoking behaviors during smoking lapses (Shiffman et al., 1997).

Results underscore how reporting of symptoms is in and of itself a dynamic rather than static behavior. Retrospective recall can be influenced by several factors (Hufford, 2007;

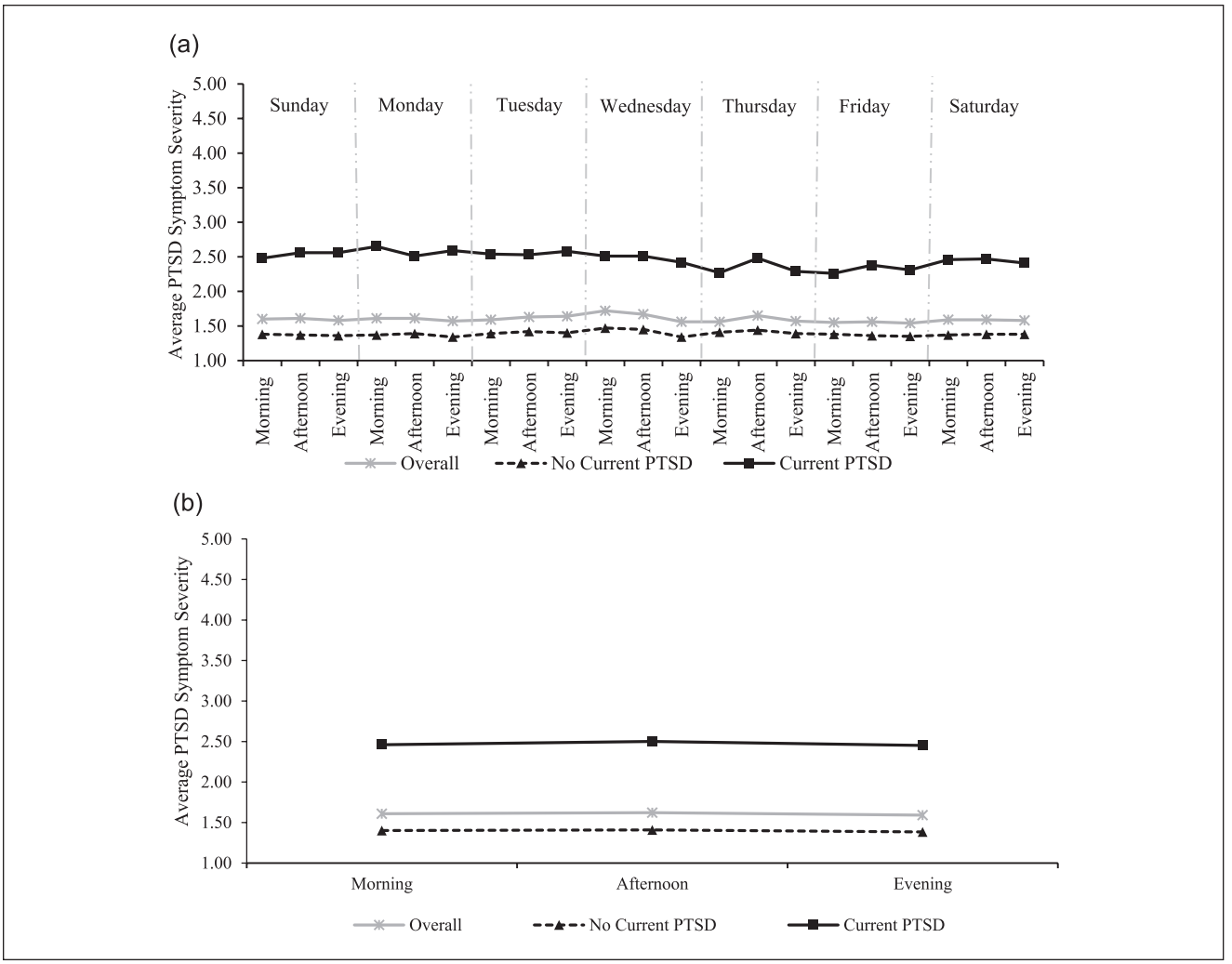

Figure 1. PTSD symptom severity (a) by day of the week and (b) by time of day.
Note. PTSD = posttraumatic stress disorder. Average PTSD symptom severity (ranged from 1 to 5) by day of the week (Figure 1A) and by the time of day in the current sample. Current PTSD diagnosis was based on *DSM-IV-TR* criteria. In the present study, Sunday was designated as the first day of week.

Van den Bergh & Walentynowicz, 2016), and a given report is not necessarily a precise mapping of the aggregate symptom burden over a given reporting period. This is particularly relevant for PTSD, where memory of the trauma plays a central role (e.g., Ono, Devilly, & Shum, 2016). Prior work has shown that memory for a trauma can change over time (Monfils & Holmes, 2018; A. Wilson & Ross, 2003) or that memory for other experiences can change (e.g., when reminded of a trauma, individuals memory for their personal attributes exhibited state-like changes over time; McFarland & Alvaro, 2000). Our work highlights how even memory for and reporting of symptoms related to that trauma are susceptible to influence (i.e., capturing peak rather than average experiences).

Importantly, there are psychometric and practical questions raised by the finding: given that EMA and retrospective reports capture different information, does EMA provide better clinical utility and is any such improvement worth the extra administrative burden? Given the ubiquity of smartphones (Pew Research Center, 2015) and the rapidly changing landscape of how clinical data can be gathered (for review, see Torous, Staples, & Onnela, 2015), these questions are becoming increasingly more relevant for the care of PTSD. Better congruence between an assessment method and a person's aggregate symptom burden though EMA is not inherently a marker of a more useful measure, or even a more valid one. For example, it might be the case that a measure that captures worst day experiences is also the one that more reliably differentiates patients' level of PTSD severity. Similarly, it is not certain which of the two measures—one capturing worst versus average symptoms—would have the better predictive

validity. That is, a retrospective measure could predict what it is supposed to in the future without necessarily corresponding to results obtained via EMA. Or, the predictive validity of one may be better for some questions (e.g., suicide risk), but worse for others (e.g., functioning). These scenarios were not tested in the present study, but findings provide impetus for future research to do so. At minimum, results point to EMA providing different information than traditional approaches, with future work needed to parse the clinical relevance of their difference to determine whether when a different assessment approach is merited.

Current findings also highlight the degree that PTSD symptoms fluctuate in the short-term. Importantly, using different indices, the present study was able to pinpoint the specific characteristics of PTSD symptom fluctuations. Results showed that PTSD symptom severity was positively related to variability, instability, and inertia. Current PTSD diagnosis was only related to variability and instability, but not inertia. Inertia reflects temporal dependency, which is the extent to which symptoms predict themselves from moment to moment (Trull et al., 2015). Previous work has suggested that PTSD symptoms could be easily triggered by environmental stimuli (e.g., loud noises; Naragon-Gainey et al., 2012), which would lead one to expect higher levels of variability in symptoms for individuals with more severe PTSD, but not high inertia. Yet the present results showed both. A few possibilities may explain this paradox: PTSD symptoms could be triggered by environmental factors, such as places, situations, or people, which may explain the high variability. However, once PTSD symptoms are triggered, they are persist. Furthermore, given the positive association between symptom severity and inertia, it might be that the persistence of PTSD symptoms is more evident in some individuals, but not all. The present study cannot speak to the mechanisms that potentially underlie variability and inertia, so future work is needed to describe the processes that may be related to variability, inertia, or both.

In addition to using different indices to understand symptom fluctuation, the present study contrasted within-person variability to between-person variability to gauge the relative magnitude of these short-term fluctuations. Findings suggested that the average variability in symptoms between assessments spanning a few hours was equivalent to a change in almost one full symptom of PTSD. However, more extreme fluctuations also occurred (e.g., from no symptoms at all to very severe symptoms of PTSD just within hours). Some of this within-person variability is likely to reflect measurement error, but a significant portion is also likely to reflect true changes in PTSD symptoms over time. Additionally, the current study adds novel evidence suggesting that this high within-person variability may be a clinical manifestation of PTSD, with fluctuations greater in those who had more severe initial PTSD symptoms. This contrasts to findings by Black et al. (2016) who reported no association between within-person variability of PTSD symptoms and symptom severity in a sample of nine veterans.

Findings regarding PTSD symptom fluctuations have potential clinical implications. The finding of symptom variability is important clinical information for providers and their patients, allowing them to anticipate and normalize these variations. They also underscore the need for work testing their possible association with environmental or contextual triggers. This type of work can pave the way for the development of effective EMA field-based interventions that capitalize on symptom occurrences in order to disrupt their maintenance and are tailored to individuals (e.g., interventions delivered during a person's peak symptom occurrence). Past work has speculated that PTSD symptoms may fluctuate the most during therapy (Naragon-Gainey et al., 2012). Implementing EMA during therapy may therefore help clinicians better understand factors that trigger or exacerbate PTSD symptoms. In fact, a fundamental shift in how the disorder is perceived may be warranted, moving from a more static view of burden to one in which symptoms are viewed as transient and episodic, with potential to mitigate the worst symptom periods. Eventually, clinicians may use information collected outside and inside the clinic to tailor treatment plans and promote treatment efficiency.

Findings should be interpreted within the context of several limitations. First, the reporting period was only 7 days, and was only for a subset of symptoms. Results may differ as the reporting period lengthens and covers broader sets of symptoms. Second, the present study primarily used analytic methods that assess differences in the mean levels between EMA and retrospective assessment to identify discrepancies. It is possible that these discrepancies also come from general concordance between measures (i.e., measurement reliability). However, given that the aim of the present study was to investigate whether retrospective assessment is an unbiased presentation of the symptoms experienced in prior week, an explicit examination of whether retrospective assessment is reliable is beyond the scope of the present study. Given the importance of this research question, further investigation is needed to understand retrospective assessment and EMA from a measurement reliability perspective. Third, it is unclear whether the correspondence between EMA and retrospective assessment revealed in the present study would hold in clinician-administered measures. Future research exploring this possibility is encouraged. Fourth, participants were instructed to report symptoms "in the past 5 hours" for each assessment, which required a certain degree of retrospection. However, past work has pointed out that recall or memory biases related to this type of retrospection is limited (Schneider & Stone, 2016). To further eliminate this type of retrospection, future

EMA work should consider combining self-report EMA with either more momentary reports or with objective symptom markers. Fifth, the present clinical sample was selected based on one specific trauma type and in a unique population. Findings from past research suggested that different trauma types might be associated with different PTSD symptomatology in daily life (e.g., Kleim et al., 2013). As such, to understand the role of different trauma types in daily presentation of PTSD symptoms, future research is needed to test these associations in diverse samples.

In conclusion, the present study using a clinical sample demonstrates that retrospective assessment of PTSD symptoms best captures PTSD symptom experienced on the worst day during a reporting period. Additionally, PTSD symptoms show substantial fluctuation in the short-term, which is more marked for those with greater PTSD symptom severity. Such findings provide a rationale and basis for future EMA-based research and for the possible clinical utility of adjunctive EMA assessments of PTSD symptoms.

Acknowledgments

We gratefully acknowledge the support of the Word Trade Center responders for generously contributing their time and energy to this project. We also thank Evelyn Bromet and Katherine Guerrera for facilitating the study, as well as study staff: Melissa Carr and Rachel Roger, Chris Ray and Peter Allen. Special thanks to Evelyn Bromet for her thoughtful comments and edits to the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was supported by the Center for Disease Control and Prevention (CDC)/the National Institute for Occupational Safety and Health (NIOSH) [Grants numbers 1U01OH010712-01 and 1U01OH011321-01] awarded to Roman Kotov. The funding sources had no involvement in conducting the research or in publication.

ORCID iDs

Camilo J. Ruggero https://orcid.org/0000-0002-8482-6269
Jennifer L. Callahan https://orcid.org/0000-0001-9190-3886

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. (5th ed.). Washington, DC: Author.

Black, A. C., Cooney, N. L., Justice, A. C., Fiellin, L. E., Pietrzak, R. H., Lazar, C. M., & Rosen, M. I. (2016). Momentary assessment of PTSD symptoms and sexual risk behavior in male OEF/OIF/OND veterans. *Journal of Affective Disorders*, *190*, 424-428.

Bromet, E. J., Hobbs, M. J., Clouston, S. A. P., Gonzalez, A., Kotov, R., & Luft, B. J. (2016). DSM-IV post-traumatic stress disorder among World Trade Center responders 11-13 years after the disaster of 11 September 2001 (9/11). *Psychological Medicine*, *46*, 771-783.

Carlson, E. B., Field, N. P., Ruzek, J. I., Bryant, R. A., Dalenberg, C. J., Keane, T. M., & Spain, D. A. (2016). Advantages and psychometric validation of proximal intensive assessments of patient-reported outcomes collected in daily life. *Quality of Life Research*, *25*, 507-516.

Clark, D. M., & Teasdale, J. D. (1982). Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. *Journal of Abnormal Psychology*, *91*, 87-95.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dasaro, C. R., Holden, W. L., Berman, K. D., Crane, M. A., Kaplan, J. R., Lucchini, R. G., . . . Udasin, I. G. (2015). Cohort profile: World trade center health program general responder cohort. *International Journal of Epidemiology*, *46*, 393-402. doi:10.1093/ije/dyv064

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in) stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology*, *118*, 195-202.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1997). *Structured clinical interview for DSM-IV clinical version*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Hufford, M. R. (2007). Special methodological challenges and opportunities in ecological momentary assessment. In A. A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-report in health research* (pp. 54-75). New York, NY: Oxford University Press.

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, *13*, 354-375.

King, D. W., Leskin, G. A., King, L. A., & Weathers, F. W. (1998). Confirmatory factor analysis of the clinician-administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment*, *10*, 90-96.

Kleim, B., Graham, B., Bryant, R. A., & Ehlers, A. (2013). Capturing intrusive re-experiencing in trauma survivors' daily lives using ecological momentary assessment. *Journal of Abnormal Psychology*, *122*, 998-1009.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155-163.

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, *13*, 1132-1141. doi:10.1037/a0033579

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984-991.

McFarland, C., & Alvaro, C. (2000). The impact of motivation on temporal comparisons: Coping with traumatic events by perceiving personal growth. *Journal of Personality and Social Psychology*, *79*, 327-343.

Mehl, M. R., & Conner, T. S. (Eds.). (2012). *Handbook of research methods for studying daily life*. New York, NY: Guilford Press.

Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: Opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *Lancet Psychiatry*, *5*, 1032-1042.

Naragon-Gainey, K., Simpson, T. L., Moore, S. A., Varra, A. A., & Kaysen, D. L. (2012). The correspondence of daily and retrospective PTSD reports among female victims of sexual assault. *Psychological Assessment*, *24*, 1041-1047.

Ono, M., Devilly, G. J., & Shum, D. H. K. (2016). A meta-analytic review of overgeneral memory: The role of trauma history, mood, and the presence of posttraumatic stress disorder. *Psychological Trauma: Theory, Research, Practice, and Policy*, *8*(2), 157-164.

Pew Research Center. (2015). *U.S. smartphone use in 2015*. Retrieved from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Priebe, K., Kleindienst, N., Zimmer, J., Koudela, S., Ebner-Priemer, U., & Bohus, M. (2013). Frequency of intrusions and flashbacks in patients with posttraumatic stress disorder related to childhood sexual abuse: An electronic diary study. *Psychological Assessment*, *25*, 1370-1376.

Schneider, S., & Stone, A. A. (2016). Ambulatory and diary methods can facilitate the measurement of patient-reported outcomes. *Quality of Life Research*, *25*, 497-506.

Schneider, S., Stone, A. A., Schwartz, J. E., & Broderick, J. E. (2011). Peak and end effects in patients' daily recall of pain and fatigue: A within-subjects analysis. *Journal of Pain*, *12*, 228-235.

Schwarz, N. (2012). Why researchers should think "real-time": A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22-42). New York, NY: Guilford Press.

Shields, A. L., Shiffman, S., & Stone, A. (2016). Recall bias: Understanding and reducing bias in pro data collection. In B. Byrom & B. Tiplady (Eds.), *EPro: Electronic solutions for patient-reported data* (pp. 5-21). New York, NY: Routledge.

Shiffman, S., Hufford, M., Hickcox, M., Paty, J. A., Gnys, M., & Kassel, J. D. (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting and Clinical Psychology*, *65*, 292-300.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1-32.

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain*, *104*, 343-351.

Stone, A. A., Schwartz, J. E., Broderick, J. E., & Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality and Social Psychology Bulletin*, *31*, 1340-1346.

Stone, A. A., & Shiffman, S. (1994). Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, *16*, 199-202.

Torous, J., Staples, P., & Onnela, J. P. (2015). Realizing the potential of mobile mental health: New methods for new data in psychiatry. *Current Psychiatry Reports*, *17*, 61. doi:10.1007/s11920-015-0602-0

Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, *7*, 355-361.

Van den Bergh, O., & Walentynowicz, M. (2016). Accuracy and bias in retrospective symptom reporting. *Current Opinion in Psychiatry*, *29*, 302-308.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD checklist for DSM-5 (PCL-5)*. Retrieved from https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp

Westermeyer, J., Shiroma, P., Thuras, P., Kattar, K., Johnson, D., & Crosby, R. D. (2015). Daily versus monthly reporting of post-traumatic symptoms: A study of reliability across time and instruments. *Psychiatry Research*, *227*, 309-312.

Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J. P., Derom, C., . . . Van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, *124*, 191-195.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *21*, 396-399.

Wilson, A., & Ross, M. (2003). The identity function of autobiographical memory: Time is on our side. *Memory*, *11*, 137-149.

Wilson, J. P., & Keane, T. M. (Eds.). (2004). *Assessing psychological trauma and PTSD*. New York, NY: Guilford Press.

Yufik, T., & Simms, L. J. (2010). A meta-analytic investigation of the structure of posttraumatic stress disorder symptoms. *Journal of Abnormal Psychology*, *119*, 764-776.