



Published in final edited form as:

Psychol Assess. 2009 September ; 21(3): 425–436. doi:10.1037/a0016869.

Clinical Assessment of Affective Instability: Comparing EMA indices, questionnaire reports, and retrospective recall

Marika B. Solhan, Timothy J. Trull, Seungmin Jahng, and Phillip K. Wood

Department of Psychological Sciences, University of Missouri-Columbia

Abstract

Traditional self-report measures of psychopathology may be influenced by a variety of recall biases. Ecological momentary assessment (EMA) reduces these biases by assessing individuals' experiences as they occur in their natural environments. This study examines the discrepancy between trait questionnaire, retrospective report, and EMA measures of affective instability in psychiatric outpatients either with a borderline personality diagnosis (BPD; $n=58$) or with a current major depressive episode or dysthymia (MDD/DYS; $n=42$). We examined the agreement of three trait measures of affective instability (*Personality Assessment Inventory-Borderline Features* scale – *Affective Instability* scale, *Affect Intensity Measure*, and the *Affect Lability Scales*) and one retrospective mood recall task with EMA indices of mood and mood instability. Results indicate only modest to moderate agreement between momentary and questionnaire assessments of trait affective instability; agreement between recalled mood changes and EMA indices was poor. Implications for clinical research and practice and possible applications of EMA methodology are discussed.

Keywords

Borderline Personality Disorder; affective instability; Ecological Momentary Assessment; retrospective report

Clinical researchers and practitioners primarily rely on retrospective self-reports of experiences such as emotions, symptoms, traits, and behaviors when evaluating an individual's functioning. Traditional questionnaire assessments have a variety of strengths, including ease of administration and cost-effectiveness. In addition, the clinical or diagnostic interview is a major tool used by clinical psychologists for evaluation and treatment planning. Both questionnaires and interviews, however, require the respondent to retrospectively integrate information from memory in order to rate items or answer questions. Unfortunately, retrospective reports are influenced by a variety of recall biases, which may introduce error and limit the usefulness of findings based on these reports (Hufford, 2007).

Alternate methodologies, such as *Ecological Momentary Assessment* (EMA; Stone & Shiffman, 1994), offer ways to circumvent (or at least minimize) many sources of recall bias and thereby provide the opportunity to assess the discrepancies between reports of recalled

Correspondence concerning this article should be addressed to Timothy J. Trull, Department of Psychological Sciences, 210 McAlester Hall, University of Missouri-Columbia, Columbia, MO 65211. Electronic mail may be sent to trullt@missouri.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/pas.

versus moment-to-moment emotional, behavioral, and cognitive experiences. EMA approaches are characterized by (1) the collection of data in real-world environments; (2) assessments that focus on individuals' current or very recent experiences; (3) assessments that may be event-based, time-based, or randomly-prompted; and (4) the completion of multiple assessments over time (Stone & Shiffman, 1994).

In the current study, we examined the degree of agreement of self-report questionnaire and retrospective recall measures of mood instability and intensity with EMA measures of affective instability. A wide variety of clinical conditions are characterized and defined by either intense negative mood states (e.g., sadness, anxiety, fear) or mood instability (marked changes in mood state over time). However, how well can patients retrospectively assess their mood states or report their trait mood instability? Several studies have documented discrepancies between real-time assessments and retrospective self-reports of mood, symptoms, traits, and behaviors, highlighting risks associated with sole reliance on retrospective accounts (Fahrenberg, Myrtek, Pawlik, & Perrez, 2007; Shiffman, Stone, & Hufford, 2007; Stone & Broderick, 2007). In fact, such discrepancies have been demonstrated consistently in a wide range of clinical problems, including chronic pain (Stone, Schwartz, Broderick, & Shiffman, 2005), coping with stress (Stone et al., 1998), smoking cessation (Shiffman, Hufford, Paty, Gnys, & Kassel, 1997), post traumatic stress disorder (Southwick, Morgan, Nicolaou, & Charney, 1997), major depression (Brewin, Andrews, & Gotlib, 1993), and personality disorders (Ebner-Priemer et al., 2006). Not only are retrospective reports on traits or experiences and momentary reports often discrepant, but the degree of error can be quite significant. For example, one study of health events found that participants accurately reported their medication use only sixty percent of the time over a three-month period, omitting from recall a wide range of prescription and over-the-counter medications they consumed (Cohen & Java, 1995).

Sources of Retrospective Bias

The cause of these discrepancies is multifaceted. When asked to reflect over a given period of time and report on particular experiences, individuals do not cognitively search for and tally all relevant instances of an experience. Rather, they engage in a complex process of estimating the frequency or intensity of a given experience automatically using a variety of heuristic strategies (Gorin & Stone, 2001). Depending on the type of recall task, individuals may attempt to mentally reproduce information about infrequent, salient, or extreme events, by recalling specific and unique details about these events. Or, they might attempt to mentally reconstruct information about past events, personality traits, or common behaviors, by constructing detailed reports based on their general beliefs about such experiences.

Commonly, individuals use some combination of reproduction and reconstruction strategies when recalling information, using reconstructed memories to fill the gaps in reproduced memories (Gorin & Stone, 2001). For example, an individual might report on her emotional functioning over the last week by recalling the intense anger she felt while fighting with her spouse and combining this reproduced memory with the belief that she always has intense moods and frequently feels angry (i.e., reconstructed generic personal memories). She may then conclude from this combination of memories that she felt intensely angry most of the week. Although useful and efficient, these strategies systematically distort the final estimation.

Significant sources of recall bias include, for example, the tendency to differentially weight certain experiences when averaging across all experiences (e.g., those consonant with current mood state; Ebner-Priemer et al., 2006; Kihlstrom, Eich, Sandbrand, & Tobias, 2000; Mayer, McCormick, & Strong, 1995; Mineka & Nugent, 1995; Southwick et al., 1997), the tendency for reports to be disproportionately influenced by the most recent or most intense aspect of an experience relevant to the construct being measured (also referred to as the peak-end rule; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Redelmeier & Kahneman, 1996),

and the tendency to reconstruct the timing, duration, or frequency of experiences based on generic personal memories and heuristics (Friedman, 1993; Gorin & Stone, 2001; Huttenlocher, Hedges, & Bradburn, 1990; Larsen & Thompson, 1995; Thompson, Skowronski, & Lee, 1988).

Of special importance to the present study, recall appears to be strongly influenced by variability in one's experience. If one's general experience is highly variable over time, specific instances of the experience will be more difficult to accurately recall. Illustrating this source of recall bias, Stone, Schwartz, Broderick, and Shiffman (2005) found greater concordance between momentary and retrospective ratings of pain in individuals with relatively stable levels of pain in comparison to individuals with greater variability in their pain. Further, Perrine and Schroder (2005) found that agreement between momentary and retrospective reports of the quantity of alcohol consumed on a given day was influenced by an individual's variability in his drinking behavior during the preceding year, such that more variable drinkers were less accurate in their reports. Finally, Fredrickson (2000) found a similar pattern of results in retrospective reports of mood states. These findings suggest that individuals with increased variability or instability may be expected to demonstrate less accurate retrospective recall for dynamic processes such as mood which may frequently change over time.

Additionally, recall biases may influence individuals' reports when they asked about the frequency of an experience over a specific amount of time or when more general tendencies or traits are assessed. When recalling experiences over time, individuals may rely on a variety of personal theories and heuristic strategies. Depending on the demands of the situation, individuals may use their current state (e.g. current mood) as a reference point and then rely upon theories of stability or change to reconstruct previous states (Gorin & Stone, 2001; McFarland, Ross, & DeCourville, 1989), regardless of the actual frequency of such states. Individuals are likely to assume that an experience happens more frequently if specific details about the experience easily occur to them, a phenomenon referred to as the availability heuristic (Tversky & Kahneman, 1974). Individuals are more likely to rely on these general guidelines when asked to retrospect over long periods of time or when asked to report on their general tendencies or traits.

The literature suggests that the recall biases discussed here apply to both disordered and healthy populations alike. Although some recall biases have been specifically examined in patient populations (e.g. negative recall biases in depression; Gotlib, Krasnoperova, Yue, & Joormann, 2004), other research examining the correspondence between momentary and retrospective reports has not found differing rates of correspondence between patient and healthy samples (e.g. Raphael, Cloitre, & Dohrenward, 1991). Additionally, some studies have not found rates of correspondence to vary by individual differences like neuroticism, depression inventory scores, or measures of state anxiety and anger (e.g. Raselli & Broderick, 2007; Stone et al., 1998), while others have (e.g. Robinson & Clore, 2002).

On the one hand, the field of clinical psychology has made significant advances over the years by using retrospective questionnaires and self-reports. However, the effects of bias on the results generated by these reports and on the conclusions drawn by researchers and clinicians are significant (Gorin & Stone, 2001). The error introduced by these recall biases leads to weaker effects in aggregated data, making the detection of between-group effects or trends over time more difficult. This also results in the need for increased sample sizes (in order to detect these diluted effects) as well as for additional time and resources to collect data from larger samples.

The correspondence between actual and recalled affective experiences is quite important to the field of clinical assessment. In clinical settings, the error inherent in patients' retrospective self-

reports may muddy the diagnostic picture, impacting treatment planning, the effective “dose” of treatment, and long-term treatment outcomes. Finally, these sources of error and the resulting difficulty of reliably detecting diluted effects can contribute to conflicting findings in the research literature.

Ecological Momentary Assessment

Given the biases inherent in recall or retrospection, a clearer understanding of the discrepancies between patients' actual and recalled experiences is essential. *Ecological Momentary Assessment* (EMA; Stone & Shiffman, 1994) offers a way to address some of the limitations of retrospective self-reports. As mentioned above, EMA is defined by the use of assessments that are both *ecological* (conducted in an individual's natural environment) and *momentary* (concerning the individual's immediate or near-immediate experience; Shiffman et al., 2007). By collecting information about an individual's experience during or immediately following the experience, EMA reduces the influence of recall biases. Because individuals are asked to reflect over a much shorter period of time (e.g. a few minutes to a few hours), they are less likely to rely on heuristics about their “typical” moods or behaviors and are more able to accurately report their current experiences. Similarly, since EMA typically assesses individuals' current experiences (e.g. current mood), individuals' tendencies to be influenced by their current mood or most recent experiences (via the mood-congruent and recency biases) are no longer problematic. Instead, these are the exact experiences EMA seeks to assess. With EMA research, individuals do not bear the burden of mentally aggregating across experiences, and researchers are free to aggregate the data in any way necessary to best address the question at hand. This provides researchers with a more “proximal” assessment of an individual's general behavioral tendencies or personality traits than is provided by a self-report questionnaire measure of these same constructs. EMA data also allow for a close examination of within-individual variation over time. This circumvents the many problems mentioned above which are associated with asking individuals to report on their own experiences over time (e.g., variability over time).

Our study focuses on the correspondence between momentary and retrospective trait and experience reports in clinical outpatients, research and clinical implications of discrepancies between these reports from patients, and problems with retrospective reports of mood that may be uniquely problematic with respect to BPD. EMA measures of mood were compared with retrospective recall of mood experience and self-report measures of trait affective instability/intensity across two psychiatric outpatient samples: 1) individuals with borderline personality disorder (BPD) who met the affective instability criterion and 2) individuals with current major depression or dysthymia (MDD/DYS) who met neither the affective instability criterion nor overall diagnostic criteria for BPD. Prominent etiological models of borderline personality disorder propose that the disorder is primarily characterized by intense mood states that fluctuate frequently in response to environmental stimuli (Linehan, 1993). Therefore, the accurate assessment of mood states is essential to an understanding of BPD symptoms. Given the dynamic nature of mood for individuals with BPD and recall biases associated with variability and instability, reliance upon retrospective recall of mood experience or trait measures of mood for this population may be particularly problematic and prone to error.

We predicted that the agreement between EMA indices and trait measures or recalled experience of affective instability would be modest. Further, based on the association between increased variability in emotional or behavioral experience and less accurate retrospection (Fredrickson, 2000; Perrine & Schroder, 2005; Stone et al., 2005), we predicted that BPD patients (characterized by frequent mood variability) would show less accuracy in reporting mood instability in negative affect than MDD/DYS patients. Finally, we expected levels of agreement for positive affect to be similar across these two groups.

Method

Participants

This study targeted two groups of psychiatric outpatients. The BPD group ($n=58$) included psychiatric outpatients who met DSM-IV-TR diagnostic criteria for BPD, including the affective instability criterion. The MDD/DYS group ($n=42$), included psychiatric outpatients who met diagnostic criteria for current major depressive disorder (MDD) or for current dysthymia (DYS), and who did not meet criteria either for BPD in general or for the specific symptom of affective instability. Individuals with major depression or dysthymia were selected in order to compare the individuals with BPD to patients that share significant mood disturbance, but not mood instability. As a result, the level and type of impairment (i.e., mood disturbance) did not vary greatly between groups, as they would if individuals with BPD were compared to healthy controls or psychiatric patients without notable mood disturbance, thereby reducing the possibility of this third-variable explanation for any group differences found.

Potential participants were made aware of the study through flyers in the waiting rooms of four local psychiatric outpatient clinics, which service community and/or university populations. Interested participants completed a form giving permission for research staff to contact them about the study, to review their medical records, and to consult with their treatment providers about the patients' current diagnoses. At this stage, patients were deemed ineligible based on factors such as a lack of documented mood disturbance in their medical records, current psychosis, age under 18 or over 65, current severe substance abuse (e.g., currently in detoxification treatment), and documented histories of severe head trauma, mental retardation, or significant neurological dysfunction or impairment. If information from a patient's medical record (e.g. diagnosis, medications, treatment history) suggested eligibility, the patient was invited to a screening interview (which included both Axis I and Axis II diagnostic interviews) to establish diagnostic eligibility. Following the screening interview, patients were excluded based on a lack of diagnostic fit with either the BPD or MDD/DYS group, a past history of severe and sustained substance dependence (e.g., leading to neurological impairment), or any of the exclusion criteria previously mentioned. Of 137 individuals excluded from participation based on the results of the diagnostic interview, only 5 met criteria for either MDD or DYS and were excluded because they also met the affective instability criterion of the SIDP-IV. Based on the results of the interview, participants that were eligible for the study were then scheduled for an orientation session (described below). All potential participants were paid \$20 for completing the screening interview.

To establish diagnostic eligibility for the study, all participants completed two DSM-IV structured interviews, the SCID-I (First et al., 1995) and the SIDP-IV (Pfohl et al., 1994). Interviews were completed by master's level clinical psychology graduate students, who received extensive training in these diagnostic assessments and who achieved high levels of reliability with the primary investigator of this project prior to conducting interviews. Audio recordings of the SCID-I and SIDP-IV interviews from a randomly selected sample of twenty participants were reviewed and scored independently by an alternative interviewer who served as a reliability checker. Agreement was excellent for the presence or absence of affective instability ($\kappa = 1.0$), a diagnosis of MDD/DYS ($\kappa = 1.0$), a diagnosis of BPD ($\kappa = .90$), and the number of BPD symptoms present (intraclass correlation coefficient [ICC] = .95).

Table 1 presents demographic characteristics of the sample, for the whole sample and by group. The sample consisted mostly of single, Caucasian women. Individuals in the BPD and MDD/DYS groups were significantly different only in the percentage of those previously hospitalized for psychiatric reasons. Of the 88 participants for whom prescription medication information was available, all but 3 (96.5%) reported taking at least one type of psychotropic medication

at the beginning of the EMA assessment period, with antidepressants and anxiolytics being most common. As for current Axis I disorders, the most prevalent disorders (other than MDD/DYS) were generalized anxiety disorder (40%), social phobia (31%), and post-traumatic stress disorder (26%). The most common Axis II disorders (other than BPD) were avoidant (33%), obsessive-compulsive (18%) and antisocial (15%) personality disorders.

Procedures

Participants were issued an electronic diary (Palm Zire 31[®] handheld computer) to record their affects, experiences, and behaviors six times a day over a 28-day period. An audible beep from the electronic diary (ED) prompted participants to complete an assessment. Prompts were scheduled automatically by a software program which stratified the participant's personalized waking hours into six equal intervals, and then randomly selected one moment within each interval to deliver a prompt. Participants were required to respond to the prompt within fifteen minutes, with the aid of two reminder beeps during the fifteen minute window. Collected data were time-stamped to determine if participants responded to prompts in a timely manner and to more accurately scale affective variability or instability as a function of time as described in the measures below. Although a number of time frames are possible, we asked participants to respond to the EMA items (e.g. mood ratings), considering their experiences since the last prompt. This was done in order to more completely sample their experiences, instead of only sampling their experience at the moment they completed the survey. Participants reported on their experiences over time periods ranging from approximately 15 minutes to a few hours.

At the initial orientation session, research staff explained use of the ED devices, including how to respond to prompts. Participants practiced completing an assessment on the ED, to ensure their ability to use the device and their understanding of the content of the items and the rating scales. Research staff emphasized timely responding to prompts and informed participants that compliance would be checked at weekly data download visits. The EDs were programmed so that the first random prompt occurred the evening of the training, and staff called participants to make sure that all had gone well. After receiving training in the use of the ED but prior to completing their first ED assessment, participants completed a battery of paper-and-pencil assessments.

At each of the four weekly data download visits, research staff met one-on-one with participants and downloaded the past week's recorded data from the ED device. These visits also provided an opportunity to check ED recording compliance and reinforce good recording practices. Uploaded ED recordings from the past week were scanned for missed prompts or unusually short or long responses to individual assessment items. If necessary, staff queried participants about the circumstances surrounding any problems identified in the data. At each weekly data download session, participants were paid \$45. At the final data download, participants again completed a battery of paper-and-pencil assessments.

Compliance with the EMA surveys was very high. Participants were asked to complete five of the six prompts per day (approximately 85%), on average, in order to receive each weekly payment. Across the entire sample, participants completed an average of 91.82% of their scheduled number of assessments. BPD participants completed 90.71% of their scheduled assessments, on average (range: 66.07%-100%). MDD participants completed 92.68% of their scheduled assessments, on average (range: 74.40%-100%). There was no significant group difference in these rates.

Analyses are based upon the second administration of the paper-and-pencil assessments. Because these assessments were administered immediately following the 28-day EMA data collection period, participants responded to the paper-and-pencil assessments after having monitored their moods and behaviors for 28 days using the EDs. This maximized the chance

for agreement between the retrospective self-report and EMA indices of affective instability because questionnaire responses were based, at least in part, on the EMA monitoring period.¹

Measures

Momentary assessment of mood—Mood descriptor items from the Positive and Negative Affect Schedule (PANAS-X; Watson & Clark, 1994) were presented to each participant on the ED during each momentary assessment. For each item, respondents consider each mood descriptor and then provide a rating that reflects the extent to which s/he felt this way (1=very slightly or not at all, 5=extremely) in the designated time period. A total of thirty-one PANAS-X items were administered in each assessment, allowing the calculation of the following subscales: *negative affect* (ten items); *positive affect* (ten items); *hostility* (six items); *fear* (six items); and *sadness* (five items). The latter three mood subscales were used to characterize specific affects relevant to BPD and to the DSM-IV-TR definition of affective instability, which highlights fluctuation in affect between states of dysphoria, anger, and anxiety.

Questionnaire assessment of trait mood instability/intensity—As mentioned, participants completed a paper-and-pencil assessment battery immediately after the 28-day EMA period. These questionnaires included several trait self-report assessments of affective experiences, including the *Personality Assessment Inventory – Borderline Features Scale*, the *Affective Lability Scale*, and the *Affect Intensity Measure*.

Participants completed the *Personality Assessment Inventory – Borderline Features Scale* (PAI-BOR; Morey, 1991), a 24-item self-report measure that assesses the major features of BPD using four subscales: affective instability, identity problems, negative relationships, and self-harm or impulsive behaviors. For the current analyses, the Affective Instability subscale (PAI-BOR-AI) was used to assess convergent validity with the EMA indices of affective instability. The Affective Instability subscale assesses mood intensity and variability, using six items (e.g., “My mood can shift quite suddenly”) rated on a 4-point scale (0-3; false, slightly true, mainly true, very true). For the complete sample, the average score and internal consistency were $M = 11.55$ ($SD = 3.34$) and $\alpha = .70$. Table 2 presents means and standard deviations, by group. As expected, individuals in the BPD group had significantly higher scores than individuals in the MDD/DYS group on the PAI-BOR-AI subscale ($t(98) = 4.60, p < .001$), indicating that the BPD individuals were more likely to endorse trait-like mood instability.

The *Affective Lability Scale* (ALS; Harvey, Greenberg, & Serper, 1989) is a 54-item questionnaire which assesses the changeability of one's moods. Items tap the subjective experiences, physiological perceptions, and behaviors related to affective lability (e.g., “I frequently shift from being able to control my temper very well to not being able to control it very well at all”). Participants rate the 54 items on a four-point scale based on the self-descriptiveness of each statement. Although the measure can be decomposed into six subscales (elation, anger, anxiety, euthymia-depression oscillation, elation-depression oscillation, and anxiety-depression oscillation), the full-scale score (average item score across all items) was used in the current analyses ($M = 1.62, SD = .48, \alpha = .95$). As indicated in Table 2, individuals in the BPD group reported significantly greater affective lability than individuals in the MDD group ($t(98) = 3.22, p < .01$).

¹Parallel analyses based on questionnaire scores collected before the EMA assessment are available from the corresponding author; similar results were found.

The *Affect Intensity Measure* (AIM; Larsen, Diener, & Emmons, 1986) is a 40-item questionnaire that assesses the strength with which an individual typically experiences emotions. Items reflect the subjective experiences of emotions and also tap physiological responses, cognitive performance, and interpersonal relations as they relate to affective experience (e.g., “The sight of someone who is hurt badly affects me strongly”). Items are rated on a six-point scale based on the self-descriptiveness of the item. The positive (AIM-pos; $M = 3.35$, $SD = .69$, $\alpha = .91$; 24 items) and negative (AIM-neg; $M = 4.22$, $SD = .68$, $\alpha = .76$; 13 items) subscales were used for the current analyses. There were no significant differences between the two groups on the positive subscale ($t(98) = 1.57$, $p = .12$) or negative subscale ($t(98) = .20$, $p = .84$), indicating that the intensity of positive and negative moods were comparable across groups (see Table 2).

Retrospective recall of mood changes—Finally, after completing the 28-day EMA period of data collection and the questionnaires, participants were asked to recall any occurrences of extreme mood change experienced over the preceding 28 days. Patients did not know they would be asked to do this until completion of the study. To aid their memory, we provided a calendar, marked with any relevant holidays and all of their study-related appointments over the 28-day period. For any mood shifts recalled, participants were asked to note the date of the mood shift, the time of the mood shift (e.g. 10:00am or “morning”), the valence of the mood shift (i.e. negative or positive), and any circumstances believed to have “triggered” the mood shift. Participants were permitted to report multiple mood shifts per day. This task was intended to simulate the informal assessment methods and recall processes typically characterizing clinical settings, in which patients are asked to recall how many mood shifts they experienced over a given period of time.

Although participants, on average, completed the EMA assessments for 28 days, some variability existed in the EMA data collection period due to the schedule of the last data download session or missed days of responding. To control for the number of days in each participant's EMA data collection period (and therefore the number of days each participant could have reported mood shifts on the retrospective calendar), the proportion of the number of mood shifts reported on the calendar to the number of days in the EMA period was calculated.

Overview of EMA Affective Instability Indices

Several features of the ED data are important to note: (1) each participant may have a different total number of usable assessments; (2) assessment times within a day were randomly selected; and (3) these times vary randomly across people. The data for each participant are therefore unbalanced in terms of the number of observations and time intervals between observations, requiring the adjustments we describe below.

Temporal instability: Mean square successive differences—A useful index of instability is the *mean square successive difference* (MSSD) because this index takes into account both variability and temporal dependency in mood scores (Jahng, Wood, & Trull, 2008). It is calculated by averaging across squared differences between successive assessment observations, in this case PANAS-X composite scales of mood. In addition to capturing variability and temporal dependency, MSSD is also an attractive index because it is robust to systematic time trends in time series data and does not require that the data be detrended (Jahng et al., 2008). *Within-day or Short-Term MSSD* (ST-MSSD) is calculated by averaging the squared successive differences within each day and then averaging those scores across all days (in this study, approximately 28 days):

$$ST - MSSD = \frac{1}{N - J} \sum_{j=1}^J \sum_{i=1}^{n_j-1} (x_{(i+1)j} - x_{ij})^2$$

where x_{ij} is the momentary mood composite score measured at i th occasion within j th day, n_j is the number of momentary assessments for j th day, J is the number of days, and N is the total number of assessments for that participant. To correct for a skewed distribution of MSSD scores, all MSSD indices were log transformed prior to analysis. Table 3 presents the means and standard deviations of the raw MSSD scores.

Temporal instability: probability of acute change—In addition to using MSSD as an index of general mood instability, we also focused on the frequency of large, extreme changes in the participants' affect scores. We defined an “acute” change in affect as a successive increase that equaled or exceeded the value for the 90th percentile of the total distribution of change for all participants in the study on a particular PANAS-X composite scale. In this way, we could compute how many acute changes occurred for each participant. Like MSSD, PAC is robust to time-related trends in the data and does not require these trends to be removed prior to calculation (Jahng et al., 2008). As with ST-MSSD, we examined *within-day (short-term) acute changes* (ST-PAC):

$$ST - PAC = \frac{1}{N - J} \sum_{j=1}^J \sum_{i=1}^{n_j-1} AC_{(i+1)j}$$

where $AC_{(i+1)j} = 1$, if $x_{(i+1)j} - x_{ij} \geq c$ (and $AC_{(i+1)j} = 0$, otherwise), where c and c' are predetermined cut points (in this case, the 90th percentile for each PANAS-X affect scale). Unevenly spaced time intervals between assessments due to random time prompts within days may produce different amounts of expected successive difference across different time intervals when mood scores are serially correlated. Therefore, successive differences were adjusted for different time intervals, as introduced in Jahng et al. (2008), to obtain valid measures of ST-MSSD and ST-PAC. To correct for a skewed distribution of PAC scores, all PAC indices were arcsine transformed prior to analysis. For the means and standard deviations of the raw PAC scores, see Table 3.

Accuracy of retrospective report of extreme mood changes—To assess the accuracy of retrospective report of extreme mood changes, we compared participants' *recalled days* of extreme mood changes with *experienced days* of extreme mood changes. Recalled days with extreme mood increases were defined as the days on which one or more moments of extreme mood increases were reported on the mood calendar. Experienced days with extreme mood increases were defined as days on which one or more moment-to-moment acute changes in mood were found in the real-time assessments of mood states from the electronic diary. As mentioned above, an acute change was defined as the successive change greater than the 90th percentile of changes across all participants for that mood score. In the present case, these represented a raw score change of .5022 for negative affect, .7507 for positive affect, .5028 for hostility, .5055 for fear, and .7450 for sadness scores, respectively, on a scale of 0 to 4. Table 4 presents the percentage of experienced days (i.e. days with EMA acute changes) and recalled days (i.e. days with recalled extreme mood changes) over the course of 28 days, by group.

Results

Agreement between trait questionnaire and momentary assessments of mood instability was first assessed using bivariate correlations between paper-and-pencil measures of affective instability or intensity and instability indices of moods from momentary reports. As indicated in Table 2, the difference in group means was significant for the PAI-BOR-AI and ALS scores. As for the inter-correlations among questionnaire scores, within the BPD group, PAI-BOR-AI scores were significantly correlated with ALS scores ($r = .44, p < .05$) and AIM-neg scores ($r = .28, p < .05$). No other scores were significantly correlated with each other. Within the MDD/DYS group, PAI-BOR-AI scores were significantly correlated with ALS scores ($r = .41, p < .05$) and with AIM-neg scores ($r = .49, p < .05$). In addition, ALS and AIM-neg scores were significantly related ($r = .40, p < .05$), as were ALS scores and AIM-pos scores ($r = .32, p < .05$) as well as AIM-neg and AIM-pos scores ($r = .39, p < .05$).

Table 5 presents the bivariate correlations between trait questionnaire scores and EMA indices of negative affect, hostility, fear, sadness, and positive affect. A Bonferroni correction was applied within each block of correlations (e.g., all mean level correlations), to control for the multiple comparisons between variables. For the entire sample (see Table 5), PAI-BOR-AI scores were significantly related to EMA negative affect and hostility scores, including those for momentary mean level, instability (MSSD) and acute changes (PAC). The PAI-BOR-AI was also significantly related to instability scores and acute changes of fear. The ALS showed significant associations with instability (MSSD) of negative affect and hostility and with acute changes (PAC) in hostility and fear. AIM-neg scores were not significantly associated with EMA mood indices. Finally, AIM-pos scores were significantly associated with mean level, MSSD, and PAC scores for positive affect.

Table 6 presents these same correlations split by diagnostic group. Within the BPD group, PAI-BOR-AI scores were moderately associated with negative affect, hostility, and fear scores (mean level, MSSD, and PAC indices). The PAI-BOR-AI was also related to the momentary mean level of sadness for BPD individuals. For individuals with MDD/DYS, PAI-BOR-AI scores were significantly related to the instability (MSSD) and acute change (PAC) scores for hostility. ALS scores were only modestly associated with instability and acute change scores for fear in BPD individuals and not associated with any EMA index for MDD/DYS individuals. AIM-neg scores were unrelated to the EMA indices in either diagnostic group. Finally, AIM-pos scores were moderately associated with all three indices of positive affect for individuals in the BPD group. Statistical tests of differences between group correlations revealed that the respective BPD and MDD/DYS correlations were significantly different only in the case of the associations between the PAI-BOR-AI and EMA mean level negative affect ($z = 2.17, p < .05$), fear ($z = 2.43, p < .05$), and sadness ($z = 2.69, p < .01$). In each instance, the correlation for the BPD group was stronger.

The accuracy of participants' retrospective recall of mood shifts was indexed by the agreement between retrospective mood calendar data (i.e., recalled) and the momentary assessment of acute mood shifts recorded with EMA methodology (i.e., experienced). Agreement between recalled and experienced reports of mood shifts was evaluated using the following indices: 1) *hit rate* (percentage of days in which participants correctly recalled either the presence or absence of a significant mood shift); 2) *kappa* (agreement between recalled and experienced report of mood shifts, after controlling for chance agreement); 3) *sensitivity* (percentage of accurately recalled days of mood shifts); 4) *specificity* (percentage of accurately recalled days without mood shifts); 5) *positive predictive value* (PPV; proportion of days with recalled mood shifts that were correct according to the momentary assessments); 6) *negative predictive value* (NPV; percentage of days without recalled mood shifts that were correct according to the momentary assessments). Agreement was assessed for the full EMA data collection period

(i.e., an average of 28 days) as well as for the last seven days of each participant's EMA data collection. The calculations of these measures of agreement were based on a two-by-two table across individuals, whereas the determination of an experienced extreme mood shift, or acute change, was based on the cut point of the 90th percentile of changes across all patients.²

As seen in Table 7, agreement between retrospective and momentary assessments of significant mood shifts was generally poor based on these agreement indices. Although specificity was quite high (range = .86 to .95) for both positive and negative mood shifts, sensitivity was poor (range = .05 to .17). This pattern was found for the full EMA monitoring period as well as for the last seven days timeframe. Even after considering the base rates of significant positive and negative mood shifts, the PPV estimates were still modest (range = .27 to .40) and the NPV estimates were moderate (range = .65 to .73). Further, kappa estimates were poor for both positive and negative mood shifts, across the 28-day and 7-day periods (range = .001 to .03). In general, agreement was comparably poor across the 28-day and 7-day periods. In addition, there were no notable differences between the BPD and MDD groups on these agreement indices.

Finally, agreement was assessed using intraclass correlation coefficients (ICCs; Shrout & Fleiss, 1979), which indexed the agreement between two measures: the retrospective and momentary assessments of significant mood shifts. For calculation of ICC, the number of days with extreme mood changes recalled on the mood calendar was compared with the number of days of extreme mood changes experienced based on EMA data, using the ST-PAC index described above. ICCs were calculated for the absolute agreement between the retrospective and momentary reports using a two-way random effects model. No ICC value was significantly different from zero, indicating poor agreement between the retrospective and momentary reports of significant mood shifts.

Across all participants, the absolute agreement ICC was -.02 (95% CI = -.12 to .11) for negative mood shifts over the study period, .01 (95% CI = -.06 to .09) for positive mood shifts over the study period, -.04 (95% CI = -.20 to .14) for negative mood shifts over the last week, and -.01 (95% CI = -.13 to .13) for positive mood shifts over the last week. Among BPD participants, the ICC was .001 (95% CI = -.13 to .17) for negative mood shifts over the study period, .05 (95% CI = -.06 to .20) for positive mood shifts over the study period, .03 (95% CI = -.12 to .21) for negative mood shifts over the last week, and -.05 (95% CI = -.24 to .17) for positive mood shifts over the last week. For MDD/DYS participants, the ICC was -.06 (95% CI = -.23 to .16) for negative mood shifts over the study period, -.07 (95% CI = -.20 to .13) for positive mood shifts over the study period, -.07 (95% CI = -.27 to .18) for negative mood shifts over the last week, and -.01 (95% CI = -.29 to .29) for positive mood shifts over the last week.

Discussion

Overall, our results indicate low to modest agreement between either retrospective or trait reports of mood instability/intensity and momentary assessments of mood instability. Results from this study indicate that scores from questionnaire measures of trait affective instability show relatively modest associations with participants' experienced affective instability and average mood intensity, as derived from EMA reports of mood state in patients' natural environments. Although the PAI-BOR-AI scores showed the greatest correspondence to EMA mood indices, the associations were variable across types of negative affect and across group

²We repeated these analyses, using a cutoff value for each individual that maximized his or her agreement level between PAC scores and retrospective reports of mood shifts over the study period and over the last week. Results using these individualized cutoffs were largely consistent with those using the 90th percentile cutoff; poor agreement was observed between momentary and retrospective data. The corresponding average kappa values using the "best cutoff scores" for each individual were: .121 (28 day NA); .119 (7 day NA); .058 (28 day PA); and .072 (7 day PA).

and were of a moderate strength, at best. PAI-BOR scores were more highly related to momentary affect for BPD individuals and to EMA indices of hostility (for both participant groups). ALS scores showed inconsistent and modest correlations with EMA mood indices, primarily for BPD individuals, and AIM-pos scores were moderately correlated with momentary indices of positive affect, but only for BPD individuals. Finally, retrospective reports of extreme mood changes, whether over the previous month or the immediately preceding week, were largely unrelated to EMA indices of acute affect changes across both patient groups. Results indicated that both groups are more accurate in recalling days without extreme mood shifts, but have great difficulty accurately recalling days with such shifts.

Contrary to our initial hypotheses, BPD individuals did not appear to be poorer retrospective reporters of their affective experiences. As mentioned above, previous research has found that individuals with greater variability or instability in their experiences were less accurate on retrospective reports of the same experiences (Fredrickson, 2000; Perrine & Schroder, 2005; Stone et al., 2005). This literature suggests poorer recall accuracy for those whose experience of a particular mood, behavior, or event is less stable or less predictable. Therefore, BPD individuals might be expected to be less accurate when retrospectively reporting on their mood, given the presence of affective instability (in contrast to MDD/DYS individuals without that symptom). However, we did not find this in our study.

Although not dramatic, results suggest that those with BPD might be slightly better at assessing their affective instability, at least by questionnaire. The few group differences that were found (see Table 6) involved the affective instability scale of the PAI-BOR. Interestingly, BPD and MDD/DYS participants did not differ on their momentary mean levels of negative affect, fear, and sadness. However, the PAI-BOR-AI, a measure specifically designed to assess instability in negative affect in individuals with BPD, was the retrospective self-report trait measure best able to distinguish between the affective experiences of these two patient groups. Further, the correlations between the PAI-BOR-AI and the EMA mean levels of negative affect, fear, and sadness for BPD individuals were of a moderate strength. These findings are interesting in two ways. First, this scale was moderately associated with the real-time affective experience of BPD individuals but not in those from another diagnostic group with comparable mean levels of these negative affect states. Second, this scale related to these real-time affective experiences more strongly than other self-report measures designed to broadly assess mood intensity, or mean level (i.e. the AIM-neg scale), and other measures specifically designed to assess affective instability (i.e. the ALS). In all, this provides evidence for the validity of the PAI-BOR-AI scale.

Despite the higher correlations among BPD patients, it must be remembered that the relationships between questionnaire trait scores and EMA indices were still modest at best. Ultimately, it will be important to evaluate the external correlates of EMA indices and questionnaire or interview-based scores for affective instability. In this way, the incremental validity of one approach over the other can be determined. This process will inform the clinical assessment of affective instability and perhaps point to the strengths and limitations of each approach. It does seem clear, however, that asking patients to remember specific details of large mood changes (e.g., timing, intensity) may be asking too much. In the present study, there was almost no relationship between retrospective reports of extreme mood changes and indices of these changes derived from EMA data. However, like other proponents of momentary assessment (e.g. Robinson & Clore, 2002), we do not mean to suggest that momentary assessments of affective instability are always superior to patient self-report. Instead, we view these sources of information as complementary and acknowledge the unique and valuable data inherent in a patient's view of his or her affective tendencies.

On the surface, results from this study appear to contradict the findings from a previous study of recall bias in a BPD sample, conducted by Ebner-Priemer and colleagues (2006). They compared the average intensity of momentary mood ratings collected over a 24-hour period with participants' retrospective account of their mood intensity during the same period and found that BPD individuals were more likely to retrospectively over-estimate the intensity of negative moods and under-estimate the intensity of positive moods, compared with healthy controls. Their study, however, examined a different aspect of emotion recall in BPD individuals than did our study. We focused on agreement (or lack thereof) between *mood fluctuations* reflected in momentary assessments and individuals' retrospective accounts of affective instability, in addition to examining mean levels of mood. In contrast, Ebner-Priemer and colleagues focused exclusively on average *mood intensity level* (versus variability or instability). Ebner-Priemer and colleagues' findings are supported in part, however, by our results indicating the poor agreement between participants' momentary indices of extreme mood changes and retrospective recall of extreme mood shifts over the immediately preceding 7-day period. Taken together, the results of these two studies suggest that BPD individuals may be inaccurate in reporting specific negative mood shifts during recent time periods and may be only moderately accurate in reporting their general tendency to experience frequent mood fluctuations using questionnaires (depending upon which questionnaire is used).

Several limitations related to the sample should be noted. First, the results of this study may not generalize beyond the demographic and diagnostic groups utilized in this study. Our sample consisted of individuals with BPD (and affective instability) and currently depressed individuals (without comorbid BPD or affective instability), most of whom were female and recruited from outpatient treatment settings. Therefore, these results may not generalize to patients with differing diagnostic profiles (e.g. individuals who meet diagnostic criteria for BPD but not affective instability) or from other treatment settings (e.g. inpatient settings). Second, as noted in Table 1, 61.4% of the BPD individuals in this study were also experiencing a current depressive episode. Although such comorbidity is typical of BPD individuals and supports the generalizability of our findings, this diagnostic overlap between the patient groups on the surface seems to complicate an examination of group differences in affect. Therefore, it is important to reiterate that the present study focused on differences in momentary and retrospective reports of affective *instability*, a symptom the two groups did not share. Finally, our findings do not provide insight into how these patient groups might compare to healthy individuals without mood dysregulation. As previously mentioned, studies have examined recall biases in healthy samples and have found discrepancies between momentary and retrospective reports in these individuals as well. However, our study focused on patient samples in order to assess the degree of discrepancy between momentary and retrospective reports that may be influencing research and clinical practice with these diagnostic groups.

Additional limitations concern the types of assessments used in this study. First, only certain trait measures of affective instability were examined in this study. Other measures might yield different correlational patterns with momentary assessments of mood fluctuations. For instance, one possible explanation of the poor correspondence between the momentary and retrospective measures could be due to poor correspondence between the PANAS (used in the EMA surveys) and the AIM, ALS, and PAI-BOR-AI. From this perspective, one could argue that the mood calendar recall task was a better test of the correspondence between momentary and retrospective assessment, as it transcends some measure effects. However, the results indicated poor agreement with momentary assessments when using that retrospective task as well. Second, the questionnaire measures of trait affective instability did not explicitly require participants to retrospectively recall specific instances of certain mood states, as did the momentary assessments and the mood recall task. However, individuals are asked to estimate the general frequency on such measures (e.g. "I frequently shift from being able to control my anger well to not being able to control it very well at all," from the ALS), perhaps activating

many of the same recall biases present in more specific retrospection tasks. Further, researchers and clinicians alike use these general trait measures as indicators of the frequency with which patients may have such affective experiences (e.g. a patient with a higher PAI-BOR-AI score may be expected to experience more frequent affective variability). Thus, the degree of agreement between these trait measures and the actual experiences assessed via EMA is important. Third, only certain types of mood were assessed in the present study. Although general negative affect, hostility, fear, and sadness are mood states with particular importance to the definition of affective instability in BPD, emerging research also suggests the importance of emotions such as guilt and shame for BPD individuals (Rusch et al., 2007). Future research should examine issues of recall accuracy with respect to these additional mood states in BPD individuals. Finally, our findings are limited by our EMA sampling method, by which we asked participants to report on their affective experience since their previous EMA survey (versus reporting on their immediate affective experience). Because participants were reporting on periods ranging from a few minutes to a few hours, recall biases could have had some effect on their mood ratings in each EMA survey.

The findings in this study point towards several future applications of EMA in clinical research. There are a variety of constructs like mood that are multifaceted and vary over time. Unlike retrospective and trait reports, EMA offers the opportunity to sample many occasions over a relative brief period of time in order to capture a detailed picture of the dynamic nature of the construct. This differs from trait measures, for example, which instead provide more global and less detailed information. An EMA approach could be especially useful as a component of a larger longitudinal study or treatment outcome study, in which EMA measures of mood are repeatedly administered intermittently over the course of many months or years. Such applications of this assessment approach would provide a greater depth to our understanding of psychopathology.

These findings have several implications for clinical practice as well. Clinicians might develop a more accurate picture of a client's mood instability by administering momentary assessments of mood during the first several weeks of treatment, or even while clients are waiting for admission to treatment. By directly assessing the client's momentary experiences, one circumvents reliance on retrospective or trait measures and informal assessments in which the information recalled by the client may be less reliable and incomplete. This more complete picture of a client's mood would greatly aid the clinician in diagnosis and treatment planning. Although traditional trait measures are still useful and important in assessing individuals' perceptions of their own symptoms and functioning, EMA measures of mood offer a level of detail that can be essential in developing and delivering a variety of clinical interventions. For example, clinicians who deliver Dialectical Behavior Therapy (Linehan, 1993) to clients with BPD use a paper-and-pencil diary, called a diary card, to monitor moods and relevant behaviors daily. Viewed as an essential part of the treatment, diary cards provide day-to-day ratings of mood intensity and instability and are used to guide the content of each individual therapy session. EMA, especially as it is conducted with electronic diaries which can prompt patients and time-stamp responses, offers a more detailed alternative approach to gathering these important data.

Clearly, EMA data collection comes with its unique challenges which may at times make it a less attractive assessment tool, compared to retrospective methods. EMA requires more time and, when using EDs to collect data, more equipment and more training (both for the patient and for the researcher/clinician). However, as this methodology becomes increasingly prevalent in psychopathology research, many "how-to" guides have been published (e.g. Fahrenberg et al., 2007; Piasecki, Hufford, Solhan, & Trull, 2007; Shiffman, Stone, Hufford, 2008), and the costs of technologies used for EMA continue to decrease. Based on the discrepancies between momentary and retrospective reports of patients' experiences, the

benefits of increased utilization of EMA methodology seem to outweigh the additional resources and time required to use this innovative and valuable form of assessment.

Acknowledgments

Supported in part by National Institute of Mental Health (MH069472).

References

- Brewin CR, Andrews B, Gotlib IH. Psychopathology and early experience: A reappraisal of retrospective reports. *Psychological Bulletin* 1993;113:82–98. [PubMed: 8426875]
- Cohen G, Java R. Memory for medical history: Accuracy of recall. *Applied Cognitive Psychology* 1995;9:273–288.
- Ebner-Priemer UW, Kuo J, Shaw Welch S, Thielgen T, Witte S, Bohus M, Linehan MM. A valence-dependent group-specific recall bias of retrospective self-reports: A study of borderline personality disorder in everyday life. *Journal of Nervous and Mental Disease* 2006;194:774–779. [PubMed: 17041290]
- Fahrenberg J, Myrtek M, Pawlik K, Perrez M. Ambulatory assessment – Monitoring behavior in daily life settings. *European Journal of Psychological Assessment* 2007;23:206–213.
- Fredrickson BL. Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion* 2000;14:577–606.
- Friedman WJ. Memory for the time of past events. *Psychological Bulletin* 1993;113:44–66.
- Gorin, AA.; Stone, AA. Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In: Baum, A.; Revenson, TA.; Singer, JE., editors. *Handbook of Health Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. p. 405-413.
- Gotlib IA, Krasnoperova E, Yue DN, Joormann J. Attentional biases for negative interpersonal stimuli in clinical depression. *Journal of Abnormal Psychology* 2004;113:127–135.
- Harvey PD, Greenberg BR, Serper MR. The Affective Lability Scales: Development, reliability, and validity. *Journal of Clinical Psychology* 1989;45:786–793. [PubMed: 2808736]
- Hufford, MR. Special methodological challenges and opportunities in ecological momentary assessment. In: Stone, AA.; Shiffman, S.; Atienza, A.; Nebeling, L., editors. *The science of real-time data capture: Self-reports in health research*. New York: Oxford University Press; 2007.
- Huttenlocher J, Hedges LV, Bradburn NM. Reports of elapsed time: Bounding and rounding processes in estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1990;16:196–213.
- Jahng S, Wood PK, Trull TJ. Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods* 2008;13:354–375. [PubMed: 19071999]
- Kahneman D, Fredrickson BL, Schreiber CA, Redelmeier DA. When more pain is preferred to less: Adding a better end. *Psychological Science* 1993;4:401–405.
- Kihlstrom, JF.; Eich, E.; Sandbrand, D.; Tobias, BA. Emotion and memory: Implications for self-report. In: Stone, AA.; Turkkan, JS.; Bachrach, CA.; Jobe, JB.; Kurtzman, HS.; Cain, VS., editors. *Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000. p. 81-99.
- Larsen RJ, Diener E, Emmons R. Affective intensity and reactions to daily life events. *Journal of Personality and Social Psychology* 1986;51:803–814.
- Larsen SF, Thompson CP. Reconstructive memory in the dating of personal and public news events. *Memory and Cognition* 1995;23:780–790.
- Linehan, MM. *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY: Guilford Press; 1993.
- Mayer JD, McCormick LJ, Strong SE. Mood-congruent memory and natural mood: New evidence. *Personality and Social Psychology Bulletin* 1995;21:736–746.
- McFarland C, Ross M, DeCourville N. Women's theories of menstruation and biases in recall of menstrual symptoms. *Journal of Personality and Social Psychology* 1989;57:522–531. [PubMed: 2778636]

- Mineka, S.; Nugent, K. Mood-congruent memory biases in anxiety and depression. In: Schacter, DL., editor. *Memory Distortions: How Minds, Brains, and Societies Reconstruct the Past*. Cambridge, MA: Harvard University Press; 1995. p. 173-193.
- Morey, LC. *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources; 1991.
- Perrine MWB, Schroder KEE. How many drinks did you have on September 11, 2001? *Journal of Studies on Alcohol* 2005;66:536–544. [PubMed: 16240561]
- Piasecki TM, Hufford MR, Solhan M, Trull TJ. Assessing clients in their natural environments with electronic diaries: Rationale, benefits, limitations, and barriers. *Psychological Assessment* 2007;19:25–43. [PubMed: 17371121]
- Raphael KG, Cloitre M, Dohrenwend BP. Problems of recall and misclassification with checklist methods of measuring stressful life events. *Health Psychology* 1991;10:62–74. [PubMed: 2026132]
- Raselli C, Broderick JE. The association of depression and neuroticism with pain reports: A comparison of momentary and recalled pain assessment. *Journal of Psychosomatic Research* 2007;62:313–320. [PubMed: 17324682]
- Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 1996;66:3–8. [PubMed: 8857625]
- Robinson MD, Clore GL. Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin* 2002;128:934–960. [PubMed: 12405138]
- Rusch N, Lieb K, Gottler I, Hermann C, Schramm E, Richter H, Jacob GA, Corrigan PW, Bohus M. Shame and implicit self-concept in women with borderline personality disorder. *American Journal of Psychiatry* 2007;164:500–508. [PubMed: 17329476]
- Shiffman S, Hufford M, Paty JA, Gnys M, Kassel JD. Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting and Clinical Psychology* 1997;65:292–300. [PubMed: 9086693]
- Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology* 2008;4:1–32.
- Southwick SM, Morgan CA, Nicolaou AL, Charney DS. Consistency of memory for combat-related traumatic events in veterans of Operation Desert Storm. *American Journal of Psychiatry* 1997;154:173–177. [PubMed: 9016264]
- Stone AA, Broderick JE. Real-time data collection for pain: Appraisal and current status. *Pain Medicine* 2007;8:S85–S93. [PubMed: 17877531]
- Stone AA, Hedges SM, Neale JM, Satin MS. Prospective and cross-sectional mood reports offer no evidence of a “Blue Monday” phenomenon. *Journal of Personality and Social Psychology* 1985;49:129–134.
- Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality and Social Psychology Bulletin* 2005;31:1340–1346. [PubMed: 16143666]
- Stone AA, Schwartz JE, Neale JM, Shiffman S, Marco CA, Hickcox M, Paty J, Porter LS, Cruise LJ. A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology* 1998;74:1670–1680. [PubMed: 9654765]
- Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine* 1994;16:199–202.
- Thompson CP, Skowronski JJ, Lee DJ. Telescoping in dating naturally occurring events. *Memory and Cognition* 1988;16:461–468.
- Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science* 1974;185:1123–1131.
- Watson, D.; Clark, LA. *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. 1994. Retrieved from University of Iowa, Department of Psychology Web site: <http://www.psychology.uiowa.edu/Faculty/Watson/Watson.html>

Table 1
Demographic features of participants

	BPD (<i>n</i> = 58)	MDD (<i>n</i> = 42)	Total (<i>N</i> =100)
<u>Age</u>			
<i>M</i>	32.3	35.4	33.6
<i>SD</i>	11.2	11.7	11.5
% female	93.1	81.0	88.0
<u>Race (%)</u>			
Caucasian	86.2	92.9	89.0
African-American	5.2	4.8	5.0
Hispanic	1.7	2.4	2.0
Other	6.8	0.0	4.0
<u>Marital Status (%)</u>			
Single, Divorced, or Separated	67.3	64.3	66.0
Married	19.0	21.4	20.0
Currently employed (%)	53.4	45.2	51.5
Income less than \$25,000 (%)	72.4	69.0	71.0
Children (none) (%)	51.1	65.8	53.5
Previous psychiatric hospitalization (%)	62.1	38.1	52.0
<u>Medications^a</u>			
Antidepressants	70.9	82.5	76.1
Anxiolytics	45.9	42.5	44.3
Mood Stabilizers	35.4	12.5	25.0
Anti-psychotics	31.2	15.0	23.8
<u>Current Axis I Diagnoses^b</u>			
Major Depressive Disorder	61.4	81.0	56.0
Dysthymia	17.5	45.2	29.0
Bipolar I Disorder	13.8	0.0	8.0
Bipolar II Disorder	1.7	0.0	1.0
Generalized Anxiety Disorder	47.4	31.0	40.0
Social Phobia	33.3	28.6	31.0
Post-traumatic Stress Disorder	31.6	19.0	26.0
Alcohol Abuse/Dependence	10.5	2.4	7.0
Cannabis Abuse/Dependence	8.8	2.4	6.0
<u>Current Axis II Diagnoses^b</u>			
Avoidant PD	25.1	31.0	33.0
Obsessive-Compulsive PD	22.8	11.9	18.0
Antisocial PD	24.6	2.4	15.0

^aBased on information from 88 of 100 participants

^bBased on information from 99 of 100 participants

Table 2
Descriptive statistics of trait measures of affective instability

	BPD (<i>n</i> =58)		MDD/DYS (<i>n</i> =42)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
PAI-BOR-AI*	12.74	2.79	9.90	3.36
ALS*	1.74	.44	1.44	.49
AIM-pos	3.44	.72	3.22	.65
AIM-neg	4.20	.71	4.23	.65

Notes:

* Difference in group means significant at $p < .01$. BPD = Borderline personality disorder, MDD/DYS = Current major depressive disorder or dysthymia. PAI-BOR-AI = Personality Assessment Inventory Borderline Features scale-Affective Instability subscale. ALS = Affective Liability Scales. AIM-pos = Affect Intensity Measure-Positive subscale. AIM-neg = Affect Intensity Measure-Negative subscale.

Table 3
Descriptive statistics of EMA indices of mood and affective instability

	BPD (<i>n</i> =58)		MDD (<i>n</i> =42)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Mean Level				
Negative Affect	1.71	.67	1.55	.37
*Hostility	1.62	.66	1.35	.30
Fear	1.70	.67	1.59	.54
Sadness	1.81	.83	1.90	.70
Positive Affect	2.12	.60	2.16	.61
MSSD				
*Negative Affect	.31	.31	.22	.20
*Hostility	.44	.44	.27	.31
*Fear	.33	.31	.25	.26
Sadness	.46	.39	.44	.37
Positive Affect	.54	.40	.41	.34
Probability of Acute Change				
Negative Affect	.11	.07	.09	.07
*Hostility	.11	.07	.08	.07
Fear	.11	.08	.10	.07
Sadness	.10	.07	.10	.07
Positive Affect	.10	.07	.10	.06

Notes: Raw scores are reported for MSSD and Acute Change values (not the log and arcsine transformed values, respectively, used in the correlational analyses).

* Difference in group means significant at $p < .05$.

Table 4
Descriptive statistics of percentage of experienced and recalled days of extreme mood changes

	BPD (<i>n</i> =58)		MDD/DYS (<i>n</i> =42)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
NA experienced days	35.15	.22	31.59	.22
NA recalled days	12.26	.14	11.56	.09
PA experienced days	34.74	.20	35.76	.20
PA recalled days	5.40	.07	6.42	.12

Notes: There were no significant group differences.

NA=extreme change in negative affect

PA=extreme change in positive affect

Table 5
Bivariate correlations between trait measure scores and within-day EMA mood indices

	PAI-BOR-AI	ALS	AIM-neg	AIM-pos
Mean Level				
Negative Affect	.33*	.25	.22	.01
Hostility	.43*	.28	.16	.00
Fear	.23	.23	.22	.02
Sadness	.17	.08	.17	-.17
Positive Affect	-.10	-.02	-.01	.36*
MSSD				
Negative Affect	.44*	.32*	.27	.14
Hostility	.51*	.32*	.23	.15
Fear	.37*	.29	.28	.06
Sadness	.24	.12	.19	.00
Positive Affect	.09	.03	-.03	.37*
Acute Change				
Negative Affect	.38*	.29	.25	.08
Hostility	.50*	.34*	.28	.11
Fear	.33*	.30*	.24	.08
Sadness	.21	.11	.08	-.05
Positive Affect	.05	.04	.00	.38*

Notes:

* Significant after Bonferroni correction ($p < .0025$).

Table 6
Bivariate correlations between trait measure scores and within-day EMA mood indices by group

	PAL-BOR-AI		ALS		AIM-neg		AIM-pos	
	BPD	MDD	BPD	MDD	BPD	MDD	BPD	MDD
Mean Level								
Negative Affect	.45*	.03	.30	.07	.20	.33	.05	-.20
Hostility	.42*	.35	.25	.23	.11	.39	-.04	-.06
Fear	.42*	-.06	.30	.10	.17	.31	.09	-.15
Sadness	.44*	-.09	.28	-.15	.21	.10	-.18	-.12
Positive Affect	-.06	-.14	-.03	.01	-.06	.08	.50*	.17
MSSD								
Negative Affect	.43*	.37	.33	.21	.18	.42	.15	.08
Hostility	.41*	.47*	.32	.19	.13	.40	.15	.05
Fear	.45*	.22	.37*	.13	.21	.39	.07	-.02
Sadness	.30	.21	.27	-.06	.20	.19	.06	-.10
Positive Affect	-.10	.23	-.10	.12	-.17	.22	.46*	.20
Acute Change								
Negative Affect	.40*	.34	.32	.21	.21	.32	.12	-.02
Hostility	.44*	.49*	.36	.21	.23	.38	.10	.04
Fear	.42*	.22	.37*	.19	.19	.33	.13	-.03
Sadness	.31	.14	.24	-.03	.09	.06	.03	-.16
Positive Affect	-.08	.21	-.02	.10	-.14	.24	.43*	.28

Notes:

* Correlation significant after Bonferroni correction ($p < .0025$). Shaded pairs of correlations denote significant group differences.

Table 7
Agreement indices between retrospective mood calendar reports and EMA indices of extreme negative and extreme positive mood shifts

	Hit Rate	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Kappa
28-day NA	.64	.13	.90	.40	.67	.035
7-day NA	.66	.17	.86	.31	.73	.028
28-day PA	.63	.05	.95	.36	.65	.003
7-day PA	.70	.07	.94	.27	.73	.001

Note. NA=extreme change in negative affect; PA=extreme change in positive affect