

## Comparing Retrospective Reports to Real-Time/Real-Place Mobile Assessments in Individuals With Schizophrenia and a Nonclinical Comparison Group

Dror Ben-Zeev<sup>1,2,\*</sup>, Gregory J. McHugo<sup>1</sup>, Haiyi Xie<sup>1</sup>, Katy Dobbins<sup>2</sup>, and Michael A. Young<sup>3</sup>

<sup>1</sup>Department of Psychiatry, Dartmouth Psychiatric Research Center, Dartmouth Medical School, Hanover, NH; <sup>2</sup>Thresholds-Dartmouth Research Center, Chicago, IL; <sup>3</sup>College of Psychology, Illinois Institute of Technology, Chicago, IL

\*To whom correspondence should be addressed; Thresholds-Dartmouth Research Center, 4101 N. Ravenswood Avenue, Chicago, IL 60613, USA; tel: 773-573-3113, fax: 773-537-3535, e-mail: dror.ben-zeev@dartmouth.edu.

**Retrospective reports are often used as the primary source of information for important diagnostic decisions, treatment, and clinical research. Whether such reports accurately represent individuals' past experiences in the context of a serious mental illness such as schizophrenia is unclear. In the current study, 24 individuals with schizophrenia and 26 nonclinical participants used a mobile device to complete multiple real-time/real-place assessments daily, over 7 consecutive days. At the end of the week, participants were also asked to provide a retrospective report summarizing the same period. Comparison of the data captured by the 2 methods showed that participants from both groups retrospectively overestimated the intensity of negative and positive daily experiences. In the clinical group, overestimations for affect were greater than for psychotic symptoms, which were relatively comparable to their retrospective reports. In both samples, retrospective reports were more closely associated with the week's average than the most intense or most recent ratings captured with a mobile device. Multilevel modeling revealed that much of the variability in weekly assessments was not explained by between-person differences and could not be captured by a single retrospective estimate. Based on the findings of this study, clinicians and researchers should be aware that while retrospective summary reports of the severity of certain symptoms compare relatively well with average momentary ratings, they are limited in their ability to capture variability in one's affective or psychotic experiences over time.**

**Key words:** experience sampling method (ESM)/ecological momentary assessment (EMA)/mobile interventions/mHealth/delusions/hallucinations

### Introduction

In order to gain insight into how clinical populations think, feel, and function, clinicians and researchers often

ask individuals to provide retrospective summary reports of their past or “typical” experiences. These can take the form of responses to questions used in well-validated and widely used diagnostic or clinical instruments and interviews (eg, Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, [DSM-IV],<sup>1</sup> Positive and Negative Syndrome Scale [PANSS],<sup>2</sup> Beck Depression Inventory [BDI]<sup>3</sup>) or as answers to less structured information gathering conducted by a practitioner who asks patients “how were your symptoms this week?” Regardless of the specific administration format, retrospective reports are the most common method of data collection in psychiatric settings and a cornerstone of clinical research and practice.

Whether such reports accurately represent an individual's past experiences or effectively summarize his or her current clinical status in the context of a serious mental illness such as schizophrenia is unclear. Retrospective summaries are typically collected in laboratory or clinic settings, not the real-world environments in which symptoms occur or daily functioning actually takes place. Consequently, their ecological validity, that is, to what extent data collected in lab conditions approximate those in real-world settings and situations, is limited.<sup>4,5</sup> In addition, retrospective reports often entail asking individuals to recollect (eg, “how many times did you leave the house last week?”) or aggregate (“on average, how often do you hear voices during the day?”) their experiences—tasks that are quite susceptible to various memory problems and recall inaccuracies.<sup>6–8</sup> These obstacles to measurement may be even more prominent when compounded by the psychopathological symptoms<sup>9,10</sup> and myriad cognitive impairments<sup>11,12</sup> that are often seen in people with schizophrenia.

In order to evaluate the accuracy and representativeness of retrospective reports, a number of studies have employed an elegant dual-methods approach in which

participants are asked to provide information on the same target construct using 2 separate strategies. One is the traditional lab or clinic-based summary in which participants describe their status on relevant items over a given period of time (eg, 2 weeks, “current”). The other strategy uses the experience sampling method<sup>13,14</sup> (ESM) or ecological momentary assessment,<sup>15</sup> in which individuals are signaled repeatedly by a mobile device to complete ratings of their immediate status on similar items, in the context of their own environment. By comparing the summary measures to a criterion based on ratings that were generated in real-time/real-place, researchers are able to determine just how representative retrospective reports really are of real-world experiences (for review, see Shiffman et al<sup>16</sup>).

Discrepancies between ESM and retrospective reports have already been identified in a number of clinical populations, including eating disorders,<sup>17</sup> panic disorder,<sup>18</sup> obsessive compulsive disorder,<sup>19</sup> borderline personality disorder,<sup>20</sup> and depression.<sup>21,22</sup> Although different studies used different methodologies and statistical approaches to examine this question, overall, results generally suggest that summary reports “overestimate” the frequency and/or intensity of symptoms and affect experienced. It is unclear whether individuals take into account all their experiences when attempting to provide a retrospective estimate or whether they are systemically biased by select events that are especially salient to them at the time of recall. A number of studies have shown that retrospective reports may be disproportionately influenced by the most intense (peak) or recent (end) moments,<sup>23–25</sup> suggesting that these select events may have heuristic value or special meaning when one attempts to summarize their longitudinal experience.<sup>26</sup> Consequently, retrospective reports may actually be most representative of “peak” or “end” experiences rather than the most typical or average of all experiences over a given period. Finally, with the growing understanding that symptoms and functioning in schizophrenia are much more dynamic than previously believed,<sup>27–29</sup> it is important to evaluate whether traditional retrospective reports can effectively capture clinical phenomena that may vary not only between but also within individuals over time.

Gaining a better understanding of how retrospective reports correspond to actual day-to-day experiences reported by individuals with serious mental illness is important, as these estimates are typically used to inform diagnostic decisions, case conceptualization, and treatment. This need is amplified in clinical practice with individuals with schizophrenia because patients’ retrospective reporting of symptoms and mood disturbances often impact subsequent administration and dosing of antipsychotic medications that may entail serious unwanted side effects.<sup>30–32</sup>

To explore the relationship between retrospective reports and real-time/real-place ratings, we conducted a dual-methods study in which 24 individuals with schizophrenia recorded daily hallucinations, delusional thoughts, and various affective experiences using a mobile device over a 1-week period. At the end of the week, participants were asked to provide an overall retrospective summary report of the same 7 days. Their retrospection patterns were examined against those of 26 nonclinical individuals who completed similar ratings of affective experiences. Specifically, we wanted to examine (a) whether momentary experiences are “overestimated” in the retrospective reports, (b) whether retrospective reports are more representative of the highest intensity or most recent experiences (ie, peak and end) compared with an average of all experiences recorded during the week, and (c) how well retrospective reports account for the overall variability of momentary experiences in both groups.

## Methods

### *Participants*

This study was approved independently by the institutional review boards of both a university and a large community psychiatric rehabilitation organization. Twenty-four individuals with schizophrenia (17 male) and 26 nonclinical individuals (9 male) provided ESM and retrospective reports for the study (see table 1 for full demographic information). Participants in the clinical sample were recruited from ads placed on bulletin boards at “Thresholds Psychiatric Rehabilitation Centers” in Chicago, in the “National Alliance on Mental Illness” newsletter, and on public bulletin boards. Participants in the comparison group were recruited for a separate ESM study that utilized a similar approach and measures, through messages on community internet websites (eg, “Craigslist,” “Chicago Reader”) and ads placed on public bulletin boards in Chicago. Participants in the clinical group met criteria for a diagnosis of schizophrenia or schizoaffective disorder. They had a mean age of 44.87 years (SD = 9.27) and were 71% male, 71% never married, and 79% African-American. The mean length of illness reported by these participants was 16.95 years (SD = 8.50; range = 4–36 y). At baseline, the clinical sample experienced on average mild psychotic symptom severity (Scales for the Assessment of Positive Symptoms [SAPS] total, M = 23.58, SD = 18.86, range = 0–74 and Scales for the Assessment of Negative Symptoms [SANS], M = 10.54, SD = 10.95, range = 0–40) and moderate depression symptom severity (BDI-2, M = 21.20, SD = 12.54, range = 0–51). Individuals in the comparison sample had no current DSM-IV diagnosis of any kind or lifetime diagnosis of psychotic or bipolar disorder, had a mean age of 34.72 (SD = 10.36), and were 65% female,

**Table1.** Descriptive Characteristics of the Study Sample

Variable	Clinical Group (N = 24)	Comparison Group (N = 26)
Age, <i>M</i> (SD)	44.87 (9.27)	34.23 (10.45)
Sex, <i>N</i> (%)		
Female	7 (29.2)	17 (65.4)
Male	17 (70.8)	9 (34.6)
Ethnicity, <i>N</i> (%)		
African-American	19 (79.2)	9 (34.6)
Caucasian	3 (12.5)	13 (50.0)
Hispanic/Latino	1 (4.2)	2 (7.7)
Other	1 (4.2)	2 (7.7)
Education, <i>N</i> (%)		
Some high school	6 (25.0)	0 (0)
High school	8 (33.3)	2 (7.7)
Some college	9 (37.5)	8 (30.8)
College degree	1 (23.9)	9 (34.6)
Graduate degree	0 (0)	7 (26.9)
Marital Status, <i>N</i> (%)		
Married	1 (4.2)	4 (15.4)
Widowed	0 (0)	1 (3.8)
Divorced	5 (20.8)	5 (19.2)
Single	17 (70.8)	16 (61.5)
Missing data	1 (4.2)	0 (0)

*Note:* The clinical group was older (independent  $t = 3.81$ ,  $P < .001$ ), comprised significantly more males ( $\chi^2 = 6.56$ ,  $P < .05$ ) and African-Americans ( $\chi^2 = 10.42$ ,  $P < .05$ ) than the comparison group. The comparison group was significantly more educated ( $\chi^2 = 23.01$ ,  $P < .01$ ).

61% never married, and 50% Caucasian. At baseline, the nonclinical group reported very minimal depression (BDI-2,  $M = 2.38$ ,  $SD = 1.90$ , range = 0–6). All participants in both groups completed the study successfully and returned the personal digital assistant (PDA) units intact at the end of the week.

### Procedures

After providing written informed consent, all participants were administered a structured diagnostic interview (Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition<sup>1</sup> [SCID-I/P]) in order to verify diagnosis of schizophrenia or schizoaffective disorder and to assess additional lifetime and current mental disorders. Both groups then completed a battery of self-report and interview measures that included demographic information and a measure of symptoms of depression (BDI-2<sup>3</sup>). Participants from the clinical group were also rated with the SAPS and SANS<sup>33,34</sup> of schizophrenia.

Following assessment, a 30-minute training session was provided concerning the meaning of all computerized ESM questions and response choices as well as training on how to operate and charge a PDA programmed to run the experience sampling program (ESP)<sup>35</sup> version 4.0, a free open-source ESM software package. Participants

were given a Tungsten E2 PDA that runs on a PALM OS Garnet (V.54) operating system and asked to carry it with them over a 1-week period. The PDAs were programmed to generate a prompting auditory signal multiple times every day in intervals that were spread out equally  $\pm 10$  minutes (determined randomly) between 9 AM and 10 PM. The clinical group was intended to receive 6 prompts per day and the comparison group 8 due to the slightly different original study protocols. The actual number of prompts participants received varied within each group due to various mechanical failures or empty PDA batteries as a result of individuals forgetting to charge the device overnight.

Participants were instructed to complete the ESM questionnaire that appeared on the PDA screen. ESM assessments took approximately 2–3 minutes to complete, and all responses were automatically stored in the PDA memory. The ESP settings chosen for the study allowed participants to complete questionnaires no later than 5 minutes after being prompted. All responses were time/date stamped electronically. Two practice ESM questionnaires were completed in the laboratory under the supervision of the research staff in order to resolve any difficulties or address questions. Each PDA had a label with the lab's contact number to call in case of technical difficulties or with any questions during the week. All messages received were responded to no later than 24 hours by research staff. Participants were also called by a research assistant 48 hours after commencing and 24 hours before their scheduled second lab visit, to troubleshoot any problems and to remind them of their return appointment.

During the second visit, all participants were asked to provide a retrospective report summarizing the passing week "overall," before returning the PDA and receiving payment. Finally, all participants were debriefed and asked whether any major or unexpected life events took place during the week of data collection (eg, unexpected health concerns, moving to a new residence) that might have contributed to response patterns that were atypical for them. Subsequently, data from 24 hours were dropped for one participant because the PDA accidentally signaled him to complete questionnaires during the night. He reported contacting the lab the next day and resolving the problem after speaking with a research assistant who instructed him on how to modify the PDA settings over the phone.

### Measures

**Computerized ESM Measures.** Both groups completed affect reports in every ESM questionnaire. Participants rated how they felt on 8 adjectives that were selected from the Positive and Negative Affect Schedule.<sup>36</sup> The clinical sample was also asked to respond to 5 psychotic symptom items that were modeled after the Psychotic

**Table 2.** Computerized ESM Question Items and Response Options

Questions	Touch-Screen Response Options
Affect	
• “How ACTIVE do you feel right now?”	0-Slightly or not at all
• “How JITTERY do you feel right now?”	1-A little
• “How SCARED do you feel right now?”	2-Moderately
• “How HOSTILE do you feel right now?”	3-Quite a bit
• “How UPSET do you feel right now?”	4-Extremely
• “How EXCITED do you feel right now?”	
• “How GUILTY do you feel right now?”	
• “How INTERESTED do you feel right now?”	
Symptoms <sup>a</sup>	
• Paranoia: “Since the last questionnaire, how strongly did you believe someone was SPYING or PLOTTING against you?”	0-Not at all
• Mind Reading: “Since the last questionnaire, how strongly did you believe you could READ others people’s THOUGHTS or they could read yours?”	1-A little
• Reference: “Since the last questionnaire, how strongly did you believe others could communicate with you through the TV or RADIO?”	2-Moderately
• Grandiosity: “Since the last questionnaire, how strongly did you believe you had SPECIAL POWERS and could do things nobody else could do?”	3-Quite a bit
• Hallucinations: “Since the last questionnaire, to what extent did you hear VOICES that others could not hear?”	4-Extremely

Note: ESM, experience sampling method.

<sup>a</sup>Administered only to the clinical group.

Symptom Rating Scales<sup>37</sup> and ESM questions that have been shown to have convergent validity with the PANSS positive symptom subscale.<sup>38</sup> The topic of focus for each item was presented in capital letters, so that after a few administrations participants could identify the question and respond quickly (see table 2). Participants could not skip over questions and needed to enter their answers by using the PDA stylus (pointer) to tap on 1 of 5 response boxes before being able to continue to the next question on the following screen. When a questionnaire was completed, the participant was notified, and the PDA would automatically enter a “hibernation mode” until the next signal.

*Retrospective Reports.* At the end of 1 week of ESM data collection, participants returned to the lab and were asked to “reflect upon the passing week” before

completing a lab-based summary questionnaire. Question items were similar to those that appeared in the ESM measures but were modified to ask for a retrospective summary rather than an immediate report (eg, “overall, how SCARED did you feel?,” “overall, how strongly did you believe someone was SPYING or PLOTTING against you?”).

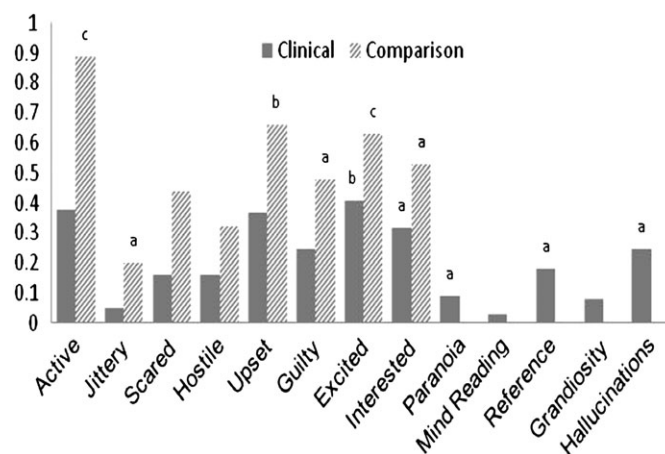
### Overview of Analyses

We calculated 3 indices of the week’s ESM data for each question item for each participant: mean ESM, peak ESM (the highest rating recorded during the week), and end ESM (the last rating provided before individuals completed the retrospective report). All available data that were collected during the specified study hours were included in the analyses. To examine whether

symptoms and affective experiences were overestimated in the retrospective reports, we compared mean ESM indices with retrospective reports for every item in each group separately, using paired samples *t* test. Effect sizes for the differences between ESM and retrospective reports were calculated as Cohen's *d*. Pearson correlation coefficients were used to examine whether retrospective reports were more representative of peak ESM or end ESM ratings rather than mean ESM, ie, the average of all symptom and affective experiences. Given the multi-level nature of the ESM data (ie, numerous observations nested within individuals), we used HLM 6.08<sup>39</sup> to generate parameters that would help to determine how much of the variability in the ESM data is explained by the retrospective ratings. We used a staged approach and modeled each affect and symptom item separately for each group. At the first stage, we estimated an unconditional mean model with no predictors to determine the intraclass correlation (ICC). ICC was calculated as the ratio of between-person variance (ie, tau of intercepts or TAU1) and the total variance in the model (TAU1 + within-person or residual variance). TAU1 in this case represents the average ESM item rating. Higher ICC scores would, therefore, indicate that more of the variance in the ESM data is due to between-person differences. At the second stage, we included the retrospective rating for each item as a level 2 predictor and determined the between-person variance that was unexplained (TAU2). BPv for this model was calculated as a ratio  $[(TAU1 - TAU2) : TAU1]$ , which represents the proportion of between-person variance that was explained by the retrospective rating. At the final stage, we calculated  $ICC \times BPv$  to create a new estimate representing the proportion of total variance that was explained by the retrospective rating (TOTv). The higher the TOTv, the more the retrospective ratings account for the total variance in the ESM data.

## Results

The schizophrenia group received an average of 41.4 prompting signals during the week, for which the ESM questionnaire completion rate was 97.7%. The majority of participants in this sample (83%) endorsed experiencing some symptoms of psychosis at some stage during the week; moderate or high severity symptoms (ie, score of "3" or "4" on momentary ESM items) were reported on 16% of paranoia, 11% of mind reading, 15% of reference, 15% of grandiosity, and 13% of hallucination ratings captured during the week. The comparison group received an average of 56.1 signals, for which the completion rate was 88.9%. On average, both groups endorsed higher positive affect (ie, feeling "active," "excited," and "interested") than negative affect (ie, feeling "jittery," "scared," "hostile," "upset," and "guilty") during the week, as reported by both ESM and retrospective measures.

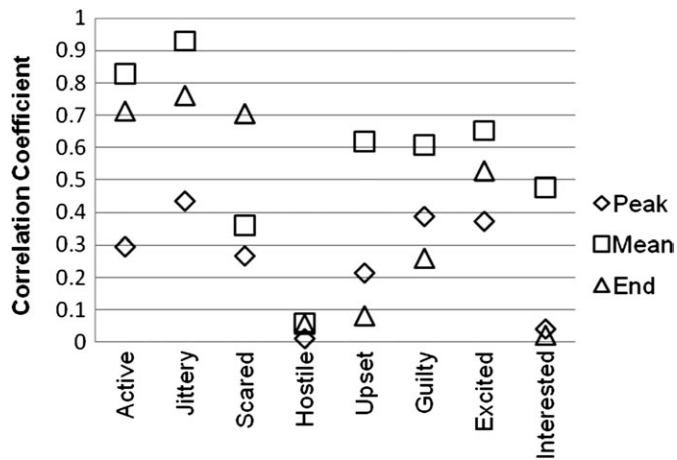


**Fig. 1.** Effect sizes (Cohen's *d*) for the difference between retrospective and mean experience sampling method (ESM) ratings for affect and symptoms of psychosis. Retrospective and mean ESM ratings significantly different within group (paired *t* test): <sup>a</sup>*P* < .05, <sup>b</sup>*P* < .01, <sup>c</sup>*P* < .001.

The first question that we addressed was whether the retrospective ratings overestimated the information captured in momentary ratings during the past week. That is, when people are asked to rate the level of a symptom or affective experience during the past week "overall," do they report a value that is greater than the average of their momentary ratings. The retrospective ratings provided by both the clinical and comparison groups exceeded mean ESM measures for all question items. However, not all these differences were statistically significant (figure 1). In both groups, the mean retrospective ratings were always significantly lower than the mean peak ESM ratings (all paired *t* test *P* values  $\leq .01$ ), suggesting that the overestimation was never out of the range of participants' actual recorded experiences during the week. To further explore the nature of the overestimation, we tested the mean retrospective ratings against the mean of the 75th percentile of the ESM ratings. Remarkably, in both groups, the 75th percentile values were not statistically different from the retrospective ratings in all but one item ("upset" for the comparison group).

The magnitude of overestimation of affect items was always substantially larger in the comparison group (Cohen's *d* range 0.20 to 0.89) than in the clinical sample (Cohen's *d* range 0.05 to 0.41), with effects for some items almost 3 times greater in the nonclinical group. In the clinical group, overestimation was typically greater in the retrospective summaries of affect items than in retrospective summaries of psychotic symptoms. Effect sizes for psychotic symptoms were all relatively small, with the largest shown for recall of hallucinations (Cohen's *d* = 0.25).

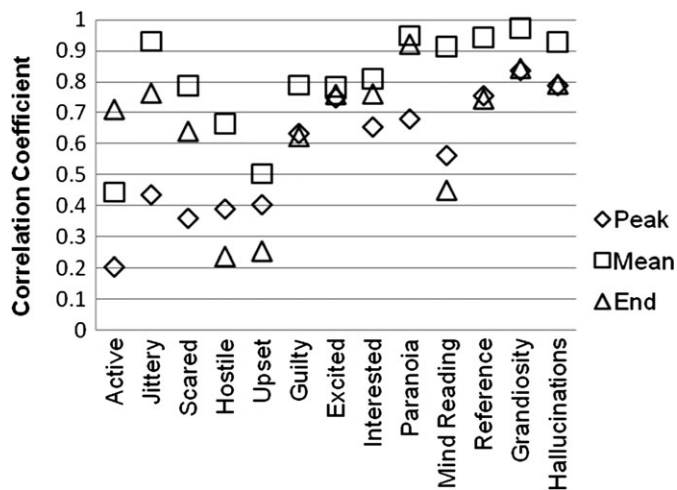
Next, we examined whether retrospective reports are more representative of the highest intensity weekly experience (peak ESM), the most recent experience (end



**Fig. 2.** Correlations of retrospective ratings with peak experience sampling method (ESM), mean ESM, and end ESM for each affect item in the comparison group; demonstrates the generally higher association with mean ESM.

ESM), or the average of all experiences (mean ESM). Pearson correlation coefficients revealed that in both groups, retrospective reports were most closely related to mean ESM (mean  $r = .77$ ; range .36–.97) and not peak ESM (mean  $r = .54$ ; range .01–.83) or end ESM ratings (mean  $r = .59$ ; range .02–.83) in all but one item (figures 2 and 3). Associations between mean ESM and retrospective reports were all significant (all  $P$ 's  $\leq .05$ ), except for “hostile” in the comparison group.

Finally, we examined the degree to which retrospective reports accounted for the variability in daily reports of experience. Parameters generated from individual item multilevel models showed that the extent to which the total variance in the week's ESM data could be explained by between-person differences (average ESM ratings)



**Fig. 3.** Correlations of retrospective ratings with peak experience sampling method (ESM), mean ESM, and end ESM for each affect/symptom item in the clinical group; demonstrates the generally higher association with mean ESM.

**Table 3.** Proportions of ESM Data Variance Explained by Average and Retrospective Ratings

	Clinical Group			Comparison Group		
	ICC	BPv	TOTv	ICC	BPv	TOTv
<b>Affect</b>						
Active	0.34	0.17	0.06	0.39	0.69	0.27
Jittery	0.29	0.42	0.12	0.40	0.88	0.36
Scared	0.60	0.61	0.37	0.14	0.10	0.01
Hostile	0.18	0.45	0.08	0.18	0.00	0.00
Upset	0.35	0.24	0.08	0.18	0.39	0.07
Excited	0.63	0.60	0.38	0.46	0.41	0.19
Guilty	0.53	0.61	0.32	0.13	0.39	0.05
Interested	0.53	0.66	0.35	0.37	0.20	0.07
<b>Symptoms</b>						
Paranoia	0.68	0.89	0.61	—	—	—
Mind reading	0.64	0.86	0.56	—	—	—
Reference	0.76	0.88	0.68	—	—	—
Grandiosity	0.91	0.94	0.86	—	—	—
Hallucinations	0.73	0.86	0.63	—	—	—

*Note:* ESM, experience sampling method; ICC, intraclass correlation. ICC is the proportion of total variance explained by between-person variance in the multilevel model, BPv is the proportion of between-person variance accounted for by the retrospective report, TOTv is the proportion of total variance in ESM data accounted for by the retrospective report.

was quite variable across items (the average item ICCs across groups was .44). Most relevant to our question, the retrospective reports never accounted fully for the between-person differences in average item ratings (the average item BPv across groups was 0.53) and did a relatively poor job at capturing the total variance in the ESM data (the average item TOTv across groups was 0.29) (table 3).

## Discussion

The current study used a dual-methods approach to compare retrospective reports and real-time/real-place ESM data capturing a week in the lives of individuals with schizophrenia and a nonclinical comparison group. The relationship between the information provided by both methods was explored in a number of ways. A pattern of some retrospective exaggeration was true for both negatively (eg, paranoia, hallucinations, feeling guilty, or upset) and positively (eg, feeling excited, interested, or active) valenced experiences. Overestimations were lowest in the clinical group's ratings of psychotic symptoms. Interestingly, in both groups, retrospective reports were closest to ESM ratings at the 75th percentile of intensity. Thus, while participants did not appear to be retrospectively averaging or counting events in an absolutely precise sense, they did not grossly misrepresent events either, providing ratings that never exceeded the intensities of ratings in their actual experiential range.

In fact, the clinical groups' retrospective estimates of psychotic symptoms compared relatively well the average of multiple momentary ratings of these symptoms. Retrospective reports were found to be more closely associated with the average of the week's ESM ratings rather than peak or end moments. In addition, multilevel modeling revealed that much of the variability in weekly assessments was not explained by between-person differences and could not be captured by a single retrospective estimate.

The retrospective "overestimation" patterns found in our study are quite consistent with findings from previous research with other clinical and nonclinical populations.<sup>17–22,40</sup> One possible explanation for the slight discrepancy between momentary and retrospective reports is that, unlike in short-term neurocognitive recall tasks, it is not feasible for individuals to retain and recall a week's worth of ongoing experiences. When asked to provide retrospective summaries of wide time-frames, especially when summarizing affective experiences, individuals don't just rely on episodic memory and likely also draw from existing semantic knowledge that may include personal beliefs or expectations one holds about their emotional life or "typical" experiential week.<sup>41</sup>

Alternatively, the differences we found between retrospective and ESM reports may be methodological in nature and linked to our use of a categorical scale. A retrospective symptom rating of "0" or "not at all" may not feel valid for an individual who had even one symptomatic moment during the week (even if arithmetically, this is the most accurate description of the majority of the time) because it does not capture the symptom exacerbations at all. The individual may opt to choose the next viable response, ie, "1" or "a little." Consequently, the retrospective estimate is slightly "bumped up" and is higher than the average of ESM reports, not truly capturing the full scope of the data. This may also explain why the retrospective estimates were closest to ESM ratings that constituted the 75th percentile for item intensity rather than mean ESM ratings.

It is difficult to determine what may have contributed to the different magnitudes of overestimation across groups as the samples were not well-matched demographically. Thus, differences in retrospection may be influenced by age, gender, or ethnicity variables rather than clinical diagnosis. For example, younger individuals may tend to retrospectively view their life as more emotionally intense than originally experienced, resulting in greater overestimations of affect in the nonclinical group. The comparison group was significantly more educated than the clinical sample, which is often the case in research with individuals with serious mental illness. Our chief concern regarding education level in this study was the participants' ability to understand the task and complete computerized ESM measures correctly during the week. In this regard, all participants in the study

showed no difficulties in completing questionnaires and demonstrated their competency levels to our research staff before commencing data collection, and study compliance rates were very high for both groups.

Our finding that much of the variance in individuals' ESM data was not fully explained by between-person differences alone is very consistent with the growing body of empirical evidence suggesting that both symptoms and affect in schizophrenia are not nearly as static as is often believed, varying not only between but also within individuals over relatively short periods of time.<sup>27–29,42</sup> Thus, clinicians and researchers should be aware that although some of the retrospective summary reports compared relatively well to average momentary ratings (of psychotic symptoms in particular), they are limited; when asking participants or patients to provide a single summary of their ongoing experiences over time, we are requesting completion of a task that may neglect variation in one's experience. Summary estimates are clearly not perfectly accurate in their ability to capture one's average, most intense, most recent, or the variability in one's emotional or symptomatic experiences over a given period and should not be viewed as such. We are not suggesting traditional retrospective reports do not have important clinical and research utility. In fact, they may be an excellent measure of one's current illness-related beliefs and personal conceptualization of past experiences that may play a pivotal role in their future decision-making and behavior.<sup>43,44</sup> For current clinical practices, retrospective reports of psychotic symptoms may be quite suitable for assessing the general clinical well-being of patients. Looking forward, if we are interested in designing interventions that can be tailored and individualized to help people cope with personal challenges as they occur in the moment, they are not the most appropriate tool as they cannot capture dynamic processes. Contemporary mobile technologies are better suited to record the ebb-and-flow of symptoms and subjective experiences as they occur within individuals longitudinally and to help identify momentary risk or preventive factors that could perhaps be targeted with immediate mobile interventions.<sup>27,28,45,46</sup>

The current study had a number of limitations. The clinical and nonclinical samples were originally recruited for different ESM projects, and the ESM questionnaires and frequency of administration were similar but not identical. This may explain in part why the clinical group was more compliant in responding to momentary prompts or other differences in the results. We asked individuals in the clinical group to report psychotic items because the last prompt was given, summarizing a number of hours, which may be susceptible to some recall inaccuracy. We decided to ask for brief summaries as opposed to momentary "snapshots" since unlike affective states that may linger for longer periods of time (and consequently easier to record with momentary questions after they commenced), automatic

thoughts or brief sensory experiences may be extremely fleeting in nature and difficult to capture as they occur. Consequently, these items had a microretrospective component to them, which may have contributed to the relatively smaller differences between the week's retrospective and momentary ratings of symptoms. Finally, completing multiple ESM assessments has been shown to have some impact on naturally occurring phenomena,<sup>22</sup> and it is possible that increased individual focus and sensitization to symptoms, or conversely, habituation and desensitization, occurred in the current study. Taken together, the evidence from this and previous studies suggests that given the opportunity, individuals with schizophrenia can and will engage in the use of mobile technologies effectively.

## Funding

National Institutes of Mental Health (P20 MH085981-02 pilot funding to D.B.Z.).

## Acknowledgments

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

## References

1. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders-Patient Edition (SCID-I/P), Version 2.0*. New York, NY: New York State Psychiatric Institute, Biometrics Research; 1995.
2. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13:261–276.
3. Beck AT, Steer RA, Brown GK. *Manual for Beck Depression Inventory II (BDI-II)*. San Antonio, TX: Psychology Corporation; 1996.
4. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. *Psychol Assess.* 2009;21:457–462.
5. Hufford MR, Shiffman S, Paty J, Stone AA. Ecological momentary assessment: real-world, real-time measurement of subject experience. In: Fahrenberg J, Myrtek M, eds. *Progress in Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies*. Seattle, WA: Hogrefe & Huber; 2001:69–92.
6. Schwartz N, Sudman S. *Autobiographical Memory and the Validity of Retrospective Reports*. New York, NY: Springer Verlag; 1994.
7. Schwartz N. Retrospective and concurrent self-reports: the rationale for real-time data capture. In: Stone A, Shiffman SS, Atienza A, Nebeling L, eds. *The Science of Real-Time Data Capture: Self-Reports in Health Research*. New York, NY: Oxford University Press; 2007:11–26.
8. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. New York, NY: Cambridge University Press; 2000.
9. Baddeley AD, Thornton A, Chua SE, McKenna P. Schizophrenic delusions and the construction of autobiographical memory. In: Rubin DC, ed. *Remembering Our Past: Studies in Autobiographical Memory*. Cambridge, UK: Cambridge University Press; 1995:384–428.
10. Sass L, Parnas J. Self, consciousness, and schizophrenia. *Schizophr Bull.* 2003;29:427–444.
11. Elvevag B, Goldberg TE. Cognitive impairment in schizophrenia is the core of the disorder. *Crit Rev Neurobiol.* 2000;14:1–21.
12. Nuechterlein KH, Barch DM, Gold JM, Goldberg TE, Green MF, Heaton RK. Identification of separable cognitive factors in schizophrenia. *Schizophr Res.* 2004;72:29–39.
13. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling Method. *J Nerv Ment Dis.* 1987;175:526–536.
14. Delespaul P, deVries M. The daily life of ambulatory chronic mental patients. *J Nerv Ment Dis.* 1987;175:537–544.
15. Stone AA, Shiffman S. Ecological momentary assessment in behavioral medicine. *Ann Behav Med.* 1994;16:199–202.
16. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1–32.
17. Stein KF, Corte CM. Ecologic momentary assessment of eating-disordered behaviors. *Int J Eat Disord.* 2003;34:349–360.
18. de Beurs E, Lange A, Van Dyck R. Self-monitoring of panic attacks and retrospective estimates of panic: discordant findings. *Behav Res Ther.* 1992;30:411–413.
19. Gloster AT, Richard DC, Himle J, et al. Accuracy of retrospective memory and covariation estimation in patients with obsessive-compulsive disorder. *Behav Res Ther.* 2008;46:642–655.
20. Ebner-Priemer UW, Kuo J, Welch SS, et al. A valence-dependent group-specific recall bias of retrospective self-reports: a study of borderline personality disorder in everyday life. *J Nerv Ment Dis.* 2006;194:774–779.
21. Ben-Zeev D, Young M, Madsen J. Retrospective recall of affect in clinically depressed individuals and controls. *Cogn Emot.* 2009;23:1021–1040.
22. Ben-Zeev D, Young MA. Accuracy of hospitalized depressed patients' and healthy controls' retrospective symptom reports: an experience sampling study. *J Nerv Ment Dis.* 2010;198:280–285.
23. Fredrickson BL, Kahanman D. Duration neglect in retrospective evaluations of affective episodes. *J Pers Soc Psychol.* 1993;65:45–55.
24. Stone AA, Broderick JE, Kaell AT, Delespaul P, Porter LE. Does the peak-end phenomenon observed in laboratory pain studies apply to real world pain in rheumatoid arthritis? *J Pain.* 2000;1:212–217.
25. Redelmeier DA, Kahanman D. Patient's memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain.* 1996;116:3–8.
26. Fredrickson BL. Extracting meaning from past affective experiences: the importance of peaks, ends, and specific emotions. *Cogn Emot.* 2000;14:577–606.
27. Ben-Zeev D, Ellington K, Swendsen J, Granholm E. Examining a cognitive model of persecutory ideation in the daily life of people with schizophrenia: a computerized experience sampling study. *Schizophr Bull.* 2011;37:1248–1256.
28. Ben-Zeev D, Morris S, Swendsen J, Granholm E. Predicting the occurrence, conviction, distress, and disruption of different delusional experiences in the daily life of people with schizophrenia. [published online ahead of print January 19, 2011]. *Schizophr Bull.* doi:10.1093/schbul/sbq167.



29. Thewissen V, Bentall RP, Lecomte T, van Os J, Myin-Germeys I. Fluctuations in self-esteem and paranoia in the context of daily life. *J Abnorm Psychol.* 2008;117:143–153.
30. Buchanan RW, Kreyenbuhl J, Kelly DL, et al. The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements. *Schizophr Bull.* 2010;36:71–93.
31. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med.* 2005;353:1209–1223.
32. Haddad PM, Sharma SG. Adverse effects of atypical antipsychotics: differential risk and clinical implications. *CNS Drugs.* 2007;21:911–936.
33. Andreasen NC. *The Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City, IA: The University of Iowa; 1984.
34. Andreasen NC, Olsen S. Negative versus positive schizophrenia: definition and validation. *Arch Gen Psychiatry.* 1982;39:789–794.
35. Barrett DJ, Barrett LF. *The Experience Sampling Program (ESP); Version 4*. 2005. <http://www.experience-sampling.org/>. Accessed September 8, 2011.
36. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol.* 1988;54:1063–1070.
37. Haddock G, McCarron J, Tarrier N, Faragher EB. Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychol Med.* 1999;29:879–889.
38. Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. *Schizophr Bull.* 2008;34:507–514.
39. Raudenbush SW, Bryk AS, Cheong YF, Cogdon RT. *HLM 6.08 Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International, Inc; 2004.
40. Wirtz D, Kruger J, Scollon CN, Deiner E. What to do on spring break? The role of predicted, online, and remembered experience in future choice. *Psychol Sci.* 2003;14:520–524.
41. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychol Bull.* 2002;128:934–960.
42. Myin-Germeys I, Delespaul PAEG, deVries MW. Schizophrenia patients are more emotionally active than is assumed based on their behavior. *Schizophr Bull.* 2000;26:847–854.
43. Kahneman D, Fredrickson BL, Schreiber CA, Redelmeier DA. When more pain is preferred to less: adding a better end. *Psychol Sci.* 1993;6:401–405.
44. Redelmeier DA, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. *Pain.* 2003;104:187–194.
45. Granholm E, Ben-Zeev D, Link PC, Bradshaw K, Holden J. Mobile assessment and treatment for schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. *Schizophr Bull.* 2012;38:414–425.
46. Oorschot M, Lataster T, Wichers M, Myin-Germeys I. Mobile assessment in schizophrenia: a data-driven momentary approach. *Schizophr Bull.* 2012;38:405–413.