

Ecologic Momentary Assessment of Eating-Disordered Behaviors

Karen Farchaus Stein* and Colleen M. Corte

School of Nursing, The University of Michigan, Ann Arbor, Michigan

Accepted 10 February 2003

Abstract: Objective: *This pilot study evaluated the feasibility of ecologic momentary assessment (EMA) to measure five eating disorder behaviors and to examine reactivity of these behaviors to this measurement approach. The pattern of correspondence between EMA and the Eating Disorder Examination (EDE) was also examined. Methods:* Women with threshold or subthreshold anorexia and bulimia nervosa (N = 16) recorded their eating disorder behaviors for 4 weeks on a hand-held computer. Upon completion of the EMA procedure, the EDE was administered. **Results:** *Eighty-eight percent of the sample completed the EMA behavioral recording. No differences in behavioral frequency were found in the first and second halves of the EMA measurement period, suggesting that behaviors were not reactive to the methodology. Binging and excessive exercise were lower when measured by EMA compared with the EDE. Discussion:* These results suggest that EMA is a feasible and valid approach to the measurement of disordered eating behaviors. © 2003 by Wiley Periodicals, Inc. *Int J Eat Disord* 34: 349–360, 2003.

Key words: *ecologic momentary assessment; eating disorders; Eating Disorder Examination; excessive exercise*

INTRODUCTION

Anorexia nervosa (AN) and bulimia nervosa (BN) are serious health problems that affect nearly 5 million American adolescent and young adult women each year. They cause life-threatening physical complications and compromised states of well-being (Becker, Grinspoon, Klibanski, & Herzog, 1999). Not surprisingly, a major focus in eating disorder (ED) research has been the development and testing of treatment approaches. Most studies that have examined the effectiveness of various treatments in reducing ED symptoms have relied on a retrospective interview to measure behavioral change. Yet, the validity and reliability of this measurement approach has not been demonstrated

Presented at the 10th Annual International Conference on Eating Disorders, Boston, MA, April, 2002.

Colleen M. Corte is now at the Addiction Research Center, The University of Michigan, Ann Arbor, Michigan.

*Correspondence to: Karen Farchaus Stein, Ph.D., School of Nursing, the University of Michigan, 400 N. Ingalls, Ann Arbor, MI 48109-0482. E-mail: kfarchau@umich.edu

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/eat.10194

© 2003 by Wiley Periodicals, Inc.

fully. Results of studies in the area of event recall and autobiographic memory suggest that behavioral frequency estimates are the product of multiple cognitive processes that are highly susceptible to biases and distortions.

Ecologic momentary assessment (EMA; Stone & Shiffman, 1994) offers a promising alternative to retrospective measurement of ED behaviors. Using an event-triggered EMA approach, target behaviors are measured as they occur during the course of the person's daily activities. Participants are asked to record the occurrence of the target behavior on a hand-held computer immediately after it occurs. Although the EMA is a valid and reliable approach to the measurement of diverse health-related behaviors (for examples, see Litt, Cooney, & Morse, 1998; Shiffman et al., 1994), the feasibility and reliability of the approach as a method of measuring ED behaviors are still being determined (for results of two small pilot tests, see Smyth et al., 2000). The purpose of the current pilot study was to examine the feasibility of using the EMA to measure five ED behaviors including binge eating, self-induced vomiting, laxative use, diuretic use, and excessive exercise and to examine the reactivity of ED behaviors to this measurement approach. In addition, the pattern of correspondence between the EMA and the Eating Disorder Examination (EDE; Fairburn & Cooper, 1993), a commonly used retrospective interview method of measuring ED behaviors, was explored.

Retrospective Recall of ED Behaviors and Associated Cognitive Processes

Most studies have relied on retrospective measures to examine changes in the frequency and intensity of ED behaviors (Safer, Telch, & Agras, 2001). The most commonly used retrospective measure is the EDE, a standardized semistructured clinical interview that measures ED attitudes and behaviors over the previous 4 weeks. Respondents indicate the number of days they engaged in one or more episodes of bingeing, vomiting, laxative use, diuretic use, and excessive exercise, as well as the total number of episodes of the behavior over the last 4 weeks.

Although studies have provided evidence to support the discriminant validity (Wilson & Smith, 1989) and interrater reliability of the EDE (Cooper & Fairburn, 1987), evidence to support concurrent validity of the measure is minimal and the results have been mixed. Rosen, Vara, Wendt, and Leitenberg (1990) asked 120 college students (20 with BN) to record their ED behaviors in a daily diary for 7 days. They also completed the EDE at the end of the diary period. Correlations between the two measures were moderate for the number of objective binge episodes ($r = .46$). However, only 17% of the sample had an ED. Therefore, this correlation probably was inflated by the high percentage of subjects who reported no binge eating episodes. The correlation for the number of vomiting episodes was high ($r = .90$ for the total sample and $r = .81$ for the bulimic sample only). The correspondence between the measures on laxative use, diuretic use, and exercise frequency was not reported.

Studies using other measures of ED behaviors have also shown significant discrepancies between levels attained using retrospective and concurrent methods of measurement. Rossiter, Agras, Telch, and Bruce (1992) found that women with nonpurging-type BN reported approximately one-half the number of binge episodes using a calendar recall ($M = 5.7$ episodes) compared with the number of binges reported using a daily food diary ($M = 10.6$), but the number of binge days was approximately equivalent. Ortega, Waranch, Maldonado, and Hubbard (1987) found that eight women with BN reported a higher frequency of binge episodes on a weekly retrospective measure ($M = 6.88$) compared with diary recordings that were rated by objective coders for binge

episodes ($M = 2.33$). Grilo, Masheb, and Wilson (2001a, 2001b) compared the measurement of the frequency of objective and subjective binge episodes using a daily written recording sheet with the EDE-Q, a self-administered version of the EDE (Fairburn & Beglin, 1994) in a sample of adults with binge eating disorder (BED). The method used to trigger the daily recording was not described. There were no significant differences between the two measures for the mean number of objective binges. However, the mean frequency of subjective binges was higher on the daily diary compared with the EDE-Q. Finally, Steiger, Gauvin, Jabalpuwala, Seguin, and Stotl (1999) used event-triggered EMA to measure dietary control, the urge to binge, and bingeing behaviors in a sample of women with BN spectrum disorders. The purpose of this study was to examine the relationship between perceptions of daily social interactions and ED behaviors. The event designated to trigger recording was the completion of a social interaction of 10 min or longer rather than the eating-disordered behaviors themselves. Results showed that the EMA-measured behaviors correlated moderately with comparable scores on the EDE, the 26-item Eating Attitude Test (EAT-26; Garner, Olmsted, Bohr, & Garfinkel, 1982), and the Dutch Eating Behavior Questionnaire (DEBQ; van Strein, Frijters, Bergers, & Defares, 1986). Mean levels of the behaviors were not reported. Consequently, the significance and direction of the differences are unknown.

The cognitive processes underlying retrospective frequency judgments provide one plausible explanation for the low concurrence between the retrospective and prospective measurement of ED behaviors. Studies on survey methods and autobiographic memory suggest that responses to behavioral frequency questions are formulated as the result of one of two general cognitive processes: retrieving and counting episodic memories of specific instances of the behavior or retrieving higher-order semantic knowledge structures that reflect common properties derived from repeated instances of a behavior (Conrad, Brown, & Cashman, 1998; Sudman, Bradburn, & Schwarz, 1996). Brown (1995) referred to these processes as enumeration and nonenumeration strategies, respectively.

Whether enumeration or nonenumeration cognitive strategies are used to respond to a behavioral frequency question is dependent on characteristics of the target behavior and the manner in which the information is stored in memory (Menon, 1993). Behaviors that occur frequently and are highly similar both in nature and contextual features are represented in memory in the form of abstract semantic knowledge. Specific episodic memories of the behaviors are more difficult to access or are inaccessible over time (Friedman & deWinstanley, 1998). For these types of behaviors, frequency responses reflect estimations derived from the generalized semantic knowledge and high rates of recall error have been observed (Means & Loftus, 1991). A nonenumerative process leads to a nonnumeric response. Therefore, an additional step in the process must occur to derive a numeric response (Brown, 1995).

Enumerative strategies refer to the process of searching the memory for specific behavioral episodes, referred to as episodic memories, and counting them. Using this strategy, the frequency estimate is equivalent to the number of episodes activated in working memory. Behaviors that occur infrequently, at irregular intervals, or are highly dissimilar are less likely to be summarized in memory (Menon, 1994). Frequency judgments for these behaviors trigger a search of episodic memory and lead to a count of discrete episodes. Some investigators suggested that the enumeration process is likely to lead to underreporting because individuals may forget or fail to access episodes (Brown, 1995). However, other studies have shown that enumeration is associated with overreporting because of the telescoping effect or the tendency of respondents to carry forward and count events that occurred before the reference time period (Bradburn, Huttenlocher, & Hedges, 1994).

Although the frequency, regularity, and similarity of behaviors influence frequency estimates, these properties of ED behaviors have not been investigated systematically and their effects on approaches to ED behavioral frequency measurement are unknown. The diagnostic criteria for BN (American Psychiatric Association, 2000) specify a minimum of two binge eating and compensatory behavior cycles per week, yet there is marked variability in the characteristics of the behaviors (Walsh & Kahn, 1997). Even within the group of women who meet criteria for BN, behaviors are highly individualistic. This problem is even more pronounced in women with AN because the diagnosis is based on the biologic outcomes of behaviors, not on the characteristics of the behaviors themselves.

EMA

The EMA uses monitoring and sampling strategies to measure target phenomena as they occur in the daily natural setting. Participants record a target experience (e.g., mood, physical symptoms, or specific behaviors) many times daily so that the time interval between the target event and recording is greatly reduced or eliminated. Three types of event sampling schemes are used with the EMA: time contingent, signal contingent, and event contingent (Stone & Shiffman, 1994). Time-contingent sampling requires recording at a specified and regular time interval (e.g., daily bedtime recording). Signal-contingent sampling refers to event recording that is triggered by a signaling device (e.g., an alarm watch or a hand-held computer). Although signals may occur at regular intervals, the methodologic risks of entrainment to the participant's schedule (oversampling of a limited variety of activities due to consistency in signaling) and alteration of activities due to anticipation of the signal have led to the more frequent use of random signaling frames.

In this study, we use event-contingent recording. With this approach, the frequency and timing of the recordings are determined by the occurrence of the target behavior itself (Stone & Shiffman, 1994). Participants are trained to identify the target behavior and instructed to record the behavior immediately after it occurs. Event-contingent sampling is used most commonly for low-frequency events and was selected for this study for several reasons. First, even at the full threshold level of BN, ED behaviors may not occur more frequently than twice a week. To signal participants repeatedly throughout the day would be highly burdensome and may undermine recording compliance. Less frequent intervals (e.g., once or twice daily) would increase the risk of significant delays between the event and the recording and undermine the basic logic of the EMA. In addition, use of either the time or signal-contingent approach may result in signals that occur while the participant is engaged in social contexts at which time completion of this highly sensitive questionnaire would be difficult or embarrassing. As a result of inopportune timing, compliance with the measure may be compromised.

An important technologic advance associated with the EMA is the use of a hand-held computer as a recording device (Stone & Shiffman, 1994). In the past, BN and BED subjects, responding to time and event-contingent signaling, recorded their behavior in paper-and-pencil diaries. The reliability of the prospective measurement of behaviors using a paper-and-pencil diary has been seriously questioned. Stone and Shiffman (2002) used a diary notebook fitted with a photosensor that detected light. Their goal was to measure the time the notebook was opened and the time the diary was completed. Actual recording time measured by the photosensor was compared with participant-documented recording date and time. Results showed that only 20% of paper-and-pencil diaries

were completed within a 90-min interval of the target time, which is a sharp contrast to the 90% compliance reported by participants. In addition, results showed that 75% of the sample had at least 1 day of “hoarding,” which was defined as a day in which the notebook was not opened according to the photosensor but a diary was completed for that day.

Although the EMA has been used to measure diverse health behaviors with normal and clinical samples, the feasibility of measuring ED behaviors using the EMA with a hand-held computer in women with AN and BN is unknown. Studies on self-monitoring reactivity suggest that the process of recording behaviors as they occur may alter systematically the frequency of the target behavior, thereby compromising the reliability of the measure. Although some studies have shown no reactivity effects (Hufford, Shields, Shiffman, Paty, & Balabanis, 2002; Sobell, Bogardis, Schuller, Leo, & Sobell, 1989), there is some evidence to suggest that this method of measurement causes, at least temporarily, an increase in positive behaviors and a decrease in negative or socially undesirable behaviors (Kazdin, 1974; Willis & Nelson, 1982).

A computerized ED behavior questionnaire was developed for this study and a 4-week period of event-triggered EMA was used to measure ED behaviors in a sample of women with symptoms of the binge-purging type of AN or BN. An exit questionnaire assessed adherence and participant reaction to the EMA. Change in the frequency of behaviors across the 4-week period was used to examine reactivity effects and correspondence with a retrospective measure of the behaviors was also explored.

METHODS

Sample

The sample included 16 women. Eleven met threshold or subthreshold criteria for BN and 5 met threshold or subthreshold criteria for AN, binge-purging type. The mean age of the women was 23.2 years ($SD = 4.8$). Eighty-eight percent ($n = 14$) of the sample was Caucasian and 2 participants (12.5%) were Asian. Fifty-six percent of the sample ($n = 9$) had some college education and 38% ($n = 6$) had a college degree or higher. The mean body mass index (BMI) was 23.7 ($SD = 4.3$). However, 1 participant with threshold AN refused to be weighed. The mean duration of symptoms was 5.1 years ($SD = 4.8$).

Participants were recruited through community-based advertising. The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID; First, Spitzer, Gibbon, & Williams, 1997) was used to establish the presence of current BN or AN, binge-purging type. The SCID was administered by a clinically experienced research assistant.

Measures

Disordered Eating Behaviors Questionnaire

A menu-driven computerized disordered eating behaviors questionnaire was developed for this study and programmed for use on a hand-held computer. The questionnaire addresses five ED behaviors (vomiting, laxative use, diuretic use, exercising, and binge eating). Items were derived from commonly used disordered eating behavior questionnaires including the EDE (Fairburn & Cooper, 1993) and the Questionnaire on Eating and Weight Patterns-Revised (QEWPR; Spitzer, Yanovski, & Marcus, 1994). In all cases, the items were written to focus on the single episode of the behavior that had just occurred.

The first screen of the computerized questionnaire comprises a list of the five ED behaviors with the written instruction, "Select a type of behavior to record." The participant used the stylus to indicate the behavior that had just occurred and based on the ED behavior selected, a brief series of three to five questions were presented. The binge menu focused on quantity, duration, and lack of control. For other behaviors, questions focused on whether the behavior was used for weight control, qualitative characteristics (e.g., the type of pill or exercise), and quantitative characteristics (e.g., the number of pills and the duration of exercise).

After recording the first behavior, the participant is asked whether she would like to record another behavior. If her response to that question is yes, the questionnaire returns to the original menu of behaviors. This process continues until the participant indicates that she has no other behavior to record, at which time the interview is terminated. Responses are entered automatically with a date and time stamp that cannot be erased by the participant.

EDE

The EDE (12th ed.; Fairburn & Cooper, 1993) was used to measure retrospectively the frequency of ED behaviors over the last 4 weeks. To adhere to the EDE administration instructions, the interview began with a calendar review to orient the individual to the time period under consideration (Fairburn & Cooper, 1993). The EDE has discriminant validity and interrater reliability (Cooper & Fairburn, 1987). Interviews were conducted by an experienced clinical research assistant and items were coded using instructions provided by Fairburn and Cooper (1993).

Exit Questionnaire

An exit questionnaire developed for this study measured participant adherence and reaction to the EMA. The questionnaire consisted of 13 close-ended questions. Five of the close-ended items focused on accuracy, thoroughness, and the timing of behavioral recordings. Items were rated on a 5-point scale. The last eight questions focused on the emotional and behavioral consequences of recording ED behaviors.

Procedures

During the first data collection session, participants provided informed consent and the SCID was administered to determine participant eligibility. Participants were informed that the purpose of the study was to examine factors that influence weight-control behaviors. They were not informed of the retrospective measurement of ED behaviors planned for the final data collection session.

During the second session, participants were oriented to the EMA procedures. They were provided with detailed instructions on the interview questions, procedures for recording their behaviors, and instructions for using the hand-held computer. Instructions were given verbally and a detailed instruction manual was given to participants. An error log was then given to the participant to record by hand any failed entries on the hand-held computer and/or to correct any incorrect entries. At the completion of the orientation session, a 24-hr practice period began. At the end of the practice period, a research assistant contacted the participant by phone to address any problems or questions and the 4-week EMA period began. Participants met with a research assistant weekly during the EMA period to backup the data from the hand-held computer to a laptop computer and to address problems/questions regarding the EMA.

At the end of the EMA period, when the hand-held computers and error logs were collected, adjustments in the data file were made to accommodate the entries on the error log. Three of the 13 participants recorded at least one error on the error log. One participant forgot her hand-held computer when she went home for 10 days and recorded behaviors on paper during that interval. Another participant, upon further reflection, questioned whether her responses to the bingeing and vomiting questions were accurate. Because she indicated on the error log that she was unsure, her data were not changed. A third participant indicated that she failed twice to record behaviors due to "laziness," and "tiredness." An additional 5 participants indicated that they never failed to record a behavior. The remaining 5 participants reported on the exit interview that they had failed to record behaviors once or twice during the entire measurement period. Because these errors were not entered on the error log, the data file was not changed to accommodate them.

A final session was scheduled within 48 hr of completing the EMA to complete the EDE. After completion of the EDE, the exit interview was administered. Participants were then debriefed and paid \$50 as compensation.

RESULTS

Feasibility of the EMA

Of the original 16 participants enrolled in the study, 2 participants (12.5%) did not complete the EMA portion of the study. They cited difficulty with the time demands of recording. A third participant completed the protocol. However, she had difficulties with the hand-held computer and the accuracy of her data was questionable. Consequently, the data provided by the remaining 13 participants were included in the analyses. Over the 4-week EMA period, participants recorded a total of 24.9 disordered eating behavior episodes ($SD = 29.7$; range = 5–101). The total number of episodes reported across the total sample was 355. Participants reported that they failed to record 1.9 behavioral episodes ($SD = 2.5$, range 0–8) over the 4-week period. Reasons cited for failure to record behaviors included embarrassment about recording the behavior in public, forgetting to carry the hand-held computer at all times, and being rushed. All participants indicated that the behaviors were recorded accurately most or all of the time and that they never or rarely failed to record a behavior. Forty-six percent of the sample reported that they recorded behaviors immediately after they occurred all or most of the time. However, 15% reported that they recorded a behavior 2 hr or more after it occurred most of the time. Because the data were date and time stamped (participants were not made aware of this), we examined the data for patterns that may indicate hoarding, but found little evidence of this. Although 51% of the behavioral recordings were instances when two or more behaviors were recorded within a 5-min interval, the majority of these were bingeing and compensatory behavior(s). There were very few instances where the same behavior was recorded more than once within 5 min. In addition, one person recorded multiple episodes at the beginning or end of the day only. Other evidence of adherence to the methodology was that slightly less than one half of the participants indicated that they did not experience the EMA methodology as an inconvenience and one third of the participants rated the methodology as only somewhat inconvenient. Nine of the participants (69%) indicated that the EMA did not increase their thoughts about food and

weight and 92% ($n = 12$) indicated that they would be willing to participate in another EMA study.

Reactivity to the EMA

To address the issue of reactivity, two sets of analyses were completed. First, a sum score of the frequency of the five behaviors was computed for the first 2 weeks of the data collection period and compared with the sum computed for the second 2-week period. Second, individual behavior frequencies were compared between the first and second 2-week periods. Table 1 shows the means and standard deviations for the behaviors for the first and second halves of the EMA measurement period. An examination of the pattern of means suggests that the frequencies for all ED behaviors, except diuretic use, were lower in the second half of the measurement period compared with the first half of the measurement period. However, as expected with frequency data, the behavioral frequency scores were not distributed normally, making the means an unreliable indicator of behavioral frequency. Therefore, the Wilcoxon signed ranks test, a nonparametric statistic, was used to determine whether there were differences between any of the behavioral frequencies in the first and second halves of the measurement period. An examination of the pattern of ranks indicates that for all behaviors, approximately one half of the time the behaviors were more frequent in the first half of the measurement period (i.e., negative ranks) and approximately one half of the time they were more frequent in the second half of the measurement period (i.e., positive ranks; Table 1). Therefore, despite the fact that on average across all cases, the means for most behaviors were lower in the second half of the measurement period compared with the first, nearly as many participants had higher frequencies in Weeks 3–4 as those who had higher frequencies in Weeks 1–2 for each behavior. Because laxative and diuretic use was so infrequent, most ranks were ties (both 0). There were no significant differences in

Table 1. Descriptive statistics and Wilcoxon signed ranks test to examine eating disorder behavioral frequency distributions in the first and second halves of the EMA measurement period ($N = 13$)

Characteristics	Weeks 1–2 <i>M (SD)</i>	Weeks 3–4 <i>M (SD)</i>	Negative Ranks (N) ^a	Positive Ranks (N) ^b	Ties (N) ^c	<i>Z</i>	<i>p</i>
Number of binge episodes	4.92 (9.12)	2.77 (4.55)	6	4	3	–.514 ^d	ns
Number of vomiting episodes	8.38 (14.40)	4.92 (7.40)	5	3	5	–1.489 ^d	ns
Number of laxative episodes	0.15 (0.55)	0.08 (0.28)	1	0	12	–1.000 ^d	ns
Number of diuretic episodes	0.85 (2.48)	1.00 (3.61)	2	1	10	0.000 ^e	ns
Number of excessive exercise episodes	1.31 (1.97)	0.69 (1.44)	4	1	8	–1.511 ^d	ns
Sum of episodes	15.62 (22.32)	9.46 (10.84)	7	6	0	–1.157 ^d	ns

Note: EMA = ecologic momentary assessment.

^aNumber of behaviors during Weeks 3–4 is less than the number of behaviors during Weeks 1–2.

^bNumber of behaviors during Weeks 3–4 is more than the number of behaviors during Weeks 1–2.

^cNumber of behaviors during Weeks 3–4 is equal to the number of behaviors during Weeks 1–2.

^dBased on positive ranks.

^eSum of negative ranks = sum of positive ranks.

the frequency distributions for any of the ED behaviors across the first and second halves of the measurement period. Because the period of time in which reactivity might be observed is unknown, the 1-month period was also divided into thirds and a non-parametric test for multiple variables, a Friedman test, was completed. Once again, no significant differences were found among the first, second, and third periods of the month for the frequency distributions of the five ED behaviors ($\chi^2 < 2.0$, all $ps = ns$).

Correspondence between the EMA and EDE Measures of Behavioral Frequencies

To examine the pattern of correspondence between EMA and EDE, we computed a sum score of the frequencies of all five behaviors. A frequency score for each behavior was also computed. Table 2 shows the means and standard deviations for the EMA and EDE total sum scores and the individual behavioral frequency scores. An examination of the pattern of means suggests that the frequencies for all ED behaviors except laxatives were lower when measured by the EMA than by the EDE. The Wilcoxon test showed that the behavioral frequency distributions were different for bingeing, excessive exercise, and the sum of all ED behaviors using the two methods of measurement (Table 2). This indicates that bingeing and excessive exercise were reported more frequently when measured by the EDE compared with the EMA. Because laxative and diuretic use was so infrequent, most ranks were ties (both 0).

Finally, Kendall's tau correlation coefficients were also computed between EMA and EDE measurements of the total sum of ED behaviors and the individual behavioral frequency scores. The correlation for the sum of all ED behaviors between the two measures was .50 ($p < .02$). For bingeing, the correlation was .60 ($p < .01$), vomiting was .75 ($p = .001$), laxative use was 1.0 (only 1 person used laxatives), diuretic use was .81 ($p = .003$), and excessive exercise was .62 ($p = .01$).

Table 2. Descriptive statistics and Wilcoxon signed ranks test to examine eating disorder behavioral frequency distributions measured by the EMA and the EDE ($N = 13$)

Characteristics	EMA M (SD)	EDE M (SD)	Negative Ranks (N) ^a	Positive Ranks (N) ^b	Ties (N) ^c	Z	p
Number of binge episodes	7.62 (11.51)	14.23 (18.77)	2	9	2	-2.273 ^d	.023
Number of vomiting episodes	13.31 (20.43)	16.38 (21.61)	3	5	5	-.981 ^d	ns
Number of laxative episodes	0.23 (0.83)	0.08 (0.28)	1	0	12	-1.000 ^e	ns
Number of diuretic episodes	1.77 (5.79)	2.92 (10.24)	1	1	11	-.447 ^d	ns
Number of excessive exercise episodes	2.00 (3.16)	8.15 (10.75)	2	6	5	-2.103 ^d	.035
Sum of episodes	24.92 (29.74)	41.77 (36.98)	10	3	0	-2.517 ^e	.012

Note: EMA = ecologic momentary assessment; EDE = Eating Disorder Examination.

^aNumber of behaviors EDE is less than the number of behaviors using EMA.

^bNumber of behaviors EDE is greater than the number of behaviors using EMA.

^cNumber of behaviors EDE is equal to the number of behaviors using EMA.

^dBased on negative ranks.

^eBased on positive ranks.

DISCUSSION

The first goal of this study was to examine the feasibility of using EMA and measure ED behaviors. The results of this pilot study provide preliminary evidence to suggest that the EMA is a feasible and nonreactive approach to the measurement of binge eating, vomiting, excessive exercise, and laxative and diuretic use in a community sample of women with threshold and subthreshold levels of AN and BN. Eighty-eight percent of our sample completed the 1-month EMA behavioral recording. Only 2 participants indicated that the protocol was too burdensome to complete and 1 participant's data were lost due to computer problems. In addition, participants generally indicated that their behaviors were recorded accurately, that they rarely or never missed a recording, and that they found the measurement approach only minimally or somewhat inconvenient. The self-report data suggest that behaviors were recorded typically within 2 hr of their occurrence and a visual inspection of the pattern of recordings suggests limited episodes of data hoarding or delayed recording. Finally, and perhaps the most impressive indicator of the feasibility of EMA measurement with this population, the majority of women indicated that they would be willing to participate in another EMA study.

Our second goal was to investigate reactivity of ED behaviors. Our failure to find differences in the frequency distributions of all five ED behaviors across the 1-month EMA period suggests that the occurrence of these behaviors in a clinical ED sample is not reactive to the measurement approach. This finding is consistent with recent EMA studies that show a lack of reactivity effects for other negative health behaviors such as alcohol consumption in clinical and subclinical samples (Hufford et al., 2002; Litt et al., 1998). The total mean frequency, as well as the mean frequencies of each behavior, was lower in the second half of the data collection period compared with the first half. This finding raises some question about whether there was inadequate power to detect reactivity effects. However, the pattern of ranks was not consistent with a reactivity effect. For example, approximately one-half of the time, all behaviors were more frequent in the first half of the measurement period and approximately one-half of the time they were more frequent in the second half of the measurement period.

The third goal of this study was to investigate the pattern of correspondence between the EMA and the EDE, a retrospective measure of ED behaviors. At first, the finding that only the frequency of binge eating and excessive exercise episodes are significantly different between the EMA and EDE approaches is not easy to explain. One plausible explanation for this finding focuses on the differences in complexity of the behaviors and the level of detail of the associated memory structure needed to evaluate a behavior as an episode or not. For example, to determine if an instance of overeating is a binge, it is necessary to distinguish that the quantity of food was unusually large, consumed within a 2-hr period of time, and was associated with feelings of loss of control. In circumstances in which the behavior has just occurred, it is possible to compare the episode against these objective criteria to determine whether the episode was a true binge. However, because detailed characteristics of the specific episodes may not be retained in memory for more than a few days (Friedman & deWinstanley, 1998), the ability to evaluate specific characteristics of the behavioral episodes a week to several weeks later is likely to be lost. In this case, even though the interviewer provides detailed information about the essential properties of an objective binge, the lack of detailed information available in memory may prevent a careful discrimination and contribute to an overestimation of episode frequency. As long as the behavior was perceived and encoded initially as a

binge, efforts to make a fine discrimination are unlikely to be successful given that this level of detail about the behavior is not available in memory. A similar process may occur in the retrospective reporting of excessive exercise episodes.

In a related study, Ortega et al. (1987) found a similar pattern of discrepancy between concurrent and retrospective measurement of ED behaviors. In their study, ED women recorded binge eating and purging episodes in a daily diary and again at a weekly treatment session. Women reported an equivalent number of purging episodes daily ($M = 10.24$) and weekly ($M = 9.17$). However, the number of binge episodes was higher on the retrospective measure ($M = 6.88$) compared with the daily diary ($M = 2.33$). In addition, the investigators reported that 41% of the disparity between the two measures was due to inconsistent reporting on the two measures and 59% of the disparity was due to the fact that an episode in the diary failed to meet criteria of an objective binge (e.g., the amount of food was not sufficiently large). These findings are consistent with the view that retrospective recall is associated with overestimates of more complex ED behaviors.

An important limitation of this study is the small sample size. Additional studies with larger clinical samples are needed to further investigate the extent to which ED behaviors are reactive to the EMA measurement approach. Given the small sample used in this pilot test, our ability to detect a reactivity effect may be limited. Therefore, a moderate or low level of behavioral reactivity cannot be ruled out. Studies that extend the EMA period beyond 1 month may be able to determine not only if the ED behaviors are reactive, but also if reactivity does occur, such studies would be able to explore in more detail the pattern of onset of behavioral reactivity, the degree of persistence, and the time period needed for behavioral levels to restabilize. In addition, studies using other methods of measurement, such as an actigraph to measure physical activity and energy expenditure, would be useful in further evaluating the validity of the EMA. Although it is possible to argue theoretically that the retrospective recall of complex behaviors such as binge eating and excessive exercise episodes is likely to lead to overestimations, related biologic indicators would provide important information about the reliability of EMA compared with the retrospective measurement approach. However, despite the limitations of this pilot study, the results provide promising preliminary evidence to suggest that EMA is both a feasible and valid approach to the measurement of ED behaviors in women with AN and BN.

The authors acknowledge Denise St. Arnault and Elise Erickson for their participation in the project, particularly their assistance in the development of the computerized menu-driven questionnaire. They also thank Qian Yang for programming the hand-held computers, Monica White and Nada Abu-Isa for data collection, and David Ronis for data management and analysis. Supported by a faculty grant from the Undergraduate Research Opportunity Program at the University of Michigan.

REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th Rev. ed.). Washington, DC: Author.
- Becker, A., Grinspoon, S., Klibanski, A., & Herzog, D. (1999). Eating disorders. *The New England Journal of Medicine*, 340, 1092–1098.
- Bradburn, N., Huttenlocher, J., & Hedges, L. (1994). Telescoping and temporal memory. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 203–215). New York: Springer-Verlag.

- Brown, N. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology*, 21, 1539–1553.
- Conrad, F., Brown, N., & Cashman, E. (1998). Strategies for estimating behavioral frequency in survey interviews. *Memory*, 4, 339–366.
- Cooper, Z., & Fairburn, C. (1987). The Eating Disorder Examination: A semistructured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders*, 6, 1–8.
- Fairburn, C.G., & Beglin, S. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders*, 16, 363–370.
- Fairburn, C.G., & Cooper, Z. (1993). The Eating Disorder Examination. In C.G. Fairburn & T.G. Wilson (Eds.), *Binge eating: Nature, assessment and treatment* (12th ed., pp. 317–360). New York: Guilford Press.
- First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J. (1997). *User's guide for the Structured Clinical Interview for DSM-IV Axis I Disorders*. Washington, DC: American Psychiatric Press.
- Friedman, W., & deWinstanley, P. (1998). Changes in the subjective properties of autobiographical memories with the passage of time. *Memory*, 4, 367–381.
- Garner, D.M., Olmsted, M.P., Bohr, Y., & Garfinkel, P. (1982). The eating attitudes test: psychometric features and clinical correlates. *International Journal of Eating Disorders*, 3, 15–34.
- Grilo, C., Masheb, R., & Wilson, G.T. (2001a). A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorders. *Journal of Consulting and Clinical Psychology*, 69, 317–322.
- Grilo, C., Masheb, R., & Wilson, G.T. (2001b). Different methods for assessing the features of eating disorders in patients with binge eating disorder: A replication. *Obesity Research*, 9, 418–422.
- Hufford, M., Shields, A., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors*, 16, 205–211.
- Kazdin, A. (1974). Reactive self-monitoring: The effects of response desirability, goal setting, and feedback. *Journal of Consulting and Clinical Psychology*, 42, 704–716.
- Litt, M., Cooney, N., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodologic problems and potential solutions. *Health Psychology*, 17, 48–52.
- Means, B., & Loftus, E. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology*, 5, 297–318.
- Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research*, 20, 431–440.
- Menon, G. (1994). Judgments of behavioral frequencies: Memory search and retrieval strategies. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 161–172). New York: Springer-Verlag.
- Ortega, D., Waranch, H.R., Maldonado, A.J., & Hubbard, F. (1987). A comparative analysis of self-report measures of bulimia. *International Journal of Eating Disorders*, 2, 301–311.
- Rosen, J., Vara, L., Wendt, S., & Leitenberg, H. (1990). Validity studies of the Eating Disorder Examination. *International Journal of Eating Disorders*, 9, 519–528.
- Rossiter, E., Agras, W.S., Telch, C., & Bruce, B. (1992). The eating patterns of nonpurging bulimic subjects. *International Journal of Eating Disorders*, 2, 111–120.
- Safer, D., Telch, C., & Agras, W. (2001). Dialectical behavior therapy for BN. *American Journal of Psychiatry*, 158, 632–634.
- Shiffman, S., Fischer, L., Paty, J., Gnys, M., Hickcox, M., & Kassel, J. (1994). Drinking and smoking: A field study of their association. *Annals of Behavioral Medicine*, 16, 203–209.
- Smyth, J., Wonderlich, S., Crosby, R., Miltenberger, R., Mitchell, J., & Roty, M. (2000). The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders*, 30, 83–95.
- Sobell, M., Bogardis, J., Schuller, R., Leo, G., & Sobell, L. (1989). Is self-monitoring of alcohol consumption reactive? *Behavioral Assessment*, 11, 447–458.
- Spitzer, R., Yanovski, S.Z., & Marcus, S. (1994). *Questionnaire on Eating and Weight Patterns-Revised (QEWP-R)*. Pittsburgh: Behavioral Measurement Database Services.
- Steiger, H., Gauvin, L., Jabalpuwala, S., Seguin, J., & Stolt, S. (1999). Hypersensitivity to social interactions in bulimic syndromes relationship to binge eating. *Journal of Consulting and Clinical Psychology*, 67, 765–775.
- Stone, A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, 16, 199–202.
- Stone, A., & Shiffman, S. (2002). Patient non-compliance with paper diaries. *British Medical Journal*, 324, 1193–1194.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Van Strein, T., Frijters, J., Bergers, G., & Defares, P. (1986). The Dutch Eating Behaviors Questionnaire (DEBQ) for the assessment of restrained, emotional and external eating. *International Journal of Eating Disorders*, 5, 295–315.
- Walsh, B.T., & Kahn, C. (1997). Diagnostic criteria for eating disorders: Current concerns and future directions. *Psychopharmacology Bulletin*, 33, 369–372.
- Willis, S., & Nelson, R. (1982). The effects of valence and nature of target behavior on the accuracy and reactivity of self-monitoring. *Behavioral Assessment*, 4, 401–412.
- Wilson, G.T., & Smith, D. (1989). Assessment of bulimia nervosa: An evaluation of the Eating Disorder Examination. *International Journal of Eating Disorders*, 8, 173–179.