Ecological Momentary Assessment to Evaluate Cognitive-Behavioral Treatment for Binge Eating Disorder

Simone Munsch, PhD
Andrea H. Meyer, PhD
Natasa Milenkovic, MSc
Barbara Schlup, MSc
Juergen Margraf, PhD
Frank H. Wilhelm, PhD

ABSTRACT
Objective: Cognitive-behavioral treatment (CBT) for binge eating disorder (BED) is traditionally evaluated using clinical interviews and questionnaires. These retrospective assessment methods are discussed to be problematic due to memory recall error. Ecological momentary assessment (EMA) might be promising for gathering ecologically valid and reliable data.

Method: We assessed the feasibility of and reactivity to EMA and compared the treatment efficacy measured by traditional vs. EMA-based instruments in 28 BED individuals participating in short-term CBT.

Results: Patients were highly compliant and we found no reactivity to EMA. Estimated treatment effects for binge eating based on EMA were comparable to questionnaire-based methods. The overall concordance between methods was moderate.

Discussion: Results suggest that binge eating over 1 week can be equally accurately assessed by EMA or by self-report questionnaires in BED treatment trials. EMA contributes to a detailed knowledge of binge eating in daily live and helps to advance treatment options.

Keywords: ecological momentary assessment; binge eating disorder; treatment efficacy; ambulatory assessment; computerized assessment

Introduction
Binge eating disorder (BED) is a common problem within the obese population with up to 30% suffering from this eating disorder. In community-based studies, prevalence rates range from 0.7 to 3.3%. BED imposes significant psychosocial burden and is often associated with impaired quality of life, obesity, and increased morbidity and mortality. According to the guidelines of the American Psychiatric Association and the National Institute of Clinical Excellence for the treatment of eating disorders, cognitive-behavioral treatment (CBT) is the best-established treatment for BED. The primary source of data in clinical trials evaluating treatment efficacy in BED are clinical questionnaires and standardized clinical interviews, both relying on retrospective memory recall. These traditional data recording methods are known to have several shortcomings including the forgetting of information when the recall period is long and/or memory retrieval occurs in a different setting from memory storage (e.g., recall of binge eating episodes during the past month in a therapist room, see for e.g.). Further, the most salient and most recent events at the time of recall will influence the ratings of behavior or mood from the period that is being recalled. Most importantly, recent studies indicate that retrospective recall favors mental representations fitting into the self-concept, social expectations, and ideals of individuals rather than representing actual experiences or the target behavior examined.

In contrast, ambulatory psychological assessment such as the experience sampling method, later also called ecological momentary assessment (EMA), allows getting closer to the real-life situations under study, both in space and time, thereby reducing retrospective recall problems.

EMA has been applied to self-assessment of behavior and psychological state in a range of

(Received 20 December 2008
Parts of the manuscript have been presented at the World Congress of Psychology, Berlin, July 2008.
*Correspondence to: Simone Munsch, PhD, Faculty of Psychology, University of Basel, Missionsstrasse 60/62a, 4055 Basel, Switzerland. E-mail: simone.munsch@unibas.ch
1 Faculty of Psychology, Department of Clinical Psychology and Psychotherapy, University of Basel, Basel, Switzerland
2 Faculty of Psychology, Department of Clinical Psychology and Psychotherapy, Division of Applied Statistics in Life Sciences, University of Basel, Basel, Switzerland
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/eat.20657
© 2009 Wiley Periodicals, Inc.)
disorders and natural settings (for an overview see Engel et al.\textsuperscript{15}), including eating disorders.\textsuperscript{13–15} Previous studies on traditional questionnaire-based self-monitoring have shown that systematic recording of behaviors can lead to temporary decrease of problematic or socially undesirable behaviors.\textsuperscript{16} Nevertheless, current findings of studies in alcohol abuse and eating disorders have shown no reactivity effects after multiple electronic assessment of the target behavior.\textsuperscript{14,17,18} Several recent diary studies evaluated concordance of EMA-data with data from retrospective questionnaires or interviews and reported notable discrepancies between near-real-time reporting of experiences and later recall of the same events. In most studies with pain disorder patients, higher pain ratings were found by retrospective assessment.\textsuperscript{19,20}

In eating disorders, a study with subthreshold bulimia and anorexia nervosa patients examined the concordance between EMA and the eating disorder examination (EDE),\textsuperscript{18,21} the common clinical interview for assessing eating disorders. Again the results indicated that binge eating and excessive exercise over the time span of 28 days were lower when measured by EMA compared to the interview. The identification of a binge eating episode refers to a complex recollection process for verifying that the quantity of food was unusually large, consumed within a 2-h period of time, and was associated with feelings of loss of control.\textsuperscript{18} It is highly likely that this detailed information was not remembered accurately for each binge-like episode when the interviewer asked for it several weeks later. The observed discrepancies matter because both researchers and clinicians base their interventions on patient reported data, and thus might be misleading. But it remains open whether the suggested retrospective memory recall bias of an especially complex behavior such as binge eating also occurs when binge eating is assessed during a shorter time-span of several days.

Studies investigating the lack of concordance between EMA-based and traditional questionnaire- or interview-based data have considered personality and clinical traits such as negative affect, severity, and stability of symptoms and characteristics of general psychopathology as possible factors that might influence lack of concordance between the two sources of data and yielded mixed results. Stone et al.\textsuperscript{22} found an overall low correspondence between retrospectively and momentarily assessed coping behavior in substance abuse, especially for complex behavior, but revealed no association between the discrepancy rates and factors such as stress level, anxiety state, depression, neuroticism, or marital satisfaction.

Despite the apparent advantages of the electronic diary method, only few studies have utilized EMA for assessment of treatment outcome in psychotherapy research.\textsuperscript{23} The results from these studies, mostly conducted with pain disorder patients, indicate that EMA is highly feasible, leads to little reactivity, and increases the accuracy of treatment outcome assessment in this patient group.

Thus, EMA appears to have the potential to increase the knowledge about patient experience and symptoms and furthermore, may reduce error variance, and thus help to reduce the number of participants required in psychotherapy research. However, the reliability and validity of this method for assessing complex behavior such as binge eating have not been addressed adequately.\textsuperscript{24} In BED, only preliminary data about the application of EMA in treatment trials have been collected. These studies focused on evaluating the effects of EMA-based self-monitoring of behavior as an additional therapeutic component and a comparison of the eating disorder characteristics of BED and obese patients during the treatment course,\textsuperscript{12,25} and did not systematically assess the feasibility and concordance of EMA vs. traditional instruments.

The aims of the present study were threefold. We, first, systematically investigated the feasibility of EMA in terms of patient acceptability, practicability, and compliance during a treatment trial for BED. We also accounted for the problem of reactivity, and assessed whether binge eating was altered due to the repeated assessment using EMA. Second, we compared EMA-based treatment outcomes with outcomes based on traditional self-report questionnaires and with outcomes based on EDE,\textsuperscript{26,27} using effect sizes. Third, we analyzed the concordance between questionnaire- or interview-based measures and EMA-based measures using indices of inter-instrument agreement, thereby accounting for the moderating influence of severity of eating disorder pathology, general psychopathology, specific mental disorders, and Body Mass Index (BMI).\textsuperscript{28,29} We expected that BED patients with higher BMI, more severe eating disorder, more comorbid mental disorders, and more negative affect would be more likely to exhibit discordance between traditional and EMA-based reports.

Method

Participants

Overweight to obese individuals with BED were recruited through newspaper advertisements for a
treatment trial at the Department of Clinical Psychology and Psychotherapy, University of Basel, Switzerland. Initially 136 individuals contacted the department and underwent telephone screening for determining preliminary eligibility. Study inclusion criteria required that participants were between 18 and 70 years old, had a BMI (kg/m²) between 27 and 40, and met full DSM-IV-TR criteria for BED according to a specialized eating disorder interview (see diagnostic assessment). All patients were free from unstable medical conditions such as diabetes, heart disease, or endocrine disorders. Participants were excluded if they met DSM-IV-TR criteria assessed by standardized interviews (see diagnostic assessment) for mental disorders warranting immediate treatment (IT) such as suicidal tendency, psychosis, mania, organic dementia, or substance use disorder. Further exclusion criteria were pregnancy, participation in a diet program or other psychotherapy, treatment with weight loss medication (current or during the last 3 months), or previous surgical treatment of obesity. The surprisingly small number of males (N = 1) resulted in their exclusion from data analysis. The final sample consisted of 28 overweight to obese females with BED (see Table 1 for sample characteristics).

### Procedure

The study was approved by the local ethics committee for medical research. All subjects were offered free treatment for participating in the study. Prior to initial assessments, the study was explained to the participants and they gave written informed consent. BED individuals were randomly assigned to either IT or a waitlist (WL) condition using a permuted block design. Each IT or WL group started whenever five to eight participants had been recruited and randomized. Electronic diaries were applied twice in the IT group, at pre-treatment (one week before treatment started) and at post-treatment (last week of treatment). Electronic diaries were applied three times in the WL group, one week before the 8-weeks waiting period prior to treatment, at pre-treatment, and at post-treatment.

**Ecological Momentary Assessment (EMA) with Electronic Diaries.** A menu-driven computerized questionnaire was developed to repeatedly assess binge eating and attributes of psychological well-being (unpublished electronic questionnaire available from the authors). Questions consisted either of a dichotomous “yes/no” format, for example, item 1: “Did you experience binge eating since your last entry?” or used a computerized Likert-type scale ranging from 0 (“not at all”) to 4 (“very much”), for example, item 11: “How do you currently feel: depressed”. Palm Tungsten E handheld computers were used as recording devices. Questionnaires were programmed and displayed using Pendragon Forms 5.0 software (Pendragon Software Corporation, Libertyville, IL). We assessed participant responses using time- and event-contingent signaling during an entire week. All software applications on the handheld computer other than the electronic diary were blocked.

During a 30-min standardized session, master students of psychology instructed the participants on how to use the handheld computers and how to fill in the electronic diary. For time-contingent monitoring, five assessment points per day were preset at individually defined times, that is, each participant could specify his/her typical awakening time on weekdays and weekends. The first alarm was preset individually at one and a half hours after awakening, the second alarm 5 h after the first, the third alarm 4 h after the second, the fourth alarm 3 h after the third and the fifth alarm 2 h after the fourth alarm. The decreasing sampling interval during the day was chosen in accordance with previous studies that demonstrated an increased likelihood of binge eating later in the day. To reduce the subject burden, individual deviations from these times of up to 1 h were allowed in order to adapt the signaling to participants’ daily routines (e.g., to avoid alarming during classes and scheduled meetings). Participants were asked to fill in the questionnaire no later than 30 min after each signaling and were reminded within this interval by repeated signals. Data recordings not corresponding to this predefined time frame were only used for calculation of compliances but removed for further analysis. For event-contingent monitoring, participants were instructed to fill in the questionnaire whenever they felt that an episode of overeating associated with a sense of loss of control occurred, within a 30-min interval. After answering the entry question (“did you experience a binge episode?”) with yes, they were asked questions relating to the criteria of binge eating according to the EDE. After each assessment week, study participants were asked to fill in an exit questionnaire (EXQ), measuring adherence and reaction to and acceptability of the EMA procedure.

---

**Table 1. Sample characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Binge Eating Group (n = 28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age in years (SD)</td>
<td>45.1 (11.1)</td>
</tr>
<tr>
<td>Mean BMI (SD)</td>
<td>32.6 (6.4)</td>
</tr>
<tr>
<td>No. (%) of participants with comorbid diagnosis</td>
<td></td>
</tr>
<tr>
<td>Current comorbidity axis I&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14 (50)</td>
</tr>
<tr>
<td>Depression</td>
<td>2 (7.1)</td>
</tr>
<tr>
<td>Anxiety disorders</td>
<td>12 (42.9)</td>
</tr>
<tr>
<td>Lifetime comorbidity axis I&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9 (32.1)</td>
</tr>
<tr>
<td>Depression</td>
<td>6 (21.4)</td>
</tr>
<tr>
<td>Anxiety disorders</td>
<td>3 (10.7)</td>
</tr>
<tr>
<td>Comorbidity axis II&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1 (3.6)</td>
</tr>
</tbody>
</table>

**Notes:** Current and lifetime diagnoses were assessed according to DSM-IV-TR, using structured clinical interviews.

<sup>a</sup> Mini-DIPS<sup>31</sup>

<sup>b</sup> SKID-II<sup>32</sup>
Diagnostic Assessment. BED diagnosis relied on the structured clinical interview EDE\textsuperscript{26,27} and was in accordance with DSM-IV-TR research criteria.\textsuperscript{30} The German language screenings for mental disorders on axis-I (Diagnostisches Kurzinterview bei psychischen Störungen, Mini-DIPS)\textsuperscript{31} and axis-II (Strukturiertes Klinisches Interview für DSM-IV, Achse-II, Persönlichkeitsstörungen, SKID-II)\textsuperscript{32} were administered to assess the current and lifetime mental disorders. Interviewers were trained in administering the EDE, Mini-DIPS and SKID-II, and were supervised weekly by two of the authors (B.S. and S.M.). In cases of uncertainties about the diagnoses of patients, supervision of the diagnostic process was supported by videotapes taken during the assessment.

Body Weight

Body Mass Index (BMI). Weight and height were measured on an electronic balance scale (Seca, Vogel and Halke, Germany) and by a stadiometer. BMI was calculated as weight in kilograms divided by the square of height in meters.

Feasibility of and Reactivity to EMA. Feasibility of EMA, that is, practicability, acceptability, representativeness, and compliances, were assessed in two ways, with the EXQ and using EMA data. The EXQ assessed several perceptions of participants relating to the electronic diary assessment according to an 11-point Likert-type scale from 0 ("not at all") to 10 ("yes, exactly"). The following items of the EXQ assessed practicability ("Did you experience any difficulties in filling in the electronic diary?"), acceptability ("How did you feel during the week with the electronic diary entries? Did the electronic diary alarm go off too often? Was your daily routine disturbed?" (highly correlated items: \(r = .47-.72\)) and representativeness ("Did the previous week correspond to your usual weekly routine?"). Signal-compliance, defined as the filling in of the electronic questionnaire within a predefined time span after being alarmed (30 min) according to Fahrenberg\textsuperscript{4} as was assessed by asking: "Was it possible for you to fill in the electronic diary 30 min after the signal?" (EXQ-based signal-compliance). Time stamps for the start and end of each data entry were automatically recorded in the electronic diary. EMA-based signal-compliance was assessed by computing the proportion of the weekly number of recordings that were started within 30 min after the signal divided by the total weekly number of recordings. To assess whether patients responded to the signaling at all, we further determined recording compliance denoting the rate of overall responses to time-contingent signaling.

Reactivity was assessed according to the EXQ ("Did the frequency of binge eating change during the diary period? Did you focus more on your psychological well-being? Did you benefit from filling in the diary? Did the previous week correspond to your usual weekly routine?"; highly correlated items: \(r = .48-.70\)). We further tested whether using EMA for one week while waiting for treatment had an effect on the number of daily binge eating episodes over the course of the week, for both groups combined, and whether the number of weekly binge eating episodes changed between the two weeks prior to treatment in the WL group.

EMA-Based Treatment Outcomes. Patients were asked to monitor binge eating according to time-contingent and event-contingent assessment modes. The electronic diary contained questions that were based on EDE questions and focused on identifying binge eating (EMA weekly binges). The number of weekly binges was obtained by summing up all recorded binge eating episodes, from both time-contingent and event-contingent monitoring, across the 1-week assessment period. Summing up binge eating episodes across a specific time interval can be problematic if compliance is not 100%, as is usually the case. Disregarding missing values may lead to an underestimation of number of binge eating episodes if they are assumed to have a value of 0 (i.e., missing value = no binge). Replacing missing values by the mean across the defined time interval, in our study, led to slightly higher values but otherwise comparable results, so unadjusted values are reported. Summing up all days for which the patient did not report any binge eating episodes assessed the number of binge-free days. The maximum possible number was six as this was the number of days for which data for a complete 24 h day were available. Abstainers were defined as having zero binge eating episodes during a 1-week EMA period of investigation.

Interview and Questionnaire-Based Treatment Outcomes. Patients recorded their number of weekly binges (self-reported weekly binges; number of "episodes of overeating during which you felt out of control" during the past week) according to DSM-IV-TR criteria during the active treatment phase.\textsuperscript{35} Using Hilbert et al.\textsuperscript{27} German version of the EDE,\textsuperscript{21} we further assessed the number of objective binge eating episodes (OBE, i.e., binge eating defined as consuming unusually large quantities of food with a subjective sense of loss of control) and abstainer rates (proportion of patients not experiencing objective binge days) during the last 28 days.

Concordance Between EMA- and Interview-Based Estimates of Signal-Compliance and Outcome Measures. To keep the subject burden low in our treatment trial we could commit our patients to fill in the electronic diary only during seven days before and after treatment. Although this procedure allowed us to directly compare the objective binge eating and abstainer rates according to EMA and according to the traditional self-report questionnaire (both covering the 7-days time span), we could...
not directly assess the concordance between EMA and EDE-based measures of objective binge eating and abstinence from binge eating (7-days versus 28-days time span). Thus, OBEs according to EDE were divided by four to make values more comparable. Concordance of binge eating symptomatology assessed by traditional questionnaire/interview or EMA was estimated as follows: EMA weekly binges versus self-reported weekly binges; EMA weekly binges versus OBE according to EDE; abstainer rates based on EMA versus abstainer rates according to self-reported weekly binges and abstainer rates based on EMA versus abstainer rates according to EDE.

Factors Influencing Concordance of Traditional and EMA-Based Measures. We analyzed to what degree eating disorder pathology (General Score of the EDE, EDE-Q), negative affect according to the BDI,36 mental disorders according to the DIPS, and BMI influenced concordances defined above.

Treatment. The CBT protocol for BED consisted of eight weekly 90-min group sessions in the active treatment phase.37 The manual in our study was a shortened version of the 16-sessions group CBT program for BED developed by Munsch,38 which has proven to be successful for patients with BED.35

Statistical Analyses

Because IT and WL group were not significantly different in the changes from pre- to post-treatment (no treatment × time interaction) we pooled the two groups to enhance power for all pre- versus post-treatment analyses.

Statistical Models. Continuously distributed measures were analyzed using a linear mixed model.39 For the analysis of the number of weekly binges (EMA-based, based on traditional self-report measures and according to EDE), we applied a generalized linear mixed model (GLMM) for Poisson-distributed data.40 To compare the abstainer rates at post-treatment/end of waiting period we used a $\chi^2$ test. To assess the changes in abstainer rates between pre- and post-treatment for pooled groups, we used a logistic-normal model that is a special case of GLMM in which the random effects structure is reduced to a random intercept only.40

Effect sizes between pre- and post-treatment are

\[ \text{Effect size} = \frac{\text{Mean pre-treatment} - \text{Mean post-treatment}}{\text{Standard deviation}} \]

Factors Influencing Concordance Between Traditional and EMA-Based Measures. Using a linear regression model we analyzed whether concordance between traditional and EMA-based measures was influenced by BMI, EDE total score, BDI, and comorbidity status (presence/absence). For these analyses, concordances between the assessment methods regarding weekly binge eating episodes were computed using the difference between the $z$-scores of the corresponding methods. As abstainer rates are based on dichotomous data, differences between corresponding methods were analyzed using a logistic regression model. In total, we tested 32 models comprised of four dependent variables (weekly binges and abstainer rates in each of two method comparisons) × two time points (pre- and post-treatment) × four predictor variables.

Results

Dropouts

Of the 26 participants starting treatment, seven (26.9%) dropped out during treatment, three (11.5%) in the IT, and four (15.4%) in the WL group ($\chi^2(1) = 1.5, p = .22$). We checked whether dropout was completely at random (i.e., MCAR pattern sensu Little and Rubin42), using $t$-tests for continuously distributed data, or Pearson’s $\chi^2$ tests for dichotomous variables for the broad range of variables. After correction for multiple comparisons, only employment status had increased the risk of dropping out of the study. Patients who were still in education or unemployed were more likely to drop out ($\chi^2(4) = 17.7, p = .001$).

Feasibility of and Reactivity to EMA

Acceptability of the electronic diary according to the EXQ was high and increased significantly between pre- and post-treatment (from 7.17 to 7.99, $p = .016$, $d = 0.34$). EXQ-based estimates of practicability and representativeness were high and remained stable (practicability: 2.00 and 2.55, $p = .58$, $d = 0.19$; representativeness: 7.17 and 6.85, $p = .78$, $d = -0.11$). EXQ-based signal-compliance significantly decreased between pre- and post-treatment (from 7.07 to 5.08; $p = .037$, $d = -0.33$) whereas EMA-based signal-compliance remained constant during that period: mean proportions of

International Journal of Eating Disorders 00:0 000–000 2009
the number of recordings starting within 30 min after the signal to the total number of recordings were 0.87 for both pre- and post-treatment. Participants' mean absolute deviations of entry from the scheduled alarms (in min) were 16.0 (median = 1.0, SD = 34.8, N = 722) before and 15.3 (median = 1.0, SD = 33.2, N = 427) after treatment, respectively. With altogether 28 patients with BED filling in the diary five times a day during 2 or 3 weeks a total of 2,485 time-contingent data entries were possible. Participants completed 1,667 of these time-contingent entries, corresponding to an overall recording compliance of 68%.

EXQ-based reactivity was moderate and remained stable (5.77 and 5.73; \( p = .95, d = -.01 \)). No indications of reactivity due to EMA emerged as the number of daily binges remained stable whereas waiting during one week before treatment start (non-significant linear temporal trend, \( p = .62 \), both groups pooled). There was also no change in the number of weekly binges between the two weeks prior to treatment in the WL group (\( p = .11 \)).

### Concordance Between EMA- and Traditional Questionnaire/Interview-Based Outcome Measures and Estimates of Signal-Compliance

Concordances between EMA-based and EXQ-based signal-compliance were significantly different from zero for pre- but not for post-treatment (Table 2). Highest concordance rates for the comparison of the core symptomatology of BED were obtained between EMA-based and self-reported weekly binge eating episodes and the corresponding abstainer rates, especially at post-treatment (Table 2). Concordance between OBE assessed by

| Table 2. Concordance between measures based on traditional interviews and questionnaires and EMA-based measures for signal compliance and objective binge eating |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                 | Pre-treatment   | Post-treatment  |                 |                 |
|                                 | \( n \) | \( r \) | ICC (C,1) | ICC (A,1) | \( n \) | \( r \) | ICC (C,1) | ICC (A,1) |
| Signal-compliance: EMA-based versus EXQ-based | 17 | .52* | 0.52* | 0.53* | 12 | .36 | 0.36 | 0.38 |
| Weekly number of binges: EMA-based versus self-report questionnaire | 24 | .46* | 0.44* | 0.45* | 18 | .80* | 0.64* | 0.64* |
| Weekly number of binges: EMA-based versus OBE according to EDE | 20 | .13 | 0.12 | 0.12 | 17, 15b | .78,* | .42b | 0.78, 0.40b | 0.79,* 0.42b |
| Weekly abstainer rate: EMA-based versus self-report questionnaire | 24 | .22c | 0.22c | 0.22c | 18 | 0.50a* | 0.50a* | 0.50a* |
| Weekly abstainer rate: EMA-based versus according to EDE | 25 | —d | —d | —d | 20 | 0.48b* | 0.48b* | 0.48b* |

ICC, intraclass correlation coefficient for consistency (C,1) and absolute agreement (A,1) according to McGraw and Wong\(^{31}\); \( r \), Pearson correlation coefficient; OBE: objective binge eating episodes according to Eating Disorder Examination, EDE; EXQ, exit questionnaire (unpublished questionnaire, available by the authors).

*Values were obtained by dividing the number of OBEs by four to make them comparable to the number of EMA-based and self-reported weekly binges.

*Values after removal of two extreme values.

*Value denotes Cohen’s kappa.

* Cohen’s kappa could not be computed as abstainer rate based on EDE at pre-treatment was equal to zero.

* \( p < 0.05 \).
International Journal of Eating Disorders 00:0 000–000 2009

Patients with BED reported significantly lower signal-compliance on the questionnaire after treatment, yet our EMA data indicate that their objective signal compliance behavior was still high and unchanged after the treatment. We found moderate concordance between the two compliance-assessment methods before and low concordance after treatment, probably due to the spuriously decreased compliance self-estimates after treatment. Because electronic diaries objectively monitor compliance behavior, the observed discrepancy between the momentary and retrospective data might indicate inferiority in the accuracy of retrospective reports in this regard.

We examined both retrospective and momentary measures of reactivity, that is, to what extent filling in the electronic questionnaire influenced the patterns of binge eating itself. Reactivity was retrospectively reported as being moderate, and this perception remained stable between pre- and post-treatment. Further, we did not find indications of reactivity in terms of a possible weekly trend in the EMA-based number of binge eating episodes prior to treatment. This result is in line with the findings from eating disorder and other clinical populations, and indicates that the problem of behavioral reactivity to self-monitoring should not be overestimated.

From a theoretical point of view, an important question addressed in our study is how electronic diary assessment of pre- to post-treatment effects compares to the traditional questionnaire and interview assessment. The numbers of weekly binge eating episodes assessed by EMA or traditional self-report questionnaire were strongly and similarly reduced between pre- and post-treatment, resulting in comparable effect sizes. With respect to abstainer rates between pre- and post-treatment we further found only small differences between EMA-based (increase from 39 to 70%) and self-report questionnaire-based (increase from 21 to 63%) effects. It seems likely that the two methods, retrospectively reporting binge eating over the past seven days and immediate assessment by EMA, both result in comparable data accuracy as it has been found in a study with pain patients by Jamison et al.

It needs to be investigated whether the observed small differences in outcomes between the two assessment methods become more pronounced when self-report questionnaires cover longer retrospective time spans than seven days. EDE-based effects of abstainer rates were somewhat lower (increase from 0 to 54%), possibly due to retrospective recall error. The ability to retain the characteristics of a complex behavior such as binge eating in memory for more than a...
few days is limited, and therefore patients are likely to forget or distort some of the information necessary to evaluate a binge eating episode correctly. This limitation might especially apply to the EDE, demanding a recall period of 28 days.

Measures of concordance ranged between low and fair to good \(^{50}\) and varied considerably between pre- and post-treatment and also among pairs of outcome measures. The correlation coefficients were lower when compared with other studies \(^{22}\) and also lower than in the Stein and Corte \(^{18}\) study with patients with subthreshold anorexia and bulimia nervosa. As the treatment turned out to be highly efficacious, the low rates of binge eating at post-treatment did not allow us to accurately identify concordance of the different outcome measures (floor effect). Thus, these findings must be interpreted with caution.

Contrary to our expectations, the concordances between EMA-based and traditional questionnaire/interview-based measures hardly depended on patients’ comorbidity status or eating disorder pathology. Yet our findings indicate that higher baseline depressiveness is associated with a higher frequency of EMA-based binge eating episodes but not of binge eating episodes assessed by traditional self-report questionnaires. Further, we showed that a higher BMI leads to higher frequencies of self-report questionnaire-based binge eating but does not influence the frequency of binge eating assessed by EMA. These results should be discussed with caution as only two out of 32 statistical tests addressing this topic were statistically significant. Thus, we cannot rule out that these two results were merely chance findings.

In interpreting our findings one should bear in mind that our study is subject to several limitations. First, although we found no indication of reactivity effects to EMA, to definitely evaluate this possible problem, future studies should control for the effects of waiting and EMA separately. A second major limitation concerns the differences in the duration of data assessment between EMA and EDE. As estimates based on EDE refer to a 28-days period we had to divide measures indicating the numbers of binge eating episodes by four to make them comparable to our measures based on a 1-week period. A presupposition for this is that binge eating episodes across the previous 28 and 7 days’ periods are reasonably highly correlated, which may or may not be true. Even though patient burden will be high, future studies should compare the concordance and treatment effects of these two methods covering the required 28 days according to EDE in order to better account for the known fluctuating nature of binge eating in BED. It would be beneficial to include direct observational data on behavior, rather than comparing two self-reported data sources. This would allow establishing the superiority of one against the other method more definitively. In a recently developed ambulatory assessment method, the electronically activated recorder \(^{51}\) may provide direct acoustic evidence of binge eating in real life. Third, as is often the case in clinical trials, our study included only a moderate sample size thereby limiting the statistical power of the hypotheses tested.

In summary, this study in patients with BED showed the electronic diary assessment to be highly acceptable and practicable, leading to high compliance rates. We found comparable treatment effects for binge eating using EMA when compared with traditional self-report questionnaires but only moderate concordance rates. Though preliminary, our results point to possible limitations of retrospective recall of complex behavior such as binge eating demonstrated previously \(^{52}\) over a sustained time period as required in the EDE.

Most importantly, our data demonstrate that the EMA method allows gathering precise and detailed information about binge eating episodes in the naturalistic environment during a treatment trial. The advantages of this assessment method in psychotherapy research include the assessment of momentary and time-stamped data, the potential to ask branching questions, the possibility to collect multiple data in a short time period, and the instant accessibility for data analysis. Awareness about the specific characteristics of BED in daily live can be utilized in the planning of individualized interventions for specific patients. However, further efforts are needed to establish the feasibility of the EMA method outside specialized research units.

References


