Long-term efficacy of psychological treatments for binge eating disorder

Anja Hilbert, Monica E. Bishop, Richard I. Stein, Marian Tanofsky-Kraff, Anne K. Swenson, R. Robinson Welch and Denise E. Wilfley

Background
The long-term efficacy of psychological treatments for binge eating disorder remains largely unknown.

Aims
To examine the long-term efficacy of out-patient group cognitive–behavioural therapy (CBT) and group interpersonal psychotherapy (IPT) for binge eating disorder and to analyse predictors of long-term non-response.

Method
Ninety people with binge eating disorder were assessed 4 years after treatment cessation within a randomised trial (trial registration: NCT01208272).

Results
Participants showed substantial long-term recovery, partial remission, clinically significant improvement and significant reductions in associated psychopathology, despite relapse tendencies in single secondary outcomes. Body mass index remained stable. While the IPT group demonstrated an improvement in eating disorder symptoms over the follow-up period, the CBT group reported a worsening of symptoms, but treatments did not differ at any time point.

Conclusions
The results document the long-term efficacy of out-patient CBT and IPT for binge eating disorder. Further research is warranted to elucidate the time course and mechanisms of change of these treatments for binge eating disorder.

Declaration of interest
None.

Participants and procedure
Participants were the first 90 patients (55.6%; the first five cohorts) out of a larger (n = 162) treatment trial for overweight people with binge eating disorder, recruited in New Haven, Connecticut, USA (for methodological detail see the main report). As this long-term follow-up study was not part of the initial two-site clinical trial, funded by the National Institutes of Health (NIH), participants were re-consented with a separate informed consent approved by the San Diego State University Institutional Review Board. Owing to the unfunded nature of this follow-up study, we opted to only enrol patients from the New Haven site. All participants met DSM-IV-TR diagnostic criteria for binge eating disorder and were, after stratification by gender, randomised to either CBT or IPT (n = 45 each). Both treatments were manual-based and consisted of 20 weekly 90-min group sessions and three additional individual sessions.

For the present study, patients were contacted for a long-term follow-up assessment approximately 4 years after treatment cessation (i.e. a mean of 46.0 months). As illustrated in Fig. 1, of the 90 patients included in this study, 77 (85.6%) could be contacted (i.e. no current contact information; 10; incapacitated or deceased: 3). Of the 77 individuals contacted, 58 (75.3%) completed the long-term follow-up assessment. Completer rates for post-treatment assessment were 89/90 (98.9%) and for the 1-year follow-up assessment 77/90 (85.6%). Eighty-three (92.2%) of the 90 patients had completed treatment. Treatment and assessment completion rates did not differ by group (all P > 0.05).

Assessments and procedures
Outcome analyses included pre-treatment, post-treatment, 1-year follow-up and long-term follow-up assessments. The long-term follow-up assessment included a telephone interview and self-report questionnaires, whereas all other assessments involved in-person diagnostic visits and self-report questionnaires. All structured clinical interviews were conducted by trained assessors (Bachelor level or higher) who received ongoing supervision to ensure standardised administration.

Eating disorder psychopathology
The diagnostic version of the Eating Disorder Examination (EDE 12.0D), a semi-structured interview with good reliability and validity, was used to assess days with objective bulimic episodes (i.e. eating an unusually large amount of food, accompanied by...
a sense of loss of control over eating). The primary outcome variables of recovered and remitted were based on the EDE assessment of days with objective bulimic episodes over the previous 28 days (see Data analysis). Secondary outcome variables included the number of days with objective bulimic episodes over the previous 28 days and a composite shape/weight concern score, both derived from the EDE. For descriptive purposes, DSM-IV-TR diagnoses of binge eating disorder, bulimia nervosa, anorexia nervosa and eating disorder not otherwise specified were determined with the abbreviated EDE. Objective bulimic episode days and episodes were comprehensively assessed over the previous 6 months, whereas purging behaviour (i.e. self-induced vomiting, laxative or diuretic misuse), fasting, intense exercising, importance of shape or weight, fear of weight gain, feelings of fation, maintained low weight, and menstruation were assessed over the previous 28 days and, if positive, over the previous 3 months.

In addition, the self-report version of the EDE, the Eating Disorder Examination Questionnaire (EDE-Q),13 was used to assess the associated eating disorder psychopathology on the four subscales of restraint, eating concern, shape concern and weight concern, and on global eating disorder psychopathology. Subscale scores were used as secondary outcomes; they ranged from 0 to 4, with higher scores indicating more severe psychiatric symptoms. Both subscales have adequate internal consistency, convergent validity and sensitivity to change.14 The EDE-Q weight/shape composite and the EDE-Q internal consistency, convergent validity and sensitivity to change.16

Body mass index
Body mass index (BMI, kg/m²), a proxy for body fat, was used as a secondary outcome. Body mass index was calculated from weight (self-reported at long-term follow-up and measured at all other time points) and height (measured at pre-treatment). For diagnosis of anorexia nervosa (see above), underweight was determined as BMI < 18.5 kg/m². Prior to treatment, measured and self-reported body weight were highly associated (r = 0.99, P < 0.001).

Healthcare utilisation
At long-term follow-up, healthcare utilisation for eating or weight problems was retrospectively assessed for the time after 1-year follow-up. Use of psychotherapy, pharmacotherapy, consultation (for example with a dietician) and/or alternative treatment (for example hypnosis) and any of these treatments was determined in a dichotomous format (0, no; 1, yes).

Data analysis
Preliminary analyses served (a) to compare patients randomised to CBT v. IPT (n = 45 each) on sociodemographic characteristics (age, gender), drop-out from treatment, adherence (i.e. number of sessions attended), and pre-treatment level of any outcome variable (univariate general linear model (GLM) analyses or χ² analyses by treatment (CBT, IPT)); (b) to compare participants who were included in the long-term follow-up assessment v. those who were not included (n = 90 v. 72) by treatment on the same variables and outcome variables at all time points (univariate GLM or generalised linear model analyses of sample (included, not-included) × treatment (CBT, IPT) for continuous or binary variables, respectively); and (c) to compare participants with completed v. non-completed long-term follow-up assessment (n = 58 v. 32) by treatment on these same variables and outcome variables at pre-treatment, post-treatment, and 1-year follow-up (univariate GLM or generalised linear model analyses of assessment (completed, not-completed) × treatment (CBT, IPT) for continuous or binary variables, respectively).

Statistical analyses were based on the generalised estimating equations (GEE) approach for the primary (categorical) outcome variables and on hierarchical linear modelling (HLM) for the
secondary (continuous) outcome variables. Both intention-to-treat approaches allow data from participants with missing data at some, but not all, time points to remain in the analyses. In addition, both approaches correct for the dependency of observations within participants, in GEE analysis by assuming an exchangeable working correlation structure, and in HLM by allowing the regression coefficients to vary between participants. For the efficacy analyses, the assessment completer sample size at long-term follow-up provided 80% power to detect a medium-to-large effect size of treatment difference ($d = 0.76$; $n = 25$ for CBT, $n = 33$ for IPT).

The primary categorical outcome variables were analysed using GEE logistic regression models (i.e. logit link function and binomial error distribution) that included treatment (CBT, IPT) $\times$ time (post-treatment, 1-year follow-up, long-term follow-up) and the respective main effects as predictors. Least significant difference tests were used for post hoc analyses in case of significant higher-order effects. Consistent with the main outcome report, the primary outcome included three variables, all determined at post-treatment, 1-year follow-up and long-term follow-up: recovered (i.e. no objective bulimic episodes in the previous month), improved to subclinical binge eating (i.e. $< 4$ objective bulimic episode days in the previous month) and being at or below a comparative level of eating disorder attitudes and behaviours. The latter rating was made based on whether the global EDE-Q score was at or below the global EDE-Q score of overweight treatment-seeking individuals without binge eating disorder who had a similar sociodemographic profile as the patients in the current study (global EDE-Q score, 2.47).17

The secondary continuous outcome variables were analysed using HLM of treatment (CBT, IPT) $\times$ time (pre-treatment, post-treatment, 1-year follow-up, long-term follow-up), with patients nested within time. In this analysis, time and treatment were treated as fixed factors and patients as a random factor. Least significant difference tests were used for post hoc analyses.

To ensure that the results from the primary and secondary intention-to-treat analyses were robust, two sensitivity analyses18 were conducted in addition, handling missing data as follows: (a) with missing data multiply imputed, creating five completed data-sets via an iterative Markov chain Monte Carlo method; (b) with last observations carried forward; and (c) with no replacement values for missing data (i.e. complete analysis only including participants who had completed all assessments). Outcome analyses were performed on all sensitivity data-sets for inspection of significance. Results were reported only if different from the primary and secondary intention-to-treat analyses described above. Completer effect sizes were computed as Pearson’s $r$ (small, $r \geq 0.10$; medium, $r \geq 0.30$; large, $r \geq 0.50$) or Cohen’s $d$ (small, $d \geq 0.20$; medium, $d \geq 0.50$; large, $d \geq 0.80$).19

All analyses were performed with PASW 18.0 for Windows. A two-tailed significance level of $\alpha < 0.05$ was applied to all statistical tests. In order to avoid $\alpha$ inflation by multiple testing, the significance level was adjusted to a two-tailed $\alpha < 0.01$ for post hoc tests.

### Results

#### Randomisation, attrition and sampling

Patients randomised to CBT v. IPT did not differ with regard to sociodemographic characteristics (Table 1) or any pre-treatment primary or secondary outcome variable (all $P > 0.05$). There were no significant differences in drop-out (CBT: 3 (6.7%); IPT: 4 (8.9%)) or adherence to treatment (CBT: 16.8 (s.d. = 3.0); IPT: 17.7 (s.d. = 3.9) sessions; all $P > 0.05$).

The 90 patients who were included in the long-term follow-up assessment did not significantly differ from those who were not included in the follow-up assessment ($n = 72$) on pre-treatment characteristics, drop-out and adherence, and no interaction effects with treatment condition were observed (sample (included, not-included) $\times$ treatment (CBT, IPT); all $P > 0.05$). The 58 patients who completed the long-term follow-up assessment did not significantly differ from the 32 assessment non-completers on sociodemographic characteristics, drop-out, adherence, missing data, or primary and secondary outcomes at pre-treatment, post-treatment, or 1-year follow-up, and no interaction effects with treatment condition occurred (assessment (completed, not-completed) $\times$ treatment (CBT, IPT); all $P > 0.05$).

#### Primary outcomes

Long-term recovery rates were 52.0% for the CBT group and 76.7% for the IPT group (Table 2, Fig. 2). The GEE logistic regression analysis showed a significant treatment $\times$ time effect on abstinence from binge eating ($P < 0.001$). Post hoc comparisons by time point did not reveal any significant between-treatment differences (all $P > 0.01$). However, for CBT, a significant decline in recovery rates from post-treatment and 1-year follow-up to long-term follow-up was observed (both $P \geq 0.002$), whereas for IPT, abstinence rates did not change over the follow-up period (both $P > 0.01$).

Long-term rates of remission to subclinical binge eating were 72.0% for the CBT group and 83.9% for the IPT group. The GEE logistic regression analysis of treatment $\times$ time did not show any significant effects on remission (all $P > 0.05$). Long-term rates of improvement to a comparative level of eating disorder attitudes and behaviours (i.e. global EDE-Q score $\leq 2.47$) were 54.5% for the CBT group and 61.5% for the IPT group. The GEE logistic regression analysis showed a significant time effect on

---

**Table 1** Pre-treatment sample characteristics ($n = 90$)

<table>
<thead>
<tr>
<th></th>
<th>Cognitive-behavioural therapy ($n = 45$)</th>
<th>Interpersonal psychotherapy ($n = 45$)</th>
<th>Test statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years: mean (s.d.)</td>
<td>45.73 (9.86)</td>
<td>44.02 (10.49)</td>
<td>$F (d.f.) = 0.64$ (1,88) $P = 0.427$</td>
</tr>
<tr>
<td>Age at onset of disorder, years: mean (s.d.)</td>
<td>17.50 (11.76)</td>
<td>18.50 (10.20)</td>
<td>$\chi^2 (n = 90) = 0.17 (1,78) P = 0.686$</td>
</tr>
<tr>
<td>Gender, female: n (%)</td>
<td>36 (80.0)</td>
<td>35 (77.8)</td>
<td>0.07</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td>0.21</td>
</tr>
<tr>
<td>White</td>
<td>41 (91.1)</td>
<td>42 (93.3)</td>
<td>0.21</td>
</tr>
<tr>
<td>African American</td>
<td>3 (6.7)</td>
<td>2 (4.4)</td>
<td>0.21</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 (2.2)</td>
<td>1 (2.2)</td>
<td>0.21</td>
</tr>
<tr>
<td>Comorbid psychiatric diagnoses, n (%)</td>
<td>19 (42.2)</td>
<td>18 (40.0)</td>
<td>0.05</td>
</tr>
<tr>
<td>Any Axis I disorder, current</td>
<td>4 (8.9)</td>
<td>5 (11.1)</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Any Axis I disorder 4 (8.9) 5 (11.1) 0.12 0.830

Any Axis II disorder 4 (8.9) 5 (11.1) 4 (8.9) 0.12 0.725
improvement \((P = 0.049)\), but post hoc analyses did not reveal any change over the follow-up period (all \(P > 0.01\)).

For all primary outcome variables, the sensitivity analysis revealed the same results with no replacement of missing data and with last observation carried forward. Multiple imputation results were largely consistent, with two exceptions: there was a significant time effect for remission and a significant treatment effect for clinically significant improvement (both \(P < 0.05\)). Between-treatment effect sizes at long-term follow-up were small (recovery, \(r = 0.26\); remission, \(r = 0.07\)).

Descriptive completer analyses revealed that 12 (27.3%) CBT patients and 10 (22.2%) IPT patients showed persistent recovery across all three follow-up assessments. From 1-year follow-up to long-term follow-up: 13 (52.0%) patients in the CBT and 13 (43.3%) in the IPT group maintained abstinence; no one (0.0%) in the CBT and 10 patients (33.3%) in the IPT group achieved abstinence; 6 (24.0%) patients in the CBT and 3 (10.0%) in the IPT group relapsed; and 6 (24.0%) patients in the CBT and 4 (13.3%) in the IPT group remained non-abstinent from objective bulimic episodes.

At long-term follow-up, 3 (12.0%) individuals in the CBT group and 3 (9.4%) in the IPT group were diagnosed with binge eating disorder. None of the participants were diagnosed with anorexia nervosa or bulimia nervosa or eating disorder not otherwise specified, and none revealed purging behaviour, fasting or intense exercising.

### Secondary outcomes

The HLM analyses of treatment \(\times\) time revealed significant time effects for all secondary outcome variables (all \(P < 0.001\)), except BMI \((P = 0.433);\) online Table DS1. Post hoc analyses showed significant improvements in the number of days with objective bulimic episodes and EDE shape/weight concern, all EDE-Q indicators (except restraint) and depression at post-treatment, 1-year follow-up and long-term follow-up when compared with pre-treatment (all \(P < 0.01\)). In contrast, the significant improvements of EDE-Q restraint and BSI anxiety present at post-treatment and 1-year follow-up when compared with pre-treatment (both \(P < 0.01\)) were no longer present at long-term follow-up (\(P > 0.01\)). Concerning the course over the follow-up period, the number of days with objective bulimic episodes and anxiety increased from post-treatment or 1-year follow-up to long-term follow-up (both \(P < 0.001\)), whereas the reduced levels of EDE shape/weight concern, all EDE-Q indicators and depression were maintained from post-treatment or 1-year follow-up to long-term follow-up (all \(P > 0.01\)).

The course of EDE-Q eating concern, shape concern and the global eating disorder psychopathology, and of EDE shape/weight concern showed significant interactions with treatment (all \(P \leq 0.03\)): for the CBT group, EDE-Q eating concern and EDE shape/weight concern worsened from 1-year follow-up to long-term follow-up (all post hoc \(P < 0.01\)), whereas for IPT, there was an improvement in EDE-Q eating concern, shape concern and global eating disorder psychopathology from post-treatment to long-term follow-up (\(P < 0.01\)).

The sensitivity analysis revealed largely consistent results with no replacement of missing data, last observations carried forward and multiple imputation. The only difference was that for EDE-Q shape concern, the interaction effect was no longer significant with multiple imputation \((P < 0.05)\). Most effect sizes between pre-treatment and long-term follow-up for eating disorder psychopathology were large \((0.97 \leq d \leq 2.10\)). Depression yielded a medium effect size \((d = 0.53)\), and EDE-Q restraint, anxiety and BMI yielded small effect sizes \((d \leq 0.37)\). Between-treatment
Hilbert et al

Abstinence from binge eating was stable over the follow-up period, although symptoms over time. The subset of patients showing reoccurrence of eating disorder symptoms appears essential for early identification of treatment-seeking following CBT and IPT, careful monitoring of long-term outcome and in light of high rates of additional psychopathology showed tendencies toward relapse. Despite the overall favourable outcomes, including objective bulimic episode days (the key symptom of binge eating disorder), restraint, and anxiety, showed tendencies toward relapse. Despite the overall favourable long-term outcome and in light of high rates of additional treatment-seeking following CBT and IPT, careful monitoring of these symptoms appears essential for early identification of the subset of patients showing reoccurrence of eating disorder symptoms over time.

There was some indication that CBT and IPT involved a differential time course over the follow-up period, although treatments did not differ in recovery rates at any time point. Abstinence from binge eating was stable over the follow-up period in the IPT group, whereas there was a significant tendency to relapse among patients in the CBT group. Concomitantly, reduction of eating disorder psychopathology in the IPT group was better maintained or further improved over the follow-up period, whereas for the CBT group psychopathology worsened from 1-year follow-up to long-term follow-up. This differential time course is similar to the ‘catching up’ effect of IPT that has previously been found in bulimia nervosa treatment. In two studies, IPT emerged as inferior to CBT at the end of treatment, but IPT patients showed continued improvement at 1-year follow-up, levelling off between-treatment differences. Based on the underlying theories of CBT and IPT, one might speculate that the focus on improving interpersonal relationships prepares individuals more comprehensively for the social challenges of daily life than the more rapidly acting, more eating disorder-focused CBT treatment, which reaches its efficacy earlier. Further research is warranted to clarify mechanisms of action of CBT versus IPT for binge eating disorder.

Body mass index

Body mass index was stable throughout the follow-up period, suggesting that the course of weight gain that is characteristic of treatment-seeking individuals with binge eating disorder has sustainably been interrupted. Given a tendency among adults to gain 400 g of weight per year, a stabilisation of body weight as documented in the current study may have led to a small amount of decreased weight gain of approximately 1600 g over the assessment period. Stabilisation of body weight as documented in the current study is a positive finding, in and of itself, as it is one of the priorities of obesity prevention.

Methodological considerations

Strengths of this study include the conduct of an additional long-term follow-up addressing a fairly large subsample of patients (55.6%) of a well-controlled clinical trial on binge eating disorder. Although the long-term assessment was not pre-planned and was unfunded, the overall participation rate of 64.4% was adequate for this study’s sample (corresponding to 35.8% of the initial study sample that, however, was not considered for participation in this study).

Regarding generalisability of findings, several aspects underscore the certainty of results. No measurable biases existed in sample selection and assessment completion regarding sociodemographic characteristics, treatment drop-out, adherence
and any outcome from pre-treatment to 1-year follow-up. In order to account for missing data, intention-to-treat analyses allowing data from assessment non-completers to remain in the analyses were conducted. The results from these intention-to-treat analyses showed the same pattern of results when compared with different treatment of missing data. The current completers sample size generated sufficient power to detect a medium-to-large effect size of treatment difference at long-term follow-up; effect sizes were reported in order to reflect smaller size differences. Of the patients who were or could have been consented for this study, 73.3% completed long-term assessment. Likely related to the longer time interval, assessment completion rates were gradually lower from post-treatment to 1-year follow-up to long-term follow-up.

A further limitation to this study is that although we collected data on interim treatments, their effects could not be separated out from long-term effects of CBT and IPT themselves. Other intervening factors (for example life events) were not controlled for. Finally, the study sample was mostly female and White, which limits generalisation of results to male and other ethnic groups.

Implications
This report on the long-term efficacy of two major out-patient treatments for binge eating disorder suggests that IPT is a viable treatment alternative to standard CBT. Both treatments yielded high rates of treatment response and long-term maintenance of therapeutic gains. To bolster this study’s findings, replication of such outcomes in larger samples and over a longer time period is warranted. Alternative evidence-based treatment options, such as more individualised or comprehensive treatment, or extended or additional treatment, should be considered for individuals with poor initial treatment response to further optimise long-term treatment effects of CBT and IPT.

References