Online Supplement: Methods

Participants

The randomized controlled trial was conducted at two eating disorder centers, the Center of Excellence for Eating Disorders (CEED) at the University of North Carolina at Chapel Hill and the Center for Overcoming Problem Eating (COPE) at Western Psychiatric Institute and Clinic, University of Pittsburgh Medical Center. The study included advertisements in university listservs, print, radio, and social media platforms (such as Facebook and Twitter), and brochures at local counseling centers, physician offices, and mental health organizations, particularly those that serve minority groups.

Inclusion criteria were Diagnostic and Statistical Manual-Fourth Edition (DSM-IV) diagnosis of bulimia nervosa (BN) [16]; age 18 years or older; body mass index (BMI) ≥ 18.5 kg/m^2; English speaking; and private access to the Internet. Patients were permitted to take psychotropic medications and were asked not to modify medications during the clinical trial. Patients did not participate in other forms of psychotherapy (e.g., individual, couples, or other group therapy). Exclusion criteria included any major medical condition that would interfere with treatment (e.g., type 1 diabetes mellitus; *N* = 6); alcohol or drug dependence in the last three months (*N* = 11); psychosis, including schizophrenia, and bipolar I disorder, or current significant suicidal ideation reported during the clinical assessment (*N* = 10).

Treatment

This form of group cognitive-behavioral therapy (CBT) for BN had been found previously to be an effective treatment (14). In an accelerated version of the treatment that included the same manual delivered over 8 sessions, 28% of patients were completely abstinent from binge eating and purging by the end of the intervention (14).

Patients in CBT4BN groups convened with the therapist via an online chat group. Each patient logged into the website using an anonymous username and a password. These private usernames were then used to identify the patient during the session. The chat room was text-based only; messages could include emoticons (e.g.,

☺) and personalized font size and color. Chats were arranged at a pre-determined time and the chat "room" was only open for the arranged 90-minute period; a message indicating that the chat room was closed was added to the screen at the end of the session. Patients completed weekly assessments and daily self-monitoring, therapy module readings, and homework via the CBT4BN portal. Therapists provided online written feedback to self-monitoring and homework exercises on pre-determined days between therapy sessions.

Patients in CBTF2F met the therapist and group members face-to-face for each session and completed weekly assessments and daily self-monitoring via pencil and paper. Patients were given a paper copy of the treatment module to read prior to their next session and completed all therapy homework worksheets on paper. Therapists provided written feedback on paper self-monitoring and homework exercises in between therapy sessions and returned these comments to the participants at their next session.

All therapists (13 total: 9 psychologists and 4 social workers) had clinical experience with eating disorders treatment, received specific training in the treatment manual used in the trial, and treated patients in both conditions. Study investigators (authors: C.B., M.M., M.L., and S.Z.) led weekly supervision meetings. CBTF2F sessions were audio-recorded and reviewed by supervisors. CBT4BN chat transcripts were also reviewed by supervisors.

Independent raters were trained to code digital audio files and chat transcripts of treatment sessions for therapist adherence to the manual. A total of 81 sessions, 7.6% (44 sessions) of CBT4BN chat transcripts and 6.4% (37 sessions) of audio CBTF2F sessions were coded. The average adherence rating was 6.2 for CBT4BN and 6.6 for CBTF2F on a 7-point rating scale (1 = not true at all; 7 = very true).

In CBTF2F, all participants were monitored for safety via a paper-and-pencil safety questionnaire completed after each session. In CBT4BN, participants completed the same safety questionnaire online. In addition, in CBT4BN participants were called every other week by the study psychiatrist to ensure safety (i.e., 3-5 minute calls to assess suicidality).

Assessment

Eating Disorder Symptoms. The *Eating Disorder Examination* interview (EDE) was administered at baseline, end of treatment, and 12-month followup by assessors blind to the patient's treatment condition [17]. The primary outcome variable was abstinence from binge eating and purging (0 episodes over the previous 28 days). Secondary outcome variables included BN diagnosis and the frequency of binge and purge episodes and tertiary outcomes included scores on the EDE global score. Independent raters coded digital audio files of the EDE and Structured Clinical Interview for DSM-IV (SCID-I/P) interviews for inter-rater reliability. The inter-rater reliability for abstinence was perfect (Krippendorff's $\alpha = 1.00$) [18]. Inter-rater reliability was high for frequency of binge eating [intraclass correlation (ICC) = 0.91], purging (ICC = 0.89), BN diagnosis (Krippendorff's $\alpha = 1.00$), and the global EDE score (ICC = 0.98).

Comorbid Psychopathology. Tertiary outcomes measured by the SCID-I/P [19] included the presence of major depressive disorder or an anxiety disorder (i.e., social phobia, generalized anxiety disorder, and specific phobia) at each time point. Inter-rater reliabilities were: SCID major depression diagnosis (Krippendorff's $\alpha = 0.47$), SCID anxiety disorder diagnosis (Krippendorff's $\alpha = 0.90$). Depression and anxiety severity were measured by the Beck Depression Inventory (BDI) and Beck Anxiety Inventory (BAI), 21-item self-report questionnaires [20, 21]. Both the BDI and BAI had good internal consistency ($\alpha = 0.92$ and $\alpha = 0.91$, respectively in the current study).

Quality of Life. Quality of life was measured by the Eating Disorders Quality of Life Questionnaire (EDQOL) and the Short-Form Health State Classification (SF-6D) [22, 23]. The EDQOL is a 26-item health-related quality of life questionnaire that was designed specifically for use with eating disorder patients, and has excellent test-retest reliability and convergent validity. Internal consistency was good ($\alpha = 0.90$ in the current study). The SF-6D is a utility-based measure of health-related quality of life and is derived from a longer version, the Short-Form (36) Health Survey (SF-36) [23].

Treatment Evaluation

Treatment Preference and Evaluation. Participants recorded their preference for treatment group at baseline (CBT4BN vs. CBTF2F). Prior to treatment, credibility (i.e., how logical the proposed treatment appeared) and expectancy (i.e., how confident participants were that treatment would succeed) were assessed with the Client Satisfaction Questionnaire (CSQ) based on Borkovec and Nau [24]. At the end of treatment, they completed a 6-item self-report measure designed for this study that assessed their satisfaction with treatment and whether they would recommend the assigned treatment to others (Likert scale, 1 = not at all to 10 = extremely).

Post-Treatment Service Utilization. Post-treatment care utilization was assessed using the McKnight Follow-up of Eating Disorders (MFED) [25]. The MFED is a structured interview designed to provide measures of eating disorder symptoms and medical and mental health care utilization. Only the health care utilization subscale was used for this study. Using the MFED, participants were interviewed at 3-, 6-, and 12-month follow-up and reported whether they had psychotherapy for their eating disorder or had taken any psychotropic medications during the post-treatment period.

Adverse Events

There were three adverse events reported during the study. All three participants were randomized to CBTF2F prior to the event. One patient dropped out of treatment but was later hospitalized due to a new diagnosis of bipolar disorder. A second patient was hospitalized for suicidal ideation with a plan after the treatment period had ended but during the follow-up period. A third patient was hospitalized due to hyponatremia during the treatment trial. All adverse events were reported to the IRB and the Data Safety Monitoring Board for the study.

Randomization, Power, and Statistical Analyses

Eligible patients were randomly assigned, according to a central computerized randomization schedule generated by a senior biostatistician (R.H.), in a 1:1 ratio to receive CBT4BN or CBTF2F using a permuted block algorithm. The research coordinators at each site assigned the participant to the next available participant number, which corresponded to the assigned condition. Randomization was blocked by site and within site to conceal the randomization via electronic concealment (locked spreadsheets for each site).

Power

Statistical power was calculated prior to analysis using PASS (version 11.00.10), a sample size software. Power calculation was based on our view of being able to recruit 180 people in the expected time frame and with the given funds, after accounting for expected attrition, we conservatively predicted a sample size of 147, the anticipated number both randomized and eligible for analysis.[1] To yield the recommended level of 80% power at this sample size, with an expected abstinence rate of 30% in the CBTF2F condition based on previous research [27, 28], and a one-sided 95% confidence interval (CI), the non-inferiority margin would have to be a 19% difference in group outcomes. But, this margin was deemed too large clinically and we wanted it to be stricter (i.e., in practical terms, this meant that if 30% achieved abstinence in CBTF2F, for CBT4BN to be considered non-inferior then their abstinence rate may be as low as 11%). We recalculated power based on different margins, seeking to balance the trade-off between power and margin size, and settled on a 15% margin: with this margin, power was calculated to be 63%, less than recommended. This margin (d = 0.38) re-parameterized as an odds ratio and converted to Cohen's *d,* was used for all outcomes in the study to determine non-inferiority.

Statistical Analyses

Analyses were conducted using Stata 12.1 and SAS 9.4. Treatment preference at baseline was compared with a one-sample binomial test, and z was converted to d [26]. To compare credibility/expectancy of treatments, a repeated measures analysis of covariance adjusting for site was used to calculate Cohen's d_{rm}

[1] Originally, we desired a non-inferiority margin of 0.06% (Bulik et al., 2012), but calculations subsequently showed that such a sample size was not feasible.

and 95% CI [27]. A repeated measures analysis was used to control for dependency in the data since each participant was asked to rate both CBT4BN and CBTF2F, which may generate a spurious correlation between these pairs of ratings. Generalized estimating equations (GEE) models were constructed for each primary, secondary, and tertiary outcome. Each model included a main effect term for treatment condition (CBT4BN and CBTF2F); a covariate term for site; dependent variables with a measure of baseline severity had baseline value included as a covariate term; a time main effect; and a treatment condition × time interaction term (end of treatment and 12-month follow up), which tested whether the effects of CBT4BN and CBTF2F differed over time. Percent reduction in binge eating and purging were log-transformed. An exchangeable correlation structure was specified. Results for the main hypothesis were interpreted with respect to the 95% CI of *d* and the non-inferiority margin (i.e., Cohen's *d* = 0.38 instead of *p* values) [28].

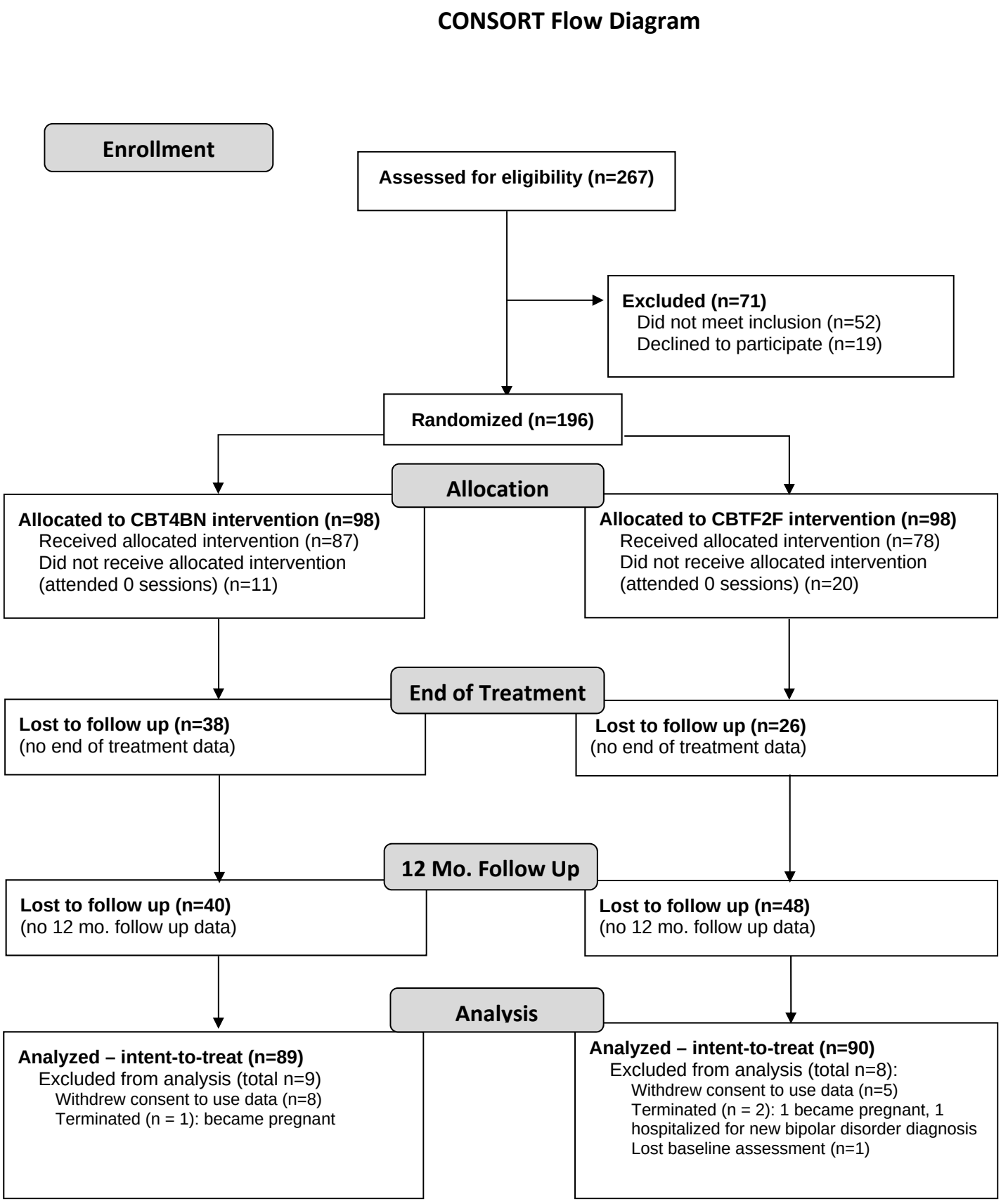
In line with other randomized controlled trials for BN, the amount of missing data at end of treatment and 12-month follow-up (*M* = 38%; 0%-51%) was unacceptable due to dropout (N.B., Although participants dropped out of treatment some still completed end-of-treatment and 12-month follow-up assessments). Participants who did not provide data on abstinence were scored non-abstinent at end of treatment, to conservatively estimate missing data on the primary outcome. Remaining missing data were imputed using multiple imputation by chained equations (MI) and maximum likelihood estimation with the expectation-maximization imputation (ML); MI results are reported. Convergence between MI and ML led to greater confidence in the results and only results that converged across imputation strategy are reported in the text. When ML produced the opposite substantive conclusion to MI this is noted in the tables with grey shading but is not reported in the Results text because our findings were inconclusive.

Sensitivity Analyses

In sensitivity analyses, we analyzed the original dataset without missing data imputed, analyzed the primary outcome as a binary variable comprising abstinence plus completion of treatment, and treated missing data using last observation carried forward [i.e. baseline scores replaced missing post-treatment

scores, and post-treatment (or baseline) scores were used in place of missing follow-up scores]. In the first analysis, the original intent-to-treat dataset was used and a complete case analysis was conducted. Results for the primary outcome, abstinence, were identical at post-treatment (original dataset, *d* = -0.21, 95% CI: -0.52, 0.11) and follow-up (original dataset, *d* = 0.24, 95% CI: -0.17, 0.65), reflecting inferiority at end of treatment and non-inferiority at the 12-month follow-up. In the second analysis, the primary outcome variable was redefined as abstinence plus completion of treatment (at least 75% of 16 sessions) and yielded findings of non-inferiority at end of treatment (MI: *d* = -0.01, 95% CI: -0.30, 0.28; ML dataset, d = -0.01, 95% CI: -0.30, 0.28) and non-inferiority at the 12 month follow up (MI dataset, *d* = 0.09, 95% CI: -0.20, 0.39; ML dataset, *d* = 0.26, 95% CI: -0.04, 0.55). In the third analysis, missing data were imputed with the last observation carried forward (LOCF) approach. This yielded identical results to the main analysis on the primary outcome at post-treatment (*d* = -0.13, 95% CI: -0.42, 0.16) and at 12-month followup (*d* = 0.03, 95% CI: -0.26, 0.33).

Online Figure 1.



Online Table 1. Descriptive statistics of study variables for individuals with bulimia nervosa in online CBT (CBT4BN) vs. face-to-face CBT (CBTF2F) (*N* = 179).

	CBT4BN		**CBTF2F**		
Characteristic	*N* = 89		*N* = 90		*p*
Sex (female)	98%	87	98%	88	0.98
Age (years)	28.5	9.3	27.5	9.1	0.47
Race					0.73
White	84%	75	86%	77	
Black or African American	7%	6	6%	5	
Asian	4%	4	1%	1	
Native Hawaiian or other Pacific Islander	-	-	1%	1	
Other	4%	4	7%	6	
Ethnicity					0.96
Latino	5%	4	5%	4	
Married/de facto	19%	17	22%	20	0.63
Employed	70%	59	67%	60	0.67
Education					0.13
Some high school or less	1%	1	-	-	
GED[a] or high school graduate	-	-	4%	4	
Some college or technical school	36%	30	44%	39	
College graduate	46%	39	36%	32	
Post graduate degree	17%	14	16%	14	
Age of BN onset (years)	18.6	5.6	18.3	5.4	0.72
Duration of illness (years)	9.5	8.9	9.5	8.8	0.95
Current psychiatric disorder					
Major depressive disorder	24%	21	19%	17	0.35
Any anxiety disorder	32%	26	20%	18	0.09
Current psychiatric medication	42%	32	41%	33	0.95
Body mass index (kg/m^2)	24.1	5.7	24.2	4.7	0.94
Frequency of binge episodes (past 28 days)	27.8	22.5	24.3	17.1	0.13
Objective	16.4	14.9	14.1	12.6	
Subjective	11.7	18.1	10.1	14.0	
Frequency of purging episodes (past 28 days)	31.7	34.2	26.8	20.7	0.19
Eating Disorder Examination (Global)	2.9	1.0	2.8	1.1	0.53

Values in the table are percentage and *N*, or mean and *SD*, and are based on non-missing data. Logistic regression was used for binary variables, negative binomial regression for count variables, and general linear modelling for continuous variables. Analyses were adjusted for site. Race was collapsed into a binary variable of White vs non-White. Education was collapsed into a binary variable of completion of college. [a]General Education Development test.

Online Table 2. Baseline demographic and clinical characteristics of treatment completers and non-completers.

Characteristic	**Completer** *N* = 74		**Non-completer** *N* = 105		*p*
Age (years)	28.3	9.4	27.8	9.0	0.67
White	88%	65	83%	87	0.45
Married/de facto	26%	19	18%	18	0.21
Education	68%	49	49%	50	0.01
Employed	68%	49	69%	70	0.91
Age of onset (years)	19.1	5.3	18.0	5.7	0.19
Duration of illness (years)	9.4	8.9	9.6	8.8	0.91
Major depressive disorder	26%	19	18%	19	0.13
Any anxiety disorder	26%	18	26%	26	0.97
Current psychiatric medication	39%	26	43%	39	0.58
Body mass index (kg/m^2)	23.0	4.1	24.9	5.8	0.02
Frequency of binge episodes (past 28 days)	26.7	20.0	25.5	20.0	0.41
Objective	15.3	13.7	15.3	14.0	0.95
Subjective	11.5	18.1	10.5	14.8	0.58
Frequency of purging episodes (past 28 days)	31.5	34.8	27.6	22.4	0.31
Eating Disorder Examination (Global)	2.8	1.1	2.81	1.0	0.84

Values in the table are percentage and *N*, or mean and *SD*, and are based on non-missing data. Logistic regression was used for binary variables, negative binomial regression for count variables, and general linear modelling for continuous variables. Analyses were adjusted for site. Education was collapsed into a binary variable of completion of college.

Online Table 3. Descriptive statistics of primary, secondary, and tertiary outcomes at end of treatment and 12-month follow-up presented in *M* (*SD*) or % (*n*): Multiple imputation analysis for all outcomes except end of treatment abstinence (*N* = 179).

		End of treatment				12-month follow-up			
	Rater	**CBT4BN[f]**		**CBTF2F[g]**		**CBT4BN**		**CBTF2F**	
Primary Outcome									
Abstinence	Clinician	14%	13	21%	19	30%	27	26%	23
Secondary Outcomes									
BN symptoms									
Binge-eating frequency	Clinician	9.6	10.6	9.0	10.2	8.4	12.5	7.4	8.7
% Reduction in binge eating[a]	Clinician	56.6	55.5	54.0	95.8	59.4	60.1	50.1	134.0
Purging frequency	Clinician	12.7	23.7	12.3	22.0	8.7	41.1	8.7	16.5
% Reduction in purging[a]	Clinician	56.6	55.5	54.0	95.8	59.4	60.1	50.1	133.9
BN diagnosis	Clinician	34%	30	28%	25	25%	22	25%	22
Treatment acceptability	Self	6.6	2.9	7.6	2.5	-	-	-	-
Tertiary Outcomes									
Eating Disorder Examination (Global Score)	Clinician	1.9	1.5	1.9	1.1	1.6	1.5	1.7	1.6
Body mass index	Self	22.7	6.6	24.0	4.7	23.7	6.7	23.7	4.5
Comorbidity									
Depressive disorder	Clinician	16%	14	20%	18	9%	8	13%	12
BDI[b]	Self	11.4	11.1	12.3	13.1	12.0	13.7	13.3	13.2
Anxiety disorder	Clinician	33%	29	26%	23	15%	13	18%	16
BAI[c]	Self	9.2	11.0	8.8	10.8	9.6	14.4	12.0	13.9
Quality of life									
EDQOL[d]	Self	1.0	0.7	1.0	0.9	0.9	0.9	1.0	0.8
SF-6D[e]	Self	0.7	0.1	0.7	0.1	0.7	0.1	0.7	0.1
Treatment									
Failure to engage	Self	12%	11	20%	19				
Dropout	Self	61%	54	57%	51	-	-	-	
Self-monitoring adherence	Self	1.6	1.2	1.8	1.2	-	-	-	

Values in the table are percentage and *N*, or mean and *SD*. [a]Non-imputed data; [b]Beck Depression Inventory; [c]Beck Anxiety Inventory; [d]Eating Disorders Quality of Life; [e]Short-Form Health State Classification; [f]CBT4BN=Online chat group therapy; [g]CBTF2F=Face-to-face group therapy