Supplementary materials

Appendix 1: Further motivation for the use of multiple imputation

Formal trial arm comparisons were carried out by multiple imputation (MI). This was necessary because non-adherence with treatment, defined as completing four of the telephone calls (not including booster sessions) for the TCBT arm and as four or more of the website sessions and at least one telephone call (not including booster sessions) for the WCBT arm, was found to be predictive of missing primary outcomes at 12 months in each of the CBT arms. To avoid unblinding the trial statisticians testing for whether the adherence to treatment was associated with missing data at the final time point was carried out by an independent statistician. The association between treatment adherence and missing data at the final time point in the TCBT or WCBT arms was tested using Fisher's exact tests. Adherence was found to be predictive in both CBT arms. Thus an MI approach was pursued to allow for a missing data generating mechanism that was missing at random (MAR), with the observed variables allowed to drive missingness including adherence with TCBT or WCBT.

To ensure that a realistic MAR process was allowed for by the MI procedure the following variables were included in the imputation model: (i) all variables of the analysis model, (ii) measures of the outcome variable at other assessment time points including baseline and (iii) known predictors of missingness. (i) is stipulated by MI theory. (ii) was done to improve the precision of the imputed values and also allow outcome measures at earlier time points to drive drop out at later time points. (iii) accommodates identified predictors of missingness including non-adherence with CBT therapy, and allowed us to make a more realistic MAR assumption. For analysis models that contained (random) therapist effects (iv) fixed effects for therapists in the TCBT arm were also added to ensure that the imputation model remained more general than the analysis model. The validity of the imputations were checked using the Stata command midiagplots (30).

Appendix 2: Selection process for identifying baseline drivers of missingness to be included in the imputation step of the MI procedure

We empirically assessed whether baseline variables were predictive of missing data using logistic regression. The following baseline variables were considered: IBS-SSS, HADS, WSAS, age at randomisation, IMD, duration of IBS symptoms, duration of IBS symptoms before diagnosis, whether the participant was registered with an IBS specialist, who they lived with, their marital status, the type of residence in which they lived, their choice of group at baseline, their highest level of education and their gender. Variables were considered to be potentially important, and later considered for inclusion in the imputation model, if an unadjusted logistic regression of missingness at 12 months on the baseline variable was statistically significant at a liberal 20% test level. The variables found to be potentially important predictors of missingness were then used in a forward selection approach to identify the best set of joint missingness predictors. At each stage, the more complex model was tested against the simpler model using a likelihood ratio test (20% level). The preferred model of baseline predictors of missingness was found to be baseline higher IBS-SSS and greater deprivation on IMD only.

Appendix 3: Selection procedure for identifying therapist effects to be included in the analysis and imputation models of the MI procedure

Since both TCBT and WCBT involved therapists delivering the intervention possible therapist effects on outcomes were investigated. For this purpose two therapists who saw only a small number of participants, all of whom were compliers, were merged into a single group which also included those participants in the TCBT and WCBT arms who were not assigned a therapist (due to failure to respond to invites to participate in therapy). This was to avoid the computational instability and perfect prediction issues which were seen when the two therapists with few participants were considered as single therapists. To select appropriate therapist effects a series of models were fitted for the participants who completed follow-up (completers) and compared using likelihood ratio tests: Model (A) allowed for therapist-varying random intercepts in each of the CBT arms with the variances of these random effects allowed to differ between TCBT and WCBT. Model (B) included therapist-varying random intercepts but only in the TCBT arm. Model (C) did not include any therapist effects. Model (A) fitting significantly better than model (B) at a liberal 10% level was interpreted as evidence for therapist effects in both CBT arms; model (B) fitting better than model (C) as evidence for therapists effects in the TCBT arm only.

Appendix 4: Results of complete case analyses

Table S1 provides the results of complete case (CC) analyses: The same analyses models as described in the statistical analysis section for the MI approach were employed. However, instead of using imputation to adjust for missing data biases a CC approach was used that further conditions on the identified drivers of missingness (by including baseline covariates IMD and IBS SSS). CC analyses simply delete cases with missing values from the analysis set. The CC analyses presented here rely on a more restrictive MAR assumption, namely that only covariates of the analysis model drive missing outcomes (trial arm, baseline values of the outcome, randomisation stratifier, IMD and baseline IBS SSS). We considered such an assumption unrealistic and hence chose the MI approach. The CC results are provided for comparison.

Comparison shows that results are similar with treatment effects estimated by CC analyses being of larger magnitude. Thus the extra adjustment employed by the MI approach to allow non-adherence with TCBT or WCBT to drive missingness leads to more conservative CBT effect estimates.

Table S1: Formal comparisons between CBT arms and TAU based on CC analysis (adjusting for baseline IMD and IBS SSS)

	TCBT vs TAU				**WCBT vs TAU**			
	Estimated difference	95% CI	Test (degrees of freedom) p-value	Standardised difference	Estimated difference	95% CI	Test (degrees of freedom) p-value	Standardised difference
IBS-SSS								
3 months	-74·0	(-93.6, -54.4)	t=-7·4 (420) p<0·001	0.78	-66·1	(-85.9, -46.2)	t=-6.5 (420) p<0·001	0·69
6 months	-63.1	(-84.0, -42.2)	t=-5.9	0.66	-45.4	(-66.9, -23.9)	t=-4.2	0·48

			(401) p<0·001				(401) p<0·001	
12 months*	-64.5	(-90.5, -38.5)	t=-4.9 (384) p<0·001	0.68	-43.1	(-64.3, -22.0)	t=-4.0 (384) p<0·001	0·45
WSAS								
3 months	-4.1	(-5.5, -2·7)	t=-5.8 (420) p<0·001	0.48	-3·2	(-4·6, -1·8)	t=-4.5 (420) p<0·001	0·38
6 months	-3.3	(-4·7, -2.0)	t=-4.8 (401) p <0·001	0.39	-3.1	(-4·5, -1·7)	t=-4.4 (401) p<0·001	0·37
12 months	-4.3	(-5·8, -2.8)	t=-5.6 (387) p<0·001	0.51	-3·6	(-5.1, -2.0)	t=-4.5 (387) p<0·001	0.42
HADS								
3 months	-2·2	(-3.3, -1.1)	t=-3·9 (420) p<0·001	0.32	-2·7	(-3.8, -1·6)	t=-4·7 (420) p<0·001	0.39
6 months	-3.2	(-4.5, -1.9)	t=-4·9 (401) p<0·001	0.46	-3.2	(-4.5, -1·9)	t=-4.9 (401) p<0·001	0.46
12 months	-3.2	(-4·5, -1·8)	t=-4·7 (387) p<0·001	0.46	-2.5	(-3.9, -1.2)	t=-3.8 (387) p<0·001	0.37
PEQ responders	Estimated OR	95% CI	Test p-value		Estimated OR	95% CI	Test p-value	
3 months	17.8	(9.1, 34.8)	z=8.4 <0·001		9.7	(5.0, 18.8)	z=6.7 <0·001	
6 months	16.5	(9.0, 29.9)	z=9.2 <0·001		6.3	(3.5, 11.2)	z=6.3 <0·001	
12 months*	11.4	(5.7, 22.8)	z=6.9 <0·001		4.0	(2.3, 6.8)	z=5.0 <0·001	
SGA Responders								
3 months	12.6	(7.2, 22.1)	z=8·8 <0·001		8.4	(4.9, 14.6)	z=7.6 <0·001	
6 months	9.8	(5.6, 17.2)	z=8.1 <0·001		6.9	(4.0, 11.9)	z=6.9 <0·001	
12 months*	8.8	(4.1, 19.1)	z=5.5 <0·001		4.4	(2.5, 7.6)	z=5.3 <0·001	

*The 12 months model included therapist effects in the TCBT arm.

Appendix 5: Sensitivity analyses

Four sensitivity analyses were conducted:
1) Purpose: To assess the impact of restricting eligibility criteria to those who were eligible at screening and also continued to meet the IBS SSS cut-off inclusion criteria at baseline, i.e. still scored ≧75 points on IBS SSS at baseline. A small number of participants had an IBS SSS of <75 at baseline and after discussion with the Trial steering committee it was felt appropriate to allow them to continue in the trial as they had met the inclusion criteria at screening. However, a sensitivity analysis was advised)
Approach: 16 participants (3 in the TCBT, 8 in the WCBT and 5 in the TAU arm) with an IBS SSS score <75 at baseline (despite IBS SSS ≧75 at screening) were dropped from the analysis set and the MI analyses repeated to check the impact on inferences. Dropping these

participants from the analysis for IBS SSS at 12 months gave an improvement of -63·0 (CI -91·9 to -34·2, p<0·001) for TCBT versus TAU and -35·5 points (CI -58·2 to -12·7, p=0·002) for WCBT versus TAU. For WSAS at 12 months differences were -3·5 points (CI -6·1 to -0·9, p=0·02) for TCBT versus TAU and -3·4 points (CI -4·7 to -2·1, p<0·001) for WCBT versus TAU. Comparisons of these findings with the original analysis results (cf Table 3) demonstrate that the analyses were robust regarding the timing of IBS SSS eligibility assessment.

2) Purpose: Participants were asked to record outcomes at 3, 6 and 12 months after randomisation within the pre-specified time window ranging from 7 days before the intended assessment date to 4 weeks afterwards. Treatment effects were expected to be reasonably constant over such a 5 week period. This sensitivity analysis assesses the impact of restricting the modelling to observations recorded within pre-specified assessment time windows.
Method: For primary outcomes at the 12 month assessment time point 98 (25·0%) IBS SSS values (TCBT 27.2% , WCBT 25.8%, TAU 22.1%) and 98 (24·9%) WSAS values (WCBT 26.8%, TCBT 25.8%, TAU 22.0%) were recorded outside this assessment window. These participants were dropped from the analysis and the co-primaries analysed in the same way as before. For IBS SSS at 12 months estimated trial arm differences were -64·3 points (CI -93·2 to -35·5, p<0·001) for TCBT versus TAU and -37·2 points (CI -64·1 to -10·4, p=0·007) for WCBT versus TAU. For WSAS at 12 months, estimated trial arm differences were -3·3 points (CI -4·9 to -1·8, p<0·001) for TCBT versus TAU and -3·3 points (CI -11.9 to 5.3, p=0·20) for WCBT versus TAU. The results demonstrate that apart from some loss of power (wider CIs) the IBS SSS analysis was not sensitive to outcomes being recorded within the specified time period. However, for WSAS power loss lead to the WCBT effect becoming non-significant (cf Table 3).

3) Purpose: Since we had not anticipated analysing PEQ scores on a binary scale we confirmed that the substantive conclusions from our analysis were not sensitive to the choice of threshold for defining a participant to be a "PEQ responder".
Method: In the main analysis we considered those reporting a PEQ score of 6 or higher to be responders. For this sensitivity analysis we considered responders to be those reporting a PEQ score of 4 or higher. The odds ratios of being a responder reported at 12 months using this criteria were estimated to be 12·7 (conditioned on therapist and stratification site, CI 4·9 to 33·0, p<0·001) for TCBT versus TAU. For the comparison of WCBT with TAU the odds ratio was estimated to be 4·0 (CI 2·3 to 6·9, p<0·001). Comparing these findings with those reported (cf Table 3) shows that while the definition of a "PEQ responder" increases the OR estimates it does not affect the substantive conclusions.

4) Purpose: The main analyses are intention-to-treat (ITT) analyses and assess the effectiveness of the CBT interventions. In the presence of non-adherence with allocated treatment an ITT analysis is known to be biased (too small) for estimating the efficacy of the treatment (effect of receiving the treatment). We aimed to estimate the efficacies of the CBT treatments.
Method: We assumed all-or-nothing adherence, that is, a participant either received a "sufficient dose of CBT" as per our predefined adherence criteria or they received TAU. Binary receipt (adherence) was operationalised in our trial protocol as completing at least 4 phone calls in the TCBT arm (84.4% or participants) and as accessing at least 4 web sessions and taking part in at least one phone session (69.2%). Participants in the TAU arm were not offered CBT for IBS (receipt=0). Efficacy was quantified by the Complier Average Causal Effect (CACE). To estimate CACE for TCBT or WCBT respectively an instrumental variables approach, namely a two-stage least squares regression, was employed. This approach used

binary randomisation to TCBT or WCBT as instruments for endogenous variables “receipt of TCBT” and “receipt of WCBT”. Further covariates in the models were baseline values of the outcome, dummy variables reflecting the randomisation stratifier and baseline variables found to predict missingness (IBS SSS and IMD). A complete case analysis approach was used, that is participants without primary outcome observations at 12 months were excluded from the analyses. 12 months efficacy as quantified by CACE was estimated to be -71.8 points (CI from -93.7 to -49.9, $p<0.001$) for TCBT compared with TAU and -50.4 points (CI from -75.2 to -25.5, $p<0.001$) for WCBT compared with TAU. For WSAS at 12 months the CACE estimates were -4.5 points (CI from -6.1 to -3.0, $p<0.001$) for TCBT compared with TAU and -4.2 points (CI from -5.9 to -2.4, $p<0.001$) for WCBT compared with TAU. We estimate that the efficacies for those who would comply with the CBT interventions were higher than the values suggested by the ITT (effectiveness) analyses.