A) QUANTITATIVE ANALYSIS PLAN

Investigators

Dr Hazel Everitt, University of Southampton

Professor Rona Moss-Morris, King's College London

Professor Trudie Chalder, King's College London

Professor Paul McCrone, King's College London

Professor Paul Little, University of Southampton

Professor Sabine Landau, King's College London

Dr Flis Bishop, University of Southampton

Dr Robert Logan, King's College Hospital

Dr Nicholas Coleman, Southampton University Hospital Trust

Chief investigator

Dr Hazel Everitt, University of Southampton

Trial manager

Dr Gilly O'Reilly, University of Southampton

Research assistants

Ms Stephanie Hughes, University of Southampton

Ms Alice Sibelli, King's College London

Ms Sula Windgassen, King's College London

Trial statisticians

Ms Rachel Holland, King's College London

Dr Kimberley Goldsmith, King's College London

Trial health economist

Professor Paul McCrone, King's College London

Sponsor

University of Southampton

1. Description of the trial

This is a pragmatic trial of cognitive behavioural therapy (CBT) and web-based CBT self-management for adults with irritable bowel syndrome. The trial will examine the treatment effect of high intensity therapist delivered CBT (TCBT) plus treatment as usual (TAU) or lower intensity web-based CBT (LIBT) plus TAU on severity of symptoms, functioning, symptom relief, distress, enablement, quality of life and health care costs. The trial will examine the effectiveness, acceptability and cost-effectiveness of the two experimental interventions in comparison with treatment as usual. The protocol has been published[29].

1.1 Principal research objectives to be addressed

To investigate whether therapist delivered CBT (plus TAU) or lower intensity web-based CBT (plus TAU) for people with irritable bowel syndrome affects symptom severity, functioning, other clinical outcomes and cost-effectiveness of health service use as compared to TAU only.

Primary objectives

1. To examine the treatment differences in symptom severity scores (as measured using the Irritable Bowel Syndrome Symptom Severity Scale (IBS-SSS)) between participants who are allocated to a course of TCBT or LIBT compared to those allocated to treatment as usual (TAU) at 12 months after randomisation.
2. To investigate the treatment differences in functioning (as measured using the Work and Social Adjustment Scale (WASAS)) between participants allocated to TCBT or LIBT and those allocated to TAU at 12 months after randomisation.

Secondary objectives

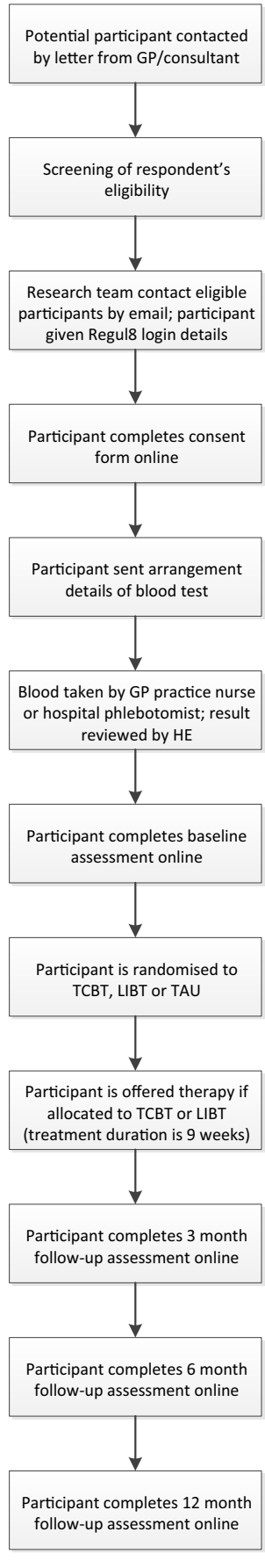
3. To examine the treatment differences in symptom severity scores (as measured using the IBS-SSS) between participants who are allocated to a course of TCBT or LIBT compared to those allocated to TAU at three and six months after randomisation.
4. To investigate the treatment differences in functioning (as measured using the WASAS) between participants allocated to TCBT or LIBT and those allocated to TAU at three and six months after randomisation.
5. To investigate the treatment differences in participants' relief from IBS symptoms (as measured using the Subject's Global Assessment of Relief scale (SGA)) between participants allocated to TCBT or LIBT and those allocated to TAU at all outcome time points (three, six and 12 months after randomisation).
6. To examine the treatment effects on patients' distress (as measured using the Hospital Anxiety and Depression Scale (HADS)) between participants allocated to TCBT or LIBT and those allocated to TAU at all outcome time points (three, six and 12 months after randomisation).
7. To investigate the treatment effects on patients' ability to cope with their illness (as measured using the Patient Enablement Questionnaire (PEQ)) between participants allocated to TCBT or LIBT and those allocated to TAU at all outcome time points (three, six and 12 months after randomisation).
8. To investigate the treatment differences in participants' quality of life (as measured using the EQ-5D scale) between participants allocated to TCBT or LIBT and those allocated to TAU at all outcome time points (three, six and 12 months after randomisation).
9. To summarise participants' perception of the acceptability of self-management treatment for participants allocated to TCBT or LIBT at all outcome time points (three, six and 12 months after randomisation).
10. To investigate the treatment differences in cost-effectiveness of health service use (as measured using the Client Services Receipt Inventory (CSRI)) between participants allocated to TCBT or LIBT and those allocated to TAU at all outcome time points (three, six and 12 months after randomisation).
11. To examine the treatment effect on number of GP contacts during follow-up between participants allocated to TCBT or LIBT and those allocated to TAU.

1.2 Trial design including blinding

The ACTIB trial is a parallel design, three-arm, multi-centre randomised controlled trial for adults with irritable bowel syndrome. Randomisation uses random block sizes stratified by treatment centre. Participants are randomised to receive either TCBT (plus TAU), LIBT (plus TAU) or TAU only. Blinding is planned for outcome assessors and the trial statistician.

Following database lock, a decision will be taken on whether to use multiple imputation (see section 2.6 assessment of outcome measures). If multiple imputation is not needed the trial statistician (RH) and outcome assessors will remain blind until after all databases are locked and the main analyses have been completed. Parts of the analyses which involve process variables or measures specific to treatment arm cannot be performed blind. These will be done at the end of the analysis in order to preserve blindness for as long as possible. If it is necessary to use multiple imputation in the main analysis, the trial statistician and assessors will be unblinded before the main analysis.

Figure 1. Trial design flow diagram



1.3 Method of allocation of groups

Once consent has been provided and baseline assessments have been completed, individuals will be randomised to one of the three treatment arms. Randomisation will be done in a 1:1:1 ratio. Randomisation is at the patient level and is performed using an online randomisation system set up by the King's Clinical Trials Unit (KCTU) at the Institute of Psychiatry, Psychology and Neuroscience at King's College London. Randomisation is stratified by type of treatment centre (Southampton GP practices, Southampton secondary care, London GP practices, London secondary care) with variable block sizes to ensure that equal numbers of patients are allocated to the three arms within each stratum. The procedure is as follows: on receipt of the baseline questionnaire, the trial manager or research assistant electronically submits details of each participant to the KCTU. This includes: participant ID number, site, initials and date of birth. The system immediately notifies the unblinded researchers and records the randomisation outcome.

1.4 Duration of the treatment period

The main treatment period for both TCBT and LIBT arms is nine weeks. Participants in the TCBT arm will have six one-hour telephone sessions with a CBT therapist during this time at approximately weeks one, two, three, five, seven and nine. They will also receive two one-hour booster sessions at four and eight months after randomisation.

Participants in the LIBT arm will undertake eight online sessions over nine weeks at home. They will receive three 30-minute telephone support calls from a therapist during this time at approximately two, four and six weeks and two booster telephone calls at four and eight months after randomisation.

1.5 Frequency and duration of follow-up

Participants will complete follow-up measures online at three, six and 12 months after randomisation using the LifeGuide website. 'Baseline complete' date is the closest date to the randomisation date which is available within the LifeGuide data, it is used as a proxy for randomisation date. Those who are unable to complete the measures online will receive a paper copy of the questionnaires (IBS-SSS, SGA, WASAS, PEQ, HADS). If this is not completed they will receive a telephone call from a researcher who will take the participant through a limited selection of the outcome measures. These are the IBS-SSS, WASAS, SGA, PEQ and HADS. Responses of participants contacted in this manner will be recorded in the MACRO database.

1.6 Visit windows

Participants will be sent an email and text one week before the questionnaire due date at each outcome time point to remind them to complete the measures. A copy of the paper questionnaire is also sent with a reminder by letter at the due date. If this is not done within a week of this email, a further two reminders by email and text will be sent. If no data have been entered one week after that, a researcher will call the participant to ask if they can collect the data over the telephone.

The acceptable time window for completion of questionnaires at any given time point is no more than 7 days before the expected due date of follow-up and no more than 28 days after the expected due date of follow-up. Treatment effects are expected to be reasonably constant over such a 5 week period.

1.7 Data collection

1.7.1 Eligibility screening

Eligibility was assessed at enrolment based on a screening questionnaire and blood tests.

Inclusion criteria

- Patient is aged 18 years old or over
- Patient has refractor IBS (clinically significant symptoms defined by a IBS-SSS > 75)
- Patient fulfils ROME III criteria
- Patient has been offered first-line therapies (e.g. anti-spasmodics, anti-depressants or fibre based medications) but still has continuing IBS symptoms for 12 months or more
- If over 60 years old, patient has had a consultant review in the previous two years to confirm symptoms are related to IBS and that other serious bowel conditions have been excluded.

Exclusion criteria

- Patient has unexplained rectal bleeding or weight loss
- Patient has diagnosis of inflammatory bowel disease
- Patient has diagnosis of coeliac disease
- Patient has diagnosis of peptic ulcer disease
- Patient has diagnosis of colorectal carcinoma
- Patient is unable to participate in CBT due to speech or language difficulties
- Patient has no access to an internet computer to be able to undertake the LIBT
- Patient has received CBT for IBS in the last two years
- Patient has had previous access to the MIBS website
- Patient is currently participating in an IBS / intervention trial

1.7.2 Measures

A detailed description of data collected is given in the Schedule of Assessments and Measures (section C of this document). What follows is a list of measures to aid understanding of the analysis plan.

Text or fields subsidiary to a free text variable will not be analysed or reported as part of the work described in this statistical analysis plan and are labelled (**).

1.7.2.1 Demographics and clinical information (measured at baseline only)

Demographics

- Sex (male; female)
- Age
- Ethnicity (Caribbean; African; other black background; British; Irish; other white background; Indian; Pakistani; Bangladeshi; other Asian background; white and black Caribbean; white and black African; white and Asian; other mixed background; Chinese; other ethnic group; not stated)
- Marital status (single; married; living with partner; separated; divorced; widowed)
- Who do you live with? (spouse/partner; spouse/partner and children; children (without spouse); parents; alone; other)
- Do you have any dependents?
    - Number of children under five years old
    - Number of children over five years old
    - Number of elderly
    - Number of other dependents
- Usual place of residence (owner occupied flat/house; privately rented flat/ house; flat/house rented from local authority; other)
- Index of Multiple deprivation IMD 2010
- Education level (no formal education; GCSE/O-level or equivalent; A-level or equivalent; degree; postgraduate; other)
- Do you have an IBS specialist/consultant?
- If you had a choice, which arm of the trial would you choose to participate in? (TCBT; LIBT; TAU)

Clinical information (measured at baseline only)

- When were you diagnosed with IBS? (year and month)
- For how long before diagnosis did you experience symptoms relating to IBS? (years and months)
- Are you, or have you ever had any experience with any of the following remedies for your IBS? (relaxation techniques; hypnotherapy; acupuncture; herbal remedies; counselling; previous CBT; any other psychological therapies; seen a dietician)
- Have you tried any specific diets? (yes; no)
- If so, which diets? (FODMAP; other) (**)
- Have you ever been treated for depression? (yes; no)
- Have you ever been treated for anxiety? (yes; no)
- Are you a member of a patient organisation/self-help group for IBS? (yes; no)

- If you are a member of a patient organisation/self-help group, how often do you participate in face-to-face meetings? (never; several times/year; monthly' fortnightly; weekly or more often)
- If you are a member of a patient organisation/self-help group, how often do you participate in online forums/groups? (never; several times/year; monthly; fortnightly; weekly or more often)

1.7.2.2 Primary outcome measures (measured at baseline, 3, 6 and 12 months)

There are two primary outcomes, IBS-SS and WASAS. These are recorded at baseline with the primary outcome recorded at 12 months. Time points of three and six months are treated as secondary outcomes.

- Symptom severity at 12 months after randomisation (as measured by the IBS Symptom Severity Scale (IBS-SSS))
- Functioning at 12 months after randomisation (as measured by the Work and Social Adjustment Scale (WASAS))

1.7.2.3 Secondary outcome measures (measured at baseline, 3, 6 and 12 months)

- Symptom severity at three and six months after randomisation (as measured by the IBS-SSS)
- Functioning at three and six months after randomisation (as measured by the WASAS)
- Relief from IBS symptoms at 3,6,12 month outcome time points (as measured by the Subject's Global Assessment of Relief (SGA))
- Distress at 3,6,12 month outcome time points (as measured by the Hospital Anxiety and Depression Scale (HADS))
- Ability to cope with illness at 3,6,12 month outcome time points (as measured by the Patient Enablement Questionnaire (PEQ))
- Quality of life at 3,6,12 month outcome time points (as measured by the EQ-5D)
- Cost-effectiveness of health service use at all outcome time points (as measured by the Client Services Receipt Inventory (CSRI))
- Number of GP contacts during follow-up

1.7.2.4 Other measures in protocol – hypothesized mediators (measured at baseline, 3, 6 and 12 months)

Mediation analysis is not covered in this document. The variables listed below were recorded for testing mediation theories, not for the purpose of clinically characterising the sample. They will therefore not be reported as part of this primary analysis plan. Summaries of these variables at baseline will be covered in later secondary analysis.

- Unhelpful cognitions related to IBS (as measured by the Cognitive Scale for Functional Bowel Disorders (CS-FBD))
- Perception of illness (as measured by the Brief Illness Perception Questionnaire for IBS (B-IPQ))
- Behaviour specific to managing IBS symptoms (as measured by the IBS Behavioural Responses Questionnaire (BRQ))

- Beliefs about the unacceptability of experiencing and expressing negative emotion (as measured by the Beliefs about Emotion Scale (BES))
- Awareness of emotional events (as measured by the Impoverished Emotions Experience (IEE) of the Emotional Processing Scale)
- Positive affect (as measured by the Positive and Negative Affect Schedule (PANAS)) (negative affect is measured by the HADS)

1.7.2.5 Adverse events (measured at time of event)

- Description of adverse event
- Body system
- Duration of AE
- Intensity
- Related to study intervention? (definite; probable; possible; remote; none)
- Outcome (resolved; resolved with sequelae)

1.7.2.6 Therapist details (recorded at therapist's start)

- Core profession of therapist (psychologist – clinical; CBT psychotherapist; other)
- Number of years working in core profession
- Number of years working with MUS
- Sex (male; female)
- Age of therapist

1.7.2.7 Process variables (measured during treatment period, not at baseline)

- Number of phone sessions
- Duration of phone sessions
- Count of web sessions accessed
- Homework task completion (completed; partially completed; not completed)
- Engagement with homework
- Continuing use of strategies learnt during treatment
- Perception of the acceptability of self-management treatment at all outcome time points

1.8 Sample size estimation (including clinical significance)

A 35 point difference between therapy groups and TAU on IBS SSS at 12 months is regarded as clinically significant (assuming a 15 point placebo response in the TAU arm in the trial[1, 15, 24]). Assuming a within-group IBS-SSS standard deviation of 76 points (taken from MIBS pilot study[15]) this equates to an effect size of 0.46. To achieve 90% power to detect such an effect or larger using a two-sided independent samples t-test at the 2.5% significance level (adjusting for 2 primary outcomes) would require 119 subjects per group. Based on each of 10 therapists delivering therapy to 17 patients within LIBT and TCBT groups and an intraclass correlation of 0.02, taken from Baldwin[38], this sample size needs to be increased by an inflation factor of 1.32 to take account of therapist effects. We will measure IBS SSS at baseline and assume that baseline values are predictive of post treatment values (correlation 0.4). Accounting for this in our statistical analysis model allows us to decrease the sample size by a deflation factor of 0.84. Finally, assuming that attrition will be less than 20% we apply a further inflation factor (factor 1.25) to allow for this. The final sample size requirement is 165 patients per group or 495 patients in total.

As the trial progressed we found that the attrition rate was closer to 30% (Nov 2014 estimate). The sample size was recalculated using the same group size of 119 subjects with inflation and deflation factors of 1.32 and 0.84 kept constant. The updated attrition rate of 30% gives a sample size of 189 patients per group and a total of 567 patients.

In terms of our second primary outcome (WASAS), this sample size would be sufficient to detect a clinically important difference between the LIBT (or TCBT) and TAU groups in the WASAS. Specifically, we can assume inflation factors of 1.32 for correlation of outcomes within therapists and of 1.25 for attrition and a deflation factor of 0.84 for correlation between baseline and follow-up measures. Therefore, a moderate effect size of 0.46 could be found with 90% power at the 2.5% significance level, given 119 participants per group. Assuming a standard deviation of 8.0 (as estimated in a study of CBT for IBS[8]) this would equate to a clinically meaningful treatment difference of 3.7 points on this scale. This is less than the difference of 5.4 points in change of means that was found in a trial of a CBT-based self-management intervention for IBS[20].

1.9 Brief description of proposed analyses

What follows is a brief introduction to the analyses. Further details are given later on in this document.

Analyses will be carried out by the trial statistician (RH). In the first instance data will be analysed under intention-to-treat assumptions (i.e. analyse all those with data in groups as randomised irrespective of treatment received).

The primary outcome measures, symptom severity (IBS-SSS) and functioning (WASAS), will be analysed using longitudinal linear mixed modelling, including outcome measures at all time points and adjusting for the stratification variable (treatment centre) and baseline measures.

The secondary outcome measures will be also analysed using longitudinal linear mixed modelling, including outcome measures at all time points and adjusting for the stratification variable and baseline measures.

In the case of missing assessments, such analysis can include these participants provided that pre-randomisation values are available for the respective scales. The analysis presumes that the drop-out mechanism is missing at random (MAR). We will use multiple imputation instead of linear mixed modelling if this assumption is found to be incorrect.

Sensitivity analyses will be performed to assess the robustness of the conclusions to the assumptions made regarding the missing value generating process.

Data summaries and analyses will be carried out in Stata 14.

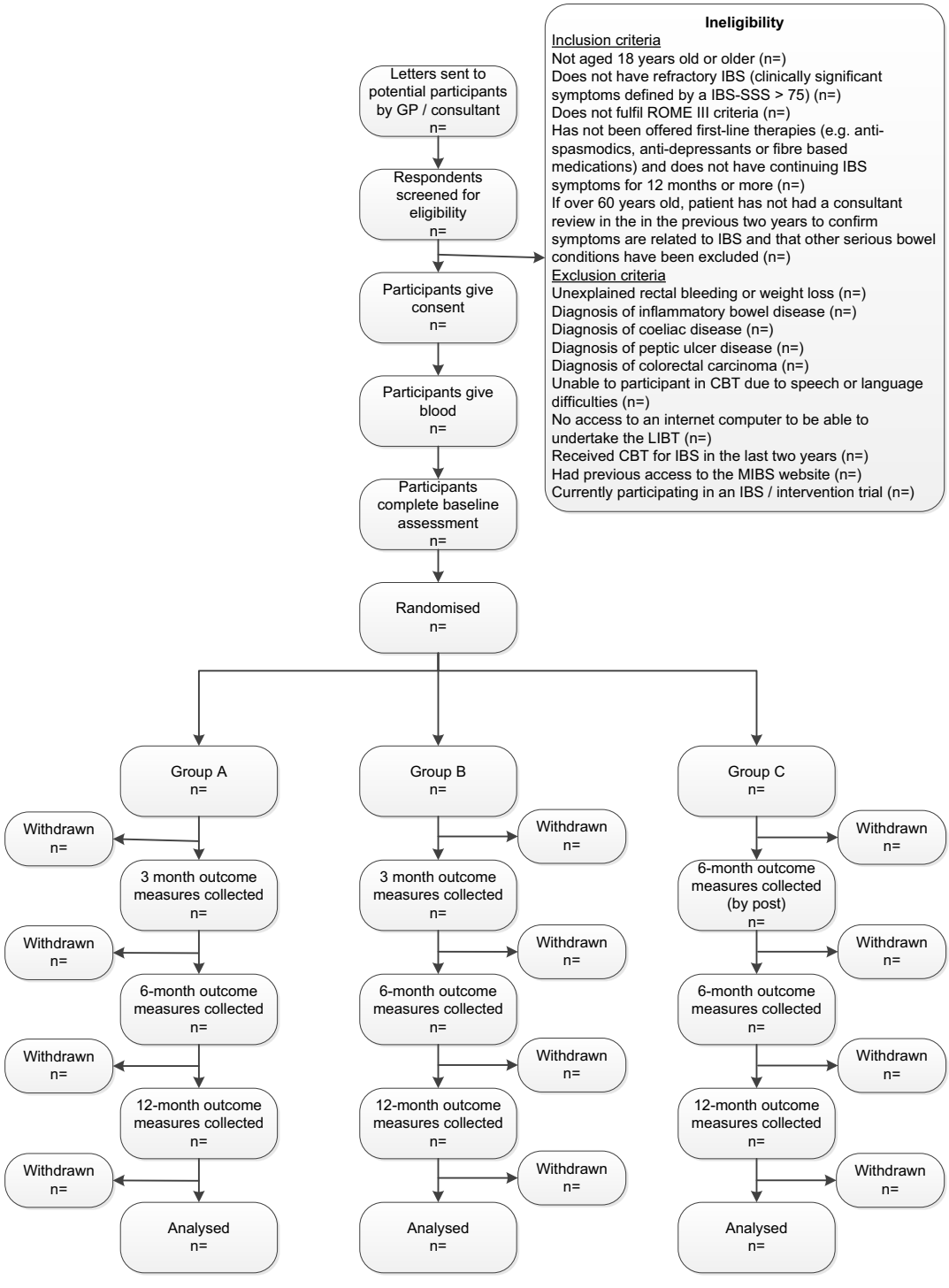
2. Data analysis plan – Data description

2.1 Recruitment and representativeness of recruited patients

A CONSORT flow chart will be constructed[42] – see Figure 2. This will include the number of potential patients contacted, number screened, number of eligible patients, number of patients agreeing to enter the trial, number of patients refusing, then by treatment arm: the number of patients adherent with treatment, the number continuing through the trial, the number withdrawing, the number lost to follow-up and the numbers excluded/analysed.

Treatment adherence is defined separately for the two active treatment arms. Participants allocated to TCBT who complete at least four of the initial telephone calls will be deemed as adherent with treatment. Those who are offered LIBT and complete four or more website sessions and at least one telephone support call will be considered as treatment adherent.

Figure 2. Template CONSORT diagram for ACTIB trial



2.2 Baseline comparability of randomised groups

Baseline descriptions of participants by treatment and overall: minimums and maximums, means and standard deviation, medians and quartiles for continuous variables as appropriate. Frequencies and proportions will be presented for categorical variables. No significance testing will be used to test baseline differences between the trial arms.

Categorical and continuous baseline variables listed in section 1.7.2 will be reported overall and by trial arm.

2.3 Adherence to allocated treatment

Binary adherence with a treatment course is defined in section 2.1. Resulting adherence rates will be calculated. The reasons for non-adherence (withdrawal from treatment) will be summarised. In addition further process variables listed in section 1.7.2 will also be summarised by trial arm. See section 2.6 for details on blinding.

Baseline characteristics of those who adhere with allocated treatment will be contrasted to those who do not adhere within treatment arms.

2.4 Loss to follow-up and other missing data

Withdrawal from trial follow-up (attrition) will be reported by intervention group. Moreover, the proportions of participants missing each variable will be summarised in each arm and at each time point.

The baseline characteristics of those missing follow-up at 12 months will be compared to those with complete follow-up. The relationship between baseline characteristics and missing data will also be investigated graphically. Factors affecting missingness will be examined using a logistic regression. This will be done by generating a binary variable for missingness for IBS_SSS and WASAS at 12 months after randomisation and regressing this on baseline variables.

The relationship between adherence and drop-out will also be assessed for the two therapy groups. This will be done using binary variables to indicate adherence to therapy as defined in section 2.1, and for drop-out at 12 months. The relationship between these variables within trial arms will be tested using a chi-squared test. The results of these analyses will inform the need to use multiple imputation in the formal analysis. This is due to the fact that these post-randomisation variables cannot be included as covariates in the model without changing the meaning of the results.

Finally, we will assess whether the assumption of MAR is reasonable. If it is found not to be reasonable, we will consider multiple imputation.

2.5 Adverse event reporting

Adverse events (AE), adverse reactions (AR), serious adverse events (SAE) and serious adverse reactions (SAR) will be summarised by treatment arm.

2.6 Assessment of outcome measures (unblinding)

Outcome assessors and the trial statistician are being kept blind to treatment allocation and will remain blind until all databases are locked for all participants for the baseline, three month, six month and 12 month time points. At this time an independent statistician will identify whether missing 12 month outcomes is associated with adherence and a decision on using multiple imputation will be taken. If it is necessary to use multiple imputation the statistician will be unblinded. If multiple imputation is not needed the statistician will stay blinded until the primary analysis is complete.

2.7 Scoring of questionnaire outcomes

IBS Symptom Severity Scale (IBS-SSS)[24]

The IBS-SSS is a visual analogue scale which can be used to rate the severity of IBS symptoms. The possible range of scores is 0-500.

The total score comprises the sum of responses to five questions, four of which are rated out of 100. Question 1c) is the frequency of pain over the last ten days – therefore this score must be multiplied by ten before the total score is computed.

Higher scores indicate greater levels of IBS symptom severity.

Scoring:

The primary publication for this questionnaire does not give instructions for how to deal with missing data. In the absence of such guidelines, the total pro-rata score will be calculated provided that the level of missingness is less than or equal to 20%. In this way, the total score will be calculated provided that at least four out of five items are present. If four items are non-missing, the mean of these four will be used to impute the fifth.

Work and Social Adjustment Scale (WASAS)[25]

The WASAS is a measure of impairment of functioning. The possible range of scores is 0-40.

The total score comprises the sum of responses to five questions, each of which is on a nine-point Likert scale (i.e. 0-8).

Higher scores indicate that ability to complete day-to-day tasks is more greatly impaired.

Scoring:

The primary publication for this questionnaire does not give instructions for how to deal with missing data. In the absence of such guidelines, the total pro-rata score will be calculated provided that the level of missingness is less than or equal to 20%. In this way, the total score will be calculated provided that at least four out of five items are present. If four items are non-missing, the mean of these four will be used to impute the fifth.

Subject's Global Assessment of Relief (SGA)[26]

The SGA is a measure of overall wellbeing, abdominal pain/discomfort, and bowel function and uses a single item with a five-point Likert scale.

Greater scores indicate deterioration of wellbeing / pain / discomfort / bowel function.

Scoring:

Patients scoring from 1–3 (completely relieved; considerably relieved; somewhat relieved) are considered responders, those scoring 4–5 (unchanged; worse) are non-responders.

Hospital Anxiety and Depression Scale (HADS)[27]

The scale asks seven questions each about depression and anxiety; the total scores for each are analysed separately. The possible range of scores for each scale is 0-21. The combined scales representing distress as an overall score range from 0 to 42.

The depression items include:

- I feel as if I am slowed down;
- I still enjoy the things I used to enjoy;
- I have lost interest in my appearance;
- I can laugh and see the funny side of things;
- I look forward with enjoyment to things;
- I feel cheerful;
- I can enjoy a good book or radio or television programme.

The anxiety items include:

- I feel tense or 'wound up';
- I get a sort of frightened feeling like 'butterflies' in the stomach;
- I get a sort of frightened feeling as if something awful is about to happen;
- I feel restless as if I have to be on the move;
- Worrying thoughts go through my mind;
- I get sudden feelings of panic;
- I can sit at ease and feel relaxed.

Higher scores indicate either greater anxiety or depression. We will consider the separate scores for anxiety and depression, plus the combined score for distress[23,24]. Studies using the combined HADS scores to determine the presence of psychological distress use scores from 13 to 19[24], we will use the midpoint 17 as our cut-off score[25] indicating any mental disorder. The HADS anxiety score will use a cut-off of 10[27], the HADS depression score will use a cut-off of 7[27].

Scoring:

Find the pro-rata mean for the seven items of the anxiety and depression scales separately and multiply by seven to give a total score for each construct, provided that at least six out of seven items are present for each of the constructs. The sum of these two scores will give the overall psychological distress score. Greater than 17 will be considered as a case of distress.

<u>Patient Enablement Questionnaire (PEQ)[21]</u>

The PEQ is a measure of self-efficacy. The possible range of scores is 0-12.

The total score comprises the sum of responses to six questions, each of which is on a three-point Likert scale (i.e. 0-2).

Higher scores indicate that belief in ability to cope / self-efficacy has improved.

Scoring:

The primary publication for this questionnaire does not give instructions for how to deal with missing data. In the absence of such guidelines, the total pro-rata score will be calculated provided that the level of missingness is less than or equal to 20%. In this way, the total score will be calculated provided that at least five out of six items are present. If five items are non-missing, the mean of these five will be used to impute the sixth.

2.8 Descriptive statistics for outcome measures

The primary and secondary outcomes listed will be summarised overall at 3, 6 and 12 months and by trial arm; IBS-SSS, WASAS, HADS, SGA, PEQ.

Each of the outcome measures will be described by treatment group and time point. Means and standard deviations or medians and interquartile ranges will be used for continuous variables; box plots, histograms and Q-Q plots will be used to assess whether the distribution of a variable is normal. Frequencies and proportions will be used to describe categorical variables.

2.9 Descriptive statistics for process variables

Process variables will be summarised for each treatment arm, including the three questions on acceptability.

For the two CBT arms, acceptability of treatment by treatment arm at 3, 6 and 12 months will be compared. Specifically, this will consider the responses to the questions; How useful did you find the telephone CBT or Regul-8 programme overall? Compared to other treatments you have received for your IBS how do rate this programme? Overall, how satisfied are you with the treatment you received? The numbers and percentages in each of the 4 categories from very useful to useless will be reported for both of the treatment groups at each time point. Fisher's exact test will be used to examine the difference between the therapy arms at the 12 month time point. See the section of handling multiple comparisons in 3.1.4.

2.10 Description of therapists/therapies

Details of the therapists will be summarised in tables using appropriate summary statistics.

Relevant variables are listed in section 1.7.2.

3. Data analysis plan – Inferential analysis

3.1 Main analysis of treatment differences

The main statistical analyses will estimate the difference in mean outcomes between patients randomised to TCBT and TAU, LIBT and TAU or TAU alone by intention to treat at the various post-treatment observation time points. This trial is not powered to provide comparisons between CBT and LIBT, thus such analyses are not included in this SAP. Group difference estimates of the IBS-SSS change between the TCBT with TAU and TAU group and between the LIBT with TAU and TAU group, associated confidence intervals and standardised effect sizes will be reported.

Missing post randomisation assessments will be dealt with by fitting adequate linear models to all the variables using maximum likelihood methods. Such an approach provides valid inferences under the assumption that the missing data mechanism is ignorable (MAR) and provided that predictors of missingness are included as covariates in the model. If the MAR assumption is found not to hold, e.g. because process variables are found to predict missing data, multiple imputation will be used instead.

Group difference estimates and associated 95% confidence intervals (or 97.5% confidence intervals for primary outcomes) will be reported. The trial statistician (RH) will remain blind as long as possible; for details regarding the unblinding procedures see sections 1.2 and 2.6.

The significance level will be 2.5% (two-sided) for each of the two primary outcomes. Secondary analyses will be carried out at the 5% level but will have to be interpreted with care as the effect of multiple testing is not taken account of.

Sensitivity analyses will be used to assess the robustness of conclusions from assumptions such as non-ignorable missing outcome data or, departures from randomised treatment.

3.1.1 Analysis of primary outcomes

The analysis population will include all patients with non-missing baseline measurements. The primary outcomes are symptom severity (IBS-SSS) and overall functioning (WASAS) at 12 months post randomisation. Outcome data at all three post treatment time points (three, six and 12 months after randomisation) will be modelled simultaneously. These outcomes will constitute the dependent variable and symptom severity / functioning at baseline, treatment centre, predictors of drop-out, trial arm, time dummy variables and a treatment*time interaction term will be included as explanatory variables. The covariance matrix of the repeated measures will be carefully modelled. An unstructured covariance matrix and the covariance matrix implied by a random intercept model will be formally compared, and the best covariance structure identified. This analysis is valid provided that outcomes are missing at random (MAR). This is to say that given the observed data, the missingness pattern does not depend on unobserved data.

The relationship between baseline variables and missing outcome data will be assessed using logistic regression with an outcome variable that represents whether outcome symptom severity / functioning data are present or missing at 12 months. Should any baseline variables be predictive of missingness then these will be included as covariates in models. Should the post-treatment variable “adherence with treatment” predict missingness then multiple imputation will be used to allow for this form of MAR. The impact of departures from MAR on treatment effects will be assessed using sensitivity analysis.

Random therapist effects will be modelled: It will be tested whether a significant interaction exists between randomisation arm and therapist. Potential clustering due to patients being treated by the same therapist in the TCBT or LIBT arms will then be allowed for by adding respective random effects for therapists in these arms.

Besides expressing effects as standardised differences (Cohen’s d), numbers needed to treat (NNT) will also be calculated for each of the two primary outcomes. We define an event as a participant improving their score across the study (between baseline and the final outcome of 12 months) by the pre-specified differences given below. The NNT for each treatment arm is calculated as the inverse of the event rate in the treatment arm minus the event rate in the control arm. This can be interpreted as the number of patients who must be treated in order for one patient to improve their score on the IBS-SSS or WASAS by at least the pre-specified point difference.

A 35 point difference between therapy groups and TAU on IBS SSS at 12 months is regarded as clinically significant (see section 1.8). Assuming an expected 15 point placebo response in the TAU arm in the trial[1, 15, 24] this translates into an expected improvement of 50 points in the therapy arms. Following on from this we define participant improvement on the IBS SSS as a decrease of 50 points or more. Thus the number of participants making at least the specified improvement of 50 points will be calculated in each arm of the trial arms.

There is no agreed clinically significant improvement in points between baseline and final outcome for WASAS. In order to calculate the NNT, we therefore apply the same threshold for defining clinical significance as used for IBS SSS: The minimal clinically important group difference in IBS SSS translates into a moderate standardised effect size of 0.46 (see section 1.8). Assuming that the standard deviation on the WASAS is 8 points (Kennedy et al[7]) this equates to a difference of 3.7 points between treatment arms and TAU on the WASAS. If we further assume an expected improvement of 2.2 points between baseline and 12 moths for the TAU arm, then this suggests an expected change of 5.9 points in the treatment arms. Thus the number of participants in each trial arm improving their score by at least 5.9 points will be calculated.

3.1.2 Analysis of secondary outcomes

Secondary patient outcomes relating to symptom severity and functioning (at three and six months after randomisation), distress (HADS) and ability to cope with illness (PEQ) will be analysed using linear mixed models in a similar method to that described above.

Relief from IBS symptoms (SGA) is a binary measure (responders; non-responders). A mixed logistic model will be considered for this outcome.

Health economic outcomes are discussed below in the section ECONOMIC ANALYSIS PLAN.

3.1.3 Sensitivity analyses

3.1.3.1 Departure from MAR assumption

If clinical input can be provided we will test sensitivity to the impact of departures from the assumption that missing data in the 12 month outcome data for IBS-SSS and WASAS is missing at random. This assumption implies that the mean change within the 3 groups would be the same regardless of whether data is missing or not. We intend to test the sensitivity to this assumption by using a range of possible mean differences in outcome between those with missing data and those with observed values in the three trial arms. This will be based on the investigators' opinion about the possible range of mean differences between those with missing data and those without.

3.1.3.2 Adherence to treatment

Adherence to treatment will be summarised by treatment arm. If there is considerable non-adherence then the primary intention-to-treat analysis might be biased for the purpose of assessing efficacy, that is estimating the effect of actually receiving treatment as defined in the protocol (see the definition of adherence in section 2.1). In order to assess this non-compliance bias we will estimate the complier average causal effects (CACE) and contrast this estimate with the ITT estimate. Methods for CACE estimation are described in the CACE estimation section below.

3.1.3.3 Eligibility

A further sensitivity analysis will be done to consider the effect of excluding those participants who were found to have refractor IBS at screening but were found to no longer have refractor IBS at the second pre-randomisation assessment time point (baseline). That is, those participants who had an IBS-SSS core of 76 or greater at screening but were found to have a score of less than 76 at baseline will be excluded.

3.1.3.4 Constancy of time effect (timeliness of questionnaire completion)

The number of questionnaires completed on time will be summarised at 3, 6 and 12 months.

A further sensitivity analysis will be done to assess the effect of failure to complete follow-up questionnaires (IBS-SSS and WASAS) within the expected time period. The main analysis assumes that the treatment effect assigned to an assessment time point (3, 6 or 12 months) is constant across the assessment window. We look at how results change if we assume that the treatment effect is constant only within the defined time window by dropping data points recorded outside of the defined time window in a sensitivity analysis.

Follow-up questionnaires are completed at 3 months, 6 months and 12 months after randomisation. The acceptable time window for a questionnaire for a given follow-up time point is no more than 7 days before the expected due date of follow-up and no more than 28 days after the expected due date of follow-up.

3.1.4 Statistical considerations

Stratification and clustering

Randomisation is stratified by type of treatment centre (which has four levels). Therefore it is important to include this variable as a covariate in the modelling process.

The structure of the majority of the data is longitudinal with repeated measurements at baseline, three months, six months and 12 months after randomisation. This correlation of observations within participants is being taken into account by a modelling process for the covariance matrix.

Correlation between repeated measures and due to sharing the same therapist will be allowed for by including subject-varying random intercepts as well as therapist-varying random intercepts for TCBT and LIBT groups in the mixed models.

Missing items in scales and subscales

The number (%) with complete data will be reported. The ideal approach is to use missing value guidance provided for scales. Where this guidance is not available, scales will be pro-rated for an individual if 20% or fewer items are missing. For example, in a scale with 10 items, prorating will be applied to individuals with 1 or 2 items missing. The average value for the 8 or 9 complete items will be calculated for that individual and used to replace the missing values. The scale score will be calculated based on the complete values and these replacements. The scoring rules for each of the questionnaire scales are listed in section 2.7.

Missing baseline data

Missing baseline data should not be a problem. However, if we encounter missing baseline values of outcome variables then these can be singly imputed according to White and Thompson[74] without incurring bias of the treatment effect estimate.

Missing outcome data

Missing post-randomisation assessments will be dealt with by fitting linear mixed models to all the available data using maximum likelihood methods. Such an approach provides valid inferences under the assumption that the missing data mechanism is ignorable (or MAR). To test the assumption of MAR we will explore whether baseline variables are associated with missing outcome data at the 12 month time point. Binary variables will be created for each individual questionnaire indicating missing/not missing if the time point is missing for that questionnaire. We will perform univariate logistic regressions using the questionnaire indicator variable as the outcome and each of the baseline variables in turn as the independent variable. Non-compliance with treatment will also be considered as a potential factor affecting the missingness of outcome data. Baseline variables with a univariate p-value of 0.2 or less will be entered into a series of manual forward stepwise logistic regression analysis to find the most important variables associated with missing outcome for each questionnaire. Variables will be retained in all models based on likelihood ratio tests.

If post treatment variables, such as adherence with treatment, are found to be predictive of drop-out, multiple imputation will be considered.

Method for handling multiple comparisons

Hochberg - Bonferroni adjustment for multiple outcomes will be used for the two primary outcomes.

Method for handling non-adherence (per protocol/CACE analyses)

In addition to the primary intention-to-treat analysis, the effect of actually receiving treatment as defined in the protocol (see sensitivity analysis 3.1.3.2) will also be estimated. If "non-adherence" with the active treatment is high, a CACE analysis will be considered (see below).

Instrumental variable (IV) methods will be used to assess the efficacy of the TCBT/ LIBT treatments. Specifically, we will use IV methods to evaluate the causal effect of TCBT/LIBT on clinical outcomes in the subpopulation who comply with intervention. The application of IV methods for explanatory evaluation of RCTs has been advocated because random allocation itself provides a strong instrument for treatment receipt.[75]

Model assumption checks

The models assume normally distributed outcomes; this will have been checked when describing the data and if substantial departures from normality occur, transformations will be considered. Residuals will be plotted to check for normality and inspected for outliers.

3.1.5 Planned subgroup analyses

No subgroup analyses are planned. The study is not powered to investigate interaction effects.

3.2 Interim analysis

No interim analyses are planned for this study.

4. Software

Data management: Two online data collection systems will be used. These are LifeGuide and MACRO (InferMed Ltd). The senior research assistant (SH) who is in charge of LifeGuide will extract the data from the main database when required. MACRO is hosted on a dedicated server at KCL and managed by the KCTU. The KCTU Data Manager will extract data periodically as needed and provide these in comma separated (.csv) format. The trial manager or trial team will extract the LifeGuide data, remove unblinding data if required or any disclosive information and provide these in comma separated (.csv) read-only format.

Statistical analysis: Stata 14 will be used for data description and the main inferential analysis.

B) ECONOMIC ANALYSIS PLAN

Heath economic objectives

As listed in section 1.1:

10. To investigate the treatment differences in cost-effectiveness of health service use (as measured using the Client Services Receipt Inventory (CSRI)) between participants allocated to TCBT or LIBT and those allocated to TAU at all outcome time points (three, six and 12 months after randomisation).
11. To examine the treatment effect on number of GP contacts during follow-up between participants allocated to TCBT or LIBT and those allocated to TAU.

Economic measures

We will measure costs and assess cost-effectiveness from both a health service and a societal perspective. To calculate the cost of TCBT the number of sessions with therapists will be recorded and combined with the unit cost of therapist time. The latter will be calculated using information on the salary band of therapists, with additional costs representing capital, overheads, training and qualifications[69]. We will ask therapists to estimate how much time during a typical working week is spent in telephone contact with patients and combine this with the total cost and total hours worked per week, in order to produce a cost per hour of direct patient contact time. For LIBT, the number of times therapist support is provided will be recorded and costed in a similar way. The LIBT development costs will be estimated and apportioned over those using the intervention. Other service use will be measured with a service receipt schedule at baseline (going back six months) and each follow-up (with measurement covering the whole period since the prior interview). The schedule will be based on other questionnaires used in similar research[28]. Services will include primary and secondary healthcare, and medication. Service costs will be generated by combining these data with appropriate unit cost information (e.g. NHS Reference Costs[69], and the British National Formulary) and these costs added to the intervention costs in order to generate total health costs per person.

Societal costs will be calculated by including family care costs and lost production. Family care costs will be recorded by asking patients to state how much time per week family members (and friends) spent providing support in specific areas *because of the IBS*. This time will be combined with average wage rates. Lost days and hours from work will be recorded on the schedule and combined with average wage rates to generate lost production costs. Cost comparisons between the three groups will be made at three, six and 12 months and over the entire follow-up period, in both cases controlling for baseline costs. Cost data are usually skewed and cost comparisons will use a bootstrapped regression model to generate appropriate 95% confidence intervals around the cost differences.

Analysis

Cost-effectiveness will be assessed (from health and societal perspectives) by combining the cost data with the change score on the IBS-SSS and QALYs. The latter will be generated from the EQ-5D combined with UK-specific tariffs. Area under the curve methods, controlling for baseline utility, will be used to calculate the number of QALYs accrued over the follow-up period. If outcomes are better for one group compared to another and costs lower then it will be defined as being 'dominant'. If outcomes are better and costs are higher then an incremental cost-effectiveness ratio will be generated to indicate the extra cost incurred to achieve an extra point reduction in symptoms or extra QALY. Cost-effectiveness planes will be produced, using 1000 cost and outcome differences (from bootstrapped regression models) for each 2-way comparison to explore the uncertainty around the results. Cost-effectiveness acceptability curves will also be produced using bootstrapped regression models with net benefit values as the dependent variables. The net benefit approach requires an assumption about the value placed on a unit improvement in outcome. For QALYs, a range from £0 to £60,000 will be used, thus including the threshold thought to influence NICE decisions. For the IBS-SSS there is no accepted threshold so a range will be chosen such that the points at which one intervention has a 60%, 70%, 80% and 90% likelihood of being the most cost-effective option can be identified.

Sensitivity analyses will be conducted by changing the intervention costs upwards and downwards by 50%, using minimum wages to value lost production, family care and travel time, and by also using the replacement cost approach to value family care with the cost of a homecare worker used as a shadow price.

Modelling beyond the trial period and making comparisons with other interventions is not in the scope of this project.

C) SCHEDULE OF ASSESSMENTS AND MEASURES

CRF	Recruitment & screening	Baseline assessment	Treatment	3-month assessment	6-month assessment	12-month assessment	Ongoing
Main database							
Screening questionnaire (M)	X						
ROME III (M)	X						
Patient registration form (M)		X					
Eligibility form (M)		X					
Randomisation form (M)		X					
Demographics (LG)		X					
About your IBS (LG)		X					
IBS Symptom Severity Score (LG; M)	X	X		X	X	X	
Work and Social Adjustment Scale (LG; M)		X		X	X	X	
Subject's Global Assessment of Relief (LG; M)				X	X	X	
Hospital Anxiety and Depression Scale (LG; M)		X		X	X	X	
Patient Enablement Questionnaire (LG; M)				X	X	X	
Thoughts on treatment (acceptability) (LG)				X	X	X	
ED-5D (LG)		X		X	X	X	
Client Services Receipt Inventory (LG)		X		X	X	X	
CS-FBD Cognitive Scale (LG) **		X		X	X	X	

B-IPQ for IBS (LG) **		X		X	X	X	
IBS Behavioural Responses Questionnaire (LG) **		X		X	X	X	
Beliefs about Emotions Scale (LG) **		X		X	X	X	
IEE factor of the Emotion Processing Scale (LG) **		X		X	X	X	
Positive and Negative Affect Schedule (LG)		X		X	X	X	
Note review form (M) ***						X	
Adverse events form (M)				X	X	X	X
Drop-out / withdrawal form (M)							X
Therapy database							
Registration form (M)		X					
Treatment session log (M)			X				
End of therapy review form (M) ***			X				
Therapist database							
Therapist registration and details (M)		X					

*LG=LifeGuide; M=MACRO

**** Mediator variables which are not used in analyses discussed in this SAP but will be used in future secondary analyses.**

***** These are not used in statistical analyses, other than those variables used in measuring adherence.**

D) QUALITATIVE ANALYSIS PLAN

A nested qualitative study will explore patients' experiences of treatments. The objectives of this study will be: to identify factors that facilitate or impede adherence to web-delivered and therapist-delivered CBT in this patient group; to provide insight into the quantitative results of this complex trial; to identify social and psychological processes of change that occur during the trial. The qualitative results can thus provide scientific value concerning understanding of change processes and practical value concerning the relative merits of each type of CBT and delivery issues to attend to in any future widespread implementation.

Semi-structured audio-recorded interviews will be conducted at three and 12 months with approximately 17 to 20 participants per arm (i.e. 10% to 12%, sampled purposively to encompass a mix of gender and ages and a range of baseline symptom severity scores). Interviewing participants from each active arm will enable us to identify factors related to adherence and change processes; including participants from the TAU arm will provide insight into the quantitative results. Interviewing the same participants at 3 and 12 months will allow us greater depth to explore change processes over time and the potential to understand better any differences in the quantitative results between 3 and 12 months.

The topic guides comprise a series of open-ended questions and prompts used by the interviewer to elicit participants' experiences of, reflections on, and thoughts and feelings about the trial within the broader context of managing IBS. The 3 month interviews explore participants' experiences of taking part in the trial and the treatment they were allocated to, their experiences of other treatments for IBS, and thoughts about the future. The 12 month interviews explore participants' reflections on taking part in the trial and the treatment they were allocated to, any other treatments they have tried since the 3 month interview, and thoughts about the future.

Interviews will be transcribed verbatim with identifying details (e.g. names) removed. Rigorous qualitative analysis techniques will be employed to ensure the objectives are addressed. Analysis will begin on completion of the first few interviews and will proceed iteratively; this will allow early insights or puzzling findings to be explored more fully in later interviews and for improvements to be made if necessary to the topic guide and interviewing technique. An inductive thematic analysis employing supplementary techniques from grounded theory[65,66] will be used to code the data and to identify themes that capture key concepts and processes. We will follow the thematic analysis procedure outlined by Braun and Clarke[67], moving (as recommended) backwards and forwards through the phases, rather than approaching the analysis in a linear fashion. We will supplement these phases with techniques from grounded theory as follows.

Phase	Thematic analysis	Supplementary techniques
1	Familiarization with the data through reading and re-reading transcripts	Listen to audio-recordings
2	Generate initial codes	Line-by-line open coding on a portion of the data; constant comparison

3	Searching for themes	Constant comparison; identifying key concepts in the data; write memos
4	Reviewing themes for fit with coded extracts and entire dataset; generate a thematic 'map'	Constant comparison; search for negative/deviant cases; generate case summaries for individual interviewees to capture whole stories and changes across the 3 month and 12 month interviews
5	Defining and naming themes and their inter-relations	Constant comparison
6	Reporting – select compelling examples, final analysis and contextualisation with the literature and research objectives	Identify the limits of the analysis

After identifying the main themes we will use a mixed methods approach to explore the relationship between the qualitative themes and two aspects of the quantitative data: adherence to the interventions and intervention outcomes. This will involve mapping the themes against adherence rates and interventions, as illustrated in the grid:

	Theme 1	Theme 2	Theme 3	Theme 4…
Adherent				
Non-adherent				
TAU				
LIBT				
TCBT				

In addition to the analytic procedures described above, we will take the following steps to enhance the trustworthiness of the analysis: multiple researchers will contribute to the analysis to avoid producing idiosyncratic interpretations; a 'member check' will be conducted whereby interviewees will be invited to comment on summaries of their interviews; an audit trail will be produced to enhance transparency, including memos and a coding manual; field notes will be written after each interview to capture initial impressions and non-verbal/contextual observations.

Amendments to SAP

List here any amendments to the SAP that were made after the document was signed off by the TSC.

RH	Nov 2016	section 3.1.1	NNT assumptions were clarified.
RH	Nov 2016	section 2.7	HADS and PEQ scoring sections were clarified
RH	Nov 2016	section 3.1.1	Additions were made to missingness of data section.
RH	Nov 2016	section 3.1.2.1	Analysis of acceptability of treatment added.
RH	Dec 2016	section 2.6	Unblinding point was stated.
RH	Nov 2016	General	Correct typos.
RH	Nov 2016	Section 3.1.3	Added sensitivity analysis on timely completion of outcomes.
RH	Nov 2016	Section 3.1.2/3/4	Re-arranged content to improve readability and clarity.
RH	Nov 2016	Section 2.4	Clarified procedure for determining baseline predictors of missingness
RH	Nov 2016	Section 3	Removed exploratory moderator analysis as agreed by the TSC July 2014.
RH	Nov 2016	Section 2.2	Clarified which variables will be used.
RH	Nov 2016	Section 3.1.1	Added treatment and variables interaction.
RH	Nov 2016	Section 1.8	Included details of sample size update to account for re-estimated drop-out rate.
RH	Nov 2016	Section 1.7	Clarified types of variable and assessment time points.
RH	Nov 2016	Section 1.6	Added definition of visit windows
RH	Nov 2016	Section 1.5	Detail on procedures added.
RH	Nov 2016	Section 1.2	Added details on blinding procedure.
RH	Nov 2016	Front page	Updated name of statistician
RH	Nov 2016	Last page	Removed NCR note. Will be kept in File Notes and NCRs document.

ACTIB: Assessing Cognitive Behavioural Therapy in Irritable Bowel

AE: Adverse event

B-IPQ: Brief Illness Perception Questionnaire

BES: Beliefs about Emotion Scale

BRQ: Behavioural Responses Questionnaire

CACE: Complier average causal effect

CBT: Cognitive behavioural therapy

CS-FBD: Cognitive Scale for Functional Bowel Disorders

CSRI: Client Services Receipt Inventory

EQ-5D: European Quality of life 5-Dimension scale

HADS: Hospital Anxiety and Depression Scale

IBS-SSS: Irritable Bowel Syndrome – Symptom Severity Scale

IEE: Impoverished Emotions Experience

KCTU: King's Clinical Trials Unit

LIBT: Lower intensity web-based cognitive behavioural therapy

LifeGuide: An open source platform that is being used for online interventions and data entry and storage.

MAR: Missing at random

NNT: Number needed to treat

PANAS: Positive and Negative Affect Schedule

PEQ: Patient Enablement Questionnaire

QALY: Quality adjusted life year

Regul8: A CBT-based self-management website for IBS developed specifically for this study using LifeGuide.

SAE: Serious adverse event

SGA: Subject's Global Assessment of Relief

TAU: Treatment as usual

TCBT: Therapist-delivered cognitive behavioural therapy

TSC: Trial steering committee

WASAS: Work and Social Adjustment Scale

IAPT: Improving Access to Psychological Therapies