A Randomised Controlled Trial of supported online self-management for symptoms of fatigue, pain, and urgency/incontinence in people with inflammatory bowel disease – the IBD-BOOST trial.

Statistical Analysis Plan

Version: 1.0
Date: 18/08/2023

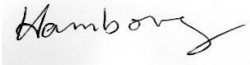
Person(s) contributing to the analysis plan	
Name(s) and position(s)	Fionn Cléirigh Büttner, Trial Statistician Thomas Hamborg, Senior Statistician Christine Norton, Chief Investigator Rona Moss-Morris, Co-Investigator, RCT co-lead. Laura Miller, Programme Manager
Authorisation	
Position	**Chief or principal investigator**
Name	**Christine Norton**
Signature	
Date	18/08/2023
Position	**Senior trial statistician**
Name	**Thomas Hamborg**
Signature	
Date	18/AUG/2023
Position	**Independent statistician**
Name	**Stephanie MacNeill**
Tick once reviewed	**X**
Date	18/08/23

Table of Contents

1. Administrative Information

1.1 Trial registration number: ISRCTN 71618461

This SAP is based on protocol version 6.0 (date: 21/03/2022)

1.2. SAP revision history

Protocol version	Updated SAP version no.	Section number changed	List of changes from previous version/protocol	Author of change	Date
4.0	0.1		New Document	Evangelia Tzorovili	24.05.2021
4.0	0.2		Updated appendices' tables and subgroup analysis	Evangelia Tzorovili	26.05.2021
4.0	0.3		Updated the text and tables based on input from CN and LM.	Evangelia Tzorovili	01.06.2021
4.1	0.4			Evangelia Tzorovili	02.08.2021
5.1	0.5		Updated all sections.	Evangelia Tzorovili	12.04.2022
5.1	0.5			Evangelia Tzorovili	20.04.2022
6.0	0.6			Thomas Hamborg	26/07/22
6.0	0.7	6.0	Primary analysis Mediation analysis Subgroup analysis Sensitivity analysis	Fionn Cléirigh Büttner	26/05/2023
6.0	0.8	1.8 6.0 All sections	Abbreviations Statistical analysis Document-wide editing	FCB & TH	06/07/2023
6.0	0.9	6.09 6.10 6.11	Random intercept for site removed from mediation analysis: CACE analysis incorporated. MAR text incorporated in step 1 of MNAR sensitivity analysis	FCB	13/07/2023

		7.02	Text about "causality" removed from safety/adverse events.		
6.0	0.10	All sections 4.2	Incorporate CI and independent statistician's comments. Move relevant text to appendix 3.	FCB	07/08/2023
6.0	0.11				
6.0	1.0				

*If the SAP has been published, indicate which version.

1.3. Members of the writing committee

The statistical analysis plan (SAP) writing committee comprises Fionn Cléirigh Büttner (FCB) and Thomas Hamborg (TH). Sally Kerry devised the initial design strategy. Evangelia Tzorovili contributed to earlier drafts. Input was provided by Christine Norton (CN), Rona Moss-Morris (RMM), and Laura Miller (LM). FCB and TH are primarily responsible for writing and implementing the statistical analysis strategy.

1.4. Timing of statistical analysis

The statistical analysis is conducted once the SAP has been signed off and the last participant has completed the last (12-month) follow-up case report form.

1.5. Timing of SAP revisions in relation to unblinding of data/results

All members of the writing committee will be blinded to trial arm allocation until the statistical analysis plan is signed off. FCB and TH will access blinded data (with trial arm assignment concealed and potentially-unblinding variables omitted) during SAP preparation and therefore before SAP sign off.

1.6. Analysis software

Statistical analyses and data presentation described in this document will be performed using Stata version 17.0 unless otherwise specified.

1.7. Remit of SAP

The document provides details of statistical analyses and presentation of results to be reported within the principal paper(s) of the IBD-BOOST randomised control trial (RCT). It specifies the statistical analysis of quantitative data for primary and secondary outcomes, mediation analysis, subgroup analyses, and sensitivity analyses, as well as

the analysis of the SWAT. Health economic analyses and the Process Evaluation are addressed in separate documents.

1.8. Abbreviations

AE	Adverse Event
CACE	Complier Average Causal Effect
CAU	Care As Usual
CBT	Cognitive Behavioural Therapy
CI	Chief Investigator
CRF	Case Report Form
FI	Faecal Incontinence
IBD	Inflammatory Bowel Disease
IBS	Irritable Bowel Syndrome
MAR	Missing at random
MNAR	Missing not at random
PCTU	Pragmatic Clinical Trials Unit (QMUL)
QMUL	Queen Mary University of London
RCT	Randomised controlled trial
SAE	Serious Adverse Event
SWAT	Study within a trial

2. Introduction

Inflammatory Bowel Disease (IBD) affects 300,000 people in the UK (Crohn's & Colitis UK, 2016), causing unpredictable bouts of gut inflammation with acute illness, diarrhoea, and pain. In remission, many people with IBD live with fatigue, chronic abdominal pain, and bowel urgency/incontinence (1). There is no current cure for IBD, which usually starts in childhood or as a young adult. Although most IBD research focuses on controlling inflammation, many patients report persistent IBD-related symptoms such as fatigue, abdominal pain, and difficulty with continence, even when IBD is in remission (1-3). These symptoms have a major impact on quality of life in people with IBD but have been largely ignored by clinicians and researchers.

IBD-BOOST is a four-stage programme grant aiming to improve the quality of life of people with IBD by reducing the burden of IBD-related fatigue, abdominal pain, and urgency/incontinence. The fourth stage of the programme is an RCT of online self-management for IBD-related symptoms of fatigue, pain, and urgency/incontinence, with an embedded pilot study, a study within a trial, a health economics evaluation, and a process evaluation. This RCT will investigate the effectiveness of a remotely delivered, self-management programme using the principles of Cognitive Behavioural Therapy (CBT) compared with standard care. The trial will provide evidence for the effectiveness of a self-management programme delivered online to improve quality of life for patients with IBD-related symptoms of fatigue, pain, and urgency, enabling clinicians and patients to make informed decisions regarding management. Eligible participants will have completed the IBD-BOOST survey (stage II). Some participants will have also participated in stage III (IBD-BOOST OPTIMISE) – a cohort study to optimise medical management of these symptoms). See Appendix 1 for a flow diagram of the IBD-BOOST programme studies.

The current document describes the SAP for the RCT of the IBD-BOOST programme (Stage IV). In accordance with good clinical practice, all members of the writing committee will be blinded to trial arm allocation until the statistical analysis plan is signed off. FCB and TH will access blinded data (with trial arm assignment concealed and potentially-unblinding variables omitted) during SAP preparation. The SAP also describes the analysis of the study within a trial (SWAT) assessing the impact of two different patient information leaflets on recruitment rates.

3. Background and trial design

Study objectives	In individuals with IBD who (i) report symptoms (i.e., ≥5/10 for the impact of one or more symptoms on an 11-point (0-10) scale), and (ii) express a desire to receive intervention, does an individually tailored, facilitator-supported, online, self-management programme for fatigue, pain, and faecal urgency/incontinence improve IBD-related quality of life and symptom relief six months after randomisation compared with usual care?
Study design	A pragmatic, multi-centre, two-arm, parallel-group, superiority RCT (with an internal pilot) of facilitator-supported, online, self-management versus care as usual (CAU) to manage symptoms of fatigue, pain, and faecal urgency/incontinence in IBD patients.
Setting	Recruitment of respondents to a previous IBD-BOOST programme survey (who were recruited via 17 NHS Trusts, the UK IBD BioResource, Crohn's & Colitis UK, and social media). Four NHS Trusts were specifically opened for the RCT: 1. London Northwest University Hospital NHS Trust 2. Nottingham University Hospitals NHS Trust 3. The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust 4. St Helens & Knowsley Teaching Hospitals NHS Trust All interventions were delivered remotely (i.e., online and by telephone).
Participants	Individuals wanting interventions for IBD-related symptoms of fatigue, pain, and/or urgency who were recruited from (i) the IBD-BOOST survey (Stage II) respondents or (ii) participants completing both the IBD-BOOST survey AND the medical symptom optimisation study (IBD-BOOST Optimise – Stage III) where the impact of at least one of these three symptoms was scored ≥5 on an 11-point (i.e., 0-10) symptom scale. Patients with IBD who meet the following: **Inclusion criteria**: • Diagnosis of IBD (self-reported as having been medically diagnosed with IBD including patients with an ileo-anal pouch or stoma) • 18 years old or older • Living in England, Scotland, or Wales • Have participated in stage II of the programme (IBD-BOOST survey) and have rated the impact of one or more IBD-related symptoms of fatigue, pain, or urgency/incontinence on their quality of life as 5 or more on a 0-10 scale when completing IBD-BOOST programme stages II (IBD-BOOST survey) or III (IBD-BOOST OPTIMISE), whichever is the more recent • No "red flags"– see below

	• Access to the online intervention via a computer or mobile device **Exclusion criteria**: • One or more "red flags" identified on pre-randomisation screening, (such as new bleeding, rapid weight loss, or vomiting) self-reported on a screening checklist • Inability to give informed consent (i.e., due to reduced mental capacity) • Insufficient command of English • No access to online materials
Interventions	Intervention arm Access to care as usual (CAU) AND (i) an online, individually tailored, interactive, self-management programme for six months (IBD-BOOST), (ii) one telephone or Skype support session for up to 30 minutes with a health care professional intervention facilitator who received training and monthly supervision from the trial team, AND (iii) access to online messaging with the intervention facilitator via the IBD-BOOST platform for the initial three months after recruitment. Control arm CAU, including usual monitoring at routine or requested clinic visits and/or via the local IBD helpline, and care from their general practitioner. Use of services outside of the trial was monitored by the IBD-Resource use questionnaire (see outcome measure below).
Primary outcome measure(s)	UK Inflammatory Bowel Disease Questionnaire (UK-IBDQ) and global rating of symptom relief at six months after randomisation (multiple primary end points).

4. Trial objectives & Outcome measures

4.1 Trial objectives

4.1.1 Primary research question

1. In individuals with IBD who report IBD-related fatigue and/or pain and/or urgency (defined as ≥5/10 impact of one or more symptoms on a 0-10 scale) and express a desire to receive relevant intervention, does an individually tailored, facilitator-supported, online, self-management programme for fatigue, pain, and faecal urgency/incontinence improve IBD-related quality of life or a global rating of symptom relief six months after randomisation compared with usual care?

4.1.2 Secondary research questions

2. Is there any difference between intervention and CAU groups in severity of fatigue, pain, and urgency/incontinence symptoms at six and 12 months after randomisation?
3. Is there any difference between intervention and usual care groups in IBD-related quality of life and global rating of symptom relief 12 months after randomisation?
4. Does prior medical optimisation of symptoms (in Stage III of the IBD-BOOST programme) moderate the treatment response as measured by the primary outcomes? That is, do participants who received medical optimisation benefit more from the intervention than participants who did not?
5. Do individuals with inactive IBD (defined as faecal calprotectin <200µg/g and IBD-control score ≥13) at trial commencement experience a better response to treatment than those with active IBD, as measured by primary outcomes?
6. Does baseline depression or the presence of irritable bowel syndrome (IBS) [using ROME IV criteria] moderate treatment response to intervention, as measured by primary outcomes? That is, do trial participants with depression at baseline or irritable bowel syndrome experience a different response to treatment compared to participants without depression at baseline or IBS?
7. Do changes in (i) cognitive (negative symptom perception, self-efficacy), (ii) behavioural (all-or-nothing and resting behaviour), and/or (iii) emotional (visceral anxiety and depression) responses after randomisation mediate the relationship between intervention and the primary outcomes at six months after randomisation?
8. Is an individually tailored, facilitator-supported, online, self-management programme for fatigue, pain, and faecal urgency/incontinence in IBD cost-effective (i.e., explored in health economics analysis)?
9. What are patients' expectations and experiences of the intervention and what factors may have influenced intervention implementation (i.e., explored in process evaluation)?

The analysis strategy for research question seven will be addressed in this statistical analysis plan but will be reported in a distinct research article, separate from primary outcomes and effect modifiers. Research questions

eight and nine, as health economic and process evaluation objectives, respectively, will not be addressed in this statistical analysis plan.

4.2 Outcome measures

4.2.1 Primary outcomes (at six months after randomisation)

1. **UK Inflammatory Bowel Disease Questionnaire** (4)
   The UK Inflammatory Bowel Disease Questionnaire (IBDQ) is a validated UK version of the McMaster IBDQ that assesses the impact of IBD-related symptoms or daily activities on quality of life. The UK IBDQ contains 30 items, each scored from 1 (i.e., best response) to 4 (i.e., worst response). The score is the sum of all individual items and ranges from 30 to 120. A higher score indicates a poorer quality of life [continuous].

   The UK IBDQ has five sub-domains: (i) bowel movements and use of facilities, (ii) general bowel symptoms, (iii) emotional, (iv) social, and (v) systemic function (mainly fatigue). If >50% of UK IBDQ items on one or more sub-domains are missing, a missing overall score will be produced. If ≤50% of UK IBDQ items are missing, the mean of complete responses within the relevant sub-domain will replace missing items. If more than four responses in total are missing the IBDQ will not be scored and set to missing (4, 5).

2. **Global rating of symptom relief** (6)
   The Global Rating of Symptom Relief is a simple eleven-point (0-10) Likert scale that measures participants' perceived change in symptoms during study participation. Higher ratings indicate higher symptom relief. Zero represents "No relief at all" and 10 represents "Completely relieved" [continuous].

4.2.2 Secondary outcome measures (all at 6 months and 12 months follow-up unless otherwise stated):

1. **UK-IBDQ** at 12 months after randomisation [continuous].

2. **Rating of satisfaction with results of IBD BOOST programme**
   Range 0-10. Higher ratings indicate greater levels of participants' perceived satisfaction with outcome from IBD-BOOST RCT [continuous].

3. **Global rating of symptom relief** at 12 months after randomisation [continuous].

4. **Numerical pain rating scale** (7)

   The numerical pain rating scale (NRS) comprises four questions assessing (i) current pain intensity, (ii) lowest pain intensity, (iii) worst pain intensity, and (iv) average pain intensity. Each item is evaluated using a simple, eleven-point (0-10) Likert scale. There are four scores – one for each question. The number that the respondent selects is the respondent's NRS score for that question item. Average pain intensity will be used as the secondary outcome representing this pain construct to estimate the difference between intervention and control arms [continuous]. Other pain constructs are assessed in exploratory analyses (7.2.1).

5. **Vaizey incontinence score** (8)

   The Vaizey Incontinence Score is a seven-item, patient-reported outcome measure that assesses the severity of faecal incontinence by evaluating aspects of bowel control including frequency, type of incontinence, and lifestyle impact. Items are scored on a scale from 0 to 4 or 0 to 2, with higher scores indicating greater severity or frequency of symptoms. The total score ranges from 0 (perfect continence) to 24 (total incontinence) and is obtained by summing individual item scores to providing an overall measure of faecal incontinence severity [continuous] (9).

6. **IBD-Fatigue score** (10, 11)

   The IBD-BOOST RCT used only Section I of the IBD-F Self-Assessment Scale to identify the level and duration of fatigue (four questions). Questions are scored on a five-point (i.e., 0-4) Likert scale, with a total score ranging from 0-16. A score of zero indicates no fatigue. Higher scores indicate higher levels of fatigue. If any items are missing, the total score cannot be computed. Results will include only participants who have answered all four scale items [continuous].

7. **IBD-Control score** (12)

   The IBD-Control-8 sub-score is calculated by summing values for eight out of nine control CRF items resulting in a range of 0–16 (i.e., 0=worst control; 16=best control). The question "Over the past two weeks, have your bowel symptoms been getting worse, getting better, or not changed?" (i.e., question three in the CRF) assesses the stability of IBD-related symptoms. It should be reported separately and not be included in the control-8 score. Note that N/A, denoted "999", is coded "1" for item "Your current treatment is useful in controlling your IBD".

8. **EQ-5D-5L general health-related quality of life** (13)

   The EQ-5D-5L questionnaire assesses participants' health-related quality of life (14). The EQ-5D-5L comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression, each rated on a scale from 1 to 5, corresponding to no problems (1), slight problems (2), moderate problems

(3), severe problems (4), and extreme problems (5). Overall QoL utility scores will be derived for all contributing study participants using the UK National Institute for Health and Care Excellence (NICE) decision support unit EQ-5D scoring algorithm (15). Briefly, this estimation algorithm (i.e., the eq5dmap Stata command) will directly map from individual-specific, EQ-5D-5L, health states to individual-specific, EQ-5D-3L, utility scores, using age and sex as necessary covariates (15). Estimated, individual-specific, EQ-5D-3L utility scores will be used as the secondary outcome during statistical analysis.

The overall score of the EQ-5D-3L index ranges from -0.594 to 1.000. A score of -0.594 represents the worst possible health status while a score of 1.000 represents the best possible health status. A score of 0.000 indicates a health status that is considered as bad as being dead (in terms of quality of life). The absolute minimum score of -0.594 indicates that an individual's health status is worse than being dead because an individual of such health status is not only experiencing significant health problems but is are also experiencing a lower quality of life compared to someone who is deceased. Due to the mapping from 5L to 3L the boundary values cannot be reached and the actual range of possible values is slightly smaller. The EQ-VAS is a patient-reported measure of perceived overall health. It is a continuous measure that ranges from 0-100, with 100 indicating “the best health imaginable” and 0 indicating “the worst health imaginable.” This score requires no further derivation.

The outcome collection timeline is summarised in **Appendix 2**. Derivations of all measurements matched with the Requirements Specification Document (RSD) are summarised in **Appendix 3**.

Unless otherwise stated above, if >20% of scale items are missing, the total/overall score will be set to missing. If ≤20% of scale items are missing, missing values will be imputed using the mean value of the present item for this participant. This approach will be applied to a domain/dimension/subscale (instead of across all items) if the outcome has different domains/dimensions/subscales. The same approach shall be used for mediator variables (section 4.3).

4.3 Putative mediators

Cognitive responses

The Brief Illness Perceptions Questionnaire (BIPQ) (minus the open-ended causal items) measures IBD-/illness-specific symptom cognitions. It is an 8-item scale, with items rated on an 11-point Likert scale (16). Each item of the BIPQ assesses one dimension of illness perceptions including consequences, timeline, personal control, treatment control, identity score, coherence score, emotional representation, and illness concern. This reflects a combination of emotional and cognitive representations scored.

To calculate a composite BIPQ score, the individual 8 domain scores are summed together (the personal control (3), treatment control (4), and coherence (7) items are reverse scored, as higher scores in these elements

represent positive illness perceptions). A higher BIPQ score indicates a greater perceived psychological burden of illness (range 0–80) (16).

The Self-Efficacy of Managing Chronic Diseases Scale (SEMCD) is a measure of how confident patients with chronic disease are in doing certain activities. The measure consists of 6 items that are rated on a 10-point scale ranging from "not at all confident" (1) to "totally confident" (10). A mean of the 6 items is calculated, where higher scores indicate greater self-efficacy for managing the chronic condition (17).

Behavioural responses

The Cognitive and Behavioural Responses to Symptoms Questionnaire (CBRQ) assesses patients' cognitive and behavioural responses to symptoms. Two subscales were measured: 1) All-or-nothing behaviour (5-25*), e.g. I tend to overdo things and then rest up for a while* and (2) Avoidance/resting behaviour (8-40), e.g. *When I experience symptoms, I rest.* Higher scores indicate higher All-or-nothing or Avoidance/resting behaviour (18). When computing the total score, a threshold of 33% missing items will be set. If >33% of items on either CBRQ subscale have missing values, that overall subscale score will be set to missing. Otherwise (i.e., ≤33% missing items), missing item-level values will be imputed using the mean of present values for that item (18).

Emotional responses

The Patient Health Questionnaire-9 (PHQ-9) is scored from a 9-item questionnaire assessing depression symptom severity (19). The total score is the sum of nine items and ranges from 0-27. For each item, responses range from 0-3 (i.e., not at all [0]; several days [1]; more than half the days [2]; nearly every day [3]). A score of 0-4 indicates no depression, 5-9 indicates mild depression, 10-14 indicates moderate depression, 15-19 indicates moderately severe depression, and 20-27 indicates severe depression (19). Items with up to two missing values are scored. If cases with more than two missing values across PHQ-9 items exist, missing values will be imputed using the mean of present values for that item (20).

The Visceral Sensitivity Index (VSI) measures gastrointestinal symptom-specific anxiety using 15 questionnaire items, with responses ranging from 1 (i.e., strongly agree) to 6 (i.e., strongly disagree) (21). Items are scored on a reversed 6-point Visceral Sensitivity Index (VSI) for anxiety scale ranging from 0 to 5, with sum scores between 0 and 75. Higher scores indicate more severe symptom-specific anxiety [continuous] (21).

5. Study methods

5.1 Sample size calculation (obtained from protocol v6.0)

The primary outcome UK-IBDQ, ranges from 30 to 120 where low values indicate poor quality of life. Using several published studies (4, 22, 23), we estimate the standard deviation of the change in score to be between 20 and 30. This would mean a standardised effect size of 0.3 would equate to a difference of between 6 and 9 points on the scale. In the validation study (4), the difference in score between those with mild disease and disease in remission was 12 points and the effect of relapse was 10 points.

An effect size of 0.3 was observed in a small study of dietary advice in patients with ulcerative colitis and deemed to be the minimal clinically important difference (MCID) for IBDQ (22). Based on the above considerations, an original sample size calculation estimated that 680 participants would need to be recruited to the trial and the original funding and ethics approval was on this basis. However, during the study the covid-19 pandemic made our anticipated recruitment of 30 facilitators from 20 NHS sites impossible. During team discussions it also became evident that we had not made appropriate statistical adjustments for having two primary end points. The sample size calculation was therefore adjusted before the end of recruitment as follows:

A minimum of 740 participants are going to be randomised, approximately 370 to each group. This allows the MCID difference to be detected with 86.4% power at a 2.5% significance level. It is anticipated that 16 facilitators are participating in the trial. Taking account of a facilitator effect (assuming a facilitator intraclass correlation of 0.04) in the intervention arm, 352 participants are required in each study arm to achieve 86.4% power (21 participants per facilitator). The sample size is decreased by a deflation factor of 0.84 assuming that baseline values of the outcome measure are predictive of post-treatment values (correlation 0.4) and inflated to account for 20% loss to follow-up resulting in the final recruitment target of 740.

The 20% drop-out assumption is based on drop-out rates from previous studies of self- management: 19.2% of 682 participants in IBD disease self-management (not online) (24); 20.4% of 333 participants for online self-management of Ulcerative Colitis disease flares (25); 18% control and 16% intervention of 1140 participants randomised for chronic disease self-management in other diseases (26). Adjustment for correlation between baseline and follow-up values of the primary outcomes is based on Walters et al who suggest a median correlation of a QoL measure with 6 months post randomisation outcome of 0.5 with a lower IQR bound of 0.41 (27). Being conservative a correlation of 0.40 is assumed.

5.2 Randomisation procedure

Participants who consent, are eligible and return the baseline questionnaire will be randomised by the central research team using an online randomisation system developed for the study by the PCTU. Participants will be

randomised using stratified blocked randomisation with block sizes of four and six and a one-to-one (i.e., 1:1) allocation ratio. The central team will then inform the participant which group they are in and inform the clinical sites of participants in the intervention group who will receive local facilitator support (phone call and online messaging). The local facilitator will be given the participant's details and access to their online tasks.

Stratification factors are:

- diagnosis (Crohn's disease vs. other IBD)
- whether or not participated in Stage 3 study (medical symptom optimisation)

5.3 Blinding (obtained from protocol v6.0)

Faecal calprotectin level will be entered into the database by a person blinded to group allocation when the result has been returned from the laboratory and the participant informed of the result. Blinding of participants or facilitators is impossible. The trial steering committee, CIs, health economics and statistics teams are blinded, that is, will not see results broken down by treatment arm during the trial or any time prior to analysis plans being signed off. Final analysis will occur once all follow up data is collected, the final statistical analysis plan has been signed off and data cleaning has occurred.

6. Analysis methods

6.1. Baseline characteristics

Baseline demographic characteristics, clinical characteristics (type of IBD classification, Rome IV IBS criteria), and baseline values of outcome variables will be presented for both study arms using descriptive statistics only. Descriptive statistics including smoking habits and alcohol consumption will be imported from the Survey dataset (Stage II of the IBD-BOOST research programme) because these variables are not included in the baseline evaluation of the RCT. Independent and outcome variable values will be summarised using the mean and standard deviation or median and interquartile range for continuous variables, as appropriate, and the absolute frequency and column-wise percentage for categorical variables. See Tables 1-3 in **Appendix 4**.

6.2. Adherence to treatment

The proportion of participants randomised to the intervention who (i) clicked on the link to the online intervention, (ii) commenced the intervention, AND (iii) completed a minimum of four online sessions will be considered to have adhered to the intervention. No applicable guideline for defining adherence was identified. The chosen criteria are based on trial team consensus and the definition for the ACTIB CBT intervention in the IBS trial (28). Adherence to the intervention will be assessed as a potential moderator of the treatment effect in a complier-averaged causal effect (CACE) analysis (section 7.1.1).

6.3. Information for CONSORT flow diagram

A dummy flow diagram is provided in **Appendix 5**.

6.4. General analysis principles

Statistical analyses will be performed according to the intention-to-treat (ITT) principle – that is, all randomised participants with a recorded outcome will be included in the analysis and analysed according to the trial arm to which they were randomised. Trial participants who withdraw consent for their data to be included in the analysis will be excluded from all statistical analyses. For the analysis of the primary outcomes and each secondary outcome, we will present the following information:

6. The number of participants included in each analysis, by treatment arm
7. A summary statistic of the outcome (e.g., number (%)), by treatment arm
8. The estimated treatment effect θ.
9. A 95% confidence interval for the estimated treatment effect.
10. A p-value for a two-sided hypothesis test of H_0: Θ=0.

6.5. Analysis of the primary outcome

The primary outcome analysis will compare (i) UK-IBDQ and (ii) global rating of symptom relief between the intervention arm and control arm six months after randomisation using a three-level, repeated measures, mixed-effects model that accounts for correlation of post-randomisation outcome measures within patients and clustering of patients within intervention facilitators in the intervention arm only. A partially nested mixed-effects model with heteroskedastic error terms, based on the model described by (29), will be fitted with the Satterthwaite approximation for degrees of freedom to avoid upward bias of the type I error rate (30). The clustering effect (of patients nested within intervention facilitators) will be modelled only in the intervention arm. Participants in the control arm will be treated as independent. In the intervention only, a random slope will be specified to allow the effect of treatment to vary between intervention facilitators. An unstructured covariance matrix will be used for the residual errors of repeated measures over time. An interaction effect for randomised treatment group and post-randomisation time point (as a categorical variable) will be fitted to achieve a saturated model that allows estimation of mean estimates at each time point in each treatment group.

Specifically, y_{ij} will be the continuous outcome for the *i*th individual participant receiving the intervention form the *j*th facilitator, *t* will be the intervention indicator (0 = control, 1 = intervention), θ will be the intervention effect, β_0 is an intercept term, and β_k represents other model covariates, and p_l is the effect of time at the lth post-randomisation timepoint:

$$y_{ij} = \beta_0 + \theta t_i + (\theta t_{ij} \times p_l) + \beta_k + u_j t_{ij} + r_{ij}(1 - t_{ij}) + \epsilon_{ij} t_{ij}$$

where $u_j \sim N(0,\sigma^2_u)$ is a random-effects term representing between-cluster variation in the clustered intervention arm, $r_{ij} \sim N(0,\sigma^2_r)$ represents individual-level variation in the non-clustered control arm, and $\epsilon_{ij} \sim N(0,\sigma^2_\epsilon)$ represents individual-level variation in the clustered intervention arm.

The following covariates will be included in the primary outcome analysis model as fixed effects:

(i) baseline value of outcome measure, included as a continuous covariate (UK-IBDQ analysis model only).

(ii) stratification factors (i.e., diagnosis type and whether participants participated in the Stage III) as binary covariates,

(iii) PROMIS fatigue, PROMIS pain, and PROMIS incontinence at baseline as continuous covariates,

(iv) participant age (continuous variable), and

(v) participant gender (categorical variable).

The code used to fit this model in Stata is provided in the first row of the table below. The estimated treatment effect is obtained via the estimation command `lincom 1.trt`

The type I error rate will be adjusted for multiple testing by applying a Bonferroni correction that divides the family-wise type I error rate (i.e., 0.05) by the number of primary outcomes (i.e., two), yielding an adjusted type I error rate of 0.025. The repeated measures, mixed-effects model imputes missing outcome values implicitly under the MAR assumption. Sensitivity analyses (section 6.10) will be conducted that assess the robustness of the primary analysis by imputing missing outcome data explicitly under Missing Not At Random (MNAR) assumptions. Missing data for baseline covariates to be included in the analysis models will be accounted for using mean imputation for continuous variables and inclusion of a 'missing' category for categorical variables (31).

Strategy for analysis of primary and secondary outcomes if model fails to converge.

If the analysis described above (i.e., row 0 in the below table) fails to converge for any outcome, the following sequential strategy will be employed for assessing between-group differences.

	Change from previous strategy	Example Stata code
0	Primary analysis	`mixed y i.time##treat y_b i.diagnosis_n /// i.opti p_b f_b i_b age i.gender || /// facilitator:treat || id: , /// nocons reml /// residuals(unstructured, t(time)by(treat)) /// dfmethod(sat)`
1	Remove fixed effects for pain, fatigue and incontinence	`mixed y i.time##treat y_b i.diagnosis_n /// i.opti age i.gender || /// facilitator:treat || id: , /// nocons reml /// residuals(unstructured, t(time)by(treat)) /// dfmethod(sat)`
2	Remove clustering of participants by intervention facilitators	`mixed y i.time##treat y_b i.diagnosis_n /// i.opti p_b f_b i_b age i.gender || /// id: , /// nocons reml /// residuals(unstructured, t(time)) /// dfmethod(sat)`
3	Remove clustering of participants by intervention facilitators and fixed effects for pains, fatigues and incontinence	`mixed y i.time##treat y_b i.diagnosis_n /// i.opti age i.gender || /// id: , /// nocons reml /// residuals(unstructured, t(time)) /// dfmethod(sat)`
4	Remove other covariates in the order gender, age, optimisation, diagnosis	`mixed y i.time##treat y_b ||/// id: , /// nocons reml /// residuals(unstructured, t(time)) /// dfmethod(sat)`
5	Fit simple between group t-test for 6 months follow-up outcome.	`regress y6_m treat`

y= outcome at 6m and 12m, y_b= outcome at baseline (UK-IBDQ only), treat = intervention arm indicator, therapist = intervention group therapist indicator, id=participant indicator, diagnosis_n=diagnosis type, optimis=whether patient participated in OPTIMISE, p_b: pain (PROMIS) at baseline, f_b: fatigue (PROMIS) at baseline, i_b: incontinence (PROMIS) at baseline

6.6. Analyses of secondary outcomes

Secondary outcomes will be analysed using the same three-level, partially nested, repeated measures, mixed-effects model as for the primary outcome. Distributional assumptions will be assessed for each secondary outcome analysis and an appropriate outcome transformation performed if necessary to fulfil model assumptions. The same stepwise strategy for simplifying the analysis model as for the primary outcome analysis will be employed should the model fail to converge. If reduced models are used for any outcomes this will be clearly stated in the statistical analysis report. Estimates of treatment effects at 12 months follow-up from the model will be obtained using the following Stata command:

```
lincom 1.treat+2.time#1.treat
```

6.7. Interim analyses

No interim analyses are planned.

6.8. Subgroup (i.e., moderation) analyses

Subgroup analyses will be performed for the primary outcomes to assess whether the effect of the intervention differs in pre-specified subgroups defined by baseline characteristics. The subgroup analysis will be performed using the same analysis model as for the primary outcome, adding an interaction term between the baseline characteristic and treatment arm. The presence of an interaction will be tested using a likelihood ratio test comparing the sub-group analysis model, including the interaction effect, and the primary analysis model, not including the interaction term. The test will be considered significant at the 5% level. All patients with complete outcome data will be included in the subgroup analysis. For each subgroup category, we will report summary statistics of the outcome by treatment arm, with treatment effect estimates and 95% confidence intervals. A p-value for the interaction test will also be reported (Table 11).

Separate models will be constructed for each subgroup treatment effect – that is, multiple interactions terms will not be included in the same model. We will assess whether treatment effects vary among levels of the following baseline characteristics:

1. **IBD in remission or not**

   IBD remission at baseline is defined as faecal calprotectin <200µg/g AND an IBD control score ≥13. Participants need to satisfy both to be defined as in remission. This subgroup analysis will investigate whether trial participants who are in remission experience a different response to the individually tailored, online, self-management intervention compared with trial participants who are not in remission.

2. **PHQ-9 measure of depression** at baseline (19)
   - Not depressed at baseline (i.e., PHQ-9 = 0-9)
   - Depressed at baseline (i.e., PHQ-9 score = 10-27)

This subgroup analysis will investigate whether trial participants with depression at baseline experience a different response to the individually tailored, online, self-management intervention compared to trial participants who are not depressed at baseline.

3. **Visceral Sensitivity Index (VSI)** (21)

The VSI measures gastrointestinal, symptom-specific anxiety using 15 questionnaire items, with responses ranging from 1 (i.e., strongly agree) to 6 (i.e., strongly disagree). The VSI measures unique aspects of fear, anxiety, and hypervigilance that can accompany misappraisals of visceral sensations and discomfort. Items are scored on a reversed, 6-point scale ranging from 0 to 5 (i.e., 1–6 becomes 5–0), with sum scores between 0 and 75. Higher scores indicate more severe, symptom-specific anxiety [continuous]. This subgroup analysis will investigate whether trial participants with symptom-specific anxiety at baseline experience a different response to the individually tailored, online, self-management intervention compared to trial participants who do not have symptom-specific anxiety at baseline.

4. **Rome IV criteria for IBS met at baseline or not** (32)

The Rome IV criteria are used for the diagnosis of IBS. IBS is present when the following items from the RSD are scored as 1:

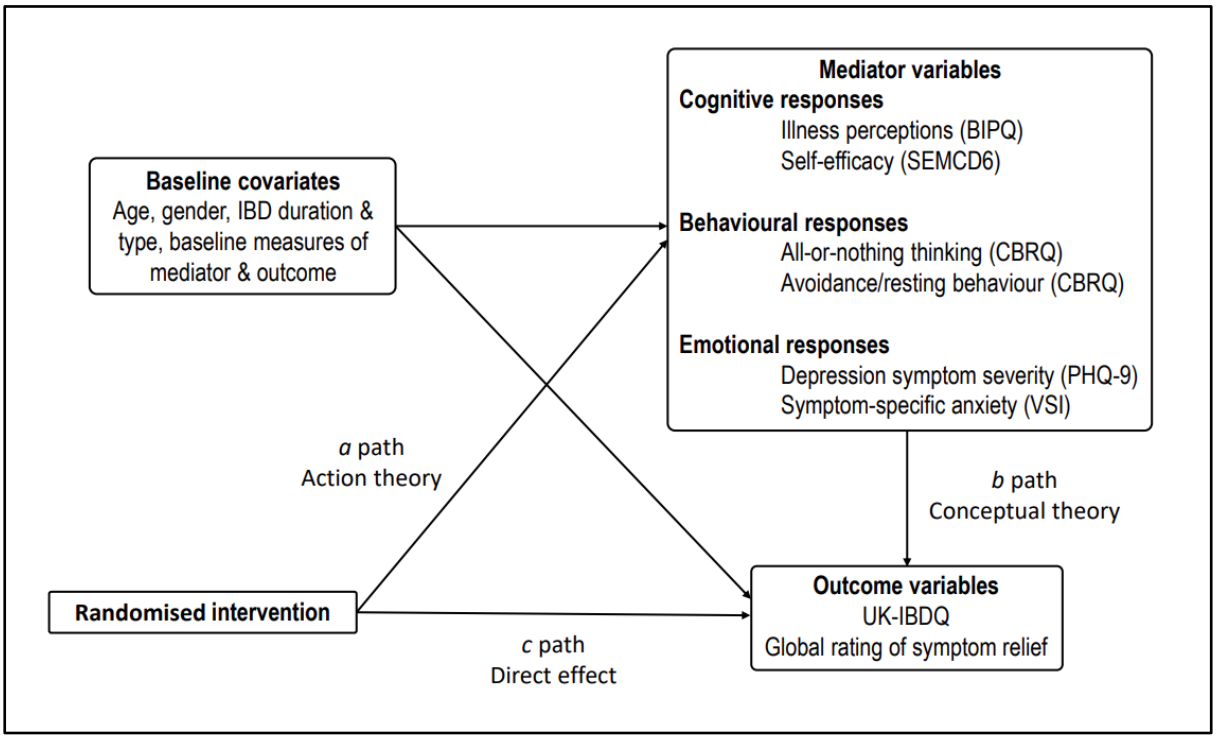
- **ibs1:** *"Do you get abdominal (tummy) pain on a weekly basis?"* **and**
- **ibs5:** "*Have you had these symptoms for at least 6 months?"* **and 2 or more of**
- **ibs2:** *"Does this pain relate in some way to opening your bowels?"*
  ibs3: "*Do your bowels change in frequency when you get this pain?"*
  ibs4: *"Do your stools change in appearance (softer, harder) when you get this pain?"*

This subgroup analysis will investigate whether trial participants with IBS at baseline respond differently to the individually tailored, online, self-management intervention compared to trial participants who without IBS at baseline.

6.9. Mediation analysis

A mediation analysis will be performed to assess whether the effect of random treatment allocation on the primary outcomes (i.e., UK-IBDQ and Global Rating of Symptom Relief) is mediated by mediators of interest (Figure 1). A multi-level, structural equations model (SEM) will be constructed by fitting explanatory, mediating, and outcome variables in a single mediator analysis to estimate natural direct, natural indirect, and total intervention effects (33). The aim of this analysis is mediation and only a 'structural' model – an analysis model with paths reflecting causal dependencies between endogenous and exogenous variables – with observed variables will be fitted (34). No 'measurement' model, no latent variables, and no correlated errors will be required or specified within the SEM framework. Random treatment allocation will be specified as an exogenous variable, and mediators and both primary outcomes will form endogenous variables.

Figure 1. Causal mediation path diagram including cognitive, behavioural, & emotional responses as mediators



IBD, inflammatory bowel disease; *BIPQ*, brief illness perceptions questionnaire; *SEMCD6*, self-efficacy for managing chronic disease; *CBRQ*, cognitive & behavioural responses questionnaire; *PHQ-9*, patient health questionnaire; *VSI*, visceral sensitivity index; *IBDQ*, inflammatory bowel disease questionnaire
Indirect effect: a path x b path (where mediators and outcomes are continuous variables)

Mediator and outcome variables, and the amount of corresponding missing data, will be summarised using mean and standard deviation, or frequency and percentage, as appropriate. Baseline and follow-up mediator and outcome variables will be standardised to baseline by subtracting the mean of the outcome variable at baseline and dividing by the standard deviation (SD) of the outcome variable at baseline (35). Thus, model coefficients will be interpreted in baseline SD units of the outcome both for direct and indirect/mediated effects. Single mediator models with contemporaneous mediation (*b*) paths – where the mediator and outcome are both measured at six months after randomisation – will be fitted (36). No parallel mediator models will be specified. Separate, single mediator models will be fitted for each mediator of interest, using full information maximum likelihood to account for missing data under the MAR assumption (37). Consequently, each fitted model will include all trial participants. Mediators of interest are reported in tables 5-10. Single mediator models will also adjust for the following confounding variables in equations for both the mediator and the primary outcome: age, gender, duration of IBD, and type of IBD (i.e., Crohn's Disease or Ulcerative Colitis), the baseline measure of the mediator, and the baseline measure of the outcome (38).

The "product of coefficients" approach will be applied to calculate the indirect (mediated) effect by multiplying the intervention regression coefficient (*a* path) by the mediation regression coefficient (*b* path) (39, 40). Percentile bootstrap 95% confidence intervals (CI) will be calculated for these effects, using 1000 repetitions. Full mediation

and partial mediation will be considered based on a change in direct/intervention and indirect/mediated effect estimates from unadjusted to adjusted analysis (41). The *gsem* command and associated options in Stata 17.0 will be used to perform mediation analyses.

Table 1. Mediators of interest for inclusion in single mediator models

Domain	#	Mediator of interest
Cognitive response	1	Illness-specific perceptions (BIPQ) (minus the open-ended causal items)
	2	Self-efficacy (SEMCD6)
Behavioural responses	3	All-or-nothing thinking (CRBQ)
	4	Avoidance/resting (CRBQ)
Emotional responses	5	Patient Health Questionnaire (PHQ-9) for depression
	6	Visceral Sensitivity Index (VSI) for anxiety

6.10. Sensitivity analyses

6.10.1 Multiple Imputation of missing data

A sensitivity analysis to assess the robustness of the analysis of primary outcomes under the assumption that missing data are Missing Not At Random (MNAR) will be performed using controlled multiple imputation (42, 43). A δ-based imputation approach will be used where an offset term, δ, is added to the expected value of the missing data to assess the impact of unobserved participants having a worse or better response than those observed. In a first step the expected values of missing data for the outcome variable of interest only (i.e., mean imputation values for baseline variables included in the model are kept the same in all imputed datasets) at all time points are generated using multiple imputation under MAR.

First, the proportion of missing values for each variable will be assessed using numerical and graphical summaries. Univariable associations between missing values of each variable and observed values of other variables will be examined to understand how reliably a missing value might be imputed. These associations will be examined by constructing separate logistic regression models after creating a binary indicator variable for each variable with missing values coded as "1" and non-missing values coded as "0".

Using clinical knowledge of inter-relationships between independent and dependent variables, and plausible reasons for missingness in the outcome data, we will assume that missing data are MAR. Multivariate Imputation using Chained Equations (MICE) will be used to impute missing data under the expectation that dependent variables will have missing values and the data will not be monotone missing (44). MICE uses fully conditional specification to impute multivariate missing data on a variable-by-variable basis after an imputation model for each incomplete variable has been specified (45), replacing missing values with a random sample of plausible, imputed values drawn from their predictive distribution (46).

Multiple imputation will be performed across three steps (47). First, an 'imputation' step will construct an imputation model to replace missing data with one set of plausible values. Assuming that missing data are MAR, the imputation

model will specify a conditional distribution for missing values of each variable given the observed values of other variables. This imputation model will repeatedly replace missing values with a random sample of plausible values, creating a completed dataset with each imputation. The number of imputations (and thus completed datasets generated) will mirror the proportion of participants with at least one missing value. For example, 25 complete datasets will be generated if 25% of study participants have at least one missing value. When $m = 100\lambda$, where m is the number of imputations produced and λ is the proportion of incomplete cases, the following properties hold for model parameter β:

1. The Monte Carlo error of β is approximately 10% of its standard error (SE)
2. The Monte Carlo error of test statistic, $\frac{\beta}{se(\beta)}$, is approximately 0.1.
3. The Monte Carlo error of the p-value is approximately 0.01 when the true p-value is 0.05.

If there is a very high percentage of incomplete cases, the average percentage of incomplete cases will be used.

If the number of imputations indicated by the average percentage of incomplete cases yields an impractically long run-time or convergence problems, the following formula will be used to estimate the total imputation variance for a given number of imputations:

$$T_m = 1 + \frac{\gamma 0}{m}$$

, where T_m is the total (i.e., between- and within-) imputation variance, $\gamma 0$ is the proportion of missingness in a single variable, and m is the number of imputed datasets produced. For example, if m = 20 and $\gamma 0$ = 0.1, T_m is only 0.5% greater than if an infinite number of imputations were produced. Perhaps more importantly, T_m is only 0.3% greater if m=20 than if m=50 and would still produce accurate point estimates while reducing computational burden.

A logistic regression model will be used for missing values of binary variables and a multinomial logistic regression model will be selected for missing values of categorical variables with three or more unordered categories. Missing values of categorical variables with three or more ordered categories will be modelled using ordinal logistic regression and a linear regression model will be specified for continuous variables with missing data. Auxiliary variables – that is, baseline characteristics variables that are not included in the primary outcome analysis but that are moderately correlated (i.e., $r > 0.4$) with (i) the observed values of the missing variable or (ii) its missingness – will be included in the imputation model. Additionally, covariates included in the primary outcome analysis model specified above will be included in the imputation model.

Next, an 'estimation' step will be undertaken, whereby the primary outcome analysis model – as described in section 6.5 – will be performed separately for each completed dataset that is generated during the imputation step. Finally, a 'pooling' step will aggregate the point estimates (e.g., sample means) and measures of precision (e.g.,

standard deviations) estimated in each dataset using Rubin's Rules to create a final estimate that accounts for between- and within-imputation uncertainty (48).

6.10.1.1 Missing outcomes worse than observed

We define $\delta_{1,2}$ as the fixed difference in the primary outcomes between observed and unobserved cases at months 6 and 12, respectively. For each participant with missing data, we then modify the MAR imputed observations at month 6 and 12 by subtracting δ_1 and δ_2. This will be done for δ values corresponding to 25%, 50%, 75%, and 100% of the estimated unadjusted, overall mean change of the outcome from baseline to 6 months (δ_1) and baseline to 12 months follow-up (δ_2), respectively. Four scenarios with the same percentage difference for δ_1 & δ_2 will be presented for each primary outcome.

6.10.1.2 Missing outcomes worse than observed in intervention and better in CAU group

It is conceivable that missing outcome values in the intervention group are worse those that are observed because participants for whom the intervention didn't work are more likely to drop out. The CAU group is a waitlist control obtaining access to the intervention after 12 months follow-up. CAU participants whose symptoms do not improve over the follow-up period or deteriorate have a higher incentive to receive the intervention and might therefore be less likely to drop-out. A second MNAR sensitivity analysis will therefore be conducted where for unobserved intervention group outcomes the same parameter as in 6.10.1.1 will be used whilst for unobserved CAU outcomes δ values corresponding to 25%, 50%, 75%, and 100% of the mean change of the outcome from baseline to 6 months (δ_1) and baseline to 12 months follow-up (δ_2) will be added to the MAR imputed observations.

6.10.2 IBD-BOOST OPTIMISE participants

This sensitivity analysis will investigate the effect of intervention in trial participants who received medical optimisation for IBD-related symptoms. Only a small number of participants received optimisation. Consequently, the analysis model will be simplified compared to the primary analysis model by removing covariates PROMIS pain, fatigue and incontinence as well as accounting for within facilitator correlation.

7 Other analyses, data summaries, and graphs

7.1 Other data summaries

7.1.1. Complier-averaged causal effect (CACE) analysis

We will perform a CACE analysis to estimate the effect of the intervention on both primary outcomes at 6 months after randomisation with a latent variable approach using structural equation modelling. The CACE treatment effect will be defined as the difference, on average, between compliant participants who were randomly assigned to the intervention arm and participants in the control arm who would have complied with treatment had they been randomised to the intervention arm (49). Using the gsem command in Stata, we will specify two regression paths within the structural equation model – a regression path for compliers and a regression path for non-compliers.

'Compliers' will only be observed in the intervention arm, where an indicator variable will identify whether the participant complied. Participants in the intervention arm will be classified as "compliant" if they were randomised to the intervention, clicked the link to the online intervention, commenced the intervention, AND completed a minimum of four online sessions. Latent mixture modelling will be used to identify participants in the control group who would have complied with treatment if they had been randomly allocated to the intervention arm (50). The latent class variable, 'compliance', will be determine using relevant predictors. Specifically, compliance among control group participants will be estimated using (i) the observed compliance data available for the participants randomized to treatment, (ii) the missing compliance data for the participants randomized to the control arm, and (iii) the distribution of the outcome variable in the sample (49). We will adjust for primary outcome scores at baseline in the analysis model and this model will also include a random intercept in the intervention arm only for clustering by intervention facilitator.

We will assume (i) monotonicity (i.e., there will be no "defiers" or "always-takers" in the study sample), (ii) stable unit treatment value (i.e., a participant's outcome depends only on their own group assignment and not on the group assignment of other participants), (iii) random assignment (i.e., exchangeability between intervention and control arms with respect to the trial outcome), and (iv) exclusion restriction (i.e., the treatment effect estimate will be fixed at zero for 'non-compliers' but freely estimated for 'compliers').

7.2. Exploratory analyses

7.2.1. Numerical pain rating scale

The numerical pain rating scale (NRS) comprises four questions assessing (i) current pain intensity, (ii) lowest pain intensity, (iii) worst pain intensity, and (iv) average pain intensity. Each item is evaluated using a simple, eleven-point (0-10) Likert scale. There are four scores – one for each question. The number that the respondent selects is the respondent's NRS score for that question item. Current, lowest, and worst pain intensity measures will be analysed as exploratory outcomes using the primary outcome analysis model specified in 6.5.

7.3 Safety analyses

The total number of events (n) and percentage (%) of adverse events (AE) and serious adverse events (SAE) that are deemed possibly related and unrelated to the trial intervention. will be reported. Details on what constitutes a (serious) adverse event can be found in the study protocol.

Categories of SAEs are shown in Tables 15 in **Appendix 6.**

7.4 Graphs

Line graphs for overall treatment effect

We will use box-and-whisker plots within violin plots to visualise between-arm differences in point estimates (i.e., group means), inter-quartile range limits, outliers, and continuous data distributions for both primary outcome measures at each post-randomisation assessment time-point.

Path diagrams for mediation analysis

A causal path diagram will be constructed that presents the natural direct, natural indirect/mediated, total effect estimates (and 95%CIs) of the intervention on each primary outcome.

Forest plot for subgroup analysis

We will construct a forest plot that presents the treatment effect estimate (and 95%CI) of both primary outcomes for pre-specified subgroups in section 6.8.

7.5 SWAT analysis

Research question:

What is the effectiveness of a brief participant information leaflet (PIL) versus standard length PIL on participant recruitment and retention rates into the IBD BOOST RCT?

Methods:

A randomised study within a trial (SWAT) embedded in IBD BOOST. Patients identified as potentially eligible and invited by the central research team to participate in IBD BOOST were randomised in a 1:1 ratio to be sent the standard length PIL or a brief PIL.

The primary outcome (i) is the proportion of invited participants that are randomised into the IBD-BOOST RCT. Secondary outcome will be retention rate at 6 months (ii) and 12 months (iii). All outcomes are compared between the brief and standard PIL groups.

Analysis:

Analyses are using observed data only. Statistical hypothesis tests are two-sided using a 5% significance level. Estimates of proportions by group, a between group difference estimate in the form of an odds ratio and its associated 95% confidence intervals and p-values will be presented (Table 18) for the following 5 analyses:

1. Logistic regression with (i) as the dependent variable and PIL group as the independent variable
2. Logistic regression with (ii) as the dependent variable and PIL group as the independent variable
3. Logistic regression with (iii) as the dependent variable and PIL group as the independent variable
4. Logistic regression with (ii) as the dependent variable, PIL group and randomised intervention group as the independent variables
5. Logistic regression with (iii) as the dependent variable, PIL group and randomised intervention group as the independent variables

8. References

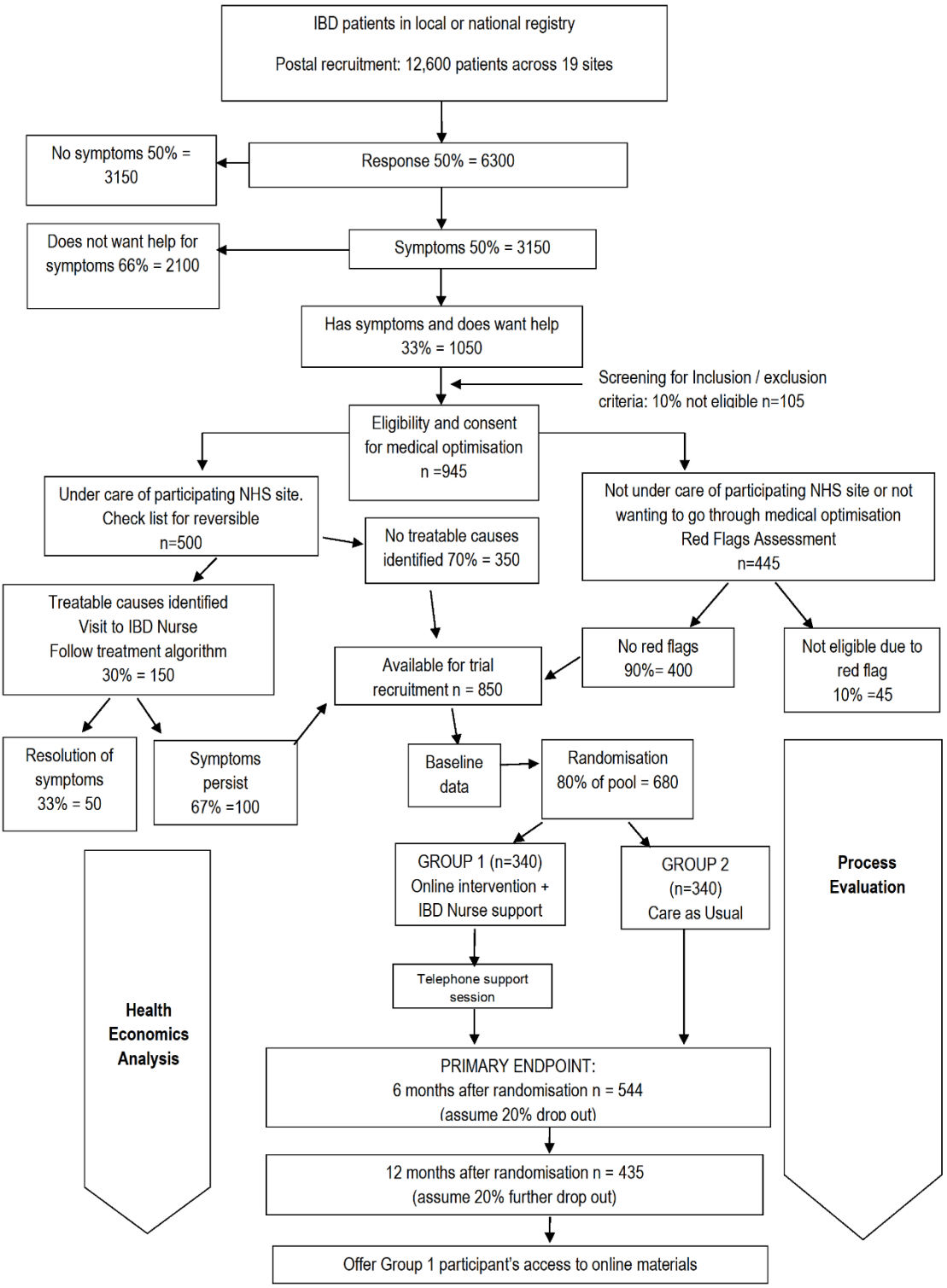
1. Lönnfors S, Vermeire S, Greco M, Hommes D, Bell C, Avedano L. IBD and health-related quality of life -- discovering the true impact. J Crohns Colitis. 2014 Oct;8(10):1281–6.
2. Czuber-Dochan W, Dibley LB, Terry H, Ream E, Norton C. The experience of fatigue in people with inflammatory bowel disease: an exploratory study. J Adv Nurs. 2013 Sep;69(9):1987–99.
3. Norton C, Dibley LB, Bassett P. Faecal incontinence in inflammatory bowel disease: Associations and effect on quality of life. Journal of Crohn's and Colitis [Internet]. 2013 [cited 2023 Aug 11];7(8):e302–11. Available from: https://academic.oup.com/ecco-jcc/article-lookup/doi/10.1016/j.crohns.2012.11.004
4. Cheung W yee, Garratt AM, Russell IT, Williams JG. The UK IBDQ—A British version of the inflammatory bowel disease questionnaire. Journal of Clinical Epidemiology [Internet]. 2000 [cited 2023 Aug 11];53(3):297–306. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0895435699001523
5. Rubin GP, Hungin APS, Chinn DJ, Dwarakanath D. Quality of life in patients with established inflammatory bowel disease: a UK general practice survey. Aliment Pharmacol Ther. 2004 Mar 1;19(5):529–35.
6. Müller-Lissner S. Subject's Global Assessment of Relief: An appropriate method to assess the impact of treatment on irritable bowel syndrome-related symptoms in clinical trials. Journal of Clinical Epidemiology [Internet]. 2003 [cited 2023 Aug 11];56(4):310–6. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0895435603000271
7. Jensen MP. Hypnosis for chronic pain management: therapist guide. New York: Oxford University Press; 2011. 283 p. (Treatments that work).
8. Vaizey CJ, Carapeti E, Cahill JA, Kamm MA. Prospective comparison of faecal incontinence grading systems. Gut [Internet]. 1999 Jan 1 [cited 2023 Aug 11];44(1):77–80. Available from: https://gut.bmj.com/lookup/doi/10.1136/gut.44.1.77
9. Trieschmann K, Chang L, Park S, Naliboff B, Joshi S, Labus JS, et al. The visceral sensitivity index: A novel tool for measuring GI-symptom-specific anxiety in inflammatory bowel disease. Neurogastroenterology Motil [Internet]. 2022 [cited 2023 Aug 11];34(9). Available from: https://onlinelibrary.wiley.com/doi/10.1111/nmo.14384
10. Czuber-Dochan W, Norton C, Bassett P, Berliner S, Bredin F, Darvell M, et al. Development and psychometric testing of inflammatory bowel disease fatigue (IBD-F) patient self-assessment scale. Journal of Crohn's and Colitis [Internet]. 2014 [cited 2023 Aug 11];8(11):1398–406. Available from: https://academic.oup.com/ecco-jcc/article-lookup/doi/10.1016/j.crohns.2014.04.013
11. Norton C, Czuber-Dochan W, Bassett P, Berliner S, Bredin F, Darvell M, et al. Assessing fatigue in inflammatory bowel disease: comparison of three fatigue scales. Aliment Pharmacol Ther [Internet]. 2015 [cited 2023 Aug 11];42(2):203–11. Available from: https://onlinelibrary.wiley.com/doi/10.1111/apt.13255
12. Bodger K, Ormerod C, Shackcloth D, Harrison M, on behalf of the IBD Control Collaborative. Development and validation of a rapid, generic measure of disease control from the patient's perspective: the IBD-Control questionnaire. Gut [Internet]. 2014 [cited 2023 Aug 11];63(7):1092–102. Available from: https://gut.bmj.com/lookup/doi/10.1136/gutjnl-2013-305600
13. Herdman M, Gudex C, Lloyd A, Janssen Mf, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res [Internet]. 2011 [cited 2023 Aug 11];20(10):1727–36. Available from: http://link.springer.com/10.1007/s11136-011-9903-x
14. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990 Dec;16(3):199–208.
15. Hernández Alava M, Pudney S, Wailoo A. Estimating the Relationship Between EQ-5D-5L and EQ-5D-3L: Results from a UK Population Study. PharmacoEconomics [Internet]. 2023 [cited 2023 Aug 11];41(2):199–207. Available from: https://link.springer.com/10.1007/s40273-022-01218-7
16. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. 2006 Jun;60(6):631–7.
17. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract. 2001;4(6):256–62.

18. Picariello F, Chilcot J, Chalder T, Herdman D, Moss-Morris R. The Cognitive and Behavioural Responses to Symptoms Questionnaire (CBRQ): Development, reliability and validity across several long-term conditions. Br J Health Psychol. 2023 May;28(2):619–38.
19. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med [Internet]. 2001 [cited 2023 Aug 11];16(9):606–13. Available from: http://link.springer.com/10.1046/j.1525-1497.2001.016009606.x
20. Arrieta J, Aguerrebere M, Raviola G, Flores H, Elliott P, Espinosa A, et al. Validity and Utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for Screening and Diagnosis of Depression in Rural Chiapas, Mexico: A Cross-Sectional Study: PHQ-9 Validity for Depression Diagnosis. J Clin Psychol [Internet]. 2017 [cited 2023 Aug 11];73(9):1076–90. Available from: https://onlinelibrary.wiley.com/doi/10.1002/jclp.22390
21. Labus JS, Bolus R, Chang L, Wiklund I, Naesdal J, Mayer EA, et al. The Visceral Sensitivity Index: development and validation of a gastrointestinal symptom-specific anxiety scale: THE VISCERAL SENSITIVITY INDEX. Alimentary Pharmacology & Therapeutics [Internet]. 2004 [cited 2023 Aug 11];20(1):89–97. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2036.2004.02007.x
22. Kyaw MH, Moshkovska T, Mayberry J. A prospective, randomized, controlled, exploratory study of comprehensive dietary advice in ulcerative colitis: impact on disease activity and quality of life. Eur J Gastroenterol Hepatol. 2014 Aug;26(8):910–7.
23. Williams JG, Alam MF, Alrubaiy L, Clement C, Cohen D, Grey M, et al. Comparison Of iNfliximab and ciclosporin in STeroid Resistant Ulcerative Colitis: pragmatic randomised Trial and economic evaluation (CONSTRUCT). Health Technol Assess. 2016 Jun;20(44):1–320.
24. Kennedy AP, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, et al. A randomised controlled trial to assess the effectiveness and cost of a patient orientated self management approach to chronic inflammatory bowel disease. Gut. 2004 Nov;53(11):1639–45.
25. Elkjaer M, Shuhaibar M, Burisch J, Bailey Y, Scherfig H, Laugesen B, et al. E-health empowers patients with ulcerative colitis: a randomised controlled trial of the web-guided 'Constant-care' approach. Gut. 2010 Dec;59(12):1652–61.
26. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care. 1999 Jan;37(1):5–14.
27. Walters SJ, Jacques RM, Dos Anjos Henriques-Cadby IB, Candlish J, Totton N, Xian MTS. Sample size estimation for randomised controlled trials with repeated assessment of patient-reported outcomes: what correlation between baseline and follow-up outcomes should we assume? Trials [Internet]. 2019 [cited 2023 Aug 11];20(1):566. Available from: https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3671-2
28. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, et al. Cognitive behavioural therapy for irritable bowel syndrome: 24-month follow-up of participants in the ACTIB randomised trial. The Lancet Gastroenterology & Hepatology [Internet]. 2019 [cited 2023 Aug 11];4(11):863–72. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2468125319302432
29. Candlish J, Teare MD, Dimairo M, Flight L, Mandefield L, Walters SJ. Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. BMC Med Res Methodol. 2018 Oct 11;18(1):105.
30. Giesbrecht FG, Burns JC. Two-Stage Analysis Based on a Mixed Model: Large-Sample Asymptotic Theory and Small-Sample Simulation Results. Biometrics [Internet]. 1985 [cited 2023 Aug 11];41(2):477. Available from: https://www.jstor.org/stable/2530872?origin=crossref
31. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med. 2005 Apr 15;24(7):993–1007.
32. Drossman DA, Hasler WL. Rome IV—Functional GI Disorders: Disorders of Gut-Brain Interaction. Gastroenterology [Internet]. 2016 [cited 2023 Aug 11];150(6):1257–61. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0016508516300488
33. Krull JL, MacKinnon DP. Multilevel Modeling of Individual and Group Level Mediated Effects. Multivariate Behav Res. 2001 Apr 1;36(2):249–77.
34. MacKinnon DP. Introduction to Statistical Mediation Analysis [Internet]. 1st ed. Routledge; 2012 [cited 2023 Aug 11]. Available from: https://www.taylorfrancis.com/books/9781136676147

35. Goldsmith K, Hudson JL, Chalder T, Dennison L, Moss-Morris R. How and for whom does supportive adjustment to multiple sclerosis cognitive-behavioural therapy work? A mediated moderation analysis. Behav Res Ther. 2020 May;128:103594.
36. Goldsmith KA, MacKinnon DP, Chalder T, White PD, Sharpe M, Pickles A. Tutorial: The practical application of longitudinal structural equation mediation models in clinical trials. Psychol Methods. 2018 Jun;23(2):191–207.
37. Enders C, Bandalos D. The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. Structural Equation Modeling: A Multidisciplinary J [Internet]. 2001 Jul 1 [cited 2023 Aug 11];8(3):430–57. Available from: http://www.informaworld.com/openurl?genre=article&doi=10.1207/S15328007SEM0803_5&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3
38. Landau S, Emsley R, Dunn G. Beyond total treatment effects in randomised controlled trials: Baseline measurement of intermediate outcomes needed to reduce confounding in mediation investigations. Clinical Trials [Internet]. 2018 [cited 2023 Aug 11];15(3):247–56. Available from: http://journals.sagepub.com/doi/10.1177/1740774518760300
39. MacKinnon DP. Mediating Variable. In: International Encyclopedia of the Social & Behavioral Sciences [Internet]. Elsevier; 2001 [cited 2023 Aug 11]. p. 9503–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/B0080430767007324
40. Mackinnon DP, Dwyer JH. Estimating Mediated Effects in Prevention Studies. Eval Rev [Internet]. 1993 [cited 2023 Aug 11];17(2):144–58. Available from: http://journals.sagepub.com/doi/10.1177/0193841X9301700202
41. Windgassen S, Goldsmith K, Moss-Morris R, Chalder T. Establishing how psychological therapies work: the importance of mediation analysis. J Ment Health. 2016;25(2):93–9.
42. Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol. 2013 Jul 16;13:92.
43. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Statistics in Medicine [Internet]. 2020 Sep 20 [cited 2023 Aug 11];39(21):2815–42. Available from: https://onlinelibrary.wiley.com/doi/10.1002/sim.8569
44. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017 Dec 6;17(1):162.
45. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007 Jun;16(3):219–42.
46. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011 Mar;20(1):40–9.
47. Austin PC, White IR, Lee DS, van Buuren S. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. Can J Cardiol. 2021 Sep;37(9):1322–31.
48. Rubin DB. Inference and missing data. Biometrika [Internet]. 1976 [cited 2023 Aug 11];63(3):581–92. Available from: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/63.3.581
49. Peugh JL, Strotman D, McGrady M, Rausch J, Kashikar-Zuck S. Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer. J Sch Psychol. 2017 Feb;60:7–24.
50. Troncoso P, Morales-Gómez A. Estimating the complier average causal effect via a latent class approach using gsem. The Stata Journal [Internet]. 2022 [cited 2023 Aug 11];22(2):404–15. Available from: http://journals.sagepub.com/doi/10.1177/1536867X221106416

11. Appendices

Appendix 1: IBD BOOST Programme Participant Flow Chart



Appendix 2. Outcome collection timeline.

Outcomes	Baseline	6 Months	12 Months
UK-IBDQ	X	X	X
Global rating symptom relief		X	X
Satisfaction of the IBD-BOOST program		X	X
Pain rating scale	X	X	X
Vaizey Incontinence score	X	X	X
IBD-Fatigue Score	X	X	X
IBD-Control Score	X	X	X
EQ-5D-5L	X	X	X
Process evaluation measures			
Patient Health Questionnaire (PHQ-9)	X	X	X
Self-efficacy	X	X	X
Brief Illness Perceptions Questionnaire (BIPQ)	X	X	X
Rome IV criteria for IBS	X		
VSI	X	X	X
CBRQ-behavioural subscales	X	X	X

Appendix 3. RSD items needed for outcome derivations.

1. UK-IBDQ

- The total UK-IBDQ score is calculated by adding 1 to all database values and summing items *ibdq1,….,ibdq31* from the RSD, omitting items *ibdq7* and *ibdq32* from the calculation.
- Items 6 (ibdq6) and 21 (ibdq21)) are reverse coded as for these items 1 is the worst response (value="No, not at all", "None") and 4 is the best response.
- Three items allow values "999" (i.e., "Not applicable to me"). Item values "999" will be re-coded to "Not at all" (score=1) to indicate that the activity in question does not impact the participant's quality of life.

Example code:

```
gen ibdq_total = ibdq1 + ibdq2 + ibdq3 + ibdq4 + ibdq5 + ibdq6_rev
+ + ibdq8+ ibdq9 + ibdq10 + ibdq11+ ibdq12 + ibdq13 + ibdq14 +
ibdq15 + ibdq16 + ibdq17 + ibdq18 + ibdq19 + ibdq20 + ibdq21_rev
+ ibdq22 + ibdq23 + ibdq24 + ibdq25 + ibdq26 + ibdq27 + ibdq28
+ ibdq29 + ibdq30 + ibdq31
```

2. Global Rating Symptom relief

- *global_relief*: range 0-10, continuous

3. Rating of satisfaction with results of IBD BOOST Program at 6 and 12 months

- *global_satif*: range 0-10, continuous

4. Numerical pain scale

- The four separate scores are obtained from *items nrs1, nrs2, nrs3* and *nrs4* in the RSD.

5. Vaizey Incontinence Score

- *vaizey_solid*: range 0-4
- *vaizey_liquid*: range 0-4
- *vaizey_gas*: range 0-4
- *vaizey_lifestyle*: range 0-4
- *vaizey_pad*, binary: 0-1 (no/yes)
- *vaizey_medicine*, binary: 0-1 (no/yes)
- *vaizey_hold,* binary: 0-1 (no/yes)

The total Vaizey incontinence score is calculated by adding each of the previous items. The score's range is 0 (perfect continence) – 24 (total incontinence) [continuous]. Before calculating the total score, the *vaizey_pad*, *vaizey_medicine* and *vaizey_hold* items need to be recoded, according to the table, in order to compute the total score. *vaizey_hold* will also be reverse-scored so that yes=0 and no=4 due to the phrasing of this item in the RSD as: "are you able to 'hold on' for 15 mins before going to the toilet?".

	Never	***Rarely***	***Sometimes***	***Weekly***	***Daily***
Incontinence for solid stool	0	1	2	3	4
Incontinence for liquid stool	0	1	2	3	4
Incontinence for gas	0	1	2	3	4
Alteration in lifestyle	0	1	2	3	4
				No	***Yes***

Need to wear a pad or plug	0	2
Taking constipating medicines	0	2
Lack of ability to 'hold on' defecation for 15 minutes	0	4

Example STATA code:

```
- recode vaizey_pad (0=0) (1=2)
-

- recode vaizey_med (0=0) (1=2)

      recode vaizey_hold (0=4) (1=0)
- gen vaizey_total = vaizey_solid + vaizey_liquid + vaizey_gas +
  vaizey_lifestyle + vaizey_pad + vaizey_med + vaizey_hold
```

6. IBD Fatigue Score

The IBD Fatigue score is calculated by summing the following RSD items:

- *ibdf_now:* range 0-4, continuous
- *ibdf_high:* range 0-4, continuous
- *ibdf_low*: range 0-4, continuous
- *ibdf_average*: range 0-4, continuous

Example code:

```
ibdf_fatigue_total = ibdf_now + ibdf_high + ibdf_low + ibdf_average
```

7. IBD Control Score

The IBD-Control score is the sum of items control1, control2, and control4 through control9 [continuous]. The range of scores is 0-16, with 0 indicating worse control and 16 indicating best control. Note that N/A, denoted 999, is counted as 1 for item control2. If control2 has the value 999, treat it as having the value 1 to sum the score.

The following level re-coding will take place for control1, control2, control4 through control9:

Item	Levels	Original coding	Recoding
control1	Yes | No | Not sure	1, 0, 2	2, 0, 1
control2	Yes | No | Not sure | N/A	1, 0, 2, 999	2, 0, 1, 1
control4	Yes | No | Not sure	1, 0, 2	0, 2, 1
control5	Yes | No | Not sure	1, 0, 2	0, 2, 1
control6	Yes | No | Not sure	1, 0, 2	0, 2, 1
control7	Yes | No | Not sure	1, 0, 2	0, 2, 1
control8	Yes | No | Not sure	1, 0, 2	0, 2, 1
control9	Yes | No | Not sure	1, 0, 2	0, 2, 1

Example code:

```
recode control1 (1=2) (0=0) (2=1)

recode control2 (1=2) (0=0) (2=1) (999=1)

foreach of varlist control4-control9 {
      recode `v' (1=0) (0=2) (2=1)
}

control_total = control1 + control2 + control4 + control5 + control6
+ control7 + control8 + control9
```

The items *ibdcontrol3* (categorical variable with 3 categories) and *ibdcontrol10* (continuous variable, range 0-10) are reported separately as they are scored. Item *ibdcontrol10* should be first converted from a 0-10 to a 0-100 scale and is usually reported as IBD control VAS score.

The IBD-Control-VAS score evaluates self-reported overall level of control and ranges from 0-10, with 0 indicating worst control. Participants with an IBD-Control-8 sub-score <13 points and an IBD-Control-VAS score ≥8 points indicate inactive IBD. IBD-Control-VAS is used to determine inactive IBD only. The secondary outcome measure is IBD-Control score [continuous].

Example code:

```
recode control10 (0=0) (1-10) (2=20) (3=30) (4=40) (5=50) (6=60) (7=70)
(8=80) (9=90) (10=100), gen control_vas
```

8. **EQ-5D-5L general health-related quality of life**

The EQ-5D-5L index questionnaire is a standardized tool that is used to measure health status and health-related quality of life. The EQ-5D-5L assesses five dimensions of health (i.e., individual health states): mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has five levels of perceived problems, ranging from no problems to extreme problems.

LEVEL 1: indicating no problem

LEVEL 2: indicating slight problems

LEVEL 3: indicating moderate problems

LEVEL 4: indicating severe problems

LEVEL 5: indicating unable to/extreme problems

9. Brief illness perceptions questionnaire (BIPQ)

To calculate a composite BIPQ score, the individual 8 domain scores are summed together (the personal control (3), treatment control (4), and coherence (7) items are reverse scored, as higher scores in these elements

represent positive illness perceptions). A higher BIPQ score indicates a greater perceived psychological burden of illness (range 0–80).

Example stata code:
```
foreach i of numlist 3 4 7 {
      recode bspq (10=0) (9=1) (8=2) (7=3) (6=4) (4=6) (3=7) (2=8) (1=9)
(0=10)
}
egen bspq_total = rowtotal(bspq1-bspq8)
```

10. Self-efficacy for managing chronic disease (SEMCD6)

Example stata code:
```
egen eff_total = rowmean(efficacy1-efficacy6)
```

11. All-or-nothing thinking of Cognitive & Behavioural Response Questionnaire (CBRQ)

CBRQ domain	Domain items	Outcome derivation (w Stata code)
All-or-nothing behaviour	AL1: I tend to overdo things when I feel energetic. AL2: I find myself rushing to get things done before I crash. AL3: I tend to overdo things and then rest up for a while. AL4: I tend to do a lot on a good day and rest on a bad day. AL5: I'm a bit all or nothing when it comes to doing things.	AL1 + AL2 + AL3 + AL4 + AL5 Stata code on IBD-BOOST dataset: `gen cbrq_total = cbrq5 + cbrq6 + cbrq7 + cbrq8 + brq12`

12. Avoidance/resting behaviour of Cognitive & Behavioural Response Questionnaire (CBRQ)

Domain	Items	Derivation
Avoidance/resting behaviour	L2: I stay in bed to control my symptoms. L3: When I experience symptoms, I rest. L4: I tend to avoid activities that make my symptoms worse. L7: I tend to nap during the day to control my symptoms.	L2 + L3 + L4 + L7 + L9 + L10 + L11 + L13 Stata code on IBD-BOOST dataset: `gen cbrq_total = crbq1 + crbq2 + crbq3 + crbq4 + crbq9 + crbq10 + crbq11 + crbq13`

	L9: I sleep when I'm tired in order to control my symptoms. L10: I avoid making social arrangements in case I'm not up to it. L11: I avoid exerting myself in order to control my symptoms. L13: I avoid stressful situations.	

13. Patient Health Questionnaire-9 (PHQ-9)

Example stata code:

```
egen phq_total = rowtotal(phq1-phq9)
```

14. Visceral Sensitivity Index (VSI)
    - Items are scored on a reversed 6-point Visceral Sensitivity Index (VSI) for anxiety scale ranging from 0 to 5, with sum scores between 0 and 75.

Example stata code:

```
egen vsi_total = rowtotal(vsi1-vsi15)
```

Appendix 4. Data completeness.

Table 1: Baseline demographic and clinical characteristics

	Intervention (n_1=_)	Control (n_2=_)
Gender, n(%)		
Male		
Female		
Prefer not to say		
Prefer to self-describe		
Age (years), mean (±sd)		
BMI (kg/m^2), n(%)		
Underweight (<18.5 kg/m^2)		
Normal weight (18.5 – 24.9 kg/m^2)		
Overweight/Obese (≥25.0 kg/m^2)		
Smoking habits, n(%)		
Never		
Ex		
Current		
Alcohol consumption (units/week), mean (±sd)		
Ethnicity, n(%)		
White		
Mixed		
Asian		
Black		
Other		
Employment, n(%)		
Full-time		
Part-time		
Student		
Retired		
Unemployed		
Self-employed		
Homemaker		
Unemployed due to illness/disability		
Education, n(%)		
No formal education		
Secondary school (GCSE)		
Sixth form (AS/A-levels)		
Further education (Vocational courses/apprenticeships/diplomas)		
Higher education – University Degrees		
Relationship status, n (%)		
Married/Civil partnership		
Living with partner		
Widowed		
Divorced/Separated		
Single		
With a partner (but not living together)		

IBD Classification, n (%)		
Crohn's disease or Crohn's colitis		
Intermediate colitis, IBD unclassified or other type of IBD		
Ulcerative colitis		
Not sure		
Rome IV IBS criteria met, n (%)		
Yes		
No		
Operation history, n (%)		
Yes		
No		
Stoma, n (%)		
Yes		
No		
Pouch, n (%)		
Yes		
No		
Fistula, n (%)		
Yes		
No		
Biologic medications, n (%)		
Yes		
No		
IBD medications, n (%)		
Yes		
No		
Mental health conditions, n (%)		
Yes		
No		
Physical health conditions, n (%)		
Yes		
No		
Pregnant, n (%)		
Yes		
No		

Table 2: Baseline outcome values

Outcomes	Intervention (n=_)	Control (n=_)
UK-IBDQ, mean (SD)		
Numeric Pain Rating Scale		
Current pain intensity, mean (SD)		

Least pain intensity, mean (SD)		
Worst pain intensity, mean (SD)		
Average pain intensity, mean (SD)		
Vaizey Incontinence Score, mean (SD)		
IBD-Fatigue Score, mean (SD)		
IBD-Control-score, mean (SD)		
IBD-Control VAS Score, mean (SD)		
IBD in remission, n (%)		
Yes		
No		
EQ-5D-5L index (utility score), mean (SD)		
EQ-5D-5L mobility, n (%)		
1 no problems		
2 slight problems		
3 moderate problems		
4 severe problems		
5 extreme problems		
EQ-5D-5L self-care, n (%)		
1 no problems		
2 slight problems		
3 moderate problems		
4 severe problems		
5 extreme problems		
EQ-5D-5L usual activities, n (%)		
1 no problems		
2 slight problems		
3 moderate problems		
4 severe problems		
5 extreme problems		
EQ-5D-5L pain/discomfort, n (%)		
1 no problems		
2 slight problems		
3 moderate problems		
4 severe problems		
5 extreme problems		
EQ-5D-5L anxiety/depression, n (%)		
1 no problems		
2 slight problems		
3 moderate problems		
4 severe problems		
5 extreme problems		

Table 3: Primary and secondary trial outcomes at 6 months post-randomisation assessment

Outcomes	Intervention		Control		Treatment effect (95%CI)	p-value
	Included in analysis, n (%)	Mean (SD)	Included in analysis, n (%)	Mean (SD)		
UK-IBDQ, mean (SD)						
Global Rating Symptom Relief, mean (SD)						
Numeric Pain Rating Scale						
Average pain intensity, mean (SD)						
Vaizey Incontinence Score, mean (sd)						
IBD-Fatigue Score, mean (SD)						
IBD-Control-score, mean (SD)						
IBD-Control VAS Score, mean (SD)						
EQ-5D-5L index (utility score), mean (SD)						

SD, standard deviation; *95%CI*, 95% confidence interval

Table 4: Primary and secondary trial outcomes at 12 months post-randomisation assessment

Outcomes	Intervention		Control		Treatment effect (95%CI)	p-value
	Included in analysis, n (%)	Mean (SD)	Included in analysis, n (%)	Mean (SD)		
UK-IBDQ, mean (sd)						
Global Rating Symptom Relief, mean (sd)						
Numeric Pain Rating Scale						
Average pain intensity, mean (SD)						
Vaizey Incontinence Score, mean (SD)						
IBD-Fatigue Score, mean (SD)						
IBD-Control-score, mean (SD)						
IBD-Control VAS Score, mean (SD)						
EQ-5D-5L index (utility score), mean (SD)						

Table 5. Exploratory outcomes at 6- and 12-months post-randomisation assessment, by treatment arm

	Six months post-randomisation				Twelve months post-randomisation			
	Intervention arm		Control arm		Intervention arm		Control arm	
	n (%)	Summary measure	n (%)	Summary measure	n (%)	Summary measure	n (%)	Summary measure
Numerical pain rating scale								

Current pain intensity, mean (SD)	
Least pain intensity, mean (SD)	
Worst pain intensity, mean (SD)	
EQ-5D-5L domains	
EQ-5D-5L mobility, n (%)	
1 no problems	
2 slight problems	
3 moderate problems	
4 severe problems	
5 extreme problems	
EQ-5D-5L self-care, n (%)	
1 no problems	
2 slight problems	
3 moderate problems	
4 severe problems	
5 extreme problems	
EQ-5D-5L usual activities, n (%)	
1 no problems	
2 slight problems	
3 moderate problems	
4 severe problems	
5 extreme problems	
EQ-5D-5L pain/discomfort	
1 no problems	
2 slight problems	
3 moderate problems	
4 severe problems	
5 extreme problems	
EQ-5D-5L anxiety/depression, n (%)	
1 no problems	
2 slight problems	
3 moderate problems	
4 severe problems	
5 extreme problems	

SD, standard deviation; *95%CI*, 95% confidence interval

Tables 6. Results of causal mediation analysis investigating direct, indirect, and total effects of the intervention on primary trial outcomes six months after randomisation (mediator: illness perceptions)

Effect type	IV → DV	Effect estimate	Std. error	95%CI	P-value
UK-IBDQ					
Direct effects	Intervention → Mediator				
	Intervention → UK-IBDQ				
	Mediator → UK-IBDQ				
Indirect effect	Intervention → UK-IBDQ				
Total effect	Intervention → UK-IBDQ				
Global Rating of Symptom Relief (GRSR)					

Direct effects	Intervention → Mediator				
	Intervention → GRSR				
	Mediator → GRSR				
Indirect effect	Intervention → GRSR				
Total effect	Intervention → GRSR				

Tables 7. Results of causal mediation analysis investigating direct, indirect, and total effects of the intervention on primary trial outcomes six months after randomisation (mediator: self-efficacy)

Effect type	IV → DV	Effect estimate	Std. error	95%CI	P-value
UK-IBDQ					
Direct effects	Intervention → Mediator				
	Intervention → UK-IBDQ				
	Mediator → UK-IBDQ				
Indirect effect	Intervention → UK-IBDQ				
Total effect	Intervention → UK-IBDQ				
Global Rating of Symptom Relief (GRSR)					
Direct effects	Intervention → Mediator				
	Intervention → GRSR				
	Mediator → GRSR				
Indirect effect	Intervention → GRSR				
Total effect	Intervention → GRSR				

Tables 8. Results of causal mediation analysis investigating direct, indirect, and total effects of the intervention on primary trial outcomes six months after randomisation (mediator: all-or-nothing thinking)

Effect type	IV → DV	Effect estimate	Std. error	95%CI	P-value
UK-IBDQ					
Direct effects	Intervention → Mediator				
	Intervention → UK-IBDQ				
	Mediator → UK-IBDQ				
Indirect effect	Intervention → UK-IBDQ				
Total effect	Intervention → UK-IBDQ				
Global Rating of Symptom Relief (GRSR)					
Direct effects	Intervention → Mediator				
	Intervention → GRSR				
	Mediator → GRSR				
Indirect effect	Intervention → GRSR				
Total effect	Intervention → GRSR				

Tables 9. Results of causal mediation analysis investigating direct, indirect, and total effects of the intervention on primary trial outcomes six months after randomisation (mediator: avoidance/resting behaviour)

Effect type	IV → DV	Effect estimate	Std. error	95%CI	P-value
UK-IBDQ					

Direct effects	Intervention → Mediator
	Intervention → UK-IBDQ
	Mediator → UK-IBDQ
Indirect effect	Intervention → UK-IBDQ
Total effect	Intervention → UK-IBDQ
Global Rating of Symptom Relief (GRSR)	
Direct effects	Intervention → Mediator
	Intervention → GRSR
	Mediator → GRSR
Indirect effect	Intervention → GRSR
Total effect	Intervention → GRSR

Tables 10. Results of causal mediation analysis investigating direct, indirect, and total effects of the intervention on primary trial outcomes six months after randomisation (mediator: depression symptom severity)

Effect type	IV → DV	Effect estimate	Std. error	95%CI	P-value
UK-IBDQ					
Direct effects	Intervention → Mediator				
	Intervention → UK-IBDQ				
	Mediator → UK-IBDQ				
Indirect effect	Intervention → UK-IBDQ				
Total effect	Intervention → UK-IBDQ				
Global Rating of Symptom Relief (GRSR)					
Direct effects	Intervention → Mediator				
	Intervention → GRSR				
	Mediator → GRSR				
Indirect effect	Intervention → GRSR				
Total effect	Intervention → GRSR				

Tables 11. Results of causal mediation analysis investigating direct, indirect, and total effects of the intervention on primary trial outcomes six months after randomisation (mediator: illness-specific anxiety)

Effect type	IV → DV	Effect estimate	Std. error	95%CI	P-value
UK-IBDQ					
Direct effects	Intervention → Mediator				
	Intervention → UK-IBDQ				
	Mediator → UK-IBDQ				
Indirect effect	Intervention → UK-IBDQ				
Total effect	Intervention → UK-IBDQ				
Global Rating of Symptom Relief (GRSR)					
Direct effects	Intervention → Mediator				
	Intervention → GRSR				
	Mediator → GRSR				
Indirect effect	Intervention → GRSR				

Total effect	Intervention → GRSR

Table 12. Subgroup analysis investigating the effect of characteristics at baseline on the relationship between intervention and each primary trial outcome

Subgroup	Numbers included in analysis		Group-level summary data		Treatment effect (95%CI)	p-value
	Intervention, n (%)	Usual care, n (%)	Intervention, mean (SD)	Usual care, mean (SD)		
UK-IBDQ						
IBD in remission						
In remission at baseline						
Not in remission at baseline						
PHQ-9						
Depressed at baseline						
Not depressed at baseline						
VSI						
IBS – ROME IV Criteria						
Yes						
No						
Global rating of symptom relief						
IBD in remission						
In remission at baseline						
Not in remission at baseline						
PHQ-9						
Depressed at baseline						
Not depressed at baseline						
VSI						
IBS – ROME IV Criteria						
Yes						
No						

IBD remission at baseline is defined as faecal calprotectin <200µg/g AND an IBD control score ≥13. IBS at baseline diagnosed when the following criteria of Rome IV classification are fulfilled: *"Do you get abdominal (tummy) pain on a weekly basis?" (ibs1)* **AND** "*Have you had these symptoms for at least 6 months?" (ibs5)* **AND 2 or more of** *"Does this pain relate in some way to opening your bowels?" (ibs2),* "*Do your bowels change in frequency when you get this pain?" (ibs3)*, **OR** *"Do your stools change in appearance (softer, harder) when you get this pain?" (ibs4)*

Table 13. Sensitivity analyses UK-IBDQ

	δ	Treatment effect est.	95% CI	p-value
Primary analysis (MAR)				
MNAR worse than observed				
MI δ_1 =25%				
MI δ_1 =50%				
MI δ_1 =75%				
MI δ_1 =100%				
MI δ_2 =25%				
MI δ_2 =50%				
MI δ_2 =75%				
MI δ_2 =100%				

MNAR worse than observed, CAU better than observed	
MI δ_1 =25%	
MI δ_1 =50%	
MI δ_1 =75%	
MI δ_1 =100%	
MI δ_2 =25%	
MI δ_2 =50%	
MI δ_2 =75%	
MI δ_2 =100%	
Sample including patients with medical optimisation	*NA*

mean imputation values for baseline variables included in the model are kept the same in all imputed datasets) at all time points are generated using multiple imputation under MAR.

Table 14. Sensitivity analyses GRSR

	δ	Treatment effect est.	95% CI	p-value
Primary analysis (MAR)				
MNAR worse than observed				
MI δ_1 =25%				
MI δ_1 =50%				
MI δ_1 =75%				
MI δ_1 =100%				
MI δ_2 =25%				
MI δ_2 =50%				
MI δ_2 =75%				
MI δ_2 =100%				
MNAR E worse than observed, CAU better than observed				
MI δ_1 =25%				
MI δ_1 =50%				
MI δ_1 =75%				
MI δ_1 =100%				
MI δ_2 =25%				
MI δ_2 =50%				
MI δ_2 =75%				
MI δ_2 =100%				
Sample including patients with medical optimisation	*NA*			

mean imputation values for baseline variables included in the model are kept the same in all imputed datasets) at all time points are generated using multiple imputation under MAR.

Table 15. Complier-averaged causal effect (CACE) analysis estimating the difference, on average, between compliant participants who were randomly assigned to the intervention arm and participants in the control arm who would have complied with treatment had they been randomised to the intervention arm

Compliance status	Included in analysis, n (%)	Effect estimate (95%CI)	p-value
UK-IBDQ			
Compliers in intervention arm			
'Would-be' compliers in control arm			
Global rating of symptom relief			
Compliers in intervention arm			
'Would-be' compliers in control arm			

mean diff, mean difference; *95%CI*, 95% confidence interval

Table 16. Frequency and percentage of adverse effects (AEs) and serious adverse effects (SAEs) considered related and unrelated to the trial intervention, stratified by trial arm

AE outcome	Intervention arm n (%)	Control arm n (%)
AEs		
X		
Y		
Z		
…		
SAEs		
A life-threatening AE		
In-patient hospitalisation or prolonged hospitalisation not related to IBD flare, which are expected events		
Persistent or significant disability/incapacity		
A congenital anomaly/birth defect in the offspring of a subject		
Is otherwise considered medically significant by the investigator		
Other medical events requiring intervention to prevent one of the above outcomes		
Total number of serious adverse events		
Number of participants experiencing one or more SAE possibly related to the trial		
Number of participants experiencing one or more SAE not related to the trial		

AE, adverse event; SAE, serious adverse event

AEs are defined as any clinical change, disease, or disorder experienced by the participant during their participation in the trial, whether or not considered related to the use of treatments being studied.

SAEs are defined as serious if it results in one of the outcomes listed in the above table.

Table 17. Number and percent of participants who gave data for each measurement, by treatment arm.

	Total (n=)	Intervention (n_1=_)	Control (n_2=_)
Baseline Characteristics			
Gender	_ (_%)	_ (_%)	_ (_%)
Age (years)	_ (_%)	_ (_%)	_ (_%)
BMI (kg/m^2)	_ (_%)	_ (_%)	_ (_%)
Smoking habits		_ (_%)	_ (_%)
Alcohol consumption (units/week)	_ (_%)	_ (_%)	_ (_%)
Ethnicity	_ (_%)	_ (_%)	_ (_%)
Employment	_ (_%)	_ (_%)	_ (_%)
Education	_ (_%)	_ (_%)	_ (_%)
Relationship status	_ (_%)	_ (_%)	_ (_%)
IBD Classification	_ (_%)	_ (_%)	_ (_%)
Rome IV IBS criteria met	_ (_%)	_ (_%)	_ (_%)
Outcome measurements			
UK IBDQ			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
Global rating symptom relief			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
Satisfaction of the IBD-BOOST program			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)

Numerical pain rating scale (NRS)			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
Vaizey Incontinence Score			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
IBD-Fatigue Score			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
IBD-Control Score			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
IBD-Control Score VAS			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
EQ-5D-5L mobility			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
EQ-5D-5L self-care			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
EQ-5D-5L usual activities			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
EQ-5D-5L pain/discomfort			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
EQ-5D-5L anxiety/depression			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
PHQ-9			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
Self-efficacy			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
BIPQ			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
VSI			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)
CBRQ			
6 months	_ (_%)	_ (_%)	_ (_%)
12 months	_ (_%)	_ (_%)	_ (_%)

Appendix Table 18. SWAT analysis

	Standard PIL		Brief PIL				
	Included in analysis, n (%)	'yes', n (%)	Included in analysis, n (%)	'yes', n (%)	Odds ratio	95%CI	p-value
Randomised (yes/no)							
Retained 6m (yes/no)							
Retained 12m (yes/no)							
Retained 6m (yes/no) adjusted							
Retained 12m (yes/no) adjusted							

PIL, patient information leaflet; *95%CI,* 95%C confidence interval

Barts and The London School of Medicine and Dentistry

PCTU

Appendix 5. Dummy CONSORT Flow Diagram