ASPECT Statistical Analysis Plan Version 2

Study Title	**A non-inferiority randomised controlled trial comparing the clinical and cost-effectiveness of one session treatment (OST) with multi-session cognitive behavioural therapy (CBT) in children with specific phobias**
Short title	**Alleviating Specific Phobias Experienced by Children Trial (ASPECT)**
Funding body:	**NIHR HTA**
Funder's reference	HTA15/38/04
Trial Registration	ISRCTN19883421

Authored by

_____ ___/___/_____
Ellen Lee Date
Statistician, Sheffield CTRU

Approved by

_____ ___/___/_____
Prof. Dawn Teare Date
Senior Statistician, Newcastle University

_____ ___/___/_____
Prof. Barry Wright Date
ASPECT Chief investigator, University of York

_____ ___/___/_____
Prof. Cathy Creswell Date
ASPECT TSC Chair, University of Reading

_____ ___/___/_____
Prof. Dankmar Böhning Date
ASPECT TSC Statistician, University of Southampton

Amendments to the SAP since version 1

Version	Date approved	Modifications	Prior to/after blind review	Prior to/aft unblind review
2.0		- Change in sample size to reflect recalculated sample size for HTA extention request (Section 3) - Refined details of the ITT analysis population (Section 6.3) to include participants found to be ineligible post randomisation - Redefined per-protocol analysis population (section 6.2and Section 11) - More details on treatment summaries and compliance (section 7.6) including investigating compliance by site - Included a section (7.7) detailing an investigation of the characteristics of the per-protocol population - A sensitivity analysis on the primary outcome has been added (section7.8.2) excluding participants who received simulated BAT stimuli - Added simple regression to assess stability of mixed effects model in light of smaller therapist clusters (Section 8.3) - Added details about scoring the ADIS (Section 9.3 - Where BAT was attended but not completed due to fear, BAT steps will be taken as 0 (section9.2) - Other small changes have been made to clarify points and can be seen in a changes tracked version	Prior to blind review, however blinded summaries (presented to trial groups) have been seen by the trial team, Data from the session recording form has been seen by the trial statistician (E Lee) but not linked to other data.	Prior to unblind review

Table of contents

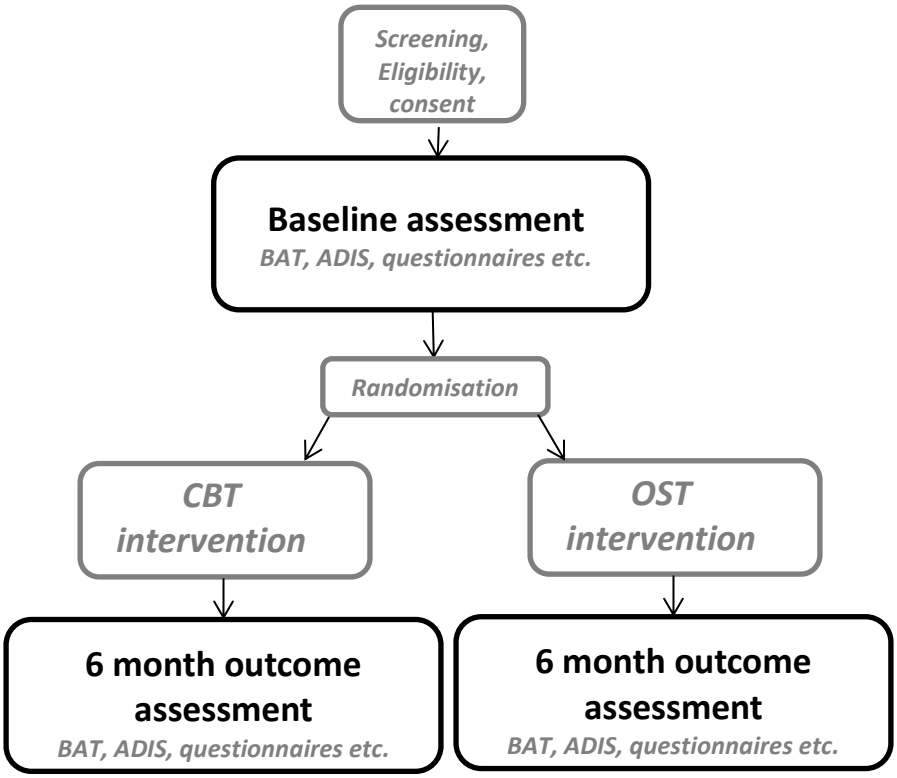
List of abbreviations used

ADIS	The Anxiety Disorder and Interview Schedule
ASPECT	Alleviating Specific Phobias Experienced by Children Trial
AE	Adverse Event
BAT	Behavioral Approach Test
CACE	Complier average causal effect
CAIS	The Child Anxiety Impact Scale
CBT	Cognitive Behavioural Therapy
CHU9D	Child Health Utility 9D
CI	Confidence Interval
CONSORT	Consolidated Standards Of Reporting Trials
CRF	Case report form
CSR	Clinician's severity rating
CTRU	Clinical Trials Research Unit
DMEC	Data Monitoring and Ethics Committee
DSM IV	Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
EQ-5D-Y	EQ-5D Youth version
FA	Functional assessment
GCP	Good Clinical Practice
HTA	Health Technology Assessment
ICC	Intraclass Correlation Coefficient
ISRCTN	International Standard Randomised Controlled Trials Number
ITT	Intention to Treat
NHS	National Health Service
NIHR	National Institute for Health Research
OST	One Session Treatment
PI	Principal Investigator
PP	Per protocol
QOL	Quality of Life
RCADS	Revised Children's Anxiety and Depression Scale
RCT	Randomised Controlled Trial
RET	Rapid exposure therapy
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOP	Standard operating procedure
SUDS	Subjective units of Distress
TMG	Trial Management Group
TSC	Trial Steering Committee

Summary Page

Trial title	A non-inferiority randomised controlled trial comparing the clinical and cost-effectiveness of one session treatment (OST) with multi-session cognitive behavioural therapy (CBT) in children with specific phobias
Trial design	Non-Inferiority, Parallel Group, Randomised Controlled Trial with internal pilot and nested qualitative component
Trial participants	Children (aged 7 to 16) with at least one DSM rated specific phobia
Sample size	286
Follow-up	6 months post randomisation. Participants will be assessed twice: at baseline and then again at 6 months post randomisation
Internal pilot	After a 9 month pilot phase the TSC will assess the feasibility of the trial based on the two outcomes: 1. 75% of the recruitment target met at 9 months (81 participants) 2. 70% retention of participants recruited in the first 3 months (anticipated 25/36 participants)
Primary analysis	BAT score at 6 months will be compared between groups using mixed effects linear regression, with adjustment for baseline BAT score and stratifying variables and site, with therapist as a random effect. Inferiority will be rejected if the 95% CI for the standardised effect size is wholly below 0.4 for CBT-OST. This will be presented for both Intention to treat (ITT) and per protocol populations and both will have to present evidence of non-inferiority to declare that OST is non-inferior to CBT.
Secondary analyses	Secondary outcomes (such as ADIS, CAIS, RCADS) will be compared between groups in a similar manner to the primary analysis.

Figure 1: Study outline



1 Introduction, study design and key trial objectives

This statistical analysis plan (SAP) is written in conjunction with the International Conference on Harmonisation topic E9 (Statistical principles for clinical trials, 2010), applicable statistical standard operating procedures from the University of Sheffield Clinical Trials Research Unit (CTRU) and trial documents (Protocol and Data Validation Specification). The trial will be conducted in accordance with Good Clinical Practice (GCP) in Clinical Trials (International Conference on Harmonisation, 2010).

This SAP will guide the Trial Statistician during the statistical analysis of all quantitative outcomes in order to answer the objectives of the study. It excludes the health economics evaluation and fidelity assessment/analysis (which will be described elsewhere).

All analysis will be performed in a statistical software package such as STATA version 14 (StataCorp, 2015).

1.1 Study outline

ASPECT is a non-inferiority parallel group randomised controlled trial with internal pilot and nested qualitative component. Its aim is to compare the clinical and cost effectiveness of one session treatment (OST) with multi-session cognitive behavioural therapy (CBT) in children with specific phobias. The study will take place across a range of health and social care settings including; Children and Young People's Improving Access to Psychological Therapies; Child and Adolescent Mental Health Services; supplementary and third sector organisations; and school based therapists.

The study is funded by the National Institute of Health Research (NIHR) Health Technology Assessment (HTA) programme. Leeds and York Partnership NHS Trust will act as the sponsor for this study.

1.2 Study objectives

Pilot objective

To assess the feasibility of trial recruitment and retention.

Primary objective

To investigate the non-inferiority of OST compared to CBT based interventions for treating specific phobias in children (aged 7 to 16 years old) at a 6 month follow up point. Non-inferiority would be demonstrated if OST is shown to produce similar, or improved, effects on the Behavioural Approach Test (BAT) (Öst et al, 1991) when compared with CBT.

Secondary objectives

1. To examine the cost effectiveness of OST, in comparison to CBT. It is hypothesised that OST, when compared with CBT, will be more cost/time effective.
2. To establish the relative impact of the interventions on the child's quality of life (QoL), school and social life as well as family functioning.
3. To establish the acceptability of OST to the children taking part in the trial as patients, their parents/guardians, and to the clinicians administering OST.

2 Outcome measures

Pilot endpoint

The endpoints for the pilot study are two recruitment/retention targets

3. 75% of the recruitment target met at 9 months (81 participants)
4. 70% retention of participants recruited in the first 3 months (anticipated 25/36 participants)

Primary outcome

1. Behavioural Approach Test (BAT) score - the number of steps the participant takes at 6 months post-randomisation (0-10)

Secondary outcomes

(Further details of which domains will be presented, how the outcomes will be scored and how missing items will be accounted for are presented in section 9.)

1. BAT subjective unit of distress measures fear/distress
2. Anxiety Disorder Interview Schedule (ADIS; Silverman & Albano, 1996) measures specific phobia symptoms and experiences in children
    a. Child version
    b. Parent version
3. Child Anxiety Impact Scale (CAIS; Langley et al, 2004) measures anxiety related functional impairment
    a. Child version
    b. Parent version
4. Revised Children's Anxiety and Depression Scale (RCADS; Chorpita et al, 2000) measures overall child mental health
    a. Child version
    b. Parent version
5. EQ-5D-Y (EuroQol, 1990) measures health related quality of life
6. Child Health Utility 9D (CHU-9D; Stevens & Ratcliffe, 2012) measures health and related quality of life

7. Goal Based Outcome Measure (Law & Jacob, 2013) measures how far participants self-report they are to achieving their goals set at baseline
8. Resource Use questionnaire collects participants' use of health, social care and community services for health economic analysis

3 Sample Size

To our knowledge, no systematic review has examined the effect of CBT on specific phobias as measured by the BAT in children. Consequently, the assumptions for the proposed sample size and non-inferiority margin are based on two separate Cochrane reviews looking at the effects of psychotherapy for those experiencing anxiety. Firstly, Wolitzky-Taylor, Horowitz, Powers, and Telch (2008) conducted a review on studies that used both behavioural measures and self-report questionnaires on adults with specific phobias and reported an overall, large effect size of d = 0.81. However, as the treatment may have a different effect on children, we also examined Reynolds, Wilson, Austin, and Hooper (2012). This review was conducted on studies of children with specific phobias but used self-report questionnaires rather than the BAT. This review also reported a large effect size (d = 0.85) for multi session CBT.

Consequently, prior meta-analyses suggest that a standardised mean difference of around 0.8 on the BAT scale is clinically important. Therefore, we set the non-inferiority margin to be half of this at 0.4 (Jones et al 1996). Assuming a correlation of 0.5 between baseline and final BAT measure, we would require 200 participants (100 in each arm) to have 90% power with a 2.5% one-sided significance level to demonstrate non-inferiority of One Session Treatment (OST) compared to cognitive behavioural therapy (CBT). The therapy is delivered by therapists who will see approximately 15 patients each and we anticipate a weak therapist effect (intraclass correlation coefficient (ICC = 0.01)). This clustering will lead to a design effect of 1.14 which increases the number required per arm to 114. We further assume a 20% dropout rate which means 286 (143 per arm) will need to be recruited to the study to demonstrate non-inferiority of OST compared to CBT.

3.1 Sample size recalculation

In April 2019, a seven month extension was requested from the funders. The sample size was recalculated and submitted with the extension request.

As of the 27th March 2019, data completeness on the primary endpoint was 64/88 (72.7%). This translates to a dropout rate of 27.3 % (95% CI; 18.3% to 37.8%). Based on this and the original timelines this would result in about 136 (68 per group) participants with 6 month primary outcome

data for analysis. We observed a correlation of 0.7 between baseline and 6 month primary outcome measures. We also observed that a therapist is now expected to treat 5 children (instead of 15). If the original assumption made about the ICC (of 0.01) is accurate, then the design effect will now be 1.04 (instead of the planned 1.14).

Given the above observations, if the study was extended to recruit the original sample size of 286 (143 per group), it would have a power of 97.7% (0.7 correlation, 27.3% drop out rate, 5 children per therapist).

Based on a conservative correlation of 0.6, observed dropout rate of 27.3%, each therapist to treat an average of 5 children, and an ICC of 0.01, then a total of 246 participants (123 per group) will be required to preserve a power of ~90% for a one-sided 2.5% test with a standardised non-inferiority margin of 0.4. This will give us 178 participants (89 per group) with primary outcome data for analysis.

4 Randomisation & Blinding

Participants will be randomised to two groups using the CTRU web-based randomisation system hosted by epiGenesys, a wholly owned subsidiary of the University of Sheffield. The randomisation allocation ratio will be 1:1 to facilitate equal group sizes across OST and CBT groups. The randomisation schedule will be generated by the trial statistician prior to the start of the study. Randomisation will be stratified according to age (7-11 years old vs. 12-16 years old) and symptom severity (as defined by ADIS clinician's severity rating (CSR) mild/moderate (scoring 4/5) vs. severe (scoring 6/7/8) phobia severity), and will use blinded random permuted blocks of variable size to ensure enough participants are allocated evenly to each arm of the trial within each stratum.

Research Assistants conducting baseline and follow up assessments will be blind to treatment group. Statisticians conducting the analysis will not be involved in the administration of the trial and will remain blind to treatment allocation during the course of the trial until data freeze. Reports to the Data Monitoring and Ethics Committee (DMEC) will be prepared by Data Management who are unblind. Unblinded statistical reports may also be provided to the DMEC on their request as guided by the DMEC charter by a Sheffield CTRU statistician who is external to the trial.

5 Interim analyses, data monitoring committees etc.

The following committees will be established:

1. **Data Management and Ethics Committee (DMEC)** – established with an independent chair that will adhere to the Standard Operating Procedure of the CTRU.
2. **Trial Steering Committee (TSC)** - consist of an independent chair, an independent subject specialist, an independent clinical academic, an independent statistician and a patient representative. The committee will meet approximately every 6 months from the start of the trial.
3. **Trial Management Group (TMG)** – oversee the day-to-day management of the trial and will comprise the core members of the team (Chief Investigator, Project Manager and direct research staff)

This trial has been designed with a fixed sample size and one formal statistical analysis at the scheduled end. Therefore there are no planned interim analyses to allow for early stopping. However, as the trial exposes children to feared objects, a Data Monitoring and Ethics Committee (DMEC) will review the data at the end of the pilot and every 6 months throughout the trial for safety. The DMEC may advise the TSC if there is evidence of harm due to the interventions or assessments and the trial may be stopped.

6 Data Sources, data evaluability and analysis populations

6.1 Data sources

The randomisation list will be held on the CTRU's randomisation system. Trial data will be extracted from source documents (including CRFs and participant questionnaires) and entered onto the CTRUs in house data management system (PROSPECT). The data management team in the Sheffield CTRU will validate and query electronic data for inconsistencies during the course of the trial (as stipulated in SOP DM005), The trial statistician will conduct any additional validation checks where appropriate before the data lock and sign off (as guided by DM005 and DM012).

Table 1: Details of data collected at each timepoint

Screening	Telephone screening
Eligibility & Consent	ADIS-P, ADIS-C
Baseline	Demographics ADIS-P, ADIS-C CAIS-P, CAIS-C RCADS-P, RCADS-C EQ-5D-Y CHU-9D

	Goal based outcome measure BAT Medications and resource use
Intervention	comorbidity assessment, OST competency , CBTS-CYP, Session recording form
6 months	ADIS-P, ADIS-C CAIS-P, CAIS-C RCADS-P, RCADS-C EQ-5D-Y CHU-9D Goal based outcome measure BAT Medications and resource use
Ongoing	Adverse Events

6.2 Study population

6.2.1 Inclusion criteria

1) Be between the ages of 7 and 16 years of age
2) Experience at least one specific phobia as defined by DSM-IV criteria, which will be assessed using the specific phobia subsection of the Anxiety Disorder Interview Schedule (ADIS). These criteria are; i) marked and out of proportion fear to a specific object or situation; ii) exposure provokes immediate anxiety; iii) the phobic situation(s) is avoided where possible; iv) the avoidance or distress interferes with the person's routine or functioning (e.g. learning, sleep, social activities); v) and present for 6 months or more.

6.2.2 Exclusion criteria

1) Children for whom exposure therapy has the potential to be unsafe or cause harm. For example, children who have severe allergies or severe asthma or cardiac problems which may be exacerbated and not possible to manage safely in the context of exposure therapy to certain stimuli (e.g. insects, furry animals) or while using associated techniques such as applied tension (e.g. hospitals)
2) Children for whom exposure therapy is not feasible in the context of this study. For example, children with learning or developmental disabilities or conduct disorders who present with a phobia may be excluded if they need specialist

support and a tailored therapy programme that cannot be routinely provided by the mainstream services delivering exposure therapy for this research project.

3) Children for whom exposure therapy is not the best first line or best available option. For example, children/young people who need to be signposted for a further assessment for possible "at risk mental state" for suspicion of psychosis that has not been recognised prior to referral into the study; or children with a phobia complicated by an eating disorder (e.g. fear of choking or swallowing with anorexia) who need a more complex intervention than exposure therapy. Another example is children/young people who self-harm in response to anxiety and whose self-harming behaviours may increase in response to exposure therapy and need an intervention to help them with minimising and managing self-harm prior to starting exposure therapy.

6.3 Definition of the analysis populations

The following analysis populations will be studied in the analyses:

Name	Participants included	Treatment group
Intention to treat (ITT)	All randomised participants according to the randomised treatment assignment with complete primary outcome data with the following exclusions: • Previously randomised • No recorded consent information • Withdrew before randomisation This includes participants found to be ineligible post randomisation (Altman, 1991)	As randomised
Per Protocol (PP)	The subset of ITT who receive their intervention in accordance with the protocol (defined below).	As randomised
Complier average causal effect (CACE)	CACE 1 –CBT delivered in accordance with protocol CACE 2 – OST delivered in accordance with protocol CACE populations are explained further in sections 6.3.1 and 8.2	As randomised

Per protocol

- A detailed definition of per protocol for the OST group is given in Section 11. In brief, a participant in the OST group is defined as per protocol if they attend: One assessment session
- One main exposure session
- An optional extra session

And all of the following have taken place during the therapy;

- An assessment

- Establishment of a fear hierarchy
- Exposure

A participant in the CBT group is defined as per protocol if they;

- Attend at least 4 CBT sessions.

If a participant is still undergoing therapy by the time they complete their 6 month follow up assessment, only sessions conducted before the 6 month follow up will be counted towards the per-protocol assessment.

6.3.1 The role of the analysis populations

Jones, Jarvis, Lewis, and Ebbutt (1996) recommend both per protocol and intention to treat (ITT) analyses for non-inferiority designs. This is because in a comparative trial, where the aim is to decide if two treatments are different, an ITT analysis is generally conservative, the inclusion of protocol violators and withdrawals will usually make the results from the two treatment groups more similar. However, for an equivalence or non-inferiority trial this effect is no longer conservative, any blurring of the difference between the treatment groups will increase the chance of declaring equivalence. We follow this recommendation with the refinement that the main analysis of the primary outcome will be per protocol (or completers only) with sensitivity analysis on the ITT population (Piaggio et al., 2012). We will require both the 'per protocol' and ITT analyses to demonstrate statistically significant evidence of non-inferiority to declare that the treatment is non-inferior. If the results of the analysis are discrepant (e.g. the ITT rejects the null of inferiority but the 'per protocol' analysis does not, or vice versa) then we will report the conflicting results from both analyses highlighting the inconclusive nature of the results.

CACE analyses are increasingly used as an attempt to remove "non-receiver" participants whilst retaining a like-for-like comparison. In brief, CACE is an attempt to compare participants who undergo their intervention in accordance with protocol to those in the comparator group who are "likely" to have had they been randomised to receive it. Further details are given in section 8.2.

Analysis will be conducted on the primary outcome (BAT at 6 months) using all analysis sets (ITT, PP, CACE). For all secondary outcomes the analysis will be reported on the ITT population unless there are important differences between results based on the ITT and PP set. As a guideline, differences between the ITT and PP estimated treatment difference of more than 0.1SD on any inventory will be assessed further.

7 Outline of analyses

Data will be reported according to the Consolidated Standards Of Reporting Trials (CONSORT) statement and the extension for non-inferiority trials(Piaggio et al., 2012; Schulz, Altman, & Moher, 2010).

7.1 General considerations

Summaries of continuous variables will comprise the number of observations used, mean, median, standard deviation, inter-quartile range, minimum and maximum as appropriate for the distributional form of the data.

Summaries of categorical variables will comprise the number of observations used, and the number and percentage of observations in each category. Tables containing the results of the statistical modelling will present the overall difference between treatment groups with 95% confidence intervals (CI).

Complete details of data derivations and methods of handling missing data are covered in sections 8.1 and 9.

7.2 Internal Pilot Analysis

The TSC will assess the feasibility outcomes at the end of the pilot phase. They will consider whether the trial should continue in light of the feasibility results against the STOP/GO criteria listed in section 2. The results and recommendations will be communicated to the funder (NIHR HTA). If the result of the internal pilot analysis is to stop the trial the trial will be written up and reported according to the updated CONSORT statement for pilot studies (Chan et al, 2016). Summary statistics and confidence intervals will be presented, however no non-inferiority hypotheses will be tested using confidence intervals.

7.3 Data Completeness

A CONSORT style diagram will be presented to summarise the flow of participants through the trial, from screening, during follow up and inclusion in to primary analysis. An example CONSORT diagram is shown in section 12, Figure 2. Data completeness will be based on the primary outcome, BAT at 6 months post randomisation. This information will be made available to the TMG, TSC and DMEC on their request during the course of the trial.

7.4 Baseline characteristics

Summaries of baseline variables relating to child and parent socio-demographics and to baseline QoI measures will be presented by treatment group and overall (as in section 12; Table 2 and Table 5). The baseline data will be assessed for comparability between groups, any noted differences will be described and considered for adjustment in primary and secondary analyses. No statistical testing will be undertaken on baseline data. The following summaries will be presented:

Child socio-demographics	Categorical variables • Delivery site • Sex • Ethnicity • School attender (yes/no) • Treatment preference • In another study/trial (yes-phobia/anxiety, yes – other, no) Continuous variables • Age • Number of adults in household • Number of children in household
Parent/carer socio-demographics	Categorical variables • Sex • Ethnicity • Highest educational qualification • Employment status • Relationship to child • Treatment preference • Marital/partner status • Partner's employment status Continuous variables • Age
Baseline assessments and Quality of life measures	ADIS-P, ADIS-C CAIS-P, CAIS-C RCADS-P, RCADS-C EQ-5D-Y CHU-9D Goal based outcome measure BAT

7.5 Therapist Summaries

It is anticipated that all therapists will deliver both OST and CBT. The following summaries will be presented for all therapists.

- Treatment preference (No preference, prefer CBT, prefer OST)
- Highest level of qualification

- Number of years delivering psychological interventions/therapies with children or young people
- Organisation and role (these are recorded as free text and will be categorised where possible)

7.6 Treatment Summaries and Compliance

Participants in the OST arm are expected to attend an initial 1-hour functional assessment (FA) session and a separate 3-hour rapid exposure therapy (RET) session. There is no recommended number of CBT sessions for specific phobias; however, it is often the case that a child receives 6-to-12 sessions of CBT.

The following compliance summaries will be presented for each treatment group for the ITT population:

- Attendance –the number of therapy sessions attended
- Early termination of therapy
    - the number and percentage of participants that withdraw from the study and from treatment
    - The reasons for termination of therapy
- The number and percentage of participants that completed further treatment for phobia

The following descriptive summaries of treatment will be presented by treatment group for the ITT population

- Total therapy duration (hours)
- Strategies used

Table 9 gives an example table of treatment compliance.

The number of sessions attended for each treatment group will also be investigated by site, if there is large amount of switch-over (for example OST participants receiving subsequent CBT sessions before 6 month follow up) at one or two sites we will perform sensitivity analyses on the ITT primary outcome analysis (described in Section 7.8.1) where the data from these sites are removed from the analysis.

7.7 Characteristics of the Per-protocol population

As the study was ongoing it became clear that a larger than anticipated amount of participants were not receiving treatment per-protocol. We will compare the per-protocol sample to the wider ITT group in order to investigate what characteristics predict whether a participant will have treatment per-protocol, this will be modelled using a multinomial logistic regression model for each treatment group. For the OST group the session attendance will be modelled using multinomial logistic regression with three possible outcomes:

- Underwent the FA and RET, (and up to one extra session) only [reference]
- Underwent FA and RET and additional sessions
- Did not attend any sessions (or attended FA only)

Correspondingly for the control arm the three possible outcomes will be:

- Attended at least four CBT sessions (per protocol) [reference]
- Attended three or fewer sessions
- Did not attend any sessions

Predictor covariates may include age, sex, measures of ethnicity, deprivation (using postcode data to calculate Indices of Multiple Deprivation), socio economic status (based on data collected for the parent/carer) , comorbidities, phobia severity (based on ADIS CSR and baseline BAT score), and phobia type.

The difference between the different compliance groups will be described.

7.8 Efficacy

The usual care arm (CBT) will be the reference group for the analysis unless stated otherwise.

7.8.1 Primary Outcome

The primary outcome (mean BAT score at six months) will be compared between groups using mixed effects linear regression with robust standard errors and exchangeable correlation to allow for the clustering of outcomes within therapist. The analysis will be conducted controlling for baseline BAT score, site and stratifying variables (age and baseline phobia severity - ADIS CSR) as fixed effects. Mean difference and 2 sided 95% confidence interval will be presented. This analysis will be performed on the per protocol population. The null hypothesis of inferiority will be rejected if the two sided 95% confidence interval (CI) for the standardised difference is wholly below 0.4 (the range of clinical non-inferiority). The standardised difference will be calculated using the pooled standard deviation. It will also be presented on the raw scale. If baseline imbalances have been observed a further analysis adjusting for these variables will be conducted and presented alongside the above analysis.

In order to display the effect of the active control (CBT), the change in BAT at 6 months will be displayed for both treatment groups with corresponding 95% confidence intervals.

7.8.2 Sensitivity analysis on primary outcome

Sensitivity analyses on the primary outcome will be undertaken based on ITT, CACE, mistimed measurements and imputation populations. Mistimed measurements exclude participants from the per-protocol set with BAT follow up measurements outside a time window of 4 weeks before to 6 weeks after the 6 month (approx. 183 days) follow up window.

The results will be presented alongside the analysis on the PP set. We will require both the 'per protocol' and ITT analyses to demonstrate evidence of non-inferiority to declare that the treatment is non-inferior. The primary analysis will be repeated for these data sets and displayed alongside the PP analysis results (see section 12 Table 10). They will also be reported using a forest plot, as illustrated in section 23, Figure 3. Imputation will be conducted on both the ITT and PP analyses of the primary outcome (see section 8.1.1 for more information on imputation analysis).

BAT score at baseline

Blinded review of the baseline data by the TMG revealed that approximately 25% of participants are recording scores on the ceiling of the BAT at the baseline. That is, participants are attaining the highest score on the BAT by completing all ten steps. Consequently, these participants cannot show imrpovement during the trial and this may dilute the treatment effect towards non-inferiority claim depending on the underlying distribution of these ceiling BAT scores (which is unknown at this stage). Noteworthy, most of these participants have phobias relating to blood-injection injury. The TMG, therefore, requested additional sensitivity analysis on the primary outcome by excluding participants who received simulated BAT stimuli. This analysis will be repeated on the per protocol and ITT analysis populations using the same statistical analysis approach described for the primary outcome in Section 7.8.1.

7.8.3 Secondary outcomes

Secondary outcomes (BAT SUDS, ADIS, CAIS, EQ-5D-Y, CHU-9D, RCADS and the goal-based outcome scores) will be compared between the groups using a mixed effects linear regression model as for the primary outcome. The mean difference and 95% CI will be presented. However, no p-values will be reported due to the non-inferiority nature of the study.

7.9 Safety and Harms

Any untoward occurrence affecting the participant will be recorded as an adverse event after each therapy session and follow up visit. An occurrence is recorded if it is suspected to be related to the intervention or an aspect of the research procedures.

An AE will be recorded as a serious adverse event (SAE) if it:

- Results in death
- Is life-threatening
- Requires hospitalisation or prolongation of existing inpatients hospitalisation
- Results in persistent or significant disability or incapacity

Descriptive statistics of Adverse Events and Serious Adverse Events will be calculated and reported by treatment group and overall. Safety data will be presented on the intention to treat population, but the events for the PP will be investigated and both presented if there is sufficient difference.

The following summaries will be presented

- Number (%) participants experiencing ≥1 AE
- Number of all AEs including repeat events
- Number of AEs by category (post hoc categorisation from free text field where possible)
- Number (%) participants experiencing ≥1 SAE
- Number of all SAEs including repeat events
- SAE by Seriousness (Death, Life threatening, Inpatient hospitalisation, Prolongs hospitalisation, Persistent or significant disability/incapacity, congenital abnormality/birth defect)
- SAE by Intensity (mild, moderate, severe)
- SAE by Relationship to study intervention (definite, probable, possible, unlikely, unrelated, not assessable)

7.10 Unblinding of outcomes

As highlighted in section 4, Research Assistants conducting outcome assessments will be blinded to intervention allocation in order to minimise operational bias. If unblinding of treatment allocation does happen the following summaries will be presented by treatment group

- Number and proportion of unblinded cases,
- Source of unblinding (therapist, participant, etc),
- Method of unblinding (face to face, phone, etc),
- Reason for unblinding,
- Number considered definite and probable unblinding,
- Recording of incorrect unblinding (suspecting the wrong treatment group)

7.11 Subgroup analysis

The objective of an exploratory subgroup analysis is to explore heterogeneity in the intervention effects across pre-defined subgroups. An exploratory subgroup analysis will be performed using mixed effects linear regression with the primary outcome, BAT score at 6 months post randomisation as the response. An interaction statistical test between the randomised treatment group and subgroup will be used to directly examine the strength of evidence for the difference between

treatment group (CBT vs OST) varying between subgroups. Four subgroups of interest have been pre-specified:

- Participant treatment preference (OST/CBT)
- Therapist preference (OST/CBT)
- BAT stimulus set up (simulated/real stimuli)
- Phobia type

Phobia type will be defined based on the groupings used in the ADIS (animal type, Natural Environment Type, Blood-injection or Injury Type, Situational Type, Other Type). Particular attention will be paid to phobia types that trigger different physiological responses (e.g. blood-injection or injury type). Noteworthy, more emphasis here is given to the subgroup analyses relating to BAT stimulus set up and phobia type.

Subgroup analysis will be performed regardless of the results of the primary analysis. The mean difference in BAT scores (95% CI) between treatment groups will be computed for each subgroup category and visually displayed using a forest plot (Cuzick, 2005). The regression coefficient for the interaction between treatment group and subgroup will be presented with the associated confidence interval and P-value. We will not calculate separate p-values within each subgroup category (Assmann et al., 2000; Pocock et al., 2002; Wang et al., 2007). Results will also be presented as shown in Table 12. The participant treatment preference is asked of the child, however it may be a joint decision between child and parent if the parent is completing the form on behalf of the child.

7.12 Assessment of post-treatment follow up time in relation to outcome

The duration of CBT (usually between 6 and 12 weeks) is longer than OST (one FE and one RET session). This means the time between finishing treatment and the 6 month post randomisation follow up may differ between treatment arms, which may affect the size of the treatment difference as estimated at six months. To investigate this further, the time between last treatment session and follow up assessment will be summarised by arm and a secondary analysis will be performed in which a covariate term for post treatment follow up time is included in the analysis. Further to this an interaction term will be fitted between treatment allocation and post treatment follow up time.

7.13 Baseline anxiety and depression in relation to outcome

The moderating effect of baseline anxiety and depression (as measured by RCADS) will be investigated. The primary analysis will be repeated with a covariate of baseline RCADS and an interaction term of RCADS*Treatment group. This will be performed for

- RCADS major depression domain

- RCADS total anxiety score

The regression coefficient for the interaction between RCADS and treatment group will be presented with its associated confidence interval and p-value.

7.14 Impact of Site and Therapist

A further sensitivity analysis will investigate whether the treatment difference is consistent across site and therapists. Site effect will be explored through an interaction test between site and treatment. The treatment difference at each site will be presented using a forest plot; estimates of the mean difference will be calculated using a mixed effects regression analysis adjusted for therapist.

We anticipate that therapists will, on average, see too few participants each to make meaningful comparisons (anticipated 50+therapists delivering the interventions). The intraclass correlation coefficient of the clustering variable (therapist) in the primary analysis model will be reported as a measure of evidence that there are differences in outcomes between therapists.

Therapists will be classified into categories to investigate whether treatment differences are consistent across therapist characteristics. The characteristics to be investigated will be;

- Setting (CAMHS or not)
- Role and grade
- Highest level of qualification
- Number of years delivering psychological interventions/therapies with children or young people

Each characteristic will be split into categories and approved by the chief investigator without seeing any outcome data. This exploratory analysis will be performed in the same way as subgroup analysis (see section 7.11) an interaction statistical test between the randomised treatment group and characteristic will be used. The regression coefficient and p value for this interaction will be presented alongside forest plots of the treatment difference in each category.

I

8 Detailed statistical methods and calculations

8.1 Missing, spurious and unused data

We anticipate some dropout/attrition so missing data may be an issue. Imputation of the primary outcome (BAT score at 6 months) will be used to reduce bias due to any missing responses in both the “per protocol” and the ITT analyses. Missing data will be imputed using multiple imputation. The proportion of missing data in each group will be reported and compared descriptively. Descriptive statistics of baseline variables will be presented by treatment group and missing data status as

illustrated in 12, Table 8 and Table 7. Baseline variables found to be predictive of missing data status will be included in multiple imputation equations.

8.1.1 Multiple imputation

One hundred multiple imputation data sets will be created using chained equations. The multiple imputation equation will include baseline data (for example age, sex, baseline BAT) and predictors of missing data to make the Missing at Random assumption as plausible as possible. A conservative approach will be adopted and treatment group will be excluded from the imputation model.

8.2 CACE analysis

In the context of a comparison between an active and inactive intervention, CACE would be carried out by the following steps:

i. Using participants in the intervention group, derive a logistic regression model to predict the probability of being a receiver. See section 7.7 for details of the model and possible predictor covariates.
ii. Apply these predictions to the control group, whereby each participant is given a probability of undergoing the intervention as planned (had they been randomised to receive it) which is based on their covariates.
iii. For each participant in the non-intervention group, calculate a re-weighted outcome defined as the original outcome multiplied by the predicted probability of receiving as planned.
iv. Compare the subset of participants in the intervention group that are deemed to have received intervention as planned to the re-weighted outcomes among participants in the control group.

CACE analyses will be conducted using 2-stage least squares regression. Since in this trial both arms contain an active intervention, a bi-directional CACE will be undertaken. The CACE1 population is defined as the complier-averaged effect for CBT, and estimated from participants who received CBT (in the CBT group) or were predicted to have undergone CBT as per the protocol (in the OST group). Likewise, the CACE2 population does the same with OST. The model used in part i. of the CACE analysis will be that described in section 7.7. Any participant that was randomised to one treatment, but received and complied with the other treatment will be removed from the CACE analysis.

8.3 Mixed effects model checks

Model goodness of fit will be investigated via graphical methods (e.g. histograms of residuals and scatterplots of residuals vs. covariates) for primary and secondary endpoints. Checks will be made on whether an exchangeable correlation assumption is reasonable. Influential observations and

outliers will also be investigated and sensitivity analyses at the discretion of the trial statistician will be undertaken and reported.

As the trial was in progress it was noted that a lot of therapists were seeing few patients, and so the clusters in the mixed effects model may be smaller than anticipated. We will perform a simple regression for the ITT and PP analysis to assess the stability of the results from the mixed effects model,

9 Data manipulation and definitions

9.1 Definitions

- BAT stimulus set up will be defined as real or simulated, this will be decided based on review by the principle investigator at each site (or a delegate).

9.2 BAT

If a visit was attended but a BAT not completed (due to fear – as recorded on the data collection forms) then a score of 0 will be imputed for the BAT.

Further to the primary outcome, the following aspects of the BAT will be reported:

- SUDS (pre avoidance task), measures anxiety/fear on 0-8 scale
- Change in SUDS (post – pre task)

Caution will be taken when interpreting change in SUDS, as the participants post-task units of distress will be dependent on the number of steps they took on the task.

9.3 ADIS

The Anxiety Disorder Interview Schedule (ADIS) specific phobia subsection is a semi-structured interview. The Composite Clinician's severity rating (CSR) will be presented only for the specific fear that is being treated. The ADIS CSR is based on a 0-8 scale and is guided by the parent's and child's ratings, total number of symptoms endorsed and the clinician's impression for each diagnostic category. A SOP was developed whilst the trial was ongoing to guide the way the research assistants score the ADIS. Any occurrences where the ADIS was scored by the research assistant in a different way to the SOP will be described. Missing item data will not be imputed as it is a single score.

9.4 CAIS

The CAIS-P (parent version) and CAIS-C (child version) are 27-item questionnaires assessing the impact of the child's anxiety symptoms on psychological function (Langley, 2004). Each question is scored on a 4 point scale ("0" not at all, "1" just a little, "2" pretty much, "3" very much). A total score for each version is calculated by summing the scores for each question. The score ranges from 0 to 81 and higher scores represent greater impact on psychological function. CAIS scores can be calculated if no more than 3 items are missing. Missing items will be imputed using the mean of the available items.

9.5 RCADS

RCADS-P (parent version) and RCADS-C (child version) are questionnaires that consist of 47 questions scored on a 4 point scale (from "0" never to "3" always). RCADS consists of 6 domains:

- Social Phobia (Q 4, 7, 8, 12, 20, 30, 32, 38, 43)
- Panic Disorder (Q 3, 14, 24, 26, 28, 34, 36, 39, 41)
- Major Depression (Q 2, 6, 11, 15, 19, 21, 25, 29, 40, 47)
- Separation Anxiety (Q 5, 9, 17, 18, 33, 45, 46)
- Generalized Anxiety (Q 1, 13, 22, 27, 35, 37)
- Obsessive-Compulsive (Q 10,16, 23, 31, 42, 44)

Domain scores are calculated by summing the scores for each domain. Domain scores can be calculated if no more than 2 items are missing (Chorpita, 2015). In the case of missing item data the following scoring approach is adopted:

- Calculate mean score for the participant from the available items in the domain
- Impute missing items with mean score
- Generate total by summing all scores
- Round to the nearest integer

Total Anxiety score is calculated by summing the anxiety related domain scores (Social Phobia, Panic Disorder, Separation Anxiety, Generalized Anxiety, Obsessive-Compulsive). Total anxiety score can be calculated if all contributing domain scores are calculated (i.e. if a maximum of 2 items are missing on each domain). The total anxiety score ranges from 0 to 111 with a higher score representing more anxiety symptoms (Choprita, 2015). Summaries on all domain scores and the total anxiety score will be presented.

9.6 EQ-5D-Y

The EQ-5D-Y consists of 5 questions (measuring mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) each with three possible answers. The scores are assigned utility values and combined to give a value index score. In the absence of specific utility value sets for the EQ-5D-Y, value sets for the EQ-5D-3L will be used. The score ranges from –0.594 to 1.00 (a score of zero means death, 1 is full health and a negative score is a state worse than death). The score will not be calculated if any items are missing. The algorithm for scoring the EQ-5D-3L can be found in the Database Specification. The EQ-5D VAS your health state today is measured on a scale from 0 (worst imaginable health state) to 100 (best imaginable health state).

9.7 CHU-9D

The CHU9D is a paediatric generic preference based measure of health related quality of life. The questionnaire consists of 9 dimensions each with 5 possible responses. A utility score is calculated by assigning utility values to each response and then summing these values. A CHU-9D score can only be calculated if all questions have been answered. The CHU-9D utility score ranges from 0.33 to 1 where higher scores represent greater health related quality of life.

9.8 Goal based outcome measure

The goal based outcome measure is used to set up to three goals at baseline. Progress towards meeting the goals are rated from 0 (goal not met) to 10 (goal reached) at baseline and then at 6 month follow up. Summaries on the number of steps towards meeting goal 1 will be presented. If sufficient numbers of participants have set goal 2 and goal 3 then these will also be presented. This will be at the discretion of the trial statistician.

9.9 Questionnaire summary table

Name	No. of items	Score range	Description	Interpretation of score
ADIS	na	0-8	Measures fear, and interference of specific phobias	Higher ADIS CSR means phobia is more disturbing 0 (none) to 8 (very severely disturbing/disabling)
CAIS	27	0-81	measures impact of anxiety symptoms on psychological function	higher score represents greater impact
RCADS	47	0-111 (total anxiety score)	measures anxiety and depression symptoms	higher score represents more anxiety/depression symptoms
CHU-9D	9	0.33-1	measures health related quality of life	Higher score represent greater health related quality of life
Goal based outcome	3	0-10	Measure of closeness to achieving goals	0(goal not at all met), 10 (goal reached)

EQ-5D-Y value index	5	-0.594-1	Measure of health status. 5 domains include mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.	A score of zero means death, 1 is full health, negative score is a state worse than death
EQ-5D-Y VAS	1	0- 100	Measure of health status.	A score of zero means worst health and 100 means best health.

10 Implementation of the original analysis plan

This SAP will be used as a work description for the statistician involved in the trial. All analyses should ideally be performed by the same statistician (under the supervision of senior trial statistician Dr Dawn Teare) and consequently, none of the investigators involved in the trial will perform any of the statistical analyses.

Initially, the data manager will provide blinded data for preliminary checks by the statistician. Following database freeze, unblinded data will be delivered to the statistician to define analysis sets and test statistical programs. Any queries will be communicated to the data manager prior to database lock, and any changes to the database during this time will be documented. The database will be locked after an agreement between the statistician, data manager and study manager. It is expected that no data amendments should be required following database lock. However, if an amendment is required, the process is documented in CTRU SOP DM012.

11 Per protocol Definition for OST

To determine whether the OST delivered adheres to that regarded as per protocol, 2 elements need to be considered – 1) the number of sessions delivered and 2) the techniques incorporated in the treatment

Session number

OST will be regarded as per protocol if it comprises:

- One assessment session
- One main exposure session (lasting up to 3 hours)
- An optional extra session – this could be a follow up session or a session where parents/carers/young person have requested a refresher session to consolidate therapeutic aims (this may be as a result of compliance to usual follow up/outcome monitoring practice within some clinical services)
- Any additional telephone contact completed as part of routine follow-up to offer reassurance and encouragement if requested or deemed necessary

If the participant receives subsequent sessions for CBT or exposure therapy (recorded on the session recording form) before 6-month outcome is collected they will not be classed as per protocol.

NB. Where only one session form has been recorded on PROSPECT this will only be deemed to adhere to per protocol if both sessions outlined in section (a) have been reported on the same form

NB. If the assessment session is conducted over two meetings this will be counted as one assessment session (e.g. where the child needed a break).

Session techniques

OST will be regarded as per protocol if all of the below have taken place during the therapy:

a. An assessment

b. Establishment of a fear hierarchy

c. Exposure

This will be evidenced through an examination of the session recording forms: OST will be regarded as per protocol if the following are each recorded on at least one of the session recording forms.

- 'Assessment and monitoring'
- 'Establishing a fear hierarchy'
- 'Exposure/behavioural experiments'

Where no session recording form has been completed the clinician will be contacted to ascertain that they have used these techniques and directed to report this on the session recording form on PROSPECT.

12 Example Tables and Figures

This section includes example tables and figures. The lists of data displayed in the tables are not comprehensive and are included only as an example.

Figure 2: CONSORT diagram: participant flow through the study

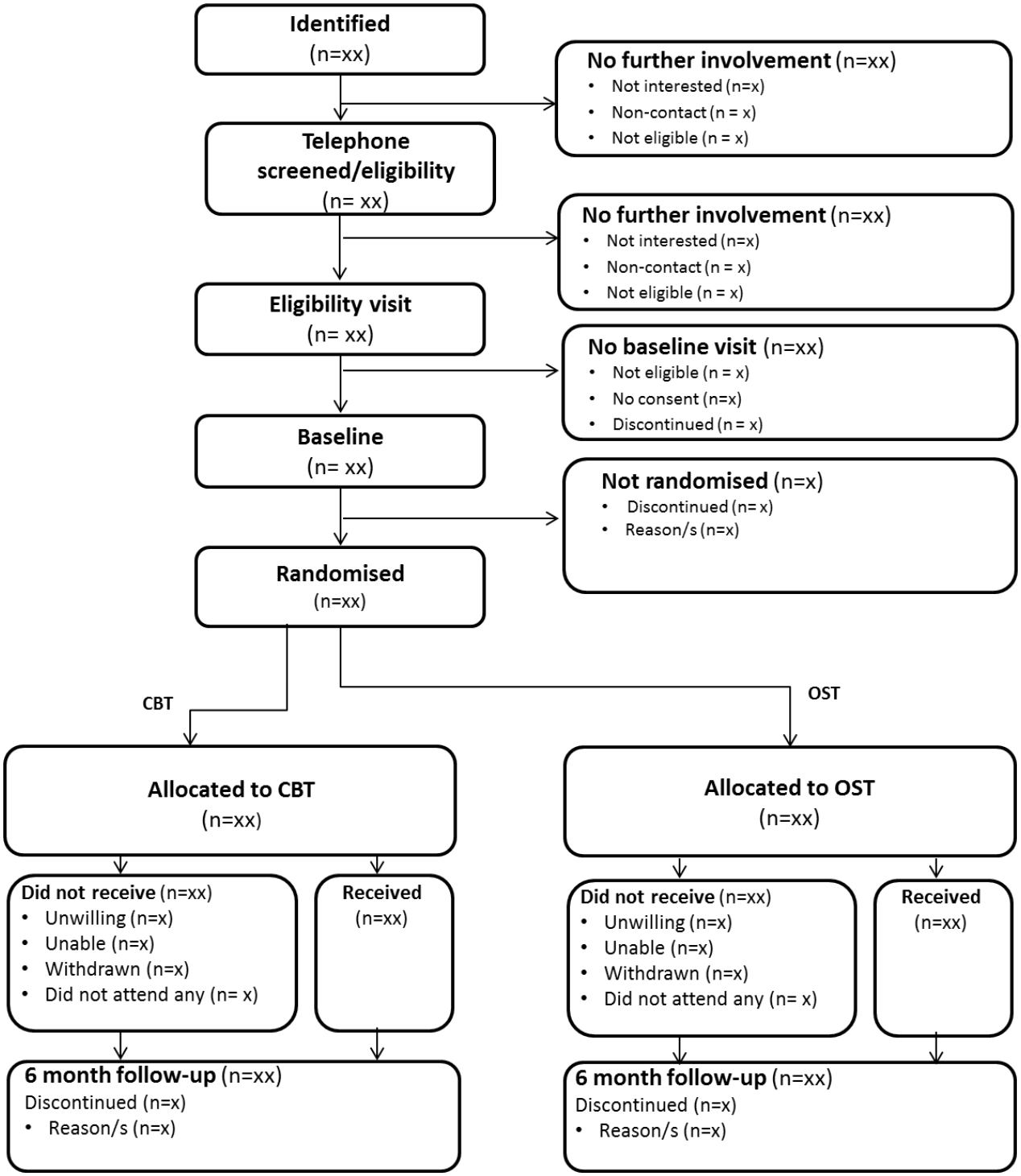


Table 2: Socio-demographics and characteristics of child participants

Variable	Scoring	CBT (n=xx)	OST (n=xx)	All (N=xx)
Site	Leeds	xx(xx%)	xx(xx%)	xx(xx%)
	York	xx(xx%)	xx(xx%)	xx(xx%)
	Sheffield	xx(xx%)	xx(xx%)	xx(xx%)
	Norfolk & Suffolk	xx(xx%)	xx(xx%)	xx(xx%)
	North east Cumbria	xx(xx%)	xx(xx%)	xx(xx%)
Sex	Male	xx(xx)	xx(xx)	xx(xx)
	Female	xx(xx)	xx(xx)	xx(xx)
Age (years)	Mean (SD)	xx(xx)	xx(xx)	xx(xx)
	Median (IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
Ethnicity [a]	White [b]	xx(xx%)	xx(xx%)	xx(xx%)
	Mixed/multiple ethnic groups [c]	xx(xx%)	xx(xx%)	xx(xx%)
	Asian/Asian British [d]	xx(xx%)	xx(xx%)	xx(xx%)
	Black/African/Caribbean/Black British [e]	xx(xx%)	xx(xx%)	xx(xx%)
	Other ethnic group [f]	xx(xx%)	xx(xx%)	xx(xx%)
	Prefer not to say	xx(xx%)	xx(xx%)	xx(xx%)
Attends school	Yes	xx(xx%)	xx(xx%)	xx(xx%)
	No	xx(xx%)	xx(xx%)	xx(xx%)
Treatment preference	OST	xx(xx%)	xx(xx%)	xx(xx%)
	CBT	xx(xx%)	xx(xx%)	xx(xx%)
In another study/trial	Yes – phobia/anxiety	xx(xx%)	xx(xx%)	xx(xx%)
	Yes – other	xx(xx%)	xx(xx%)	xx(xx%)
	No	xx(xx%)	xx(xx%)	xx(xx%)
Number of adults in household[g]	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	3 or more	xx(xx%)	xx(xx%)	xx(xx%)
Number of children in household[g]	Mean (SD)	xx(xx)	xx(xx)	xx(xx)
	Median (IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx

. [a] Main ethnic groups could be collapsed depending on the observed distribution. [b] White: English/Welsh/Scottish/Northern Irish/British, Irish, Gypsy or Irish Traveller, and Any other White background; [c] Mixed/multiple ethnic groups: White and Black Caribbean, White and Black African, White and Asian, and Any other mixed/multiple ethnic groups background; [d] Asian/Asian British: Indian, Pakistani, Bangladeshi, Chinese, and Any other Asian background; [e] Black/African/Caribbean/Black British: African, Caribbean, and Any other Black/African/Caribbean/Black British background; [f] Other ethnic group: Arab, and Any other ethnic group. [g]May be presented continuously or in categories depending on distribution

Table 3: Socio-demographics and characteristics of parent/carer

Variable	Scoring	CBT (n=xx)	OST (n=xx)	All (N=xx)
Sex	Male	xx(xx)	xx(xx)	xx(xx)
	Female	xx(xx)	xx(xx)	xx(xx)
Age (years)	Mean (SD)	xx(xx)	xx(xx)	xx(xx)

	Median (IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
Ethnicity [a]	White [b]	xx(xx%)	xx(xx%)	xx(xx%)
	Mixed/multiple ethnic groups [c]	xx(xx%)	xx(xx%)	xx(xx%)
	Asian/Asian British [d]	xx(xx%)	xx(xx%)	xx(xx%)
	Black/African/Caribbean/Black British [e]	xx(xx%)	xx(xx%)	xx(xx%)
	Other ethnic group [f]	xx(xx%)	xx(xx%)	xx(xx%)
	Prefer not to say	xx(xx%)	xx(xx%)	xx(xx%)
Highest educational qualification	Primary or less	xx(xx%)	xx(xx%)	xx(xx%)
	Secondary	xx(xx%)	xx(xx%)	xx(xx%)
	Higher	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Employment status	Employed full-time	xx(xx%)	xx(xx%)	xx(xx%)
	Employed part-time	xx(xx%)	xx(xx%)	xx(xx%)
	...	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Relationship to child	Parent	xx(xx%)	xx(xx%)	xx(xx%)
	Other relative	xx(xx%)	xx(xx%)	xx(xx%)
	Other adult in loco parentis	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Marital/partner status	Single – living in a couple	xx(xx%)	xx(xx%)	xx(xx%)
	Single – not living in a couple	xx(xx%)	xx(xx%)	xx(xx%)
	Married – co-habiting	xx(xx%)	xx(xx%)	xx(xx%)
	...	xx(xx%)	xx(xx%)	xx(xx%)
	prefer not to say	xx(xx%)	xx(xx%)	xx(xx%)

[a] Main ethnic groups could be collapsed depending on the observed distribution. [b] White: English/Welsh/Scottish/Northern Irish/British, Irish, Gypsy or Irish Traveller, and Any other White background; [c] Mixed/multiple ethnic groups: White and Black Caribbean, White and Black African, White and Asian, and Any other mixed/multiple ethnic groups background; [d] Asian/Asian British: Indian, Pakistani, Bangladeshi, Chinese, and Any other Asian background; [e] Black/African/Caribbean/Black British: African, Caribbean, and Any other Black/African/Caribbean/Black British background; [f] Other ethnic group: Arab, and Any other ethnic group

Table 4: Baseline assessments

Variable	**Scoring**	**CBT**	**OST**	**All**
		(n=xx)	**(n=xx)**	**(N=xx)**
BAT		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Media(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
BAT SUDS post-task		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to max	xx to xx	xx to xx	xx to xx
BAT SUDS change (pre to post task)		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
ADIS- CSR		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
		(n=xx)	(n=xx)	(N=xx)

Goal-based outcome: steps towards goal 1	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx

Table 5: Baseline Quality of life measures

Variable	Scoring	CBT (n=xx)	OST (n=xx)	All (N=xx)
CAIS-C (total score)		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to max	xx to xx	xx to xx	xx to xx
CAIS-P (total score)		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
RCADS-C				
Social Phobia		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Panic Disorder		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Major Depression		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Separation Anxiety		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Generalized Anxiety		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Obsessive-Compulsive		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Total Anxiety score		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
RCADS-P				
Social Phobia		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Panic Disorder		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Major Depression		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Separation Anxiety		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Generalized Anxiety		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Obsessive-Compulsive		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
Total Anxiety score		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
EQ-5D-Y		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
EQ-5D-Y VAS		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx

CHU-9D		(n=xx)	(n=xx)	(N=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx

Table 6: Summary of therapists involved in delivering trial treatment

Variable		(n=xx)
Treatment preference	OST	xx(xx)
	CBT	xx(xx)
	No preference	xx (xx)
Number of years delivering psychological interventions/therapies with children/young people	Mean (SD)	xx(xx)
	Median (IQR)	xx(xx to xx)
	Min to Max	xx to xx
Qualification (highest level)	GBC	xx(xx%)
	BACP	xx(xx%)
	...	xx(xx%)
	...	xx(xx%)
	...	xx(xx%)
	Other	xx(xx%)

Table 7: Continuous baseline characteristics by treatment group and missing primary outcome data status

NB: variables included in the table are given as an example and are not a comprehensive list of the variables that will be included in the analysis

Variable	Summary Statistic	Completers			Non-completers		
		CBT (n=XX)	OST (n=XX)	All (n=XX)	CBT (n=xx)	OST (n=xx)	All (n=xx)
Age (yrs)	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
BAT	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
CAIS-C	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
…	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
EQ-5D-Y	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
EQ-5D-Y VAS	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)

Table 8: Categorical baseline characteristics by treatment group and missing primary outcome status

Variable	Scoring	Completers			Non-completers		
		CBT (n=xx)	OST (n=xx)	All (n=xx)	CBT (n=xx)	OST (n=xx)	All (n=xx)
Sex	Male	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	Female	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
Ethnicity	White	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	Mixed/multiple ethnic groups	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	Asian/Asian British	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	Black/African/Caribbean/Black British [e]	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	Other ethnic group	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	Prefer not to say	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
Attends school	Yes	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	No	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
Treatment preference	CBT	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	OST	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)

NB: variables included in the table are given as an example and are not a comprehensive list of the variables that will be included in the analysis

Table 9: Treatment summaries and compliance

Variable	Scoring	CBT (n=xx)	OST (n=xx)	All (N=xx)
CBT Attendance, n(%)	No Sessions	xx(xx)		
	1 session	xx(xx)		
	...	xx(xx)		
	8 or more sessions	xx(xx)		
Attend at least 4 sessions, n (%)		xx(xx.x%)		
OST Attendance, n (%)	No sessions		xx(xx%)	
	FA only		xx(xx%)	
	Both FA and RET sessions		xx(xx%)	
Total number of hours attended	Mean (SD)	xx(xx)	xx(xx)	
Withdrew from treatment	Total	xx(xx%)	xx(xx%)	xx(xx%)
	Participant request	xx(xx%)	xx(xx%)	xx(xx%)
	Too busy	xx(xx%)	xx(xx%)	xx(xx%)
	Not happy with allocated treatment	xx(xx%)	xx(xx%)	xx(xx%)
	Clinician decision	xx(xx%)	xx(xx%)	xx(xx%)
	...	xx(xx%)	xx(xx%)	xx(xx%)
	Lost to follow up	xx(xx%)	xx(xx%)	xx(xx%)
Completed further treatment for phobia	Yes	xx(xx%)	xx(xx%)	xx(xx%)

Table 10: Primary and sensitivity effectiveness analysis: BAT score at 6 months

Primary outcome BAT score at 6-months	**CBT** n	**CBT** Mean(SD)	**OST** n	**OST** Mean(SD)	**Adjusted mean difference (95% CI)** [a]	**Adjusted mean difference** [b] **(95% CI)**
Per-Protocol	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	xx(xx to xx)
Intention to treat	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	xx(xx to xx)
CACE	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	xx(xx to xx)
excluding mistimed measurements	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	xx(xx to xx)
Multiple imputation	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	xx(xx to xx)

CBT is the reference group; [a] adjusted for baseline BAT score, age and phobia severity (baseline ADIS score) (; [b] adjusted for baseline BAT score, age and phobia severity baseline ADIS score and additional baseline covariates *decided based on baseline imbalances*.

BAT score	**CBT** n	**CBT** Mean change (95% CI)	**OST** n	**OST** Mean(SD)	**All** n	**All** Mean(SD)
change from baseline (at 6 months)	xx	xx (xx-xx)	xx	xx(xx-xx)	xx	xx(xx-xx)

Table 11: BAT score change over time

Figure 3: Forest plot of primary and sensitivity effectiveness analysis: Difference in BAT between OST and CBT

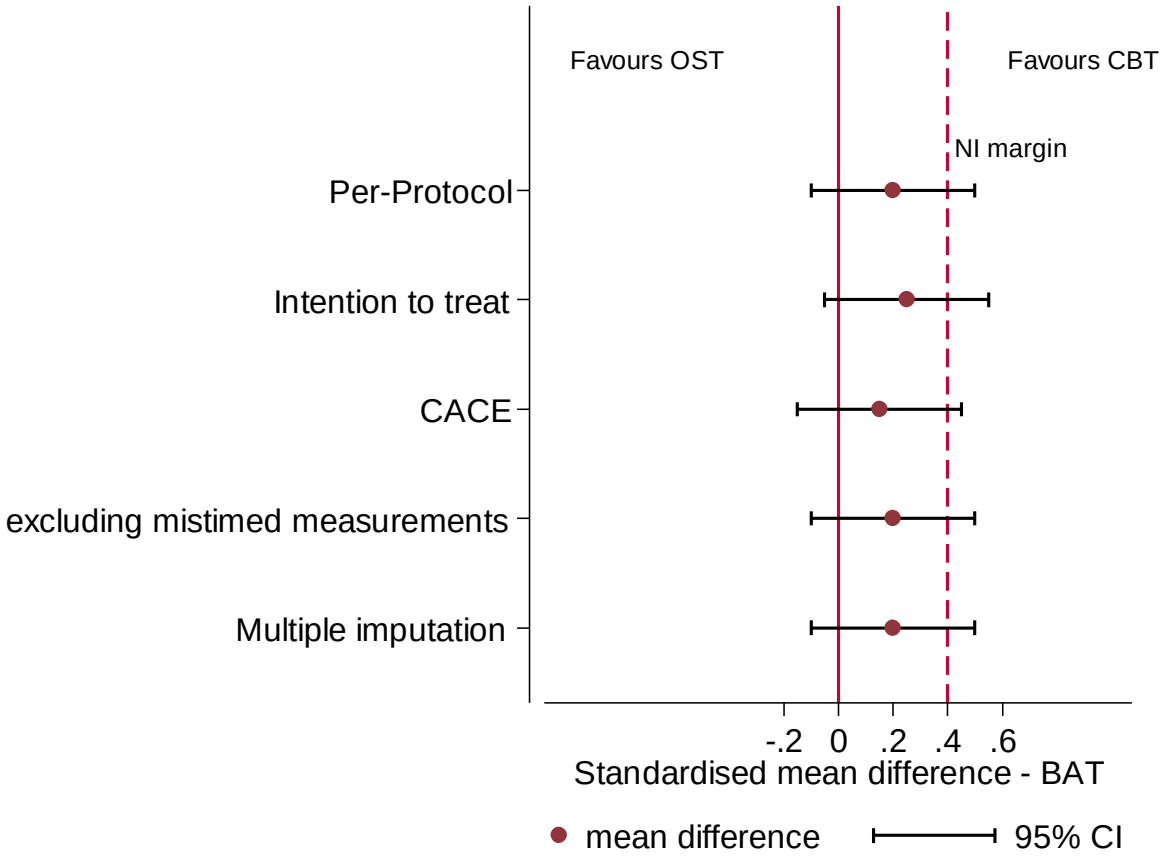


Figure 4: Mean (95% CI) BAT score over time

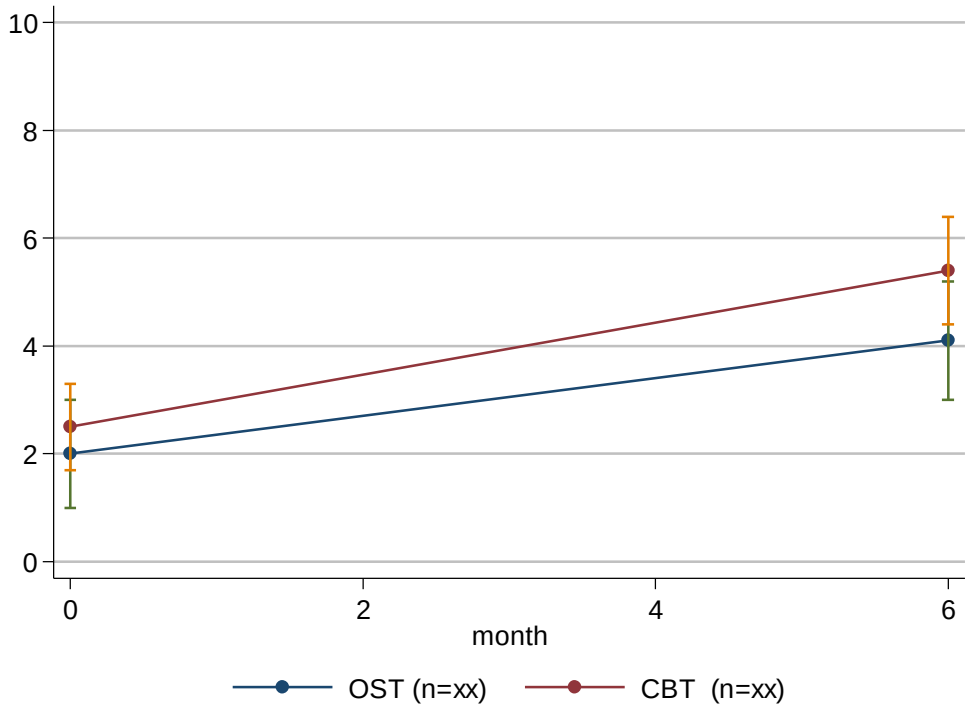


Table 12: Exploratory effect of OST intervention by subgroup: BAT at 6 months

Subgroup	Classification	CBT		OST		Adjusted mean difference (95% CI) [a]	P-value [b]
		n	Mean(SD)	n	Mean(SD)		
Participant treatment preference	CBT	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
	OST	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	
Therapist treatment preference	CBT	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
	OST	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	

Table 13: Effectiveness analysis: secondary outcomes at 6 months

Secondary outcomes at 6 months		CBT		OST		Adjusted mean difference (95% CI) [a]
		n	Mean(SD)	n	Mean(SD)	
BAT SUDS pre-task		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
BAT SUDS post-task		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
BAT SUDS change (pre to post task)		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
ADIS CSR		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
Goal Based outcome		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
CAIS-C		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
CAIS-P		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
RCADS-C	Social Phobia	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Panic Disorder	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Major Depression	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Separation Anxiety	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Generalized Anxiety	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Obsessive-Compulsive	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Total Anxiety score	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
RCADS-P	Social Phobia	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Panic Disorder	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Major Depression	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Separation Anxiety	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Generalized Anxiety	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Obsessive-Compulsive	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
	Total Anxiety score	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
EQ-5D-Y value index		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
EQ-5D-Y VAS		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)
CHU-9D		Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)

Table 14: Safety outcomes by treatment group

Safety outcomes	CBT (n=XXX)	OST (n= XXX)	All (n= XXX)

Number (%) of participants who experienced ≥1 AE	XXX (xx%)	XXX (xx %)	XXX (xx %)
Number (%) of participants who experienced ≥1 SAE	XXX (xx%)	XXX (xx %)	XXX (xx %)
Number of all AEs (including repeated events)	XXX	XXX	XXX
Number of all SAEs (including repeated events)	XXX	XXX	XXX

SAE Details			
Seriousness			
Death	xx (xx%)	xx (xx%)	xx (xx%)
Life threatening	xx (xx%)	xx (xx%)	xx (xx%)
Inpatient hospitalisation	xx (xx%)	xx (xx%)	xx (xx%)
Prolongs hospitalisation	xx (xx%)	xx (xx%)	xx (xx%)
Persistent or significant disability/incapacity	xx (xx%)	xx (xx%)	xx (xx%)
Total	xx (xx%)	xx (xx%)	xx (xx%)
Intensity			
Mild	xx (xx%)	xx (xx%)	xx (xx%)
Moderate	xx (xx%)	xx (xx%)	xx (xx%)
Severe	xx (xx%)	xx (xx%)	xx (xx%)
Total	xx (xx%)	xx (xx%)	xx (xx%)
Relationship to study intervention			
Unlikely	xx (xx%)	xx (xx%)	xx (xx%)
Unrelated	xx (xx%)	xx (xx%)	xx (xx%)
Total	xx (xx%)	xx (xx%)	xx (xx%)

13 References

Altman, D. G. (1991). Practical statistics for medical research. CRC press.

Assmann, S.F., Pocock, S.J., Enos, L.E. and Kasten, L.E. (2000), "Subgroup analysis and other (mis)uses of baseline data in clinical trials.", Lancet (London, England), Vol. 355 No. 9209, pp. 1064–9.

CHAN, C., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A. (2016). CONSORT 2010 Statement: extension to randomised pilot and feasibility trials.

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of dsm-iv anxiety and depression in children: A revised child anxiety and depression scale. Behaviour Research and Therapy, 38(8), 835-855.

Chorpita, B. F. Ebesutani, C,, Spence, S.H., (2015) RCADS User guide. http://www.childfirst.ucla.edu/RCADSUsersGuide20150701.pdf Accessed Dec 2016

Cuzick, J. (2005), "Forest plots and the interpretation of subgroups.", Lancet (London, England), Vol. 365 No. 9467, p. 1308.

EuroQol, G. (1990). Euroqol--a new facility for the measurement of health-related quality of life. Health policy (Amsterdam, Netherlands), 16(3), 199-208.

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for human use; ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials E9. (2010).

Jones, B., Jarvis, P., Lewis, J. A., & Ebbutt, A. F. (1996). Trials to assess equivalence: The importance of rigorous methods. British Medical Journal, 313(7048), 36-39.

Langley, A. K., Bergman, R. L., McCracken, J., & Piacentini, J. C. (2004). Impairment in childhood anxiety disorders: Preliminary examination of the child anxiety impact scale-parent version. Journal of Child and Adolescent Psychopharmacology, 14(1), 105-114. doi:10.1089/104454604773840544
Law, D., & Jacob, J. (2013). Goals and goal based outcomes (gbos) (3rd ed.). London: CAMHS Press.
Öst, L. G., Salkovskis, P. M., & Hellstrom, K. (1991). One-session therapist-directed exposure vs self-exposure in the treatment of spider phobia. Behavior Therapy, 22(3), 407-422. doi:10.1016/s0005-7894(05)80374-0
Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J., Altman, D. G., & CONSORT Group. (2012). Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. Jama, 308(24), 2594-2604.
Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine*, *21*(19), 2917–30. http://doi.org/10.1002/sim.1296
Reynolds, S., Wilson, C., Austin, J., & Hooper, L. (2012). Effects of psychotherapy for anxiety in children and adolescents: A meta-analytic review. Clinical Psychology Review, 32(4), 251-262. doi:10.1016/j.cpr.2012.01.005
Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*(mar23 1), c332–c332. http://doi.org/10.1136/bmj.c332
Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in Medicine*, *13*(17), 1715–1726. http://doi.org/10.1002/sim.4780131703
Silverman, W. K., & Albano, A. M. (1996). Anxiety disorders interview schedule (adis-iv) child and parent interview schedules. Albany, NY: Greywind Publications.
StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.
Stevens, K., & Ratcliffe, J. (2012). Measuring and valuing health benefits for economic evaluation in adolescence: An assessment of the practicality and validity of the child health utility 9d in the australian adolescent population. Value in Health, 15(8), 1092-1099. doi:10.1016/j.jval.2012.07.011
Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios L et al (2010). A tutorial on pilot studies: the what, why and how. BMC Medical Research Methodology; 10(1):1.
Wang, R., Lagakos, S.W., Ware, J.H., Hunter, D.J. and Drazen, J.M. (2007), "Statistics in medicine--reporting of subgroup analyses in clinical trials.", The New England Journal of Medicine, Massachusetts Medical Society, Vol. 357 No. 21, pp. 2189–94.
Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. Clinical Psychology Review, 28(6), 1021-1037. doi:10.1016/j.cpr.2008.02.007