Collaboration for Leadership in Applied Health Research and Care East Midlands

NHS National Institute for Health Research

Additional Supporting Information for: The impact of a computerised test of attention and activity (QbTest) on diagnostic decision-making in children and young people with suspected ADHD: single-blind randomised controlled trial – by Chris Hollis et al.

Appendix S3. Statistical Analysis Plan.

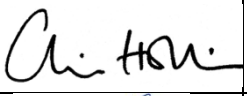
A randomised controlled trial comparing the effects of providing clinicians and patients with the results of an objective measure of activity and attention (QbTest) versus usual care on diagnostic and treatment decision making in children and young people with ADHD

AQUA (Assessing QbTest Utility in ADHD)

Statistical Analysis Plan

Version 0.1.2 (01 Aug 2017)

Based on Protocol version 1.4 (dated 15th December 2015)

The following people have reviewed the Statistical Analysis Plan and are in agreement with the contents			
Name	**Role**	**Signature**	**Date**
Prof Chris Hollis	Chief Investigator		16-JAN-2018
Dr Boliang Guo	Trial Statistician		16-JAN-2018

Dr Charlotte Hall	Lead Research Fellow		16-JAN-2018

Contents

Abbreviations

Abbreviation	Description
ADHD	Attention Deficit/Hyperactivity Disorder
AE	Adverse Event
CAMHS	Child and Adolescent Mental Health Services
C-GAS	Children's Global Assessment Scale
CI	Chief Investigator overall
CPT	Continuous Performance Test
CRF	Case Report Form
CSRI	Client Service Receipt Inventory
DAP	Data Analysis Plan
DAWBA	Development and Wellbeing Assessment
DMC	Data Monitoring Committee
EQ-5D-Y	EuroQol Five Dimensions Heath Questionnaire
GCP	Good Clinical Practice
ICF	Informed Consent Form
ITT	Intention to treat
NHS	National Health Service
PI	Principal Investigator at a local centre
PIS	Participant Information Sheet
REC	Research Ethics Committee
R&D	Research and Development department
SAE	Serious Adverse Event
SC	Scientific Committee (TSC/DMC)
SDQ	Strengths and Difficulties Questionnaire
SNAP-IV	Swanson, Nolan & Pelham questionnaire

TMG	Trial Management Group
TSC	Trial Steering Committee
QbB	QbBlind
QbO	QbOpen

1. INTRODUCTION & PURPOSE

This document details the rules proposed and the presentation that will be followed, as closely as possible, when analysing and reporting the main results from "A randomised controlled trial comparing the effects of providing clinicians and patients with the results of an objective measure of activity and attention (QbTest) versus usual care on diagnostic and treatment decision making in children and young people with ADHD" which was funded National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) East Midlands (EM).

The purpose of the plan is to:

1. Ensure that the analysis is appropriate for the aims of the trial, reflects good statistical practice, and that interpretation of a priori and post hoc analyses respectively is appropriate.

2. Explain in detail how the data will be handled and analysed to enable others to perform the actual analysis in the event of sickness or other absence

Additional exploratory or auxiliary analyses of data not specified in the protocol are permitted but fall outside the scope of this analysis plan (although such analyses would be expected to follow Good Statistical Practice).

The analysis strategy will be made available if required by journal editors or referees when the main papers are submitted for publication. Additional analyses suggested by reviewers or editors will, if considered appropriate, be performed in accordance with the Analysis Plan, but if reported the source of such a post-hoc analysis will be declared.

Amendments to the statistical analysis plan will be described and justified in the final report of the trial

2. SYNOPSIS OF STUDY DESIGN AND PROCEDURES

2.1. Trial aims and objectives

The purpose of the trial is to investigate whether using the QbTest in routine NHS settings can facilitate care for children with ADHD. The primary purpose of the study is to assess whether the QbTest can accelerate diagnostic assessment of ADHD without compromising accuracy.

Additionally, the study will provide knowledge on whether the QbTest can improve ADHD treatment by reducing the time to medication normalisation.
Together these findings will assess whether the QbTest can improve practice and patient outcome

for children with ADHD. Additionally, information will be gained on whether the QbTest is clinically (patient and clinician opinion on QbTest) and financially viable in routine settings. These aims will be investigated through a series of primary and secondary objectives.

2.1.1. Primary objective

The primary objective of this study is to investigate whether the QbTest can accelerate time to a correct diagnostic decision. It is hypothesised that providing clinicians with the QbTest reports (QbO arm) will reduce the time to confirm/reject a diagnosis of ADHD in comparison to clinicians who do not receive QbTest reports (QbB arm). The primary outcome measure will be number of clinic sessions until a confirmed correct diagnosis. Additionally, we shall record time to diagnosis (in days) and the duration of visits (in minutes) and include these as secondary outcomes to this primary objective.

2.1.2. Secondary objectives

1) To assess whether QbTest increases the clinician's degree of confidence in their diagnostic decision. It is hypothesised that clinicians in the QbO arm will have greater confidence in their diagnosis in comparison to clinicians in the QbB arm. This will be measured through a 6-point Likert scale (1 = definitely ADHD, 6 = definitely not ADHD). This will aid our understanding of how QbTest facilitates the diagnostic decision process.
2) Proportion who receive correct definitive diagnosis (including definitive ADHD exclusion). This will provide evidence of the amount of correct diagnoses received in each arm.
3) To assess the stability in diagnosis between the QbO and QbB arms. It is hypothesised that patients/clinicians in the QbO arm will have greater stability in diagnosis than patients/clinicians in the QbB arm. This will be measured by asking clinicians to re-rate their diagnostic decision and confidence at 6 months. This will aid our understanding of whether QbTest helps facilitate a diagnosis of ADHD.
4) To assess the diagnostic accuracy of the QbTest by comparing diagnosis with diagnosis made through the 'gold standard' assessment; DAWBA. It is hypothesised that more diagnoses in the QbO arm will be in agreement with the diagnostic decision from DAWBA results than those made in the QbB arm. This will provide evidence for the diagnostic accuracy of the QbTest.
5) QbTest scores (Q scores for attention, impulsivity and activity) will be compared with DAWBA ADHD categorisation to obtain the best predictive model based on QbTest scores that discriminates between ADHD 'positive' and ADHD 'negative' gold standard DAWBA diagnoses.
6) To check that QbTest does not compromise quality of decision making (assessed via side-effects). If QbTest facilitates medication titration, it is hypothesised that patients in the QbO arm should report fewer medication side-effects than those in the QbB arm. This will be assessed by measuring side-effects through a well validated scale at 4-8 weeks after medication initiation. This will provide evidence as to whether the QbTest can facilitate medication judgements and improve patient outcome.

7) To investigate whether the QbTest improves treatment response. **A)** It is hypothesised that if the QbTest facilitates decisions to medicate, more patients in the QbO arm will be normalised at 6 months than in the QbBlind arm (assessed via scores of SNAP-IV). **B)** It is hypothesised that if the QbTest facilitates treatment, patients in the QbO arm will report less symptom impact at 6 months than those in the QbB arm. This will be measured via SDQ scores collected at baseline and 6 months and C-GAS scores collected at baseline, 3 months and 6 months. This will provide initial evidence on the utility of QbTest to inform medication decisions and improve patient outcome. We shall also compare changes in SNAP-IV scores at baseline and 4-8 week follow-up with changes in QbTest scores, to see how objective (QbTest) measures of change relate to subjective (SNAP-IV) measures of change.
8) To investigate whether QbTest improves patient quality of life. It is hypothesised that if the QbTest improves treatment decision making patients in the QbO arm will report better quality of life than those in the QbB arm. This will be measured by quality of life scores gained from the EQ-5D-Y measured at baseline, 4-8 weeks after medication titration and 6 month follow-up. This will further our understanding on whether QbTest improves patient outcome.
9) To investigate the cost effectiveness of the QbTest. Parents will be asked to complete the CSRI from baseline, 3 months and 6 months. This information will aid our understanding on the economic impact of the QbTest.
10) To investigate opinions of QbTest. Clinician and patient opinion of the QbTest will be assessed via questionnaires or interviews. This will provide information of the acceptability of QbTest in routine NHS settings.

2.2. Trial design and configuration

All participants will receive the same intervention. Specifically, this will be assessment as usual, plus a QbTest. The research team will also collect the additional research measures (DAWBA, SDQ, C-GAS, side-effects scale, SNAP-IV, CSRI, EQ-5D-Y, clinic pro-forma, QbTest questionnaire and/or interviews). Clinicians complete the C-GAS at baseline, 3- and 6-months, but the research team will be responsible for ensuring this is completed. The C-GAS and clinic pro-forma are the only clinician completed measures.

After obtaining consent, participants (n=270) will be randomised on a 1:1 ratio into either the QbOpen (QbO) arm or QbBlind (QbB) arm of the trial. For participants randomised to QbO, clinicians will have immediate access to the QbTest report to inform their diagnostic decision making. For participants randomised to the QbB arm, clinicians and patients will not have access to their QbTest reports until 6 months after the baseline assessment. With the exception of withholding QbTest reports, the procedures are identical for the two arms.

Participants will be sent a welcome pack once consented into the study explaining the process.

There are 2 phases to this study:

Phase 1 - Assessment: The first phase will investigate QbTest as a tool to aid the diagnostic process. We shall recruit 234 patients into the study (117 in each of the 2 arms).

Participants will undergo their standard ADHD assessment as usual practice. This varies from clinic to clinic, and between cases and clinicians, however, it would typically involve an interview with the parent and young person and collection of questionnaires and outcome measures from the parent/young person and teacher. Whilst attending the clinic, the participant will be asked to complete QbTest at some point during their initial clinic visit. The RA or technician will usually test the participant whilst they are waiting to see their clinician. This will avoid increasing participants' time at the clinic. However, in some cases (i.e. if the RA/technician is already testing another participant) QbTest may be completed after the clinician's appointment. There are no anticipated effects regarding completing QbTest pre/post clinician appointment. The participants will complete QbTest in a designated room within the clinic.

All clinicians will be trained by QbTech in how to use and interpret the QbTest.

The QbTest is a computerised assessment which takes approximately 20 minutes to complete. The test combines a continuous performance test (CPT) alongside an infrared camera which measures the participant's movements. The infrared camera measures motor activity of the participant whilst they undertake the task. During the test the participant has to wear a soft headband with a "marker" on the front, this allows the infra-red camera to track the participant's head movement. This headband is non-invasive and comfortable to wear. No images of the participant are recorded, only head movements from the infra-red marker. During the test, the participant is presented with continuously changing stimuli. Embedded within these stimuli is a given target. Participants have to respond by pressing a hand-held button only when the target appears. Attention is measured through omission errors (non-responses to a target) and reaction time to response. Impulsivity is assessed through commission errors (pressing to a non-target) and anticipatory errors (pressing before target). Motor activity is measured through head movements during the task. The researcher or clinician will be present throughout the test.
Parents will also be asked to complete the SNAP-IV regardless of their child's age (Swanson et al., 1983) at baseline assessment. The SNAP-IV is a short 26-item questionnaire designed to assess ADHD symptoms, with established validity, reliability and use in clinical and research settings. In cases where the researcher is present at the clinic, this will be done at the initial appointment. Where the researcher is not present, the SNAP-IV will be posted to the participant to complete before the child receives medication or offered the option to complete it online. Alternatively, parents will be offered the opportunity to complete the SNAP-IV over the phone with the researcher. The SNAP-IV should take no more than 10 minutes to complete. Teachers will also be contacted by the researcher to complete the SNAP-IV. Initial contact with teachers will be made as soon as written consent is received. Again, teachers will be posted the questionnaire or offered the opportunity to complete over the phone or online.

The parents will be asked to complete the CSRI (Beecham & Knapp, 1992) at baseline assessment. The CSRI asks questions pertinent to economic factors that may alter as a result of the intervention. The questionnaire can be completed at clinic, via the phone, posted to the participant's home address or completed online. Young people together with their parents will be asked to complete an EQ-5D-Y (Wille et al., 2010) The EQ-5D-Y is a short 15-item questionnaire which assesses health

related quality of life. The questionnaire takes approximately 10 minutes to complete and can be completed at the clinic (where researcher is present), via the phone or posted to the participants home address or completed online. For all postal questionnaires, participants will be provided with a pre-paid return envelope.

After the first appointment and before the young person starts medication the parent and teacher will be asked to complete the DAWBA (Goodman et al., 2000). The DAWBA is a well validated semi-structured diagnostic assessment interview for common child mental health problems, including ADHD, which includes the SDQ (Strengths and Difficulties Questionnaire; Goodman, 1997) as part of its initial screen. The DAWBA is similar to a semi-structured interview and gathers information from the parent and teacher via a series of questions with prompts and skip rules. Research diagnoses are allocated following a consensus diagnostic conference (CDC) involving two experienced consultant child and adolescent psychiatrists who review the DAWBA assessment scores and verbatim supporting text. It can be completed online or via the telephone. DAWBA completion time varies according to individual case characteristics and responses to questions. Completion time can vary from ½ hour – 3 hours. Participants will be given the option of either online or phone completion. To encourage online completion, and compensate participants for any additional time the online version may take, participants will be offered £15 vouchers for their time for online completion. Although the DAWBA can (in some cases) take a long while to complete (up to 3 hours), this is not for every case and the DAWBA interview is essential as it provides a "gold standard" assessment for ADHD, thus on the recommendation of the scientific committee its inclusion in this study is pivotal for assessing the accuracy of the QbTest.

Phase 2 - Treatment: Patients who receive a diagnosis of ADHD and are allocated by clinicians to receive ADHD medication initiated within 3 months of their baseline assessment will be asked to complete a 2nd QbTest (Qb2) 4-8 weeks after medication initiation. This timeframe was chosen to ensure that all participants can complete their second QbTest before the 6-month follow-up. The QbTest should coincide with their clinic appointment; however, where necessary the test can be completed at the patient's house or school.

At 4-8 weeks post medication initiation, parents and teachers will also be asked to complete the SNAP-IV (either at clinic where researcher is present, posted, telephone or online) and the parent/young person will be asked to complete a side-effects scale (Hill & Taylor, 2001), containing 15-itemes and taking approximately 5 minutes to complete. Patients who do not receive a diagnosis of ADHD or start medication within 3 months of baseline assessment will be asked to complete the SNAP-IV at 3 months. Again, this could be done at clinic (by the researcher) or via telephone, posted to their home address or completed online. ALL Parents will be asked to complete the CSRI again at 3 months. ALL young people alongside their parents will also be asked to complete a EQ-5D-Y. Clinicians will complete the C-GAS for all patients at 3 months.

At 6 months after baseline assessment, all participants will be asked to complete the following measures: Parents will be asked to complete a SNAP-IV and CSRI and SDQ (Goodman, 1997) to assess changes in symptoms and symptom impact. Teachers will be asked to complete a SNAP-IV and SDQ.

Young people alongside their parents will be asked to complete a EQ-5D-Y. We will ask families to complete the EQ-5D-Y collaboratively (between the child and parent/carer).

Participants will also be asked to complete a short questionnaire and/or take part in a short interview to gain their opinion on the acceptability/feasibility/utility of QbTest in routine NHS clinical practice. This will be conducted at a time and place to suit the participant.

Throughout the study, clinicians will be required to keep a pro-forma record for each clinic visit the young person and/or parent attends. The clinic pro-forma will document diagnosis and medication/ treatment decisions. Clinicians will not have access to the research measures (SNAP-IV, CSRI, side effects scale, DAWBA, SDQ, EQ-5D-Y). Clinicians will be asked to complete the C-GAS at baseline, at 3 months and 6 months.

All participants entering phase 1 regardless of diagnosis or medication will receive the following measures:

- CSRI (completed by parent at baseline, 3 months and 6 months);
- SDQ (completed by parent and teacher at 6 months) and DAWBA (baseline)
- C-GAS (completed by the clinician at baseline, 3 months and 6 month follow-up);
- EQ-5D-Y (completed by parent/young person at baseline, 3 months and 6 months).
- SNAP-IV (completed by parent and teacher at baseline, 4-8 weeks after medication or 3 months from baseline if not medicated and 6 months)

Participants who receive ADHD diagnosis and medication within 3 months

- QbTest 2 (receive 2nd QbTest at 4-8 weeks after medication initiation)
- Side-effects scale (at 4-8 week appointment)

Patients who receive a diagnosis of ADHD but who are not offered (or reject) medication will receive parent and teacher rated SNAP-IV at baseline, 3- and 6 months.

During an assessment, if the QbTest technician perceives that a participant is not performing QbTest properly, the test will be stopped and the QbTest technician will discuss with the patient whether or not they wish to continue. If they wish to continue, the QbTest technician will explain to the participant that it is important they perform the test to the best of their ability. Otherwise, or if they still don't perform properly, the test will be stopped. If this occurs at their first assessment, they will still be invited to complete the test at a follow-up assessment and the same procedure will be applied.

Attendance to clinic sessions will be recorded and taken as a measure of compliance to the treatment regime. The questionnaire/interview will assess the drivers and barriers to completing the QbTest in routine NHS settings.

2.3. Trial centres

The study is a multi-centre RCT recruiting participants from the caseloads of CAMHS and community paediatric clinicians. The protocol lists the following sites:

Medway NHS Foundation Tryst
Alder Hey Children's NHS Foundation Trust
Nottinghamshire Healthcare NHS Trust
Nottingham University Hospitals Trust
Leicestershire Partnership NHS Trust
Sussex Partnership NHS Foundation Trust
United Lincolnshire Hospital Trust (ULH)
Central Manchester University Hospitals NHS Foundation Trust (CMFT)
Bridgewater Community Healthcare NHS Trust
Cumbria Partnership NHS Foundation Trust

However, Bridgewater and Cumbria did not recruit any participants into the trial. Instead, an additional site was added in the form of:

North East London NHS Foundation Trust (NELFT)

Clinicians working within these Trusts will be recruited by the study team to participate in the research. The clinical team will invite all new cases referred for a diagnostic evaluation of ADHD to participate in the research.

2.4. Eligibility criteria

2.4.1. Inclusion criteria

- Age 6 – 17 years (at the time of consent)
- Referred to CAMHS or Community Paediatrics for an ADHD diagnostic assessment
- Capable of providing written informed consent (over 16)
- Parental consent (under 16)

2.4.2. Exclusion criteria

- Severe learning disability (to be assessed by clinical judgement)
- Non-fluent English speaking
- Previous or current confirmed diagnosis of ADHD
- Currently receiving ADHD medication

2.5. Description of interventions

All participants will receive the following interventions:

1) Clinical assessment as usual. The participant will undergo standard diagnostic assessment as usual for ADHD.
2) SNAP-IV (Swanson, Nolan & Pelham questionnaire; Swanson et al., 1983). The SNAP-IV is a short 26-item questionnaire designed to assess ADHD symptoms.

3) DAWBA (Development and Well-being Assessment; Goodman et al., 2000). The DAWBA is a semi-structured, investigator-based diagnostic interview for child mental health problems, including ADHD, which includes the SDQ (Strengths and Difficulties Questionnaire; Goodman, 1997) as an initial screen.
4) QbTest. The QbTest is a computerised assessment of attention, impulsivity and activity.
5) CSRI (Client Service Receipt Inventory, Beecham & Knapp, 1992).
6) Side-effects scale (Hill & Taylor, 2001).
7) SDQ (Strengths and Difficulties Questionnaire; Goodman, 1997).
8) EQ-5D-Y (EuroQol Five Dimensions Heath Questionnaire; Wille et al., 2010). The EQ-5D-Y is a short 15-item questionnaire which assesses health related quality of life.
9) Feedback questionnaire or interview: service-users will be asked to fill in a short questionnaire assessing their opinion of QbTest. A sub-sample of 20 participants will also be asked to participate in an interview.

All participants will receive QbTest as part of their baseline clinical assessment; however, the timing of feedback of the QbTest results to the clinical team will be randomised. For half the sample (Qb Open study arm), the clinician will received the QbTest result immediately following the test during the baseline assessment. The clinician will incorporate the QbTest result into their overall diagnostic assessment and decision making. For the other half (Qb Blind study arm) the clinical team will not receive QbTest until 6 months after the initial baseline assessment. Clinicians will also be required to keep a short pro-forma documenting for each participating patient their clinic appointment schedule, diagnostic status at each appointment (including diagnostic confidence), medication and other treatments offered. Clinicians will also be asked to complete the Children's Global Assessment Scale (C-GAS; Shaffer et al., 1983)

If medication is initiated for the treatment of ADHD within 3 months of the baseline assessment for participants in either arm, those participants will be assessed by QbTest and the SNAP-IV questionnaire (parent and teacher) within 4-8 weeks of medication initiation. Only a proportion of participants at baseline will have a confirmed diagnosis of ADHD and only a sub-group of those (approximately 70%) will receive medication within 3 months. Participants who do not receive a diagnosis of ADHD and medication within 3 months will not receive the 2nd QbTest, but will be asked to complete the SNAP-IV questionnaire and other measures at 3 and 6 months.

2.6. Randomisation procedures

The study consists of two arms, QbOpen (QBO) whereby the patient and clinician have immediate access to the QbTest report which can be used to inform clinical decision making and QbBlind (QbB) whereby the patient and clinician do not have access to the QbTest report until 6 months after study initiation.

Once consent has been obtained from participants they will be allocated a unique trial identity code number (including a code for the person, site, age and gender) this information will be entered onto a web-based randomisation system (set up by University of Nottingham Clinical Trials Unit; CTU). **Randomisation** should occur at or before the first appointment. QbTest will be informed (via email from the CTU) of each participant randomised into the study. They will then manually control which

QbTest reports to release to the clinician (if in the QbOpen arm) or withhold (if in the QbBlind arm). The arm to which a participant is assigned will be determined by a computer generated pseudo-random code using random permuted blocks of varying size, created by the Nottingham CTU in accordance with their standard operating procedure and held on a secure server. Participants will be allocated with equal probability to each treatment arm with **stratification by site**. All participants will undergo the same research measures, including the QbTest. It is the time at which the report is made available to the clinician and patient that is randomised.

The statistician, outcome assessors and the CI will be blind to which arm participants are in.

2.7. Sample size and justification

The sample size calculation was based on an audit study data from the department of community pediatrics service of the Medway NHS Trust (Hall et al., 2016). Calculations based on this audit study data showed that the mean number of visits needed to achieve an ADHD diagnosis before introduction of the QbTest (control rate) for children aged 6-14-year olds was 2.94 visits and following the introduction of QbTest a diagnosis was reached in a mean of 2.18 visits. Following consultation with stakeholders, it was agreed that this difference (2.94-2.18) represented the minimum clinically important difference (MCID), with any smaller difference in mean clinic visits being of debatable value. Therefore, 71 patients in each study group will be required to detect a mean count difference of the above magnitude with 80% power at two tailed 0.05 significance level [1, 2], assuming the count of visit following Poisson distribution. Given that evidence shows the Intra-class Correlation Coefficients (ICC) of mental health measures across GP centers were extremely low[3-5], plus results from audit data also indicated that count of visit needed to achieve an ADHD diagnosis were homogeneous across centers, we will therefore assume the center effects will not influence sample size calculation in our study. After taking into account a 20% withdrawal rate, the final total sample size will be 178. The same calculation performed with 90% power would require a total sample of 234 participants. We aim to recruit 178 participants as a minimum and 234 participants as a maximum. Software Stata 13 was used for power analysis.

However, in September 2015, half way through the trial we (TMG group) observed that approximately 35% of our cases do not have a diagnosis by six months. An additional 6% of participants have missing data/withdrawn from the study. This figure was calculated based on the 145 participants that had reached 6 months at that moment in time. We had anticipated a significantly larger proportion would have received a confirmed diagnosis by 6 months. As a result we were concerned that our analysis may be under powered. To allow for this we sought sponsor and ethical approval to recruit additional 15% more participants, resulting in a maximum sample of 270 participants.

The CLAHRC EM Scientific Committee (Independent Trial Steering Group) had approved our request on Dec-4-2015 and suggested a new sample size calculation would be presented as response to the proposed new analysis. In line with Scientific Committee statistician suggestion,

we rerun power analysis and rewrote the sample size justification section as follow:

The sample size calculation was based on an audit study data from the department of community pediatrics service of the Medway NHS Trust (Selby, 2013). Based on person-period dataset for discrete survival analysis using logistic regression modelling, results showed the percentage of ADHD diagnosis were 34.07% and 45.98% for group without Qb test result and with Qb test result respectively, with 19.45% outcome variability due to time variables. Being slightly conservative, one hundred ninety six participants will be requested to detect the difference between proportion 34% and 45% with 80% power at two tailed 0.05 significance level, assuming 20% total variability to be explained by time. With 90% power at two sided 0.05 significance level, the number of participant required will be 268. To check the robustness of sample size calculation sensitive to different assumption, a Cox regression was performed with same data yielding hazard ratio (HR) as 1.67. By performing power analysis with log rank test assuming 30% censor rate in QB blind arm and 10% withdrawal, 238 (250, 284) participant would be required to detect such an effects of HR=1.67(1.65, 1.60) with 90% power at two sided 0.05 significance level. Stata powerlog and power logrank command were used to perform power analysis.

2.8. Blinding and breaking of blind

Personal data, which will consist of a list of names, birth dates, postcodes, NHS number and unique ID codes of participants, will be kept by the Local Collaborator (LC) at each site for the duration of the study. At the end of the study, the lists which decode the participants' IDs will be stored securely in accordance with the University of Nottingham's Code of Research Conduct, under the custodianship of the Chief Investigator.

Investigators may identify the arm of the trial to which participants have been allocated through password protected access; this will be done at the end of the trial or in the event of a medical emergency or serious adverse event.
The clinician performing the QbTest at baseline will be blind to the randomisation until after the first QbTest. They will be aware of allocation arm when they conduct interviews about the acceptability/utility of Qb at the end of the study. Researchers collecting outcome measures will be blind to trial arm allocation. DAWBA consensus diagnoses will be made blind to trial arm allocation

Any case of un-blinding will be reported to the trial coordinator. Following data collection an analysis will be completed to determine if incidences of un-blinding were equal in the two treatment groups. Clinicians and patients will not be 'blind' to which arm (QbO/QbB) they have been allocated.

2.9. Trial committees

A number of committees will be assembled to ensure the proper management and conduct of the trial, and to uphold the safety and well-being of participants. The general purpose, responsibilities and structures of the committees are described in the protocol. However each

committee will develop its own rules and procedures which may evolve with time.

Trial Management Group: The Trial Management Group (TMG) will oversee the operational aspects of the trial. The TMG will meet regularly to review the progress of the trial and address any issues arising. The trial management group consists of the study team. All members are blind to arm allocation except the Clinical Trials Manager, Catherine Kaylor-Hughes.

Trial Steering Committee:

The CLAHRC EM scientific committee act as an independent trial steering committee. There is no assigned data monitoring committee (DMC). However, the scientific committee take on some roles of the DMC in the form of independent statistical review.

2.10. Outcome measures

2.10.1. Primary outcome

Number of clinic appointments until ADHD diagnosis is confirmed or excluded for the QBOpen (delivery of immediate QbTest feedback report) and QbBlind (delayed feedback of QbTest report) groups.

2.10.2. Secondary outcomes

- Time to confirmation or exclusion of ADHD diagnosis (in days) and duration of clinic visits (in minutes). This will provide supporting evidence that a reduced number of clinic visits is associated with shorter overall time to diagnosis and reduced clinic time required for assessment.
- Proportion who receive correct definitive diagnosis (including definitive ADHD exclusion). This will provide evidence of the amount of correct diagnoses received in each arm
- Differences in degree of confidence of diagnosis between cases in QbOpen (QbO) and QbBlind (QbB) group. These variables will assess the utility of QbTest in aiding diagnosis (diagnostic certainty).
- Stability in diagnosis and confidence between the QbO and QbB group measured as number of patients where the primary diagnosis (ADHD confirmed or excluded) was changed at their 6 month visit to the clinic.).
- Correct classification rate on QbTest vs gold standard classification rate (assessed via DAWBA completed by the parent and teacher) (Diagnostic accuracy).
- QbTest scores (Q scores for attention, impulsivity and activity) will be compared with DAWBA ADHD categorisation to obtain the best predictive model based on QbTest scores that discriminates between ADHD 'positive' and ADHD 'negative' gold standard DAWBA diagnoses.
- Side effects scales will be used as a control check to ensure greater speed to diagnosis/medication normalisation is not off-set by greater side-effects.
- Proportion of participants achieving 'normalisation', measured via SNAP-IV scores 6 months post baseline assessment. This will assess whether QbTest increases the rate of

normalisation. We will also compare changes in SNAP-IV scores at baseline and 4-8 week follow-up with changes on QbTest scores, to see how objective (QbTest) measures of change relate to subjective (SNAP-IV) measures of change.

- SDQ scores (parent & teacher completed) conducted at baseline (through DAWBA) and follow-up will be used to compare outcomes between the two study arms. C-GAS scores conducted at baseline, 3 months and 6 month follow-up will provide a clinician rated measure of outcome.
- EQ-5D-Y will compare quality of life between QbOpen and QbBlind groups at baseline, 3 months and 6 month follow up.
- Cost effectiveness of the QbTest (assessed via CSRI scores).
- Ratings of QbTest on the client feedback questionnaire and interview: this will provide a measure of acceptability to service users.

Ratings of QbTest on the clinician feedback questionnaire and interview: this will provide a measure of feasibility and acceptability to clinicians and patients of adding QbTest to routine clinical care.

2.11. Interim analysis

No interim analysis or analyses for safety or efficacy are planned. All analysis will be backed up to the UoN servers

3. GENERAL ANALYSIS CONSIDERATIONS

3.1. Analysis samples

The analysis will be conducted on an Intention to treat (ITT) basis. However as per ICH 9 guideline [6], participants who failed to satisfy major entry elegibility criteria will be exclude from the analysis set; Children who was randomised but not had Qb test will be excluded as well.

3.2. Derived variables

Children's age will be calculated as time between date of randomization and birthday:

Age=: date of randomization - DOB

3.3. Procedures for missing data

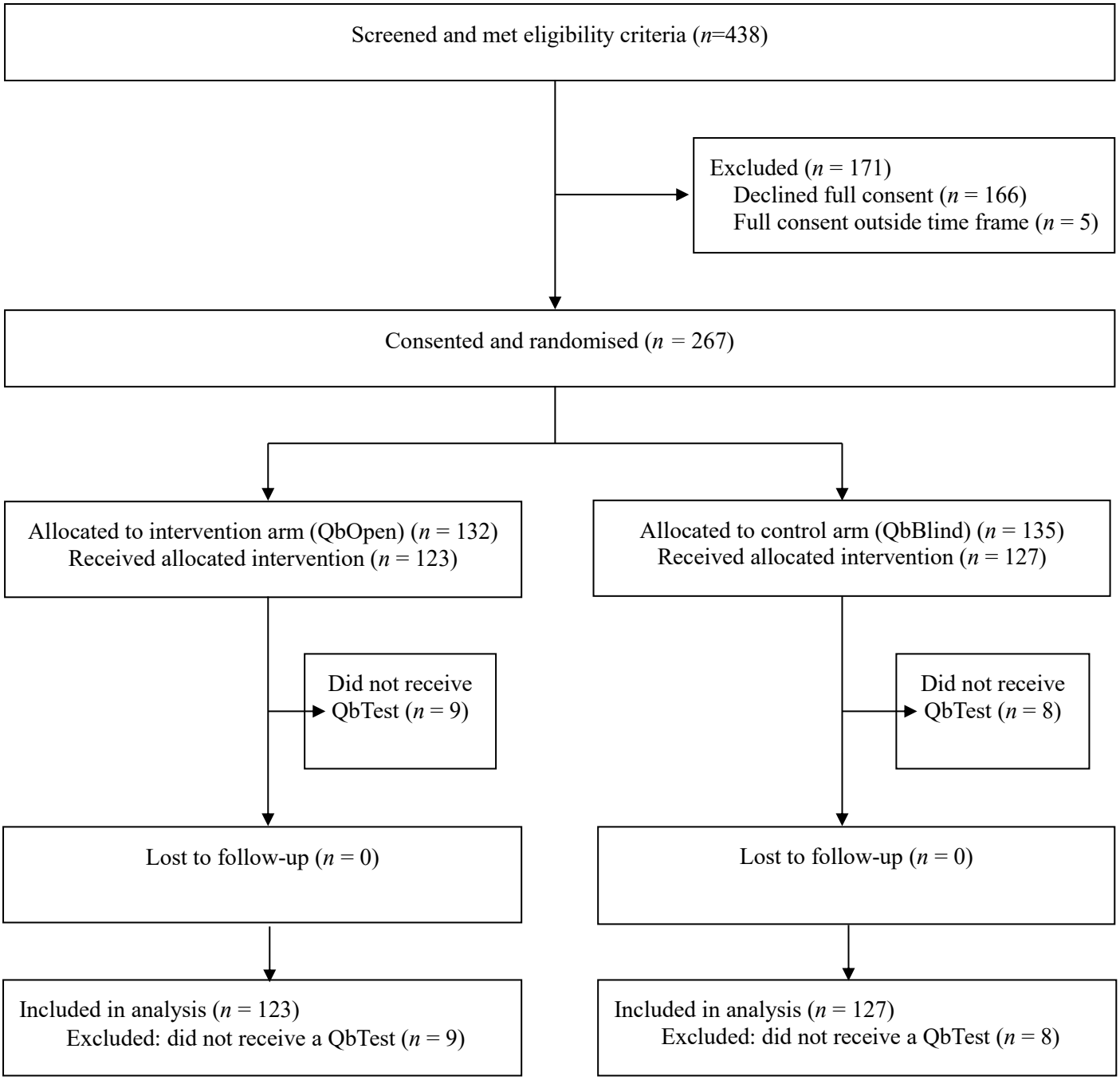
Missing values will be checked and reported across group and follow up time. For the primary outcome the number of visits needed for a confirmed diagnosis up to 6 month and for the secondary outcome variable, the time to diagnosis in days, patients who do not have a confirmed diagnosis up to 6 month or have withdrawn from the study will have their time to event censored on the date of last follow-up. Logistic regression will be used to explore the influence of group status on outcome missingness. For secondary outcomes having both baseline and one follow up measure, logistic regression will be used to explore the influence of group status and baseline measure on outcome missingness. For secondary outcome having two follow-up measures, multilevel logistic regression with patient as level 2 unit will be used to test the impact of group status and baseline measure on

outcome missingness. These results will be used to inform missing value imputation. Multiple imputation will be used to handle missing values in secondary outcome variables under missing at random (MAR) assumption (Carpenter & Kenwood, 2013); imputation based on analytical multilevel modelling approach will be performed to impute missing values in repeated measures under MAR assumption (Carpenter et al., 2011). Results of analytical model with observed data will be examined as sensitivity analysis. STATA14 and REALCOM software will be used to run multiple imputations.

4. Description of participant characteristics

4.1. Participant flow

Figure 1: **Trial profile**



4.2. Baseline characteristics

Continuous data that are approximately normally distributed will be summarised in terms of the mean, standard deviation, median, minimum, maximum and number of observations. Skewed data will be presented in terms of the maximum, upper quartile, median, lower quartile, minimum and number of observations. Categorical data will be summarised in terms of frequency counts and percentages.

Both cluster level and participant-level summary data will be recorded. This will include:

- Demographic variables (e.g. age, sex, ethnicity, height and weight) by treatment group and overall
- Disease factors (e.g. specific entry criteria, severity of disease, baseline values for critical clinical measurements) by treatment group and overall
- Comorbidity: concomitant illness by treatment group and overall
- Medication: concomitant medication by treatment group and overall.

Table: Summary of Participant demographic background information

Variable	QbO	ObB

5. ASSESSMENT OF STUDY QUALITY

5.1. Randomisation

Participants will be randomised on a 1:1 ratio into either the QbOpen (QbO) arm or QbBlind (QbB) arm of the trial.

5.2. Adherence

Adherence to protocol is defined as the participant conforming to the inclusion/exclusion criteria and having received a QbTest as part of their diagnostic assessment. Their clinician must complete accurate pro-formas reflecting their consultation for the six-months the participant remains in the trial. The clinician must record a pro-forma when an ADHD diagnosis is confirmed or excluded. If allocated to the blind arm of the trial, the clinician should not see the QbTest report.

Any one not (1)conforming to the inclusion/exclusion criteria and (2)having received a QbTest as part of their diagnostic assessment will not be included in analysis, as per ICH9 document suggestion on data analyses set[6].

5.3. Protocol deviations

There are no protocol deviations.

5.4. Changes made to the planned statistical analyses

The main analyses will compare the number of visit needed to achieve an ADHD diagnosis (either confirmed or excluded) between the two groups. The original protocol stated the primary outcome would be analysed using Poisson regression. However, given that not all participants would be diagnosed within 6-months, we sought expert independent advice (not based on any interim analysis or access to blinded or unblinded data) on the suitability of the analysis. Based on this advice, we changed the analysis of the primary outcome measure from regression analysis to survival analysis. This approach is more suitable for our data and allows all participants to be included in the analysis of the primary outcome (regardless of whether they reached a diagnosis), which reflects a more appropriate analysis than the original regression analysis, and fits with ITT.

6. ANALYSIS OF EFFECTIVENESS/EFFICACY

6.1. Summary of primary and secondary outcomes

Primary and secondary outcome measures have been described above.

6.2. Primary analysis

As there are lots of ties within each time point, so the discrete time survival analysis using complementary-log-log regression was recommended to be performed to quantify the treatment effects [7]. As the center effects should be adjusted, so multilevel complementary-log-log regression will be performed to quantify the treatment effect.

The model will be

$$Pr(y_i=1)= 1\text{-}exp(\text{-}exp(\beta_0+ \beta_1 group_{ij}+ \Sigma\beta_{k-1}(dummy\ time)+\mu_j))$$

Exponentiated β_1 (95%CI) is the treatment effect estimate expressed as **Hazard Ratio**, k= 1~ max visit time[1], i =center ID , j =center ID, μ_j is center site level deviance. The variability at center level will be explored with non-statistically significant variance to be fixed for model parsimony purpose.

STATA code:

mecloglog visit i.group (dummys) || center: , eform

center to be replaced with center ID variable name in dataset.

[1] https://www.iser.essex.ac.uk/resources/survival-analysis-with-stata

6.3. Secondary/sensitivity analysis of primary outcome

Secondary analysis of primary outcome

Secondary analysis 1: using ML Weibull modelling, for HR and time ratio estimates. HR estimated from ML Weibull modelling will be used as sensitivity analysis and time ratio used to aid interpreting results.

Secondary analysis 2: using ML logistics model for discrete survival model. (sensitivity analysis)

Secondary analysis 3: using complementary log log regression without adjusting site effects, as some center only have few children following suggestion by EMA document[8]. (sensitivity analysis)

Secondary analysis 4: using ML complementary log log model with log time baseline hazahrd model for discrete survival model. (sensitivity analysis)

6.4. Secondary outcomes

All secondary outcome measures have been described above.

As the trial is a multicenter trial with randomization stratified by clinical site, Site effects will be adjusted by Multilevel modelling (MLM) where possible and needed. Outcome variability at centre level will be checked first, if non-significant center effects shown, modelling without adjusting centering effect will be performed[9]. Shape of each continuous variables will be checked, skewed variable will be transformed using appropriate method prior linear regression modelling. If skewness can't be corrected, quantile regression will be considered.

- Same survival analysis modelling will be applied for the **days needed for a diagnosis up to 6 months**. Median survival time will be derived.
- Same survival analysis modelling will be applied for the **total minutes needed for a diagnosis up to 6 months**. Median survival time will be derived.
- Proportion who receive correct definitive diagnosis: multilevel logistic regression with binary arm as explanatory variable.

The ML logistics used is $\text{logit}\,(\pi_{ij}) = \ln\dfrac{\pi_{ij}}{(1-\pi_{ij})} = \beta_0 + \beta_1 group + u_j$, with $u_{0j} \sim N(0, \sigma_u^2)$,

Where σ_u^2 is the site level variance. β_1 is treatment effect, fixed parameter. j: site ID

Stata code: melogit i.arm || siteid:,

- Differences in degree of confidence of diagnosis, measured by a 6-point ordered single question: Ordinal logistic regression with binary arm as explanatory as explanatory variable.

  The ML ordinal Logistics model is $\text{logit}\,(\gamma_{ij}^s) = \alpha^s + \beta_1 group_{ij} + u_{0j}$, u_{0j} is a random effect for site, $u_{0j} \sim N(0, \sigma_u^2)$, σ_u^2 is the site level variance, β_1 is treatment effect, fixed parameter. Assuming outcome has t categories, indexed by s (s = 1; … ; t-1), j: site ID. (Ref MLwiN manual).

  Stata code: meologit

- The stability of diagnosis between 1st and re-rated will be conducted by Kappa statistics for each group. The Arm difference of Stability in diagnosis between first and re-rated diagnosis:

stability coded as 0/1 with same diagnosis as 1, otherwise as 0, ML logistic regression with binary arm as explanatory variable.

Stata code: kap diag0 diag2

Kappa:0~1, with 1 indicating perfect agreement. A rough guide to interpreting kappa is as follows:

Value	Strength of agreement
<0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very Good

- The stability of confidence between 1st and re-rated will be conducted by Kappa statistics for each group. The Arm difference of Stability in confidence between first and re-rated diagnosis: stability coded as 0/1 with same diagnosis as 1, otherwise as 0, logistic regression with binary arm as explanatory variable. Code and formulae ref stability for diagnosis.
- Diagnostic accuracy: Correct classification rate on QbTest vs gold standard classification rate (assessed via DAWBA completed by the parent and teacher). Overall and group specific result will be presented.

Each child will have ADHD/not diagnosis1 based on Qb test results and ADHD/not diagnosis2 based on DAWBA. Following diagnosis accuracy indicator will be presented:

		DAWBA	
		adhd	No
Qb	adhd	a	B
	no	c	D

Sensitivity= $a/(a+c)$, Sensitivity= $d/(b+d)$

Positive Predictive Value ppv**+=** $a/(a+b)$, negative Predictive Value ppv**-=** $d/(d+c)$

The “positive likelihood ratio” (LR ve+) tells us how much to increase the probability of disease if the test is positive. On the other hand, the “negative likelihood ratio” (LR ve-) tells us how much to decrease it if the test is negative

LR ve+

=

(probability that an individual **with** disease has a positive test)/(probability that an individual **without** disease has a positive test)

=sensitivity / (1-specificity)

LR ve-

=

(probability that an individual **with** disease has a negative test)/(probability that an individual **without** disease has a negative test)

=(1-sensitivity) / specificity

LR not change as disease prevalence change, but PPV change with prevalence.

Stata code for proportion and 95%CI.

cii #obs #succ

QbTest scores (Q scores for attention, impulsivity and activity) will be compared with DAWBA ADHD categorisation to obtain the best predictive model based on QbTest scores that discriminates between ADHD 'positive' and ADHD 'negative' gold standard DAWBA diagnoses. ROC analysis for overall data. Stata code: rocfit and rocfit postestimation such as rocplot. Ref Stata rocfit methods & formula

- Side effects scales will be used as a control check to ensure speed in diagnosis/medication normalisation is not achieved at the cost of greater adverse-effects.

Side effects was measured by questionnaire as continuous variable so ML linear regression will be used, the model will be $y_{ij} = \beta_0 + \beta_1 group + u_j + e_{ij}$, with $u_{0j} \sim N(0, \sigma_u^2)$, Where σ_u^2 is the site level variance and e_{ij} is residual term. β_1 is treatment effect as fixed parameter, j is site ID.

- Proportion of patients achieving normalisation, at 6 months measured via SNAP-IV scores. This will assess whether QbTest improves normalisation rate.

Normalsation will be defined SANP<=1[10]. ML logistics regression with site as level 2 ID will be performed. Model information ref above ML logistic regression formula and stata code.

- Changes in scores of QbTest and SNAP-IV will also be compared from baseline to 4-8 week follow-up.
- SDQ scores (parent & teacher completed) conducted at baseline (through DAWBA) and follow-up will look at outcome of patients between the two groups. C-GAS scores conducted at baseline, 3-months and 6 month follow-up will provide a clinician rated measure of outcome.

  Repeatedly measured SNAP-IV, SDQ, C-GAS will be compared using ML linearly regression with site as level 3 unit and child as level 2 unit. The model will be
  $y_{ijk} = \beta_0 + \beta_1 group + \beta_2 time + \beta_3 group * time + \beta_4 baseline + v_k + u_{jk} + e_{ijk}$,
  with $v_k \sim N(0, \sigma_v^2)$, where σ_v^2 is the site level variance, $u_{jk} \sim N(0, \sigma_u^2)$ and σ_u^2 is the patient level variance, e_{ij} is residual term. β_1 is treatment effect as fixed parameter, k is site ID and j is children ID. y_{ijk} is outcome measured at 1st and 2nd follow up time.

  SDQ have only baseline measure and one followup, so two level model with site as level 2 unit to be performed.
  Stata code: mixed for MLM, margin to derive group comparison at each time.

- EQ-5D-Y will compare quality of life between QbOpen and QbBlind group at baseline, 3- month sand 6 month follow up.

6.5. Other analysis

The clinical lead for the trial at each of the ten sites will be invited to interview about their experience of the QbTest and being part of the AQUA-Trial. A random selection of parents and young people who had participated in the trial will also invited to interview. The interviews will be semi-structured and guided

by separate interview schedules for each participant group (HCP or families). By utilising a semi-structured approach, the researcher was able to ask additional questions based on the interviewees responses. Questions for discussion center around their experience of participating in the trial, their opinion on QbTest clinical utility, issues with current ADHD practice (clinician's only) and the future of the QbTest. A series of prompts have been generated for each question to aid discussion if required. Audio recordings were anonymised and transcribed verbatim. Half of the transcripts will be first analysed by CLH and half first analysed by HC using the guidelines of Braun and Clarke (2006) using an inductive approach. Each coding unit will be coded exclusively into just one category rather than multiple categories to created well defined coding categories (Joffe et al., 2004). To ensure validity and reliability of data interpretation, the transcripts will then re-coded by the other researcher (CLH or HC). The coders' epistemology will be that of an essentialist/realist paradigm (Braun & Clarke, 2006), which seeks to understand the opinions of QbTest through the words of the participants, as opposed to the researchers' co-created meaning.

7. ANALYSIS OF SAFETY

7.1. Adverse events

There are no anticipated adverse effects of the QbTest that would result in exacerbation of a pre-existing illness, increase in frequency or intensity of a pre-existing episodic event or condition, condition detected or diagnosed after medicinal product administration even though it may have been present prior to the start of the study or continuous persistent disease or symptoms present at baseline that worsen following the start of the study.
There are no anticipated side effects as result of the QbTest.
There are no anticipated Serious Adverse Event (SAE) as a result of this study that would result in death, a life-threatening adverse event, inpatient hospitalisation or prolongation of existing hospitalisation, disability/incapacity or a congenital anomaly in the offspring of a participant.
Any adverse event that occurs will be assessed for seriousness, expectedness and causality.

Reporting of adverse events

Participants will be asked to contact the study site immediately in the event of any serious adverse event. All serious adverse events will be recorded and closely monitored until resolution, stabilisation, or until it has been shown that the study treatment / intervention is not the cause. The Chief Investigator, Professor Chris Hollis, shall be informed immediately of any serious adverse events and shall determine seriousness and causality in conjunction with any treating medical practitioners.
All treatment related serious adverse events will be recorded and reported to the REC as part of the annual reports. Unexpected serious adverse events will be reported within the timeframes to the REC as stated below. The Chief Investigator, Professor Chris Hollis shall be responsible for all adverse event reporting.

Trial Treatment / Intervention Related SAEs

A serious adverse event that is unexpected in its severity and seriousness *and* deemed directly related to or suspected to be related to the trial treatment or intervention shall be reported to the ethics committee that gave a favourable opinion as stated below.

The event shall be reported immediately of knowledge of its occurrence to the Chief Investigator.

The Chief Investigator will:

- Assess the event for seriousness, expectedness and relatedness to the trial treatment or intervention.
- Take appropriate medical action, which may include halting the trial and inform the Sponsor of such action.
- If the event is deemed related to the trial treatment or intervention shall inform the REC using the reporting form found on the NRES web page within 7 days of knowledge of the event.
- Shall, within a further eight days send any follow-up information and reports to the REC.
- Make any amendments as required to the study protocol and inform the REC as required

Participant removal from the study due to adverse events

Any participant who experiences an adverse event may be withdrawn from the study at the discretion of the Investigator. No adverse events are anticipated as a result of taking part in this study.

8. FINAL REPORT TABLES AND FIGURES

9. REFERENCES

REFERENCES

1. Feiveson A. Power by simulation. *Stata J* 2002; **2**(2): 107-24.
2. Van Belle G. Statistical rules of thumb: John Wiley & Sons; 2011.
3. Smeeth L, Ng ES-W. Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC Trial of the Assessment and Management of Older People in the Community. *Controlled Clinical Trials* 2002; **23**(4): 409-21.
4. Elley CR, Kerse N, Chondros P, Robinson E. Intraclass correlation coefficients from three

cluster randomised controlled trials in primary and residential health care. *Australian and New Zealand journal of public health* 2005; **29**(5): 461-7.
5. Sayal K, Daley D, James M, et al. Protocol Evaluating the effectiveness of a school-based group programme for parents of children at risk of ADHD: the 'PArents, Teachers and CHildren WORKing Together (PATCHWORK)'cluster RCT protocol. *BMJ open* 2012; **2**(5).
6. European Medicines Agency. Statistical principles for clinical trials (E9): ICH tripartite guideline: European Medicines Agency; 1998.
7. Jenkins SP. Survival Analysis. Colchester, UK: Institute for Social and Economic Research, University of Essex; 2005.
8. European Medicines Agency. Guideline on adjustment for baseline covariates. London, UK: European Medicines Agency; 2013.
9. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practiceand problems. *Statistics in Medicine* 2002; **21**(19): 2917-30.
10. Swanson JM, Kraemer HC, Hinshaw SP, et al. Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001; **40**(2): 168-79.