ASTAR Trial

A randomised controlled pilot trial comparing a parent training intervention (ASTAR B) to attention control (ASTAR A) in parents/carers of a child aged 4-8 years diagnosed with ASD

Statistical Analysis Plan

Version 1.0 27/04/2019

ISRCTN: ISRCTN91411078

Trial Statistician: Dominic Stringer

Senior Statistician: Professor Andrew Pickles

Signature Date 28/05/2019

Chief Investigator: Professor Tony Charman

Signature Date 27/04/2019

Trial Steering Committee Chair: Professor Alan Stein

Signature Date 10/05/2019

ASTAR Statistical Analysis Plan

ASTAR Statistical Analysis Plan

1 QUANTITATIVE ANALYSIS PLAN

Principal investigators: Professor Tony Charman and Professor Stephen Scott

Trial project manager: Dr Melanie Palmer

Trial statisticians: Dominic Stringer, Professor Andrew Pickles

1.1 Description of the trial

This analysis plan should be read as a supplement to the trial protocol; as such a description of the trial design and other aspects of the trial will not be duplicated here. This analysis plan was written according to Trial Protocol V1.4 04.02.2019.

1.1.1 Measures covered by this analysis plan

Primary, secondary and exploratory outcome measures are described in the protocol. This section will only serve to outline which of these measures will be covered by this statistical analysis plan (SAP) and who is responsible for the analysis of each of these measures.

Primary and secondary outcomes listed below will be reported in the main paper and analysed by the trial statistician(s). Health economics outcomes will be analysed by the Health Economists. Any other outcomes are not covered by this SAP.

Primary outcome:

- Observed child behaviour as measured using the OSCA-AB. The frequency of a range of child behaviours that challenge (destructive behaviour, aggression towards themselves and others, frustrated vocalisations, non-compliance, avoidance and reassurance seeking) during the assessment are coded from video by blinded researchers and summed. This produces a semi-continuous score, potentially subject to a floor reflecting how some children show little challenging behaviour during observation. As the length of the observations vary, the rate of child behaviours that challenge per minute will be calculated.

Secondary outcomes:

- Observed child compliance during the OSCA-AB. The frequency of child compliance will be coded from video by blinded researchers. The rate of child compliance per minute will be calculated.
- Observed parent behaviour during the OSCA-AB. The frequency of child-centred parenting behaviours (positive comments, clear commands, praise, and supportive physical guidance) and child-directive parenting behaviours (negative comments, unclear commands, no opportunity to comply, and physical handling) are coded from video-recordings by blinded raters and summed. The proportion of child centred parenting behaviour / child centred and child directive parenting behaviours will then be calculated and reported alongside the frequency of child centred and child directed behaviours per minute.

- Aberrant Behaviour Checklist (ABC) Irritability and Hyperactivity subscales Total scores – parent and teacher rated.
- Assessment of Concerning Behaviour Scale (ACB) Internalising and Externalising Total scores – parent and teacher rated.
- Home Situations Questionnaire (HSQ)-ASD Mean Per-Item Severity score and Mean Per-Item Severity scores for the Demand Specific and Socially Inflexible subscales
- Preschool Anxiety Scale Revised (PASR) Total Scores
- Parent Defined Target Symptoms score
- Clinical Global Impressions-Improvement Scale (CGI-I) score, based on Parent-Defined Target Symptoms and parental views of global improvement in their children
- Autism Parenting Stress Index Total Stress score
- Child Adjustment and Parent Efficacy Scale-Developmental Disability Parent Efficacy subscale (CAPES-DD Parent Efficacy) Total score
- Parenting Scale (PS) Mean score and Means for the Laxness and Overreactivity subscales
- Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS) Total score
- Adverse events

Health Economic outcomes:

- Client Service Receipt Inventory (CSRI)
- EQ-5D 5L
- Office of National Statistics (ONS) Personal Wellbeing
- Therapist time use

Exploratory/other measures

- Bespoke measures of acceptability/satisfaction and fidelity of the interventions

1.1.2 Sample size estimation (including clinical significance)

We expect that retention will be approximately 90%, as reported by other trials of psychological intervention trials among young children with ASD that included recruits from our service areas. We expect a more modest effect size than the 1.3 reported by Sofronoff et al (1) as this was for a parent-reported and thus unblinded measure. For the comparison of ASTAR A and ASTAR B, power was calculated by a non-central chi-square method using a linear mixed model with baseline (baseline-outcome correlation assumed 0.7) as covariate for two-tailed p=.05 and intraclass correlation for within intervention group of 0.02 and 10% drop-out. For an ES of 0.5, our study gives an expected 95%CI of 0.08, 0.92 and power of 64%, while for an ES of 0.6 the expected 95%CI is 0.18, 1.02 and 79% power.

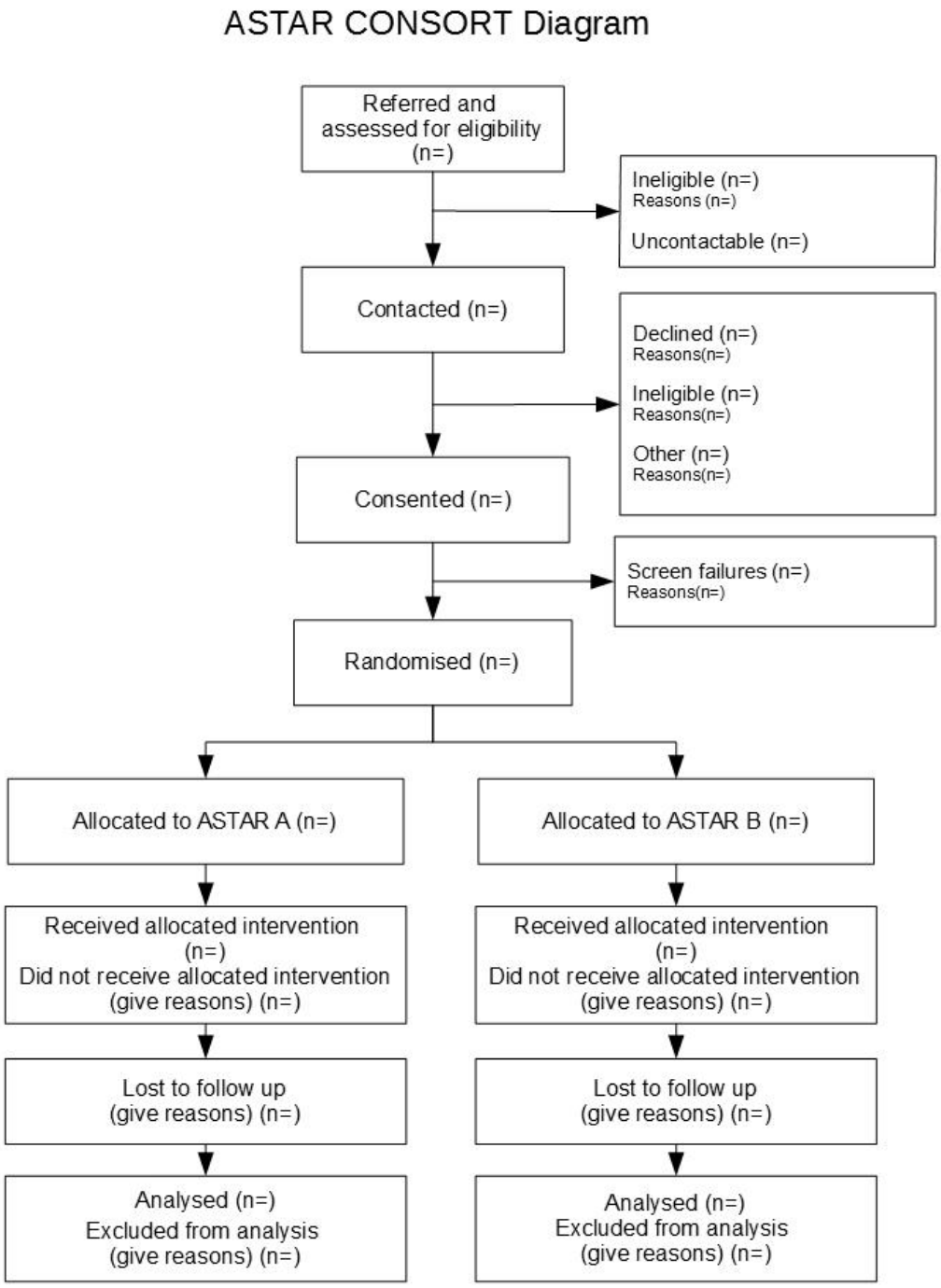
1.2 Data analysis plan – Data description

1.2.1 Recruitment and representativeness of recruited participants

ASTAR Statistical Analysis Plan

A CONSORT flow chart will be constructed (2) – see Figure 2. This will include the number of eligible participants, number of participants agreeing to enter the trial, number of participants refusing, then by treatment arm: the number continuing through the trial, the number withdrawing, the number lost to follow-up and the numbers excluded/analysed.

Figure 1. Template CONSORT diagram for ASTAR trial



1.2.2 Baseline comparability of randomised groups

Baseline descriptions of participants by trial arm and overall: means and standard deviation or numbers and proportions as appropriate. No significance testing will be used to test baseline difference between the randomised treatment groups.

1.2.3 Adherence to allocated treatment and treatment fidelity

Compliance with the therapy will be described in terms of number of therapy sessions attended. Reasons for withdrawal from therapy will be summarised.

Treatment fidelity will be summarised by trial arm. Checklists measuring intervention fidelity have been developed and are completed by the therapists after each session. The measure explores fidelity related to session content (6 items) and 'group process' (6 items), scored on a scale of 0-2 (0=not covered; 1=partially covered; 2=fully covered). The mean fidelity of content and group process ratings summed over all sessions will be reported per arm.

Parents provided satisfaction ratings of their experience of the groups via a 6-item, self-report questionnaire, completed anonymously at post-intervention. Questions asked about overall satisfaction with the course, whether parents felt supported, whether they would recommend the course to a friend and whether they felt the course had helped with their child's behaviours and emotions.

1.2.4 Loss to follow-up and other missing data

The proportions of participants missing each variable will be summarised in each arm and at each time point.

The baseline characteristics of those missing follow up will be compared to those with complete follow up.

The reasons for withdrawal from the trial will be summarised.

1.2.5 Adverse event reporting

Serious adverse events (SAE) and serious adverse reactions (SAR) will be summarised by arm. The total adverse events (AE) and adverse reactions (AR) grouped by (1) deterioration in child behaviour/wellbeing and (2) deterioration in parental or other family member wellbeing will be reported by arm, and any individual events/reactions that are experienced by more than 10% of the sample.

1.2.6 Assessment of outcome measures (unblinding)

This is a single blind study; the outcome assessors and the trial statisticians will be blind until the primary and secondary outcomes are analysed (prior to analysis of compliance).

1.2.7 Descriptive statistics for outcome measures

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

Parental satisfaction and fidelity of implementation of the intervention will be reported descriptively.

1.3 Data analysis plan – Inferential analysis

1.3.1 Main analysis of treatment differences

Analyses will be carried out blinded; trial statisticians will only be unblinded upon completion of all inferential analyses. The main analyses will use the intention-to-treat population, including all participants who were randomised according to the group they were allocated to. Statistical tests and confidence intervals will be two-sided. Between-group comparisons will be calculated and presented with 90% confidence intervals.

1.3.2 Analysis of primary outcomes

We will test for a between-group change in the primary outcome at the post-intervention. The distribution of the primary outcome at baseline will be examined for evidence of floor effects. Where present an appropriate transformation and if required a cut-point will be chosen, prior to extraction of an unblinded randomisation file, to allow an assumption of left-censored Gaussian distribution to be applied. Dummy variables will be used to account for randomisation stratification and group clustering. Where floor effects are absent, the analysis will be an ANCOVA regression predicting outcome where the endpoint is also covaried for baseline. Where floor effects are present, a generalized mixed model/structural equation modelling setup in which both baseline and endpoint are modelled as potentially censored response variables will be used, with a covariance between equations that yields in the complete data case the ANCOVA estimate of treatment effect in the absence of censoring.

1.3.3 Analysis of secondary outcomes

Secondary outcomes will be analysed in the same way as the primary outcome.

Statistical considerations

1.3.4 Time points

There are only 2-time points, baseline and post-intervention; deviation of measurement of the planned post-intervention timepoint will be summarised by treatment group.

1.3.5 Stratification and clustering

Randomisation has been stratified by ADOS module and Site. Therefore, these will be included as covariates in the analysis. As interventions occur in groups, there is expected to be a group effect. The model will therefore account for clustering within group.

1.3.6 Missing items in scales and subscales

The number (%) with complete data will be reported. Where available the missing value guidance provided by authors of scales will be used. In its absence, scales will be pro-rated for an individual if 20% or fewer items are missing. For example, in a scale with 10 items, prorating will be applied to individuals with 1 or 2 items missing. The average value for the 8 or 9 complete items will be calculated for that individual and used to replace the missing values. The scale score will be calculated based on the complete values and these replacements.

1.3.7 Missing baseline data

Missing baseline data should not be an issue for the primary analysis. Some extensions to this analysis may use other baseline variables; if these contain missing data, the number with complete data will be reported and they will be imputed using a method suitable to the variable as per the recommendations of White and Thompson (3).

1.3.8 Missing outcome data

If there is missing outcome data, we will investigate to see if any baseline variables are predictors of outcome missingness. Such variables could then be included as covariates in the model if deemed suitable for adjustment.

If post treatment variables such as compliance (to therapy), or baseline variables that it would not be suitable to adjust for in the main analyses, are found to be predictive of drop out, multiple imputation will be considered.

1.3.9 Method for handling multiple comparisons

No adjustment will be considered for multiple comparisons/testing, allowing reviewers to make their own adjustment to significance, precision and bias if they wish. However, care will be given to the interpretation of inference for the numerous secondary outcomes.

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

None; non-compliance will be described only (in terms of number of therapy sessions attended).

1.3.11 Model assumption checks

The models assume (potentially censored) normally distributed outcomes; this will have been checked when describing the data and if substantial departures from normality occur, transformations or bootstrapping of the standard errors will be considered. Residuals will be plotted to check for normality and inspected for highly influential observations.

1.3.12 Sensitivity analyses

Where it has been necessary to analyse the primary outcome within a model accounting for floor effects, the analysis will be repeated for a range of plausible thresholds for the instrument floor.

1.3.13 Planned subgroup analyses

None planned.

1.3.14 Exploratory analyses

None planned for the main paper and so not covered in this SAP.

1.3.15 Exploratory mediator and moderator analysis

None planned for the main paper and so not covered in this SAP.

1.3.16 Interim analysis

No interim analyses are planned of post-randomisation data. An analysis of baseline data examined the reliability and validity of the OSCA-AB outcome, and with the agreement of both DMC and TSC the OSCA-AB child behaviour that challenges was confirmed to be the nominated primary outcome measure.

1.4 Software

Data management: Data will be collected in SPSS databases and in the Delosis Psytools online system.

Analyses will be performed in Stata (4). R (5) may additionally be used for descriptives, report generation and/or production of graphs.

2 HEALTH ECONOMICS ANALYSIS PLAN

2.1 Definition of Outcome Measures (inc. trial periods)

Service Use
Comprehensive data are being collected on all health, social care and other relevant services used by individual study members using a tailored version of the Client Service Receipt Inventory (CSRI). The CSRI was used at baseline and post-intervention follow-up each time covering resource use. The CSRI covers the following broad categories of information.

- Use of health, social care public, private and voluntary services used by the child, as well as services used by other family members that are linked to the child's ASD (insofar as the link can be made)
- Unpaid family support
- Employment impacts due to time off work

Cost calculation
The costs of each resource item will be calculated using best available unit cost estimates (6). The cost of parent training intervention and attention control will be estimated using information on the core resource inputs involved in delivering the interventions.

QALY measurement
The EQ-5D consists of five domains (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). Each of these will receive a score of 1, 2, 3, 4, 5 corresponding to no problems, slight problems, moderate problems, severe problems, and extreme problems. Utility scores will be attached to each health state based on these scores (a table of utility values has been produced by the Centre for Health Economics, University of York). These utility scores will be used to generate QALY gains over the follow-up period.

2.2 Primary Analysis (incl. method of analysis)

Cost comparisons
Regression analysis will be used to compare service costs and total costs between the two intervention arms (ASTAR A and ASTAR B) which will each be represented by dummy variables (1 indicating group membership, 0 representing membership of another group).

Cost data are usually skewed and if this results in similarly skewed residuals then the standard linear model is inappropriate. The distribution of the regression residuals will be checked visually and if the distribution is non-normal we will use bootstrapping with 10,000 resamples to estimate 95% confidence intervals around the cost differences. (Confidence intervals will be based on the percentile or bias-corrected method depending on the level of bias observed in the model.)

Economic evaluation

The primary economic evaluation post-intervention (i.e., a cost-effectiveness analysis) will include cost and outcomes for the two randomization groups—ASTAR A and ASTAR B — from three perspectives: a health and social care perspective (which will include health and social care services used by the child, as well as health and social care services used by other family members that are linked to the child's ASD); public sector perspective (which will include health, social care public, private and voluntary services used by the child, as well as services used by other family members that are linked to the child's ASD) and a wider societal perspective (which will include health, social care public, private and voluntary services used by the child, as well as services used by other family members that are linked to the child's ASD, unpaid support and employment impacts from time off work). The cost-effectiveness analysis will focus on two outcomes: (1) Observed child behaviour that challenges as measured using the OSCA-AB, and (2) QALYs.

Cost-effectiveness will be assessed by linking data on service cost differences OSCA-AB differences and QALYs gains in turn. If any group has significantly lower costs and significantly better outcomes, then it will be deemed to be more cost-effective. If costs are significantly higher and outcomes significantly better or if there is uncertainty in these findings (indicated by the confidence intervals) then we will use the net benefit approach and cost-effectiveness acceptability curves to assess cost-effectiveness.

A secondary analysis will use quality adjusted life years (QALYs) (derived from the EQ-5D) as the outcome measure.

2.3 Models

We will run a multilevel mixed effect model. In the cost and outcome models, we include the treatment variable as a random effect at the cluster level, account for stratification and covarying for baseline.

The cost-effectiveness of ASTAR A and ASTAR B will be evaluated by calculating incremental cost-effectiveness ratios (ICERs) if appropriate. The ICER will be defined as the between group difference in mean costs divided by the between group difference in mean outcome: (1) Observed child behaviour as measured using the OSCA-AB and (2) QALYs. Using these ratios, we will be able to determine a cost–per unit improvement in QALYs and a cost per child behaviour that challenges.

The main focus of the analysis will be to plot data on the cost-effectiveness plane and their associated cost-effectiveness acceptability curves (CEACs) under each of the perspectives. These plots will be based on bootstrapped sample means, generated from cost–outcome pairs from the primary outcome child behaviour that challenge and EQ-5D data in turn. Interpretation of the CEACs will be based around the probability of cost-effectiveness in the £20,000–30,000 per QALY range, to reflect the thresholds typically used by NICE to identify which interventions to fund. We will present probability of cost-effectiveness using an appropriate cost per improvement in parent and child behaviour range, as there are no established thresholds typically used.

Bootstrapping will be conducted with the mixed-effects model to generate confidence intervals (CIs) to capture parameter uncertainty around the estimates in the cost-effectiveness analysis. The 1000 treatment effect replications from this bootstrapping process were plotted on cost-effectiveness planes and used to construct cost-effectiveness acceptability curves (CEACs), which are also a standard practice in trial-based economic evaluations. Dummy variables will be used to account for randomisation stratification (ADOS module and site), clustering effects of group, baseline costs and outcomes.

2.4 Other Analyses Supporting the Primary Analysis (inc. sensitivity analyses)

The main analyses will use an informal care unit cost based on the replacement method (where the cost of a homecare worker is used as a proxy for informal care). We will alternatively use a zero cost and a cost based on the national minimum wage for informal care. We will also conduct sensitivity analyses around the costs attached to lost employment.

We will explore other impacts on the sensitivity of the results based on what is presented in the data.

The ONS personal wellbeing questionnaire will not form part of the cost-effectiveness analysis but will be used to compare our sample data with those from the general population.

3 SCHEDULE OF ASSESSMENTS AND MEASURES

Month	0/1	2	3	4	5/6
Research/home setting					
Baseline	Demographics; ADOS–2; SCQ; ABAS–3; OSCA–AB; ABC; ACB; HSD-ASD; PASR; APSI; CAPES-DD PE; PS; PTS; SWEMWBS; ONS wellbeing; EQ-5D 5L; CSRI				
Post-					OSCA–AB; ABC; ACB; HSD-ASD; PASR; APSI; CAPES-DD PE; PS; PTS; SWEMWBS; ONS wellbeing; EQ-5D 5L; CSRI; CGI-I; Adverse events
Education setting					
Baseline	ABC-T; ACB-T				
Post-					ABC-T; ACB-T
Clinic/home setting					
Intervention		12 group sessions 2 individual sessions (ASTAR B only)			

ABAS–3=Adaptive Behaviour Assessment System – 3rd edition; ABC=Aberrant Behaviour Checklist; ACB=Assessment of Concerning Behaviour; ADOS–2=Autism Diagnostic Observation Schedule – 2nd edition; APSI=Autism Parenting Stress Index; CAPES-DD=Child Adjustment and Parent Efficacy Scale-Developmental Disability; CGI-I=Clinical Global Impression-Improvement; CSRI= Client Service Receipt Inventory; HSQ-ASD=Home Situations Questionnaire-Autism Spectrum Disorders; ONS=Office of National Statistics; OSCA–AB=Observation Schedule for Children with Autism – Anxiety and Behaviour; PASR= Preschool Anxiety Scale Revised; PS= Parenting Scale; PTS=Parent Target Symptoms; SWEMWBS=Short Warwick-Edinburgh Mental Wellbeing Scale; SCQ=Social Communication Questionnaire.

4 REFERENCE LIST

1. Sofronoff K, Attwood T, Hinton S. A randomised controlled trial of a CBT intervention for anxiety in children with Asperger syndrome. *J Child Psychol Psychiatry*. 2005 Nov;46(11):1152–60.

2. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010 Mar 24;340:c332.

3. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med.* 2005 Apr 15;24(7):993–1007.

4. StataCorp. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC; 2017.

5. R Development Core Team. R: A language and environment for statistical computing [Internet]. *R Foundation for Statistical Computing*. Vienna, Austria; 2008. Available from: http://www.R-project.org.

6. Curtis LA, Burns A. Unit Costs of Health and Social Care 2017 [Internet]. 2017 [cited 2019 Mar 28]. Available from: http://dx.doi.org/10.22024/UniKent/01.02/65559