IMPACT TRIAL Statistical and Health Economics Data Analysis Plan
Version 1.0
Chris Roberts, Sarah Byford, Fiona Holland, Barbara Barrett
18th November 2014
Approved By the IMPACT Investigator Committee
Chief Investigator Ian M Goodyer

This document considers the final statistical and health economic analyses for the IMPACT trial. The purpose of the SAP is to document the confirmatory statistical analyses of the trial thereby controlling for statistical analyses bias. The statistical analyses follow the principles of ICH E9.

1. Aims

Improving Mood with Psychoanalytic and Cognitive Therapies, the IMPACT Study, will determine whether both medium intensity Cognitive Behavioural Therapy (CBT) [up to 20 sessions] and high intensity short term psychoanalytic therapy (STPP) [up to 28 sessions] are superior in reducing relapse compared to low intensity specialist clinical care (SCC) that is primarily advice and support [up to 12 sessions] in adolescents with moderate to severe depression attending routine child or adolescent mental health clinics. An additional aim is to establish the cost-effectiveness of CBT and STPP compared to SCC.

2. Study Design

Originally it was proposed that the trial will run in six CAMH clinics in each of three centres, giving 18 clinics with a minimum of one therapist for each treatment modality in each clinic and ten patients per treatment modality recruited in each clinic. This gives a total sample size of 540.

The ADAPT trial gave an SD of 14·6 at 28 weeks follow-up and correlation between baseline and follow-up of 0·41 for MFQ, proposed primary outcome of this study. We have assumed five points on the MFQ to be the minimum clinically important difference. This is approximately 25% of the change in the MFQ scale from baseline to 28 weeks. It is equivalent to a one point improvement on five of the 34 items of the scale. It is a standardize effect size of 0·34 (small to medium) and corresponds to non-overlap between treatments of approximately 25% (Cohen, 1988). Table 1 below gives estimates of power for Superiority, Non-Inferiority, and Equivalence designs for an intra-therapist correlation coefficient of 0·0, 0·025, or 0·05. Provided that the intra-cluster correlation is less than 0·025 a superiority analysis comparing CBT with STPP will have a power of over 80%. By virtue of the increased sample size specialist comparisons of the specialist treatments (CBT and STPP) with treatment as usual (SCC) will have substantial power. These power calculations assume a cross-sectional analysis, but statistical analysis will be based on longitudinal data using a linear mixed effects model (LME, see Section 8·2). Use of such a model will increase the power of the statistical analysis as data is in effect shared across follow-up time-points. This power calculation assumed a 90% follow-up as 92% follow-up at 28 weeks was achieved in ADAPT.

Allocation to treatment group was by minimisation controlling for severity (defined by MFQ score), sex, age, and recruiting region.

2·1 Statistical Hypotheses

The study was designed with a two level hypothesis: i) Both CBT and STPP will show superiority effects compared to SCC in the primary outcomes at 52 and 86 weeks; ii) CBT will show non-inferiority effects to STPP at 52 weeks; iii) STPP will show superiority effects compared to CBT at 86 weeks.

Table 1. Power assuming 18 therapists for each treatment modality, and ten patients per therapist

Intra-therapist correlation	Superiority	Inferiority	Equivalence
CBT vs STPP			
0	88%	93%	87%
0·025	80%	88%	75%
0·05	73%	82%	64%
(CBT+STPP) vs SCC			
0	96%	98%	96%
0·025	91%	95%	90%
0·05	85%	91%	82%

3. Outcome Measures

3·1 Primary outcome measure
Depressive symptoms over 36 to 86 weeks measured by the adolescent self report (Mood and Feelings Questionnaire, MFQ).

3·2 Secondary outcome measures

Along with the primary outcome, the secondary outcome measures are shown in the last column of Table 2 by frequency of collection and type of report. All are assumed to be continuous variables.

3·3 Hierarchy of young person versus parent reporting on various scales

When a young person and the parent both complete a particular questionnaire then the young person's data will form the basis of the main inference. The results from the parent will be supplementary.

4. Data analyses

Data analyses will be carried out by a statistician based in Biostatistics, Institute of Population Health, Manchester University, under the supervision of the trial statistician (CR) in conjunction with the IMPACT trial coordinator and trial centre in Cambridge. Economic data analyses will be carried out by a health economist based in the Centre for the Economics of Mental and Physical Health, Institute of Psychiatry, Kings College London, under the supervision of the trial economist (SB).

5. Data Sources

5·1 Pre randomisation data
Data are required for completion of the CONSORT diagram pre randomisation. These include:

- Numbers of potential participants assessed
- Numbers excluded after initial assessment by reason
- Numbers invited to baseline research interview
- Numbers excluded after baseline research interview by reason
- Numbers consenting and randomised by treatment arm.

5·2 Demographic data and patient characteristics prior to randomisation of randomised patients

The information collected at baseline consisted of basic demographic data (gender, age, twin, adopted/fostered, complications with pregnancy/labour/delivery, ethnicity, living arrangements, education, family employment) and clinical data (current medical problems, current medication of subject and other family member currently or in the past suffering any medical, emotional, or behavioural problems), plus the standard schedules as shown in Table 2.

5·3 Therapist/Care Provider and Treatment data

As suggested in the CONSORT guidance extension for trials of non-pharmacological interventions, information will be gathered regarding the characteristics of all therapists and care providers for each intervention. A statistical summary of this data will be prepared.

For each patient the intended therapist assigned at randomisation will be recorded along with the number of sessions attended.

The following minimum data will be collected for each trial therapy session (SCC, CBT, STPP):

- Therapist id
- Type of therapy delivered
- Individuals present in session

This will be aggregated to determine the number and type of therapy sessions received from each therapist.

A Kaplan-Meier plot and the associated log-rank test will be presented for time from randomisation to start of trial therapy and time to completion of trial therapy by a) treatment arm and b) by regional research centre (EA, NL, NW).

Summary statistics will be provided for duration of therapy by treatment arm.

5·4 Follow-up Assessments

The follow-up schedule is shown in Table 2.

6 Handling Missing Data and Slotting of Assessments

6·1 Item Non-response in Scale Measures

For questionnaire instruments, item non-response will be dealt with using a pro-rating strategy. Provided that at least 50% of items are available the observed total (for the completed items) and the number of items completed will be used to calculate an adjusted total as follows:

Adjusted total = Observed total * Total number of items in scale/Number of items completed

Note, this is equivalent to replacing the missing item by the average of available data for that dimension. The extent of pro-rata estimation will be reported for each scale for each treatment arm.

The NEO-FFI with five subscales and DSC with two subscales (global and affective) will have each subscale pro-rated and analysed separately.

6·2 Missing baseline covariate data

Subjects will not be excluded from outcome analyses due to missing baseline data. Where baseline covariate data (current not lifetime) cannot be obtained across different questionnaires, simple imputation (White & Thompson, 2005) which is based on multiple regression will be used. The following covariates will be used (see Section 8·2): region, comorbid behaviour disorders (CD+ODD), all anxiety disorders combined, SSRI use at baseline, age at randomisation, and sex. In addition baseline severity (MFQ score) will also be used. Substitution or imputation will not be used for post-baseline outcomes (see Section 8·3 for reasons).

6·3 Slotting of assessment measures

In this study there is often a delay in starting the therapy post randomisation and the visits (research assessments) are scheduled relative to the start of therapy rather than from time since randomisation. To provide summary statistics we need to assign each actual assessment to a target assessment week based on pre-defined intervals from time since randomisation. The assessment will be assigned to one of the following scheduled visit weeks based on the interval it falls into for the time (in weeks) from randomisation:

Scheduled Visit	Interval in Weeks since randomisation
0	<= rand date
6	>0 - 11
12	12 - 25
36	26 - 46
52	47 - 64
86	65+

Note, these bands may have to be modified when the data is inspected. To avoid bias this will be carried out blind to outcome scores by calculating summary statistics on the completeness of the primary outcome and will take account of the results of analyses on time to start and completion of trial therapies.

If, for a given assessment window, there is more than one measurement in the band then the measurement nearest to the week from randomisation will be used for descriptive statistics.

7 Descriptive Analyses of randomised patients

7·1 Baseline Characteristics

Patients in the three treatment groups (SCC, CBT, STPP) will be described separately with respect to the characteristics given in Section 5·2.

Numbers (with percentages) for binary and categorical variables, and ordered categories plus means, standard deviation, median plus minimum and maximum values for continuous variables will be presented. Consistent with CONSORT guidance there will be no tests of statistical significance or confidence intervals for differences between randomised groups for any baseline variable.

All baseline measurement scales will be summarised separately for adolescent and parent responses, by treatment arm.

7·2 Follow-up

All measurement scales in Table 2 taken during follow-up will be summarised separately for adolescent and parents, by visit and each treatment arm. Note, the assignment of data to a specific assessment visit will use slotting as described in Section 6·3.

7·3 Missing follow-up data

For the primary outcome measure (MFQ) the frequencies (with percentages) of patient losses to follow-up at 6, 12, 36, 52, and 86 weeks after randomisation will be reported and compared between arms. For each subject, the provision of a measurement at each time point will be based on the slotting procedure given in Section 6·3.

Treatment arm and selected baseline characteristics (see Section 5·2) of subjects providing an adolescent outcome measure at the week 6 visit and those with missing data will be compared using a logistic regression model. Similarly, separate logistic regression models will be used to investigate patterns of failure to provide outcome measures at the later follow-up weeks.
These analyses will be used to develop an understanding of the missing data mechanism. These models will be repeated for the primary parent MFQ measure. The reasons for end of treatment and study discontinuation will be tabulated by treatment arm.

Table 2 Summary of Baseline and Follow-up Assessments

	Weeks						Type
	0	6	12	36	52	86	
Adolescent Self report							
MFQ: Mood and Feelings Questionnaire	✓	✓	✓	✓	✓	✓	P
RCMAS: Revised Children's Manifest Anxiety Scale	✓	✓	✓	✓	✓	✓	S
LOI: Leyton Obsessional Inventory	✓	✓	✓	✓	✓	✓	S
Behaviours Checklist	✓	✓	✓	✓	✓	✓	S
RSES Rosenberg's Self Esteem Scale	✓	✓	✓	✓	✓	✓	
DEQ: Depressive Experiences Questionnaire ***(two subscales)***	✓				✓	✓	
DES-IV: Differential Emotion Scale-IV	✓		✓		✓		
DSC: Depressed States Checklist ***(two subscales)***	✓		✓		✓		
NEO-FFI: NEO-Five Factor Inventory ***(five subscales)***	✓				✓	✓	
DEEP	✓						
RTSHIA: Risk-Taking & Self-Harming Inventory for Adolescents ***(two subscales)***	✓	✓	✓	✓	✓	✓	
RRS: Ruminative Responses Scale	✓	✓	✓	✓	✓	✓	
EQ-5D: EuroQol measure of Health Related Quality of Life	✓	✓	✓	✓	✓	✓	
Interviewer Completed Measures							
K-SADS-PL: Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime	✓	✓	✓	✓	✓	✓	
ZAN:BPD: Zanerini Rating Scale for Borderline Personality Disorder	✓				✓		
C-SSRS: Classification Suicide Severity Rating Scale	✓	✓	✓	✓	✓	✓	
CA-SUS: Child and Adolescent Service Use Schedule	✓	✓	✓	✓	✓	✓	
HoNOSCA: Health of the Nation Outcome Scale for Children and Adolescents	✓	✓	✓	✓	✓	✓	S
EECI*: Expectations and Experience of Therapy Interview	✓			✓		✓	
Clinician Completed							
CGI: Clinical Global Impressions Scale	✓	✓	✓	✓	✓	✓	
WAI-S[×]*: Working Alliance Inventory-Short*		✓	✓	✓			
Family							
FAD: Family Assessment Device	✓		✓		✓		
APQ: Alabama Parenting Questionnaire	✓		✓		✓		
Life Events Questionnaire	✓		✓		✓		
Friendships Questionnaire	✓		✓		✓		
Parent self report							
SCL:90: Symptom Checklist 90 (global severity index)	✓	✓	✓	✓	✓		

Week 0 refers to baseline i.e. prior to randomisation

Type: P=Primary; S=Secondary outcome measure

Due to rationalisation of some of the scales there may be insufficient data for formal analyses. However, summary statistics will be provided for each scale.
**Analysed as part of IMPACT-ME substudy (at weeks 36 and 86 follow-up only for London participants)*
[×]Completed by adolescent and parent at same time points.

7·4 Quality Control of Measures

Observer Reliability between and within research sites

Intra and inter observer reliability will be considered using graphical methods and relevant summary statistics including intra-class correlation coefficients and kappa coefficients.

8 Statistical analysis of outcome comparing treatments

Extensive data cleaning of outcome and baseline data will be conducted without the treatment group allocations attached to the dataset. Results of these preliminary analyses will be reviewed by the trial research team to identify data errors and carry out preliminary checks regarding distributional assumptions prior to linking the treatment allocation to the follow-up data.

The analyses comparing treatments will be conducted applying the principle of intention to treat (ITT). No interim analyses of outcome data will be carried out unless specifically requested by the trial data monitoring and ethics committee.

8·1 Statistical inference between treatments

Within the protocol we considered both superiority and non-inferiority as potentially relevant hypotheses. The following hypotheses are stated in the protocol:

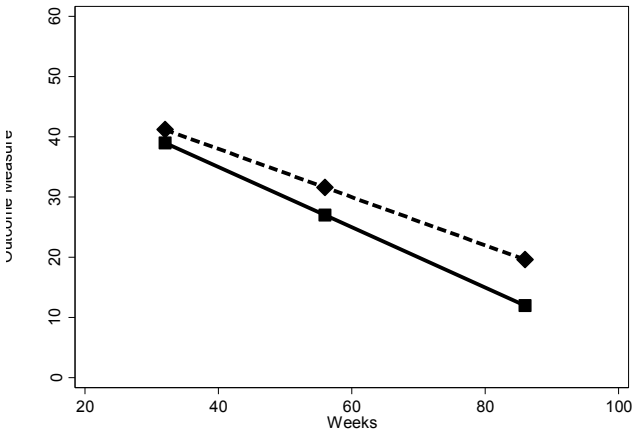
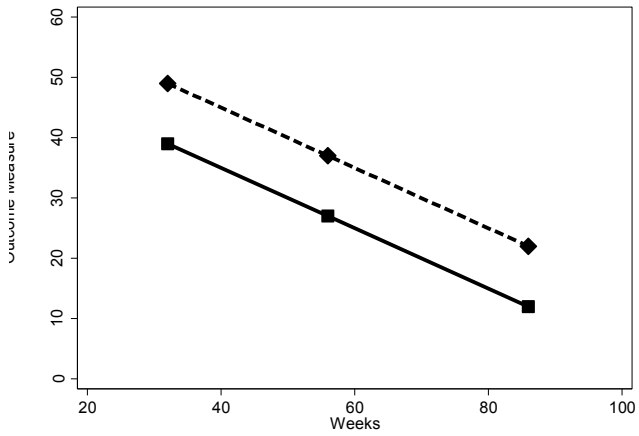
i) CBT will show superiority effects compared to SCC in the primary outcomes at 52 and 86 weeks
ii) STPP will show superiority effects compared to SCC in the primary outcomes at 52 and 86 weeks
iii) CBT will show non inferiority effects to STPP at 52 weeks
iv) STPP will show superiority effects compared to CBT at 86 weeks.

The hypotheses will be addressed using a linear mixed models analysis.

8·2 Treatment Effect Estimation for the Primary Outcome and other Continuous Outcome Measures

The intervention may influence outcome in two ways. Firstly, there may be a faster rate of recovery by 36 weeks and/or reduced clinically meaningful symptom recurrence between weeks 36 to 86 in one group than the other. Differential changes in symptoms over time can be estimated using a time with intervention group interaction. Secondly, there may be a systematic difference between intervention groups during follow-up, which is measured by a main effect. The statistical analysis of the primary outcome measure (MFQ) and the secondary continuous measures (see Table 2) will estimate the treatment effect using linear mixed effects models (LME, also know as random effects or random coefficient models). For all models, time (from randomisation considered as a continuous variable) will be centred based on the available data for the particular analysis being undertaken.

Because the aim of this study is to establish the longer term benefits of therapy we will consider only data over the post-treatment period for the primary analyses. All measures from week 36 onwards will be used for the statistical analyses for this purpose. By using data from week 36 onwards this should yield up to three measures per subject. This and the fact that time is continuous rather than discrete will reduce the potential for model identifiability problems given the number of random effects, time points and interaction terms.

Models with and without a time with treatment interaction	
	

Each model will adjust for baseline values of the outcome under consideration and the pre-specified prognostic variables as shown in Table 3:

Table 3 Fixed Baseline Covariates for the Each Outcome Measure

Primary: MFQ	SR baseline MFQ, plus RCMAS, LOI, and BC scores at baseline Other baseline*: arm, region, sex, age at randomisation in years, and use of SSRI at baseline
Secondary: RCMAS	SR baseline RCMAS, plus MFQ, LOI, and BC scores at baseline Other baseline: see MFQ outcome
Leyton Obsessional Inventory	SR baseline LOI, plus MFQ, RCMAS, and BC scores at baseline Other baseline: see MFQ outcome
Behaviours Checklist	SR baseline BC, plus MFQ, RCMAS, and LOI scores at baseline Other baseline: see MFQ outcome
HoNOSCA	Co-morbid behaviour disorders (i.e., a diagnosis of oppositional defiant disorder or conduct disorder) and all anxiety disorders combined+ Other baseline: see MFQ outcome

SR= Self-report

*These covariates will be used in all lme models

+ For these two disorders a binary variable will be created for absent (coded as 0) versus a diagnosis of "Yes" or a "high clinical index" (coded as 1).

Models will also include a subject level random intercept and correlated random coefficient for time. In addition, therapist will be included as a random effect subject to model fitting constraints.

First, a model with a time with intervention group interaction will be fitted. If there is a significant treatment by time interaction, inference for the interaction will be reported and separate adjusted treatment effects for the three pairs of treatments will be estimated for 52 and 86 weeks from the model. The hypothesis of non-inferiority of CBT relative to STPP at 52 weeks will be addressed by considering the 95% confidence interval of the treatment effect.

If the interaction between time and treatments is not significant, this term will be omitted from the model. Adjusted treatment effects will be estimated and tested using this simplified model. Non-inferiority will be considered using the 95% confidence interval of the treatment effect.

To assess the treatment effect while receiving therapy, random intercept LME models will be fitted to data prior to week 36 post randomisation, made up mostly of notional week 6 and 12 assessment data.

8·3 LME Inference and missing data

Of note, by using maximum likelihood for these models, "Missing At Random" is assumed for drop-out i.e., missing outcome data is conditional on observed data. Under this assumption it is assumed that future behaviour, given the past, is the same for all, whether a subject drops out or not. This allows distributional information to be "borrowed" from those who remain on the trial and applied to those who drop-out given they have the same covariate set up until the time of dropout. Therefore, the estimand of the treatment effect is what would be seen if all subjects had remained on the study until the end.

8·4 LME Model Diagnostics

Normal probability plots will be used to check distributional assumptions of the model for residuals of within and between subject variance terms. Where there is evidence of non-normality outcome data may be transformed.

8·5 Longitudinal Models for the parent MFQ and other Continuous Outcome Measures

The analysis of the parent data and the secondary outcome measures will be essentially the same as the primary analyses of the adolescent data.

8·6 Models for binary and ordinal outcome data

Binary data will be analysed using longitudinal logistic regression and ordered categorical secondary outcome measures such as the CGI scale will be analysed using an ordinal logistic regression model with random intercept and gradient terms on the log-odds scale.

9 Further Analyses

9·1 Adherence to therapy

Summary statistics on the number of trial therapy sessions attended by each subject will be tabulated by arm. A frequency distribution of number of sessions will also be presented by arm. In addition, the percentage of target total sessions will be summarised by arm.

Based on input from specialists, a binary variable for adherence is defined as follows for the three modalities:

STPP: Eight sessions is considered as the minimum therapeutic dose: thus adherence = 0 when seven or fewer sessions in total are completed, otherwise adherence = 1.

CBT: Six sessions is considered as the minimum therapeutic dose: thus adherence = 0 when five or fewer sessions are completed, otherwise adherence = 1.

SCC: Three sessions is considered as the minimum therapeutic dose: thus adherence = 0 when two or fewer sessions are completed, otherwise adherence = 1.

Adherence will be summarised by arm and this may be used in a secondary causal analysis of treatment effects which will be investigated separately from the main statistical analysis following a proposed discussion of causal pathways. Analyses will estimate the propensity to receive/adhere to treatment, accounting for SSRI usage at baseline as a dichotomous variable.

9·2 Moderator Analyses

The following are the pre-specified moderators of treatment which will be investigated, one at a time, to determine whether they interact statistically with therapy group based on MFQ outcome data over the short term (i.e. >0 and < 36 weeks post randomisation) analysed using linear random intercept models:

- MFQ score at baseline
- Age at randomisation
- Sex
- Region
- SSRI prescribing at baseline

Based on the MFQ outcome data from week 36 onwards the same treatment interactions will be tested using random effects (intercept and slope) models.

In addition, an interaction between MFQ score and SSRI usage at baseline will be examined.

10 Economic evaluation

10·1 Perspective

In the first instance the economic evaluation will take a service perspective, which will include the use of all hospital, community health, costs in addition to mainstream education and social services. Secondly, we will undertake analyses from a societal perspective, which in addition to the service costs will include the out of pocket costs of travel to treatment that fall to carers and any productivity losses for the study participant or their carer as a result of illness.

10·2 Calculation of total costs

For each piece of service use information collected in the CA-SUS, a unit cost (for example a cost per hour with a professional, a cost per inpatient night, a cost per unit of a drug) will be applied and the total costs calculated. The total cost per participant is calculated by summing all costs. All unit costs will be for the financial year 2012-2013. Costs between 52 and 86 weeks will be discounted at a rate of 3·5% because cost-effectiveness results should reflect the present value of costs and benefits, as recommended by the National Institute for Health and Clinical Excellence (NICE) 2013. Sensitivity analysis using rates of 1·5% will also be presented in additional analyses.

All NHS hospital contacts will be costed using NHS reference costs (Department of Health 2011). Unit costs of community health and social services will be taken from national publications (Curtis 2011) and education costs from government published statistics (https://www.gov.uk/government/statistics/la-and-school-expenditure-financial-year-2012-to-2013). Medications will be costed using information in the British National Formulary (British Medical Association and Royal Pharmaceutical Society of Great Britain 2010). Contacts with criminal justice sector services using available data from published sources (e.g. HM Prison Service 2009). Where necessary, costs will be inflated to 2012-2013 rates using the Hospital and Community Health Services inflation indices or the Retail Price inflation indices, as appropriate (Curtis 2011).

The cost of the CBT, STPP, and SCC interventions will be calculated on the basis of the salary of the therapist plus overhead expenses (administrative, managerial, and capital). Calculation of the indirect time, including preparation and supervision of therapists, will be based on information provided by the trial therapists on the ratio of direct face-to-face contact compared with other intervention-related activities using the bottom-up approach (Drummond et al. 2005) used in similar research (Byford et al. 2007) to generate a cost per hour with each study therapist and clinician. Sensitivity analyses will vary the assumptions used in generating the intervention unit costs to investigate the impact of low and high cost estimates on the results of the study.

Productivity losses will be calculated for the adolescent (if they are in employment) and the parent or carer using the human capital approach, which involves multiplying the individual's salary by reported days off work due to illness.

10·3 Calculation of QALYs

QALYs will be calculated on the basis of the EuroQol EQ-5D health state classification instrument which has five domains: mobility; self-care; usual activities; pain/discomfort; and anxiety/depression. For each domain the respondent chooses one of three levels of functioning, good to poor. The three levels for each of the five domains are used to define 243 health states (Glick et al. 2007). The health states will then be given a utility score using responses from a representative sample of adults in the UK (Dolan et al. 1995). QALYs in the second year will be discounted at a rate of 3·5%, as recommended by NICE (National Institute for Health and Clinical Excellence 2013). QALYs will be calculated as the area under the curve as defined by the utility values at baseline, six, 12, 36, 52, and 86 weeks follow-up and it will be assumed that changes in utility score over time will follow a linear path (Richardson and Manca 2004).

10·4 Service use

Differences in the use of services between randomised groups will be compared descriptively. No statistical comparisons will be made.

10·5 Costs

Total cost per participant over follow-up will be calculated and analysed for both a service and a societal perspective. Although costs are not expected to be normally distributed, analysis will compare mean costs using standard t-tests/analysis of co-variance with covariates as described in Section 8·2. The robustness of the parametric tests will be confirmed using non-parametric, bias-corrected bootstrapping (Barber & Thompson, 2000). The following comparisons will be made:

1. CBT v SCC at 52 and 86 weeks
2. STPP v SCC at 52 and 86 weeks
3. CBT v STPP v SCC at 52 and 86 weeks

10·6 Cost-utility analysis

Cost-utility analysis will be undertaken using quality adjusted life years (QALYs) calculated from the EQ-5D as the measure of effect. Cost-utility will be assessed through the calculation of incremental cost-effectiveness ratios (ICERs) –the ratio of the additional cost of one intervention compared with another over the additional effects of one intervention over another. The following primary cost-utility analyses will be carried out using a service perspective:

1. CBT vs SCC at 86 weeks
2. STPP vs SCC at 86 weeks
3. A three-way analysis which will involve pair-wise comparisons between CBT, STPP, and SCC and a three-way comparison at the 86-week follow-up. When more than two strategies are compared, ICERs are calculated using rules of dominance and extended dominance (Johannesson & Weinstein, 1993). Strategies will be ranked by cost, from the least expensive to the most expensive, and if a strategy is more expensive and less effective than the previous strategy, it is said to be dominated and is excluded from the calculation of ICERs. This process compares strategies in terms of observed differences in costs and effects, regardless of the statistical significance of the difference.

In addition, a secondary analysis will make the same comparisons using a societal perspective and also using data from the 52 weeks follow-up.

Uncertainty around the costs and effectiveness estimates will be represented by cost-effectiveness acceptability curves, which will be calculated using the net benefit approach (Briggs, 2001). Net benefits for the sample using values for λ (willingness to pay for an additional QALY) ranging from £0 to a maximum value of £50,000 will be calculated. A bootstrap replication of 5000 means for each net benefit estimate will be created, adjusted for baseline covariates outlined in Section 8·2.

The proportion of these replications that are greater than zero will indicate the probability that the intervention is cost-effective for each value of λ. Plotting these probabilities on a graph creates a cost-effectiveness acceptability curve, which depicts graphically the probability that the estimated cost-effectiveness ratio falls below the specified willingness to pay values (Van Hout et al., 1994).

10·7 Sensitivity analyses

A number of one-way sensitivity analyses will be undertaken to test the robustness of the results to the assumptions made in the economic evaluation. These will include, but will not be limited to:

- Variation of the cost of the interventions, dependent on seniority of therapists, time in direct contact with patients and other assumptions.
- Variation in the rate used for discounting of costs and outcomes in the second year to 1·5% as recommended by NICE (2013).

Appendix: Potential Additional Analyses

Mediator Analyses

Analysis of treatment mediators will depend on there being evidence of a treatment effect and will therefore be part of a later exploratory analysis. The effect of treatment on the mediators will be investigated separately from the main statistical analysis following the proposed discussion of causal pathways. The proposed mediators are:

STPP:
STPP involves reflective and dynamic processes directly with the patient focussing on potential underlying unconscious abnormalities stemming from experience dependent learning. Parent support is a key element in this therapy and we hypothesise that improvements in parent well-being will mediate the efficacy of STPP by 86 weeks. This will be expressed as:
Lower Global Severity Index scores of the SCL-90 over the course of treatment will be associated with better response to STPP revealed as lower self reported depression scores by 86 weeks.

CBT:
CBT involves a very clear focus on current abnormalities and distortions of thinking processes and their ruminative style that serves to maintain and potentially amplify the pathological cognitive reasoning about the self, the future, and the world. We hypothesise that self-reported ruminations about negative cognitions will mediate the efficacy of CBT by 86 weeks. This will be expressed as:
Lower self reported total rumination score over the course of treatment will be associated with a better response to CBT revealed as lower self reported depression scores by 86 weeks.

SCC:
SCC is a pragmatic treatment involving here and now advice and support to aid understanding of illness and remedy clear-cut maladaptive behaviours in the environment such as social withdrawal and solitariness. We hypothesise that reducing solitariness and increasing behavioural sociability through a focus on well-being will mediate the efficacy of SCC by 86 weeks. This will be expressed as:
High friendship scores over the course of the treatment will be associated with a better response to SCC revealed as lower self-report depression scores by 86 weeks.

References

Barber, J. A. & Thompson, S. G. (2000). Analysis of cost data in randomised trials: an application of the non-parametric bootstrap. Statistics in Medicine, vol. 19, pp. 3219-3236.

Briggs, A. (2001). A Bayesian approach to stochastic cost-effectiveness analysis. International Journal of Technology Assessment in Health Care, vol. 17, pp. 69-82.

British Medical Association & Royal Pharmaceutical Society of Great Britain (2010). BNF 59 BMJ Books/Pharmaceutical Press, London.

Byford, S., Barrett, B., Roberts, C., Wilkinson, P., Dubicka, B., Kelvin, R. G., White, L., Ford, C., Breen, S., & Goodyer, I. (2007). Cost-effectiveness of selective serotonin reuptake inhibitors and routine specialist care with and without cognitive behavioural therapy in adolescents with major depression. Br J Psychiatry, vol. 191, no. 6, pp. 521-527.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, New Jersey.: Lawrence Erlbaum Associates, Inc.,.

Curtis, L. (2011). Unit costs of health and social care 2010 PSSRU, University of Kent, Canterbury.

Department for Education and Skills (2002). LEA and School Information Service, The Stationery Office, London, LEA Level information - Standard Report 06.

Department of Health (2011). Reference costs 2010, Department of Health, London.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1995). A social tariff for EuroQoL: results from a UK general population survey, University of York: Centre for Health Economics, York.

Drummond, M., Sculpher, M., Torrance, G. L., O'Brien, B. J., & Stoddart, G. L. (2005). Methods for the economic evaluation of health care programmes, Third Edition edn, Oxford University Press, Oxford.

Glick, H. A., Briggs, A., & Polsky, D. (2007). Economic evaluation in clinical trials Oxford University Press, Oxford.

HM Prison Service (2009). Prison Service Annual Report and Accounts 2008, TSO, London.

Johannesson, M. & Weinstein, M. C. (1993). On the decision rules of cost-effectiveness analysis. Journal of Health Economics, vol. 12, no. 4, pp. 459-467.

National Institute for Health and Clinical Excellence (2013). Guide to the methods of technology appraisal, NICE, London.

Richardson, G. & Manca, A. (2004). Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. Health Economics, vol. 13, pp. 1203-1210.

Van Hout, B. A., Al, M. J., Gordon, G. S., & Rutten, F. H. (1994). Costs, effects and cost-effectiveness ratios alongside a clinical trial. Health Economics, vol. 3, pp. 309-319.

White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Statistics in Medicine 2005; 24: 993-1007

Addendum to IMPACT SAP (31st March 2016)

Section 1 and throughout document Replace specialist clinical care (SCC) with the term brief psychological intervention (BPI).
Section 6.2 The SSRI covariate used for imputation and analyses was "SSRI prescribed before trial entry" (where if missing information then it was assumed not to be prescribed) and not "SSRI use at baseline".
Section 6.3 The time of assessment caused some duplication in the slotting procedure where if more than one measurement was assigned to a band then only the nearest to the week from randomisation was to be used for the summary statistics. As the summary statistics from this approach were difficult to interpret we adopted the standard approach of reporting by researcher assessment instead.
Section 7.2 Only the primary and secondary outcome measurement scales were summarised.
Behaviours checklist (BC) is also known as the Antisocial Behaviours questionnaire (ABQ). Since the distribution of ABQ total score was highly skewed with the standard deviation larger than the mean at many time-points and medians of zero at weeks 52 and 86 for each group we considered ABQ as a binary outcome coded as one if the ABQ score was ≥1.
Section 8.2 Because MFQ and RCMAS are correlated the latter was omitted where both were originally listed as baseline covariates in an outcome model (except for the RCMAS outcome where MFQ was dropped). The covariates in Table 3 of the SAP were replaced by the ones shown in Table 1 of this Addendum with ABQ entered on the 3 point scale.
The SAP states:
"First, a model with a time with intervention group interaction will be fitted. If there is a significant treatment by time interaction, inference for the interaction will be reported and separate adjusted treatment effects for the three pairs of treatments will be estimated for 52 and 86 weeks from the model. The hypothesis of non-inferiority of CBT relative to STPP at 52 weeks will be addressed by considering the 95% confidence interval of the treatment effect."

Based on these interaction models the marginal effect of treatment was estimated at 36, 52 weeks and 86 weeks post randomisation for the following two comparisons rather than three to match the protocol hypotheses:

(i) STPP against CBT and
(ii) (CBT and STPP) against BPI

Note, the sample size calculation used a significance level of 2.5% to allow for this multiplicity. A Bonferroni correction was not applied to the p-values, but it is suggested that readers use a 2.5% significance level to maintain the family-wise 5% significance level at a particular point of assessment.
Section 9.1 Adherence is now redefined as therapeutic dose. For STPP this was changed from ≥8 to ≥6 sessions based on consultation with experts in this therapy field.

Section 9.2 Following detailed discussion between the PI's the original list of moderators were replaced by the following as they were deemed relevant:
Hypotheses for the DEQ at baseline:
1) Elevated relatedness/dependent scores will be associated with a relatively better response in the STPP group compared to BPI or CBT groups.
2) Elevated self-critical/identity scores will be associated with a relatively better response in the CBT group compared to BPI or STPP groups.

Hypotheses for the RRS at baseline
1) Higher scores will show a better treatment response in the CBT compared to the BPI and STPP arms.
Additional Analyses not specified in the SAP
In order to gain a better understanding of patterns over time in diagnosis, medication prescription and adverse events the following were undertaken. The results are presented in the HTA report.

1. We investigated change over time using GEE longitudinal analyses for Unipolar major depressive disorder (MDD) and MFQ total score >25 outcomes. The analysis of ABQ was changed from an mixed model to GEE since the data was not normally distributed.
2. Summaries on SSRI prescription prior to trial entry and also during follow-up overall and also split by <36 weeks and ≥36 weeks post randomisation to match the two analyses time periods were provided.
3. Adverse event reporting.

Table 1 Fixed covariates for each model

Measure	Type of Measure	Data Collection Method	Fixed Covariates
Primary			
MFQ	Continuous	SR	Baseline MFQ, LOI, ABQ scores, treatment allocation, region, sex, age at randomisation, co-morbid behaviour disorder+ , prescription of SSRI before trial entry.
Secondary			
RCMAS	Continuous	SR	Baseline RCMAS, LOI, ABQ scores, treatment allocation region, sex, age at randomisation, co-morbid behaviour disorder+ , prescription of SSRI before trial entry.
LOI	Continuous	SR	Baseline LOI, MFQ, ABQ scores, treatment allocation region, sex, age at randomisation, co-morbid behaviour disorder+ , prescription of SSRI before trial entry.
HoNOSCA	Continuous	IR	Baseline HoNOSCA ,MFQ, LOI and ABQ scores, treatment allocation, region, sex, age at randomisation, co-morbid behaviour disorder+ , prescription of SSRI before trial entry
ABQ	Binary	SR	Baseline ABQ, plus MFQ score, treatment allocation, region, sex, age at randomisation, co-morbid behaviour disorder+ , prescription of SSRI before trial entry.
K-SADS MDD	Binary	IR	See MFQ outcome above
MFQ >25	Binary	SR	See MFQ outcome above

SR= Self-report, IR=interviewer rated

+ co-morbid behaviour disorder i.e., a diagnosis of oppositional defiant disorder or conduct disorder. Note, this was added as a binary variable at the analysis stage since it was found to be significantly predictive of missing data