IY Pooling Project

Statistical Analysis Plan

Version 1

Version 1 started: 14/01/2015

Contents

1. Description of the trials

1.1 Principal research objectives to be addressed

All research objectives addressed in this analysis plan relate to modification of the effect of the Incredible Years parenting intervention on children's disruptive behaviour.

Primary objectives

Treatment effect modification by social economic status:

- To determine whether the effectiveness of IY intervention for reducing children's disruptive behaviour differs across social economic status, as represented by five indicator variables. The indicator variables of interest are whether a family has or is at risk of low income (as defined by various measures including being in receipt of means tested benefits), education level of the primary parent, unemployment (defined by whether there is no employed individual in the household), whether the primary parent is a lone parent (defined as not living with a partner) and whether the primary parent was a teen parent (defined as aged less than twenty at the birth of the target child).

Secondary objectives

Further IY effect modification by baseline variables:

- To determine whether the effectiveness of the IY intervention varies with level of baseline symptom severity. In particular whether there is an increased reduction of children's disruptive behaviour from baseline to post-intervention in response to the IY intervention for those with higher levels of disruptive behaviour at baseline.
- To determine whether the effectiveness of the IY intervention differs between those from an ethnic minority and those not from an ethnic minority.
- To determine whether IY is equally effective at reducing disruptive child behaviour for families with co-morbid ADHD problems, as for families with no co-morbid ADHD problems.
- To determine whether IY is equally effective at reducing disruptive child behaviour for families with co-morbid child emotional problems, as for families with no co-morbid child emotional problems.
- To assess whether the IY is equally effective at reducing disruptive child behaviour for different levels of primary parent depressive symptomatology

- To determine whether IY is equally effective at reducing disruptive child behaviour across families with children of different ages.
- To determine whether IY is equally effective at reducing disruptive child behaviour across families with children of different genders.
- Is IY equally effective at reducing disruptive child behaviour across families with different levels of harsh and inconsistent parenting at baseline?
- Is IY equally effective at reducing disruptive child behaviour across families with different levels of positive parenting at baseline?
- To determine whether the effectiveness of the IY intervention at reducing child disruptive behaviour differs across geographical regions at a trial level, in particular between UK and non-UK trials and between primarily urban and primarily rural trial
- To determine whether the effectiveness of the IY intervention differs across different types of service provider. Is IY equally effective in the NHS and similar clinical settings, as in non-clinical settings?
- Is IY equally effective when delivered when more staff is IY certified than when fewer staff is IY certified?
- Is IY equally effective when delivered when more staff is clinically trained than when fewer staff is IY clinically trained?
- To determine whether there is a difference in the effectiveness of the IY intervention across controlled efficacy settings versus non-controlled effectiveness settings.
- To determine whether the effectiveness of the IY intervention differs by number of IY sessions offered in the trial.

IY effect modification by post-treatment variables

- To determine whether there is a dose response effect within the intervention arm, in particular whether the intervention is more effective in families where parents attended a higher proportion of the sessions offered.
- Are effects of IY on the reduction of disruptive child behaviour stronger in families of which 2 parents participated in IY, instead of 1 parent?
- Is IY equally effective when staff in trial received regular supervision?

Possibility of confounding

- To assess whether any detected effect modification by a variable can be attributed to another observed moderator variable.

Higher order IY effect modification:

- To determine whether the IY effect modification by social economic status varies with levels of baseline symptom severity as suggested by Leijten et al. (2013), which found in a meta-analysis that disadvantaged samples benefited less from IY, but only when they had low levels of initial problem severity. When initial problems were severe,

disadvantaged and non-disadvantaged families benefited equally, but when initial problems were mild, disadvantaged families benefited less.

1.2 Trials included in pooling study

1.2.1 Overview of trials

Criteria for including a trial in this pooling study were that they were randomised controlled trials of the IY intervention targeted at reducing conduct problems in children. The primary outcome must be child conduct problems as measured by either Eyberg Child Behavior Inventory (ECBI) or Parent Account of Child Symptoms (PACS) and this measure must have been taken at between 0 and 2 months after the end of treatment.

Table 1 shows a list of the trials included in the pooled dataset and their design features. Across the trials some offered the IY intervention only in the treated group, whilst others offered a literacy intervention alongside IY. Similarly the control condition differs across trials, with some offering waitlist and treatment as usual and others offering only minimal treatment. Since this study is focused upon the IY training programme participants randomised to a literacy only intervention were excluded from the pooled dataset. Table 1 shows the location where the trial was carried out. The column labelled N gives the total sample size from the trial used in the IY pooling dataset. The active group in most trials is the IY intervention, although some used a literacy intervention in addition to IY. The column control group details the type of control condition and the column arms used details the arms included in the pooled sample, although the trial may have included arms that have not been used in the pooled sample. In the control type care as usual/no care refers to the fact that no support or services were provided in the control arm other that what was normally accessible to the patient during their daily life (in particular in NLBS where mothers recently released from incarceration may not have access to the services that would normally be available). Minimal intervention means that some non-intensive intervention was provided to parents in the control arm, e.g. a telephone helpline. Parents in the waitlist control condition were crossed over to the intervention after 6 months. A full list of the trials included in the pooled sample and corresponding references is given in section 5.1.

Trials included are:

- Nor: a treatment of oppositional defiant and conduct problems in young Norwegian children (4 to 8 years).
- Swed: an evaluation of the incredible years programme in Sweden.
- Port: a trial of middle class families in Portugal.
- Irel: a trial involving disadvantaged communities in Ireland.
- NLBS: Netherlands (better start), a trial involving mothers being released from incarceration.

- NLSES: a trial of low socio-economic status families in the Netherlands.
- WlsSS: Wales sure start. An intervention for children at risk of developing conduct disorder.
- Birm: a Birmingham run trial of pre-school children.
- SPOKES: London trial investigating child antisocial behaviour using both the IY intervention and a literacy intervention.
- PALS: London trial involving high risk families using both the IY intervention and a literacy intervention.
- HCA: London trial of the IY intervention and literacy intervention, which compared both the IY intervention alone and the IY+literacy intervention.
- Oxfrd: Oxford based trial investigating mechanisms of change for conduct problems in children.
- VTST:

Table 1: List of trials included in the study and design features.

Number	Trial Acronym	Location	N	Active Group	Control Group	Duration randomisation to end of intervention (months)	Average Number of IY Sessions Offered	Number of IY sessions was changed	Booster sessions offered	Arms used
1	Nor	Norway	75	Incredible Years	Waitlist	5	12.09	Yes	0	2
2	Swed	Sweden	62	Incredible Years	Waitlist	5	13	No	0	2
3	Port	Portugal	124	Incredible Years	Waitlist	5	14	No	1	2
4	Irel	Ireland	149	Incredible Years	Waitlist	5	13.28	Yes	0	2
5	NLBS	Netherlands	99	Incredible Years	Care as usual/ no care	5	12	No	4	2
6	NLSES	Netherlands	156	Incredible Years	Waitlist	5	14.46	Yes	0	2
7	WlsSS	Wales	153	Incredible Years	Waitlist	5	12	No	0	2
9	Birm	England	161	Incredible Years	Waitlist	5	12	No	0	2
10	SPOKES	London	112	Incredible Years + Literacy	care as usual/ no care	8	12	No	0	2
11	PALS	London	174	Incredible Years + Literacy	minimal intervention	8	12	No	0	2
12	HCA	London	214	Incredible	minimal intervent	5--8	12	No	0	3

				Years; Incredible Years + Literacy	ion					
13	Oxfrd	England	76	Incredible Years	Waitlist	5	14	No	0	2
14	VTST	London	141	Incredible Years	Waitlist	5	14.11	Yes	0	2

1.2.2 Further trial design features (cluster and stratified randomisation)

Table 2 contains details of the included trials, including follow-up time points, randomisation type, planned randomisation ratio, whether the randomisation ratio was changed and which stratifiers were used in the randomisation. In addition the IY therapy was delivered as a group intervention, which induces a further clustering effect.

Table 2: Features of trial design including time at which measures were recorded, randomisation type, randomisation ratio by design and whether the randomisation ratio changed over the course of the trial. For some trials there was no second follow-up.

Number	Trial Acronym	Duration randomisation to first assessment (months)	Duration end of intervention to second assessment (months)	Duration end of intervention to third assessment (months)	Randomisation unit	Stratified randomisation	Stratifiers used in randomisation	Randomisation ratio by trial design	Variable randomisation ratio
1	Nor	1	0—2	na	Individual	Yes	child age, child gender, child scored > 97th percentile on ECBI intensity scale, site	1:1	Yes
2	Swed	1	0—2	na	Individual	Yes	site	2:1	Yes
3	Port	1	0—2	na	Individual	Yes	child age, child gender	1:1	Yes
4	Irel	1	0—2	na	Individual	Yes	site	2:1	No
5	NLBS	1	0—2	4	Individual	No	none	2:1	Yes
6	NLSES	1	0—2	na	Individual	No	none	2:1	No
7	WlsSS	1	0—2	na	Individual	Yes	child age, child gender, site	2:1	No
9	Birm	1	0—2	na	Individual	Yes	child age, child gender, children's centre attachment	2:1	No
10	SPOKES	1	0—2	na	Individual	Yes	school-year (10 strata formed	1:1	No

							from 8 schools over 3 years)		
11	PALS	1	0—2	na	cluster - classrooms	No	none	1:1	Yes
12	HCA	1	0—2	5--7	Individual	Yes	recruitment cohort	1:1:1	Yes
13	Oxfrd	1	0—2	na	Individual	No	none	1:1	Yes
14	VTST	1	0—2	na	cluster - time period	No	none	2:1	Yes

1.3 Variables

1.3.1 Trial design variables

Table 2 shows a number of design features by trial, in particular the timings at which measures were taken and the type of randomisation. Several of the trials used stratified randomisation and one trial used a cluster randomised design. The table also shows the randomisation ratio included in the trial design, however some of the trials varied the randomisation ratio during the course of the study. The primary reason for this is that the IY intervention is a group treatment and so the allocation ratio was adjusted in order to construct sufficiently large group sizes. The table records for each trial whether the randomisation ratio differed across (identifiable) subsets of participants.

A number of variables are included in the dataset in order to describe the type of trial design. These are not variables whose effect on outcome is of interest but denote those variables that may have to be conditioned on in the final analysis and so are included for the purpose of describing the dataset. These variables are:

- Trial ID: a variable denoting which trial an observation is taken from.
- Unique family ID: a unique variable (both within and between trials) denoting the family number for each observation.
- Treatment condition: whether the family was randomised to receive the active treatment or the control.
- IY only or IY plus reading: type of active intervention used. Only active arms which include IY only or the IY+literacy intervention are included in the pooled dataset.
- Control type: type of control used in the trial. This variable has 4 levels: waitlist, minimal intervention, no care or care as usual. This variable is constant within trials but varies between trials.
- IY sessions offered: The number of sessions of IY offered to the participant. The number of sessions offered varies between trials, as different manuals were used, and in some cases this also varies within a trial as the IY programme was changed during the course of the trial.
- Boosters offered after post: In some of the trials booster sessions were offered post treatment. This variable is coded as missing by design if no booster sessions were offered.
- Randomisation ratio applied to each participant: for variable randomisation ratios a variable denoting the batch of randomisations the participant was in.

- Clusters used for randomisation within trial: if clustered randomisation was used within a trial (trials 11 and 14) this variable denotes the cluster a participant belonged to.
- Stratification variables used within trial: if stratified randomisation was used this variable denotes which variables were used in the stratification. This is generally a set of variables applied to stratify all trial participants, the exception being Norway where participants were first randomised to receive active and control treatment and then within the active arm stratified randomisation was used to assign participants to IY or literacy.
    - Trial site: if randomisation was stratified by site (for trials 2 and 4) variable denotes which site a participant belonged to within the trial.
    - School year (for stratification trial 10): school year of the target child, used as a stratifier in trial 10.
    - Recruitment cohort (for stratification trial 12): trial 12 randomised within batches and so the recruitment cohort is included for this trial.
    - ECBI 97th percentile stratifier (for stratification trial 1): whether or not the target child is in the 97th percentile at baseline on the ECBI scale. Used as a stratifier in trial 1.
    - Stratification categories used for child age: where age of the target child was categorised for use in stratification (trials 1, 3 and 7) the categories of age have been included in this variable.

Table 3 shows the trial design variables and which trials information on the variable is available and for which the data is missing. Family ID, treatment condition, type of active treatment and type of control are available for all trials. The randomisation ratio batch is only relevant for those trials in which the randomisation ratio was changed. However, this information was not recorded in one of those trials (trial 3) and thus this feature cannot be accounted for in the analysis. Stratification variables used in different trials include trial site, school year, recruitment cohort, ECBI percentile and child age. For the stratification variables the numbers of the trials in which they were used are noted in the column titled applicable trials and the missing trials column indicates trials in which those variables were used as stratifiers but data is not available on the. Two trials used a cluster randomised design and data on cluster membership is available for one of these trials. The number of IY sessions offered and attended and the IY group are variables that are only relevant to those in the treatment condition.

Table 3: Information available on trial design variable.

Variable Name	**Applicable to trials**	**Relevant information missing for trials**
Trial ID	All	
Unique family ID	All	
Treatment condition	All	

IY only or IY plus reading	All	
Control type	All	
Randomisation ratio which applied to each participant	1,2,3,5,13	3
Number of IY sessions offered	All	
Number of booster sessions offered	3, 5	
Clusters used for randomisation	11, 14	
Stratification variables used:	1,2,3,4,7,9,10,12	
- Trial site	2,4	4
- Recruitment cohort (for stratification trial 12)	12	
- ECBI 97th percentile stratifier (for stratification trial 1)	1	
- Stratification categories used for child age	1,3,7	

1.3.2 Baseline Measures

A number of demographic and clinical variables were measured within the trials at baseline (before randomisation) for the purpose of describing the population.

At the individual level these are:

- Child gender: a binary coded variable denoting whether the target child is male or female.
- Child age: age in months of the target child at baseline.
- Primary parent gender: a binary variable coding whether the primary parent is male or female.
- Primary parent age: age in years of the primary parent at baseline.
- Primary parent age at birth of target child: age in years of the primary parent when the target child is born.
- Second parent gender: a binary coded variable denoting whether the second parent is male or female. Coded as not applicable if the primary parent is a lone parent.
- Second parent age: age in years of the second parent, coded as not applicable if the primary parent is a lone parent.
- Child was referred: binary variable denoting whether or not the child was referred to the trial for behaviour problems.
- Low income: binary variable denoting whether the family has or is at risk from low income.

- Education level: highest level of education attained by the primary parent.
- Lone parent: binary coded variable denoting whether or not the primary parent is living without a partner.
- Teen parent: binary coded variable denoting whether the primary parent was under the age of twenty at the birth of the target child.
- Primary parent unemployed: binary coded variable denoting whether the primary parent is unemployed.
- SES Unemployed: binary variable denoting whether or not there is no employed parent in the household.
- SES occupation: highest occupation level in the household.
- SES benefits: binary variable denoting whether or not the family is in receipt of benefits. The interpretation of this variable may vary depending on the country in which the trial was conducted. In particular some benefits are not means tested in certain countries.
- Ethnic minority: binary coded variable denoting whether or not the primary parent is non-white.
- Ethnic background: more detailed variable giving the ethnicity of the primary parent.
- Ethnic country: binary variable denoting whether or not the primary parent was born outside the country of residence.
- ADHD Comorbidity
- Emotional problems comorbidity
- Parental depression
- Positive parenting
- Negative parenting

A number of variables were also recorded at a trial level at baseline. These are baseline characteristics that vary between trials but not within a trial. Trial level baseline variables include:

- Geographical region
    - UK versus non-UK: whether or not the trial was conducted in the UK. There are 7 UK trials from England and Wales within the pooled dataset and 6 from Ireland and other European countries.
    - Urban versus rural: whether the trial was carried out in a mostly urban or mostly rural setting.
- Service provider: variable denoting the type of service provider organisation, i.e. clinical versus non-clinical settings.
- Efficacy or effectiveness setting: level of control within the trial of the efficacy versus effectiveness. An efficacy trial is conducted in a controlled setting, whereas an effectiveness trial is conducted in a “real world” setting.
- Number of IY sessions offered by trial design.
- Number of booster sessions offered by trial design.
- % staff certified: Percentage of staff at a trial level that were certified.
- % staff clinically trained: Percentage of staff at a trial level that was clinically trained.

- Supervision: Variable denoting whether staff in the trial received supervision.

Table 4 shows a list of demographic variables included in the pooled dataset, along with the available sample size for each of these variables. The table also shows the proportion of the applicable data that is available (discounting observations where the value is recorded as not applicable, e.g. variables pertaining to the second parent where the primary parent is a lone parent). The amount of missing data varies substantially across these trials. Whilst some of this data is missing, there are also variables which may be not applicable for certain participants, in particular variables pertaining to the second parent where the primary parent is a lone parent.

Table 4: List of available clinical and demographic variables at baseline.

Variable Name	**N**	**Proportion of applicable data available (%)**	**Relevant information missing for trials**	**Type**
Child gender	1696	100		Binary
Child age	1682	99.1		Continuous (in months)
Primary parent gender	1674	98.7		Binary
Primary parent age	1539	90.7	2	Continuous (in years)
Primary parent age at birth of target child	1527	90.0	2	Continuous (in years)
Second parent gender	854	74.4	2,5,7	Binary
Second parent age	746	65.1	2, 5, 7	Continuous (in years)
Child was referred	1156	68.2	4, 7, 9, 13	Binary
Low income	1614	95.2		Binary
Education level	1573	92.7		Ordinal
Lone parent	1606	94.7		Binary
Teen parent	1550	91.4	2	Binary
Primary parent unemployed	1127	66.5	6, 7, 9	Binary
SES	1303	76.8	6, 7	Binary

Unemployed				
SES occupation	430	25.4	4,5,6,7,9,11	Ordinal
SES benefits	1310	77.2	2, 5 ,6	Binary
Ethnic minority	1611	95.0		Binary
Ethnic background	1602	94.5		Binary
Ethnic country	650	38.3	4, 7, 11, 12, 13, 14	Categorical
ADHD Comorbidity				
Emotional problems comorbidity				
Parental depression				
Positive parenting				
Negative parenting				
Trial level variables:				
Non UK or UK	14	All		
%_rural	14	All		
Service provider	14	All		
Efficacy or effectiveness	14	All		
% staff certified		All		
% staff clinically trained		All		
Supervision		All		

Commented [KCL1]: To fill this bit of table

1.3.3 Aspects of Parenting Training

Treatment Condition

A number of variables were measured in the active arm to describe the training programme received by the parents. These are:

- IY sessions attended: The number of IY sessions that at least one parent was present for.
- Booster attended after post: The number of booster sessions at least one parent was present for.
- Number of Parents in IY: Whether one or both parents were involved in the IY intervention.

- IY group: IY group is coded for the therapy group the participant was in, since the IY intervention is delivered as a group therapy.

Table 5 shows variables that are applicable in the treatment condition. The total number of participants in the treatment condition is 1046 across the pooled sample.

Table 5: List of variables in the treatment condition.

Variable Name	N	Relevant information missing for trials	Proportion of applicable data available (%)
IY sessions attended	855	2, 11	81.7
Booster attended after post	49	3	34.5
Number of Parents in IY	730	1, 7, 9	69.8
IY group	774	1, 2, 11	74.5

Control

650 participants in the pooled sample are included in the control condition. The variable control type denotes the control condition within the trial and is available for all participants in the control condition. This variable varies across trials but not between trials.

1.3.4 Child Outcome Measures

Table 6 shows the child outcome measures at baseline, post-treatment and 12 month follow up. There is a single outcome measure: the ECBI scale (harmonised from the PACs data in those trials where the PACs scale was used in place of ECBI). This is measured at up to three time points, which are:

- Baseline: all trials measured child conduct disorder at pre-treatment and this is available in both the treated and control arms, although there is some missing data.
- Time window 1: the first follow-up point for most of the trials. The time window for this measure is between 0 and 2 months post-treatment, which is defined as the end of the intervention. .

- Time window 2: The second follow-up time point, except in the time window for this measure was between 4 and 7 months after the end of the intervention. In NLBS and HCA this is available for both treated and controls, as the whole sample was followed up twice. In other trials, where there is no second follow-up, this available for only the active arm.

Table 6: Available child outcome measures at each time point.

Variable Name	N	Proportion of applicable data available (%)
Baseline ECBI	1555	70.3
ECBI in time window 1	1317	67.2
ECBI in time window 2	782	55.8

1.3.4.1 Child outcome data at post-treatment time of interest

The primary outcome of interest in the pooled sample is the first follow up time point, which was collected within time window 1. The time from randomisation to first follow-up varied within different trials due to variations in the length of the intervention (see Table 1), however this measure was taken within 0-2 months of the end of the intervention.

1.3.4.2 Child outcome data at further times

The child outcome at the second follow-up will not be used in the main analysis but may be useful as an auxiliary variable, particularly if the previous child measure is missing.

1.3.5 Putative Moderators

1.3.5.1 Individual Level Moderators

A number of individual-level moderators will be investigated. These are moderators that vary by individual within the study. The majority of these are pre-randomisation baseline clinical or demographic variables but there are in addition some post-randomisation variables that will be investigated (in particular treatment fidelity). The individual level moderators are:

At baseline (pre-randomisation):

- Socio-economic status (SES) indictors
    - Low income: whether or not the family is or is at risk from low income (see table 2).
    - Lone parent: primary parent is a lone parent (see table 2).
    - Teen parent: primary parent was aged less than twenty at birth of target child (see table 2).
    - SES unemployed: there is no employed individual in the household (see table 2)
    - Primary parent education level: highest education level attained by the primary parent (see table 2).
- Baseline child outcome: child conduct problems as measured by the ECBI or harmonised ECBI scale at baseline (see table 5).
- Child age at baseline: age in months of target child at baseline (see table 2).
- Child gender: gender of the target child (see table 2)
- Baseline parental depression: level of depression of the primary parent at baseline
- ADHD Co-morbidity: child's ADHD co-morbidities (to be added).
- Emotional problems co-morbidity: Level of child's co-morbid emotional problems.
- Ethnic minority: whether or not the primary parent is from an ethnic minority (see table 2).
- Baseline parental depression: Primary parent's depression level at baseline.
- Level of positive parenting
- Level of negative parenting

Post-randomisation

- Treatment adherence: proportion of sessions offered that were attended by at least one parent (see table 6).
- Number of parents in IY.

1.3.5.2 Trial Level Moderators

Trial level moderators are variables which may influence treatment effect that vary between trials but not within trials. These are:

At baseline (pre-randomisation):

- Geographical region
    - UK versus non-UK: whether or not the trial was conducted in the UK. There are 7 UK trials from England and Wales within the pooled dataset and 6 from Ireland and other European countries.
    - Urban versus rural: whether the trial was carried out in a mostly urban or mostly rural setting.

Post-randomisation:

- Staff supervision

1.3.6 Data Harmonisation

1.3.6.1 Complex definitions

SES low income is a derived variable that uses information on participants, such as whether the family receives means tested benefits.

1.3.6.2 Transformation of measures based on population norms

In a number of the trials the PACs measure was used for child outcome rather than the ECBI. These values have been transformed using population norms to the ECBI scale and where both the PACs and the ECBI data are available the harmonised ECBI is also calculated as a comparison.

1.4 Sample size estimation

Power calculations for the total sample size give >97% power for the interaction term when compared with the treatment and covariate main-effects-only model for a treatment arm difference in the covariate effect size on outcome of .15 SD (significance level 0.05).

Commented [HV2]: Taken from the initial proposal.

1.5 Brief description of proposed analyses

The objective of the analysis is to assess inequality in treatment effectiveness across five social economic status variables. In particular the primary goal is to understand whether the IY intervention is less effective for individuals of lower social economic status. The proposed social economic status variables are: low income, binary coded and based on a combined measure of variables that represent whether a family has or is at risk of low income; education level, coded on an ordinal scale from 1-7; lone parent, binary coded for if the primary parent lives alone; teen parent, binary coded for if the primary parent was younger than 20 at the birth of the target child and unemployment status, binary coded representing whether there is no employed parent in the household. Additionally ethnicity, which represents whether the primary parent is non-white, and baseline level of symptom severity at baseline will be analysed in order to assess whether they influence the effectiveness of the intervention. It is hypothesised that the intervention may be more effective for those children with higher baseline scores on the level of problems scale.

Commented [KCL3]: Revisit when have first draft of inferential analysis

Initially each of these seven variables will be considered individually as a moderator effect, i.e. as a two-way interaction with the treatment effect in the

analysis model. As an extension of this we will consider the combined effect of all the social economic status variables. Finally in order to assess whether any of the primary moderator effects are influenced by baseline differences in problem severity, we will explore three way interactions between treatment condition, moderator and symptom severity at baseline.

2. Data analysis plan – Data description

2.1 Describing the population

Descriptive statistics will be initially used to summarise baseline clinical and demographic variables listed under section 1.3.2. We will construct descriptive statistics for each of the baseline demographic variables, including baseline ECBI and potential moderators of treatment effects, both at a trial level and at the individual participant level. In general summaries will be provided by trial and for the pooled data set.

For binary coded demographic variables (e.g. child gender, treatment condition, SES low income) proportions within each trial and in the pooled sample will be displayed graphically using bar graphs, with bar height being given by percentages in each group. The distribution of continuous demographic variables, such as child age at the start of the study and parent age at the birth of the target child, will be displayed using box plots. A box plot is a display of the data distribution where the axis is the range of the data and the data is plotted as a box whose boundaries are the upper and lower quartiles of the data and is bisected by a line representing the median value. Lines either side of the box display the range of the data and statistical outliers, i.e. extreme values that lie outside the normal range of the data, are plotted as points. By displaying the demographic variables across trials graphically it will be possible to see the variability in the population across trials.

Commented [KCL4]: Please update your descriptive analyses to include any new variables since Xmas.

Trial level variables will be displayed in a similar way for the pooled dataset only (i.e. with percentage bar charts for binary coded variables and box plots for continuous variables).

2.2 Aspects of treatment

Aspects of treatment, such as the number of IY sessions attended (for a full list see Section 1.3.3), will be shown graphically. For the number of sessions offered this will be displayed as a bar chart of the mean in each trial, as there is little within trial variability. For the number of sessions attended this will be displayed as box plots. Number of parents in IY will be displayed in the same way as binary variables, as this can be dichotomised as one or two parents.

2.3 Treatment effects on child outcome

Cohen's d will be estimated as an unadjusted measure of treatment effect size. Cohen's d is a standardised measure of the difference in means between two groups. It is computed by taking the difference between the means in each trial arm and dividing by the pooled standard deviation, where the pooled standard deviation is estimated from the standard deviation within each group under the assumption that the standard deviation is the same across the population. A Cohen's d of 0.2 is typically considered a "small effect", 0.5 is considered "medium" and 0.8 or larger is considered a "large effect size". For the child outcome the change between baseline and the first follow up is calculated and the Cohen's d between treatment and control conditions will be computed for the change scores. The baseline measure occurs in all trials before treatment has occurred and the first follow-up in all trials is the measure taken in time window 1, which takes place between 0 and 2 months after the end of the intervention (see section 1.3.4). The Cohen's d for each trial and in the pooled sample will be displayed graphically along with a 95% confidence interval.

2.4 Treatment effect moderation

Moderation by individual level baseline variables

We will use descriptive statistics to provide some preliminary exploration of potential treatment-effect moderation by participant-level or trial-level variables listed in section 1.3.5 Using the pooled dataset for individual level binary coded moderators (e.g. SES low income, SES teen parent, SES lone parent, SES unemployed and ethnic minority) box plots of change in ECBI from baseline to post-treatment for both treatment and control conditions will be used to assess the difference in treatment effect for different levels of the moderator. For example for SES low income there will be four box plots, treatment and control conditions for low income families and treatment and control conditions for non-low income families. If the difference between median change scores differs between levels of the moderator then this may be indicative of a moderation effect, although inferential analysis will be required to determine whether this effect is statistically significant accounting for trial variability.

For individual level continuous variables (e.g. baseline values of ECBI) we will plot change in ECBI against the potential moderator for both treated and control conditions, including a smooth fitted line for interpretation. If the difference between treated and controls varies across values of the moderator then this may indicate a moderating effect, although as with the binary coded moderators further inferential analysis will be required to determine the significance of this effect.

Moderation by trial level baseline variables

For binary coded trial level variables we will plot the range of Cohen's d, as calculated as the standardised mean difference between the change score from baseline to post-treatment between treatment and control groups, as boxplots across levels of the putative moderator. For continuous trial level variables we will create scatter plots of Cohen's d against the putative moderator. These plots will demonstrate when the treatment effect, as estimated by Cohen's d, differs across values of the trial level moderator.

Moderation by post-randomisation variables

Post randomisation variables include number of parents attending IY, number of sessions at least one parent attended and at the trial level of supervision refer to the intervention arm only and so no comparison can be made between treated and controls. Instead we will plot change scores from baseline to post-treatment (or average change scores for trial level variables) across levels of the putative moderators.

2.5 Correlation structure

We will use Pearson correlations (tetrachoric correlations for two binary variables) to empirically identify variables that are associated with putative moderator variables. For each hypothesized baseline moderator listed in Section 1.3.5 we will calculate correlations with observed baseline variables (listed in Section 1.3.2). For each baseline moderator we will then rank the covariates by their level of association with the moderator and thus produce lists of potential confounders of moderator effects.

For each hypothesized treatment aspect moderator (listed in Section 1.3.3) we will proceed in a similar fashion to identify correlations with baseline variables in treated arms of trials only.

2.6 Missing data patterns

Using the pooled dataset we will summarize the missingness patterns of (long-format) variables to be included in the basic analysis models (see section 3.1); these are:

- Trial identifier
- Trial arm
- Respective randomisation stratifiers for trials that used stratified randomisation (design feature)
- Respective cluster identifiers for trials that used cluster randomisation (design feature)
- Respective randomisation batch identifier for trials that varied the randomisation ratio (design feature)
- Training group identifier within active treatment arms of trials
- ECBI at baseline
- ECBI at the post-randomisation time window of interest
- Putative participant-level moderator; see list in section 1.3.5.1.

- Putative trial-level moderator; see list in section 1.3.5.2.

We will identify prominent missing value patterns and also summarize the amount of missingness for each by trial and across the pooled data set. The command mvpatterns in stata can be used to describe patterns of missingness, i.e. those variables that are missing together.

Binary logistic regression will be used to identify baseline demographic variables that predict the probability of being missing for putative moderators (where the outcome is a binary coded variable that is coded 1 for missing and 0 for non-missing on the moderator of interest).

3. Data analysis plan – Inferential analysis

3.1 Assessing treatment effect modification by baseline variables (moderation)

The research questions covered under the primary objectives and under further IY effect modification by baseline variables (section 1.1) will be formally addressed using moderation analysis. The goal of the moderation analysis is to assess whether the effectiveness of treatment on the primary outcome (ECBI in time window 1) is modified by each of the following individual level variables:

- Socio-economic status
  - Lone parent
  - Teen parent
  - Low income
  - Education level
  - Unemployment
- Baseline ECBI
- Child age
- Child gender
- ADHD co-morbidities
- Emotional problems co-morbidities
- Ethnic minority
- Parental depression at baseline
- Positive parenting
- Negative parenting

Additionally trial level variables at baseline will be explored using the same analysis. Putative trial level baseline moderators are:

- Geographical region
  - UK versus non-UK: whether or not the trial was conducted in the UK. There are 7 UK trials from England and Wales within the pooled dataset and 6 from Ireland and other European countries.

    - Urban versus rural: whether the trial was carried out in a mostly urban or mostly rural setting.
- Service provider: variable denoting the type of service provider organisation.
- "Efficacy setting": level of control within the trial of the efficacy versus effectiveness.
- % Certified: variable denoting the percentage of individuals delivering the therapy who are professionally certified.
- % clinically trained: variable denoting the percentage of therapists delivering the intervention who have been clinically trained.
- Average number of sessions offered by trial design.

Each of these moderators will be explored using the same analysis model. Specifically a regression model will be fitted to the primary outcome with treatment condition and the moderator of interest as explanatory variables. The moderator effect will be modelled as a two-way interaction between treatment condition and the putative moderator. An interaction term is a product of two or more predictors, which is used as an additional predictor in the regression model. The parameter of interest is the regression coefficient of the interaction term, as this will be informative as to whether effect modification is occurring.

3.1.1 Basic analyses models for pooled dataset

In order to assess intervention effect modification (moderation), child outcomes of the combined sample (*n*=1696) will be modelled. We will consider both putative moderators measured at the individual child or parent level, and at the trial level. The advantage of this individual level analysis over conventional aggregate data meta-regression is that it enables the assessment of intervention effect moderation by both trial-level and individual-level variables (Brown et al, 2011).

The primary outcome in this analysis is the child ECBI score taken at the first follow-up, defined as between 0 to 2 months after the end of treatment. The treatment condition is the IY intervention contrasted to the control condition. This is represented by a set of dummy variables coding for the trial arm. We envisage categorising trial arms by three dummy variables (choosing waitlist as the reference group): IY intervention (y/n), addition of literacy (y/n), minimal intervention (y/n)). We will assess empirically whether there is any evidence for differences between IY or control categories and combine trial arms accordingly. For each hypothesised moderator the main effect(s) will be included in the model in addition to the product terms between the moderator(s) and the trial arm dummy variables. (For putative moderator variables measured at the individual level two variables will be constructed to represent potentially differing trial-level and individual-level effects: The trial mean will represent trial-level effects and the deviation from that mean will capture the effects of individual scores relative to their sample mean.) The parameters of interest will be the coefficients relating to the product terms. We will test the statistical significance of the interaction and if detected we will

describe its nature by estimating intervention effects within subgroups defined by the moderator.

There are some putative moderators, such as specific treatment options offered by the trial (e.g. number of IY sessions offered) which are assumed not to have an effect in the control arms of the trials. The moderation effect of such variables is assessed by including the trial arm x trial-level moderator term in the model but not any main effect of the moderator. The parameters of interest are again the regression coefficient of the product terms.

Random effects modelling assuming normally distributed outcomes will be used to separate individual-level variation from trial-level variation.

Fixed explanatory variables (fixed effects) in the model will be:-

- Dummy variable coding the trial arm effect (three variables),
- the putative moderator under investigation (two variables for individual-level moderators),
- interactions moderator x trial arm dummy (two variables for individual-level moderators),
- pre-randomisation values of the outcome (i.e. baseline ECBI),
- further baseline variables known to be predictive of post-treatment ECBI (trial-specific),
- further baseline variables necessary to define relevant conditional treatment effects (trial specific).

As tends to be standard practice in psychosocial RCT analyses, pre-randomisation values of the outcome variable are included in the model to gain precision for the intervention effect estimate. When further baseline variables were known to be predictive of child outcome within a trial these are also included as explanatory variables. For example, stratified randomisation is motivated by the stratifier being a predictor of outcome. So randomisation stratifiers will be included as explanatory variables (for more details see Section 3.1.2.). In addition, conditioning on baseline variables might be necessary in some trials to define conditional effects that can be estimated without bias. For example, the randomisation ratio was changed over time in some trials opening up the possibility that the marginal treatment effect is confounded by factors that change over the duration of the trial. In such situations the conditional trial arm effect – that is conditional on the time period during which the randomisation ratio was held constant – can still be estimated without bias. We assume that the conditional effect does not vary over time in such trials (i.e. the conditional effect is the marginal effect) and parameterise the model such that the regression coefficients of the trial arm dummy variables (and their interaction terms) represent conditional effects (for more details see Section 3.1.2.).

Random effects in the model will be:-

- Random intercepts representing the cluster structure of the pooled data, i.e.
    - varying at the level of trial (13 levels) to account for predictive effects of trial characteristics (e.g. differences in trial target

Commented [VH5]: Trial 8 is going to be added in for the depression data but I believe is still missing the child outcome.

So can only use 13 trials?

populations or general service organisation contexts affecting control groups) on child outcome under the control condition;
    - varying at the level of treatment cluster when cluster randomisation was used (trial specific, for more see Section 3.1.2.)
    - varying at the level of IY training group within the IY arm of a trial only to account for predictive effects of the training group/therapist environment within the active treatment arm.
- Random coefficients representing effect heterogeneity; specifically:
    - The regression coefficients representing treatment effects (of trial arm dummy variables) are allowed to vary with trial to model treatment effect heterogeneity (e.g. due to differences in treatment implementation or target population) not already captured by fixed baseline x trial arm interaction terms.
    - The regression coefficients of the interaction terms are allowed to vary between trials to model heterogeneity in the moderation effects of individual-level baseline variables (e.g. due to differences in treatment implementation).

The random intercept and coefficients were chosen such that the hierarchical structure of the pooled data is represented and to model heterogeneity in trial arm effects due to differences in treatment implementation and trial participants. Specifically, here we will compare the observed variability in treatment effects between putative trial-level moderators (e.g. between rural and urban trials) with the residual trial variability in treatment effects to formally assess moderation by trial-level variables. We will also allow two-way interaction effects between trial arm and individual-level variables to vary by trial should such higher-order treatment effect heterogeneity be present. Effects representing universal mechanisms (e.g. of baseline variables) are assumed constant across trials.

3.1.2 Acknowledging trial design features in the analysis models

The pooled dataset has a hierarchical structure with families (level 1 units) nested within therapy groups (level 2 units) within the intervention arm and therapy groups nested within trials (level 3 units).

In addition features of the trial designs need to be reflected in the analysis models: A number of the trials used stratified randomisation and two trials (PALs and YTST) used a cluster randomised design (Table 2). Where available these stratification variables should be conditioned on in the analysis so that trial arm effects are estimated within subpopulations defined by stratifiers. Cluster variables are available and need to be included as random intercepts to acknowledge the possible correlation between outcome values from individuals of the same randomisation cluster. Table 3 gives information on the trial design variables and for which trials stratification variables are available. Finally, in a number of the trials the randomisation ratio varied over the duration of the trial (see Table 2, 3), which could lead to confounding of the treatment effect on outcome by time

at which the family was randomised. To avoid such bias dummy variables variable which code the randomisation batch, where available, will be conditioned on in the analysis model.

Another additional complication is the fact that some trials used IY only and others used the IY+literacy intervention, including some trials using arms with both IY only and the combined intervention. This means that the type of active treatment will need to be included in the model as an additional covariate, as this may influence the treatment effect. Likewise different control conditions were used in different trials, meaning this may additionally need to be adjusted for in the model. The control condition can take four values: no care, care as usual, minimal intervention and waitlist. The control condition remains constant within each trial and so without replication differences in the type of control cannot be separated from other trial-level effects (e.g. differences between trial populations).

Thus additional variables that will be included in the analysis models to account for the design features of the different trials are:

- Varying randomisation ratios: For trials which randomised batches of people at different ratios condition on batch by including this variable as fixed effects in the model.
- Stratification variables - child age, child gender and area (trial site): As area does not have a consistent meaning across different trials it will be included only in those trials that used it as a stratifier. Baseline ECBI has been used as a stratifier in certain trials and this variable has already been included in the model for all trials. Similarly, child age and gender are available in all trials and so will be conditioned on in all trials as potential predictors of child outcome. All of these variables will be treated as fixed effects.
- Clusters for trials that used a cluster randomised design: This will be accounted for using a cluster-varying random intercept to acknowledge this extra source of variability. This will be included as two variables - one for each trial that used a cluster randomised design.
- We will account for differences in the type of intervention received by including a fixed effect, modelling as a dummy coded binary variable, which denotes whether the participant received the additional reading intervention.
- A dummy variable coding for the control types used. Since care as usual/no care was only used in two trials they will need to be combined with another category, since otherwise it will be impossible to distinguish the effect of control type from the trial level effect. We chose the following control categorization: care as usual/no care/waitlist vs minimal intervention.

3.1.3 Investigating possible confounding bias in moderating effects

We will investigate possible confounding bias in the moderating effects. Initially we will investigate one moderator at a time but since we do not know the true causal moderators it is possible that a statistically significant effect of one moderator may be actually due to another moderator effect. This can occur when one of the putative moderators has a causal effect on another/both have a common cause but also is the causal moderator of the treatment effect on the outcome, meaning that the magnitude of the causal interaction for the second putative moderator could be overestimated. We can explore this by empirically identifying potential confounders and then conditioning on them by including both, the confounder and its interaction with trial arm, in the analysis model. If the interaction effect of interest is reduced it means that the moderation effect that was detected originally could be explained by a causal moderating effect of a correlated variable. For each variable that has a statistically significant moderating effect at the 5% we will investigate the effect of adding additional interaction terms to the model based on other putative moderators that correlate highly with the moderator under investigation. To compare the sizes of moderation effects across variables and also before and after adjustment we will calculate standardised moderation indices as the change in treatment effect per unit standard deviation of a putative moderator.

Note that in the context of this research project (and perhaps shared with most stratified medicines applications) we ideally would want to identify a causal treatment effect moderator (a variable that is the cause of the treatment effect heterogeneity) and not simply a predictive marker (a variable that predicts treatment effect heterogeneity in the current target population) since we might want to further develop interventions for those families for whom they are currently not effective. The latter requires us to identify such families outside the context of the current study where observed correlations between the causal moderator and the predictive marker might be different.

3.1.4 Dealing with missing values

In the absence of missing values in the explanatory variables of the analysis models maximum likelihood (ML) will be used to estimate respective moderation parameters. Such estimates remain consistent in the presence of missing values in the response variable (post-treatment ECBI) provided missingness in ECBI is missing at random (MAR), that is that the probability of a missingness pattern depends only upon observed variables. (The ML approach is less restrictive than traditional analysis methods such as repeated measures ANOVA, which require the data to be missing completely at random (MCAR), i.e. there are no variables that drive the probability of data being missing.) In the context of our analysis model this means that the probabilities of the missingness patterns in multivariate trial-level ECBI observations depend only on the explanatory variables (baseline ECBI, trial arm, moderator etc.) and observed ECBI values from the same trial.

Table 4 shows the proportion of relevant data that is available for each demographic variable recorded at baseline. Clearly, our analysis models will

also suffer from missing values in the explanatory variables of our analysis models. These will be accounted for using multiple imputation (MI, White (2011), Royston (2004)). MI also relies on the assumption that the data are missing at random (MAR), with the observed variables predicting missingness patterns being specified during the imputation step of the procedure. To provide valid imputations of missing values and consistent parameters estimates after combining analyses results of imputed data sets according to Rubin's rules, the imputation model needs to be more general than the analysis model (White (2011). Thus at the minimum all variables included in the analysis model also need to be included in the imputation model. In addition, the imputation model can contain extra variables to relax MAR assumptions and/or to generate more precise predictions and so increase precision of estimates (White (2011)). In our context this is helpful in that it allows us to exploit the extra information provided by longer term ECBI follow-up (see Table 6) as well as by other predictors of missingness patterns (e.g. primary parent demographics).

Thus the following list of variables will be included in the imputation model:

- Analysis models variables:
    - All individual and trial level putative moderators:
        - Socio-economic status
            - Lone parent
            - Teen parent
            - Low income
            - Education level
            - Unemployment
        - Baseline ECBI
        - Child age
        - Child gender
        - ADHD co-morbidities
        - Emotional problems co-morbidities
        - Ethnic minority
        - Parental depression at baseline
        - Positive parenting
    - Negative parenting
        - Geographical region
            - UK versus non-UK: whether or not the trial was conducted in the UK. There are 7 UK trials from England and Wales within the pooled dataset and 6 from Ireland and other European countries.
            - Urban versus rural: whether the trial was carried out in a mostly urban or mostly rural setting.
        - Service provider: variable denoting the type of service provider organisation.
        - "Efficacy setting": level of control within the trial of the efficacy versus effectiveness.
        - % Certified: variable denoting the percentage of individuals delivering the therapy who are professionally certified.

        - % clinically trained: variable denoting the percentage of therapists delivering the intervention who have been clinically trained.
        - Average number of sessions offered by trial design.
    - Baseline ECBI
    - ECBI at window 1
    - Trial arm
    - Variables used for stratification as fixed effects: child gender, child age and area only within those trials that stratified by area
    - Randomisation ratio batch: as a fixed effect within those trials that altered the randomisation ratio
- Auxiliary variables:
    - Further follow-up ECBI measures
    - Primary parent demographic variables if shown to be predictive of missing values
- Dummy variables representing further fixed effects (see below)
    - Dummy variables for trial
    - Dummy variables for clusters within cluster randomised trials only
    - Dummy variables for training group within the IY arms of trials only

Reflecting the hierarchical structure in the imputation step
For the imputation model to be at least as general as the analysis model it must account for the hierarchical structure of the pooled data. This means that it must account for the trial level random intercepts. Additionally it must incorporate cluster effects within those trials that used cluster randomisation and IY group within the treatment arm of each trial. Accounting for these effects can be achieved by adding fixed effects for trials, for clusters within cluster randomised trails and for training group within the IY trial arms. Methods exist for imputing multilevel models with random effects in the imputation model but currently these are restricted to two levels (Van Buuren, 2011; Mistler, 2013). We therefore opt for the fixed effects representation, although this may lead to an overestimation of the variances of the point estimates for the fixed effects in the model (Reiter, Raghunathan and Kinney, 2006). Unless there is a very large proportion of missing data per variable or the intraclass correlations are very low, then it is likely that the bias of the fixed effects estimates will be relatively small (Drechsler, 2015) and in the IY pooling study it is the fixed effects that are primarily of interest.

Allowing for effect heterogeneity and interactions in the imputation model
Since the analysis models are set up to assess treatment effect interactions with individual-level of trial-level variables such treatment effect heterogeneity needs to be allowed for in the imputation model. We suggest imputing separately within each arm within each trial. First, separately imputing by trial ensures that the imputed data can be generated by a treatment x trial interaction, i.e. the imputed data reflects treatment-effect heterogeneity (moderation by trial-level variables). Second, separately imputing within trial arms the effect of individual-level baseline variables is allowed to vary between treatments, and thus implying the presence of a baseline x treatment

interaction within that trial. Third, separately imputing by trials arms within trials ensures that the imputed data can be generated by a baseline x treatment x trial interaction, i.e. the imputed data reflects heterogeneity across trials in the moderation effects of individual-level baseline variables.

There are a number of variables in the pooled IY data set that are known not to affect outcome under the control conditions, such as the number IY sessions offered to those allocated to the IY arm. Ideally, this restriction should be enforced in the imputation step by setting its effect on ECBI to zero.

Commented [KCL6]: Not sure that is possible in mice?

Would not want to exclude the variable from control imputation because it could still be predictive of baseline vaiables.

Alternative imputation approaches have been suggested for dealing with interactions: The first is the just another variable (JAV) approach (White, 2011), in which interactions are computed and added to the imputation model as an additional predictor. In the IY pooling study this is complicated by the large number of putative individual-level moderators, and the additional trial x treatment and moderator x treatment x trial interaction terms. The second is a linear passive approach using multiple imputation by chained equations (White, 2011). This approach includes interactions by computing them from the imputed variables and as such is likely to underestimate the strength of the interaction effect in the final model. This approach can be improved upon by allowing interactions to predict incomplete variables in the imputation step. If the outcome is to be modelled as a response to treatment, a baseline variable and the interaction between the baseline variable and treatment in the final analysis then it will be modelled as such in the imputation step. Additionally it is necessary to include an outcome x treatment interaction in the imputation model for the moderator. Imputing separately by randomisation group may provide a simpler approach when randomisation group is complete (White, 2011). Thus we preferred the flexibility of the “separate imputation” approach.

Multiple imputation by chained equations

We opted to impute missing values from respective multivariate distributions using the multiple imputation by chained equations approach (MICE, White (2011). The MICE approach involves generating imputations from a set of equations, one for each variable with missing values. For continuous variables a regression model is used, whilst for logistic regression, ordinal logistic regression or multinomial logistic regression are used for binary, ordered categorical and categorical variables respectively. In the first step missing values are replaced by random sampling with replacement from observed values of that variable. The first variable with missing values is regressed restricted only to observed values of that variable on all other variables and missing values are subsequently replaced by simulated draws from the posterior predicted distribution of that variable, that is to say the predicted values of the missing observations conditional on the observed data. The next variable with missing values is regressed restricting only to observed values of that variable on all other variables, including the first variable that has been filled in with imputed data. Missing values of the second variable are then filled in with draws from the posterior predicted distribution of that variable. This process repeats until all variables with missing values have been

imputed, completing a cycle. The process is then repeated over several cycles to stabilise the data.

3.2 Assessing treatment effect modification by aspects of treatment (moderation by post-randomisation variables)

A number of post-randomisation variables exist that may influence the effectiveness of the IY intervention (see section 1.1).

Post randomisation putative moderators at an individual level are:

- Number of IY sessions attended (if IY offered)
- Number of parents in IY (if IY offered)

At a trial level post randomisation moderators are:

- Whether therapists received regular supervision (if IY offered)

These variables are not observed in the control group. Treatment aspects such as “sessions attended if they were offered” are counterfactual in that they can only be observed for those for whom the condition “session offered” is true. The variables will be included as product terms only in the analysis models, as they and so cannot be used to model main effects. Importantly, this affects the meaning of the regression coefficients of the product terms: They now represent a combination of both, the interaction effect of interest and the main effect of the treatment aspect. In addition, the effect of the latter post-randomisation variable might well be confounded by other hidden prognostic baseline variables. Thus despite randomisation in trials we are not able to estimate interaction effects of (partly observed) aspects of treatment. While we will fit respective analysis models, resulting estimates of effects of interaction terms will need to be interpreted with care. They are subject to biases and at best will provide “some indication” of the treatment effect moderation potential of these treatment receipt variables.

3.3 Investigating higher order effects

Three way interactions involving SES and baseline severity

In order to determine whether the effect of SES on IY effectiveness differs across levels of baseline symptom severity we will explore three way interactions between SES low income, baseline symptom severity and treatment. Fixed effects) in the respective analysis model will be:-

- dummy variable coding the trial arm effect,
- baseline ECBI
- SES low income
- interaction baseline ECBI x trial arm dummy,
- interaction low income x trial arm dummy,
- interaction baseline ECBI x low income,

- three-way interaction baseline ECBI x low income x trial arm dummy
- further baseline variables known to be predictive of post-treatment ECBI (trial-specific),
- further baseline variables necessary to define relevant conditional treatment effects (trial specific).

Otherwise this model will be identical to those used for the basic analyses, including all the trial design features that were incorporated in the basic analysis models. The parameter of interest now is the regression coefficient of the three-way interaction. We will carry out a significance test for this coefficient and describe the nature of this interaction should it exist.

For assessing the three-way interaction the imputation model will need to be adapted to reflect the hypothesised three-way interactions. To achieve this we will impute separately for high and low income families within each arm within each trial.

4. Software

Statistical analysis: Stata will be used for data description and the main inferential analysis. SAS may be used for random effects modelling.

5. Reference list

Drechsler, J. (2015). Multiple Imputation of Multilevel Missing Data—Rigor Versus Simplicity. *Journal of Educational and Behavioral Statistics*, *40*(1), 69-95.
Mistler, S. A. (2013). A SAS macro for applying multiple imputation to multilevel data. In *Proceedings of the SAS Global Forum*.
Reiter, J. P., Raghunathan, T. E., & Kinney, S. K. (2006). The importance of modeling the sampling design in multiple imputation for missing data. *Survey Methodology*, *32*(2), 143.
Van Buuren, S. (2011). Multiple imputation of multilevel data. *Handbook of advanced multilevel analysis*, 173-196.
White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377-399
Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, *4*, 227-241.

5.1 Trials Included in the Pooled Sample

References for the included trials:

i) England and Wales:

1. Morpeth, L, Berry, V, Blower, S, Tobin, Taylor, R, Edwards, RT,

Linck, P, Bywater, T, Axford, N, & Lehtonen, M (2013). The effectiveness of the Incredible Years (IY) pre-school parenting programme in Birmingham, UK: A randomised controlled trial. (Investigators based at Dartington Social Research Centre).

2. Gardner F, Burton J, Klimes I. (2006) Randomised controlled trial of a parenting intervention in the voluntary sector for reducing conduct problems in children: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry* 47:1123-32. (Site, Oxford)
3. Hutchings J, Bywater T, Daley D, Gardner F, et al. (2007). Parenting intervention in Sure Start for children at risk of developing conduct disorder: randomised controlled trial. *BMJ*:334-678.
4. Scott S, O'Connor T, Futh A, Price J, Matias C & Doolan M. (2010). Impact of a parenting program in a high-risk, multi-ethnic community: The PALS trial. *Journal of Child Psychology and Psychiatry* 51, 1331-134.
5. Scott S, Sylva K, Doolan M, Price J, Jacobs B, Crook C, et al. (2010). Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: the SPOKES project. *Journal of Child Psychology & Psychiatry* 51:48-57
6. Scott S, Sylva K, Beckett C, Doolan M, Kallitsoglou A, Beecham J & Ford, T with the HCA study teams (2013, in press). Which type of parenting programme best improves child behaviour and reading? The Helping Children Achieve trial. Final Report to DCSF.
7. Scott S, Spender Q, Doolan, M, Jacobs, B, & Aspland, H (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal, 323*, 194-200.

ii) Ireland and other European countries:

8. McGilloway, S., Ni Mhaille, G., Bywater, T., Leckey, Y., Kelly, P., Furlong, M., Comiskey, C., & Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: A randomised controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology, 80*, 116-27 (Ireland)
9. Axberg, U., & Broberg, A. G. (2012). Evaluation of the Incredible Years in Sweden: The transferability of an American parent-training program to Sweden. *Scandinavian Journal of Psychology,* 53, 224–232.
10. Larsson, B., &. Fossum, S, et al. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children: results of a randomized controlled trial. *European Child & Adolescent Psychiatry*,18, 42-52
11. Gaspar, M. &. Seabra-Santos, M. Randomised trial of the Incredible Years Basic Parenting Programme in Portugal (to be completed Spring 2013)
12. Leijten, P., Raaijmakers, M. A. J., Orobio de Castro, B., Van den Ban, E., & Matthys, W. (2013). Effectiveness of the Incredible Years parent training for the reduction of disruptive behavior in

migrant children. (In preparation). (Netherlands)
13. Menting, A. T. A., Orobio de Castro, B., Matthys, W., & Wijngaards-de Meij, L. D. N. V. (2013). A randomized trial of parent training for mothers being released from incarceration. (Submitted for publication). (Netherlands)