Could scale-up of parenting programmes improve child disruptive behaviour and reduce social inequalities? Using individual participant data meta-analysis to establish for whom programmes are effective and cost-effective

Frances Gardner, Patty Leijten, Joanna Mann, Sabine Landau, Victoria Harris, Jennifer Beecham, Eva-Maria Bonin, Judy Hutchings and Stephen Scott

NHS

National Institute for Health Research

Could scale-up of parenting programmes improve child disruptive behaviour and reduce social inequalities? Using individual participant data meta-analysis to establish for whom programmes are effective and cost-effective

Frances Gardner,[1]* Patty Leijten,[1] Joanna Mann,[1] Sabine Landau,[2] Victoria Harris,[2] Jennifer Beecham,[3] Eva-Maria Bonin,[3] Judy Hutchings[4] and Stephen Scott[2]

[1]Department of Social Policy and Intervention, Centre for Evidence-based Intervention, University of Oxford, Oxford, UK
[2]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[3]Personal Social Services Research Unit, London School of Economics and Political Science, London, UK
[4]School of Psychology, Bangor University, Bangor, UK

*Corresponding author

This report should be referenced as follows:

Gardner F, Leijten P, Mann J, Landau S, Harris V, Beecham J, *et al*. Could scale-up of parenting programmes improve child disruptive behaviour and reduce social inequalities? Using individual participant data meta-analysis to establish for whom programmes are effective and cost-effective. *Public Health Res* 2017;**5**(10).

Public Health Research

Criteria for inclusion in the *Public Health Research* journal
Reports are published in *Public Health Research* (PHR) if (1) they have resulted from work for the PHR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Public Health Research* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

PHR programme

The Public Health Research (PHR) programme, part of the National Institute for Health Research (NIHR), evaluates public health interventions, providing new knowledge on the benefits, costs, acceptability and wider impacts of non-NHS interventions intended to improve the health of the public and reduce inequalities in health. The scope of the programme is multi-disciplinary and broad, covering a range of interventions that improve public health. The Public Health Research programme also complements the NIHR Health Technology Assessment programme which has a growing portfolio evaluating NHS public health interventions.

For more information about the PHR programme please visit the website: http://www.nets.nihr.ac.uk/programmes/phr

This report

Abstract

Could scale-up of parenting programmes improve child disruptive behaviour and reduce social inequalities? Using individual participant data meta-analysis to establish for whom programmes are effective and cost-effective

Frances Gardner,[1]* Patty Leijten,[1] Joanna Mann,[1] Sabine Landau,[2] Victoria Harris,[2] Jennifer Beecham,[3] Eva-Maria Bonin,[3] Judy Hutchings[4] and Stephen Scott[2]

[1]Department of Social Policy and Intervention, Centre for Evidence-based Intervention, University of Oxford, Oxford, UK
[2]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[3]Personal Social Services Research Unit, London School of Economics and Political Science, London, UK
[4]School of Psychology, Bangor University, Bangor, UK

*Corresponding author   frances.gardner@spi.ox.ac.uk

Background: Child disruptive behavioural problems are a large and costly public health problem. The Incredible Years® (IY) parenting programme has been disseminated across the UK to prevent this problem and shown to be effective in several trials. It is vital for policy to know for which families IY is most effective, to be sure that it helps reduce, rather than widen, socioeconomic inequalities. Individual trials lack power and generalisability to examine differential effects; conventional meta-analysis lacks information about within-trial variability in effects.

Objectives: To overcome these limitations by pooling individual-level data from the IY parenting trials in Europe to examine to what extent it benefits socially disadvantaged families. Secondary objectives examine (1) additional moderators of effects on child behaviour, (2) wider health benefits and potential harms and (3) costs, cost-effectiveness and potential long-term savings.

Design: Individual participant data meta-analysis of 14 randomised trials of the IY parenting intervention.

Settings: UK (eight trials), the Netherlands, Ireland, Norway, Sweden and Portugal.

Participants: Data were from 1799 families, with children aged 2–10 years (mean 5.1 years; 63% boys).

Interventions: IY Basic parenting programme.

Main outcome measures: Primary outcome was disruptive child behaviour, determined by the Eyberg Child Behavior Inventory Intensity scale (ECBI-I). Secondary outcomes included self-reported parenting practices, parenting stress, mental health, children's attention deficit hyperactivity disorder (ADHD) and emotional symptoms.

Results: There were no differential effects of IY on disruptive behaviour in families with different levels of social/socioeconomic disadvantage or differential effects for ethnic minority families, families with different parenting styles, or for children with comorbid ADHD or emotional problems or of different ages. Some moderators were found: intervention effects were strongest in children with more severe baseline disruptive behaviour, in boys, and in children with parents who were more depressed. Wider health

benefits included reduced child ADHD symptoms, greater parental use of praise, and reduced harsh and inconsistent discipline. The intervention did not improve parental depression, stress, self-efficacy or children's emotional problems. Economic data were available for five UK and Ireland trials (maximum $n = 608$). The average cost per person of the IY intervention was £2414. The probability that the IY intervention is considered cost-effective is 99% at a willingness to pay of £145 per 1-point improvement on the ECBI-I. Estimated longer-term savings over 20 years range from £1000 to £8400 per child, probably offsetting the cost of the intervention.

Limitations: Limitations include a focus on one parenting programme; the need to make assumptions in harmonising data; and the fact that data addressed equalities in the effectiveness of, not access to, the intervention.

Conclusions: There is no evidence that the benefits of the IY parenting intervention are reduced in disadvantaged or minority families; benefits are greater in the most distressed families, including parents who are depressed. Thus, the intervention is unlikely to widen socioeconomic inequalities in disruptive behaviour and may have effects in narrowing inequalities due to parent depression. It was as likely to be effective for older as for younger children. It has wider benefits for ADHD and parenting and is likely to be considered to be cost-effective. Researchers/funders should encourage data sharing to test equity and other moderator questions for other interventions; further research is needed on enhancing equality of access to interventions.

Funding: The National Institute for Health Research Public Health Research programme.

Contents

List of tables

List of figures

List of abbreviations

A&E	accident and emergency department
ADHD	attention deficit hyperactivity disorder
APQ	Alabama Parenting Questionnaire
BDI	Beck Depression Inventory
CAMHS	child and adolescent mental health service
CBCL	Child Behavior Checklist
CC	complete case
CEAC	cost-effectiveness acceptability curve
CI	confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CONSORT	Consolidated Standards of Reporting Trials
CONSORT-SPI	Consolidated Standards of Reporting Trials Statement for Social and Psychological Interventions
CSRI	Client Service Receipt Inventory
ECBI-I	Eyberg Child Behavior Inventory Intensity scale
GP	general practitioner
IPD	individual participant data
IY	Incredible Years®
MAR	missing at random
MI	multiple imputation
NGO	non-governmental organisation
NICE	National Institute for Health and Care Excellence
ONS	Office for National Statistics
PACS	Parental Account of Children's Symptoms
PaPI	Parenting Practices Inventory
PI	principal investigator
PPI	public and patient involvement
PRISMA-IPD	Preferred Reporting Items for a Systematic Review and Meta-analysis of individual participant data
PS	Parenting Scale
PSI-SF	Parental Stress Index Short Form
PSOC	Parental Sense of Competence
PSQ	Parent Satisfaction Questionnaire
RCT	randomised controlled trial
SD	standard deviation
SDQ	Strengths and Difficulties Questionnaire
SES	socioeconomic status
T1	time point 1
T2	time point 2
WTP	willingness to pay

Plain English summary

Behavioural problems in children are a major and costly public health issue. Parenting programmes that aim to reduce these problems, such as Incredible Years® (IY), have been found to be effective in the UK and elsewhere. However, little is known about whether or not the most socially disadvantaged and troubled families benefit from these programmes, and whether or not they are cost-effective. Up to now, trials and systematic reviews of parenting programmes have mostly been too small to answer these questions. This study combined individual child and parent data from all of the high-quality (i.e. randomised controlled trial) evaluations of the IY parenting in Europe into a single data set. This was then large enough to answer important questions about who benefits from this parenting programme, in terms of reduction in child behaviour problems, whether it narrows or widens social inequalities in these problems, whether or not it has other benefits (or harms) and whether or not it is likely to be cost-effective. The study pooled data for 1799 families, from 14 trials of IY in the UK, the Netherlands, Ireland, Norway, Sweden and Portugal. It found that for families who are socially disadvantaged or from an ethnic minority, the programme is just as likely to be beneficial. However, for the most distressed families, including children with marked behaviour problems and depressed parents, it was more beneficial than for average families. It was just as effective for older as younger children. The data could not address potential inequalities in programme access. In addition, the programme appears to have wider benefits, improving parenting and child attention deficit hyperactivity disorder, and is likely to be cost-effective.

Scientific summary

Background

Disruptive behaviour problems in children are a large and costly public health problem, which, if not remediated, raise the risk of drug use, criminality, unemployment and poor (mental) health later in life. Many interventions have been developed to prevent these problems but only a few have strong UK evidence of effectiveness. There is growing government interest across the UK and globally in finding solutions to these problems. The Incredible Years® (IY) parenting programme has been disseminated across the UK with government support and shown to be effective in seven randomised trials. However, it is vital for policy-making to know for which families these interventions are most effective, in order to be sure that they are helping to reduce, rather than widen, social inequalities. This is a realistic concern, as we know that for some public health and social interventions (e.g. Sure Start), more advantaged families make better use of these facilities. Surprisingly, there are no clear answers to these questions about social inequality from existing trials of parenting interventions, because each individual trial is quite small and thus lacks the power to test these questions.

Research questions

Underlying population, intervention, comparison, outcome, study design (PICOS) question for main effects in the pooled trials:

- Population: families with children aged 2–10 years
- Intervention: IY parenting programme
- Comparison: waiting list, minimal intervention or care as usual
- Outcome (benefit): child disruptive behaviour post test
- Study design: randomised controlled trial (RCT).

Specific questions for this individual participant data (IPD) meta-analysis:

- To what extent does the IY parenting intervention benefit:

  - the most socially disadvantaged compared with average families?
  - families from ethnic minorities compared with from ethnic majority?
  - children with different characteristics, including age, gender, severity of conduct problems and comorbid problems at baseline?
  - children whose parents have different levels of depression and parenting skill at baseline?

- To what extent do trial-level effects predict outcome, including contextual variables (country, rural vs. urban), factors indexing intervention fidelity (staff training, certification or supervision) and sessions offered and taken up?
- What are the wider public health benefits and potential harms of the IY parenting intervention?
- What are the costs, cost-effectiveness and potential longer-term savings of the IY parenting intervention?

Methods

Protocol and trial registration

The protocol is available on the National Institute for Health Research Public Health Research website (project number 12/3070/04).

Eligibility criteria

Completed randomised trials of the IY parenting programme in Europe. Non-RCTs were excluded. No restrictions were placed on when trials were conducted or included outcome measures.

Identifying studies: information sources

Studies were identified in 2013 based on (1) systematic literature search in databases [Cumulative Index to Nursing and Allied Health Literature (CINAHL), EMBASE, Global Health, MEDLINE and PsycINFO], (2) the IY website, which lists trials evaluating IY, (3) the European IY mentors' network and (4) the advice of experts. Searches in January 2015 revealed no further completed trials.

Identifying studies: search

EMBASE, Global Health, MEDLINE (< 1946–present) and PsycINFO were searched via OvidSP using the following search terms:

1. incredible year$.mp
2. webster-stratton.mp
3. 1 or 2.

CINAHL was searched via EBSCO*host*.

Study selection processes

Eligibility was assessed by the first author and double-checked by additional authors. There were no differences of opinion.

Data collection process

Anonymised data for 15 randomised trials were requested for all families randomised. Data for 14 trials were supplied; investigators for the 15th reported that the data were no longer available. Raw (i.e. not recoded) individual item-level data were supplied in IBM SPSS Statistics version 21 (IBM Corporation, Armonk, NY, USA) and checked for missing items, internal consistency and consistency with trial protocols and reports. Copies of the original questionnaires were requested to check for consistent use across trials. Investigators were contacted for additional information for interpretation of IPD.

Data items

All available data on socioeconomic disadvantage, demographics, moderators, wider benefits and service use were requested. The instrument for the primary outcome (disruptive child behaviour) was chosen based on the measures used most frequently. If trials did not include this measure (three trials), data from a similar measure were requested and converted to a score on the primary measure using norm deviation scores. Instruments for secondary outcomes were chosen based on those that were available in each trial.

Results

Moderators of intervention effectiveness (maximum n = 1696)

Social and socioeconomic disadvantage as moderators

We considered five variables as capturing socioeconomic disadvantage: low income, low education, unemployment, lone parent or teenage parent. Results were robust across variables: there was no evidence to suggest that any indicator of socioeconomic disadvantage affected the benefit of the IY programme. More specifically, there was no moderation by low income compared with high-income status [effect modification index 1.9 points, Eyberg Child Behavior Inventory Intensity scale (ECBI-I), 95% confidence interval (CI) –4.8 to 8.6 points; $p = 0.58$], by low or high education status (modification index 4.4 points, 95% CI –2.2 to 10.9 points; $p = 0.49$), by being employed or unemployed (modification index 4.88 points, 95% CI –2.7 to 12.4

points; $p = 0.21$), by lone parent status (modification index 0.5 points, 95% CI –6.1 to 7.1 points; $p = 0.88$) or by teenage parent status (modification index 7.3 points, 95% CI –2.2 to 16.9 points; $p = 0.13$).

Ethnicity as moderator
There was no evidence of moderation by ethnicity (modification index –1.4 points, 95% CI –9.8 to 7.1 points; $p = 0.75$).

Child characteristics as moderators

Age and gender
There was no evidence of moderation by child age (modification index 0.04 points, 95% CI –0.1 to 0.2 points; $p = 0.65$), across the range 2–10 in the pooled trials. There was a large moderation effect by gender (modification index –6.6 points, 95% CI –13.0 to –0.3 points; $p = 0.04$), such that boys benefited more than girls.

Disruptive behaviour
Effectiveness of the intervention was moderated by baseline level of disruptive behaviour, with children with more severe problems benefiting more. This effect was found both at an individual level (modification index –4.3 points, 95% CI –7.9 to –0.7 points; $p = 0.02$) and at a trial level (modification index –18.3 points, 95% CI –24.6 to –12.0 points; $p < 0.001$).

Attention deficit hyperactivity disorder and emotional problems
There was a suggestion that the putative moderator effect by attention deficit hyperactivity disorder (ADHD) was non-linear; thus, a quadratic term was included in the model ($p = 0.02$). The test for effect moderation did not reach significance at the 5% level ($p = 0.07$). We conclude that there is insufficient evidence for moderation by level of ADHD at baseline, either linear or quadratic. We found no evidence of moderation by emotional problems (modification index –2.3 points, 95% CI –6.7 to 0.9 points; $p = 0.13$).

Parent characteristics as moderators

Depression
We found a large and significant effect moderation by depression (modification index –4.8 points, 95% CI –8.4 to –1.1 points; $p = 0.01$), in the direction that children whose parent was more depressed benefited more from the intervention.

Parenting behaviour
None of the parenting variables (e.g. monitoring, praise or punishment) showed a moderation effect, suggesting no differential benefit for children by level of positive or negative parenting at baseline.

Individual-level moderators after adjusting for confounding
We adjusted significant moderation effects for potential confounders. The finding of a treatment effect by baseline ECBI-I was relatively robust to adjustment, in terms of the size of the moderation index and statistical significance. In contrast, moderation by parent depression was affected by adjustment for praising, with the effect size reduced and non-significant. It is therefore unclear what causes this detected effect moderation: parental clinical characteristics or parenting approach.

Trial-level moderators
We assessed whether variables characterising the setting of the trial (UK/Ireland vs. other locations) or affecting programme delivery (e.g. % staff clinically trained/% IY certified) moderated outcome. Note that, as in aggregate-level metaregression, we effectively have a maximum of only 13 programme effects on which to base these analyses. Thus, we could only investigate trial-level variables with a reasonable number of replicates in the data set (e.g. rural locations were rare, 4% of intervention sites). The number of sessions offered moderated the programme effect ($p < 0.001$), with treatment benefits decreased for

trials offering more sessions. This finding, although potentially interesting, is hard to interpret, as variation between trials is very low (range 12–14 sessions) and unlikely to make much difference to outcome. No other trial-level variables moderated the intervention effect.

Wider health benefits

Children's mental health (attention deficit hyperactivity disorder symptoms and emotional problems)

The intervention had wider benefits on child ADHD symptoms ($\beta = -0.30$, 95% CI -0.44 to -0.17). The percentage of intervention children scoring above the borderline threshold on ADHD fell from 54% to 42%, a 12% reduction, compared with a 2% reduction in control participants.

No difference in child emotional symptoms was detected ($\beta = -0.06$, 95% CI -0.18 to 0.06; $p = 0.303$); these were reduced in both groups. In both groups the percentage scoring above the borderline threshold on emotional problems fell from 40% to 32%.

Parental mental health

The intervention did not affect parental mental health, with a non-significant trend towards reduced symptoms of depression ($\beta = -0.08$, 95% CI -0.17 to 0.01). Neither parenting stress ($\beta = -0.18$, 95% CI -0.44 to 0.07) nor self-efficacy ($\beta = 0.32$, 95% CI -0.77 to 0.13) improved because of the intervention.

Harsh and inconsistent parenting

The intervention reduced harsh and inconsistent parenting practices. The results were robust across three of the indicators of harsh and inconsistent parenting practices: corporal punishment ($\beta = -0.22$, 95% CI -0.42 to -0.01), threatening ($\beta = -0.21$, 95% CI -0.36 to -0.06) and shouting ($\beta = -0.31$, 95% CI -0.61 to -0.01). IY did not reduce parental laxness ($\beta = -0.15$, 95% CI -0.37 to 0.07). Parents allocated to IY thus reported fewer harsh parenting practices post test, although not fewer inconsistent practices.

Positive parenting

The IY programme improved some aspects of positive parenting practices, including praising the child more frequently ($\beta = 0.26$, 95% CI 0.01 to 0.51). Parents did not report using more tangible rewards ($\beta = 0.15$, 95% CI -0.16 to 0.45) and monitoring of child behaviour did not increase ($\beta = 0.05$, 95% CI -0.08 to 0.18).

Possible harms

We checked the direction of effects of all secondary outcomes for signs of harmful effects of IY. All effects pointed in the direction of benefits, and none in the direction of harm.

Costs, cost-effectiveness and longer-term savings

Economic data were available for five trials in the UK and Ireland (maximum $n = 608$). The cost of the intervention ranged from £1733 to £2586 per trial. The probability that the intervention will be considered cost-effective reaches 80% at a willingness to pay of £121 per 1-point improvement on the ECBI-I. There were no cost variations associated with social disadvantage, ethnicity, ADHD or child emotional problems at baseline. In contrast with the main effectiveness analysis, there were also no cost variations associated with baseline levels of disruptive behaviour or parental depression. Potential longer-term savings are estimated at £1400–8400, probably offsetting the cost of the intervention. Thus, the intervention is likely to be cost-effective.

Risk of bias within and across studies

This was assessed within studies using the Cochrane risk of bias tool; the risk of bias was judged low in most trials. Risk of bias was also low with regard to availability of studies, as all but one eligible trial supplied data. Risk of bias may be higher with regard to availability of data, and how some constructs were harmonised and synthesised.

Parent involvement

Despite our findings being at odds with much of the literature, they resonated with the parents' experiences; indeed, parents in our focus groups expected many of the findings.

Discussion

The underpinning question for this study concerns equity effects of parenting interventions. Successful synthesis of data from almost all randomised trials of the IY parenting intervention in Europe led to a uniquely large (maximum $n = 1799$) and diverse sample that allowed for the most stringent and well-powered tests of equity effects to date, of wider health benefits and of cost-effectiveness of parenting interventions for children.

The IY parenting intervention does not increase social or socioeconomic inequalities in children's disruptive behaviour problems. Families with a range of social and socioeconomic disadvantages, and those from ethnic minorities, are just as likely to benefit. If anything, the intervention reduces inequalities because some of the more distressed families (those with more severe disruptive child behaviour and parental depression) benefit more. In addition to reducing disruptive child behaviour, the intervention reduces the ADHD symptoms of children and improves several aspects of positive parenting (e.g. parents praised their children more) and negative parenting (e.g. parents used less harsh and inconsistent discipline).

Why do our results show such an encouraging lack of adverse equity effects? The content of IY is much like that of several other programmes derived from social learning theory and attachment theory. However, the collaborative and culturally sensitive approach of IY may be of paramount importance to reach similar effectiveness across a wide range of families with different social and socioeconomic status and ethnic backgrounds. Perhaps the careful focus on individual needs and values in the IY programme helps to offset any adverse effect that might otherwise result from education-based programmes.

The study has many strengths that increase our confidence in the findings. First, our unique pooled data set means that we have a large sample, which is potentially generalisable across countries, service settings and level of child problems. Second, plausibility of the findings is enhanced by our analytical strategies. We included only RCTs, controlled for the possibility of confounding effect by the type of trial (e.g. whether it was selective prevention compared with indicated prevention or treatment trial, which makes a considerable difference in terms of the initial severity of child disruptive behaviour) and we were able to check for non-linear effects. Finally, we account for missing data in moderator analyses using multiple imputation, which requires fewer restrictive assumptions than using only cases with no missing values.

Limitations of this study include, first, that we focused on only one parenting programme, albeit one of the programmes that is most established in the UK. Second, several assumptions had to be made in the process of data harmonisation, for example that different measures from different trials actually measured the same construct. Third, we had data only on equalities in effectiveness and were unable to analyse potential inequalities in access to the intervention, a topic needing further research.

To conclude, the IY parenting intervention does not increase social or socioeconomic inequalities in children's disruptive behaviour problems and is likely to be cost-effective, especially for families with higher levels of child behaviour problems or parent depression. Families with a wide range of social and socioeconomic disadvantages, and those from ethnic minorities, are just as likely to benefit from the intervention as families without these disadvantages. Older and younger children are equally likely to benefit. Moreover, children with more disruptive behaviour or with parents who are more depressed benefit more than their less distressed counterparts. This study is potentially a valuable model for other public health policy questions, which we suggest could mine existing data, using IPD meta-analysis, at relatively low cost in order to enhance understanding of cost-effectiveness and equity effects of different commissioning strategies and promote social mobility.

Funding

Chapter 1 Introduction

In this report, particularly in the methods and results, the headings are structured and, when appropriate, numbered according to the most relevant and up-to-date guidelines, the 'Preferred Reporting Items for a Systematic Review and Meta-analysis of individual participant data' (PRISMA-IPD Statement).[1]

Disruptive behaviour in children

Persistent disruptive or antisocial behaviour is a major public health issue, not least because it is the most common mental health problem in children. Two terms are used when employing diagnostic criteria to apply a cut-off point for the level of disruptive behaviour: oppositional defiant disorder, which is more commonly seen in younger children (defying requests, tantrums, blaming other people for their mistakes, physical aggression and so on), and conduct disorder, usually seen in adolescents, which includes, for example, more serious behaviour such as assault, theft and forcing people to have sex. Together, oppositional and conduct disorders affect 5% of the population.[2] In this report we use the term 'disruptive behaviour' synonymously with oppositional and conduct problems, as they refer to the same phenomena. We prefer this to the terms oppositional/conduct problems/disorder, as the these are often not widely understood beyond the relatively narrow confines of the mental health field, and so may not have meaning for teachers, social workers and the general public. This grant was for parenting trials primarily outside the UK NHS child and adolescent mental health services (CAMHSs), so terminology is particularly important.

There are high health burdens into adulthood. For example, for the most disruptive 5% of 7-year-olds, by the age of 25–27 years there is a 5- to 10-fold risk of alcoholism, drug abuse, criminality, domestic violence, sexually transmitted infections, unemployment, psychosis and early death.[3–5] As with poor parenting, there is a strong association with social and socioeconomic disadvantage, with a four- to fivefold higher rate of disruptive behaviour in the most disadvantaged groups in the population.[2] The extra public cost of early-onset disruptive behaviour is £225,000 (USD 335,000) per person by the age of 27 years, which is 10 times that of control participants.[6] Cost savings appear to apply to children with mild and moderate problems, as well as to those with more severe disruptive behaviour who are at greatest risk for long-term problems.[7]

Parenting interventions

Parenting interventions potentially form an important public health strategy for preventing disruptive behaviour and other poor outcomes in children for a number of reasons. First, poor parenting skills are strongly predictive of youth disruptive behaviour.[8,9] Second, because the public health and financial burden of child disruptive behaviour and its later consequences are very high, it provides an excellent opportunity for early preventative intervention. Serious, enduring disruptive behaviours in adulthood nearly all always begin in childhood, particularly in early childhood: in < 10% of persistent cases does the disruptive behaviour begin after the age of 18 years.[10] Third, the National Institute for Health and Care Excellence (NICE) and Cochrane reviews of randomised controlled trials (RCTs)[11–13] clearly show that parenting interventions help prevent child disruptive behaviour problems and enhance parent and child mental health. Many policy bodies worldwide have recognised this (e.g. World Health Organization,[14] United Nations Office on Drugs and Crime[15] and US Centers for Disease Control[16]). The National Research Council and Institute of Medicine[17] has issued a strong call for early preventative intervention trials to explore how mental health disorders can be prevented. These calls are echoed in UK policy. Thus, the Department of Health states the need to promote evidence-based parenting programmes in several 2011 policy documents (e.g. *No Health Without Mental Health: A Cross-Government Mental Health Outcomes Strategy for People of all Ages - a Call to Action*;[18] *Talking Therapies: A Four-Year Plan of Action*[19]). The NICE report[13] recommends parenting interventions. Its meta-analysis found an effect size of 0.6 standard deviation (SD) on child problem behaviour in the 3–8 years

age range, with good long-term effects, which is an extremely worthwhile effect in public health terms. On 27 February 2013, NICE launched its full guideline for prevention and management of antisocial behaviour and conduct disorders, a further recognition of the public health importance of this problem. The guidance confirmed the previous health technology assessment analyses and recommended the use of high-quality evidence-based parenting programmes to prevent the development of antisocial behaviour and conduct disorders. Similar to the economic modelling study by members of our own group,[20] and by Lee et al.,[21] the NICE guidance[13] shows good financial returns from investment in evidence-based programmes.

Moderators: understanding for whom interventions are effective and their 'equity' implications

Given a strong body of evidence showing beneficial main effects of parenting interventions on child outcomes in various trial populations, it is important to understand any effect of heterogeneity between and within study populations. Investigations that aim to establish whether or not there are differential effects for different subgroups in the population are referred to as moderator, effect modifier or subgroup analyses. Importantly, we distinguish moderator from 'predictor' analyses, which we define as those that do not examine interaction effects, but instead analyse predictors of outcome in a treated group only, making no comparison with change in the control group. This is a significant distinction, as without this comparison it is not possible to know whether or not any differential effects are related to the intervention per se, or if they merely reflect naturally occurring subgroup differences in prognosis. Understanding treatment effect heterogeneity is vital for a number of reasons, including (1) assessing equity effects of interventions, and whether or not they work for those at greatest risk of poor outcomes, (2) ensuring that interventions are targeted appropriately, (3) understanding for whom intervention strategies may need to be improved or altered and (4) exploring possible differential intervention mechanisms.

Assessing equity effects of interventions

Health inequities have been defined as unfair and avoidable differences or inequalities in health between subgroups in populations. Subgroups might be defined by social or socioeconomic disadvantage, gender or ethnicity (e.g. Tugwell et al.,[22] Welch et al.[23] and Whitehead[24]). Most health and well-being outcomes are strongly patterned by social and socioeconomic disadvantage, and there are multiple biological and environmental reasons for these observed inequalities. However, an over-riding concern for public health policy and practice is to ensure that interventions that may be effective at improving the mean level of a health outcome across the population do not, at the same time, have the unintended effect of increasing inequalities between groups. For several reasons interventions may serve to increase inequalities, for example if there is differential screening, diagnosis, access, uptake, compliance or effectiveness of interventions[22] by different groups in the population. Such effects were inferred from observational epidemiological data relating to a child public health programme in rural Brazil,[25] suggesting that in this context, despite good programme access and coverage for the very poorest people, health outcomes for this group were slow to change and occurred only subsequent to improvements in wealthier children. It is important to note that, although inequalities may be magnified (or reduced) at any of these stages on the pathway from screening to intervention effectiveness, moderator analyses generally deal with differential effectiveness, subsequent to intervention access and uptake. It is this aspect of health equity that the current study is able to investigate; different approaches and methods are needed to examine differential access.

These equity considerations are highly relevant to parenting interventions; poor parenting and disruptive behaviour are patterned by social and socioeconomic disadvantage and are linked to diminished life chances in key areas of schooling, employment and health.[5,26] The question then arises whether or not disadvantaged children are also at higher risk of poorer outcomes from parenting interventions than more advantaged children. With the increasing availability of parenting interventions as a public health measure across the population, this becomes a hugely important question. It is a risk well recognised in public health, albeit often poorly defined,[27] and which is termed the 'inverse care effect', that interventions may sometimes have greater benefits for more advantaged families.[28] This 'inverse' effect was found in the early Sure Start evaluations.[29] If it were the case that more disadvantaged families were failing to access or benefit from parenting interventions, then roll-out of these programmes would have the unintended

consequence of increasing social inequalities in parenting and disruptive behaviour, and potentially in subsequent life chances, in the very groups at highest risk for these problems. Such an unequal effect could occur despite average beneficial effects of an intervention across the population. Conversely, if an intervention were to have differential effects that conferred greater benefit on the most disadvantaged families, then it would potentially serve to narrow social inequalities in the intended outcome.

Equity effects by ethnicity are also important to understand for several reasons. First, it is important not to increase any social inequalities that may arise from ethnicity, if there were to be differential effects of parenting intervention by ethnicity. Second, investigating whether or not there are ethnicity effects can help to illuminate questions about the generalisability of interventions developed (and often delivered) by people from an ethnic majority, and then applied to minority families. Questions such as these about transportability of interventions across cultures and countries are of wider global importance, as countries seek to enhance child outcomes through parenting and other psychosocial interventions.[14,30,31]

Ensuring that interventions are targeted appropriately

A second reason for examining moderator effects is to establish for which subgroups interventions might be most efficiently targeted. In the parenting literature, there is mixed evidence and opinion about the effectiveness of interventions for children at different levels of risk for conduct disorder. Yet this evidence is vital for prevention policy and determining the most appropriate targets for scarcer-indicated prevention and treatment delivery. When interventions are targeted at children showing early signs (indicated prevention) or diagnoses (treatment) of conduct disorder, are parenting interventions most effective for those at higher or lower levels of severity?

Understanding for whom interventions strategies may need to be improved or altered

If it were found to be the case that parenting interventions are less effective for children with high levels of attention deficit hyperactivity disorder (ADHD), in addition to conduct problems, or for those whose parents are depressed, then it would be vital to develop and test improved versions of parenting programmes or to provide other effective interventions that are more suited to the needs of these families. For example, if parents who are depressed show less benefit, then it might be important to modify the approach or add additional strategies that address the ways in which depression may affect parenting and behaviour change.[32] Equally, if low-income or ethnic minority families were found to benefit less, then content or delivery features would need to be modified to ensure fairer access to effective interventions, for example by use of cultural or practical adaptations.[33]

Moderator analyses may also provide pointers towards differential intervention processes

When there are differential effects on child behaviour outcomes by subgroup, this may be a sign of potential subgroup differences in the underlying mechanisms of change. Hypotheses about mechanisms, drawn from existing theory and process evaluations, can then be tested in analyses of moderated mediation.[34–37]

Prior literature on predictors and moderators of parenting intervention effects

Existing literature points to a number of putative treatment effect moderators. The main categories are:

- social and socioeconomic disadvantage
- ethnicity
- child characteristics: behavioural and emotional problems; age and gender
- parents' clinical and parenting practices
- contextual factors.

We will review these in turn.

Social and socioeconomic disadvantage

The wide dissemination of parenting interventions in the UK makes it urgent to determine the effectiveness of these interventions across a range of social groups. Early poor parenting is patterned by social and socioeconomic disadvantage and in the absence of intervention is predictive of child behaviour problems and diminished educational and health outcomes, which suggests that it is a key mechanism for perpetuating health and social inequalities across generations (e.g. UK birth cohort data, Ermisch[8] and Kelly *et al.*[38]). Prospective studies[39,40] of high-risk samples support the hypothesis that poor parenting mediates intergenerational transmission of adverse child outcomes. Furthermore, UK cohort analyses suggest that social inequalities in both child and parent mental health appear to be widening over time.[41,42]

These observational studies suggest that parenting may be an important mechanism in promoting inequalities and that, with current policies, inequalities in mental health do not appear to be lessening over time. Instead, intervention studies are needed to examine whether improving parenting will have an adverse or beneficial effect on inequalities in child outcomes. Despite a large number of high-quality trials and systematic reviews on the topic, conclusions from the literature are mixed on the question of whether or not there are differential benefits of parenting interventions by family social and socioeconomic disadvantage. A number of trials and systematic reviews have found weaker effects of parenting interventions for more disadvantaged families, including two of the largest meta-analyses of predictor effects.[43,44] On the other hand, other reviews draw more uncertain conclusions.[12,45] Some of the few individual trials testing moderator effects[46–48] find that these parenting interventions are equally or sometimes more effective for the most disadvantaged families, which, without intervention, tend to do worse, suggesting the potential for reversing some of the poorer child outcomes associated with family poverty. It is worth noting that most trials have not used their data to ask these questions. However, this limitation aside, there are a number of methodological reasons for these conflicting results, and we argue (in *Methodological limitations of current moderation literature*) that such moderation questions cannot adequately be answered from individual trials or aggregate-level meta-analyses alone.

The qualitative literature on parents' experiences can also help identify putative moderators. A systematic review by Kane *et al.*[49] found five qualitative studies of the views of (mainly disadvantaged) parents who had participated in parenting interventions; our searches found several more recent studies of the Incredible Years® (IY) programme, two of which were embedded within a trial.[50,51] Barriers to uptake and success in intervention included stresses related to time pressure, financial pressure, the influence of antisocial neighbourhoods, reluctance to share problems with others and lack of support from family members. These findings show congruence between parents' views of barriers and those factors drawn from the literature on moderators and risk factors for child disruptive behaviour.[52]

Ethnicity

Few trials in the UK or elsewhere have been able to examine effects of parenting interventions on families from different ethnic backgrounds. This is vital in order to assess whether such services are likely to reduce or widen inequalities by ethnicity in child and maternal outcomes.[53] This is becoming increasingly important in the UK, and, for example, recent population data in London suggest that in some of the more disadvantaged boroughs one-third of children or more (over half in Tower Hamlets) belong to an ethnic minority. When there is evidence from other countries, mainly from the USA, the picture is quite mixed about whether or not there are differential effects of parenting interventions. Measuring a very wide range of parent and child outcomes, as well as parent engagement and satisfaction, Reid *et al.*[54] found surprisingly little evidence of differential effects of the IY parenting intervention by ethnicity in a predictor analysis; when there were differences by ethnicity, they tended towards greater engagement and uptake by some minority groups. On the other hand, much theoretical and prevention literature from the USA focuses on the need for interventions to be specially adapted for different ethnic groups.[55,56] Even leaving aside the controversial question of which is most effective,[57] approaches involving substantial adaptation by culture would imply the need to run parenting groups that are separated by ethnicity, and this raises critical questions about what would be appropriate service delivery patterns for multiethnic UK inner

cities.[58] There have been very few studies of outcome differences in parenting trials by ethnicity in the UK or other European countries, or qualitative studies alongside trials. One exception is the trial by Scott *et al.*,[59] conducted in a highly deprived London borough. They found considerable baseline differences in parenting practices by ethnicity, but, intriguingly, no ethnic differences in attendance, or in intervention effects on parenting skills. This trial therefore suggested that despite large initial differences in parenting style by ethnicity, parenting programmes apparently based on 'Western' family values are equally effective with ethnic minority parents, when sensitively delivered, using a programme with an underlying philosophy that is collaborative and parent centred.[60] A Manchester study of parents' views of a similar programme, Triple P,[61] suggested that Asian and African parents are most inclined to take up a parenting intervention, but that white and African Caribbean parents are somewhat less likely to do so.

Child characteristics: behavioural and emotional problems – age and gender

Relatively little is known about how child characteristics, such as age, gender and initial severity and comorbidity of problems, and parent characteristics, such as depression and parenting style, influence the effects of parenting interventions.

Age

Whether or not intervention effects (and cost-effectiveness) vary by age of the child is a particularly salient policy issue. Current policy thrust is towards earlier interventions being thought to having more powerful and longer-term effects on child outcomes, based apparently on evidence from neuroscience[62] and from intervention studies. For example, Heckman's[63] broadly conceived synthesis of a range of youth interventions at different ages concluded that early interventions are more cost-effective than later ones for improving subsequent 'human capital' outcomes. In the UK, the 2011 Allen Report on early intervention[62] strikingly proposed that resources should be taken away from later intervention and redeployed to earlier age groups. Yet there are surprisingly few conclusive data on the most effective age for targeting preventative interventions for disruptive behaviour, with small trials providing conflicting results. For example, one recent trial found no age effects,[48] whereas another found a slight advantage of younger age[46] but it was limited by including only a narrow preschool age range in the trial. Systematic reviews have also produced mixed findings; two reviews of age effects on parenting interventions found no differential advantage of young age,[12,43] and two found greater effects for older children.[64,65] However, most of these reviews are not up to date, and are severely constrained by lack of data on age at an individual (rather than trial) level. Moreover, the number of trials is rarely large enough to be able to control for baseline severity of child problems, which is important because age is often confounded by severity (as it is with gender, such that older children, and boys, tend to have more severe problems), and there is evidence that children with more severe behaviour problems may gain more from these interventions.

Gender

The picture from existing literature is complex: when girls present with severe disruptive behaviour problems, they often show more marked comorbidity than boys. However, in prevention samples, they often have less severe behaviour problems, and this might contribute to finding stronger intervention effects in boys in some studies (e.g. moderator analyses in the Wales Sure Start Trial[46]). However, other prevention trials find no such gender effects.[54,66,67] Therefore, it is important that studies are able to investigate whether or not gender moderates intervention outcomes, while also controlling for initial severity and comorbidity, as it may be likely that these are associated with gender. If there are weaker effects for girls, programmes may need adjusting to take account of their needs.

Initial severity of behaviour problems

Systematic reviews of parenting interventions again provide conflicting results, albeit based on contrasting meta-analytic methods for synthesising moderator effects. One review found that children with higher levels of behaviour problems did better,[43] another that they did worse[44] and a third review[50] found no difference. As with other moderators with mixed findings, this may be because of programme differences, and especially methodological weaknesses and differences. These are vital issues for making public health decisions about the most appropriate targeting of parenting intervention by level of severity.

Comorbid child problems

Linked to initial severity is the question of whether or not child comorbid problems moderate intervention effects. Some studies suggest that children with high levels of other mental health problems (e.g. ADHD) do less well in parent training, but others have found as good a response in these children.[68,69] If severe ADHD does moderate treatment response adversely, then this might suggest that, before parent training is undertaken with this population, stimulant medication (as recommended by NICE) should be considered. The impact of comorbid child emotional problems, such as anxiety and depression, which, for example, reduce the effectiveness of some interventions for ADHD, is worth examining in the context of disruptive behaviour, as these problems have been found to sometimes diminish intervention effects (e.g. Beauchaine *et al.*[66]).

Parents' clinical and parenting characteristics

Parent depression

Parental depression is related to children's behavioural problems,[70] with as many as 50% of mothers of children with disruptive behaviour showing clinical levels of depression. Thus, many parents who participate in a parenting intervention to reduce children's conduct problems may suffer from mild to moderate levels of depression. Policy-makers and practitioners may worry that these families are harder to treat because of the complexity of the family problems. Earlier findings about the extent to which parental depression actually impacts parenting intervention effectiveness are highly inconsistent. Some suggest that families with parental depression are harder to treat,[44] whereas others suggest that parenting interventions particularly benefit families with higher levels of parental depression.[46,71] For example, skill deficits such as poor problem-solving and an inability to recall specific events are commonly associated with both depression and inadequate parenting of children with disruptive behaviour problems.[72,73]

Parenting behaviour

Parenting interventions aim to reduce children's conduct problems through improvement of parenting behaviour. Parents' knowledge and skills at the start of the intervention may impact the extent to which their parenting behaviour improves as a result of the intervention. Perhaps surprisingly, therefore, baseline levels of parenting behaviour are rarely studied as putative moderators of parenting intervention effectiveness. Alternatively, parenting behaviour may have been included in moderator analyses but not reported because of non-significant outcomes (i.e. reporting bias[74]). The present study seeks to determine more conclusively if parents' baseline levels of parenting skills impact the extent to which families benefit from a parenting intervention.

Contextual factors

Relevant contextual factors that the literature suggests are likely to affect implementation and effectiveness of parenting interventions can be coded at the level of the trial, because they are the same for all families within the same trial.[75,76] These include type of service provider organisation, level of professional training of the staff delivering the intervention, level of attention to fidelity of implementation, university efficacy compared with 'real-world' service effectiveness setting, and geographic factors (e.g. UK vs. other countries, city vs. rural). It is largely unknown whether any reduction in effectiveness is better accounted for by family-level variables (e.g. income and lone parenthood) or, even after taking these into account, whether there are still contextual factor effects. As contextual factors are measured at the trial level, power will be limited for these analyses. In addition, variation between trials on these factors may be limited. For example, the IY parenting intervention was mainly delivered in towns, with few groups in rural areas.

Methodological limitations of current moderation literature

Across these studies of child- and family-level risk factors as moderators of parenting intervention effects, there are common methodological issues arising. Data come from secondary moderator analyses of individual trials, from narrative synthesis of such findings across trials or from metaregression as part of an

aggregate-level meta-analysis of parenting trials. We first describe the current methods used in the literature, and then discuss how pooled individual-level data can overcome these drawbacks.

The authoritative paper by Lambert et al.[77] comparing these usual meta-analytic methods with pooling individual data concluded that: '[m]eta-analysis of summary data may be adequate when estimating a single pooled treatment effect or investigating study-level characteristics. However, when interest lies in investigating whether patient characteristics are related to treatment, individual patient data analysis will generally be necessary to discover any such relationships'. A conventional aggregate/summary data analysis approach is likely to be less powerful and miss important moderating effects, as it can detect effect moderation only by trial-level summaries rather than individual-level variables. For example, the mean age of children in a range of trials may be similar, and the average effect sizes may also be similar, so that using trial-level comparisons, such as metaregression or subgroup analysis of effect sizes, age would not be predictive of intervention effect. However, often within trials there is age variation, and, by combining them at an individual case level, we will have requisite information to see whether or not there is effect modification by age, and by the socioeconomic variables that we plan to investigate.

Second, as is well established from the statistical multilevel modelling literature, effects operating at the (aggregate) trial level need not be the same as those operating at the level of the individual. Basing inferences on the equality of between- and within-trial effects when in reality this is not the case is known as the ecological fallacy.[78] Thus, aggregate-level metaregression can inform us only about between-trial effects. We should not use these trial-level results to infer effect moderation by individual-level variables, such as child age, maternal depression, and so on; however, this is what most meta-analyses do. Our study will serve to illustrate the extent of this fallacy, as we will be able to directly compare trials and individual-level moderator effects.

A further problem that stems from the use of trial-level predictors of treatment effects is that predictors are prone to confounding by a common cause,[79] making it hard to interpret their meaning. An example of this can be seen in the subgroup analyses in the Furlong et al. review.[12] There, trials conducted in research settings, or with more affluent parents, were also those more likely to be conducted by the developer: all factors that tend to produce larger effect sizes. Thus, it is unclear which of these factors is the cause of the observed treatment effect moderation.

Finally, a second but less commonly used meta-analytic approach to investigating moderators is one that synthesises the published findings of predictor analyses from trial data across trials,[44] that is, a meta-analysis of predictor effects. This approach has the advantage of making use of within-trial variability in socioeconomic characteristics, therefore not committing the ecological fallacy and mitigating somewhat the problem termed as 'those confounded moderators' by Lipsey,[79] and does not require labour-intensive pooling of individual data. However, this approach also suffers from serious drawbacks, leading to researchers recommending against its use.[80,81] A key drawback is that most trials do not report predictor or moderator data, raising the possibility of reporting bias, or, at best, resulting in meta-analyses that can summarise only an incomplete picture. In addition, trial outcome data are rarely broken down by equity factors; when trials do test socioeconomic or other predictors of treatment effects, statistical models are specified in varying ways,[80] for example, some calculating interaction effects but others only within-group predictors, which renders synthesis meaningless.[23,82] These problems apply no less to parenting intervention trials,[46,50] and can be overcome by use of pooled data. Pooling individual-level data is an exciting new approach to data synthesis,[80,83] increasingly common in medicine in recent years,[84] but rarely used in public health or psychosocial fields. This study is the first to pool individual-level data from multiple independent trials of a parenting intervention (the IY Basic parenting series) in order to investigate moderator effects in a large, well-powered sample, and making use of individual rather than aggregate trial-level measurement of moderator variables.

Wider health benefits and potential harmful effects

In addition to assessing child disruptive behaviour, typically the primary outcome, parenting interventions may have wider benefits for family well-being. First, these include improving parenting skill and parent–child relationships, with increases in positive involvement with children, and reductions in harsh parenting and abusive practices. Although these are termed secondary outcomes in most trials, they are also seen as crucial mediators between intervention and outcome.[46] Second, programmes have been shown to improve adult mental health and well-being, including parental depression, confidence in their ability to be a successful parent and partner relationships. Third, some studies show generalisation to improved behaviour of other children in the family.[85]

The inclusion of these wider health benefit measures, however, is far from systematic across trials. It is often unclear why certain trials include some measures of wider health benefits, whereas others do not. Moreover, reporting bias may exist in that authors report only wider health benefit outcomes that were significantly altered by the intervention,[74] although the recent trend to publish trial protocols (e.g. Chhangur *et al.*[86]) will hopefully diminish this. Reporting bias is problematic because it may overestimate the effects that parenting interventions have on wider health outcomes. Alternatively, if relevant wider health benefits are not assessed, the policy impact of parenting interventions on family well-being may be underestimated. There is, thus, a need for a systematic investigation of the extent to which parenting interventions designed to reduce conduct problems improve family well-being more broadly. A systematic investigation would require (1) authors to share data on all measures of wider health benefits they included in the study and (2) sufficient numbers of participants for greater power and precision to estimate the magnitude of effects. This would allow for more conclusive results that include a wide range of possible health benefits (e.g. more elaborate than previous reviews, such as Barlow *et al.*[11]) and, that is, more up to date (e.g. building on Furlong *et al.*[12]).

Wider health benefits

Negative (i.e. harsh and inconsistent) parenting, and a lack of positive parenting (i.e. positive reinforcement and monitoring) are of particular importance for child well-being and quality of life; in prevention trials, in which many of the children show quite low levels of behaviour problems, these interventions impact public health by reducing levels of harsh or abusive parenting and family stress. This has been found in universal and selective prevention trials,[32,87] and in studies of parents at high risk for abusive parenting.[88,89] The Triple P trial[87] showed that widespread implementation of a similar parenting programme reduced admissions to hospital for abuse, measured by county-level indicators. Recent reviews confirm that even mildly harsh parenting is associated with harmful biological effects on children, including, for example, dysfunctional cortisol secretion patterns and raised C-reactive protein, which in turn are associated with increased cardiovascular disease and mortality.[90]

Parental well-being (i.e. depression or stress) has been shown in some trials to be improved by parenting interventions, and this will be an important public health benefit to document. Parents with young children spend many or most of their waking hours caring for them, and qualitative studies suggest that failure to succeed in controlling child behaviour is a major source of lack of confidence and depressive cognitions.[51] Improving parent–child dynamics, including more positivity and improved communication, may contribute to parents' sense of well-being and fulfilment in the parenting role. Some previous work suggests that parenting interventions designed to improve the parent–child relationship and children's conduct problems also reduce parental symptoms of depression.[12]

Although parenting interventions primarily aim to reduce children's conduct problems, there is evidence to suggest that they may have a wider impact on children's mental health, including children's ADHD symptoms, ADHD being the most prevalent comorbidity of conduct problems.[91] Recent findings about the extent to which parenting interventions reduce ADHD symptoms in children are inconsistent, and may in part depend on the type of instrument used (e.g. Daley *et al.*,[92] Jones *et al.*[68] and Sonuga-Barke *et al.*[93]). The extent to which parenting interventions designed for reducing conduct problems may benefit children's emotional

well-being, however, remains understudied. Symptoms of anxiety and depression are common in children with conduct problems.[94] Some studies evaluating the IY parenting intervention indicate effects of the intervention on reduced emotional problems in children (e.g. Herman *et al.*[95]). Others, however, have failed to replicate these findings (e.g. Leijten *et al.*[96]) or emotional problems are not measured. These inconsistent findings suggest that more thorough investigation is needed of the extent to which parenting interventions for conduct problems also impact children's emotional problems.

Potential harmful effects of parenting interventions

As well as benefits, it is important to consider potential harms, especially as they are rarely studied in parenting intervention trials. When harms have been studied in psychosocial trials, they have mainly involved youth-focused interventions and have been defined as main effects in the unintended direction.[97,98] Recent systematic reviews of parenting interventions have not found evidence of harmful effects defined in this way.[12] Parents rarely report potentially harmful outcomes in qualitative studies; Morch *et al.*[51] mention none, despite interviewing many parents for whom intervention was not successful. A few parents in the study by Furlong and McGilloway[50] were concerned about increased conflict with partners related to trying new parenting techniques and the lack of privacy in group interventions that discuss family problems. The Cochrane review by Furlong *et al.*[12] planned to examine two potential adverse effects, namely the burden on families in attending (e.g. child-care issues) and increased family conflict, but found no studies reporting these outcomes. Given this weak state of evidence on harms from parenting interventions, cautious, exploratory (hypothesis generating, rather than hypothesis driven) analyses are needed to see if there is evidence of main effects in the adverse direction.

Parent satisfaction data

Data on parents' views are rarely presented in detail in trial reports or reviews[12] and the rate of missing data is often high. These are available only for parents in the intervention group, and not the control group, but if numbers are sufficient, and instruments comparable, they could potentially be synthesised.

The Incredible Years parenting programme

The IY Basic parenting programme is a 12- to 14-session programme, delivered to groups of between 6 and 15 parents, in weekly sessions of 2–2.5 hours.[32] It has been widely rolled out throughout the UK, as well as in some other European countries. It has been identified in many systematic reviews[11,12,99] as effective for preventing disruptive behaviour in children, and improving parenting quality and parental mental health. The programme has received UK government funding for training as part of the Pathfinder Early Intervention programme and Welsh Government funding as part of the Parenting Action Plan for Wales, and as a result of its widespread dissemination, eight community-based RCTs have been completed [sited in London,[59,100,101] Plymouth,[101] Oxfordshire,[102,103] Wales[85,104] and Birmingham.[105]

The content of the programme[106,107] is derived from social learning theory and attachment theory. More specifically, the techniques that parents learn are designed to break coercive cycles of parent–child interaction in which parents and children reinforce negative and aggressive behaviour in each other.[108] The following topics are covered: relationship-building through playing or spending special time with the child, providing praise and rewards as reinforcements of positive behaviour, effective limit-setting, adequate disciplining techniques (e.g. ignore and time-out techniques), and coaching children in social, emotional and academic skills. The content of the IY parenting programme is very similar to the contents of programmes based on social learning theory (e.g. Parent–Child Interaction Therapy and Parent Management Training – the Oregon model).

What is different from other programmes, however, and potentially important for reaching and benefiting socioeconomically disadvantaged families, is the approach of the IY programme. As opposed to a didactic style in which the therapist teaches the parent how to change his or her behaviour, IY therapists (i.e. 'group leaders') use a collaborative approach in which parents are seen as the experts on their own children. Parents are guided to set weekly goals, which fit with their cultural and personal needs and values. Moreover, as opposed to the therapist 'talking' about the kind of parenting behaviour that is considered to be appropriate, video-taped scenes showing examples of parent–child interactions are central in the sessions and parents are guided to identify key parenting behaviours or principles that might be useful for their own family context. There are also discussions about programme-driven and parent-initiated topics, brainstorms about different parenting techniques and role-plays in which parents practise different techniques. There is also an important focus on home practice, and parents receive weekly check-in telephone calls. The group format is essential here, because the group leaders try to let most of the ideas and solutions come from the parents themselves. The group format further allows for normalisation and social support.

Rationale for the current study (PRISMA-IPD #3)

This study addresses the following broad question: how far could widespread dissemination of parenting programmes improve child disruptive behaviour and reduce social inequalities? Combining data sets from trials in different communities to establish for whom programmes are effective: an individual participant-level meta-analysis.

To answer this question, we combine individual participant data (IPD) from 14 randomised trials of the IY parenting intervention across Europe. Our study brings to bear four important design features that increase power, value and generalisability, compared with single trials and compared with conventional meta-analysis. First, we analyse RCTs, thus overcoming the risks of biased estimates of treatment effectiveness that may arise from observational studies of public health strategies, in which differences seen between populations do not necessarily translate into benefits when these are rigorously tested in intervention experiments.

Second, rather than combining trials at the study level as is usual in meta-analytic studies using aggregate data, we combine data from 14 trials at an individual participant level, thus greatly enhancing the opportunity to detect moderating (interaction) effects of social and socioeconomic disadvantage and other risk factors. A particularly key advantage is that combining individual-level data makes full use of the rich variation within (as well as between) trials. This information is completely lost when testing moderators in a conventional meta-analysis, for which participants can be characterised only at the aggregate trial level and not as individuals. Our design represents a unique opportunity to overcome the problems of low power and reporting bias that beset subgroup analyses in most individual trials, and hence to enhance our understanding of how parenting interventions may reduce or widen health and social inequalities.[82,109,110] Moreover, our study will illustrate the extent of this 'ecological fallacy', which can arise when aggregate-level data are used to infer individual-level effects, as we will be able to directly compare trials and individual-level moderator effects.

However, it is important to note that pooling data from multiple trials, although bringing unique advantages, also brings challenges and compromises, primarily in that different trials do not use the same measuring instruments, and variables thus need harmonising across trials. In conventional, aggregate-level meta-analysis, combining outcome data by converting aggregate scores to standardised mean differences is relatively straightforward but this is not applicable to individual-level data meta-analysis.[109]

Third, we draw upon qualitative research and public involvement, soliciting parents' views on factors affecting intervention success, to inform hypotheses for testing and help interpret the findings.

Fourth, by accessing a complete set of trial data, irrespective of whether or not findings from secondary outcomes have previously been published, and analysing using a consistent and preplanned strategy, we aim to obtain a less biased and more precise estimate of the wider benefits and potential harms of the intervention.

Fifth, we apply cost and cost-effectiveness approaches by health economists to enable potential benefits to public health to be predicted as accurately as possible.

This unique study based on pooled individual data is the largest of its type in the world and will considerably advance our understanding of differential intervention effects, and cost-effectiveness of parenting interventions for families with differing levels of social and socioeconomic disadvantage and child risk factors. This will help to determine whether or not such programmes are likely to reduce or widen social inequalities, which is an important public health question because of the damaging and expensive effects of disruptive behaviour. It is of direct relevance to the NHS, which is investing heavily in programmes to prevent disruptive behaviour. It will also generate a more precise estimate of the wider benefits of the IY parenting programme, which may be at least in part generalisable to the effects of other parenting programmes with a similar background in social learning theory. Perhaps most importantly, if there are groups for which programmes work less well, it will stimulate change in working practices to try to improve availability and effectiveness of these programmes for such groups.

By pooling all available baseline and outcome variables across trials, our analyses will help to reduce selective reporting and publication bias, whereby positive secondary outcomes are reported more than those showing null or harmful effects, or non-significant moderator analyses (if conducted) may potentially not be published.[80] Reporting bias is known to be a considerable problem in many areas of health care; systematic reviews find it to be linked to higher effect sizes.[111,112] It is especially problematic in this field, in which there are typically multiple secondary outcomes and multiple measures of the same construct within and between trials.[12]

Our analyses will also allow for wider generalisability across community service contexts, regions and countries. As the trials were conducted in a range of service settings (non-governmental organisations, Sure Start services, day nurseries and primary schools), samples and regions, inferences will be generalisable. This will potentially allow us to examine contextual effects on outcome, and their interaction with individual-level factors. However, it should be noted that power for examining contextual effects, as these operate largely at trial level, is likely to be very limited as the sample size is only as large as the number of trials.

Finally, our evidence is up to date. It is unlikely that we have missed any relevant European trials of IY, as we conducted extensive literature searches and contacting of experts, which revealed many trials, which, at the outset of the study, were recently completed and not yet published.

Research questions (PRISMA-IPD #4)

Underlying population, intervention, comparison, outcome, study design (PICOS) question for main effects in the pooled trials:

- Population: families with children aged 2–10 years
- Intervention: IY parenting programme
- Comparison: waiting list, minimal intervention or care as usual
- Outcome (benefit): child disruptive behaviour post test
- Study design: RCT.

Specific questions for this IPD meta-analysis:

- To what extent does the IY parenting intervention benefit the most socially disadvantaged families compared with average families?
- To what extent does the IY parenting intervention benefit families from ethnic minorities compared with those from the ethnic majority?
- To what extent does the IY parenting intervention differentially benefit children with different levels of characteristics, including age, gender, severity of conduct problems and comorbid problems, at baseline?
- To what extent does the IY parenting intervention differentially benefit children whose parents have different levels of depression and parenting skill at baseline?
- To what extent do trial-level effects predict outcome, including contextual variables (country, rural vs. urban) and intervention factors that indicate level of intervention fidelity (level of staff training, certification and supervision); and number of sessions offered?
- What are the wider public health benefits and potential harms of the IY parenting intervention?
- What are the costs, cost-effectiveness and potential longer-term savings of the IY parenting intervention?

Chapter 2 Methods

In this report, the numbered headings are structured according to the most relevant and up-to-date guidelines, the PRISMA-IPD Statement.[1]

Protocol and trial registration (PRISMA-IPD #5)

The protocol for this study is available on the National Institute for Health Research Public Health Research website (project number 12/3070/04).

Eligibility criteria (PRISMA-IPD #6)

We sought to include all completed RCTs of the IY parenting programme in Europe for children aged 1–12 years. Non-RCTs were excluded because no causal inference about the effects of the IY programme can be drawn from non-randomised designs. No restrictions were placed on the years in which trials were conducted, required minimum follow-up or included outcome measures. Within each RCT we included individuals who had received the IY (or a combination of IY and a reading intervention that focused on similar parenting behaviour) and individuals in control conditions. We excluded trials with additional non-parenting programmes, as well as excluding individuals who had received additional treatment for disruptive behaviour such as the IY child programme, because the focus of this project was to examine the effect of the parenting programme as a sole intervention. We excluded programmes that were much more minimal than the standard IY programme of 12–14 sessions, for example highly abbreviated non-standard versions. We also excluded individuals who received only a reading intervention.

Identifying studies: information sources (PRISMA-IPD #7)

Studies were identified in 2013 through (1) a systematic literature search in the following databases, Cumulative Index to Nursing and Allied Health Literature (CINAHL), EMBASE, Global Health, MEDLINE and PsycINFO; (2) the IY website, which provides information on trials evaluating IY; (3) the European IY mentors' network; and (4) asking experts. Searches in January 2015 revealed no further completed trials.

Identifying studies: search (PRISMA-IPD #8)

EMBASE, Global Health, MEDLINE (< 1946 to present) and PsycINFO were searched via Ovid using the following search terms:

1. incredible year$.mp
2. webster-stratton.mp
3. 1 or 2.

Cumulative Index to Nursing and Allied Health Literature was searched via EBSCO*host* using the phrase 'incredible years'.

Study selection processes (PRISMA-IPD #9)

Eligibility was assessed by the first author and double-checked by four additional authors (SS, JH, PL, JM). There were no differences of opinion.

Data collection process (PRISMA-IPD #10)

Anonymised data for 15 RCTs were requested for all families randomised. Investigators were first e-mailed to ask whether or not they would be willing to collaborate and share their data for this project. They were then sent a detailed guideline on how to anonymise their data and an overview of the variables we were hoping to collect. Investigators uploaded their data set to a University of Oxford web service that supports the exchange of large data files. The data transfer was encrypted and password protected. Files were deleted from the web server once the transfer had taken place. Raw (i.e. not recoded) individual-item level (i.e. not total scale scores) were supplied in SPSS format (SPSS statistics version 21; IBM Corporation, Armonk, NY, USA) and were checked for missing items and consistency with trial protocols and published reports. Copies of the original questionnaires were requested and received to check for consistent use of similar questionnaires across trials and the order in which questions were asked.

Individual participant data were available and received for all randomised participants in 14 trials (see flow diagram, *Figure 1*). Investigators were contacted in cases for which additional information was needed
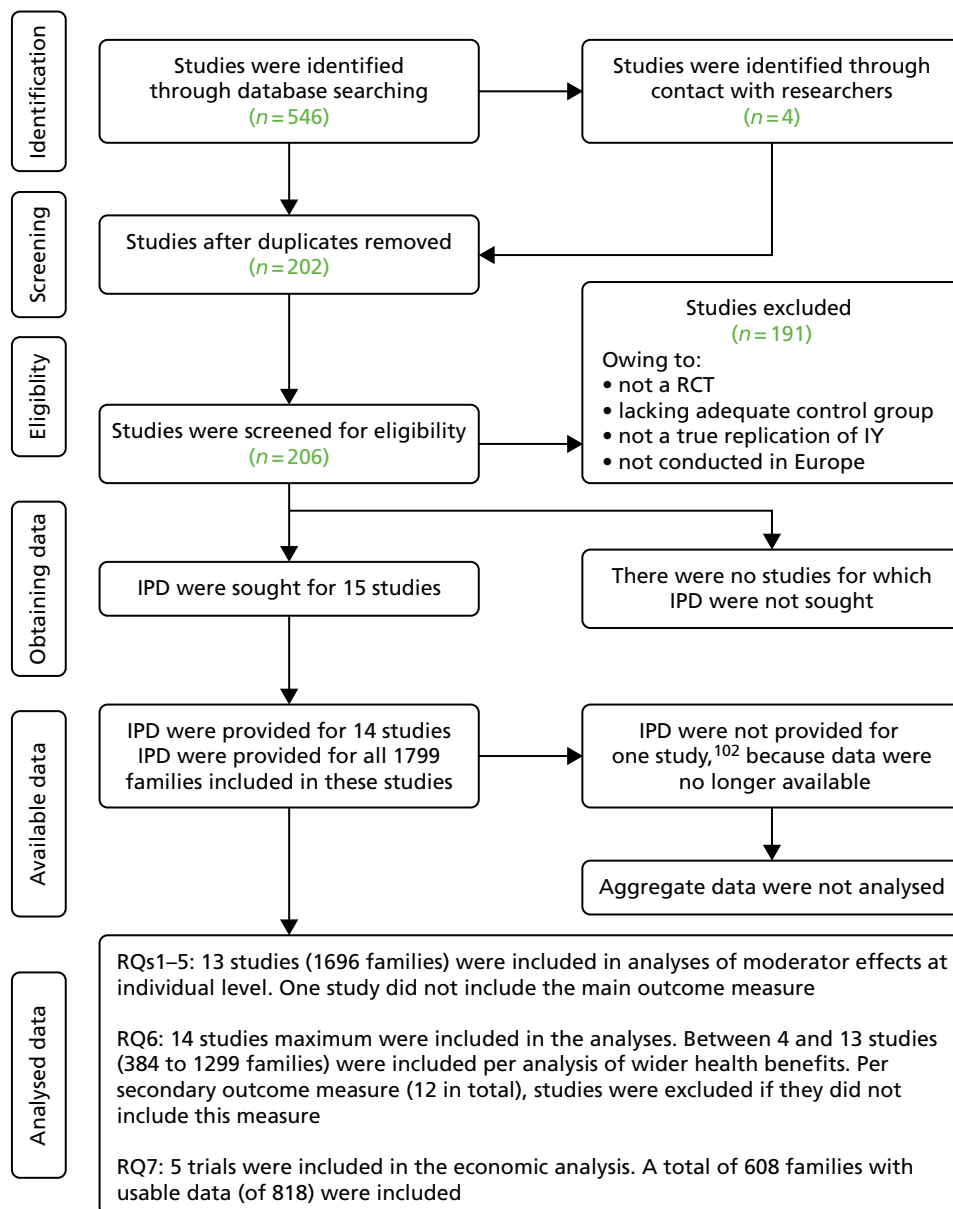
FIGURE 1 Flow diagram of search results. RQ, research question.

about the interpretation of the IPD. All investigators signed a data sharing agreement (see *Appendix 1*). IPD for the 15th trial[102] were reported by the investigators to be no longer available. The pooled data set consisted of records on 1799 families.

Overview of trials

An overview of the trials is listed in *Table 1*. Detailed information about trial characteristics is provided in the results (see *Chapter 3*, *Description of included trials*), including fuller tables (see *Table 7* and *Appendix 2*).

TABLE 1 Brief overview of trials

Trial number	Country	Trial acronym	n	Brief description	Reference
1	Norway	NOR	75	Referred children in psychiatric clinics	Larsson *et al.*[113]
2	Sweden	SWED	62	Referred children in psychiatric clinics	Axberg and Broberg[114]
3	Portugal	PORT	124	Parent-referred children in university clinics, screened for conduct problems	Homem *et al.*[115] and Azevedo *et al.*[116,117]
4	Ireland	IRE	149	Mixed community services, children screened for conduct problems	McGilloway *et al.*[48]
5	The Netherlands	NL-BS	99	Mothers released from incarceration (non-governmental organisation for former incarcerated mothers)	Menting *et al.*[118]
6	The Netherlands	NL-SES	156	Socioeconomically disadvantaged and immigrant families (clinics and community services)	Leijten *et al.*[96]
7	Wales	WL-SS	153	Sure Start services; preschoolers, screened for conduct problems	Hutchings *et al.*[85]
8	Wales	WL-FS	103	Flying Start services, toddlers in highly socioeconomically disadvantaged areas	Hutchings *et al.*[104]
9	England	BIRM	161	Mixed community services in Birmingham, children screened for conduct problems	Morpeth *et al.*[105] and Little *et al.*[119]
10	England	LON-SPO	112	Socioeconomically disadvantaged primary schools in London, children screened for conduct problems	Scott *et al.*[100]
11	England	LON-PAL	174	Socioeconomically disadvantaged primary schools in London	Scott *et al.*[59]
12	England	LON-HCA	214	Primary schools in London and Devon, children screened for conduct problems	Scott *et al.*[101]
13	England	OXF	76	Referred children in voluntary sector service, children screened for conduct problems	Gardner *et al.*[103]
14	England	LON-NHS	141	Referred children in NHS psychiatric clinics	Scott *et al.*[120]
15	England	–	116	Data not available, children screened for conduct problems in general practice and NHS	Patterson *et al.*[102]

Individual patient data integrity (PRISMA-IPD A1)

We checked each trial data set to identify missing data and assess data validity (e.g. we double-checked individual item formulation and scores for all constructs). To assess randomisation integrity, we checked patterns of treatment allocation and balance of baseline characteristics by treatment group. Any queries were resolved in collaboration with each trial investigator.

Risk of bias assessment in individual studies (PRISMA-IPD #12)

We used the Cochrane risk of bias assessment tool.[121]

Harmonisation of individual-level data

Three different data harmonisation strategies were used.

1. Combining similar classification systems to harmonise data for different socioeconomic status (SES) indicators. Indicators of SES were screened for comparability. Fortunately, most trials used similar indicators that were operationalised in similar ways. For example, whether or not a family had low income was defined by receiving financial benefits (10 trials), receiving financial benefits and having below-median income (one trial), scoring below the low SES threshold on the Hollingshead Index (one trial) or living in social housing or with family/friends (two trials). Indicators of families' (risk for) low income were dichotomised. Educational-level categories used across trials were compared with the United Nations Educational, Scientific and Cultural Organization Institute for Statistics *International Standard Classification of Education 2011*.[122] Although some categories had to be combined (e.g. less than primary education and primary education) because these had already been combined in some trial data sets, five main categories (primary education or less, lower secondary education, upper secondary education, postsecondary education and university degree) were present in data from all trials and used in the final pooled data set.
2. Using norm deviation scores to harmonise scores on child behaviour problems (disruptive, ADHD and emotional problems) and parental depression. A primary measure (i.e. the most frequently used measure) was selected for each construct. If data on the primary measure were unavailable, data from similar measures were converted into scores on the primary measure using norm deviation scores (i.e. number of SDs the individual scores are above or below the population mean). This approach assumes that both instruments measure the same construct with the same measurement error on different instruments and thus the scores can be converted using known population characteristics. The advantages of using norm deviation scores are that (1) absolute scores can easily be interpreted, because they are on the original scale of the primary measure – this allows for interpretation of clinical significance of intervention effects – and (2) the scores of individuals remain the same after adding data from new trials, because harmonisation is done on an individual family level.
   In contrast, integrative data analysis using latent variables based on the different measures included across trials, although strong in its use of multiple measures within trials, has the disadvantage of scores that are hard to interpret on an absolute level, as well as scores that depend on the model tested and thus the trials or families included in the model. Norm deviation scores were used to harmonise data from the following constructs: parent-reported disruptive child behaviour, children's comorbid ADHD symptoms, children's comorbid emotional problems and parental depression.
3. Using item-level harmonisation based on face validity and correlations to harmonise scores on self-reported parenting practices. When no norm scores were available for measures, scores on similar items across scales that aimed to measure the same construct were selected. Response scales were harmonised to reflect the response scale used most frequently. For example, if most measures used a 1–7 Likert scale, scores from a measure using a 0–3 Likert scale were converted such that 0 = 1, 1 = 3, 2 = 5, 3 = 7. This approach makes assumptions based on face validity: items that are formulated similarly will measure a similar construct.

Relatively similar response scales were used across the instruments. Response scales varied between 'never and always' or 'not at all likely' and 'extremely likely'. For the items in relation to the last 2 days, response scales vary from 'never' to 'not with my child in the last 2 days' and from 'none' to 'more than 4 hours'.

Somewhat different time periods were covered by the different measures. The Parenting Scale (PS) asks parents to reflect on the last 2 months; the Parenting Practices Inventory (PaPI) asks parents 'how often do you do each of the following?' or 'within the last 2 days how many times did you?' or 'about how many hours in the last 24 hours did . . .?' or 'within the last 2 days, about how many total hours was your child . . .?' and 'what percentage of the time do you know . . .?'. The Alabama Parenting Questionnaire (APQ) asks parents to reflect on what typically occurs at their home. The interview used in trials 10–12 asks parents to reflect on what occurred in the last week (rewards) or yesterday (praise). The interview used in trial 14[120] did not define a time point. We were able to check actual correlations between scores from different measures for a few constructs based on a small sample ($N = 44$). These varied between 0.30 and 0.87. We used item-level harmonisation for the following constructs: self-reported parenting practices (seven constructs).

Data items (PRISMA-IPD #10)

Design variables: trial level

Trial identifier
Indicating the trial from which the observation is taken.

Family identifier
Denoting the family number for each observation.

Type of control condition
Whether or not the trial used a waiting list (10 trials), no treatment control condition (two trials) or a minimal intervention control condition (two trials).

Incredible Years sessions offered
Number of sessions of the intervention offered to the participant.

Design variables: individual level

Randomisation ratio applied to each participant
Where the randomisation ratio varied with the trial, this variable denotes the batch of randomisations in which the participant was randomised.

Cluster used for randomisation
Trials 11 and 14 used cluster randomisation. This variable denotes the cluster to which the participant belonged.

Stratification variables used within the trial
If stratified randomisation was used, this variable denotes those variables that were used in the stratification [e.g. trial site, sex, age, recruitment cohort and/or score above the Eyberg Child Behavior Inventory Intensity scale (ECBI-I) 97th percentile].

Baseline measures: individual family level

Child gender
Gender of the target child was coded as male or female.

Child age

Age of the target child at baseline was described in months.

Primary parent gender

Gender of the child's primary caregiver was coded as male or female.

Primary parent age

Age in years of primary caregiver at birth of target child was described in years. Primary parent is defined as the parent responsible for the majority of the care of the target child.

Secondary parent gender

Gender of the child's secondary caregiver was coded as male or female.

Secondary parent age

Age in years of secondary caregiver at birth of target child was expressed in years.

Child was referred

Denoting whether or not the child was referred to the service for behaviour problems.

Low income

Indicators of families' (risk for) low income were dichotomised. Low income was defined as receiving income-dependent financial benefits (10 trials: trials 3, 4 and 7–14); receiving financial benefits or having below-median income if financial benefits were not income dependent (one trial: trial 1); scoring below the low SES threshold on the Hollingshead Index (one trial: trial 2); or living in social housing or with family/friends (two trials: trials 5 and 6). Categorised as a binary variable (0 = no; 1 = yes).

Educational level

Highest educational level of the primary parent. Educational level was categorised according to an amended version of the *International Standard Classification of Education 2011*:[122] 1 = primary education or less, 2 = lower secondary education, 3 = upper secondary education without qualifications, 4 = post-tertiary education and 5 = bachelor-, master's- or doctoral-level education. In addition, a binary variable was created because of low numbers in some of the educational-level categories in some of the trials: 0 = upper secondary-, tertiary- or university degree-level education; 1 = primary or lower secondary educational status.

Lone parenthood

The primary parent does not live with a partner or spouse. Categorised as a binary variable (0 = no; 1 = yes).

Teenage parenthood

Primary parent was younger than 20 years at birth of the target child. Categorised as a binary variable (0 = no; 1 = yes).

Parental unemployment

No employed parent in the household. Categorised as a binary variable (0 = no; 1 = yes).

Ethnic minority status

Binary categorisation of the primary parent's ethnic background into ethnic majority status or ethnic minority status based on the adapted Office for National Statistics (ONS)[123] classification (see *Ethnic background*). Value 1 (white) was scored as 0 = ethnic majority status. All other values (2–10) were scored as 1 = ethnic minority status.

Ethnic background

Categorisation of the primary parent's ethnicity according to adapted ONS[123] classification. We adapted the ONS classification to include categories that are relatively common in our pooled data set, such as Mediterranean, but were not included in the original ONS classification. This variable included the following categories: 1 = white, 2 = black, 3 = Middle Eastern, 4 = South-East Asian/Chinese, 5 = Indian, 6 = Pakistani, Bangladeshi, 7 = Arabian (North African), 8 = Mediterranean (Turkish, Greek and Italian), 9 = Latin American and 10 = other.

One trial (trial 12) had many missing data on parent ethnicity. We therefore used child ethnicity to approach parent ethnicity (e.g. if a child was coded as white, we coded the primary parent as white). However, this was done only if (1) the child was coded as white or black (e.g. not when child was coded as 'other' or 'mixed') and (2) the child and parent were known to be biologically related.

Baseline measures: trial level

Urban or rural

The percentage of therapy groups within the trial that were held in a rural setting.

Country

Whether the trial was held in the UK or Ireland (coded as 0) or in another European country (coded as 1).

Efficacy or effectiveness

Type of clinical setting. Whether the trial was carried out under optimal conditions (efficacy, coded as 1) or under a real-world context (effectiveness, coded as 0).

Percentage of staff who were clinically trained

Percentage of the staff within the trial who provided the therapy groups and had formal clinical training and education, for example in clinical psychology, psychiatry or mental health nursing.

Percentage of staff Incredible Years certified

Percentage of the staff within the trial who provided the therapy groups and were formally certified as an IY group leader.

Service context

Whether the therapy groups were held within health services (e.g. NHS or a similar service in other countries) (coded as 1) or in community or voluntary sector services (coded as 0).

Type of trial

Whether the trial was a selective prevention trial (i.e. families targeted were at risk of conduct problems but not necessarily currently experiencing problems, coded as 0) or a indicated prevention or treatment trial (i.e. children with reasonably high levels of conduct problems were targeted, coded as 1).

Checklist

Did staff complete the IY checklist after sessions (coded 1 for yes and 0 for no)?

Mentor

Whether or not a mentor was part of the trial (coded 1 for yes and 0 for no).

Video

Were sessions video-taped (coded 1 for yes and 0 for no)?

Video supervision

Were video-taped sessions used in supervision (coded 1 for yes and 0 for no)?

Supervision
Was there weekly/fortnightly supervision (coded 1 for yes and 0 for no)?

Fidelity
Did independent ratings of session fidelity take place (coded 1 for yes and 0 for no)?

Workshop
Whether or not any group leaders in the trial attended an international workshop (coded 1 for yes and 0 for no).

Number of Incredible Years sessions offered
Number of IY sessions offered in the active arm.

Primary outcome: individual family level

Disruptive child behaviour
The ECBI-I[124] was used to assess disruptive child behaviour, primarily conduct problems. The ECBI-I is a widely used 36-item measure that rates parent-reported frequency of disruptive child behaviour on a 7-point scale. The ECBI-I has shown good convergent[125] and discriminant validity.[126,127] If a trial did not include the ECBI-I we chose the measure that best captured the same construct and used norm scores to convert scores on the alternative measure to ECBI-I scores (see *Harmonisation of individual-level data*). Three trials did not include the ECBI-I (trials 3, 8 and 14). For two of these trials (trials 3 and 14) scores on the Parental Account of Children's Symptoms (PACS[128]) were converted into ECBI-I scores using norm deviation scores. One trial (trial 8) did not include a measure of parent-reported disruptive child behaviour because of the young age of the children.

Parental Account of Children's Symptoms scores and ECBI-I scores correlated ($r = 0.71$) in our sample, based on data from four trials (trials 10–13) that included both the ECBI-I and PACS. The internal consistency of ECBI-I scores was $\alpha = 0.94$ at time point 1 (T1) and $\alpha = 0.95$ at time point 2 (T2). The internal consistency of the PACS scores was 0.82 at T1 and 0.79 at T2. Data were available from 13 trials (all except trial 8).

Secondary outcomes and baseline moderators: individual family level

Comorbid attention deficit hyperactivity disorder symptoms
The Strengths and Difficulties Questionnaire (SDQ) was used to assess parent-reported comorbid ADHD symptoms in children. If a trial did not include the SDQ hyperactivity/inattention subscale, we chose the measure that best captured the same construct (child ADHD symptoms) and used norm scores to convert scores on the alternative measure to SDQ scores.

Original SDQ scores were used for trials 2–4, 6, 7, 9–11 and 14. Scores were converted for trials 1 (from the Child Behavior Checklist, CBCL[129]), 12 (PACS) and 13 (CBCL). Original US male aged 4–18 years norm scores[129] were used to convert CBCL scores into norm deviation scores. Original UK norm scores[130] were used to convert PACS scores into norm deviation scores. American male aged 4–17 years norm scores were used to convert norm deviation scores into SDQ scores.[131] Trials 5 and 8 did not have a parent-reported measure of child ADHD.

Converted SDQ scores above the maximum or below the minimum possible score on the SDQ scale were altered to the maximum or minimum possible score, respectively. If, after harmonising, scores were outside the theoretical range (0–10), they were changed to either the theoretical minimum ($n = 0$ at all time points) or maximum ($n = 68$ at T1 and $n = 40$ at T2).

Comorbid emotional problems

The SDQ was used to assess parent-reported comorbid emotional problems in children. If a trial did not include the SDQ emotional problems subscale, we chose the measure that best captured the same construct (children's emotional problems) and used norm scores to convert scores on the alternative measure to SDQ scores.

Original SDQ scores were used for trials 2–4, 6, 7 and 9–11. Scores were converted for trials 1 (from the CBCL[129]), 12 (PACS), 13 (CBCL) and 14 (PACS – for participants for whom SDQ scores were missing). Original US male aged 4–18 norm scores[129] were used to convert CBCL scores into norm deviation scores. Original UK norm scores[130] were used to convert PACS scores into norm deviation scores. American male aged 4–17 years norm scores were used to convert norm deviation scores into SDQ scores.[131] Trials 5 and 8 did not have a parent-reported measure of children's emotional problems.

If, after harmonising, scores were outside the theoretical range (0–10), scores were changed to either the theoretical minimum ($n = 21$ at T1; $n = 7$ at T2) or maximum ($n = 12$ at T1; $n = 4$ at T2).

Parental depression

The Beck Depression Inventory (BDI)[132] was used to assess parental depressive symptoms. The BDI is a 21-item measure of depressive symptoms and has shown good concurrent and convergent validity.[133]

If a trial did not include the BDI we chose the measure that best captured the same construct (parental depression) and used norm scores to convert scores on the alternative measure to BDI scores. BDI scores were included for trials 1, 3, 4, 7, 8, 13 and 14. More specifically, BDI version IA scores were included for trial 13 and BDI version II scores were included for trial 8. Scores were converted using norm deviation scores for trials 2 (from the Brief Symptom Inventory – depression subscale;[134] see Francis *et al.*[135] for norm scores), 5 and 6 (from the Symptoms checklist – depression subscale;[136] see same reference for norm scores), 10 and 11 (from the General Health Questionnaire;[137] see Booker and Sacker[138] for norm scores) and 12 (from the Depression Anxiety Stress Scale;[139] see Crawford and Henry[140] for norm scores).

Internal consistency for the BDI was 0.93 at T1 and 0.93 at T2. Internal consistency of the General Health Questionnaire was 0.86 at T1 and 0.87 at T2. Internal consistency of the Symptoms checklist – depression subscale was 0.93 (T1 only).

Parental stress

The Parental Stress Index Short Form (PSI-SF)[141] was used to assess symptoms of parental stress. The PSI-SF is a 36-item measure of parental stress. Data were available from trials 1, 4, 7, 8 and 10. Internal consistency of the PSI-SF was 0.95 at T1 and 0.96 at T2. Trials 3 and 6 used a subset of the items of the PSI-SF. Data from these trials were therefore not included.

Parental self-efficacy

Parental Sense of Competence (PSOC) scale[142] was used to assess parental self-efficacy. The PSOC scale is a widely used 16-item measure of parental self-reported self-efficacy scored on a 6-point scale. The PSOC scale was used in trials 3, 8, 11 and 13.

Trials 3, 8 and 11 used a 5-point rating scale for the PSOC scale and trial 13 used a 6-point rating scale. Trials 3, 8 and 11 were therefore recoded to a 6-point scale using the following recoding: $1 = 1$, $2 = 2.25$, $3 = 3.5$, $4 = 4.47$ and $5 = 6$.

Trial 12 included parental self-report data on 'confidence in managing child behaviour' but it was decided not to include these data, as they consisted of one item on a 6-point scale. Internal consistency on the PSOC scale total score was 0.84 at T1 and 0.90 at T2.

Self-reported positive parenting practices (use of praise, use of tangible rewards, and monitoring)

Across measures, items theoretically fitted three different constructs of positive parenting (*Table 2*). Praise was defined as any verbal compliment in response to the child's behaviour. Tangible rewards were defined as any rewards for the child that are not verbal or physical, for example privileges, stickers on a chart, special food, small toy or money. Monitoring was defined as parental supervision and knowledge of the child's whereabouts when the child is out of the parents' sight, including knowing the child's friends.

To assess positive parenting practices, four different instruments were used: PaPI (trials 1, 3, 6 and 10), APQ (trials 5 and 12), PS (one item: trials 3, 7, 9 and 13) and interview version 1 (trials 10–12 and 14, although items often differed across these four trials). Please see *Appendix 3* and *Harmonisation of individual-level data* for full details on items and how data were harmonised at an item level.

Several trials included multiple instruments: trial 3 had data on both the PS and the PaPI, although > 50% of data on the PaPI were missing. PS data were therefore used when available. Trial 10 had data on both the PaPI (selected items only) and the interview. PaPI data were used when available. Trial 12 had data on both the APQ (selected items only) and the interview. APQ data were used when available.

The most frequently used instrument was the PS. This instrument provides scores on a 7-point Likert scale. Scores from other instruments were therefore converted to a 7-point Likert scale. For the APQ, for example, scores are on a 5-point scale. These were converted into a 7-point scale using $1 = 1$, $2 = 2.5$, $3 = 4$, $4 = 5.5$, $5 = 7$.

Whenever possible, items selected were based on the original subscales of the instruments. Details are provided in *Table 2* regarding which items were included from which instrument. Internal consistency was sometimes low (the lowest was 0.34), often when there was a limited number of items. When more items were included, internal consistency went up to 0.75 on the PaPI and 0.99 on the APQ.

Self-reported negative parenting practices (corporal punishment, harsh threatening, laxness and shouting)

Across measures, items theoretically fitted four different constructs of negative (i.e. harsh and/or inconsistent) parenting (*Table 3*). Corporal punishment was defined as any physical punishment. Threatening was defined as threatening to punish the child (but not really punishing him/her). Laxness was defined as when the parent intended to punish the child but did not follow through or let the child get

TABLE 2 Overview of included positive parenting constructs and original instruments' subscales

Instrument	Construct	Original subscales
PS	Monitoring (one item)	Items was an extra item for the total score on the PS and was not part of any subscale
PaPI	Praise (two items)	Items are part of 'praise and incentives' subscale
	Tangible reward (four items)	Items are part of 'praise and incentives' subscale
	Monitoring (five items)	Items are part of the 'monitoring' subscale
APQ	Praise (four items)	Items are part of 'positive parenting' subscale
	Tangible reward (one item)	Items are part of 'positive parenting' subscale
	Monitoring (ten items)	Items are part of 'poor supervision' subscale
Interview	Praise (one item)	No subscales
	Tangible reward (one item)	No subscales
	Monitoring (one item)	No subscales

TABLE 3 Overview of included negative parenting constructs and original instruments' subscales

Instrument	Construct	Original subscales
PS	Corporal punishment (one item)	Item is from the 'hostility' subscale
	Threatening (two items)	Items are not part of any subscale
	Laxness (five items)	Items are from 'laxness'. Only items included that matched items from other measures
	Shouting (five items)	Items are from the 'over-reactive' and 'hostility' subscale. Some items are not part of any subscale
PaPI	Corporal punishment (PPI – six items)	Identical subscale
	Threatening (PPI – three items)	Items are part of 'harsh and inconsistent discipline' subscale
	Laxness (five items)	Items are part of 'harsh and inconsistent discipline' subscale
	Shouting (five items)	Items are part of 'harsh and inconsistent discipline' subscale
APQ	Corporal punishment (three items)	Identical subscale
	Threatening (one item)	Item is part of 'consistency subscale'
	Laxness (five items)	Items are part of 'consistency subscale'. Only items included that matched items from other measures
	Shouting (two items)	Items were not part of any subscale
Interview		N/A

N/A, not applicable.

away with the disruptive behaviour. Shouting was defined as any raising of the voice, shouting, scolding, use of bad language, swearing or saying mean things.

To assess negative parenting practices, four different instruments were used: PaPI (trials 1, 3, 6 and 10), APQ (trials 5 and 12), PS (trials 3, 7, 9 and 13) and interview version 1 (trials 10, 11, 12 and 14). See *Harmonisation of individual-level data* for how data were harmonised at an item level.

Several trials included multiple instruments: trial 3 had data on both the PS and the PaPI, although > 50% of data on the PaPI were missing. PS data were therefore used when available. Trial 10 had data on both the PaPI (selected items only) and the interview. PaPI data were used when available. Trial 12 had data on both the APQ (selected items only) and the interview. APQ data were used when available.

The most frequently used instrument was the PS. This instrument provides scores on a 7-point Likert scale. Scores from other instruments were therefore converted to a 7-point Likert scale. For the APQ, for example, scores are on 5-point scale. These were converted into a 7-point scale using $1 = 1$, $2 = 2.5$, $3 = 4$, $4 = 5.5$, $5 = 7$.

Whenever possible, items selected were based on the original subscales of the instruments. Details are included below regarding which items were included from which instrument. Internal consistency was sometimes low (lowest was $\alpha = 0.41$), often when there was a limited number of items. When more items were included internal consistency went up to $\alpha = 0.69$ on the PS, $\alpha = 0.84$ on the PaPI and $\alpha = 0.61$ on the APQ.

We were able to compute correlations between the PS and the PaPI based on a small sample ($n = 44$) from one trial (trial 3). Correlations were small for pretest scores on threatening and laxness (0.30 and 0.34). However, all correlations on other time points and for all other constructs of negative parenting were more satisfactory, ranging from 0.53 to 0.87.

Parent satisfaction data

See *Appendix 4* for full details. There were many missing data, at both trial and family levels. Five trials used comparable instruments, but with a high percentage of missing data (45%); hence these were not further analysed or discussed.

Overview

In summary, our data resource provided more or less complete information on a large number of baseline variables acting as putative treatment effect moderators. Some of these variables targeted the same domain. *Tables 4* and *5* provide an overview of moderator variables and their domains, and also for each domain identify a variable that might be considered as a representative on the basis of completeness.

Availability of data across trials

The number of data available varied across constructs (*Table 6*). For the demographic variables, data were available from all of the trials and almost all of the families (e.g. 100% of the data available on child gender and 95% of the data available on low income), except for data about the secondary parent

TABLE 4 Individual-level moderators at baseline by domain for 13 trials used in moderator analyses

Domain	Variable	Representative	Sample size (maximum 1696)	Number of trials with information available	Scale: binary vs. continuous
SES	Low income	Yes	1614	13	Binary
	Parental education level	No	1561	13	Ordinal
	Unemployment	No	1303	11	Binary
	Lone parent	No	1606	13	Binary
	Teenage parent	No	1609	12	Binary
Child age	Child age	Yes	1682	13	Continuous
Child gender	Child gender	Yes	1696	13	Binary
Child problem severity	Baseline ECBI-I	Yes	1622	13	Continuous
	Baseline ADHD	No	1532	11	Continuous
	Baseline emotional problems	No	1340	11	Continuous
Parental mental health	Baseline parental depression	Yes	1395	11	Continuous
Negative parenting	Corporal punishment	Yes	1393	10	Continuous
	Threat	No	999	9	Continuous
	Laxness	No	978	9	Continuous
	Shouting	No	967	9	Continuous
Positive parenting	Praise	Yes	630	6	Continuous
	Tangible rewards	No	625	6	Continuous
	Monitoring	No	1088	9	Continuous
Ethnicity	Ethnic minority	Yes	1651	13	Binary

TABLE 5 Trial-level moderators by domain

Domain	Trial sample size	Comment	Scale	Mean (SD) or number
Any rural sites	13		Binary	2 rural, 11 not
Geographical region	13	UK vs. non-UK	Binary	5 non-UK, 8 UK
Efficacy or effectiveness	13	Effectiveness (yes/no)	Binary	12 efficacy, 1 effectiveness
Percentage staff IY certified	13		Continuous	Mean 29.57 (SD 32.89)
Percentage of staff clinically educated	13		Continuous	Mean 55.571 (SD 38.392)
Service provider	13	Non-clinical setting (yes/no)	Binary	11 non-clinical, 3 clinical
Type of trial	13	Treatment or prevention	Binary	10 treatment/indicated prevention, 3 selective prevention
Did staff complete IY checklist after sessions?	13		Binary	13 yes
Was IY mentor part of the trial?	13		Binary	9 yes, 4 no
Were sessions video-taped?	13		Binary	13 yes
Were video-taped sessions used in supervision?	13		Binary	13 yes
Was there weekly/fortnightly supervision?	13		Binary	9 yes, 4 no
Did independent ratings of session fidelity take place?	13		Binary	3 yes, 10 no
Did any of team attend international IY workshop/training?	13		Binary	9 yes, 4 no

TABLE 6 Summary of variables and missing data

Variable name	Sample size (maximum $n = 1799^a$)	Per cent of applicable data available	Trials with relevant information missing[b]	Type
Child gender	1799	100		Binary
Child age	1785	99.2		Continuous (in months)
Parent gender	1777	98.2		Binary
Parent age (years)	1693	94.1		Continuous
Parent age at birth of target child	1680	93.4		Continuous
Second parent gender	916	75.4	2, 5 and 7	Binary
Second parent age	746	59.7	2, 5, 7 and 8	Continuous
Low income	1717	95.4		Binary
Education level	1664	92.5		Ordinal
Lone parent	1709	95		Binary
Teenage parent	1708	94.9		Binary

TABLE 6 Summary of variables and missing data (*continued*)

Variable name	Sample size (maximum *n* = 1799[a])	Per cent of applicable data available	Trials with relevant information missing[b]	Type
SES unemployed	1393	77.4	6 and 7	Binary
Baseline ECBI-I	1622	90.2	8	Continuous
ADHD	1532	85.2	5, 8 and 10	Continuous
Emotional problems	1340	74.5	5, 8, 10 and 12	Continuous
Parental depression	1395	82.3	9	Continuous
Positive parenting				
Praise	630	66.0	7, 9, 13 and 14	Continuous
Tangible rewards	625	65.5	7, 9, 13 and 14	Continuous
Monitoring	1088	90.7	10	Continuous
Negative parenting				
Corporal punishment	1393		2, 4 and 8	Continuous
Threatening	999	92.5	2, 4, 8, 11 and 14	Continuous
Laxness	978	90.5	2, 4, 8, 11 and 14	Continuous
Shouting	967	78.5	2, 4, 8, 10, 11 and 14	Continuous
Ethnic minority	1754	97.4		Binary
Ethnic background	1705	94.8		Categorical
Ethnic country	753	41.9	4, 7 and 11–14	Binary
Trial-level factors				
Urban or rural	14	100		Binary
UK/non-UK	14	100		Binary
Efficacy or effectiveness	14	100		Binary
% staff certified	14	100		Continuous
% staff clinically trained	14	100		Continuous
Service context	14	100		Binary
Type of trial (treatment vs. prevention)	14	100		Binary
Supervision	14	100		Binary
Checklist				Binary
Mentor	14	100		Binary
Video	14	100		Binary
Video supervision	14	100		Binary
Supervision	14	100		Binary
Fidelity	14	100		Binary
Workshop	14	100		Binary
Number of sessions offered	14	100		Continuous

a *n* refers to the number of participants/trials for which the data are available, and the proportion of individuals/trials for whom the data are available, of those to whom it is applicable (e.g. information on the second parent is applicable only if the primary parent is not a lone parent).

b Trials in which the variable is missing for all participants are also recorded in the table, in column 4. When this column is blank, then all trials had data on this variable.

(e.g. 75% of the data on the secondary parent's gender were available). For the main outcome variable (reduced disruptive child behaviour as measured with the ECBI-I), after harmonisation, data were available for 92% of the families.

The number of data available on the wider health benefits was more varied because most trials included only some of these. Nevertheless, data from almost all trials were available on parental depression and parental use of corporal punishment. Constructs on which data were scarcer include parental feelings of self-efficacy and parental laxness. *Table 17* in *Chapter 4* on wider health benefits of the IY programme includes a detailed table on the constructs that were available in each of the trials.

Risk of bias across studies (PRISMA-IPD #15)

See *Chapter 3*, *Risk of bias across studies*.

Analytic plan part 1: moderation analyses

Specification of outcomes and (interaction) effect measures (PRISMA-IPD #13)

The primary objective of the moderation analyses was the evaluation of putative moderation effects. We use a single outcome measure: disruptive behaviour as measured by the ECBI-I scale (harmonised from the PACS data in those trials for which the PACS scale was used in place of ECBI-I). This is available at up to three time points:

- Baseline: all trials measured child conduct disorder pretreatment and this information is available in both the treated and control arms, although there are some missing data.
- T1: 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.
- T2: The second follow-up time point. The time window for this measure was between 4 and 7 months after the end of the intervention. In NL-BS and LON-HCA data are available for both treated and control participants, as the whole sample was followed up twice. In other trials, when there is a second follow-up, data are available only for the active arm. T2 data were used only to improve imputation in the multiple imputation (MI) analyses, and not for evaluating outcomes, because the randomised comparison group was at that point no longer retained.

Notably, the child outcome is missing at all time points for trial 8 because these data were not collected in the trial. Therefore, this trial had to be excluded from all moderator analyses, although it was included in the wider benefits analysis. In particular, our goal was to understand whether the IY intervention is less or more effective for individuals who vary on one of the following putative moderator variables:

Individual-level moderators at baseline (pre randomisation):

1. SES indicators:

   - Low income: whether or not the family has or is at risk for low income.
   - Lone parent: primary parent is a lone parent.
   - Teenage parent: primary parent was aged < 20 years at birth of target child.
   - SES unemployed: there is no employed individual in the household.
   - Primary parent education level: highest education level attained by the primary parent. Dichotomised as lower secondary or lower and upper secondary and higher.

2. Ethnic minority: whether or not the primary parent is white.

3. Child clinical and demographic variables:

  - Child age: age in months of target child at baseline.
  - Child gender: gender of the target child.
  - Baseline severity: child conduct problems as measured by the ECBI-I or harmonised ECBI-I scale at baseline.
  - ADHD comorbidity: child's ADHD symptoms as measured by the SDQ subscale of hyperactivity/inattentiveness.
  - Emotional problems comorbidity: level of child's emotional problems as measured by the SDQ subscale of emotional problems.

4. Parent clinical and demographic variables:

  - Parental depression: level of depression of the primary parent at baseline.
  - Negative parenting measures:

    - corporal punishment
    - threatening behaviour
    - laxness
    - shouting.

  - Positive parenting measures:

    - praise
    - tangible rewards
    - monitoring.

Trial-level moderators at baseline (pre randomisation):

5. Contextual variables:

  - UK versus non-UK: whether or not the trial was conducted in the UK. There are seven UK trials from England and Wales within the pooled data set and six 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.

6. % certified: variable denoting the percentage of individuals delivering the therapy who are professionally certified.
7. % clinically trained: variable denoting the percentage of therapists delivering the intervention who have been clinically trained.

  - Average number of sessions offered by trial design.
  - % staff certified: percentage of staff delivering the IY intervention at a trial level that were IY certified.
  - % Staff clinically trained: percentage of staff at a trial level that were clinically trained (i.e. social worker, psychologist, nurse or psychiatrist).
  - Trial mentor: whether or not an official IY mentor was part of the staff of the trial.
  - Type of trial: binary coded for treatment or prevention trial.

Initially each of these putative moderator variables will be considered individually as a moderator effect, that is, as a two-way interaction with the treatment effect in the analysis model. We then proceeded to condition the analyses of significant treatment effect moderators on moderation effects of variables that were associated with target moderator to shed some light on possible confounding effects.

Descriptive analyses to inform moderation assessment (PRISMA-IPD #13)

Describing the target population

Descriptive statistics were used to summarise individual-level baseline clinical and demographic variables listed under *Specification of outcomes and (interaction) effect measures (PRISMA-IPD #13)*. We constructed descriptive statistics for each of the baseline demographic variables, including baseline ECBI-I and potential moderators of treatment effects, both at a trial level and at the individual participant level. In general, summaries are provided by trial and for the pooled data set.

Summarising treatment effects on child outcome (Eyberg Child Behavior Inventory)

Cohen's *d* was computed for each trial 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 SD, for which the pooled SD is estimated from the SD within each group under the assumption that the SD is the same across the population. A Cohen's *d* of 0.2 is typically considered a 'small' effect size, 0.5 is considered a 'medium' effect size and $\geq$ 0.8 is considered a 'large' effect size. For the child outcome, the change between baseline and the first follow-up was first calculated and then the Cohen's *d* between treatment and control conditions was 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 at T1, which takes place between 0 and 2 months after the end of the intervention [see *Specification of outcomes and (interaction) effect measures (PRISMA-IPD #13)*]. The Cohen's *d* for each trial and in the pooled sample will be displayed graphically (*Figure 2*) along with a 95% confidence interval (CI).

Visualising moderation by individual-level baseline variables

We used descriptive statistics to provide some preliminary exploration of potential treatment effect moderation by participant-level variables listed in *Specification of outcomes and (interaction) effect measures (PRISMA-IPD #13)*. For individual-level binary-coded moderators (e.g. SES low income) box plots of change in ECBI-I from baseline to post treatment for both treatment and control conditions were used to assess the difference in treatment effect between the two 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 the two levels of the moderator, then this may be indicative of a moderation effect, although inferential analysis will be required to determine whether or not this effect is statistically significant after accounting for trial variability.

For individual-level continuous variables (e.g. baseline values of ECBI-I) we plotted change in ECBI-I against the potential moderator for both treated and control conditions, including a smooth line for interpretation. If the difference between treated and control conditions 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.

Visualising moderation by trial-level baseline variables

For binary-coded trial-level variables, we plotted 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 box plots across levels of the putative moderator. For continuous trial-level variables we created scatterplots 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.

Understanding relationships between putative moderator variables

We used Pearson's correlation coefficients (tetrachoric correlations for two binary variables) to empirically identify variables that are associated with putative moderator variables. For each hypothesised baseline
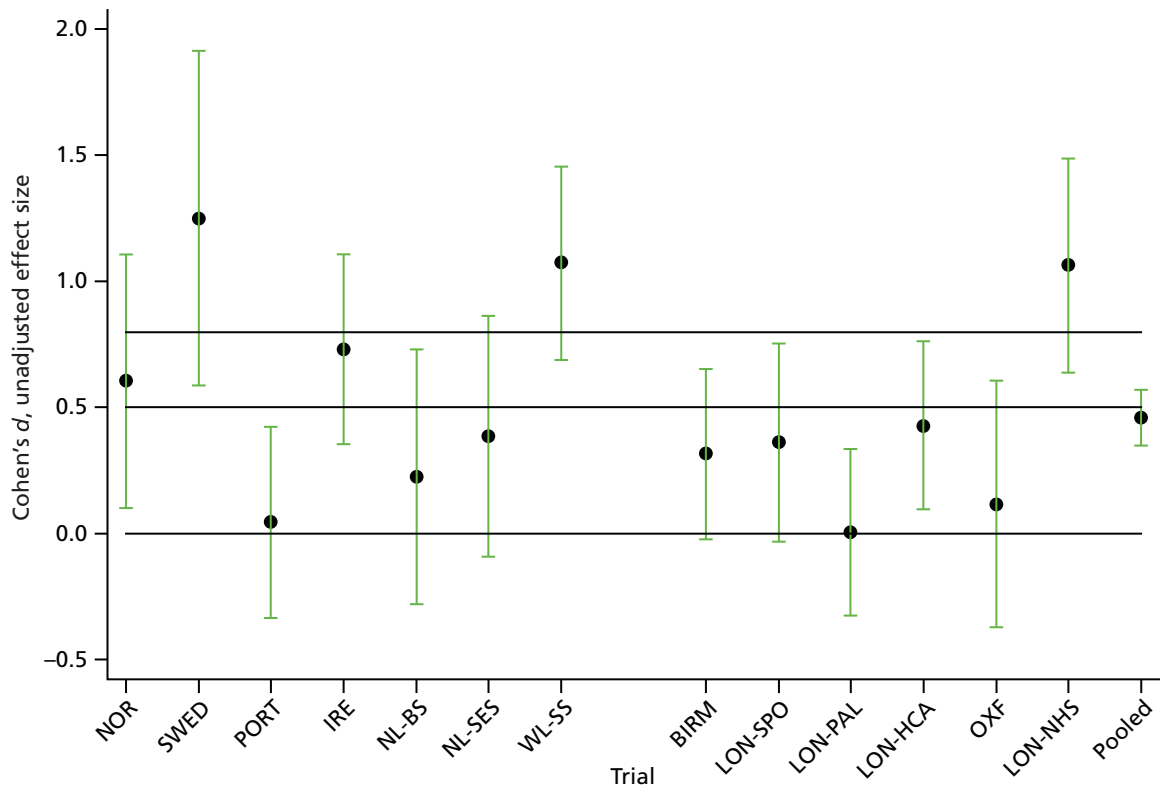
FIGURE 2 Standardised effect sizes for each trial on disruptive behaviour (ECBI-I). Cohen's *d*'s provide standardised treatment effects in terms of ECBI-I. Error bars represent 95% CIs calculated from change scores. The horizontal lines represent no ($d = 0$), a moderate ($d = 0.5$) or a large IY pooling benefit ($d = 0.8$), respectively. Trials references: NOR,[116] SWED,[117] PORT,[104,105,118] IRE,[49] NL-BS,[119] NL-SES,[98] WL-SS,[86] BIRM,[101,103] LON-SPO,[120] LON-PAL,[60] LON-HCA,[121] OXF[122] and VTST.[123]

moderator listed above, we calculated correlations with observed baseline variables. We then ranked these covariates by their level of association with the moderator and thus produced lists of potential confounders of moderator effects. For the purpose of identifying potential confounders, baseline variables were considered in domains as described in *Tables 4* and *5*, and a single domain representative used for the purpose of conditioning of analyses (for more see *Formal inferences*).

Missing data patterns

Using the pooled data set, we assessed predictors of missingness for (long-format) variables to be included in the basic analysis models (see *Formal inferences*). Binary logistic regression was used to identify baseline demographic variables that predict the probability of being missing for putative moderators (when the outcome is a binary-coded variable that is coded 1 for missing and 0 for non-missing on the moderator of interest). We control for trial, trial arm, child gender, child age and baseline ECBI-I in all models, as these variables are included in the imputation step for every moderator. After conditioning for these variables, only lone parent predicted missingness for unemployment, emotional problems predicted missingness for ADHD and teenage parent predicted missingness for parental depression. We included these predictors in the imputation step so as to ensure that the imputation model was valid under the 'missing at random' (MAR) assumption.

Formal inferences

The goal of the inferential analyses was to test each hypothesised moderation effect, adjust for confounding when this was considered a possibility, describe the nature of the moderation effect and quantify the size of the effect.

General modelling approach

Multilevel modelling with MI was used to assess effect moderation based on our pooled data set. Initially, each of these separate moderator variables was considered individually as a moderator effect, that is, as a two-way interaction with the treatment effect in the analysis model. We then investigated possible confounding bias in the moderating effects. This can occur when a baseline variable has an effect on the target moderator or when both have the same cause and the baseline variable is also a moderator of the treatment effect on the outcome. Under these scenarios the magnitude of the causal interaction for the target 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 continuous variables and also before and after adjustment we will calculate standardised moderation indices as the change in treatment effect per unit SD of a putative moderator. For binary-coded variables, where there is no difference between the between- and within-trials effect, the effect size is the difference in treatment effect between one group and another. For binary-coded variables that have different between- and within-trials effects, then the trial-level moderator can be interpreted as a proportion, and therefore the modification index now represents the change in treatment effect between 0% and 100% of participants in a trial belonging to '1' category.

Analysis model for pooled data set

To assess intervention effect modification (moderation), child outcomes of the combined sample ($n = 1799$, decreasing to $n = 1696$ when trial 8 is excluded) were modelled. We considered both putative moderators measured at the individual child or parent level, as well as at the trial level. As we were concerned with assessing (differential) effectiveness of the IY intervention, all statistical analyses were based on the intention-to-treat principle, that is, participants were analysed in the groups to which they were randomised irrespective of whether or not they received a full dose of the intervention.

We start by describing the analysis models used to derive inferences for each putative moderator singly: the dependent variable of our analysis model was child ECBI-I score taken at the first follow-up. The treatment condition is the IY intervention contrasted with the control condition. This is represented by a set of dummy variables coding for the trial arm. Trial arms were coded by three dummy explanatory variables (choosing waiting list as the reference group): IY intervention (yes/no), addition of literacy (yes/no) and minimal intervention (yes/no). We assessed empirically within the complete case (CC) analysis whether or not there was any evidence for differences between IY or control categories and combined trial arms accordingly. If the inclusion of the dummy variable for trial arm is not statistically significant at the $p = 0.1$ level, then it is dropped from the analysis model. For each hypothesised moderator the main effect(s) were 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 were constructed to represent potentially differing trial- and individual-level moderation effects: the variable consisting of trial mean values represents the trial-level aggregate information and the deviations from those means capture the variability in individual scores within a trial. We tested whether or not there is a difference between the trial-level and within-trial individual moderation effects. If this test was not statistically significant at the liberal $p = 0.1$ level, we assumed that there was no evidence of separate between- and within-trial effects and analysed the moderator using a single individual-level variable. The strength of this approach is that trial-level moderation effects are more likely to be subject to confounding and so by separately assessing the within-trial effect, when this is different, we hope to remove some of these potential biases.

The functional form of the relationship between ECBI-I and continuous moderators was not known theoretically. We therefore assessed this empirically and chose the most parsimonious model formulation. Specifically, we tested for a non-linear relationship between the moderator in question and child ECBI-I by adding a quadratic term and an interaction between the quadratic term and trial arm to the model. We then tested whether the fit of this extended model was better than that of a model that assumes a simple linear relationship with the moderator. If the extra terms significantly improved the fit (at the liberal $p = 0.1$ level), they were added to the model; otherwise we stayed with the more parsimonious linear relationship.

As tends to be standard practice in psychosocial RCT analyses, pre-randomisation values of the outcome variable ECBI-I were included in the model to gain precision for the intervention effect estimates. Furthermore, because the pooled data set included a combination of treatment and prevention trials, a dummy variable coding for this trial status was included, as this may be a confounder of trial-level treatment effects (e.g. prevention trials included more low SES and ethnic minority families on average and are hypothesised to have, on average, a smaller treatment effect size). Finally, the design features of specific trials necessitated the inclusion of further explanatory variables for those trials.

First, any randomisation stratifier that was used in a trial was included as further predictive explanatory variable (for a list of such trial-specific stratifiers see *Appendix 2*, *Table 26*). Second, further conditioning on variables may be necessary in some trials to define conditional effects that can be estimated without bias. For example, in some trials the randomisation ratio was changed over time, opening up the possibility that the marginal treatment effect was confounded by factors that change over the duration of the trial. In such situations the conditional trial arm effect, which is conditional on the time period during which the randomisation ratio was held constant, can still be estimated without bias. We assumed that the conditional effect does not vary over time in such trials (i.e. the conditional effect is the marginal effect). Third, child gender and child age were always included as explanatory variables, as these variables were known predictors of outcome and were also used as stratifiers in several trials.

So far we have specified the explanatory variable and associated fixed effects that were of interest to us. We now proceed to extend this model to also include random effects. Random-effects modelling/multilevel modelling was necessary to account for the hierarchical structure of our pooled trials data set. We assumed normally distributed outcomes and random effects. The following random effects were included to (1) acknowledge the clustered structure of the pooled data and (2) represent effect heterogeneity. The pooled data set 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):

1. Cluster structure of the pooled data:

   i. Random intercept varying at the level of trial (13 levels) to account for predictive effects of trial characteristics (e.g. differences in trial target populations or general service organisation contexts affecting control groups) on child outcome under the control condition.
   ii. Random intercept varying at the level of treatment cluster when cluster randomisation was used (trial-specific, for more see *Appendix 2*, *Table 26*).
   iii. Random intercept 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.

2. Random coefficients representing effect heterogeneity:

   i. The regression coefficients representing treatment effects (of trial arm dummy variables) are allowed to vary with trial to model treatment effect heterogeneity (e.g. owing to differences in treatment implementation or target population) not already captured by fixed baseline × trial arm interaction terms.

Our analyses then compared 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 empirically assessed the necessity of including random effects for IY training groups or randomisation clusters when appropriate by calculating intraclass correlation coefficients. When the intraclass correlation coefficient estimate was smaller than 0.01 the effect was assumed to be negligible and the corresponding random effect removed.

Adjusting for potential confounders

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) as we might want to further develop interventions for those families in whom they are currently not effective. This requires us to identify such families outside the context of the current study for which observed correlations between the causal moderator and the predictive marker might be different.

To address this, we adjust any statistically significant moderator effects for possible confounders. Potential confounders are identified from the correlation tables. For each domain we have selected a representative variable, which we use in our adjusted analysis. If a moderator that is statistically significant is significantly correlated with the representative of a domain, then we will adjust for this representative variable by adding both a main effect and interaction with treatment condition of the potential confounder to the analysis model. We will not adjust for variables within the same domain, as these are thought to measure the same concept. In addition, we will adjust only for variables that may be considered as a common cause.

Dealing with missing values

The pooled data set suffered from some missing values in the child ECBI-I outcome and considerable missingness in putative moderator variables. For a summary of data availability within and across trials, see *Table 6*. In particular, child age and gender have complete or near complete data. Low income, education level, lone parent, teenage parent, ethnic minority and the parenting measures have very few (< 10%) missing data. Unemployment, ADHD comorbidity, emotional problems comorbidity and parental depression have a greater proportion of missing data.

We initially carried out CC analyses, that is, we included in the analyses only participants for whom we had values on ECBI-I and all explanatory variables of the analysis model. We also carried out any model section procedures based on the CC approach. In other words, we made decisions regarding combination of between- and within-moderator effects, functional form of the relationship between ECBI-I and continuous moderators and need to include certain random effects based on the CCs only.

Complete case analyses cannot fully exploit the data resource; for example, a participant with missing baseline ECBI-I cannot contribute the moderator analysis for that variable. In addition, and more important, CC analyses make the restrictive assumption that given the level of the covariates, data are missing completely at random. Under such an assumption processes such as post-randomisation ECBI-I predicting missingness in baseline ECBI-I are not allowed. We therefore employed MI to produce inferences that are based on all the available information (precision gain) and are valid under a less restrictive MAR assumption. Specifically, we produced analyses that allow for missingness in a variable to be predicted by any observed variable included in the analysis model, response variable or covariate. We later present results from both approaches to assess the robustness of our results to varying assumption. However, MI is expected to provide less biased and more precise inferences.

Multiple imputation approach

Missing values were accounted for using MI.[27,143,144] MI relies on the assumption that the data are 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.[143] 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.[143]

We therefore included the following variables in the imputation step of a MI analysis:

- trial arm dummy variables
- baseline ECBI-I
- ECBI-I measured at T1
- the putative moderator of interest
- dummy variables coding for therapy group
- dummy variable coding for trial design features (randomisation ratio batch and variables used as stratifiers)
- child gender
- child age.

We opted to impute missing values from multivariate distributions using the MI by chained equations approach.[143] The MI by chained equations approach involves generating imputations from a set of equations, one for each variable with missing values. A regression model is used for continuous variables, whereas logistic regression, ordinal logistic regression or multinomial logistic regression are used for binary, ordered categorical and categorical variables, respectively.

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. In addition, 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 trials 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.[145,146] We therefore opted 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.[147] Unless there is a very large proportion of missing data per variable or the intraclass correlations are very low, it is likely that the bias of the fixed-effects estimates will be relatively small,[148] and in the IY pooling study it is the fixed effects that are primarily of interest. When categories of design variables include only a small number of participants, it is necessary to drop the relevant dummy variables from the imputation step because otherwise this may lead to overfitting of the imputation model, that is, very small numbers of replications can lead to a category being perfectly predicted by other variables in the imputation model.

As the analysis models were set up to assess treatment effect interactions with individual- or trial-level variables, such treatment effect heterogeneity also needed to be allowed for in the imputation model. We opted for imputing separately within each arm within each trial. First, separately imputing by trial ensures that the imputed data can be generated by a treatment × trial interaction, that is, the imputed data reflect treatment effect heterogeneity (moderation by trial-level variables). Second, by 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 × treatment interaction within that trial. Third, separately imputing by trials arms within trials ensures that the imputed data can be generated by a baseline × treatment × trial interaction, that is, the imputed data reflect heterogeneity across trials in the moderation effects of individual-level baseline variables.

Alternative imputation approaches have been suggested for dealing with interactions: the first is the 'just another variable' approach,[143] in which interactions are computed and added to the imputation model as an additional predictor. The second is a linear passive approach using MI by chained equations.[143] 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. In addition, it is necessary to include an outcome × treatment interaction in the imputation model for the moderator. Imputing separately by randomisation group may provide a simpler approach when randomisation group is complete.[143] Thus, we preferred the flexibility of the 'separate imputation' approach.

Synthesis methods (PRISMA-IPD #14)

We used a one-stage approach in which analyses were conducted in a pooled data set of harmonised data from each trial. Clustering of patients within studies was accounted for by analysing data in a hierarchical structure. The pooled data set 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, our one-stage approach allowed us to reflect features of the trial designs in the analysis models: a number of the trials used stratified randomisation and two trials (trials 11 and 14) used a cluster randomised design. In the analysis, these stratification variables were conditioned on, so that trial arm effects are estimated within subpopulations defined by stratifiers. Cluster variables were also available and included as random intercepts to acknowledge the possible correlation between outcome values from individuals of the same randomisation cluster. Finally, in a number of the trials, the randomisation ratio varied over the duration of the trial, which could lead to confounding of the treatment effect on outcome by the time at which the family was randomised. To avoid such bias, dummy variables that code the randomisation batch, when available, will be conditioned on in the analysis model.

Exploration of variation in effects (PRISMA-IPD #14)

The assessment of variation in effects of the intervention by participant characteristics was the main goal of this project. These putative interactions between treatment effects and covariates were all prespecified and discussed in *Specification of outcomes and effect measures (PRISMA-IPD #13)*.

Additional analyses (PRISMA-IPD #16)

Moderation analyses were carried out by CC analysis as well as after MI. By contrasting the results from these two alternative approaches we can assess the sensitivity of findings to changing the assumptions regarding the missing data generating process (from missing completely at random to MAR) with a change in the size of the moderation effect implying some bias removal by the MI analyses.

Analytic plan part 2: wider health benefits and potential harms

Specification of outcomes and effect measures (PRISMA-IPD #13)

To assess potential harms and benefits of the IY intervention we consider the main effect of the intervention (IY vs. control) on each secondary outcome within T1 (post treatment). As an initial step in a series of independent, univariate models (i.e. multilevel analyses of covariances), we tested main effects of the intervention on 12 prespecified outcomes: children's ADHD symptoms, children's emotional problems, self-reported positive parenting behaviour (use of praise, rewards and monitoring), self-reported negative parenting behaviour (corporal punishment, threatening, laxness and shouting), parental depression, parenting stress and parental self-efficacy. In the future, we plan to carry out a multivariate analysis that simultaneously assesses the effects on the 12 variables.

For positive variables (i.e. those for which a higher value is a positive outcome) a statistically significant positive effect of treatment can be seen as a benefit, whereas a negative coefficient may be interpreted as a harmful effect. For negative variables (i.e. those for which a higher value represents a more negative outcome), the regression coefficients are in the opposite direction for harms and benefits. We condition on

baseline values of the outcome of interest because these are known to influence post-treatment outcomes. In addition, random effects are added to the analysis models to account for between-trial heterogeneity both in the intercept and in the coefficient of treatment. Additional random effects for IY group are added only in the IY arm, to allow for the fact that individuals within the same therapy group may be more correlated than individuals in different therapy groups. As for the moderator modelling, fixed effects are added to account for trial design features, including variables used for stratification and randomisation batch when the randomisation ratio was varied within a trial.

Fixed effects included in the analysis model for harms and benefits are:

- baseline values of the outcome of interest
- a dummy variable coding for treatment condition (IY vs. control)
- child age and child gender (used for stratification in some trials)
- dummy variables coding for relevant trial design features (i.e. other stratification variables and randomisation ratio batches).

Random effects included in the model are:

1. Cluster structure of the pooled data:

    i. Random intercept varying at the level of trial (up to 14 levels) to account for predictive effects of trial characteristics (e.g. differences in trial target populations or general service organisation contexts affecting control groups) on outcome under the control condition.
    ii. Random intercept varying at the level of treatment cluster when cluster randomisation was used (trial specific).
    iii. Random intercept varying at the level of IY training group within the IY arm of a trial to account only for predictive effects of the training group/therapist environment within the active treatment arm.

2. Random coefficients representing effect heterogeneity:

    i. The regression coefficients representing treatment effects (of trial arm dummy variables) are allowed to vary with trial to model treatment–effect heterogeneity (e.g. because of differences in treatment implementation or target population) not already captured by fixed baseline × trial arm interaction terms.

To estimate a standardised effect size, we ran the analysis for each variable using a standardised outcome. The outcome was standardised by dividing it by the SD of the variable at baseline. This expresses the treatment effect size in units of baseline SDs, meaning that the size of this effect can be compared across outcomes. In addition, we analysed ECBI-I in the same way, so as to compare the treatment effect on the secondary outcomes with the effect size on the primary outcome.

Most of the secondary outcomes are available within only a subset of the trials and so the sample size is limited for this analysis.

We addressed these issues by using multilevel modelling (random-effects modelling) to capture the hierarchical structure of each variable, accommodated design features by trial-specific fixed effects (e.g. conditioned the model for trial 10 on school year strata) and fitted resulting models by maximum likelihood which is valid under a MAR assumption regarding the process that generates the missing data. In our context MAR implies that variables that are included in the analysis model can drive missingness without this leading to bias. We also intended to analyse the outcome variables in a single multivariate model to further relax missingness assumptions, by allowing missingness of the values of one variable to be driven by other observed outcome values. However, it was not technically feasible to simultaneously model the hierarchical structure as well as fitting a covariance matrix for the multiple outcome measures

per participant. We thus ran two sets of analyses, each having to make some further assumptions to become feasible: (1) a set of univariate analyses that fully capture the hierarchical structure but require the more restrictive MAR assumption and (2) a simplified multivariate analysis that deals with all the outcome variables in one model but does not account for all the trial features (in particular IY group and cluster effects have to be assumed absent).

The single multivariate model (2) includes dummy variables for each outcome and interactions between these dummy variable and each covariate. We estimated the model with an unstructured covariance structure, which allows correlations to vary between each pair of outcomes. We also included our measure of disruptive child behaviour (i.e. the ECBI-I) to the model, the primary outcome of the trials, to allow for further relaxation of the missing data assumptions.

Synthesis methods (PRISMA-IPD #14)
Similar to analytic plan part 1, we used a one-stage approach (this is PRISMA-IPD terminology). This means that analyses were conducted in a pooled data set of harmonised data from each trial, rather than analyses being conducted within the individual trials and aggregate trial results combined afterwards. Clustering of participants within studies was accounted for by analysing data in a hierarchical structure. The pooled data set 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).

Exploration of variation in effects (PRISMA-IPD #14)
Exploring variation in effects of the intervention by participant characteristics was not part of this research question.

Additional analyses (PRISMA-IPD #16)
No additional analyses were conducted.

Chapter 3 Results and discussion of moderator analyses

Descriptive data

Description of included trials

Study selection and individual participant data obtained (PRISMA-IPD #17)

Individual participant data were available and received for all randomised participants in 14 trials (see flow diagram, *Figure 1*). Investigators were contacted in cases when additional information was needed about the interpretation of the IPD. All investigators signed a data sharing agreement (see *Appendix 1*). IPD for the 15th trial[102] were no longer available. The pooled data set consisted of records on 1799 families. However, for all moderator analyses 13 trials were analysed, as one trial (trial 8; Wales toddler trial) did not include a measure of the primary outcome, child disruptive behaviour.

Study characteristics (PRISMA-IPD #18)

Tables 1 and *7* provide an overview of the 14 trials included in the individual participant-level analyses. Trials were conducted in seven different countries: six trials in England, two in Wales, two in the Netherlands and one trial each in Ireland, Norway, Sweden and Portugal.

Ten trials were either indicated prevention trials or treatment trials and therefore included mostly children with clinical levels of disruptive child behaviour. More specifically, level of disruptive child behaviour was the main selection criterion for inclusion of families in these trials, typically as assessed by standardised parent questionnaire. In some trials children were referred for treatment in a mental health clinic because of concern about behavioural problems; in others they were screened for behavioural problems in various community settings. In contrast, four trials were selective prevention trials (trials 5, 6, 8 and 11). These trials selected participants based on risk factors others than baseline levels of disruptive child behaviour (e.g. socioeconomic disadvantage of family or school). To illustrate, mean baseline score on the ECBI-I in indicated prevention or treatment trials was 145.95 (SD 34.26) and mean baseline score on the ECBI-I in selective prevention trials was 112.36 (SD 33.75).

Whether or not a trial was a selective prevention trial was strongly related to the number of families with social or socioeconomic disadvantage and ethnic minority backgrounds included in the trial. In three selective prevention trials (excluding Wales Flying Start toddler trial, WL-FS, which is not part of the moderator analyses), on average 70% (range 44–93%) of the families had low income and 73% (range 65–78%) of the families came from ethnic minority backgrounds. In contrast, in the 10 indicated prevention or treatment trials, in which levels of behaviour problems were higher, on average 45% (range 0–80%) of the families had low income and only 12% (range 0–52%) of the families came from ethnic minority backgrounds.

Interventions took place in a range of settings, including child mental health services and community children's settings, such as primary schools (trials 10–12), and local authority children's services (trials 4 and 7–9), for example in children's centres within Sure Start services (trial 7). In three trials the intervention was delivered wholly or partly by voluntary sector organisations (trial 4, 7 and 13). One took place in a university clinic set up for research (trial 3), and trialled a specialist service that was set up by the researchers for mothers being released from prison (trial 5). We were asked to examine only interventions delivered outside the NHS; however, with the agreement of the funder, one trial (trial 14) in a NHS CAMHS was included because the intervention is identical and the sample is highly similar in terms of problem severity and demographic background to many of the other trials. One further trial took place in the NHS, in general practice, but was not included as data were unavailable.[102]

TABLE 7 Trial design features

| Trial number | Country/city | Trial acronym | n | 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 |
				Active	Control					
1	Norway	NOR[113]	75	IY	Waiting list	5	12.09	Yes	0	2
2	Sweden	SWED[114]	62	IY	Waiting list	5	13	No	0	2
3	Portugal	PORT[115,116]	124	IY	Waiting list	5	14	No	1	2
4	Ireland	IRE[48]	149	IY	Waiting list	5	13.28	Yes	0	2
5	The Netherlands	NL-BS[118]	99	IY	Care as usual/no care	5	12	No	4	2
6	The Netherlands	NL-SES[96]	156	IY	Waiting list	5	14.46	Yes	0	2
7	Wales	WL-SS[85]	153	IY	Waiting list	5	12	No	0	2
8	Wales	WL-FS[104]	103	IY – toddler version	Waiting list		12	No	0	2
9	England	BIRM[105,119]	161	IY	Waiting list	5	12	No	0	2
10	England	LON-SPO[100]	112	IY and literacy	Care as usual/no care	8	12	No	0	2
11	England	LON-PAL[59]	174	IY and literacy	Minimal intervention	8	12	No	0	2
12	England	LON-HCA[101]	214	IY; IY and literacy	Minimal intervention	5–8	12	No	0	3
13	England	OXF[103]	76	IY	Waiting list	5	14	No	0	2
14	England	LON-NHS[120]	141	IY	Waiting list	5	14.11	Yes	0	2
15 – no IPD available	England	Patterson *et al.*[102] in primary care setting	116	IY	No intervention	Not known	10	No	0	2

Children's ages varied between 10 and 133 months (mean 60.76 months, SD 19.84 months). Most trials included children between 36 and 96 months. One trial (trial 8) included very young children (toddlers aged 10–36 months) and one trial (trial 5) included children from a very wide age range (22–133 months). Most other trials included children between 36 and 96 months. Importantly, we note that the trial involving toddlers (trial 8) was included in the wider benefits analysis but was omitted from all moderator analyses, as it did not include data on the primary outcome (parent-reported conduct problems).

In 11 included trials the majority of families were social disadvantaged, in terms of having a low income or a lone parent (all except trials 1–3). Some trials specifically targeted low-income areas (trials in Welsh Sure Start and Flying Start areas; trials 7 and 8), low-income families (trials 6) or schools in low-income wards (trials 10–12). Five trials included substantial numbers of families from ethnic minorities (33–75% of the trial sample; trials 5, 6 and 9–11) and five had very low numbers (0–3%; trials 1, 3, 7, 8 and 13).

There was some trial-level variation in what was planned to be delivered to the intervention and control groups (see *Table 7*). In most trials, the standard 12- to 14-week IY Basic parenting programme was delivered. Three trials (trials 10–12) offered a parent–child literacy intervention alongside IY. The trial with mothers released from prison (trial 5) offered four additional home visits to support the group parenting work. Similarly, the control condition differs across trials, with some offering waiting list or treatment as usual and others offering minimal treatment. As this study is focused on the IY training programme, participants randomised to a literacy-only intervention were excluded from the pooled data set, as well as participants who were randomised to an IY parent and IY child programme (trial 1). The median number of sessions offered to parents across all trials was 12 (range 11–19), and the median attended was 10 (range 0–18).

Table 7 shows the total sample size from the trial that was used in the IY pooling data set, the type of control condition and the arms included in the pooled sample. When the trial included arms that were not used in the pooled sample (e.g. reading intervention only or IY child programme), these are excluded from the table. 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 than what was normally accessible in the community. Minimal intervention means that some non-intensive intervention was provided to parents in the control arm, such as a telephone helpline. Most trials used a waiting list design, in which parents in the waiting list condition were offered the intervention after 6 months; however, these families, although some were followed up, were no longer compared with a control group, and, therefore, these follow-up data did not form part of the analysis sample in the present study. A full list of the trials included in the pooled sample and corresponding references is given in *Appendix 2*.

The trials differed in other aspects of study design. A few employed cluster randomisation, most employed stratified randomisation and some varied the randomisation ratio during the course of the trial (e.g. because of low recruitment). Further details regarding these trial design variables are provided in *Appendix 2*.

Risk of bias within studies (PRISMA-IPD #19)

Table 8 summarises the risks of bias identified in each trial. We focused on the main outcome of child conduct problems as measured by parent-reported ECBI-I.

Participants were not able to be blind to condition in these trials, as participants in the intervention group received the IY parenting intervention and participants in the control group did not. Some of the sources of bias in *Table 8* are minimised through the pooling of individual participant-level data, including missing data and selective outcome reporting. Outcome data were based primarily on parent-reported child behaviour and other outcomes; thus, although assessors were blind to intervention status of families, reporters (parents) were not. Observational data on child and parent behaviour, coded by blinded assessors, helped compensate for this problem in many trials; however, these data were not suitable for harmonising.

TABLE 8 Summary of risk of bias within trials

Trial number	Trial acronym	Sequence generation	Who assigned	Assigned concealed	Blind assessors	Address incomplete data	Select outcome reporting	Analysing dropouts	Any other bias
1	NOR[113]	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Low risk
2	SWED[114]	Low risk	Low risk	Unclear risk	High risk	Low risk	Low risk	Low risk	Low risk
3	PORT[115,116]	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Low risk
4	IRE[48]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
5	NL-BS[118]	High risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
6	NL-SES[96]	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk
7	WL-SS[85]	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
8	WL-FS[104]	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	Unclear risk	Low risk
9	BIRM[105,119]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
10	LON-SPO[100]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
11	LON-PAL[59]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
12	LON-HCA[101]	Low risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk
13	OXF[103]	Low risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk
14	LON-NHS[120]	High risk	Low risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk

Results of individual studies (PRISMA-IPD #20)

Table 9 shows a standardised moderation index for each trial and putative individual-level moderator at baseline. For binary moderators, this index was calculated for each trial as the difference in treatment effects (in terms of post-test ECBI-I) between the two moderator levels. For continuous moderators the index was constructed for each trial as the change in post ECBI-I per one (pooled sample) SD change in the baseline moderator. The indices are expressed on the original ECBI-I scale and their size can be compared across trials for each moderator variable (but only across continuous variables or binary variables within a trial). *Table 9* shows that for binary moderators the largest mean indices were observed for SES low income, low education and unemployment. For continuous variables the largest mean indices were for baseline ECBI-I, monitoring, praise and laxness. These summaries show considerable heterogeneity about trials, in particular for binary moderators such as child gender. This volatility in estimates can be explained by some individual trials having few participants in one of the moderator categories and the moderation index is very sensitive to the values of a few observations. For example, trial 1 included very few girls.

Descriptives for the pooled data set

Individual-level variables
Tables 10 and *11* provide demographic and clinical summaries for the pooled data set. The majority of children were boys (63%) and their mean age was 63 months. They had high levels of behavioural problems, with a mean score of 137 points (SD 37 points) on the ECBI-I, compared with a clinical cut-off point of 127 points. A small majority of families (58%) had a low income: 35% were unemployed and 35% were headed by a lone parent. Thirty per cent of families were from an ethnic minority. Comparison of summaries between treatment arms confirms that randomisation successfully avoided any imbalances between trials arms at baseline.

Table 12 shows intercorrelations between baseline variables. The table shows that some variables were strongly correlated (correlation coefficient > 0.4). As expected, larger correlations were found between variables from the same domain, for example between SES variables. But there were also some large correlations between theoretical domains, for example between praise and corporal punishment. These correlations will need to be taken into account when interpreting unadjusted moderation effects later.

Trial-level variables
Table 13 shows trial-level characteristics, and shows that the majority of trials took place in the UK and Ireland, and the IY intervention was delivered by non-clinical organisations. There was little variability between trials in indices of fidelity of intervention implementation, which were mostly at ceiling level. The exception to this was the percentage of staff certified in the IY intervention variable, which applied to only a minority of staff (30%). However, all staff were trained in IY and regularly supervised using video of their own groups, and they self-monitored implementation of the programme using materials provided within the programme.

The lack of variability between trials meant that we were able to investigate moderation of IY effects by only a few trial-level variables, namely geographical region, percentage staff clinically education, percentage of staff IY certified and type of trial. It is important to note that any findings on trial-level moderators should be interpreted with extreme caution. Effectively they have only 13 replicates and thus are likely to be correlated with each other in our pooled sample. Furthermore, they are likely to be correlated with unobserved trial-level variables, which opens up the possibility of their observed moderation effects being subject to confounding.

TABLE 9 Treatment effect modification indices by trial and putative baseline moderator

Moderator variable	Trial number													Mean
	1	2	3	4	5	6	7	9	10	11	12	13	14	
SES low income	19.6	0.4		0.9	3.1	–0.8	–6.0	0.2	22.0	0.05	8.2	22.9	–15.0	4.7
Low education	–22.7		17.1	–2.2	24.1	–8.6	2.6	0.9	14.4	12.8	17.1	6.1	2.5	5.3
SES lone parent	3.1	–21.8	–0.5	–9.7	12.2	–41.0	–9.3	15.6	1.5	1.7	6.3	4.9	3.0	–2.6
SES teenage parent	6.0		–25.5	–30.9	4.1	–10.8	–0.4	26.2	59.4	–1.7	22.7	–24.5	–12.0	1.1
SES unemployed	32.5			5.2	–0.3			7.3	27.7	4.7	15.2	12.8	0.8	11.8
Ethnic minority				30.2	1.8	8.2		–6.3	–1.8	–0.8	–17.8		4.2	2.2
Child gender	21.5	23.2	–7.5	–8.3	–4.4	–11.8	–27.1	–2.9	–23.7	7.0	11.7	–29.7	9.3	–3.3
Child age	1.9	–5.9	–5.7	2.2	–0.8	–0.4	32.3	–1.4	–15.3	2.8	–6.8	–0.2	–4.9	–0.2
Baseline ADHD	2.9	–6.7	11.4	8.7		–0.4	–3.5	9.4	–1.6	4.8	–3.5	–7.3	4.4	1.6
Baseline emotional problems	0.05	–4.3	–6.6	–2.5		–5.4	–0.2	–6.5	–3.7	–2.4		2.7	–1.2	–2.7
Baseline ECBI-I	4.8	–12.9	7.5	–4.8	–1.9	–11.4	–1.7	–16.4	–2.7	4.6	–7.6	2.9	–0.5	–3.1
Baseline parental depression	–8.4	–9.39	0.5	–4.2	–3.0	–3.3	–13.7		–1.4	–1.5	7.9	2.0	8.3	–2.2
Monitoring	15.7		–4.4		39.5	6.9	8.9	–7.7			–2.9	–1.4	–1.8	5.9
Tangible rewards	0.6		–12.4		2.1	3.7			–6.5		–3.7			–2.7
Praise	15.1		6.8		7.0	–3.4			0.2		–3.9			3.7
Corporal punishment	–33.4		7.3		–0.2	–7.0	2.9	1.5	1.9	1.6	–9.4	11.4	–0.3	–2.2
Threatening	–19.2		–0.8		1.8	–1.9	3.3	0.3			–0.26	8.2		–1.1
Laxness	–2.5		–9.1		–9.5	–4.1	–4.5	–5.5			–2.52	8.2		–3.7
Shouting	–5.8		5.4		18.6	–6.6	–9.4	–8.0			0.15	18.7		1.7

TABLE 10 Summaries for demographics by randomised group (pooled sample)

	Control		IY	
Variable	n	Mean (SD) (%)	n	Mean (SD) (%)
Child gender (male)	650	63.8	1046	63.1
Child age (months)	643	64.2 (16.9)	1039	62.4 (18.3)
SES low income	615	57.9	999	57.6
Low education	650	35.5	1046	40.5
SES lone parent	606	33.0	1000	36.8
SES teenage parent	605	12.6	1004	11.7
SES unemployed	522	30.3	781	37.5
Ethnic minority	629	30.0	1022	30.9

TABLE 11 Summaries for clinical variables at baseline plus ECBI-I outcome post by randomised group for pooled sample

	Control		IY	
Variable	n	Mean (SD)	n	Mean (SD)
ECBI-I total baseline	611	135.5 (37.0)	1011	139.4 (37.0)
ECBI-I total post treatment	567	125.5 (37.9)	878	116.2 (34.7)
SDQ ADHD	589	5.8 (2.7)	943	5.9 (2.7)
SDQ emotional	491	3.2 (2.4)	849.0	3.4 (2.7)
Monitoring	394	5.2 (1.7)	694	5.3 (1.7)
Tangible rewards	243	3.3 (1.3)	382	3.3 (1.2)
Praise baseline	244	5.4 (1.2)	386	5.4 (1.2)
Corporal punishment	533	2.2 (1.4)	860	2.1 (1.4)
Threat	361	3.6 (1.5)	638	3.5 (1.6)
Laxness baseline	355	3.3 (1.3)	623	3.3 (1.3)
Shouting	349	3.3 (1.5)	618	3.1 (1.3)
Parent depression BDI total	543	10.1 (9.7)	852	12.2 (10.9)
PSI-SF total	181	89.0 (28.4)	361	92.1 (28.4)
PSOC scale total	181	54.1 (7.6)	236	54.0 (7.6)

TABLE 12 Correlations between putative individual-level moderators

	Low income[a]	Parental education	Unemployed	Lone parent	Teenage parent	Child age[a]	Child gender[a]	ECBI-I[a]	ADHD	Emotional problems	Parent depression[a]	Corporal punishment[a]	Threat	Laxness	Shouting	Monitoring	Tangible rewards	Praise[a]	Ethnicity
Low income[a]	1																		
Parental education	0.22[b]																		
Unemployed	0.54[b]	0.27[b]																	
Lone parent	0.32[b]	0.15[b]	0.44[b]																
Teenage parent	0.16[b]	0.15[b]	0.22[b]	0.19[b]															
Child age[a]	−0.01	−0.02	−0.07[b]	0.03	0.00														
Child gender[a]	−0.04	0.01	−0.02	0.00	−0.03	0.01	1												
ECBI-I[a]	0.09[b]	0.10[b]	0.18[b]	0.06	0.06	0.03	0.10[b]	1											
ADHD	0.00	0.13[b]	0.05	0.02	0.00	0.05[b]	0.11[b]	0.36[b]	1										
Emotional problems	0.00	0.00	−0.02	0.02	0.01	0.05	−0.03	0.04	0.22[b]	1									
Parent depression[a]	0.13[b]	0.18[b]	0.21[b]	0.14[b]	0.06[b]	−0.07	0.01	0.28[b]	0.16[b]	0.26[b]	1								
Corporal punishment[a]	−0.04	0.08	−0.02	0.03	0.00	−0.10[b]	0.05	0.09[b]	0.11[b]	0.00	0.11[b]	1							
Threat	0.17[b]	0.12[b]	0.06	0.07[b]	0.04	−0.04	−0.04	0.05	−0.02	0.04	0.06	0.05	1						
Laxness	−0.01	0.01	0.05	−0.04	−0.04	−0.12[b]	−0.02	0.13[b]	0.00	0.06	0.09[b]	0.00	0.43[b]	1					
Shouting	−0.01	0.02	0.00	0.04	0.02	0.01	0.03	0.20[b]	0.17[b]	0.05	0.24[b]	0.28[b]	0.14[b]	0.20[b]	1				
Monitoring	0.05	0.11[b]	−0.03	0.05	−0.01	0.01	−0.06	−0.18[b]	−0.16	0.03	−0.20[b]	−0.17[b]	−0.01	−0.15[b]	−0.17[b]	1			
Tangible rewards	0.12[b]	0.04	−0.07	0.05	0.00	−0.06	−0.01	−0.09	−0.10[b]	−0.08	0.01	0.04	0.07	−0.04	−0.07	0.15[b]	1		
Praise[a]	0.12[b]	−0.02	0.08	−0.10[b]	−0.01	0.01	−0.08[b]	−0.02	−0.16[b]	−0.18[b]	−0.21[b]	−0.12[b]	0.07	0.04	−0.20[b]	0.10[b]	0.12[b]	1	
Ethnicity	0.07[b]	−0.12[b]	0.05	0.02	−0.03	−0.07[b]	−0.03	−0.17[b]	−0.14[b]	0.07[b]	−0.07	−0.03	0.10[b]	0.10[b]	−0.17[b]	0.02	0.16[b]	−0.04	1

a Denotes a representative variable within a domain.
b Denotes a statistically significant correlation ($p < 0.05$).
Note
All correlation coefficients are partial correlation coefficients controlling for the type of trial (prevention trial yes/no).

TABLE 13 Summaries of putative trial-level moderators

Trial-level moderator	Mean (SD) or number true (%)
Percentage of sites rural	4.1% (10.9)
Geographical location UK and Ireland (vs. non-UK and Ireland)	9 trials (64.3)
Effectiveness trial (vs. efficacy)	13 trials (92.9)
Percentage of staff IY certified before trial	29.6% (32.9)
Percentage of staff clinically educated	55.6% (38.3)
Service provider for trial non-clinical (vs. clinical)	11 trials (78.6)
Type of trial treatment (vs. prevention)	10 trials (71.4)
Did staff complete IY checklist after sessions?	14 trials (100)
Was IY mentor part of the trial?	10 trials (71.4)
Were sessions video-taped?	14 trials (100)
Were video-taped sessions used in supervision?	14 trials (100)
Was there weekly/fortnightly supervision?	10 trials (71.4)
Did independent ratings of session fidelity take place?	3 trials (21.4)
Did any of team attend international IY workshop/training?	10 trials (71.4)
Number of IY sessions offered	14 trials (100)

Results of moderator analyses

Overall effect of the parenting intervention

We know from the published trial literature that the IY parenting intervention significantly reduced parent-reported disruptive child behaviour. The majority of studies indicated that the intervention was beneficial in terms of ECBI-I. The overall unadjusted Cohen's *d* effect size for our pooled sample was 0.46. Effect sizes varied between trials, ranging from 0.01 to 1.25 (see *Figure 2*). Trials with larger effect sizes tended to be trials that were indicated prevention or treatment trials (e.g. trials 2 and 7), as opposed to selective prevention trials (i.e. trials 5, 6, 8 and 11). As mentioned earlier, these four trials also study selected subpopulations that differed on various baseline variables. We therefore controlled for the type of prevention trial in all formal analyses to ensure that any trial-level moderator results were not a result of confounding by this selection variable.

Moderators of parenting intervention effectiveness (PRISMA-IPD #21 cont'd)

In the analyses that follow, we report findings on within-trial moderators, based on MI of missing data, as described in the analysis plan, *Analytic plan part 1: moderation analyses*. Full results of all moderator analyses can be found in *Appendix 5*, *Tables 29* and *30* including both the MI and, for comparison, the CC analysis findings for individual-level moderators. A summary of significant moderator effects, using MI, is shown in *Table 14*. There were generally few differences in results between these two approaches for handling missing data. For individual-level moderators we therefore present the less bias-prone MI results. Between-trial moderators results are also provided, using CC analysis, in *Trial-level moderators*.

We start by presenting the unadjusted moderation results by different types of putative moderators. As a preliminary model-building step, and before investigating any interaction effects, we tested whether or not we needed to distinguish between four treatment arms in our basic model using CC analyses. We found no evidence that this was necessary (ECBI-I did not differ significantly between waiting list/no care and minimal intervention control groups; $p = 0.11$; or between IY pooling arms with and without the addition of a reading intervention; $p = 0.46$) and thus we always treated all control arms as equivalent in our

TABLE 14 Results from moderator analyses of parenting variables

| Moderator | Test of | | Split between/within | Analysis results | | |
	Between-/ within-effect	Quadratic term		Moderator index	95% CI	*p*-value
Positive parenting						
Parenting: monitoring	*p* = 0.34	*p* = 0.36		1.8	–2.5 to 6.0	0.42
Parenting: tangible rewards	*p* = 0.83	*p* = 0.34		–3.0	–7.7 to 1.6	0.20
Parenting: praise	*p* = 0.68	*p* = 0.55		–3.7	–8.7 to 1.4	0.16
Negative parenting						
Parenting: corporal punishment	*p* = 0.42	*p* = 0.18		0.4	–3.0 to 3.8	0.83
Parenting: threatening	*p* = 0.16	*p* = 0.43		0.7	–3.5 to 4.9	0.74
Parenting: laxness	*p* = 0.39	*p* = 0.09	Linear	–4.3	–8.6 to 0.1	0.12
			Quadratic	1.61	–1.4 to 4.6	
Parenting: shouting	*p* = 0.8893	*p* = 0.2475		–0.02	–4.1 to 4.0	0.99

All analyses results shown here were obtained by MI.

analyses. Furthermore, as part of our preliminary model-building we report the specific model selected for each putative moderator variable based on CC analyses: for all variables we report whether or not there was any evidence for differential between- and within-trial moderation effects and describe our moderation findings accordingly. In addition, for continuous variables we include a test of whether or not non-linear effects needed to be allowed for, in order to capture the relationship between the variable and ECBI-I. We quantify the strength of any interaction effects by moderation indices. We tested for the necessity of adding random effects to account for trial design features, such as training group in the IY arm or cluster for those trials that used cluster randomisation. The intraclass correlation was < 0.01 for training group in all trials with the exception of trials 3 and 7, and so random intercepts for the IY group were included in the final analysis for only those trials. The intraclass correlation for cluster in trial 14 was < 0.01 and so this random effect was also excluded.

Family-level moderators

Socioeconomic and social disadvantage as moderators
We considered five variables as capturing social and socioeconomic disadvantage: low income, low education, unemployment status, lone parent or teenage parent.

Low income was the SES moderator with the fewest missing values (CC sample size: *n* = 1614). *Figure 3* illustrates the moderation effect of this variable graphically and indicates very similar IY intervention effects in the low- and high-income groups. Formally, we found no evidence that any IY effect moderation by low income varied between the trial and individual level (*p* = 0.286). We therefore modelled a single interaction effect for low income. This effect could not be shown to be statistically significant (effect modification index 1.9 points on ECBI-I, 95% CI –4.8 to 8.6 points; *p* = 0.58). There was therefore no evidence to suggest that low or high income status affected the benefit of the IY intervention.

Low or high education status was recorded for a slightly smaller subsample (CC sample size: *n* = 1573). *Figure 4* illustrates this moderation effect, and also suggests similar-sized IY intervention benefits in the two education groups. We found no evidence that any IY effect moderation by low education varied between
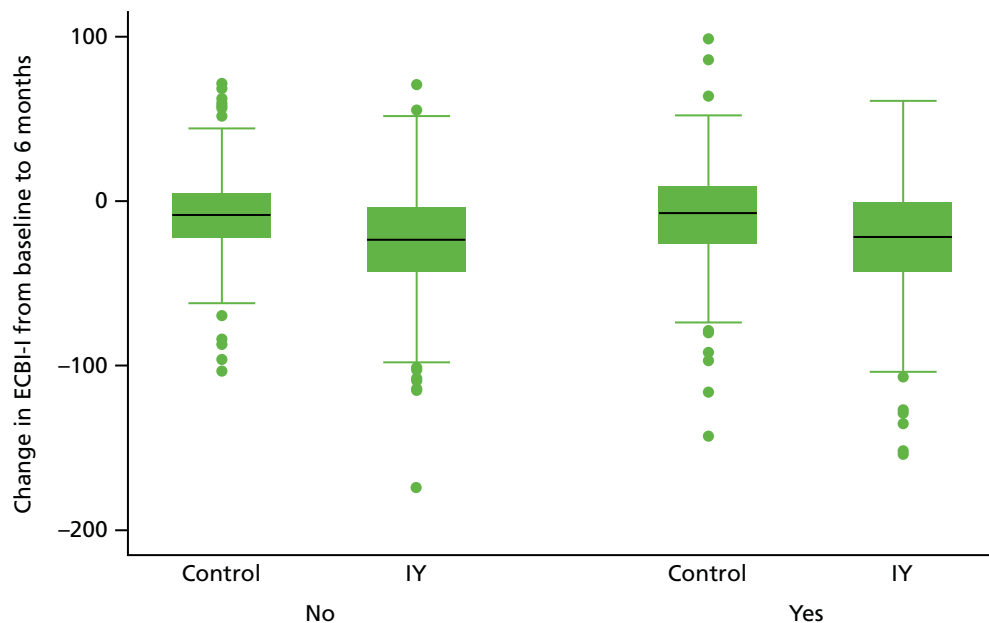
FIGURE 3 Box plots illustrating low income (SES) as a moderator of the effects of the IY programme on child conduct problems.
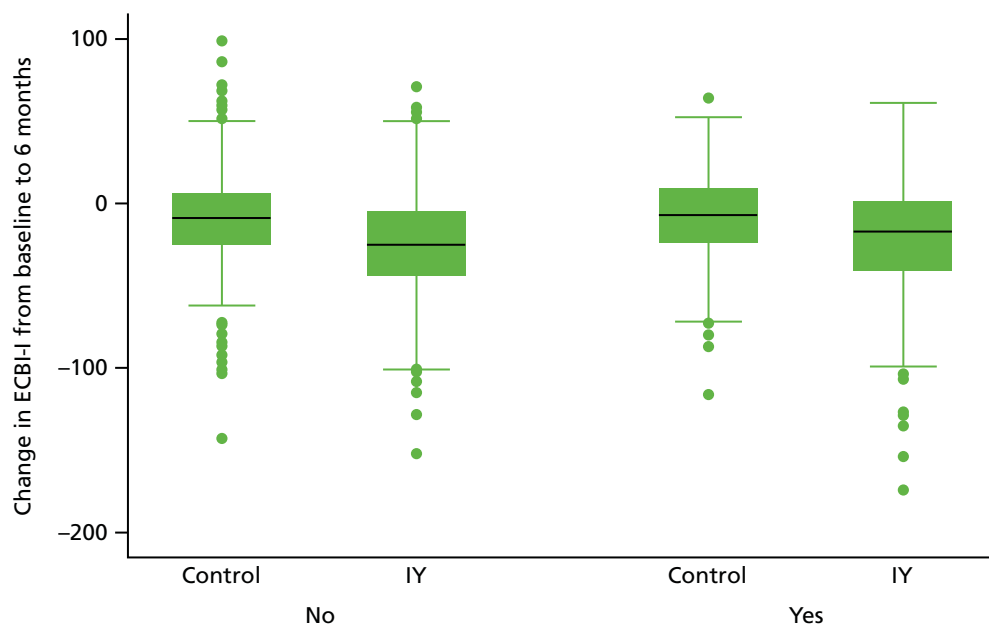
FIGURE 4 Box plots illustrating low education (SES) as a moderator of the effects of the IY programme on child conduct problems.

the trial and individual level ($p = 0.27$). We therefore again modelled a single interaction effect for low education. Albeit larger in size, this effect could not be shown to be statistically significant in the formal analysis (modification index 4.4 points, 95% CI –2.2 to 10.9 points; $p = 0.49$). There was therefore no evidence to suggest that low or high education status affected the benefit of the IY intervention.

Unemployment was recorded for a smaller subsample (CC sample size: $n = 1303$). *Figure 5* illustrates this moderation effect, and also suggests similar-sized IY intervention benefits in the two employment groups. We found no evidence that any IY effect moderation by unemployment varied between the trial and individual level ($p = 0.67$). We therefore again modelled a single interaction effect for unemployment.
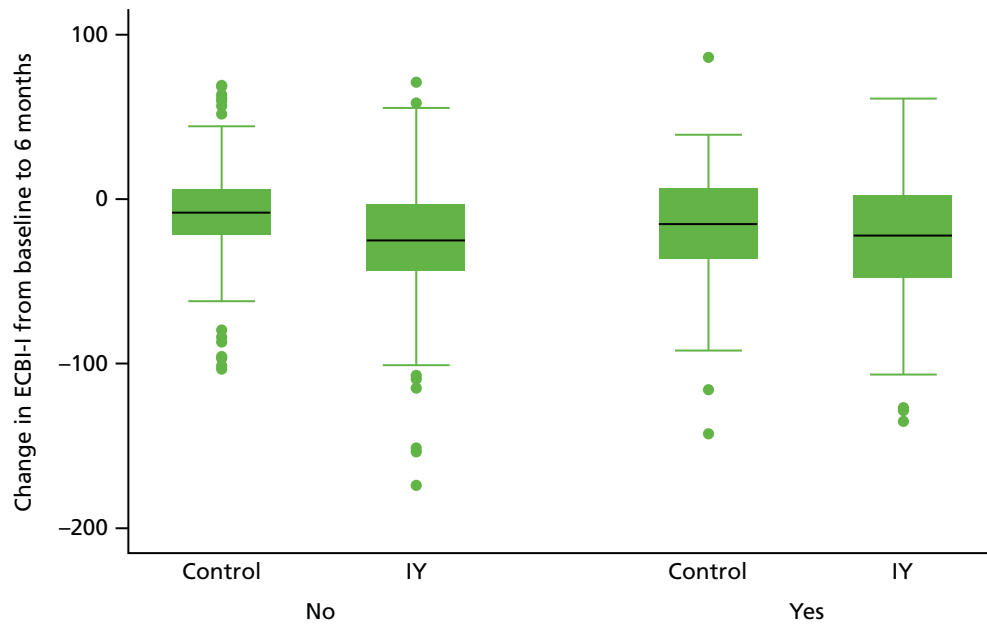
FIGURE 5 Box plots illustrating unemployment status (SES) as a moderator of the effects of the IY programme on child conduct problems.

Albeit larger in size, this effect could not be shown to be statistically significant in the formal analysis (modification index 4.88 points, 95% CI –2.7 to 12.4 points; $p = 0.21$). There was therefore no evidence to suggest that unemployment status moderated the intervention effect.

Lone parent status was recorded for a subsample of size $n = 1606$. *Figure 6* illustrates this moderation effect, again suggesting similar sizes of IY intervention benefits in the two groups. We found no evidence that any IY effect moderation by lone parent status varied between the trial and individual level ($p = 0.840$). We therefore again modelled a single interaction effect for lone parent. This moderation effect was small
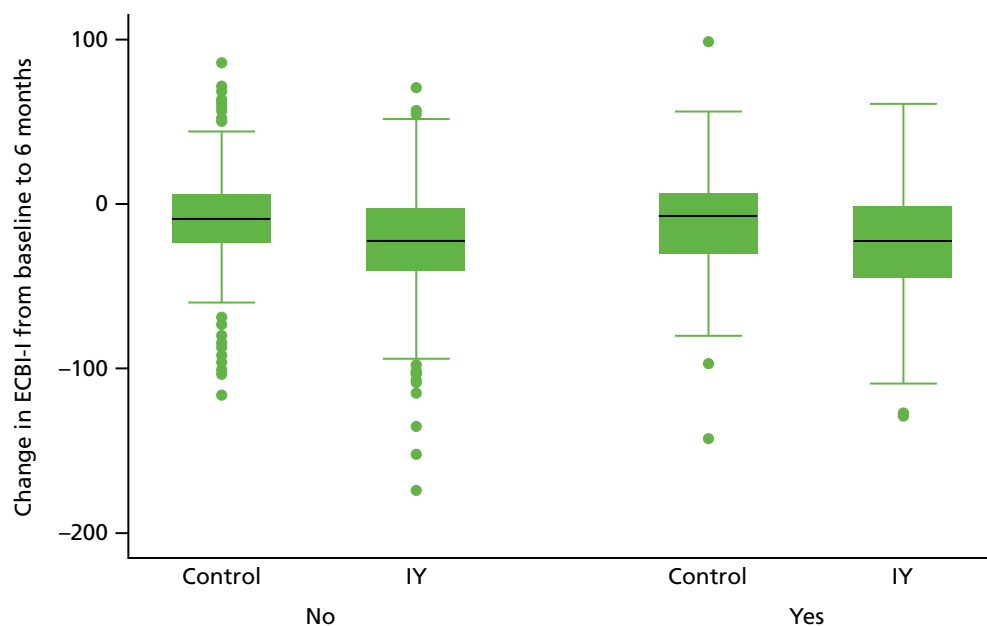
FIGURE 6 Box plot illustrating lone parent (SES) as a moderator of the effects of the IY programme on child conduct problems.

in size and not statistically significant in the formal analysis (modification index 0.5 points, 95% CI –6.1 to 7.1 points; $p = 0.88$). There was therefore no evidence to suggest that lone parent status moderated the benefit of the IY intervention.

Teenage parent was recorded for a subsample of size $n = 1550$. *Figure 7* illustrates this moderation effect, and also suggests similar sizes of IY intervention benefits in the two age groups. We found some evidence that any IY effect moderation by teenage parent status varied between the trial and individual level ($p = 0.051$). We therefore modelled separate interaction effects for teenage parent at the between- and within-trial level. At the trial level there was a trend for a moderation effect, suggesting that children in trials with a higher proportion of teenage parents benefited more from the intervention (modification index –76.1 points, 95% CI –166.1 to 14.0 points; $p = 0.10$), whereas at the individual (within-trial) level there was no evidence for a moderation effect and the effect direction was reversed (modification index 7.3 points, 95% CI –2.2 to 16.9 points; $p = 0.13$). Given that the modification index at the trial level represents the effect of all trial participants shifting from not being teenage parents to all being teenage parents and that between-trial effects are liable to hidden confounding, we treat this result as insufficient evidence for effect modification by teenage parent status.

Ethnicity as moderator

Most of the data are drawn from six trials (trials 5, 6 and 9–12), which included 97.7% of the ethnic minority families in the pooled sample. We found some evidence that any IY effect moderation by ethnicity varied between the trial and individual level ($p = 0.042$). We therefore modelled separate interaction effects for ethnicity at the between- and within-trial levels. At the trial level there was a significant effect (modification index 19.5 points, 95% CI 1.0 to 38.1 points; $p = 0.04$), suggesting that children in trials with a higher proportion of ethnic minorities benefit less from the intervention. However, no effect modification could be detected within trials and the effect was reversed (modification index –1.4 points, 95% CI –9.8 to 7.1 points; $p = 0.75$). Given that the between-trial effects are susceptible to confounding we considered that there was no evidence to suggest that the effectiveness of the IY parenting intervention to reduce disruptive child behaviour was influenced by the parent's ethnic background (*Figure 8*).
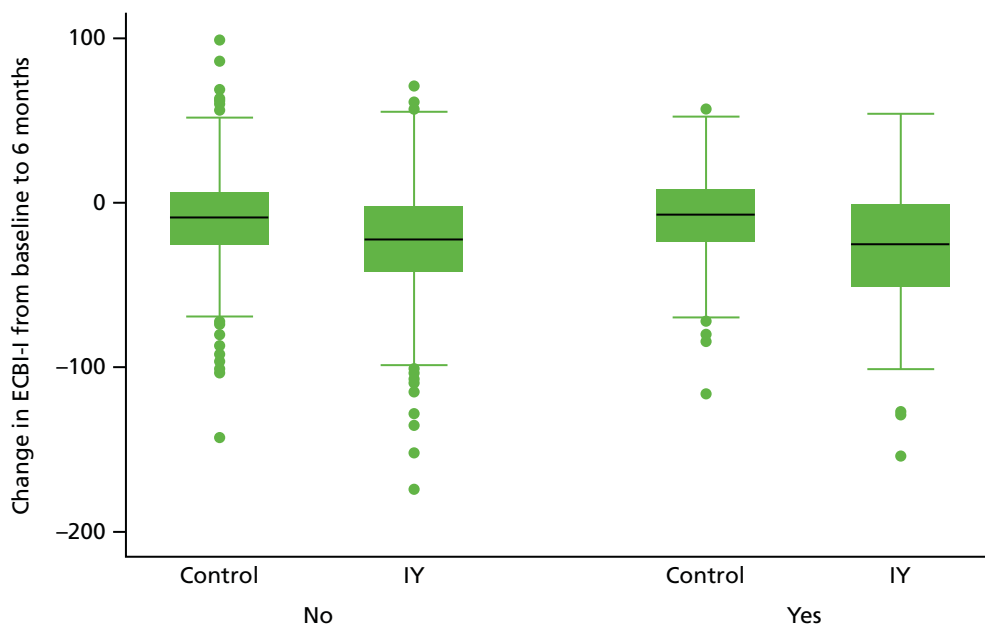
FIGURE 7 Box plots illustrating teenage parent (SES) as a moderator of the effects of the IY programme on child conduct problems.
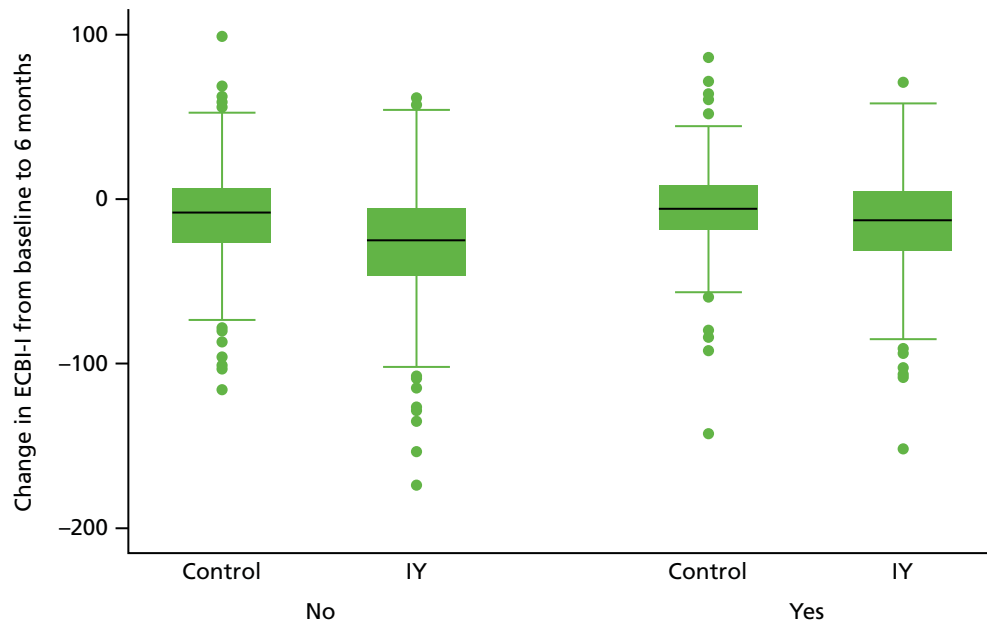
FIGURE 8 Box plots illustrating ethnic minority as a moderator of the effects of the IY programme on child conduct problems.

Child characteristics as moderators

Figure 9 illustrates the observed moderation by child age. We found no evidence that any IY effect moderation by age varied between the trial and individual level ($p = 0.45$) or that the functional relationship between ECBI-I and age was not linear ($p = 0.89$). We therefore again modelled a single interaction effect for age. This moderation effect was small in size and not statistically significant in the formal analysis (modification index 0.04 points, 95% CI –0.1 to 0.2 points; $p = 0.65$). There was therefore no evidence to suggest that child age moderated the benefit of the IY intervention.

Figure 10 illustrates the observed effect moderation by child gender. We found no evidence that any IY effect moderation by gender varied between the trial and individual level ($p = 0.21$). We therefore again modelled a single interaction effect for gender. This moderation effect was large in size and statistically
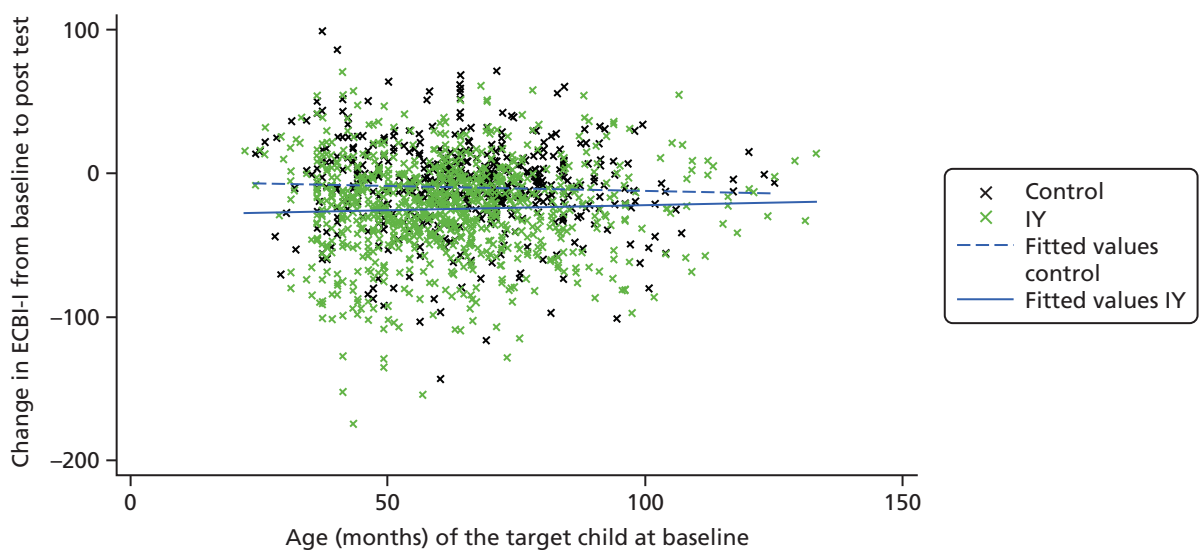
FIGURE 9 Scattergraph illustrating child age as a moderator of the effects of the IY programme on child conduct problems.
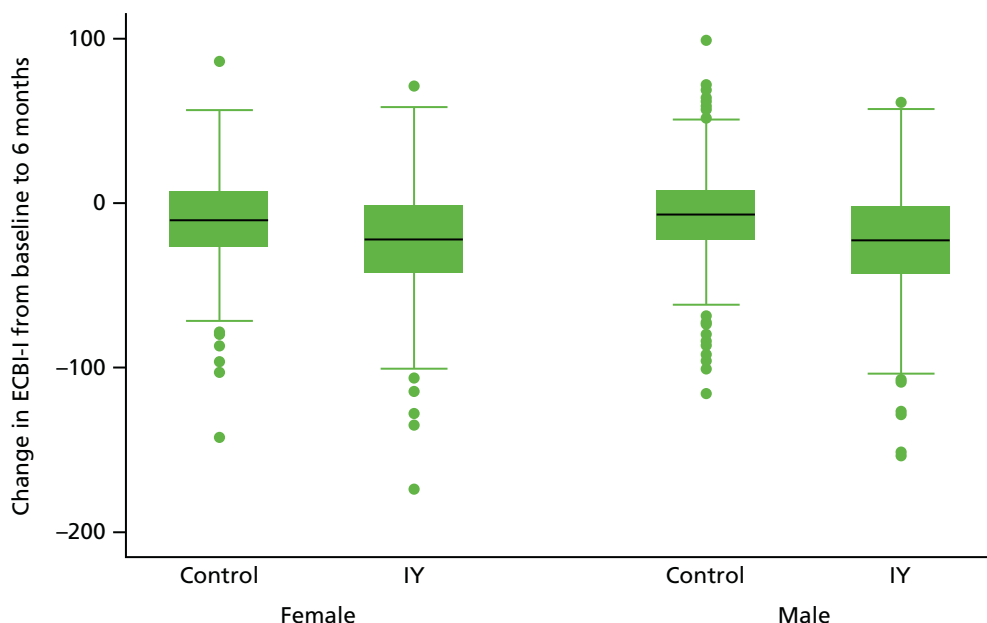
FIGURE 10 Box plots illustrating child gender as a moderator of the effects of the IY programme on child conduct problems.

significant in the formal analysis (modification index –6.6 points, 95% CI –13.0 to –0.3 points; $p = 0.04$). The direction of this effect was such that boys benefited more from the IY programme than girls.

Child disruptive behaviour

Figure 11 illustrates the observed moderation by baseline ECBI-I. We found evidence that any IY effect moderation by baseline ECBI-I varied between the trial and individual level ($p = 0.004$) but no evidence that the functional relationship between pre- and post-ECBI-I was not linear ($p = 0.09$). We therefore assessed the moderation effect separately at the trial and individual levels. At the trial level there was a large-sized and statistically significant moderation effect (modification index –18.3 points, 95% CI –24.6 to –12.0 points; $p < 0.001$). At the individual level the moderation effect remained statistically significant and
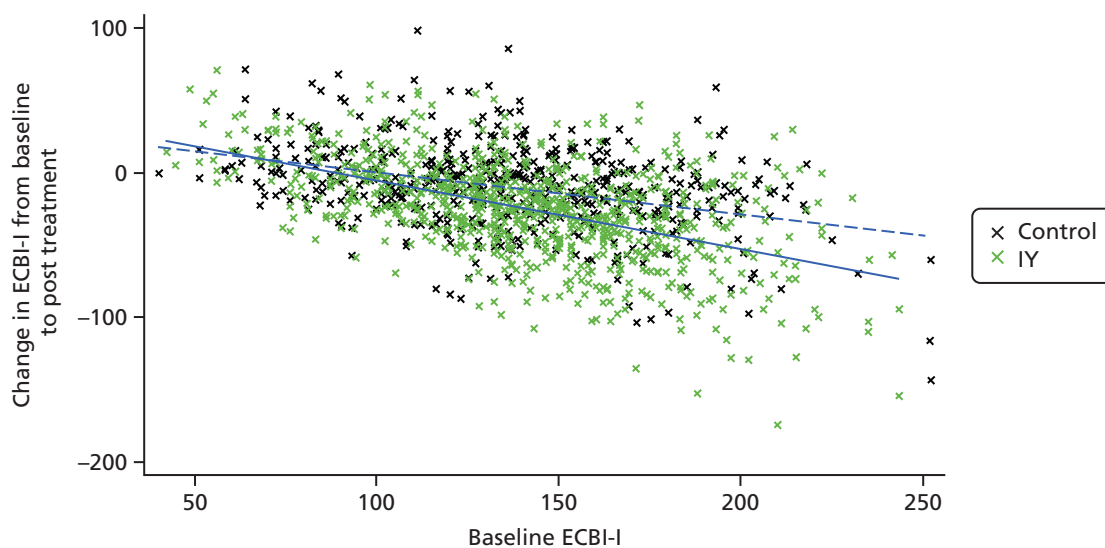
FIGURE 11 Scattergraph illustrating baseline ECBI-I as a moderator of the effects of the IY programme on child conduct problems.

in the same direction. However, its size was reduced (modification index –4.3 points, 95% CI –7.9 to –0.7 points; $p = 0.02$). We therefore concluded that the effectiveness of the IY parenting intervention to reduce disruptive child behaviour was moderated by child's level of disruptive behaviour at baseline, in the direction that children who had more severe behavioural problems at baseline benefited more from the intervention.

The solid line describes the trend over increasing baseline ECBI-I in the IY arm and the dashed line that in the control arm.

Figure 11 shows the slope of change in disruptive behaviour by level of baseline disruptive child behaviour, for control and intervention groups.

Child attention deficit hyperactivity disorder

Figure 12 illustrates the observed moderation by child ADHD. We did not find any evidence that any IY effect moderation by ADHD varied between the trial and individual level ($p = 0.58$), but there was a suggestion that the functional relationship was not linear and a quadratic term needed to be included in the model ($p = 0.02$). The test for effect moderation did not reach statistical significance at the 5% level ($p = 0.07$). We conclude that there is insufficient evidence that the effectiveness of the IY parenting intervention to reduce disruptive child behaviour was moderated by child's level of ADHD behaviours at baseline.

Figure 12 shows the change in disruptive behaviour by level of baseline child ADHD, for control and intervention groups.

Child emotional problems

Figure 13 illustrates the observed moderation by child emotional problems. We found no evidence that any IY effect moderation by these problems varied between the trial and individual level ($p = 0.28$) or that the functional relationship was not linear ($p = 0.38$). We therefore again modelled a single interaction effect for child emotional problems. This moderation effect was not statistically significant in the formal analysis (modification index –2.3 points, 95% CI –6.7 to 0.9 points; $p = 0.13$). Therefore, there was no evidence that the effectiveness of the IY parenting intervention to reduce disruptive child behaviour was moderated by child's level of emotional problems at baseline. Thus, having a high or low level of emotional problems at baseline did not significantly predict any greater or lesser benefit from IY.
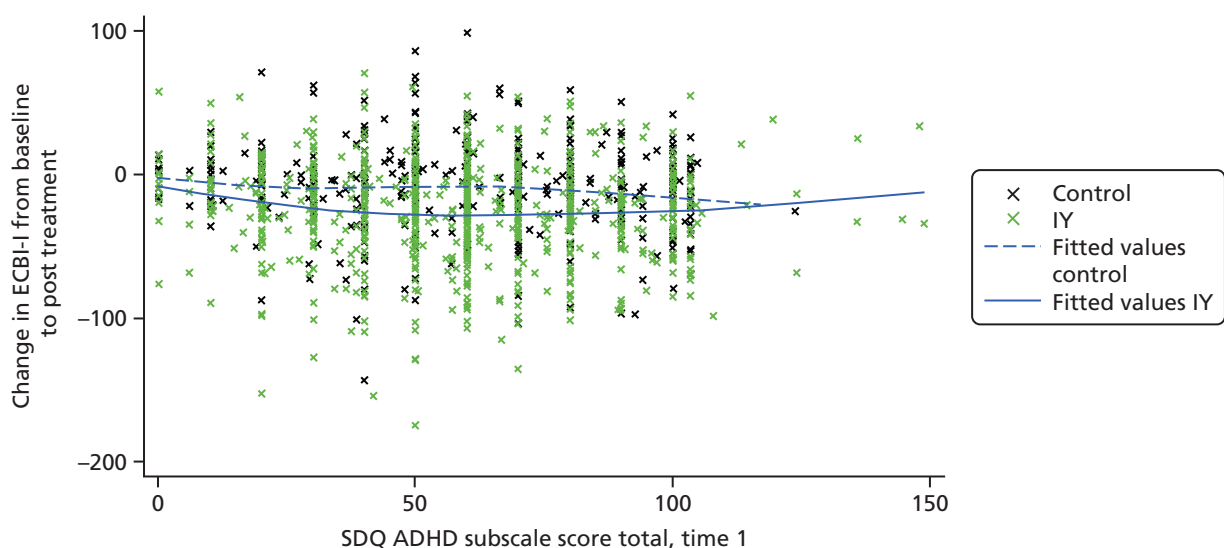
FIGURE 12 Scattergraph illustrating child ADHD as a moderator of the effects of the IY programme on child conduct problems.
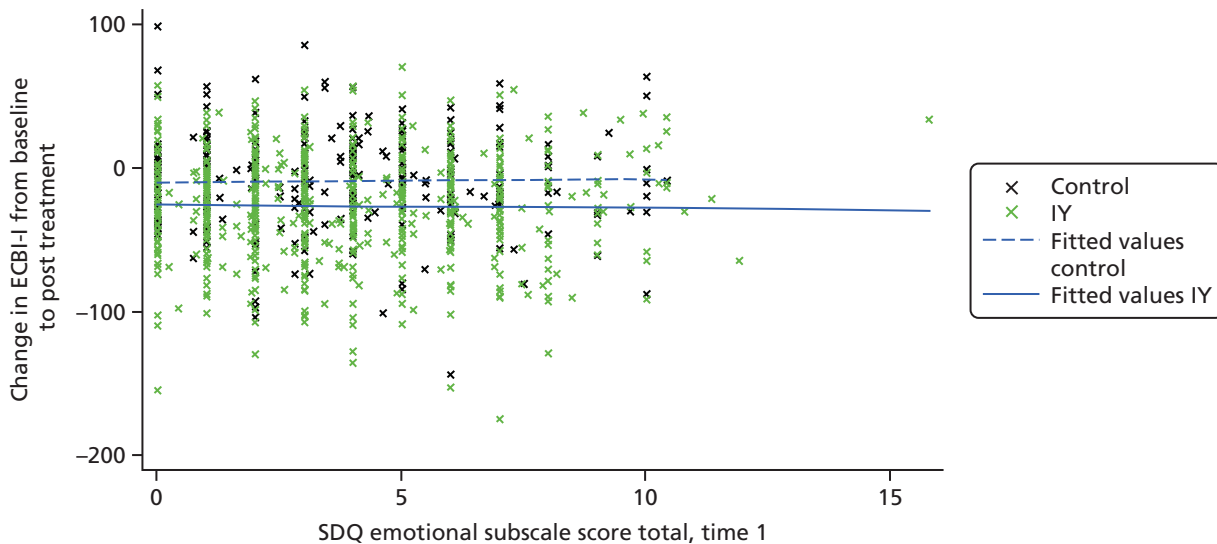
FIGURE 13 Scattergraph illustrating child emotional problems as a moderator of the effects of the IY programme on child conduct problems.

Parent characteristics as moderators

Parent depression

Figure 14 illustrates the observed moderation by parent depression. We found no evidence that any IY effect moderation by depression varied between the trial and individual levels ($p = 0.30$) or that the functional relationship was not linear ($p = 0.31$). We therefore again modelled a single interaction effect for depression. This moderation effect was of larger size and statistically significant in the formal analysis (modification index –4.8 points, 95% CI –8.4 to –1.1 points; $p = 0.01$). Thus, the effectiveness of the IY parenting intervention to reduce disruptive child behaviour was moderated by the parent's level of depression, in the direction that children whose parent was more depressed benefited more from the intervention. *Figure 14* shows the slope of change in disruptive behaviour by level of depression, for control and intervention groups.
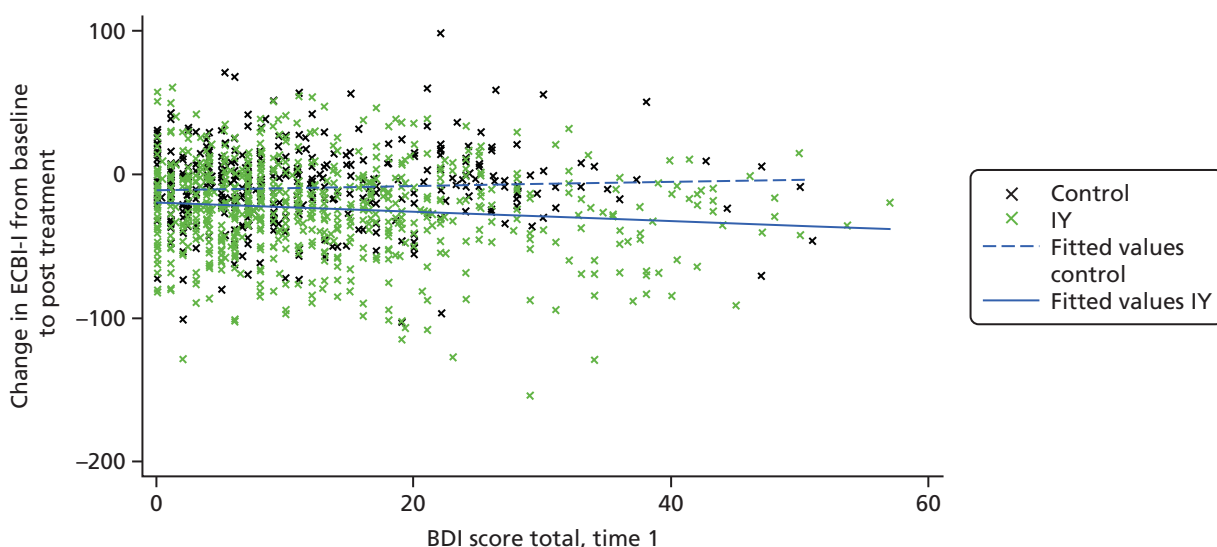
FIGURE 14 Scattergraph illustrating parent depression as a moderator of the effects of the IY programme on child conduct problems.

Parenting behaviour

All baseline, parenting variables, that is, use of monitoring, rewards, praise, corporal punishment, threatening, laxness or shouting, were tested as moderators. The results are shown in *Table 14*. None showed a significant moderation effect, suggesting that there is no differential benefit of the intervention for children whose parent has higher or lower levels of positive or negative parenting behaviour at baseline.

Individual-level moderators after adjusting for confounding

We adjusted significant moderation effects for variables that might have confounded their effect. Thus, we investigated how the moderation indices for baseline ECBI-I and depression changed after conditioning on potential confounders. The moderation effect of child gender was not adjusted, as baseline variables cannot be the underlying cause of this effect. Potential confounders considered were those variables that have shown a significant correlation with moderator under investigation in *Table 11*, choosing one representative variable per domain to adjust for when multiple variables within a domain were significantly correlated with the moderator. *Table 15* shows the results after conditioning analyses on one potential confounder at the time. This shows that the finding of a treatment effect by baseline ECBI-I is relatively robust to adjustment. The size of the moderation index estimates is little affected and the results remain statistically significant. In contrast, the detected effect moderation by parent depression is affected by an adjustment for praising. After adjustment the effect size is reduced and the effect becomes non-significant. It is therefore not clear what causes this detected effect moderation, parental clinical characteristics or parenting approach.

TABLE 15 Adjustment of significant moderators for potential confounders

Moderator	Adjusting for	Between or within trials	Moderator effect size	95% CI	*p*-value
Baseline ECBI-I	Unadjusted	Between	−18.29	−24.64 to −11.95	0.00
		Within	−4.30	−7.87 to −0.73	0.02
	Gender	Between	−17.71	−23.89 to −11.53	0.00
		Within	−4.33	−7.92 to −0.73	0.02
	Low income	Between	−16.76	−25.33 to −8.19	0.00
		Within	−4.54	−8.13 to −0.96	0.01
	Depression	Between	−14.66	−23.22 to −6.10	0.00
		Within	−1.74	−5.47 to 1.99	0.36
	Corporal punishment	Between	−17.43	−23.86 to −11.00	0.00
		Within	−4.07	−7.64 to −0.49	0.03
	Child age	Between	−18.13	−24.39 to −11.88	0.00
		Within	−4.33	−7.96 to −0.69	0.02
	Monitoring	Between	−15.36	−24.97 to −5.74	0.02
		Within	−4.93	−9.26 to −0.59	0.03
	Ethnic minority	Between	−20.41	−27.45 to −13.37	0.00
		Within	−4.47	−8.04 to −0.89	0.01
Depression	Unadjusted		−4.79	−8.43 to −1.14	0.01
	Low income		−4.62	−8.44 to −0.80	0.02
	ECBI-I		−3.40	−7.03 to 0.24	0.07
	Praise		−2.92	−8.10 to 2.26	0.27
	Corporal punishment		−3.50	−7.64 to 0.64	0.10

Trial-level moderators

We assessed whether variables characterising the setting of the trial (UK/Ireland trials vs. other trial locations) or variables affecting the delivery of the IY programme moderated outcome (i.e. percentage of staff clinically trained, percentage of staff IY certified, whether or not IY mentor was part of the trial team and whether or not any member of staff attended an international IY workshop).

Note that, as in aggregate-level metaregression, we effectively have a maximum of only 13 IY programme effects on which to base these analyses. Thus, we are able to empirically investigate only trial-level variables that have a reasonable number of replicates in the pooled data set (see *Table 13*). For example, trial sites located in rural areas were very rare, at only 4% of sites. *Table 16* shows the results for the trial-level variables for which moderation assessment was possible. The number of sessions offered was found to moderate the IY programme effect ($p < 0.001$), with the treatment benefit estimated to decrease for trials which offered more sessions. This finding, although potentially interesting and counterintuitive, is hard to interpret, as the variation between trials in number of sessions is very low. *Table 7* shows that they range (at trial level) from 12 to 1, and this would not be expected to make a substantial difference to outcome. Furthermore, some trials had booster sessions or home visits, which were not included in the analysis, and which focused on the number of IY sessions offered. None of the other trial-level variables considered was found to moderate the IY pooling effect.

Risk of bias across studies (PRISMA-IPD #22)

Risk of bias was low with regard to availability of studies, as all but one eligible trial supplied data to the pooled study. The study for which data were no longer available took place in a general practice setting in Oxfordshire.[102] There were 116 children allocated to IY or to no intervention, in a block randomised design, with eligibility based on a screening survey of all parents of 2- to 8-year-olds registered in three practices. Children scoring in the top 50% of disruptive behaviour were invited into the trial. This is a lower criterion of severity than the other indicated prevention studies in the pool. In Patterson's trial,[102] 34% were in the clinical range, whereas in many other trials 100% were in this range. Parents came from a range of social backgrounds, although generally the sample was less disadvantaged than other samples in the pool, with 14% of the families headed by a lone parent, 75% in non-manual employment and 9% from an ethnic minority.

The primary outcome was the same: disruptive child behaviour (ECBI) and the intervention yielded significant effects on this variable at 6 months' follow-up, but not post test. Effect sizes were not reported, and subgroup analyses were not planned, owing to low power. However, it was noted that there appeared to be more change in the children with high levels of behaviour problems at baseline than in those with closer to average levels. In terms of wider benefits, this study found no significant effect on

TABLE 16 Results from moderator analyses of trial-level variables

Moderator (13 replicate values maximum)	Analysis result		
	Moderation index	95% CI	*p*-value
Eight UK and Ireland vs. five non-UK and Ireland	−13.1	−37.6 to 11.3	0.29
Percentage staff certified	2.2	−38.6 to 43.1	0.91
Percentage staff clinically trained	−22.0	−52.6 to 8.7	0.16
Presence of mentor (nine trials present, four trials absent)	15.6	−10.5 to 41.7	0.24
Workshop (nine trials yes, four trials no)	12.9	−13.7 to 39.5	0.34
Number of IY sessions offered (trial-level mean, 12–14)[a]	6.1	3.6 to 8.6	< 0.001

a In trials 1, 4, 6 and 14 the number of sessions offered was varied over time; for all other trials this variable did not vary between participants. At the individual level, the range of sessions offered was 12–19.
All analyses results shown were obtained by CC analysis.

parent depression, child emotional problems or ADHD symptoms. Although there are clearly some differences between this trial and the others in the pool, there is no particular reason to think that our findings would be affected by inclusion of this trial. We note that at trial level, the findings of Patterson *et al.*[102] echo the broad pattern found in our analyses, whereby samples with lower levels of conduct problems showed weaker effect sizes.

Risk of bias was low with respect to the intervention evaluated, as all trials evaluated the same intervention. Moreover, the IY parenting intervention is strictly protocolled, including training and supervision of the staff. The intervention as evaluated is therefore expected to be similar, broadly speaking, across trials.

Risk of bias may be higher with regard to the data available, and how data for some of the constructs were synthesised. Not all trials included the same measures for the same constructs. This means that for some constructs (e.g. self-reported parenting practices) data came from up to four different instruments. Although these measures individually showed good individual psychometric properties and explicitly aimed to measure the targeted construct, we were not always able to test the extent to which scores across measures correlated with each other. A potential bias of this procedure is that the instruments may not have measured exactly the same construct. Please note, however, that in cases for which we were able to examine correlations between measures targeting the same construct, these were reassuring. For example, correlation between the two measures used for the primary outcome was around.[70]

Relatedly, it was inevitable to make several assumptions in the data harmonisation process. For example, we used the same norm scores across trials from different countries, because using different norm scores for different samples would have differentially changed the distribution of scores within the samples. Moreover, norm scores were often not available for every measure from every country (e.g. Portugal and Ireland). We therefore chose to use one set of norm scores (i.e. the instruments original and validated norm scores) for all samples. A potential bias of this procedure is that the original norm scores may not reflect the actual norm scores in another country.

Some relevant variables (e.g. SES and ethnicity) were not evenly distributed across trials. We controlled for trial-level variance in all the analyses, which was essential because several variables were confounded on the trial level (e.g. trials with more ethnic minority families were more often secondary prevention trials that include families with less severe disruptive behaviour problems, rather than indicative prevention or treatment trials that include families with more severe disruptive behaviour problems.

Discussion of moderator results

Our IPD meta-analysis included a near complete sample of European randomised trials of the IY parenting programme, with data pooled for 14 out of the 15 trials that were completed by 2015. Trial publication dates ranged from 2001 to 2015, with some still to be published. With close to 1800 participants, it represents the largest ever data set on randomised evaluation of a parenting programme, providing a uniquely large sample and considerable heterogeneity in terms of countries, settings, ethnicity, and socioeconomic and clinical characteristics. Thus, it is able to yield high power and generalisability for testing moderators of change in child disruptive behaviour. It should be noted that 1800 represents the maximum possible sample size; for moderator questions, there were suitable data from a maximum of 13 trials (maximum $N = 1696$), as one trial involving young toddlers lacked data on the primary outcome.

Social and socioeconomic disadvantage
The primary question for this individual participant meta-analysis was whether or not there are differential effects of the IY parenting programme on child disruptive behaviour for families with higher levels of social and socioeconomic disadvantage. The findings were quite clear: in our primary analyses, based on pooled individual-level data (i.e. within-trial analyses), there were no moderator effects, meaning that families who were socioeconomically disadvantaged were just as likely to benefit from the intervention as those who

were not. This finding applied across a range of indices of social and economic disadvantage, including low income, low educational level, being a lone parent or a teenage parent and having no employed person in the household. We can be reasonably confident in these findings in several respects. First, for three of the moderator variables, there were near-complete data from all 13 trials (5–7% missing); data for teenage parenthood were unavailable for one trial, and for unemployment for three trials. Second, the primary analyses used MI to account for missing data but closely similar results were obtained from CC analysis (see *Appendix 5*, *Table 30*). On the other hand, moderator effects were not particularly consistent across trials; this is likely to be a result in large part of the problem of small sample size in each trial overall, with this being especially marked for some moderator variables. Thus, here and in other analyses, we do not attempt to interpret between-trial variability in size of the moderator indices.

For each moderator variable, we examined whether or not there was any difference between the moderator effects at the between- and within-trial levels. When a difference was found, we then modelled both effects. This was the case for one variable: teenage parenthood. There was no within-trial moderator effect for teenage parent, but a trend towards a between-trial effect ($p = 0.1$) in the direction that trials with more teenage parents tended to have higher effect sizes. However, this direction of effect was not seen in the within-trial findings, which trended non-significantly in the opposite direction. It should be noted that between-trial effects are unlikely to be important: they are based on a very small sample size, representing the number of trials included in the analysis, which in this case was 12. The small sample size also means that they are likely to be subject to hidden confounding, related to particular characteristics of trials that happen to covary with percentage of teenage parents.

These pooled data findings help to clarify a previously very mixed picture in the literature, in which highly cited reviews have concluded that children from socially disadvantaged families benefit less from parenting interventions to reduce disruptive behaviour.[43,44] On the other hand, one recent narrative review[45] and some meta-analyses[12,149] concluded that there were few differential effects of social and socioeconomic disadvantage. However, these studies suffer from important drawbacks, which are overcome by our study. First, many studies are limited by examining predictors rather than moderators of outcome.[43,44] Second, most reviews index moderator variables only at the aggregate trial level.[12,43,149] Finally, many studies, although testing moderation at individual level, are instead limited by analysing or synthesising data from small trials, which lack power to detect interaction effects.[45,46] Even when trials are much larger, findings can be hard to interpret or replicate; for example, one US trial found no moderator effects of low income or teenage parenthood on outcome but found stronger effects for low educated parents along with diminished effects for lone parents.[47] Given a prominent body of literature concluding that socially disadvantaged families fare less well in parenting interventions, but which is based on less than ideal methods, our findings are of particular significance. Our study overcomes these three major limitations and points to a clear and more optimistic conclusion, namely that families disadvantaged by low income and low educational level, or by lone parenthood, are just as likely to benefit. Furthermore, there are clear implications for equity effects of parenting interventions: if taken to scale, these interventions are unlikely to result in further widening of existing social inequalities in child disruptive behaviour.

It is worth noting that many of the qualitative studies in the field[49] point to parents feeling that family social and economic stresses contribute to low levels of engagement and success in a parenting intervention. Although not invalidating individual experience, which of course is hugely variable, our data suggest that at a group average level, the stresses associated with low income and unemployment do not systematically affect the outcome of these interventions, at least once families have been recruited in to the intervention. Generally, the families in our focus groups (see *Appendix 6*) were also unsurprised about the results; they were of the view that factors such as educational level of the parent should not make a difference to outcome, as the IY materials, they felt, would be accessible to a wide range of parents. However, in one qualitative study of low-income families attending IY in Ireland,[50] parents perceived different social barriers as more salient, including antisocial behaviour in the neighbourhood, and disagreement with their partner about implementing the programme. This may illuminate a possible reason why lone parenthood did not function as a moderator. Although lack of a partner may lead to feeling unsupported in a new

parenting role, having a partner who is unsupportive may have a similar effect. Rather than perceiving life stressors as straightforward barriers, many parents stressed their commitment to implementing parenting skills in the face of difficulties, helping them to maintain positive child outcomes in the longer term.[150]

Ethnicity

Our findings showed no moderation effect by ethnicity; thus, children from an ethnic minority family were just as likely to benefit as those from an ethnic majority family. There is relatively small literature with which to compare these findings, as there have been few studies examining ethnicity as a moderator of parenting interventions, especially in Europe. However, the findings are consistent with data from a large predictor study in the USA,[54] a moderator analysis in a large trial of the Family Check-Up in the USA[151] and in a smaller trial in London, UK.[59]

Although our data are largely consistent with the modest body of prior work on predictors and moderators, our large pooled study design means that we can have considerably more confidence in this important finding of no moderation by ethnicity. Our finding is also likely to be generalisable across ethnic groups, as families were drawn from a wide range of ethnic groups typical of the UK and Dutch cities where the trials were conducted. The groups in the four London trials were particularly variable, with the largest ethnic groups being African and African-Caribbean families. In Birmingham the largest group was Asian, and in the Netherlands Utrecht trial, it was North African and Middle Eastern families. In the Dutch prisoner's trial, however, most mothers were Caribbean or Latin American, reflecting the female prison population there.

However, it should be noted that our findings are not consistent with much current practice and policy. There is a frequent assumption that interventions developed by (and often for) people from the majority ethnic group need to be adapted or redesigned for other ethnic groups. This is reflected in an extensive literature on cultural adaptation of parenting interventions.[55,56] There are many good reasons for altering interventions for particular cultural groups, including that fact that parenting values and practices vary across cultures, and the importance of involving communities and families in the design and ownership of new interventions.[33] The assumption pervading this literature is that these differences will result in poor engagement with, or poor effectiveness of, interventions for minority groups. However, this is not borne out by our study, and the parents in our focus groups did not agree with this view. Furthermore, in the absence of evidence for moderation by ethnicity, it becomes more important to consider a potential harm from developing adapted parenting interventions for different cultural groups, namely that parenting services would then need to be organised separately by ethnicity. In some contexts, this would be impractical, and arguably is not desirable. Instead, others have argued that parenting interventions should be designed not as different versions for different groups, but with flexibility built in, in order to respond to the range of diverse family values and goals found in our communities.[152] These differences may be based on ethnicity, or on socioeconomic, educational, regional and personal differences. This is very much in keeping with the collaborative processes that pervade the IY programme training and delivery [see *Research questions (PRISMA-IPD #4)*], and which explicitly address cultural diversity.[60]

Our findings potentially have implications for understanding the extent to which parenting programmes can be transported across countries as well as cultures. This is a programme that has been transported from the USA to several different parts of Europe, which differ from the USA in terms culture and ethnicity, as well as service infrastructure and values in the area of public health and family services. Yet despite such cross-country transportation, the programme appears equally effective with ethnic minority families within these new countries. This is in keeping with a recent meta-analyses exploring applicability of parenting programmes across cultures by examining their transportability across countries,[30] which found that effect sizes were generally just as strong as in their country of origin when these programmes were taken to new countries. Moreover, Gardner *et al.*[30] found that effect sizes did not vary with country-level characteristics, such as cultural values around parenting or level of child welfare provision.

A limitation of our study is that moderator analyses can address whether or not families engage in and benefit from the intervention only after they have been recruited into the study. It cannot address whether or not ethnicity (or SES) affects access and initial recruitment into an intervention. However, some trials of IY partially address this issue by reporting data on families enrolled in a trial compared with those who were not. For example, Patterson et al.[102] found that families randomised did not differ in SES or level of problem behaviour from those who declined the intervention but who attended the same universal general practice service and were screened as eligible for the trial.

It might be that versions of a programme explicitly created by and for ethnic groups would increase the chances of minority families accessing programmes. However, it is important to be cautious about the effects of creating new versions of programmes, and to test these properly in new trials, as the trial by Gottfredson et al.[153] in Washington, DC, found that an adapted parenting programme improved engagement and retention of minority families, but intervention effects were much reduced compared with trials of the original intervention.

Child characteristics

Gender
Our analyses found significant moderation by gender, such that boys benefited more than girls. This is in keeping with some other smaller studies, with some authors suggesting that this might be explained by the fact that boys show higher baseline levels of disruptive behaviour and that they are likely to benefit more. We did not formally test for disruptive behaviour as a confounder of the moderation effect, as, strictly speaking, disruptive behaviour could not be a confounder for (cause of) the gender effect. However, we might view disruptive behaviour as mediating the effect of gender on outcome, and indeed this appeared to be the case, as there was a substantial gender difference in disruptive behaviour at baseline (mean of 10 points on ECBI-I), and when we controlled for disruptive behaviour, the moderation effect by gender disappeared ($p = 0.26$). We may conclude that it is unlikely that there is anything about the intervention that is less well suited to helping parents to deal with their girls; rather, it looks as if this greater benefit to boys may be related to their higher levels of disruptive behaviour.

Child age
Our analyses found no moderation by age, such that children were equally likely to benefit in all parts of the age range, from age 2 to 10 years. It should be noted that, although there were a good number of children outside the age range 3–8 years, they nevertheless represent < 10% of the pooled sample, suggesting that we can be more confident about the findings within the 3–8 years range. This finding of no age effects on child outcomes is of particular note, given the strong policy thrust towards early intervention, primarily in the preschool years.[62] However, there is no evidence from our data that preschool children are any more likely to change their behaviour in response to this parenting intervention than older children of primary school age. Almost half of the sample were under 5 years of age, which is the statutory school age in the UK. This helps to clarify a very mixed pattern of conclusions in the parenting literature, based on small trials and often out-of-date aggregate data meta-analyses. It suggests that services should focus on a wider age range for parenting interventions, as they are just as likely to be useful for reducing disruptive behaviour in primary school-aged children as in children of preschool age. Age was not correlated with severity of disruptive behaviour, meaning that severity cannot help us to explain these findings.

Child disruptive behaviour at baseline
We found evidence of moderation by baseline level of disruptive behaviour, such that children with more severe problems at baseline were more improved post test. This effect appeared robust in the face of potential confounders; when we adjusted for low income, gender, parent depression, praise, corporal punishment and child age, the pattern of findings remained significant and closely similar. This suggests that these other factors are not explaining the relationship between baseline disruptive behaviour level and outcome. It is worth noting that there was also a strong moderator effect at the between-trial level, such that trials with children with high level of behaviour problems also tended to have stronger average effect

sizes. We note, however, as before, that between-trial analyses are based on a very small trial sample of 13, and are more likely than within-trial analyses to be confounded by other trial-level variables. In this case, the between-trial moderation is consistent with that at the within-trial individual level.

This finding of greater benefit to children with more severe problems is consistent with much of the existing literature.[45] Parents in our focus groups also suggested there might be greater benefit to these families, owing to parents with more difficult children having higher levels of motivation to change, a factor explored to some extent in the literature.[45,154] This moderating effect of initial problem severity is sometimes dismissed as being merely 'regression to the mean'; however, this cannot explain the finding, as such an effect would apply equally to the control group. The scatterplot for level of disruptive behaviour shows quite clearly that reduction in ECBI-I score is greater for all children, that is, for those in the control and intervention groups if their score is greater at baseline. This could reflect regression to the mean. However, over and above this effect in the control group, there is nevertheless a differentially greater effect of baseline ECBI-I in the intervention group.

This finding is potentially reassuring to parenting group leaders, who may feel that they struggle to achieve change in the most difficult children. Rather than endorsing an expectation of possible failure with these families, these results in fact show that (on average) the intervention is making a greater difference to outcomes in the neediest families than in children with fewer problems. It is often the case that older children who present to indicated prevention and treatment services may have somewhat more severe problems (although interestingly this was not the case in our data set), and our findings show that on the grounds of neither age nor severity should we be pessimistic about intervening with older children or fail to target these groups.

Child attention deficit hyperactivity disorder and emotional problems at baseline

Evidence for moderation by baseline ADHD was less clear. There was a borderline moderation effect of ADHD on disruptive behaviour outcome ($p = 0.07$), suggesting that the evidence was insufficient to be conclusive. There was also evidence that the effect was not linear; hence a quadratic term was included. If we were to very cautiously interpret the trends that appear in the chart, then we might suggest that across much of the middle range of ADHD scores (score 4–10 on SDQ) IY is equally effective. However, it appears that the intervention might be less effective at very high and low levels of ADHD. It is important to note, however, that these effects do not reach significance and would need to be replicated. We found no evidence for moderation by baseline emotional problems, suggesting that IY is equally likely to be effective for children who have high or low levels of emotional problems.

There is limited literature on ADHD or emotional problems as moderators of parenting interventions for disruptive behaviour, so there is little with which to compare these data; in general, it is often thought that children with comorbid problems may be harder to treat, partly because, without interventions, they generally have worse outcomes. This pessimism persists despite literature suggesting the contrary, albeit based on analysing on predictor and not moderator effects.[155] Hence these findings should be reassuring to practitioners and other decision-makers, in that, from these data, this is clearly not the case. If it were the case that ADHD rendered parenting interventions for disruptive behaviour less effective, then it might be important to consider separately treating the ADHD, for example with medication. There is evidence that these interventions directly improve ADHD as well,[68,116,117] which we test in our data set in *Chapter 4*.

Parent characteristics

Parent depression

We found evidence of moderation of disruptive behaviour outcomes by baseline level of parent depression, such that the children of parents who were more depressed were more improved post test than the children of parents who were less depressed. This was a marginally stronger effect than for baseline disruptive behaviour, and was robust in the face of adjusting for some of the potential confounders: when we adjusted for low income, laxness and child disruptive behaviour, the pattern of findings remained

largely similar. However, adjusting for parental praise removed the moderator effect, suggesting that both of these variables (low praise and high depression at baseline) may potentially contribute to the moderation effect. We would suggest, however, that it is more likely that depression would contribute to low levels of praise than the other way around, suggesting that praise mediates rather than confounds this relationship. It has been found in some studies that depression is associated with parents showing low levels of praise to their child,[156] and it is plausible that, following intervention, parents who are depressed are especially able to increase their levels of praise, from a low base rate, which in turn promotes change in their child's problem behaviour. This could be tested in future studies in a mediated moderation model.[157]

Our findings of moderation by parent depression are helpful in clarifying a small and mixed literature on this topic. A recent review found only four studies of parent depression as a moderator of parenting intervention effects on child behaviour,[45] of which two found the same moderator effect as ours (Gardner et al.,[46] based on the Welsh Sure Start trial, part of our pooled data set; Shaw et al.[35]), and two found no moderation by depression. This is in contrast with the findings of many trials and systematic reviews in which depression was examined as a predictor rather than moderator, which have tended to conclude that there was reduced benefit for children with a parent who is depressed.[44] Parents in our focus groups also took the view that the intervention would be helpful to parents who are depressed, as well as to those who were not depressed. Interestingly, a couple of parents who had been depressed themselves thought that IY would have differentially greater benefit for parents who are depressed, because depression creates such challenges in dealing with everyday parenting difficulties, and because they found that small successes in the programme could be very mood lifting.

Again, the implications for practice and policy are important. Parental depression is a significant risk factor for a host of poor outcomes in children, including problem behaviour, and it is reassuring that not only is this parenting intervention helpful for parents who are depressed, but also it appears to be especially beneficial for their children, and as such may help to reduce some of the inequalities in outcomes for these children compared with the children of parents who are not depressed. It suggests that it may not be necessary for depression to be treated or improved before there can be any beneficial effect of a parenting intervention. Indeed, it is possible that the intervention itself may improve depression, as has been found in a number of trials and systematic reviews,[11,85] a question that we test in *Chapter 4*.

Parenting skills

Our findings show that none of the parenting variables at baseline, negative or positive, moderates the effect of the parenting intervention on child disruptive behaviour, suggesting that the intervention is suitable and beneficial for parents across a range of levels and types of parenting skill. It suggests that parents who are very harsh with their children, or who lack warmth or positive parenting skills, are just as likely to benefit as those who are at lower risk in terms of their parenting when they begin the intervention. There have been very few prior studies testing parenting skill as a moderator, indeed a review by Shelleby and Shaw[45] found only one such study that concerned a variable allied to parenting skill, namely parent–child relationship quality. A study by Tein et al.[36] of a preventative parenting intervention for families going through divorce found that poor parent–child relationship quality at baseline was associated with greater child improvement following intervention. Poor parent–child relationship quality is likely to partly reflect poor parenting skills but also to reflect the child's level of behavioural difficulties. It is possible that the latter helps explain this moderator effect. In any case, these findings are consistent in the sense that in neither study did children in the highest-risk families, in terms of parenting skill, appear to fare any less well following a parenting intervention than those from lower-risk families.

Trial-level variables

The question of how contextual and implementation factors affect outcome of parenting interventions is clearly of huge importance. However, most of these variables vary only at the level of the trial, not the individual, and hence our findings are based on 13 trials at most. These analyses suffer from low power, and have much potential for unmeasured confounding at trial level. The only significant finding concerned the variation between trials in the number of IY sessions offered, with more sessions associated with lower

effect sizes. It seems of limited interest, because the variation between trials is very limited and a causal relationship is not especially plausible; it seems unlikely that offering two extra sessions would make child outcomes worse. Thus, this finding is probably due to some other feature of the trials that had more sessions. Other trial-level variables showed either little or no variability, and were not analysed, or showed no relationship with outcome. There are over 200 randomised trials of parenting interventions in this age group, so, in this field, large conventional meta-analyses are likely to be more practical and better powered to answer contextual questions.[158]

Chapter 4 Results and discussion of wider health benefits and possible harms

Preliminary analyses

All included secondary outcome measures of the individual trials were examined to test for wider health benefits and possible harms of the intervention. Measures with too few data (e.g. from < 350 of the 1799 families or from fewer than three trials) were excluded, as these would be able to provide only very limited information. A total of 12 secondary outcome measures were included. These were in the domains of children's wider mental health (ADHD symptoms and emotional problems), parental mental health (depression, parenting stress and feelings of self-efficacy), self-reported harsh and inconsistent parenting practices (corporal punishment, threatening, laxness and shouting) and self-reported positive parenting practices (use of praise, tangible rewards and monitoring). *Table 17* provides an overview of the number of available data across trials and families on these constructs, and *Table 18* provides descriptive data.

TABLE 17 Available data per trial for wider health benefits and potential harm analysis (numbers post test)

Secondary outcome variable	Number of Trials	Families	Trial number 1	2	3	4	5	6	7	8	9	10	11	12	13	14
Children's mental health																
ADHD symptoms	11	1219	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Emotional problems	10	1055	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
Parental mental health																
Depressive symptoms	11	1131	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	No	No	No	No
Parenting stress	5	502	Yes	No	No	Yes	No	No	Yes	Yes	No	No	Yes	No	Yes	No
Feelings of self-efficacy	4	384	No	No	Yes	No	No	No	No	Unclear	No	Yes	Yes	Yes	Yes	Yes
Harsh and inconsistent parenting practices																
Corporal punishment	10	1038	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Threatening	9	987	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Laxness	9	945	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Shouting	9	882	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Positive parenting practices																
Praise	6	460	Yes	No	Yes	No	Yes	Yes	No	No	No	Yes	No	Yes	No	No
Tangible rewards	6	544	Yes	No	Yes	No	Yes	Yes	No	No	No	Yes	No	Yes	No	No
Monitoring	9	959	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes

TABLE 18 Descriptive data pre and post intervention by randomised group

Outcome variable	Control		IY	
	n	Mean (SD)	*n*	Mean (SD)
ECBI-I total baseline	611	135.5 (37.0)	1011	139.4 (37.0)
ECBI-I total post test	567	125.5 (37.9)	878	116.2 (34.7)
SDQ ADHD baseline	589	5.8 (2.7)	943	5.9 (2.7)
SDQ ADHD post test	483	5.8 (2.6)	736	5.2 (2.7)
SDQ emotional baseline	491	3.2 (2.4)	849.0	3.4 (2.7)
SDQ emotional post test	396	2.8 (2.2)	659	2.8 (2.3)
Monitoring baseline	394	5.2 (1.7)	694	5.3 (1.7)
Monitoring post test	368	5.3 (1.7)	591.0	5.4 (1.6)
Tangible rewards baseline	243	3.3 (1.3)	382	3.3 (1.2)
Tangible rewards post test	229	3.4 (1.3)	315.0	3.6 (1.2)
Praise baseline	255	4.5 (1.3)	399	4.7 (1.3)
Praise post test	187	4.8 (1.1)	273.0	5.2 (1.2)
Corporal punishment baseline	527	2.2 (1.6)	853	2.1 (1.5)
Corporal punishment post test	396	2.4 (1.6)	642.0	2.0 (1.4)
Threatening baseline	400	3.6 (1.6)	682	3.5 (1.6)
Threatening post test	381	3.2 (1.5)	606.0	2.8 (1.5)
Laxness baseline	392	3.3 (1.3)	667	3.3 (1.3)
Laxness post test	369	3.3 (1.2)	576.0	3.1 (1.2)
Shouting baseline	475	2.9 (1.7)	751	2.9 (1.4)
Shouting post test	325	3.0 (1.3)	557.0	2.7 (1.3)
Parent depression BDI total baseline	571	10.2 (9.8)	909	12.2 (10.9)
Parent depression BDI total post test	453	8.7 (9.0)	678.0	8.7 (9.1)
PSI-SF total baseline	181	89.0 (28.4)	361.0	92.1 (28.4)
PSI-SF total post test	180	82.9 (34.9)	322.0	80.5 (33.6)
PSOC scale total baseline	181	54.1 (7.6)	236.0	54.0 (7.6)
PSOC scale total post test	165	59.4 (13.2)	219.0	55.7 (14.1)

Baseline values of children's ADHD symptoms were on average around the threshold for borderline ADHD symptoms. More specifically, 53% of the children scored above the threshold for borderline problems and 29% of all children scored above the clinical threshold for ADHD symptoms. The data showed a similar pattern for children's emotional problems, with 28% of the children scoring above the clinical threshold for emotional problems. Compared with ADHD symptoms, however, a larger proportion of children scored within the normal range: only 40% of the children scored above the borderline threshold of emotional problems.

In line with gender differences in the severity of disruptive behaviour in this sample, boys showed higher levels of ADHD symptoms (mean 6.07) than girls (mean 5.36). No such pattern existed for emotional problems. Boys and girls showed similar levels of emotional problems (mean 3.34 and mean 3.38, respectively). Neither ADHD symptoms nor emotional problems correlated with children's age. ADHD symptoms and emotional problems were significantly, but weakly, correlated with each other ($r = 0.26$; $p < 0.001$).

Overview of findings

Table 18 shows descriptive data and *Table 19* shows an overview of the results of the formal analyses of wider health benefit outcomes of the intervention. For these analyses, effect sizes are denoted as 'β'. These represent standardised group differences that express the estimated difference in units of baseline SDs, thus allowing comparison of effect sizes across different variables in *Table 19*. The intervention significantly improved children's ADHD symptoms, increased parental use of praise and reduced the use of corporal punishment, such as threatening and shouting, as assessed by parent report. The intervention did not significantly affect children's emotional problems, parental depression, parenting stress or feelings of self-efficacy, parental laxness or parental use of tangible rewards and monitoring. We report the results of the univariate analyses. Univariate and multivariate analyses results did not differ substantially.

Children's mental health (attention deficit hyperactivity disorder symptoms and emotional problems)

The intervention had wider health benefits on children's ADHD symptoms (β = –0.30, 95% CI –0.44 to –0.17; see *Table 19*). Children whose parents had participated in the parenting intervention showed fewer ADHD symptoms than children whose parents had not participated in the parenting intervention. In families that received the intervention, the percentage of children who scored above the borderline threshold on ADHD symptoms fell from 54% to 42%, which is more than a 10% reduction. In families that did not receive the intervention, the percentage of children who scored above the borderline threshold on ADHD symptoms fell from 52% to 50%, a reduction of only 2%.

TABLE 19 Wider health benefits of the IY parenting intervention on children's mental health, parental mental health and self-reported parenting practices

| Secondary outcome variable | *n* at post test | Analysis | | | | | |
| | | Univariate | | | Multivariate | | |
		Standardised difference estimate (β)	*p*-value	95% CI	Standardised difference estimate (β)	*p*-value	95% CI
Children's mental health							
ADHD	1219	–0.30	0.000	–0.44 to –0.17	–0.28	0.000	–0.41 to –15
Emotional problems	1055	–0.06	0.303	–0.18 to 0.06	–0.01	0.933	–0.13 to 0.11
Parental mental health							
Depression	1131	–0.08	0.095	–0.17 to 0.01	–0.08	0.158	–0.19 to 0.03
Parenting stress	502	–0.18	0.164	–0.44 to 0.07	–0.08	0.280	–0.21 to 0.06
Parental self-efficacy	384	–0.32	0.165	–0.77 to 0.13	–0.31	0.083	–0.66 to 0.04
Positive parenting practices							
Praise	460	0.26	0.045	0.01 to 0.51	0.28	0.001	0.12 to 0.44
Tangible rewards	544	0.15	0.347	–0.16 to 0.45	0.17	0.077	–0.02 to 0.35
Monitoring	959	0.05	0.434	–0.08 to 0.18	0.03	0.625	–0.10 to 0.16
Harsh and inconsistent parenting practices							
Corporal punishment	1038	–0.22	0.004	–0.42 to –0.01	–0.19	0.005	–0.32 to –0.05
Threatening	987	–0.21	0.007	–0.36 to –0.06	–0.20	0.003	–0.34 to –0.07
Laxness	945	–0.15	0.174	–0.37 to 0.07	–0.15	0.045	–0.29 to 0.00
Shouting	882	–0.31	0.041	–0.61 to –0.01	–0.22	0.001	–0.35 to –0.08

Earlier findings about the extent to which parenting interventions designed to reduce conduct problems can reduce children's symptoms of ADHD are inconsistent. Some review studies have suggested that effects may depend on the type of instrument or informant used (e.g. parents vs. teachers[93]). Even within the same informant, however, results are strikingly inconsistent. Some individual trials find that the IY parenting intervention can successfully reduce ADHD symptoms,[68,69] whereas others found no effects on children's ADHD symptoms (e.g. Leijten *et al.*[96]). Combining individual family-level data from > 1200 families from 11 trials led us to conclude that one of the wider health benefits of the IY parenting intervention is that it does reduce parent-reported ADHD symptoms in children.

The intervention did not have wider health benefits on children's emotional symptoms ($\beta = -0.06$, 95% CI -0.18 to 0.06; $p = 0.303$). There seemed to be a reduction in children's emotional problems, regardless of intervention status. In families that received the intervention, the percentage of children who scored above the borderline threshold on emotional problems fell from 40% to 32%, which was an 8% reduction. In families who did not receive the intervention, the percentage of children scoring above the borderline threshold on emotional problems also fell, from 40% to 32%.

Evidence of the extent to which parenting interventions designed to reduce conduct problems reduce children's emotional problems is limited. Some trials have reported a measure of children's emotional problems as a secondary outcome, but most did not find any effects of the intervention on this measure (e.g. Leijten *et al.*[96]). However, some trials have specifically focused on the extent to which there are wider health benefits in relation to children's emotional problems and these do show the hypothesised effects (e.g. Herman *et al.*[95]). Our finding that the intervention does reduce ADHD symptoms, and not emotional problems, may not be surprising. ADHD symptoms more often co-occur with conduct problems than emotional problems.[91] When the intervention successfully reduced conduct problems, ADHD symptoms may therefore have been more easily reduced than the more unrelated emotional problems.

These findings have important implications for policy-makers and practitioners. They suggest that an intervention for one type of externalising behaviour (i.e. conduct problems) may have wider benefits on other types of externalising behaviour (in this case ADHD symptoms). They also indicate, however, that the intervention should not be expected to have wider benefits for children's internalising behaviour (i.e. emotional problems).

Parental mental health

The intervention did not affect parental mental health. There was a trend that the intervention reduced parental symptoms of depression ($\beta = -0.08$, 95% CI -0.17 to 0.01), but this effect did not yield significance. Neither parenting stress ($\beta = -0.18$, 95% CI -0.44 to 0.07) nor feelings of parental self-efficacy ($d = 0.32$, 95% CI -0.77 to 0.13) improved as a result of the intervention (see *Table 19*).

This finding is surprising. Parental mental health is not consistently included as an outcome measure of parenting interventions but is often reported to improve as a result of parenting interventions,[11,159] including IY (e.g. Hutchings *et al.*[75,85]). Moreover, several studies suggest that improvements in parental mental health may be one of the mechanisms through which the intervention affects children's behaviour problems (e.g. Hutchings *et al.*[71] and Shaw *et al.*[159]).

Harsh and inconsistent parenting practices

The intervention successfully reduced harsh and inconsistent parenting practices. Importantly, results were robust across three of the indicators of harsh and inconsistent parenting practices: corporal punishment ($\beta = 0.22$, 95% CI -0.42 to -0.01), threatening ($\beta = -0.21$, 95% CI -0.36 to -0.06) and shouting ($\beta = -0.31$, 95% CI -0.61 to -0.01) (see *Table 19*). In the univariate analysis the intervention did not reduce parental

laxness ($\beta = -0.15$, 95% CI –0.37 to 0.07), but in the multivariate analysis the treatment effect on laxness reached statistical significance. Parents who had participated in the intervention thus consistently reported less harsh parenting practices, although not necessarily less inconsistent parenting practice.

These findings correspond with earlier work on the effects of parenting interventions on negative parenting practices. Earlier work, however, often showed inconsistencies and some elements of harsh and inconsistent parenting practices, especially corporal punishment, have not always been shown to be affected by the intervention (e.g. Posthumus et al.[160]). Some of these inconsistencies may have been because of the instruments used. Importantly, we combined data across four types of instruments on self-reported parenting practices. Our results thus do not hinge on a specific instrument.

That the IY parenting intervention reduces harsh parenting is an important wider health benefit outcome. Even mildly harsh parenting is associated with harmful biological effects on children.[90] These harmful effects include dysfunctional cortisol secretion patterns and raised C-reactive protein. These in turn are associated with increased cardiovascular disease and mortality.[90] Reducing harsh parenting thus is an important wider health benefit outcome, above and beyond its role in reducing conduct problems in children. Our finding that the parenting intervention indeed reduced harsh parenting, consistent across three different indicators, may be encouraging for policy-makers and practitioners who aim to reduce negative parenting practices and their harmful consequences.

Positive parenting practices

The intervention improved some aspects of positive parenting practices. Parents who had received the intervention reported praising their children more frequently ($\beta = 0.26$, 95% CI 0.01 to 0.51). They did not report, however, using more tangible rewards for their child's positive behaviour ($d = 0.15$, 95% CI –0.16 to 0.45), and neither did their level of self-reported monitoring of their child's behaviour increase ($d = 0.05$, 95% CI –0.08 to 0.18; see *Table 19*).

It has been consistently found that parenting interventions increase positive parenting behaviour (e.g. Dishion et al.[151] and Leijten et al.[149]). This is in line with the focus of most parenting interventions, including IY, which explicitly teach parents these behaviours. However, the specific aspects of positive parenting behaviour that the intervention does change, and the aspects that the intervention does not change, have not been sufficiently studied. Most previous studies focused on the effects of parenting interventions on increased parental use of positive reinforcement strategies as a whole, such that praise and rewards were combined (e.g. see the Furlong et al.[12] review). This is not surprising, given that most of the frequently used instruments (e.g. the Parenting Practices Inventory and the APQ) include a combined subscale of positive reinforcement strategies. In addition to combining individual family-level data across trials and across measures, our study disentangled the different strategies parents use to reinforce positive behaviour in their children. It indicates that parents make more use of positive reinforcement strategies, but that these seem to pertain to praising children, not to using tangible rewards.

In addition to parental use of praise and rewards, we examined the extent to which the intervention influenced parental monitoring of the child. This aspect of positive parenting is less well studied in younger children. Monitoring tends to be especially important in adolescence: as a predictor of youth behaviour problems[161] and as a mediator of the effects of parenting interventions on youth behaviour problems.[162] The absence of an effect of the IY parenting intervention on parental monitoring of their young children's behaviour contributes to the perception that monitoring is not the main aspect of positive parenting that is changed by interventions with young children. Instead, increased use of praise by parents seems central.

Possible harms

We checked the direction of effects of all possible secondary outcome measures for signs of harmful effects of the parenting intervention to children's wider mental health, parental mental health and parenting practices. There were no signs of harmful effects. All effects pointed in the direction of benefits; none pointed in the direction of harm.

Harmful effects are rarely studied in parenting intervention evaluation trials. However, as we know some interventions can do harm (e.g. some youth interventions[97]), we consider it vital to always check for possible adverse effects of interventions. The absence of any signals pointing towards harmful effects is consistent with a traditional (i.e. trial level, not individual patient level) meta-analysis on the effects of parenting interventions.[12] At least with regard to children's wider mental health (e.g. ADHD symptoms and emotional problems), parental mental health (e.g. depression) and parenting behaviour (e.g. corporal punishment) there were no signs of harmful effects.

Chapter 5 Economic evaluation

Several UK studies have established that the costs associated with conduct disorder, both to the public sector and to wider society, are high. Costs incurred in childhood are borne by the public sector as well as by families.[163] The annual costs of mental health problems in children to the public sector have recently been estimated to be at least £1.47B (2008 prices), with the costs per child with conduct problems surpassed only by the costs associated with hyperkinetic disorders.[164] Using the same British Child and Adolescent Mental Health Surveys data set, a higher level of mental health difficulties measured on the SDQ was associated with a higher likelihood of service use and higher costs of mental health services.[165]

The impact of childhood conduct disorder reaches into adulthood. Costs for a small cohort with childhood conduct disorder were 10 times higher in adulthood than for those with no behaviour problems.[7] In a more recent study, the costs incurred in early adulthood were two to three times higher among those with high levels of childhood conduct problems than in those without mental health problems, driven mainly by contacts with the criminal justice system.[166]

The most comprehensive estimate of the costs of providing evidence-based parenting programmes to date comes from a database of information provided by intervention developers.[167] The costs of parenting programmes were estimated from details of five evidence-based and commonly used programmes, and include staff costs, overheads, materials and additional items such as catering and childcare as well as the costs of training and supervision.[167] The median cost of a group intervention is estimated at £952 (range £282–1486) per participant, whereas the median cost of an individual intervention is £2078 (range £769–5642).

Given the high costs associated with conduct disorders, it is reasonable to assume that preventing such problems could result in substantial societal savings, but few cost-effectiveness analyses have been undertaken in a European context and existing studies often lack the statistical power needed to detect cost differences. A recent review of the evidence[168] located two UK-based cost-effectiveness analyses of parenting programmes.[169,170] For participants receiving the IY intervention,[169] the cost of bringing a child below the clinical cut-off point was £1344. A RCT of IY in Ireland similarly found significant improvements in behaviour, alongside a reduction in service use in both groups but with a greater reduction in the intervention group.[171]

There is no evidence of the longer-term impact of parenting programmes on costs from RCTs in Europe because a waiting list design is common for these interventions. Two studies of IY with longer-term follow-ups found that improvements in ECBI-I score were sustained and accompanied by decreased costs of health and social services or a reduction in the likelihood that formal services were used.[172,173] The US-based evaluation of the multifaceted Perry Preschool Program suggests a possible long-term impact on conduct disorder, criminal behaviour and employment.[174] Evidence from model-based analyses suggests that parenting programmes are likely to provide substantial savings (e.g. Aos *et al.*[175] and Bonin *et al.*[20]). However, it is important to note that these analyses are relying on the limited empirical evidence currently available.

Although there is some indication that parenting programmes are likely to be cost-effective in the short term, and it is reasonable to assume that effective programmes will reduce the prevalence of severe behaviour problems and the associated costs, there is currently little empirical evidence. This study will strengthen the evidence base for short-term cost-effectiveness of parenting programmes for behaviour problems, and use this evidence to re-examine the likely longer-term savings. No cost-effectiveness analysis to date has been able to identify differential cost-effectiveness for subgroups or account for potential moderators of cost-effectiveness. Here we identify characteristics that are associated with cost variation at follow-up, which will guide further analysis to determine whether or not the IY intervention is more likely to be cost-effective for specific groups of children.

Methods

Service use data

Central to the estimation of costs and cost-effectiveness are the records of service use for each person in the trial, including use made of the IY programme by the intervention groups. Of the 14 studies, five were undertaken in countries other than the UK or Ireland, for which the service array and public sector financing systems are very different. Two UK trials did not collect any service use information, and one collected it in a way that was not suitable for use in economic evaluation. Another UK trial was excluded because no baseline service use information was collected. Five trials therefore met the criteria for inclusion.

All service use data were merged and cleaned by the economics team, ensuring that categories of service use and measures of intensity are comparable across all studies. A detailed description of our data harmonisation approach can be found in *Appendix 7*.

Service use

Information on self-reported service use was obtained from resource use questionnaires completed by trial participants at baseline and follow-up. The baseline period covered 6 months for three trials and 12 months for two trials. Follow-up periods covered 6 months for three trials, 3 months for one trial and 12 months for one trial.

Intervention cost

For the cost-effectiveness analyses, unit costs for six sites that implemented IY interventions were estimated using standard methods informed by economic theory. Data were requested from all collaborating centres using a standardised Service Information Schedule. To ensure comparability with other unit costs used in these analyses, we have excluded set-up costs, such as initial training or amendment of the standard IY programme to meet the needs of the study. These are identified in our paper exploring the costs of the IY intervention and their relationship with fidelity characteristics (J Beecham, 2016, personal communication). Staff time forms the major part of the costs; we have used nationally applicable (all England, i.e. London and the rest of England) salaries and on-costs taken from the *Unit Costs of Health and Social Care*[176] or estimated using a commensurate method. Staff costs are estimated according to their professional background; average salaries from group leaders have been estimated to reflect the staff mix employed in each site. Although we have included the costs of supervisors' time over the course of the programme, we have excluded their travel and subsistence costs, as the two sites that purchased supervision directly from IY were subject to high (and variable) travel costs, relative to the other four sites. We also excluded the costs of reimbursing participants' travel costs, as insufficient information was available. Venue costs have been estimated using data from the *Unit Costs of Health and Social Care*.[176] We also include costs accruing for project management, administrative assistance, materials used in the sessions, snacks and provision of a crèche facility.

Service costs

For each service that participants reported using on a service use questionnaire (Client Service Receipt Inventory; CSRI),[177] an appropriate unit cost was obtained from publicly available sources[176,178] or calculated using an equivalent approach.[179] Unit costs used in this study are detailed in *Appendix 8*, *Table 31*. The total costs associated with service use, as well as subtotals for costs for community health services (including primary care), hospital services, specialist mental health services, social care, accommodation away from home (e.g. foster care) and services provided by the voluntary sector, were calculated for each participant by multiplying the reported number of service contacts by the corresponding unit cost. When the number of service contacts was missing but it was indicated that a participant used the service, the number of contacts was imputed using the mean of service contacts for the specific trial.

All costs are presented in 2014 prices. Given that costs cover a 6-month period, no discount rate is applied.

This re-estimation of unit costs and per participant costs improves the comparability of costs across the trials by ensuring that any cost differences are not a result of the individual approaches to estimating unit costs or total costs, without deviating from the original data collection.

Statistical analyses

Analyses were performed using CCs. Participants were included in the cost analysis if they had completed a CSRI[177] at baseline and follow-up, and in the cost-effectiveness analysis if they also had completed the ECBI-I assessments.

Unless stated otherwise, all statistical models account for (1) the skewed distribution typical of cost data and (2) the hierarchical structure of the data and potential clustering by drawing bootstrap samples[180] (10,000 replications) by cluster (family and trial) and strata (treatment group, per cent of sites in rural locations, whether or not the trial was conducted in England and whether or not the CSRI data collection period differed from 6 months).

Analysis of service use and costs

We present the number and percentage of trial participants reporting use of services by randomisation group at baseline and follow-up for the pooled sample.

We show the average costs associated with reported service use by service category for the two treatment groups at both time points for the pooled sample, alongside their SDs and range. The *p*-values for the differences in costs between treatment group at baseline and follow-up were derived using a clustered regression model (described in *Statistical analyses*).

Cost variations

An exploratory analysis was undertaken to determine whether or not costs varied based on participant baseline characteristics. Characteristics considered were the moderators also used in the main analysis (note that this did not include parenting variables) (presented in *Chapter 3*) as well as baseline service costs and treatment condition.

For each potential predictor, a multivariate model was fitted, which controlled for baseline costs. From the results, a final model was fitted, which retained only significant predictors of total costs. This model was used to estimate cost differences for the cost-effectiveness analyses.

Cost-effectiveness analyses

We tested whether or not IY is likely to be considered cost-effective compared with the control condition. Cost-effectiveness analysis was conducted on the pooled data. The cost-effectiveness analysis takes a public sector perspective, focusing on the children's use of services and support. The primary analysis uses change in ECBI-I as the outcome measure. A secondary analysis uses a binary variable indicating whether or not a child is moved below the ECBI-I clinical cut-off point (score $\geq$ 131 points) following the intervention.

Using seemingly unrelated regression, separate regression models were fitted for (1) costs at follow-up, based on our analysis of cost variations and (2) each outcome measure. The outcome measures of interest were (1) change in ECBI-I (linear regression) and (2) whether or not child is below the clinical cut-off point on the ECBI-I (logistic regression).

For each combination of costs and outcome, 10,000 bootstrap replications of the treatment effect were generated. The probability that the intervention is cost-effective was assessed by calculating the proportion of incremental cost-effectiveness ratios that indicated the intervention would be judged cost-effective, given a range of values that a funder or wider society may place on an improvement in outcome. This value is commonly known as 'willingness to pay' (WTP). The probability that the intervention would be considered cost-effective was then plotted against its corresponding value of WTP, resulting in a cost-effectiveness acceptability curve (CEAC).[181]

Economic modelling

To estimate potential longer-term savings from the IY intervention, we updated our economic model of the long-term costs associated with persistent conduct disorder.[20] The original model included parameters for intervention take-up and dropout. No data for take-up are available in our data set, and, as results for the present study were obtained on an intention-to-treat basis, dropout is already accounted for in the overall effectiveness figure.

As in our original model, we estimate a trajectory of remission in the absence of intervention to use as a simulated control group. Based on the literature, we assume that about 60% of children with behavioural problems at age 3 years still show problems at age 8 years, and that 50% of children with behaviour problems develop adult antisocial personality disorder.[182] Without intervention, we therefore estimate that the chance that behaviour problems at age 5 years persisting beyond age 16 years is approximately 60%.

In this analysis, the estimated intervention effect is calculated in relation to the observed control group trajectory. We used the information from the pooled sample to construct a sensitivity analysis, varying the presumed control group trajectory.

Longer-term cost estimates were updated to reflect more recent research. In addition to uprating the existing cost parameters to 2014 prices, costs were taken from two recent studies of UK cohorts.[164,166] This allowed us to create a 'high-cost' and a 'lower-cost' scenario for each trajectory.

The following cost categories are considered in the model: NHS, social services departments, Department for Education, voluntary sector, criminal justice system, health impacts of crime and benefits payments. Note that all costs are calculated as additional costs due to conduct disorder, over and above what would be spent on a person who did not experience behaviour problems at age 5 years. Costs are calculated from age 5 years to age 30 years, and discounted to present value using a discount rate of 3.5%.

The intervention effect is based on the results of our cost-effectiveness analysis, in which we calculate the odds ratio of a child falling into the non-clinical range of the ECBI-I (post treatment) for the intervention compared with the control group. In this analysis, we assume that the intervention is delivered at age 5 years, which is approximately the average age of our study sample.

Results

Sample for economic analysis

Data from five trials were merged to conduct the economic analyses. They were trial 4, IRE; trial 7, WL-SS; trial 9, BIRM; trial 10, LON-SPO; and trial 12, LON-HCA. *Table 20* shows baseline demographic data for the 608 participants included in the analysis (control group, *n* = 236; intervention group, *n* = 372).

There are some notable differences between the full analysis sample and the sample for economic analysis. Children in the latter tend to be younger, on average, and there are differences in the sociodemographic composition of the samples. Indicators of SES tend to be less favourable in the economic sample, whereas the proportion of participants identifying as members of an ethnic minority is smaller.

Trial-level characteristics for the trials included in the economic sample and in the full analysis sample are shown in *Table 21*. Although the small number of trials makes a comparison difficult, there do not appear to be striking differences between the full set of trials and the sample for economic analysis, with the exception of geographic location.

TABLE 20 Baseline demographic data, economic sample

Variable	Control		IY	
	n	Mean (SD) (%)	*n*	Mean (SD) (%)
Child gender (male)	236	63	372	60
Child age (months)	235	59.23 (16.63)	370	56.17 (16.63)
SES low income	230	72	359	62
Low education	215	40	361	44
SES lone parent	222	33	355	38
SES teenage parent	222	11	365	16
SES unemployed	192	36	297	43
Ethnic minority	228	18	368	20
Baseline ECBI-I score	234	144.99 (32.59)	368	149.76 (31.29)
Baseline ADHD (SDQ)	221	6.18 (2.47)	322	5.88 (2.47)
Baseline emotional problems (SDQ)	145	3.88 (2.53)	252	3.84 (2.53)
Baseline parental depression	182	11.00 (10.37)	268	13.48 (10.37)

TABLE 21 Trial-level characteristics

Trial characteristic	Full analysis sample Mean (SD) or number true (%)	Sample for economic evaluation Mean (SD) or number true (%)
Geographical location (England vs. other)?	6 trials	3 trials
Service provider for trial (non-clinical or clinical)?	11 trials (78.6)	1 clinical, 4 non-clinical
Efficacy or effectiveness trial?	13 trials effectiveness (92.9)	5 (100)
Did staff complete IY checklist after sessions?	14 trials (100)	5 (100)
Was IY mentor part of the trial?	10 trials (71.4)	4 (80)
Were sessions video-taped?	14 trials (100)	5 (100)
Was there weekly/fortnightly supervision?	10 trials (71.4)	5 (100)
Did independent ratings of sessions take place?	3 trials (21.4)	2 (40)
Did any of team attend international IY workshop/training?	10 trials (71.4)	3 (67)

Service use

Service use at baseline and follow-up is shown by treatment group in *Table 22* (a fully expanded service use table can be found in *Appendix 9*, *Table 32*). At baseline, > 70% in each treatment group had seen their general practitioner (GP). In the control group, 23% reported contacts with a health visitor, whereas in the intervention group, this percentage was higher, at 32%; this was the only service for which there was a difference between groups that exceeded 5%. Outpatient hospital services were used by 21% in each treatment group, whereas 17% in the control and 19% in the intervention group utilised the accident and emergency department (A&E). Five per cent of participants reported inpatient stays at baseline.

TABLE 22 Service use

Service	Baseline				Follow-up			
	Control (*n* = 236)		IY (*n* = 372)		Control (*n* = 236)		IY (*n* = 372)	
	Number using service	Per cent using service	Number using service	Per cent using service	Number using service	Per cent using service	Number using service	Per cent using service
Hospital								
A&E/casualty	40	17	69	19	29	12	58	16
Ambulance	3	1	7	2	4	2	2	1
Outpatient service	49	21	77	21	38	16	51	14
Inpatient stay	11	5	20	5	5	2	17	5
Other hospital	8	3	17	5	3	1	5	1
Community health care								
GP	170	72	265	71	116	49	196	53
GP nurse	29	12	51	14	18	8	41	11
Health visitor	55	23	120	32	23	10	37	10
Speech and language therapist	27	11	48	13	19	8	29	8
Other community health	29	12	46	12	18	8	31	8
Mental health								
CAMHS	3	1	8	2	4	2	2	1
Other mental health	16	7	20	5	8	3	17	5
Social care								
Social worker	11	5	30	8	12	5	20	5
Other social care	1	0	6	2	1	0	0	0
Accommodation								
Child placements[a]	2	1	3	1	0	0	1	0
Voluntary sector/self-help								
Voluntary sector support	10	4	18	5	8	3	10	3
Self-help	1	0	3	1	1	0	1	0

A&E, accident and emergency department; GP, general practitioner.
a Includes respite foster care, foster care and children's homes.

At follow-up, there were no differences between treatment groups in the proportion of participants reporting service contacts of 5% or more for any service, and there were no notable increases in the proportion reporting contacts from baseline. Contact with GPs reduced to around 50%, whereas the proportion in contact with health visitors decreased to 10%. Use of A&E and outpatient services also declined in both groups.

Few participants reported use of services beyond primary care and hospital services. Notably, use of mental health services was low. Placements for children away from their family are worth noting as they tend to be costly but these were not frequent occurrences in our sample.

Costs and cost variations

Table 23 shows costs at baseline and follow up by intervention group.

Intervention cost

The cost per IY session for the five trials included in the economic analysis ranged from £228 to £352. On average, intervention participants in the economic analysis sample were offered 12.7 sessions (range 11–19 sessions, at individual level), and attended 8.7 sessions. The cost of the intervention 'as offered' ranged from £1733 to £2586 at the trial level, and the cost of the intervention 'as provided' ranged from £1792 to £1496. At the individual level, the range was £0 to £4675, with an average of £2414 (SD £1248).

Other service costs

At baseline, costs associated with hospital contacts and community-based health care were the largest contributor to total costs for both treatment groups. This reflects the patterns of service use described in *Table 23*, and is also seen at follow-up. There are no significant cost differences between groups for any service category or total costs at baseline. At follow-up, total costs for the control group are nearly 40% lower than at baseline, driven largely by decreased hospital and other health-care costs. Although average total costs for the intervention group are significantly larger than for the control group (£338 compared with £2766; $p < 0.001$), the difference is accounted for by the cost of the IY intervention (average £2414). Although there is a significant difference in the cost associated with use of mental health services ($p = 0.046$), the amount is small (difference of £18).

Cost variations

Variations in costs at follow-up were explored using linear regression models, drawing bootstrap replications from clusters and strata as described above. *Table 24* shows models investigating the effect of a single variable on costs, adjusting for baseline costs (model 1). The only predictor of costs that is significant at the usual 95% level is the treatment condition, reflecting the higher costs incurred by the IY group because of the intervention. It has been argued that, when it comes to cost differences, avoiding a type II error may not be as critical as it is when it comes to health outcomes. Predictors significant at the 80% level were therefore considered for a multivariate model (note that this did not include parental unemployment status, as this was available for only four of the five trials) constructed by stepwise removal of non-significant predictors. The final model (model 2) included only child age and gender alongside the treatment condition. On average, costs were £340 higher for boys than for girls. Average costs decreased as children got older by £8 per month.

Cost-effectiveness analyses

Intervention effect

The effect of the intervention on outcome measures is an integral part of any cost-effectiveness analysis. On average, unadjusted ECBI-I scores in the control group ($n = 214$) improved by 8.1 points (SD 28.6 points). In the intervention group ($n = 357$), the improvement was larger with 30.7 points (SD 34.1 points). At baseline, 33% of the control group and 26% of the intervention group were in the subclinical range of the ECBI-I, increasing to 45% and 69% at follow-up, respectively. The (bootstrapped) odds ratio for a child in the intervention group being in the non-clinical range compared with the control group was 1.54 (SD 0.22) in our regression model, accounting for clustering and controlling for baseline ECBI-I score. The chance that a child in the intervention group is in the non-clinical range post treatment is therefore estimated to be 1.54 times as high as for a child in the control group.

Cost-effectiveness acceptability curves

Figure 15 shows the CEAC for improvement in disruptive behaviour (ECBI-I) in the intervention group compared with the control group, for WTP from £0 to £250. The probability that the IY intervention will be considered cost-effective exceeds 50% at a WTP of £109 per 1-point improvement on the ECBI-I. This is equivalent to a coin toss. An 80% chance of cost-effectiveness is exceeded at a WTP of £121, whereas 95% is reached at £134 and 99% at £145.

TABLE 23 Service costs by service category and intervention group, baseline and follow-up

| Service category | Baseline | | | | | Follow-up | | | | |
| | Control | | IY | | | Control | | IY | | |
	Mean (SD) (£)	Range (£)	Mean (SD) (£)	Range (£)	p-value[a]	Mean (SD) (£)	Range, (£)	Mean (SD) (£)	Range (£)	p-value[a]
Hospital	292 (1938)	0–28,650	406 (2065)	0–23,184	0.497	193 (1182)	0–17,676	215 (857)	0–8799	0.792
Community health	131 (191)	0–1123	159 (283)	0–3365	0.184	84 (155)	0–1400	115 (238)	0–2034	0.072
Mental health	25 (145)	0–1550	62 (455)	0–6200	0.224	29 (161)	0–1416	11 (67)	0–620	0.046
Social services	10 (96)	0–1430	31 (169)	0–1980	0.086	22 (139)	0–1430	23 (184)	0–2860	0.989
Accommodation	37 (47)	0–8400	470 (6618)	0–112,700	0.316	0 (0)	N/A	2 (36)	0–700	0.426
Voluntary sector	5 (37)	0–470	6 (54)	0–879	0.967	9 (85)	0–1200	5 (55)	0–967	0.499
IY intervention	N/A		N/A			N/A		2414 (1248)	0–4675	N/A
Total costs	501 (2064)	0–29,220	1135 (6971)	0–112,816	0.175	338 (1261)	0–18,468	2766 (1594)	0–13,381	< 0.001

N/A, not applicable.
a   Difference between intervention and control group.

TABLE 24 Variation in costs at follow-up (in £)

Predictor	Model 1			Model 2		
	Coefficient (standard error)	95% CI	*p*-value	Coefficient (standard error)	95% CI	*p*-value
Total costs at baseline	0.02 (0.01)	0.00 to 0.11	0.161	0.01 (0.02)	–0.04 to 0.06	0.714
Child gender	225.05 (157.64)	–92.95 to 540.89	0.154	341.03 (103.26)	2373.41 to 2778.19	< 0.001
Child age	–9.58 (6.77)	–22.21 to 2.43	0.158	–7.99 (4.59)	–16.99 to 1.01	0.082
Low income	–231.53 (170.60)	–540.44 to 76.51	0.175			
Education level 1 or 2 (low education)	–140.45 (159.23)	–435.11 to 167.08	0.378			
Lone parent	19.64 (164.29)	–286.14 to 329.54	0.905			
Teenage parent	302.40 (226.92)	–137.77 to 832.86	0.183			
Unemployed[a]	249.08 (188.88)	–140.77 to 698.32	0.188			
Ethnic minority	116.15 (222.48)	–277.39 to 496.11	0.602			
Baseline ECBI-I score	2.56 (2.53)	–2.13 to 6.95	0.312			
Baseline ADHD (SDQ)	–17.08 (35.96)	–87.60 to 50.74	0.635			
Baseline emotional problems (SDQ)	34.67 (30.63)	–25.67 to 94.40	0.258			
Baseline parental depression	0.51 (9.73)	–15.44 to 17.03	0.958			
Treatment condition	2573.10 (119.14)	2351.20 to 2760.70	0.000	2575.80 (103.26)	2373.41 to 2778.19	< 0.001

a Data available for only four out of five trials.
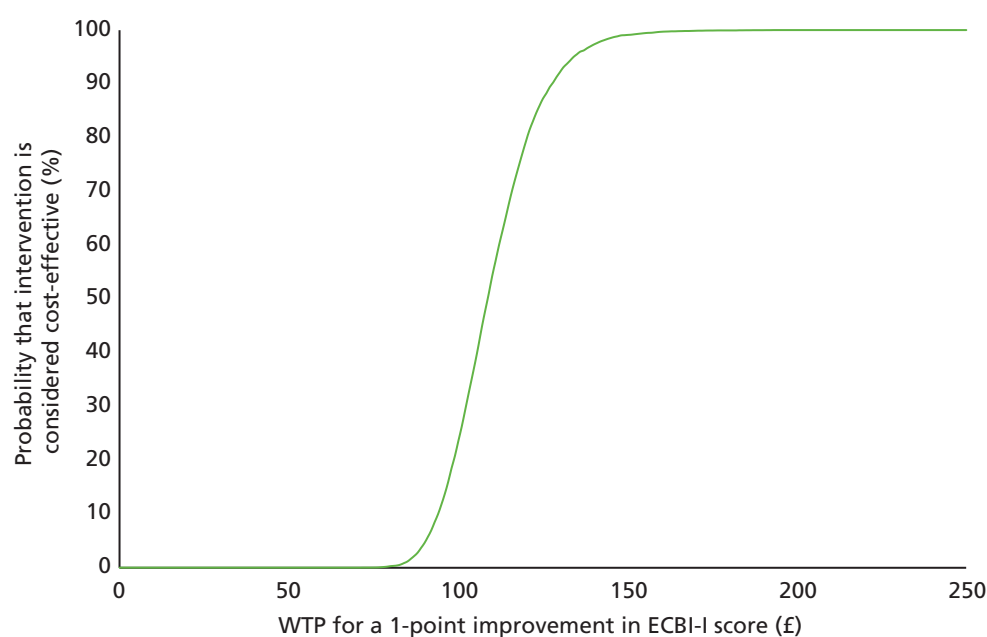
FIGURE 15 Cost-effectiveness acceptability curve for the IY intervention: ECBI-I score.

Subgroup analysis

The main analysis was repeated for subgroups by gender (male vs. female), baseline ECBI-I score (< 131 vs. ≥ 131), child age (< 5 years vs. ≥ 5 years) and parental depression at baseline (parental BDI score of ≥ 20 points vs. < 20 points). The corresponding CEACs are shown in *Figure 16*, together with the curve for all children. Note that these curves were estimated in separate models.

Based on our analysis, the IY intervention is less likely to be considered cost-effective for children with an ECBI-I score below the clinical cut-off point at baseline and who are < 5 years old. It is more likely to be considered cost-effective for boys than for girls, and for children whose parent's BDI-II score at baseline indicates at least a moderate level of depression.

When the outcome is increasing by 1% the probability of scoring below the clinical threshold on the ECBI-I scale as a result of the intervention (*Figure 17*), there is a 50% chance that the IY intervention will be
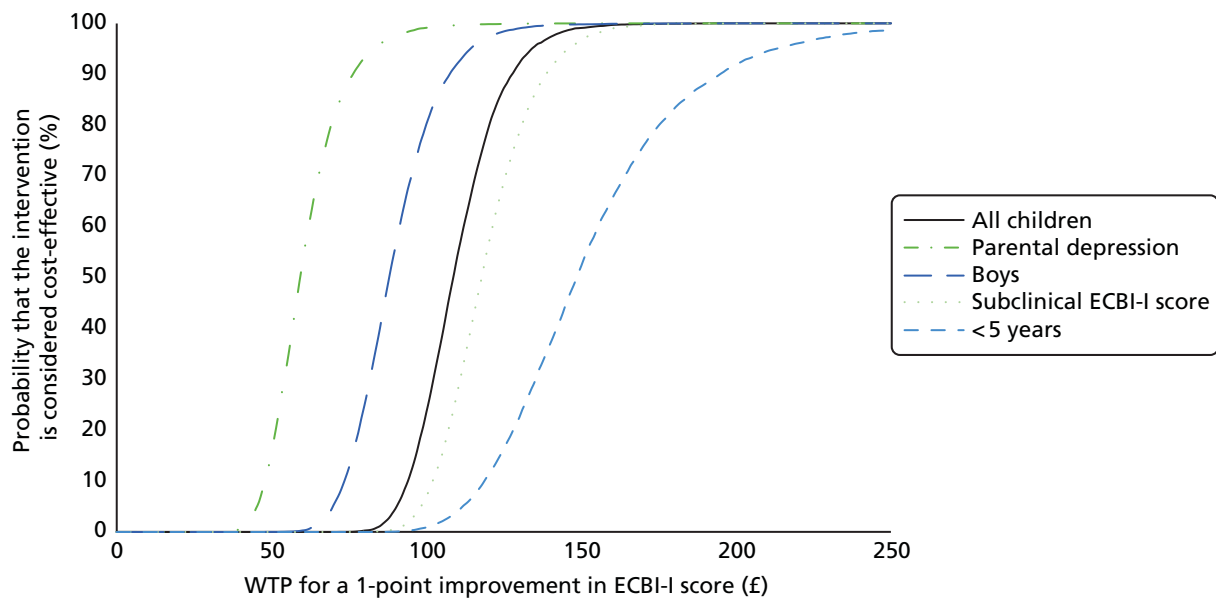
FIGURE 16 Cost-effectiveness acceptability curves (subgroup analyses): ECBI-I score.
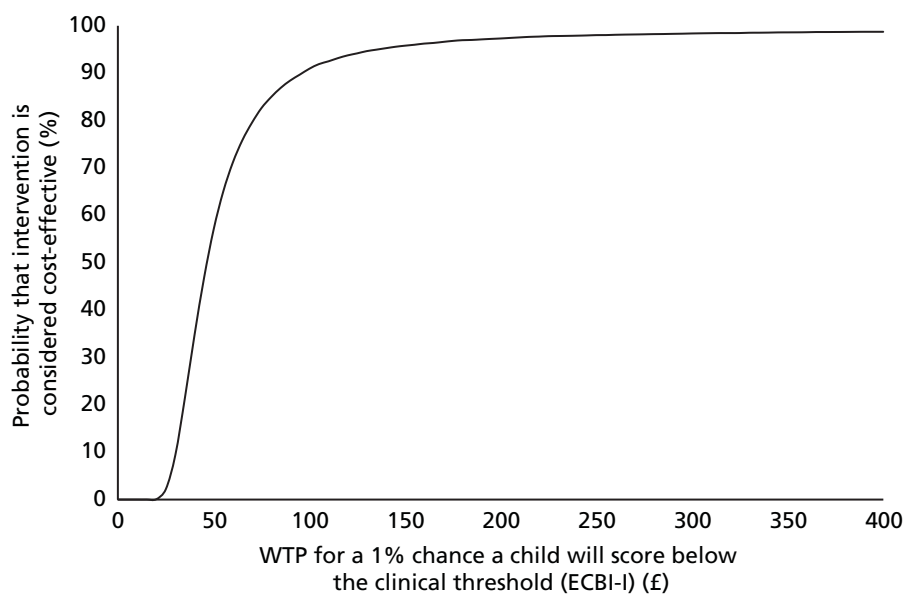
FIGURE 17 Cost-effectiveness acceptability curve for the IY intervention: ECBI-I score below clinical threshold post intervention.

considered cost-effective at a WTP between £45 and £50. An 80% chance is reached at a WTP of £70, whereas 95% is reached at £95 and 99% reached at a WTP of £370.

Longer-term savings from the Incredible Years intervention: economic modelling

Figures 18 and *19* show the probability that a child still experiences behaviour problems over time, with year 0 being the year of the intervention and year 1 the time when the intervention effect manifests. The 'no intervention' trajectory in *Figure 18* follows the data from the literature, with a 60% chance that a child with behaviour problems at age 5 years will still experience problems after the age of 16 years (11 years post intervention). In the intervention scenario, this probability is 54%.

Figure 19 shows scenario 2 and the effect of the intervention compared with a trajectory in the absence of intervention that is modelled on the change observed in the pooled IY sample for the control group, with a 12% decrease in the probability of scoring above the ECBI-I cut-off point in year 1. On this trajectory, the probability of behaviour problems persisting past age 16 years is again 60%, but with a steeper decrease in year 1. As the intervention effect is estimated as a multiple of the reduction in probability in
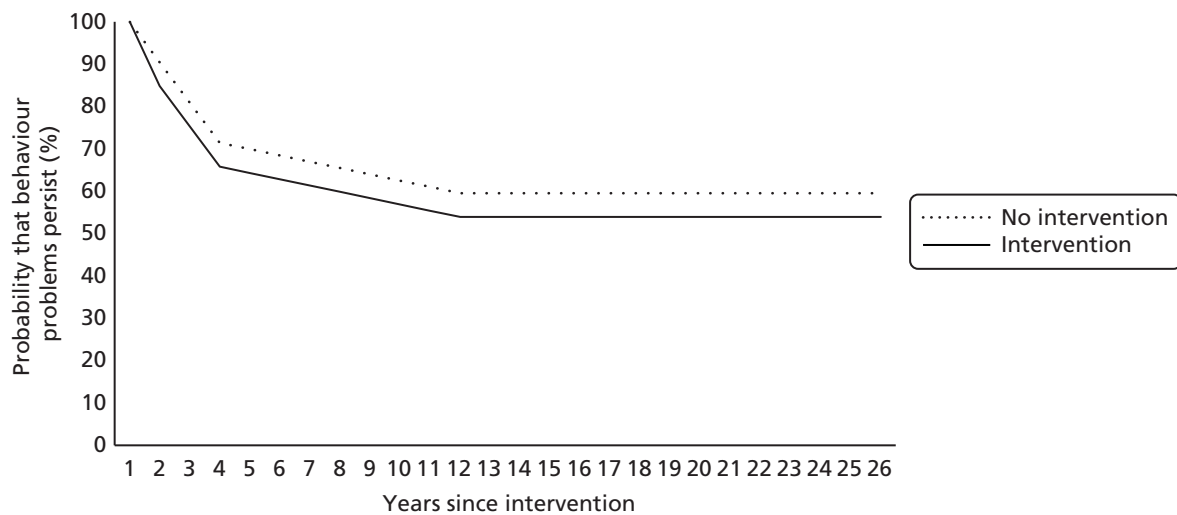
FIGURE 18 Trajectory of probability of behaviour problems persisting from age 5 years: scenario 1.
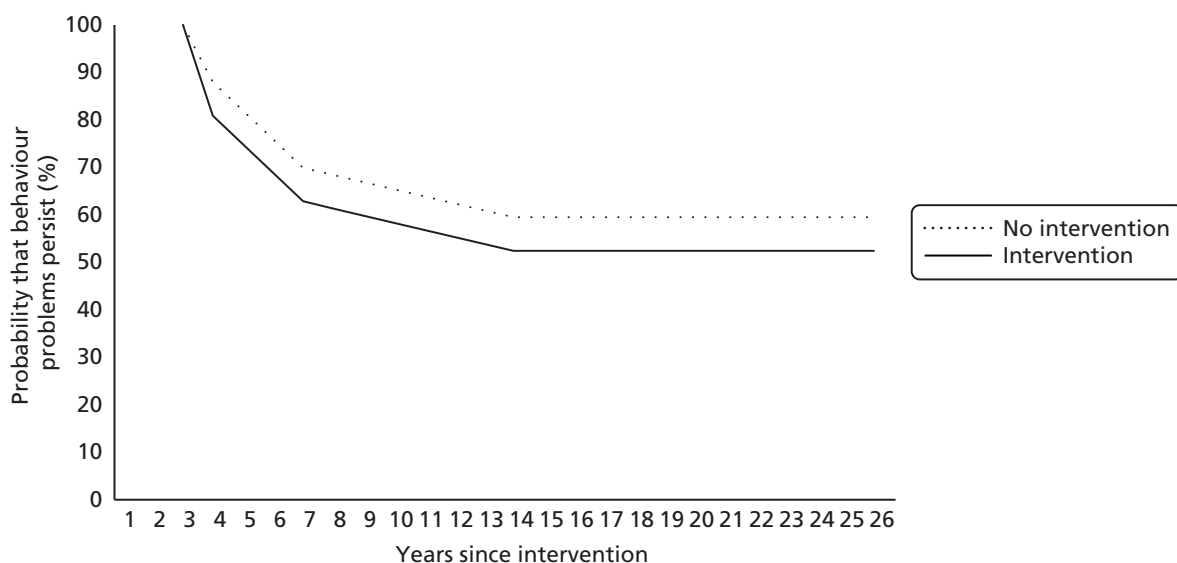
FIGURE 19 Trajectory of probability of behaviour problems persisting from age 5 years: scenario 2.

the control group, in this scenario, the probability that conduct problems persist beyond age 16 years in the intervention group is reduced to 52%.

Table 25 shows the present value (i.e. discounted to 'today's money' using a 3.5% discount rate) of savings from the intervention based on the assumptions above.

In scenario 1, savings range from £1023 to £7565, whereas in scenario 2, they range from £1254 to £9408. In the low-cost scenarios, no savings are estimated to be achieved once the intervention cost is subtracted. Savings are largest in the areas of highest potential costs, the criminal justice system and the Department for Education.

Discussion

Our analysis of the five IY trials from the UK and Ireland with economic data identified service use for the intervention and control groups, calculated associated costs (including the cost of the IY intervention) and showed the probability that the intervention will be considered cost-effective. In addition, we modelled potential longer-term savings from the intervention.

The pooled sample for economic analysis ($n = 608$) in this study is larger than for any previous individual trial of the IY intervention. There are obvious advantages to having a larger sample, most notably increased power to detect differences in costs. Often, individual trials are not powered to the larger standard variations found in cost data. At the same time, harmonising data from different trials meant that choices regarding the level of detail to retain in individual data sets had to be made to ensure comparability (see *Appendix 7*). The differences between study designs, coverage of services and level of detail on the CSRI, and the different follow-up periods mean that there may be inherent differences between the trials that could affect results. As much as possible, this has been accounted for by utilising statistical methods that allow for clustering.

Use of services and service costs
The findings regarding service use are broadly in line with other findings from studies of children with mental health problems. Most support comes from primary care and hospital services, with little involvement from mental health services. Although our data do not allow us to determine whether this is because of a lack of help seeking or a gap in provision, the fact that only one in four children and young people with mental health problems is in contact with mental health services has long been recognised as a significant problem, and increasing access is a priority.[183]

TABLE 25 Present value of savings from providing the IY intervention at age 5 years, per child and by agency

Agency/budget	Savings 'scenario 1'		Savings 'scenario 2'	
	Low (£)	High (£)	Low (£)	High (£)
NHS	106	2445	129	3039
Social services departments	72	182	88	221
Department for Education	613	949	746	1148
Voluntary sector	0	26	0	32
Criminal justice system	232	3583	290	4493
Benefits payments	0	381	0	475
Totals	1023	7565	1254	9408
Net of intervention cost	−1391	5151	−1160	6994

Although we observed significant differences in total costs between treatment groups, this was driven by the cost of the IY intervention. The cost of the intervention was estimated for each trial, using the same principles and cost categories. Notably, the average cost of the intervention for this sample was £2414 (2014 prices), and therefore more than twice the previous 'best estimate' of £952 (2009 prices);[167] it should be noted, however, that this is a median, whereas we report the mean. However, our average cost is similar to the estimate by Edwards et al.[169] of £1934 for IY delivered through Sure Start Centres once we account for inflation.

Mirroring the findings from the analysis of differential treatment effects presented in *Chapter 3*, we do not find any cost variations associated with social disadvantage, ethnicity, ADHD or child emotional problems at baseline. However, unlike the main analysis, we do not see cost variations associated with baseline levels of disruptive behaviour or parental depression. However, we did find significant variations in costs at follow-up for child gender and age, suggesting that there is both a differential effect and a differential impact on costs from the IY intervention for these characteristics. Girls in our study incurred lower costs on average but they also did not benefit from the intervention as much as boys. The evidence on the relationship between gender and costs is mixed. Although Romeo et al.[163] found costs to be lower for girls than boys aged 3–8 years referred to specialist mental health services, Knapp et al.[165] found higher mental health service costs for girls. We saw (modestly) decreasing costs with age, in line with previous findings,[165,184] even though mental health problems generally increase with age.[185] We did not find an association between costs and ethnicity, and again the previous evidence is mixed, with one study finding higher mental health-related special education costs,[165] although others have found no link.[184]

Cost-effectiveness and longer-term savings

The CEACs presented in this study may appear unfamiliar to some readers. Although cost-effectiveness analyses used to be presented as incremental cost-effectiveness ratio, summarising information on cost and effect differences in a single figure is problematic: a negative ratio can indicate higher costs coupled with worse outcomes, as well as lower costs and better outcomes. Similarly, a positive ratio can represent both higher costs and better outcomes in the experimental group, and lower costs and less favourable outcomes than the control group. Therefore, the incremental cost-effectiveness ratio on its own does not provide the information needed to judge the cost-effectiveness of an intervention. For this reason, CEACs are the current standard method of presenting cost-effectiveness information. For a discussion of the statistical issues associated with the incremental cost-effectiveness ratio and the derivation of the CEAC, see van Hout et al.[181]

Our cost-effectiveness analyses show that at a WTP of £109 per point improvement on the ECBI-I is 50%, increasing to 99% at a WTP of £145. Our subgroup analysis showed that the intervention is less likely to be considered cost-effective for children who scored below the clinical threshold on the ECBI-I at baseline, and for children under the age of 5 years. At the same time, it would be more likely to be considered cost-effective for boys than for girls and for children whose parents had a BDI-II score at baseline indicating at least a mild level of depression. It should be noted that this analysis was limited to cases with full data, and results should therefore be interpreted with caution.

The results of cost-effectiveness analyses are often difficult to compare because of different methods used and different coverage of cost categories. In the study by Edwards et al.[169] a 50% probability of cost-effectiveness was reached at a WTP of approximately £75 (2004 costs; note that the data in the Edwards et al. study are not presented in such a way that this comparison can be made easily), whereas a 99% chance is reached around £150. Overall, our findings appear reasonably similar.

Willingness to pay is hard to interpret when there is no established societal threshold for the outcome in question (in this case, the ECBI-I). The only outcome with a well-established WTP threshold is quality of life as measured by quality-adjusted life-years. No quality-of-life measure was included in any of the trials that form part of our analysis here and, arguably, would not be appropriate for these young children with mental health problems.[186] Our findings suggest that a 1% increase in the chance that a child scores below the clinical threshold following intervention can be achieved with reasonable certainty (80%) at a

WTP of £70. This figure needs to be evaluated in the context of the likely long-term societal costs associated with persistent conduct disorder.

It should also be noted that the treatment effects for three of the trials included in the economic analysis were below the mean for the pooled sample, and therefore our findings may underestimate the potential for cost-effectiveness for the IY sample as a whole (assuming costs in the other trials would have been similar to our economic sample).

Longer-term savings

There is a lack of empirical evidence on the long-term savings from parenting programmes, especially in a European context. In this study, we adapted our previous model estimating the longer-term savings from parenting programmes intended to prevent persistence of behaviour problems.[20] We drew on additional literature and created a 'low-cost' and a 'high-cost' scenario. In addition, we varied the assumptions around the natural course of conduct disorder. These differences in the underlying model mean that our results are not directly comparable. Instead, they should be viewed as the next iteration of the model.

The previous study assumed an intervention effect of 34%, that is, the proportion of children who, following the intervention, were below the clinical cut-off point, was 34% higher in the intervention group than in the control group. Note that this study included both group-based parenting programmes and those delivered to individual families. This is a different approach to estimating effectiveness, and our present estimates result in a much more conservative assumption about the intervention effect. Combined with a higher intervention cost (£2414 in 2013–14 prices vs. £1177 in 2008–9 prices).

Our results indicate that if the costs associated with persistent conduct disorder are in fact lower than previously thought (see the 'low-cost' scenarios), at the current price and effectiveness, the IY intervention would not result in longer-term savings. However, in the 'high-cost' scenario, the return on investment is substantial – threefold in scenario 1 and nearly fourfold in scenario 2. They also indicate that any model-derived findings are very sensitive to even slight changes to the assumptions about the control group trajectory, here a change in the assumed chance of behaviour problems persisting, from 8% to 12%.

Conclusions/implications

Children in our sample had little contact with specialist mental health services, highlighting a potential gap in provision. As few services were used, costs were low and, thus, the potential for immediate savings from the IY intervention is reduced. This raises the question of whether or not children with mental health problems are adequately supported by mainstream services, and how access and engagement can be improved.

Our model of the longer-term savings highlights the need for better information about what happens to children with behaviour problems who do not receive effective interventions, as model results are sensitive to these assumptions. Anecdotally, children with behaviour problems grow up to be troubled adults. However, previous research[187] suggests that, for example, those with childhood behaviour problems who are employed in adulthood actually report higher wages than their peers. Research into the trajectory of children with conduct problems will provide better information to support economic modelling.

A resulting more robust estimate of the longer-term costs of childhood behaviour problems will not just inform policy but also help discussions around how much society should be willing to pay for improvements in early behaviour problems. This will help in decision-making about cost-effectiveness.

Our analyses suggest that there are variations in costs by gender and age, and, consistent with the main moderator analysis (see *Chapter 3*), by level of parent depression and child disruptive behaviour, suggesting that the intervention is more likely to be considered cost-effective for children with clinical levels of disruptive behaviour, for those with parents who are depressed, for boys and for children aged > 5 years.

Chapter 6　General discussion

Are there differential effects of the Incredible Years parenting intervention on different subgroups of families?

The question underpinning this study concerns equity effects of parenting interventions. It may be preferable to base public health policy decisions not only on whether or not an intervention is effective, on average, across the population for whom it is intended, but also on whether it is likely to have the effect of widening or narrowing the substantial social inequalities that exist in child and family outcomes. Hence the aim of this study was to test whether or not there are differential effects of parenting interventions aimed at reducing child disruptive behaviour, in different groups of families, defined by socioeconomic or psychosocial disadvantage. IY is a strongly evidence-based and well-disseminated programme, and, by pooling data from a near-complete set of trials of this programme across Europe, we aimed to overcome the marked limitations of other studies. The first limitation is that of analysing moderators in single trials, which tend to be underpowered, and, as they are not tested in most trials, are likely to suffer from substantial reporting and publication bias. The second is the limitation of conventional meta-analysis, which can test moderator or subgroup effects only at aggregate trial level, thereby losing power and discarding all information about individual-level variability in characteristics. As well as overcoming these limitations through IPD meta-analysis, our data set is unique in its size; in its potential generalisability; in that it was drawn from multiple countries, investigators and service settings, with a sizable number of families from ethnic minorities; and in the fact that all trials were conducted independently of the programme developer.

The findings concerning equity effects can be seen as falling into three main parts. First, we found that, when it comes to social and socioeconomic disadvantage, there were no differential effects. Families living in poverty, in which the parents were unemployed, or headed by a lone or teenage parent or with a parent with a low educational level, were just as likely to have children whose behaviour benefited from this intervention as families without these disadvantages. Thus, there is no evidence that social inequalities in child problem behaviour would be increased by this intervention. We found the same for families from ethnic minorities: their children were just as likely to benefit as those from the ethnic majority groups.

Second, we saw a rather different picture with respect to psychosocial disadvantages, as we did find differential effects of some of the mental health factors tested, including parental level of depression and children's level of disruptive behaviour. Interestingly, the differential effects were all in favour of the more distressed families, such that families with higher levels of problems showed greater improvement in child disruptive behaviour. Both these findings were expected by the parents in our focus groups and, consistent with our findings, they felt that socioeconomic factors would be of less importance in predicting outcome. It is noteworthy that these psychosocial factors are ones which, without intervention, function as risk factors for poor child behavioural outcomes, yet they predict differentially greater benefits from this parenting intervention. Thus, there is no evidence of any adverse equity effects; on the contrary, these data show that the intervention may be likely to reduce inequalities in child disruptive behaviour between groups with and without these risk factors. We also tested other aspects of children's mental health as moderators of outcome, that is, their levels of ADHD and emotional symptoms, but did not find any differential effects. Thus, even children with troubling comorbid problems are likely to benefit from the intervention. We also examined another type of psychosocial risk, namely the parent's level of parenting skill. Again, there was no moderator effect, suggesting that the parents and children at most high risk because of poor parenting skill are just as likely to benefit.

Finally, we examined key child demographic characteristics of gender and age. Gender is significant in that it is perhaps the strongest risk factor for disruptive behaviour; hence, the greatest inequalities in this problem are found between boys and girls. Thus, the finding that boys benefit more than girls means that

these interventions could serve to reduce one major source of inequality in disruptive behaviour. On the other hand, age effects tend to be viewed differently. Age is not seen as a particular risk factor for disruptive behaviour; rather, from a developmental perspective, some problem behaviours are more common in younger children and some emerge in adolescence. Instead, age is seen as a predictor of intervention effects, with younger children thought to be more malleable and older children harder to change. This is underpinned by basic neuroscience, and reflected in much policy globally. However, for parenting, and for many other interventions, there is remarkably little clear evidence from randomised trials comparing children of different ages. Our data show a clear lack of any age effects for this intervention, echoing a small body of older, but inconclusive, data from meta-analyses and small trials, and suggesting that, for these interventions, and this age range, 2–10 years, there is no advantage to intervening in younger compared with older children. A high percentage of the sample (> 90%) was aged between 3 and 8 years, so we can have greater confidence in this finding for the 3–8 years age range.

These findings are important in helping to clarify a literature that is beset with conflicting conclusions for many moderator effects and a lack of data for others. To some extent these mixed results are likely to be a function of weak methods, but they may also reflect real differences between programmes or in their implementation in different contexts. Our data show a quite large number of null findings on moderator effects: we would see these as very important for policy for several reasons. First of all, this is the first time a study is likely to have had adequate power to show null findings. Second, null findings may have been underplayed in the past; they are less likely to be published, leading to bias in the evidence base on which most systematic reviews are based. Third, from an equity point of view, these null findings are important for policy, in that they present the optimistic message to stakeholders that there are no adverse equity effects of this intervention on the most disadvantaged groups in society, and that groups that may be underserved, or about which there is therapeutic pessimism, are just as likely to benefit from this kind of parenting intervention. Our study cannot address the question of the extent to which different groups access parenting interventions, as it addresses its effects only after families have signed up for (but not necessarily completed) the intervention. A different study design would be needed for this question, examining factors that predict initial access to and enrolment in parenting programmes. Data bearing on these questions are included in some trial reports (e.g. Patterson *et al.*;[102] see *Chapter 3*, *Ethnicity*), as well as in studies of service delivery, outside trials.

What are the wider public health benefits of parenting interventions?

The secondary question for this study concerned the extent to which this parenting intervention has potential wider benefits for public health, beyond reducing disruptive behaviour. In terms of child outcomes, we found that the intervention improved child ADHD, as well as disruptive behaviour, but that it did not improve emotional problems. It reduced several aspects of harsh parenting, including corporal punishment, threatening and shouting, and increased parents' use of praise. However, it did not increase their use of rewards or monitoring. Surprisingly, it did not improve parents' depression, given that this has been found in other studies, including a Cochrane review by Barlow *et al.*[11] and some of the trials in our pool (e.g. Gardner *et al.*[103] and Hutchings *et al.*[85]). It is possible that publication bias may have been at play here, as our pooled data set included some previously unpublished data on parent depression. The same may apply to parenting stress and sense of competence, which also showed no change, yet have been found to improve in some small trials, including some in our pool. Even though parents who are depressed may not benefit from the intervention in terms of their depressive symptoms (there was a marginal effect in a positive direction and no evidence of harm), we know from the moderator analyses that, despite depression, these parents tend to be effective at helping their children achieve behaviour change.

Our study did not aim to investigate the equity implications of these findings. It might well be that there is a similar pattern of moderator effects for these wider benefits, especially as we would expect changes in child disruptive behaviour to be mediated partly through changes in parenting skill.[46]

Why no adverse equity effects?

Why do our moderator results show such an encouraging lack of adverse equity effects? Our data pertain only to the IY parenting programme, but are they likely to be generalisable to all similar parenting programmes? With regard to what the programme targets, that is, its content, IY is much like several other programmes derived from social learning theory and attachment theory. Most of these programmes focus first, and similarly to IY, on improving parental sensitivity to the child's needs and the parent–child relationship more generally. They then guide parents through learning techniques for increasing positive child behaviour (e.g. providing praise and rewards for positive behaviour) and reducing negative child behaviour (e.g. ignoring trivial misbehaviour and providing time-out for more severe misbehaviour). However, with regard to how the programme targets changes in parenting behaviour, IY is meaningfully different from most programmes. IY has a strong collaborative and, relatedly, a strong cultural sensitivity focus. Rather than a didactic approach, in which the therapist teaches the parents which techniques to use and how to use them, the IY therapist asks parents to set their own goals for the programme overall, and for their weekly home practice targets. This collaborative and culturally sensitive approach may be of paramount importance to reach similar effectiveness across a wide range of families with different SES and ethnic backgrounds. Cultural differences that exist between families of different socioeconomic backgrounds, even within the same ethnic group, often appear to be underestimated. Programmes designed in university or medical centre settings may vary in the extent to which they are able to match the needs of families with these different cultural backgrounds. Two factors that may make the IY programme suitable for a wide range of families[60] include its inbuilt collaborative and flexible style, and the efforts it makes to remove barriers to accessing for families, including provision of child care and meals during the group sessions.

Examining the wider public health literature may also help explain why no adverse equity effects were identified. A recent review by Lorenc et al.[28] asked whether or not there are particular types of public health interventions that appear more likely to produce intervention-generated inequalities or, conversely, to reduce inequalities. Their 'rapid overview' of systematic reviews containing data on differential effects of public health interventions found 12 relevant reviews. They concluded that the limited evidence suggested that more 'upstream', structural and policy-level interventions were more likely to have positive equity effects. Examples included differential pricing policies for tobacco, providing free nutrient supplements and structural workplace interventions. On the other hand, more 'downstream' interventions,[27] focusing on individual and behavioural factors, such as media campaigns and school-based interventions, appeared to be more likely to increase inequalities. Leaving aside the difficult question of what might count as upstream or downstream from among these examples, it seems that, by any definition, parenting interventions would count as 'downstream' yet do not appear to widen socioeconomic inequalities, although neither do they serve to reduce them. Perhaps the close focus on individual needs and values in the IY programme helps offset any adverse effect that otherwise might result from education-based programmes, or maybe the upstream–downstream distinction is not the crucial feature of interventions that is likely to affect equity.

Is the Incredible Years intervention cost-effective?

The economic analysis found the level of service use among study participants to be low, with most support coming from primary care and hospitals, but with little involvement of specialist mental health services. There were no differences in costs between the treatment and the control groups, other than the cost of the IY intervention. The average cost per family of the intervention for this sample was £2414 (2014 prices). Although this is higher than the previous 'best estimate' of £952 (2009 prices),[167] it is in line with previously reported costs of the IY intervention.[169]

Costs varied based on age and gender, but no variations were associated with social disadvantage, ethnicity, ADHD, child emotional problems, levels of disruptive behaviour or parental depression.

Given that 'cost-effectiveness' is driven by both costs and effects, further analyses explored cost-effectiveness of the IY intervention based on age and gender (cost differences identified) and baseline level of disruptive behaviour and parental depression (effect difference identified). These analyses indicate that the IY intervention is less likely to be considered cost-effective for children who scored below the clinical threshold on the ECBI-I at baseline and children aged < 5 years. At the same time, it would be more likely to be considered cost-effective for boys than for girls, and for children whose parents had a BDI-II score at baseline indicating at least a mild level of depression. It should be noted that this analysis was limited to cases with full data, and results should therefore be interpreted with caution.

Although comparisons between cost-effectiveness studies are difficult because of methodological differences, our findings appear to be similar to a previous cost-effectiveness analysis of the IY intervention,[169] in which the chance that the intervention would be considered cost-effective was 99% at a WTP of around £150 (2004 prices), whereas we estimate it to be 99% at £145 (2014 prices). However, these figures needs to be evaluated in the context of the likely long-term societal costs associated with persistent conduct disorder.

The findings from our longer-term economic modelling indicate that if the costs associated with persistent conduct disorder are in fact lower than previously thought, at the current price and effectiveness, the IY intervention would not result in longer-term savings. However, in our 'high-cost' scenario, in which the assumptions are more in line with previous literature, the return on investment is substantial, with average net savings of between £5000 and £7000 per child.

Strengths and limitations of the study

Strengths
The study has many strengths that increase our confidence in the findings. Our unique pooled data set means that we have a much larger sample than other studies in this field, and findings that are potentially generalisable across countries, service settings, families and level of child problems. The use of pooled individual-level data is vital for testing moderators and has many advantages compared with traditional meta-analysis, in which data are analysed at a trial level by producing average treatment effects weighted by sample size. A potential problem with this traditional approach is that effects are more likely to be influenced by common causes that occur at a trial level. In this study we analyse data at an individual level rather than a trial level. Individual-level meta-analysis allows the separation of between- and within-trial effects,[80] meaning that individual-level effects can be distinguished from trial-level effects that are more likely to be confounded. Individual-level meta-analysis also allows for the analysis of a greater sample size (1799 individuals compared with 14 trials, if this had been a traditional meta-analysis), meaning that there is greater statistical power to detect effects.

There are a number of statistical challenges in pooling data from different trials at an individual level, which our methods were able to address. First, the analysis model must account for the fact that there are trial-level differences in both outcome and treatment effect. Second, different trials used different designs, such as clustered randomisation and different therapy groups in a trial for participants in the intervention arm, and these needed to be accounted for in the analysis model. A multilevel model,[188] which accounts for differences between trials, is used to address these issues. Finally, we account for missing data in the moderator analysis using MI,[143] an approach that requires less restrictive assumptions than using only cases with no missing values.

Furthermore, plausibility of the findings is enhanced in other ways by our analytical strategies. Randomisation appeared adequate, as we found no evidence for baseline imbalances across the pooled data set. We controlled for the possibility of confounding by the type of sample selection into the trial, that is, whether or not it is a selective prevention compared with an indicated prevention or treatment trial, as this can make a considerable difference to the initial severity of child disruptive behaviour. Sensitivity analyses found no differences according to the type of control group used, such as treatment-as-usual

compared with no intervention. We were able to check for non-linear effects, to take account of the possibility that there may be more complex relationships between moderator and outcome. For example, it could be that mild to moderate levels of parent depression enhance the effects of IY on child outcomes, compared with low depression, but, that at very high levels, parents are unable to effectively engage with the programme once they have signed up for it. We found no evidence for this kind of non-linear effect in the depression moderator data.

Plausibility of the findings is also enhanced by reanalysis of complete trial data sets, using a preplanned data-analytic strategy, applied consistently across all trials. This approach is highly likely to reduce the risk of reporting and publication bias, compared with conducting moderator analyses in small trials, which are designed and powered to test main effects but are generally underpowered for moderator analyses. Low power combined with absence of prespecified protocols for moderator analyses in the existing literature would tend to lead to a high risk of publication and reporting bias.

Our data also allowed us to examine within- vs. between-trial moderation effects; these make for an interesting comparison, as they show us how the moderator findings would have looked if we had conducted a conventional metaregression analysis, compared with a meta-analysis with IPD. In some cases, there was evidence of moderator effects at an individual level and none at the trial level (e.g. for parent depression), meaning that a conventional metaregression would have led us to miss significant moderator effects. In other cases, there was evidence of a trial-level moderator effect, which in the individual-level analysis was non-significant and, in the case of ethnicity, was reversed in direction. Thus, with conventional metaregression, we would have concluded that families from ethnic minorities did less well, whereas the IPD analysis suggested that they in fact tended to do slightly (non-significantly) better. We attributed this discrepancy to confounding at trial level by severity of child disruptive behaviour, such that trials with higher proportion of ethnic minority families in them were more likely to be selective prevention trials, and to have lower levels of disruptive behaviour. Interestingly, in one case, trial- and individual-level analyses concurred in showing significant moderation, by baseline level of child disruptive behaviour, although again the magnitude of this modification effect appeared larger in the trial-level analysis.

Our data also allow us to compare the conclusions that might be drawn from predictor compared with moderator analyses. Examining the charts for binary moderator findings, it appears as if the same conclusions would have been drawn from analysing predictors in the intervention group only, without examining the interaction with treatment status. However, we did not plan to make this comparison formally and, in any case, we could not have known this without doing full moderator/interaction analysis. It would not be recommended to conduct predictor analyses in order to understand equity effects, as in some studies, predictors and moderators do show different patterns of results.[46]

Limitations
Alongside these substantial strengths of the study, there are a number of limitations. In particular, assumptions and compromises inevitably had to be made in data harmonisation. First, there is the problem of variation in instruments used. Although many trials used similar instruments to assess the primary constructs in this study, we had to harmonise data across instruments for every construct. Whenever data were available to cross-validate our harmonisation procedure, for example by computing correlations between the different instruments, results were reassuring that the instruments indeed seemed to tap into the same constructs. Nevertheless, we had to make several assumptions in combining data across instruments, for example that the different instruments measured the same construct with the same measurement error. Combining data across instruments assumes measurement of the same construct in each instrument and the same measurement error. We were, to a limited extent, able to check whether or not the different instruments indeed measured the same construct. For our primary outcome measure, for example, some trials included data on both instruments used and correlation analysis revealed satisfactory association between the instruments of $r = 0.70$. Because for some constructs (e.g. parental depression) no overlap in instruments within trials was available, these checks were unfortunately not possible for all

constructs for which we harmonised data. Similar problems arose from the need to combine economic data collected using different instruments. These are discussed in *Appendix 7*.

Second, assumptions had to be made in using norm scores to harmonise data across trials on child disruptive behaviour and parent depression. For example, we used the same norm scores across trials from different countries, because using different norm scores for different samples would have differentially changed the distribution of scores within the samples. Moreover, norm scores were often not available for every measure from every country. We therefore chose to use one set of norm scores, namely the instruments' original and validated norm scores, for all samples. A potential bias of this procedure is that the original norm scores may not reflect the actual norm scores in another country.

Third, social status and SES were the primary moderators of interest in this study. Availability of different indicators of social and socioeconomic disadvantage (e.g. income, educational level) across trials allowed us to approach the putative moderator role of SES in multiple ways, a recommended approach for studying SES. The individual indicators, however, were all binary variables. Although most individual trials provided data for some variables at an ordinal level, harmonisation often required us to simplify these data into binary variables. From an applied perspective, it may be especially interesting to understand the extent to which interventions benefit the most disadvantaged families. From a statistical view, however, binary variables ignore much of the meaningful variation between families who range on a dimension of SES.

Fourth, there was a great deal of variation between trials in the measures of self-reported parenting used, presenting greater challenges in harmonising these variables than most others. In order to find items that were common across trials, we often had to rely on very small numbers of items for some parenting constructs, and some items were measured in only a small number of trials, leading to considerable reduction in the sample size for some of the analyses, as can be seen in *Tables 3* and *4* (see *Chapter 2*).

Fifth, we were unable to meaningfully synthesise the parent satisfaction data, owing to high levels of missing data. Five trials used the same instrument, the Parent Satisfaction Questionnaire (PSQ), which is administered only to parents in the intervention group. There were 379 families in the intervention groups in these trials; however, with a response rate of 55%, this yielded data on only 210 families, representing only 19% of all families in the intervention condition (the total sample size in the intervention condition is 1116).

Sixth, our study, and the underlying trials, were limited by many of the key outcomes being measured by parent report. It should be noted that most of the trials in our pool also included systematic direct observational measures. Observational measures, although not without technical and other challenges, have many advantages;[189,190] in particular, they allow for coding of data by researchers who are blinded to the intervention status of the families. In many cases observational data showed positive effects on child behaviour and on parenting skill outcomes (e.g. Gardner *et al.*[103] and Hutchings *et al.*[85]), increasing our confidence in the validity of the main findings, based on parent self-report. However, our ability to incorporate observational data in the pooled set depended upon the possibilities for harmonisation of measures in the individual trials. Data on observed child and parenting behaviour were too diverse in terms of content (e.g. compliance vs. negative affect in children) and structure (e.g. Likert scales vs. frequency counts) to harmonise across trials, within the scope of this project. Teacher-reported child behaviour would potentially provide a more independent source; however, these measures were available only in a few trials.

Seventh, although there were many important questions for policy that we would like to have asked about the effects of trial-level moderators, these analyses were of relatively limited value, owing to low power ($n = 13$) and much potential for confounding at trial level. Furthermore, some trial-level (or site-level) variables showed little variation, such as rural compared with urban intervention site, with most interventions being delivered in urban settings.

Eighth, our moderator findings can assess factors predicting success only once parents have signed on to the programme; we cannot tell if different factors might affect equity of initial access to the programme.

Ninth, our study was unable to inform us about the longer-term effects of IY, because most of the studies used a waiting list design, with families in the control group being offered the intervention ≥ 6 months later. However, it should be noted that a recent systematic review of longer-term effects of largely similar social learning theory-based parenting programmes found that the effects of parenting programmes on disruptive child behaviour are generally highly stable. This review found that the mean effect size of change between post test and follow-up was 0.01, indicating strong maintenance of change compared with control groups, at least up to 3 years after the end of the programme.[191]

Finally, it is possible that our findings may not be generalisable to other parenting interventions. On the one hand, the content and principles of the IY programme are very similar to other social learning theory-based parenting programmes. On the other hand, it is possible that IY differs in process more than content; the marked and explicit emphasis in IY on a collaborative and flexible approach, in which parents design goals and strategies to suit their own family situation, may make it more likely to be beneficial to a wide range of families. Partly this may be a function of how carefully these skills are trained and supervised in the IY programme compared with other programmes. At the same time, some group leaders may acquire these kinds of strong clinical skills from their generic training or from training in other parenting programmes rather than from the IY programme. Further IPD meta-analyses would be needed to establish if a similar pattern of moderator effects is found in other programmes.

Recommendations for research

First, we strongly suggest that greater use is made of existing resources, by sharing and pooling IPD from intervention trials. The many advantages of IPD for understanding moderator effects, and in particular health equity effects, are clear from this study. In addition, there are advantages for understanding wider benefits and harms of interventions. There are a number of ways in which trial conduct and design could be improved, with future IPD in mind, which would also bring benefits in terms of rigour and transparency. In order to facilitate data sharing, we suggest that it is vital that funders, journal editors and other stakeholders enhance their policies on open data, or, when these already are in place, further monitor the implementation of these policies, as recommended by recent transparency initiatives such as the *British Medical Journal* 'AllTrials' campaign (www.alltrials.net/) and Open Science Framework Transparency and Openness guidelines (https://cos.io/top/) (both accessed 20 December 2016). To facilitate data pooling, a number of factors are important, including full and accurate reporting of trials, in line with CONSORT (Consolidated Standards of Reporting Trials) guidelines, and appropriate extensions, including the new CONSORT-SPI (CONSORT Statement for Social and Psychological Interventions), developed for social, psychological and public health interventions.[192] Ethics approval and participant consent for trials needs to allow for, and not hinder, future data sharing, provided that data are carefully and fully anonymised. Investigators and ethics committees should ensure that data are retained rather than lost or destroyed.

Second, owing to the potential for bias from parent reports, trials should continue also to use direct observational measures of child and parent behaviour, for which researchers can be blinded to the parenting intervention condition.[93,189] Although harmonising of observational data would be challenging, future IPD work should explore the feasibility of this. In order to overcome this limitation in the field, triallists should aim to use common observational instruments when possible.

Third, in order to better understand the longer-term effects of parenting programmes, triallists should refrain from using waiting list control group designs.

Fourth, we recommend further work on moderators of intervention effects, using IPD meta-analysis. Future IPD analyses should test whether or not these findings generalise to other types of parenting interventions

(see *Strengths and limitations of the study*). Although we have found from a well-powered study that few of the potential moderators, especially in the SES domain, have an effect, this still leaves us not knowing what does explain the substantial variability found in outcome, between and within trials. It might be that there are individual-level differential susceptibilities to intervention effects, such as genetic or temperamental factors, that might apply to parents or to children (e.g. Scott *et al.*[193] and Chhangur *et al.*[86]). For these studies to further our understanding, they would preferably be factors that were not simply markers for more easily observable moderators, such as depression or disruptive behaviour. Alternatively, it might be that variation in outcome depends more on factors that are hard to assess at baseline, such as shifting family circumstances, or the relationship between group leader and parent.

Fifth, harmonisation of all kinds of trial data, on outcomes, moderators and mediators, would be greatly facilitated by use of common outcome measures, as well as common indices of SES and implementation. At the same time, it needs to be recognised that different studies may have different scientific purposes and constraints, and that uniform measurement may not necessarily be a completely attainable or desirable goal.

Sixth, we suggest that IPD should also be used to investigate mediators or mechanisms of change in interventions, overcoming problems of lower power inherent in many single trials. Similarly to the moderators literature, it is unclear if there is substantial reporting bias in the published mediation literature, as most parenting trials measure potential mediating variables (e.g. parenting skill) but many trials do not test for them. Trials ideally need to be designed with three or more assessment points, in order to conduct these analyses, so that baseline, mediator and outcome can all be separated in time.[194]

Finally, further primary research, and up-to-date research synthesis, is needed on factors predicting equity of initial access to parenting programmes, as our study could examine outcomes only for those families who had signed up to the IY programme.

Public health policy and service delivery implications

Owing to its power to detect robust findings across a wide range of contexts, this study has important implications for the commissioning and delivery of services. It is worth re-emphasising the extremely poor outcomes for children with disruptive behaviour and that, if these are untreated, they are considerably worse for those who come from disadvantaged backgrounds. The increased risk of poor outcomes in adulthood is very large for children with disruptive behaviour, of the order of an 800–1000% increase on measures such as serious substance abuse, teenage pregnancy, leaving school without any qualifications, domestic violence, unemployment, criminal conviction and premature death.[195] The increased prevalence in disadvantaged groups is also striking; thus, in careful analyses of the definitive ONS surveys, the prevalence of conduct disorder (persistent serious antisocial behaviour) was 600% greater in the lowest income quintile (poorest fifth of the population) than in the highest income quintile (richest fifth of the population).[196] All of the family factors tested in this study predict worse outcomes in untreated groups, and it may have been thought that they would make it harder for such families to benefit from the intervention: with poorer living conditions it is harder to concentrate and make mental space for the parenting task; as a lone parent there are fewer hands to help; and as a teenage parent one has had less life experience and less experience of parenting. Many ethnic minority families still experience discrimination and may also feel that the intervention or the people delivering it were not sufficiently attuned to their cultural needs. In this context, the findings of this study are heartening and encouraging: children whose parents fall into these disadvantaged groups do just as well in terms of their improvement in behaviour, and there is reasonable hope that the good effects will be maintained over the longer term, thus reducing social inequity. A recent 8- to 10-year follow-up of IY parenting programmes still showed significant effects in reducing the prevalence rates of oppositional defiant and conduct disorders, with half the rate in those who were allocated to IY.[197] We note, however, that these findings should be interpreted cautiously, as the full randomised sample was not retained at follow-up. Thus, because disruptive behaviours are so much more common in disadvantaged groups, making these interventions available to all populations whose children have this difficulty offers the opportunity of

reducing social inequity. This is in contrast with some other health improvement programmes, such as anti-smoking programmes, which are more effective in more advantaged groups, so have the paradoxical effect of increasing social inequity.

The major Department of Health strategy document *Future In Mind*[198] concerns 'Promoting, protecting and improving our children and young people's mental health and well-being', and three major themes are the need for increased access, for interventions to be effective and for interventions to be socially equitable. Our study found successful outcomes in settings such as schools, Sure Start and other children's services, in both local authority and non-governmental organisation (NGO) sectors. This shows that making IY parenting programmes available in the general community, outside formal CAMHSs, is potentially a good way to meet these Department of Health goals. However, it is important to stress that all parenting programmes are not the same, and that the IY is notable, first in terms of its content and processes, whereby there is an extensive training and certification process and ongoing supervision to maintain fidelity, and, second, that independent replications, on which this study is based, uphold the programme developer's findings, which is not always the case for other parenting interventions, for which either the original evaluation by the developers showed no effect (e.g. the Family Links Programme[199]) or independent evaluations found much smaller effects (e.g. Triple P in Birmingham[119]). Therefore, it may be preferable for commissioners to ensure, first, that they are commissioning formal parenting programmes and, second, that they are high-quality ones with a robust evidence base.

The particular findings of the individual moderation analyses are also important for policy and commissioning, as they suggest that the current programme does not need significant modification to deal with particular subgroups. This is interesting in the current zeitgeist towards personalised interventions and personalised medicine, for which modifying interventions to be suitable for certain subgroups is proposed as the way forward to increase effectiveness.[200] Thus, for example, if the findings had found significant differences for groups from ethnic minorities, then the commissioning process would become more complicated, as it may suggest that a different type of programme should be used.

It is also noteworthy that parents with harsher parenting did just as well, so that, again, a different approach does not seem indicated. And the fact that parents who are depressed seemed to do better than those were not depressed is encouraging, and suggests that in the first instance this programme on its own is suitable and does not require modification for parents who are depressed. Indeed, on the contrary, our findings suggest that in such cases it may be a more effective and more cost-effective method of reducing child disruptive behaviour. The finding that ADHD problems improve and that they do not worsen the outcome of disruptive behaviour supports the recommendation of NICE[201] that the first-line intervention for ADHD should be parenting programmes.

The fact that older age did not lead to reduced effectiveness is also pertinent. The age range was wide, from 2 to 10 years, and it has been argued that late intervention is ineffective and all intervention should be early, with more resources targeted at the younger age range. However, this study shows that later intervention, at least up to the age of 10 years, is just as effective and, according to the subsample with economic analyses, may be more cost-effective, so that it is possible that resources could be equally well (and possibly better) deployed in school-aged children.

Finally, there could be many benefits for population health policy of this kind of IPD approach to studying intervention equity effects. Our study suggests that children can be successfully screened or referred into this programme, parents are prepared to attend, and the programme achieves good outcomes for children and appears to be cost-effective. There is clear evidence from the data that this programme would not widen socioeconomic inequalities in outcomes. Moreover, it might well narrow inequalities in terms of the psychological well-being of children and parents, as well as being more cost-effective for the more distressed families. Our study is potentially a model for other public health policy questions, which could mine existing data, using IPD meta-analysis, at relatively low cost, and high value, in order to enhance

understanding of effectiveness, cost-effectiveness and equity effects of different commissioning strategies, and to promote social mobility.

Public and patient involvement

The public and patient involvement (PPI) survey for the study (see *Appendix 6*) agreed remarkably with the quantitative findings. Thus, the parents interviewed, who included those with low income and from ethnic minorities, believed that programme was likely to be just as effective for ethnic minorities, those who were from disadvantaged backgrounds or those who were depressed. This is very important, as if parents believed differently then they are either less likely to attend in the first place or more likely to drop out. These findings therefore underline the suitability of providing the programme to families from the range of disadvantage indicators.

Chapter 7 Conclusions

Successful synthesis of data from almost all randomised trials of the IY parenting intervention in Europe led to a uniquely large ($n = 1799$) and diverse sample that allowed for the most stringent and well-powered tests of equity effects to date, wider health benefits and cost-effectiveness of parenting interventions for children with, or at risk for, disruptive behaviour problems aged 2–10 years. The IY parenting intervention does not increase social or socioeconomic inequalities in children's disruptive behaviour problems. Families with a range of social and socioeconomic disadvantages and those from ethnic minorities are just as likely to benefit. If anything, inequalities are reduced because some of the more distressed families (those with more severe disruptive child behaviour and those with higher levels of parental depression) benefit more. This is important, because these are a group of children who tend to be quite disadvantaged, and who, without intervention, tend to have poorer outcomes than the rest of the population across all domains. Contrary to current emphasis in neuroscience and policy on early intervention, there were no effects of child age on intervention benefit, with older children just as likely to benefit as those in the early years. Furthermore, the economic analysis found a significant negative association between costs at follow-up and age, indicating that the IY intervention may be more likely to be considered cost-effective for older children. In addition to reducing disruptive child behaviour, the intervention reduces children's ADHD symptoms and improves several aspects of positive parenting (e.g. parents praised their children more) and negative parenting (e.g. parents used less harsh and inconsistent discipline).

Our cost-effectiveness analyses using the pooled data supports previous findings from individual trials, and it is therefore likely that the IY intervention can provide savings to the public sector in the longer term.

Future use of this unique data set will increase insight into the mechanisms (i.e. mediators) through which the intervention operates. We encourage future work to focus on the extent to which different groups access parenting interventions because our work suggests that once families participate, the IY parenting intervention successfully reduces disruptive child behaviour in families across wide ranges of social and socioeconomic backgrounds.

Acknowledgements

We are most grateful to the following:

Anna Goodman, London School of Hygiene and Tropical Medicine, London, UK (project advisory group).

George Howe, The George Washington University, Washington, DC, USA (project advisory group).

Wendy Knerr, Oxford University, Oxford, UK (assistance with editing and writing the report).

Andrew Pickles, King's College London, London, UK (advised on grant proposal).

Three anonymous reviewers, whose thoughtful comments strengthened the report.

We are most grateful to the following, who were key in providing access to data and invaluable information about included studies:

Ulf Axberg, Göteborg University, Sweden [principal investigator (PI), Swedish trial data site].

Vashti Berry, Plymouth University, Plymouth, UK (co-PI, Birmingham trial data site).

Maria Filomena Gaspar, University of Coimbra, Coimbra, Portugal (PI, Portuguese trial data site).

Bjørn Helge Handegård, The Arctic University of Norway, Tromsø, Norway (Statistician, Norwegian trial data site).

Grainne Hickey, National University of Ireland Maynooth, Maynooth, Ireland (Researcher, Irish trial data site).

Maria Joao, University of Coimbra, Coimbra, Portugal (PI, Portuguese trial data site).

Sinead McGilloway, National University of Ireland Maynooth, Maynooth, Ireland (PI, Irish trial data site).

Ankie Menting, Utrecht University, Utrecht, the Netherlands (Researcher, Dutch prisoners trial data site).

Willy-Tore Mørch, The Arctic University of Norway, Tromsø, Norway (PI, Norwegian trial data site).

Louise Morpeth, Dartington Social Research Unit, Dartington, UK (PI, Birmingham trial data site).

Bram Orobio de Castro, Utrecht University, Utrecht, the Netherlands (PI, both Dutch trials).

Margiad Elen Williams, Bangor University, Bangor, UK (Researcher, Welsh trials data sites).

Cheng Zhang, King's College London, London, UK (Researcher, King's College London trials data site).

Contributions of authors

Frances Gardner (Professor of Child and Family Psychology) wrote the grant, led the study and wrote the report.

Data sharing statement

We shall make data available to the scientific community with as few restrictions as feasible, via the corresponding author, Gardner, within 2 years of publication. This is because (1) we reserve the right to retain exclusive use until the publication of major outputs, and (2) these are secondary data, and there are intellectual property and legal issues around further sharing for some of the original data holders.

References

1. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, Tierney JF, PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015;**313**:1657–65. http://dx.doi.org/10.1001/jama.2015.3656

2. Office for National Statistics (ONS). *Ethnic Group Statistics: A Guide for the Collection and Classification of Ethnicity Data*. London: ONS; 2004. URL: www.ons.gov.uk/ons/guide-method/ measuring-equality/equality/ethnic-group-statistics/ethnic-group-statistics--a-guide-for-the-collection-and-classification-of-ethnicity-data.pdf (accessed October 2014).

3. Fergusson DM, Horwood LJ, Ridder EM. Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *J Child Psychol Psychiatry* 2005;**46**:837–49. https://doi.org/10.1111/j.1469-7610.2004.00387.x

4. Odgers CL, Caspi A, Broadbent JM, Dickson N, Hancox RJ, Harrington H, *et al.* Prediction of differential adult health burden by conduct problem subtypes in males. *Arch Gen Psychiatry* 2007;**64**:476–84. https://doi.org/10.1001/archpsyc.64.4.476

5. Piquero AR, Shepherd I, Shepherd JP, Farrington DP. Impact of offending trajectories on health: disability, hospitalisation and death in middle-aged men in the Cambridge Study in Delinquent Development. *Crim Behav Ment Health* 2011;**21**:189–201. http://dx.doi.org/10.1002/cbm.810

6. Sainsbury Centre for Mental Health (SCMH). *The Chance of a Lifetime: Preventing Early Conduct Problems and Reducing Crime*. London: SCMH; 2009.

7. Scott S, Knapp M, Henderson J, Maughan B. Financial cost of social exclusion: follow up study of antisocial children into adulthood. *BMJ* 2001;**323**:191. https://doi.org/10.1136/bmj.323.7306.191

8. Ermisch J. Origins of social immobility and inequality: parenting and early child development. *Natl Inst Econ Rev* 2008;**205**:62–71.

9. Hoeve M, Dubas JS, Eichelsheim VI, van der Laan PH, Smeenk W, Gerris JR. The relationship between parenting and delinquency: a meta-analysis. *J Abnorm Child Psychol* 2009;**37**:749–75. http://dx.doi.org/10.1007/s10802-009-9310-8

10. Moffitt T, Caspi A. Early Prevention of Adult Antisocial Behaviour. In Farrington D, Coid J, editors. *Early Prevention of Adult Antisocial Behaviour*. Cambridge: Cambridge University Press; 2003.

11. Barlow J, Smailagic N, Huband N, Roloff V, Bennett C. Group-based parent training programmes for improving parental psychosocial health. *Cochrane Database Syst Rev* 2012;**6**:CD002020. https://doi.org/10.1002/14651858.cd002020.pub3

12. Furlong M, McGilloway S, Bywater T, Hutchings J, Smith SM, Donnelly M. Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database Syst Rev* 2012;**2**:CD008225. http://dx.doi.org/ 10.1002/14651858.CD008225.pub2

13. National Institute for Health and Care Excellence (NICE). *Antisocial Behaviour and Conduct Disorders in Children and Young People: Recognition, Intervention and Management, NICE Clinical Guideline 158*. London: NICE; 2013. URL: www.nice.org.uk/guidance/cg158 (accessed October 2014).

14. World Health Organization (WHO). *Violence Prevention: The Evidence*. Geneva: WHO; 2010.

15. United Nations Office on Drugs and Crime (UNODC). *Guide to Implementing Family Skills Training Programmes for Drug Abuse Prevention*. New York, NY: United Nations; 2009.

16. Centers for Disease Control and Prevention (CDC). *Parent Training Programs: Insight for Practitioners*. Atlanta, GA: CDC; 2009.

17. National Research Council/Institute of Medicine. *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Washington, DC: The National Academies Press; 2009.

18. Department of Health (DH). *No Health Without Mental Health: A Cross-Government Mental Health Outcomes Strategy for People of all Ages – a Call to Action*. London: DH; 2011.

19. Department of Health (DH). *Talking Therapies: a Four-Year Plan of Action*. London: DH; 2011.

20. Bonin EM, Stevens M, Beecham J, Byford S, Parsonage M. Costs and longer-term savings of parenting programmes for the prevention of persistent conduct disorder: a modelling study. *BMC Public Health* 2011;**11**:803. http://dx.doi.org/10.1186/1471-2458-11-803

21. Lee S, Aos S, Pennucci A. *What Works and What Does Not? Benefit-cost Findings from WSIPP (Doc. No. 15-02-4101)*. Olympia, WA: Washington State Institute for Public Policy; 2015.

22. Tugwell P, de Savigny D, Hawker G, Robinson V. Applying clinical epidemiological methods to health equity: the equity effectiveness loop. *BMJ* 2006;**332**:358–61. https://doi.org/10.1136/bmj.332.7537.358

23. Welch V, Petticrew M, Tugwell P, Moher D, O'Neill J, Waters E, White H, PRISMA-Equity Bellagio group. PRISMA-Equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. *PLOS Med* 2012;**9**:e1001333. http://dx.doi.org/10.1371/journal.pmed.1001333

24. Whitehead M. The concepts and principles of equity and health. *Int J Health Serv* 1992;**22**:429–45. https://doi.org/10.2190/986L-LHQ6-2VTE-YRRN

25. Victora CG, Vaughan JP, Barros FC, Silva AC, Tomasi E. Explaining trends in inequities: evidence from Brazilian child health studies. *Lancet* 2000;**356**:1093–8. https://doi.org/10.1016/S0140-6736(00)02741-0

26. Waylen A, Stallard N, Stewart-Brown S. Parenting and health in mid-childhood: a longitudinal study. *Eur J Public Health* 2008;**18**:300–5. http://dx.doi.org/10.1093/eurpub/ckm131

27. White M, Adams J, Heywood P. How and Why Do Interventions that Increase Health Overall Widen Inequalities Within Populations? In Barbones S, editor. *Health, Inequality and Public Health*. Bristol: Policy Press; 2009. pp. 65–83. https://doi.org/10.1332/policypress/9781847423207.003.0005

28. Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. *J Epidemiol Community Health* 2013;**67**:190–3. http://dx.doi.org/10.1136/jech-2012-201257

29. Rutter M. Is Sure Start an effective preventive intervention? *Child Adolesc Ment Health* 2006;**11**:135–41.

30. Gardner F, Montgomery P, Knerr W. Transporting evidence-based parenting programs for child problem behavior (age 3-10) between countries: systematic review and meta-analysis. *J Clin Child Adolesc Psychol* 2016;**45**:749–62. http://dx.doi.org/10.1080/15374416.2015.1015134

31. Ward C, Sanders MR, Gardner F, Mikton C, Dawes A. Preventing child maltreatment in low- and middle-income countries: parent support programs have the potential to buffer the effects of poverty. *Child Abuse Negl* 2016;**54**:97–107. https://doi.org/10.1016/j.chiabu.2015.11.002

32. Webster-Stratton C, Reid J. The Incredible Years. In Weisz J, Kazdin A, editors. *Evidence-based Psychotherapies*. 2nd edn. New York, NY: Guilford; 2010.

33. Cardona JP, Holtrop K, Córdova D, Escobar-Chew AR, Horsford S, Tams L, *et al.* 'Queremos aprender': Latino immigrants' call to integrate cultural adaptation with best practice knowledge in a parenting intervention. *Fam Process* 2009;**48**:211–31. https://doi.org/10.1111/j.1545-5300.2009.01278.x

34. Reid MJ, Webster-Stratton C, Baydar N. Halting the development of conduct problems in head start children: the effects of parent training. *J Clin Child Adolesc Psychol* 2004;**33**:279–91. http://dx.doi.org/10.1207/s15374424jccp3302_10

35. Shaw DS, Dishion TJ, Supplee L, Gardner F, Arnds K. Randomized trial of a family-centered approach to the prevention of early conduct problems: 2-year effects of the family check-up in early childhood. *J Consult Clin Psychol* 2006;**74**:1–9. https://doi.org/10.1037/0022-006X.74.1.1

36. Tein JY, Sandler IN, MacKinnon DP, Wolchik SA. How did it work? Who did it work for? Mediation in the context of a moderated prevention effect for children of divorce. *J Consult Clin Psychol* 2004;**72**:617–24. http://dx.doi.org/10.1037/0022-006X.72.4.617

37. Chamberlain P, Price J, Leve LD, Laurent H, Landsverk JA, Reid JB. Prevention of behavior problems for children in foster care: outcomes and mediation effects. *Prev Sci* 2008;**9**:17–27. http://dx.doi.org/10.1007/s11121-007-0080-7

38. Kelly Y, Sacker A, Del Bono E, Francesconi M, Marmot M. What role for the home learning environment and parenting in reducing the socioeconomic gradient in child development? Findings from the Millennium Cohort Study. *Arch Dis Child* 2011;**96**:832–7. http://dx.doi.org/10.1136/adc.2010.195917

39. Belsky J, Conger R, Capaldi DM. The intergenerational transmission of parenting: introduction to the special section. *Dev Psychol* 2009;**45**:1201–4. http://dx.doi.org/10.1037/a0016245

40. Kim HK, Capaldi DM, Pears KC, Kerr DC, Owen LD. Intergenerational transmission of internalising and externalising behaviours across three generations: gender-specific pathways. *Crim Behav Ment Health* 2009;**19**:125–41. http://dx.doi.org/10.1002/cbm.708

41. Schepman K, Collishaw S, Gardner F, Maughan B, Scott J, Pickles A. Do changes in parent mental health explain trends in youth emotional problems? *Soc Sci Med* 2011;**73**:293–300. http://dx.doi.org/10.1016/j.socscimed.2011.05.015

42. Langton EG, Collishaw S, Goodman R, Pickles A, Maughan B. An emerging income differential for adolescent emotional problems. *J Child Psychol Psychiatry* 2011;**52**:1081–8. http://dx.doi.org/10.1111/j.1469-7610.2011.02447.x

43. Lundahl B, Risser HJ, Lovejoy MC. A meta-analysis of parent training: moderators and follow-up effects. *Clin Psychol Rev* 2006;**26**:86–104. https://doi.org/10.1016/j.cpr.2005.07.004

44. Reyno SM, McGrath PJ. Predictors of parent training efficacy for child externalizing behavior problems – a meta-analytic review. *J Child Psychol Psychiatry* 2006;**47**:99–111. https://doi.org/10.1111/j.1469-7610.2005.01544.x

45. Shelleby EC, Shaw DS. Outcomes of parenting interventions for child conduct problems: a review of differential effectiveness. *Child Psychiatry Hum Dev* 2014;**45**:628–45. http://dx.doi.org/10.1007/s10578-013-0431-5

46. Gardner F, Hutchings J, Bywater T, Whitaker C. Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *J Clin Child Adolesc Psychol* 2010;**39**:568–80. http://dx.doi.org/10.1080/15374416.2010.486315

47. Gardner F, Connell A, Trentacosta CJ, Shaw DS, Dishion TJ, Wilson MN. Moderators of outcome in a brief family-centered intervention for preventing early problem behavior. *J Consult Clin Psychol* 2009;**77**:543–53. http://dx.doi.org/10.1037/a0015622

48. McGilloway S, Ni Mhaille G, Bywater T, Furlong M, Leckey Y, Kelly P, *et al.* A parenting intervention for childhood behavioral problems: a randomized controlled trial in disadvantaged community-based settings. *J Consult Clin Psychol* 2012;**80**:116–27. http://dx.doi.org/10.1037/a0026304

49. Kane GA, Wood VA, Barlow J. Parenting programmes: a systematic review and synthesis of qualitative research. *Child Care Health Dev* 2007;**33**:784–93. https://doi.org/10.1111/j.1365-2214.2007.00750.x

50. Furlong M, McGilloway S. The Incredible Years parenting program in Ireland: a qualitative analysis of the experience of disadvantaged parents. *Clin Child Psychol Psychiatry* 2012;**17**:616–30. http://dx.doi.org/10.1177/1359104511426406

51. Morch W-T, Clifford G, Larsson B, Rypal P, Tjeflaat T, Lurie J, *et al. The Norwegian Webster-Stratton Program 1998–2004*. Trondheim: University of Trondheim; 2004.

52. Murray J, Farrington DP. Risk factors for conduct disorder and delinquency: key findings from longitudinal studies. *Can J Psychiatry* 2010;**55**:633–42. https://doi.org/10.1177/070674371005501003

53. Brooks-Gunn J, Markman LB. The contribution of parenting to ethnic and racial gaps in school readiness. *Future Child* 2005;**15**:139–68. https://doi.org/10.1353/foc.2005.0001

54. Reid MJ, Webster-Stratton C, Beauchaine TP. Parent training in head start: a comparison of program response among African American, Asian American, Caucasian, and Hispanic mothers. *Prev Sci* 2001;**2**:209–27. https://doi.org/10.1023/A:1013618309070

55. Kumpfer KL, Alvarado R, Smith P, Bellamy N. Cultural sensitivity and adaptation in family-based prevention interventions. *Prev Sci* 2002;**3**:241–6. https://doi.org/10.1023/A:1019902902119

56. Castro FG, Barrera M, Holleran Steiker LK. Issues and challenges in the design of culturally adapted evidence-based interventions. *Annu Rev Clin Psychol* 2010;**6**:213–39. http://dx.doi.org/10.1146/annurev-clinpsy-033109-132032

57. Huey SJ, Polo AJ. Evidence-based psychosocial treatments for ethnic minority youth. *J Clin Child Adolesc Psychol* 2008;**37**:262–301. http://dx.doi.org/10.1080/15374410701820174

58. Moran P, Ghate D, van der Merwe A. *What Works in Parenting Support? A Review of the International Evidence*. London: Her Majesty's Stationery Office; 2004.

59. Scott S, O'Connor TG, Futh A, Matias C, Price J, Doolan M. Impact of a parenting program in a high-risk, multi-ethnic community: the PALS trial. *J Child Psychol Psychiatry* 2010;**51**:1331–41. http://dx.doi.org/10.1111/j.1469-7610.2010.02302.x

60. Webster-Stratton C. Affirming diversity: multi-cultural collaboration to deliver the incredible years parent programs. *Int J Child Health Hum Dev* 2009;**2**:17–32.

61. Patel A, Calam R, Latham A. Intention to attend parenting programmes: does ethnicity make a difference? *J Child Serv* 2011;**6**:45–58.

62. Allen G. *Early Intervention: The Next Steps. Independent Report to Her Majesty's Government*. London: HM Government; 2011.

63. Heckman JJ. Skill formation and the economics of investing in disadvantaged children. *Science* 2006;**312**:1900–2. https://doi.org/10.1126/science.1128898

64. Serketich WJ, Dumas JE. The effectiveness of behavioral parent training to modify antisocial behavior in children: a meta-analysis. *Behav Ther* 1996;**27**:171–86.

65. Weisz JR, Weiss B, Han SS, Granger DA, Morton T. Effects of psychotherapy with children and adolescents revisited: a meta-analysis of treatment outcome studies. *Psychol Bull* 1995;**117**:450–68. https://doi.org/10.1037/0033-2909.117.3.450

66. Beauchaine TP, Webster-Stratton C, Reid MJ. Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: a latent growth curve analysis. *J Consult Clin Psychol* 2005;**73**:371–88. https://doi.org/10.1037/0022-006X.73.3.371

67. Bierman K. Predictor variables associated with positive fast track outcomes at the end of third grade. *J Abnorm Child Psychol* 2002;**30**:37–52.

68. Jones K, Daley D, Hutchings J, Bywater T, Eames C. Efficacy of the Incredible Years Programme as an early intervention for children with conduct problems and ADHD: long-term follow-up. *Child Care Health Dev* 2008;**34**:380–90. http://dx.doi.org/10.1111/j.1365-2214.2008.00817.x

69. Webster-Stratton CH, Reid MJ, Beauchaine T. Combining parent and child training for young children with ADHD. *J Clin Child Adolesc Psychol* 2011;**40**:191–203. http://dx.doi.org/10.1080/15374416.2011.546044

70. Goodman SH, Rouse MH, Connell AM, Broth MR, Hall CM, Heyward D. Maternal depression and child psychopathology: a meta-analytic review. *Clin Child Fam Psychol Rev* 2011;**14**:1–27. http://dx.doi.org/10.1007/s10567-010-0080-1

71. Hutchings J, Bywater T, Williams M, Lane E, Whitaker CJ. Improvements in maternal depression as a mediator of child behaviour change. *Psychology* 2012;**3**:795–801. https://doi.org/10.4236/psych.2012.329120

72. Wahler RG, Sansbury LE. The monitoring skills of troubled mothers: their problems in defining child deviance. *J Abnorm Child Psychol* 1990;**18**:577–89. https://doi.org/10.1007/BF00911109

73. Hutchings J, Nash S, Williams JM, Nightingale D. Parental autobiographical memory: is this a helpful clinical measure in behavioural child management? *Br J Clin Psychol* 1998;**37**:303–12. https://doi.org/10.1111/j.2044-8260.1998.tb01387.x

74. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;**303**:2058–64. http://dx.doi.org/10.1001/jama.2010.651

75. Hutchings J, Bywater T, Daley D. Early prevention of conduct disorder: how and why did the North and Mid Wales Sure Start study work? *J Child Serv* 2007;**2**:4–14.

76. Scott S. Nationwide dissemination of effective parenting interventions: building a parenting academy for England. In Weisz J, Kazdin A, editors. *Evidence-Based Psychotherapies for Children and Adolescents*. New York, NY: Guilford; 2010. pp. 500–18.

77. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002;**55**:86–94. https://doi.org/10.1016/S0895-4356(01)00414-0

78. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health* 1994;**84**:819–24. https://doi.org/10.2105/AJPH.84.5.819

79. Lipsey MW. Those confounded moderators in meta-analysis: good, bad, and ugly. *Ann Am Acad Pol Soc Sci* 2003;**587**:69–81.

80. Brown CH, Sloboda Z, Faggiano F, Teasdale B, Keller F, Burkhart G, *et al.* Methods for synthesizing findings on moderation effects across multiple randomized trials. *Prev Sci* 2013;**14**:144–56. http://dx.doi.org/10.1007/s11121-011-0207-8

81. Shadish WR, Sweeney RB. Mediators and moderators in meta-analysis: there's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *J Consult Clin Psychol* 1991;**59**:883–93. https://doi.org/10.1037/0022-006X.59.6.883

82. Petticrew M, Tugwell P, Kristjansson E, Oliver S, Ueffing E, Welch V. Damned if you do, damned if you don't: subgroup analysis and equity. *J Epidemiol Community Health* 2012;**66**:95–8. http://dx.doi.org/10.1136/jech.2010.121095

83. Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychol Methods* 2009;**14**:165–76. http://dx.doi.org/10.1037/a0015565

84. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;**340**:c221. http://dx.doi.org/10.1136/bmj.c221

85. Hutchings J, Gardner F, Bywater T, Daley D, Whitaker C, Jones K, *et al.* Parenting intervention in Sure Start services for children at risk of developing conduct disorder: pragmatic randomised controlled trial. *BMJ* 2007;**334**:678. https://doi.org/10.1136/bmj.39126.620799.55

86. Chhangur RR, Weeland J, Overbeek G, Matthys W, Orobio de Castro B. ORCHIDS: an observational randomized controlled trial on childhood differential susceptibility. *BMC Public Health* 2012;**12**:917. http://dx.doi.org/10.1186/1471-2458-12-917

87. Prinz RJ, Sanders MR, Shapiro CJ, Whitaker DJ, Lutzker JR. Population-based prevention of child maltreatment: the U.S. Triple P system population trial. *Prev Sci* 2009;**10**:1–12. https://doi.org/10.1007/s11121-009-0123-3

88. Webster-Stratton C, Reid M. Adapting The Incredible Years, an evidence-based parenting programme, for families involved in the child welfare system. *J Child Serv* 2010;**5**:25–42.

89. Barlow J, Johnston I, Kendrick D, Polnay L, Stewart-Brown S. Individual and group-based parenting programmes for the treatment of physical child abuse and neglect. *Cochrane Database Syst Rev* 2006;**3**:CD005463. http://dx.doi.org/10.1002/14651858.CD005463.pub2

90. Scott S. Parenting quality and children's mental health: biological mechanisms and psychological interventions. *Curr Opin Psychiatry* 2012;**25**:301–6. http://dx.doi.org/10.1097/YCO.0b013e328354a1c5

91. Thapar A, Harrington R, McGuffin P. Examining the comorbidity of ADHD-related behaviours and conduct problems using a twin study design. *Br J Psychiatry* 2001;**179**:224–9. https://doi.org/10.1192/bjp.179.3.224

92. Daley D, van der Oord S, Ferrin M, Danckaerts M, Doepfner M, Cortese S, Sonuga-Barke EJ, European ADHD Guidelines Group. Behavioral interventions in attention-deficit/hyperactivity disorder: a meta-analysis of randomized controlled trials across multiple outcome domains. *J Am Acad Child Adolesc Psychiatry* 2014;**53**:835–47, 847.e1–5. http://dx.doi.org/10.1016/j.jaac.2014.05.013

93. Sonuga-Barke EJ, Brandeis D, Cortese S, Daley D, Ferrin M, Holtmann M, *et al.* Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *Am J Psychiatry* 2013;**170**:275–89. http://dx.doi.org/10.1176/appi.ajp.2012.12070991

94. Levy F, Hawes D, Johns A. Externalizing and Internalizing Comorbidity. In Beauchaine T, Hinshaw S, editors. *The Oxford Handbook of Externalizing Spectrum Disorders*. Oxford: Oxford University Press; 2015. pp. 443–60.

95. Herman KC, Borden LA, Reinke WM, Webster-Stratton C. The Impact of the Incredible Years Parent, Child, and Teacher Training Programs on children's co-occurring internalizing symptoms. *Sch Psychol Q* 2011;**26**:189–201. http://dx.doi.org/10.1037/a0025228

96. Leijten P, Raaijmakers MA, Orobio de Castro B, van den Ban E, Matthys W. Effectiveness of the Incredible Years Parenting Program for families with socioeconomically disadvantaged and ethnic minority backgrounds. *J Clin Child Adolesc Psychol* 2017;**46**:59–73. http://dx.doi.org/10.1080/15374416.2015.1038823

97. Dishion TJ, McCord J, Poulin F. When interventions harm. Peer groups and problem behavior. *Am Psychol* 1999;**54**:755–64. https://doi.org/10.1037/0003-066X.54.9.755

98. Petrosino A, Turpin-Petrosino C, Hollis-Peel ME, Lavenberg JG. 'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database Syst Rev* 2013;**4**:CD002796. http://dx.doi.org/10.1002/14651858.CD002796.pub2

99. Dretzke J, Frew E, Davenport C, Barlow J, Stewart-Brown S, Sandercock J, *et al.* The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder, in children. *Health Technol Assess* 2005;**9**(50). https://doi.org/10.3310/hta9500

100. Scott S, Sylva K, Doolan M, Price J, Jacobs B, Crook C, Landau S. Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: the SPOKES project. *J Child Psychol Psychiatry* 2010;**51**:48–57. http://dx.doi.org/10.1111/j.1469-7610.2009.02127.x

101. Scott S, Sylva K, Kallitsoglou A, Ford T. *Which Type of Parenting Programme Best Improves Child Behaviour and Reading? Follow-Up of the Helping Children Achieve Trial*. London: Nuffield Foundation; 2014.

102. Patterson J, Barlow J, Mockford C, Klimes I, Pyper C, Stewart-Brown S. Improving mental health through parenting programmes: block randomised controlled trial. *Arch Dis Child* 2002;**87**:472–7. https://doi.org/10.1136/adc.87.6.472

103. Gardner F, Burton J, Klimes I. Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. *J Child Psychol Psychiatry* 2006;**47**:1123–32. https://doi.org/10.1111/j.1469-7610.2006.01668.x

104. Hutchings J, Griffith N, Bywater T, Williams ME. Evaluating the Incredible Years Toddler Parenting Programme with parents of toddlers in disadvantaged (Flying Start) areas of Wales. *Child Care Health Dev* 2017;**43**:104–13. http://dx.doi.org/10.1111/cch.12415

105. Morpeth L, Blower S, Tobin K, Taylor RS, Bywater T, Edwards RT, *et al.* The effectiveness of the Incredible Years pre-school parenting programme in the United Kingdom: a pragmatic randomised controlled trial. *Child Care Pract* 2017;**25**:1–21. http://dx.doi.org/10.1080/13575279.2016.1264366

106. Hutchings J, Gardner F. Support from the start: effective programmes for three- to eight-year-olds. *J Child Serv* 2012;**7**:29–40.

107. Webster-Stratton C. *The Incredible Years: Parents, Teachers and Children's Training Series*. Seattle, WA: Incredible Years Inc.; 2011.

108. Patterson GR. *Coercive Family Process*. Eugene, OR: Castalia; 1982.

109. Stewart L, Tierney J, Clarke M. Reviews of Individual Patient Data. In Higgins J, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley; 2011. pp. 547–59.

110. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* 2005;**365**:341–6. https://doi.org/10.1016/S0140-6736(05)70200-2

111. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, *et al.* Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLOS ONE* 2008;**3**:e3081. http://dx.doi.org/10.1371/journal.pone.0003081

112. Sterne J, Egger M, Moher D. Addressing Reporting Biases. In Higgins J, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley; 2008. pp. 297–334. https://doi.org/10.1002/9780470712184.ch10

113. Larsson B, Fossum S, Clifford G, Drugli MB, Handegård BH, Mørch WT. Treatment of oppositional defiant and conduct problems in young Norwegian children: results of a randomized controlled trial. *Eur Child Adolesc Psychiatry* 2009;**18**:42–52. http://dx.doi.org/10.1007/s00787-008-0702-z

114. Axberg U, Broberg AG. Evaluation of 'the incredible years' in Sweden: the transferability of an American parent-training program to Sweden. *Scand J Psychol* 2012;**53**:224–32. http://dx.doi.org/10.1111/j.1467-9450.2012.00955.x

115. Homem TC, Gaspar MF, Santos MJS, Azevedo AF, Canavarro MC. Incredible Years parent training: does it improve positive relationships in Portuguese families of preschoolers with oppositional/defiant symptoms? *J Child Fam Stud* 2014;**24**:1861–75.

116. Azevedo AF, Seabra-Santos MJ, Gaspar MF, Homem TC. The Incredible Years Basic parent training for Portuguese preschoolers with AD/HD Behaviors: does it make a difference? *Child Youth Care Forum* 2013;**42**:403–24.

117. Azevedo AF, Seabra-Santos MJ, Gaspar MF, Homem T. A parent-based intervention programme involving preschoolers with ad/hd behaviours: are children's and mothers' effects sustained over time? *Eur Child Adolesc Psychiatry* 2013;**23**:437–50.

118. Menting AT, de Castro BO, Wijngaards-de Meij LD, Matthys W. A trial of parent training for mothers being released from incarceration and their children. *J Clin Child Adolesc Psychol* 2014;**43**:381–96. http://dx.doi.org/10.1080/15374416.2013.817310

119. Little M, Berry V, Morpeth L, Blower S, Axford N, Taylor R, *et al.* The impact of three evidence-based programmes delivered in public systems in Birmingham, UK. *Int J Conf Violence* 2012;**6**:260–72.

120. Scott S, Spender Q, Doolan M, Jacobs B, Aspland H. Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *BMJ* 2001;**323**:194–8. https://doi.org/10.1136/bmj.323.7306.194

121. Higgins J, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011). London: Cochrane; 2011.

122. United Nations Educational, Scientific and Cultural Organisation (UNESCO). *International Standard Classification of Education 2011*. Montreal, QB: UNESCO Institute for Statistics; 2011.

123. Office of National Statistics (ONS). *Ethnic Group Statistics: A Guide for the Collection and Classification of Ethnicity Data*. London: ONS; 2003. URL: www.ons.gov.uk/ons/guide-method/measuring-equality/equality/ethnic-group-statistics/ethnic-group-statistics--a-guide-for-the-collection-and-classification-of-ethnicity-data.pdf (accessed October 2014).

124. Robinson EA, Eyberg SM, Ross AW. The standardization of an inventory of child conduct problem behaviors. *J Clin Child Psychol* 1980;**9**:22–8.

125. Achenbach T. *The Direct Observation Form of the Child Behavior Checklist*. Rev. edn. Burlington, VT: University of Vermont, Department of Psychiatry; 1986.

126. Baden AD, Howe GW. Mothers' attributions and expectancies regarding their conduct-disordered children. *J Abnorm Child Psychol* 1992;**20**:467–85. https://doi.org/10.1007/BF00916810

127. Eyberg SM, Ross AW. Assessment of child behavior problems: the validation of a new inventory. *J Clin Child Psychol* 1978;**7**:113–16.

128. Taylor E, Schachar R, Thorley G, Wieselberg M. Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. *Br J Psychiatry* 1986;**149**:760–7. https://doi.org/10.1192/bjp.149.6.760

129. Achenbach T. *Manual for Child Behavior Checklist/ 4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.

130. Taylor E, Sandberg S, Thorley G. *The Epidemiology of Childhood Hyperactivity*. Vol. 33. Oxford: Oxford University Press; 1991.

131. Centers for Disease Control and Prevention (CDC). *National Health Interview Survey 2001 Public Use Data Release: NHIS Survey Description (December 2002)*. National Center for Health Statistics; 2002. URL: www.cdc.gov/nchs/nhis.htm (accessed 12 November 2015).

132. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;**4**:561–71.

133. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev* 1988;**8**:77–100.

134. Derogatis L, Spencer M. *The Brief Symptom Inventory (BSI): Administration, Scoring, and Procedures Manual-1*. Baltimore, MD: Johns Hopkins University School of Medicine, Clinical Psychometrics Research Unit; 1982.

135. Francis VM, Rajan P, Turner N. British community norms for the Brief Symptom Inventory. *Br J Clin Psychol* 1990;**29**:115–16. https://doi.org/10.1111/j.2044-8260.1990.tb00857.x

136. Arrindell W, Ettema J. [*SCL-90: Manual to a Multidimensional Psychopathology-indicator*.] 2nd edn. Amsterdam: Pearson; 2003.

137. Golderberg D, Williams P. *A User's Guide to the General Health Questionnaire*. Windsor: NFER-Nelson; 1988.

138. Booker C, Sacker A. Health over the Life Course: Associations Between Age, Employment Status and Well-being. In McFall S, Garrington C, editors. *Understanding Society: Early Findings from the First Wave of the UK's Household Longitudinal Study*. Colchester: Institute for Social and Economic Research; 2011. pp. 75–86.

139. Lovibond S, Lovibond P. *Manual for the Depression Anxiety Stress Scales*. 2nd edn. Sydney, NSW: Psychology Foundation; 1995.

140. Crawford JR, Henry JD. The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. *Br J Clin Psychol* 2003;**42**:111–31. http://dx.doi.org/10.1348/014466503321903544

141. Abidin R. *Parenting Stress Index: Manual, Administration Booklet, [and] Research Update*. Charlottesville, VA: Pediatric Psychology Press; 1983. p. 86.

142. Johnston C, Mash EJ. A measure of parenting satisfaction and efficacy. *J Clin Child Psychol* 1989;**18**:167–75.

143. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;**30**:377–99. http://dx.doi.org/10.1002/sim.4067

144. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Stat Med* 2004;**23**:2509–25. http://dx.doi.org/10.1002/sim.1815

145. Mistler S. *A SAS Macro for Applying Multiple Imputation to Multilevel Data. The SAS Global Forum: 2013*. SAS Global Forum, San Francisco, CA, 28 April–1 May 2013.

146. Van Buuren S. Multiple Imputation of Multilevel Data. In Hox J, Roberts J, editors. *Handbook of Advanced Multilevel Analysis*. Oxford: Routledge; 2011. pp. 173–96.

147. Reiter JP, Raghunathan TE, Kinney SK. The importance of modelling the sampling design in multiple imputation for missing data. *Surv Methodol* 2006;**32**:143–9.

148. Drechsler J. Multiple Imputation of multilevel missing data – rigor versus simplicity. *J Educ Behav Stat* 2015;**40**:69–95.

149. Leijten P, Raaijmakers MA, de Castro BO, Matthys W. Does socioeconomic status matter? A meta-analysis on parent training effectiveness for disruptive child behavior. *J Clin Child Adolesc Psychol* 2013;**42**:384–92. http://dx.doi.org/10.1080/15374416.2013.769169

150. Furlong M, McGilloway S. Barriers and facilitators to implementing evidence-based parenting programs in disadvantaged settings: a qualitative study. *J Child Fam Stud* 2014;**24**:1809–18.

151. Dishion TJ, Shaw D, Connell A, Gardner F, Weaver C, Wilson M. The family check-up with high-risk indigent families: preventing problem behavior by increasing parents' positive behavior support in early childhood. *Child Dev* 2008;**79**:1395–414. http://dx.doi.org/10.1111/j.1467-8624.2008.01195.x

152. Yasui M, Dishion TJ. The ethnic context of child and adolescent problem behavior: implications for child and family interventions. *Clin Child Fam Psychol Rev* 2007;**10**:137–79. http://dx.doi.org/10.1007/s10567-007-0021-9

153. Gottfredson D, Kumpfer K, Polizzi-Fox D, Wilson D, Puryear V, Beatty P, Vilmenay M. The Strengthening Washington D.C. Families project: a randomized effectiveness trial of family-based prevention. *Prev Sci* 2006;**7**:57–74. http://dx.doi.org/10.1007/s11121-005-0017-y

154. Nock MK, Kazdin AE. Randomized controlled trial of a brief intervention for increasing participation in parent management training. *J Consult Clin Psychol* 2005;**73**:872–9. https://doi.org/10.1037/0022-006X.73.5.872

155. Ollendick TH, Jarrett MA, Grills-Taquechel AE, Hovey LD, Wolff JC. Comorbidity as a predictor and moderator of treatment outcome in youth with anxiety, affective, attention deficit/hyperactivity disorder, and oppositional/conduct disorders. *Clin Psychol Rev* 2008;**28**:1447–71. http://dx.doi.org/10.1016/j.cpr.2008.09.003

156. Lovejoy MC, Graczyk PA, O'Hare E, Neuman G. Maternal depression and parenting behavior: a meta-analytic review. *Clin Psychol Rev* 2000;**20**:561–92. https://doi.org/10.1016/S0272-7358(98)00100-7

157. Morgan-Lopez AA, MacKinnon DP. Demonstration and evaluation of a method for assessing mediated moderation. *Behav Res Methods* 2006;**38**:77–87. https://doi.org/10.3758/BF03192752

158. Leijten P, Melendez-Torres GJ, Knerr W, Gardner F. Transported Versus homegrown parenting interventions for reducing disruptive child behavior: a multilevel meta-regression study. *J Am Acad Child Adolesc Psychiatry* 2016;**55**:610–17. http://dx.doi.org/10.1016/j.jaac.2016.05.003

159. Shaw DS, Connell A, Dishion TJ, Wilson MN, Gardner F. Improvements in maternal depression as a mediator of intervention effects on early childhood problem behavior. *Dev Psychopathol* 2009;**21**:417–39. http://dx.doi.org/10.1017/S0954579409000236

160. Posthumus JA, Raaijmakers MA, Maassen GH, van Engeland H, Matthys W. Sustained effects of incredible years as a preventive intervention in preschool children with conduct problems. *J Abnorm Child Psychol* 2012;**40**:487–500. http://dx.doi.org/10.1007/s10802-011-9580-9

161. DiClemente RJ, Wingood GM, Crosby R, Sionean C, Cobb BK, Harrington K, *et al.* Parental monitoring: association with adolescents' risk behaviors. *Pediatrics* 2001;**107**:1363–8. https://doi.org/10.1542/peds.107.6.1363

162. Dishion TJ, Nelson SE, Kavanagh K. The family check-up with high-risk young adolescents: preventing early-onset substance use by parent monitoring. *Behav Ther* 2003;**34**:553–71.

163. Romeo R, Knapp M, Scott S. Economic cost of severe antisocial behaviour in children – and who pays it. *Br J Psychiatry* 2006;**188**:547–53. https://doi.org/10.1192/bjp.bp.104.007625

164. Snell T, Knapp M, Healey A, Guglani S, Evans-Lacko S, Fernandez JL, *et al.* Economic impact of childhood psychiatric disorder on public sector services in Britain: estimates from national survey data. *J Child Psychol Psychiatry* 2013;**54**:977–85. http://dx.doi.org/10.1111/jcpp.12055

165. Knapp M, Snell T, Healey A, Guglani S, Evans-Lacko S, Fernandez JL, *et al.* How do child and adolescent mental health problems influence public sector costs? Inter-individual variations in a nationally representative British sample. *J Child Psychol Psychiatry* 2015;**56**:667–76.

166. D'Amico F, Knapp M, Beecham J, Sandberg S, Taylor E, Sayal K. Use of services and associated costs for young adults with childhood hyperactivity/conduct problems: 20-year follow-up. *Br J Psychiatry* 2014;**204**:441–7. http://dx.doi.org/10.1192/bjp.bp.113.131367

167. Puig-Peiro R, Stevens M, Beecham J. *The Costs and Characteristics of the Parenting Programmes in the NAPP Commissioners' Toolkit*. London: Personal Social Services Research Unit, the London School of Economics and Political Science; 2010.

168. Stevens M. The cost-effectiveness of UK parenting programmes for preventing children's behaviour problems – a review of the evidence. *Child Fam Soc Work* 2014;**19**:109–18.

169. Edwards RT, Céilleachair A, Bywater T, Hughes DA, Hutchings J. Parenting programme for parents of children at risk of developing conduct disorder: cost effectiveness analysis. *BMJ* 2007;**334**:682. https://doi.org/10.1136/bmj.39126.699421.55

170. Muntz R, Hutchings J, Edwards RT, Hounsome B, O'Céilleachair A. Economic evaluation of treatments for children with severe behavioural problems. *J Ment Health Policy Econ* 2004;**7**:177–89.

171. O'Neill D, McGilloway S, Donnelly M, Bywater T, Kelly P. A cost-effectiveness analysis of the Incredible Years parenting programme in reducing childhood health inequalities. *Eur J Health Econ* 2013;**14**:85–94. http://dx.doi.org/10.1007/s10198-011-0342-y

172. Bywater T, Hutchings J, Daley D, Whitaker C, Yeo ST, Jones K, *et al.* Long-term effectiveness of a parenting intervention for children at risk of developing conduct disorder. *Br J Psychiatry* 2009;**195**:318–24. http://dx.doi.org/10.1192/bjp.bp.108.056531

173. McGilloway S, NiMhaille G, Bywater T, Leckey Y, Kelly P, Furlong M, *et al.* Reducing child conduct disordered behaviour and improving parent mental health in disadvantaged families: a 12-month follow-up and cost analysis of a parenting intervention. *Eur Child Adolesc Psychiatry* 2014;**23**:783–94. http://dx.doi.org/10.1007/s00787-013-0499-2

174. Heckman JJ, Moon SH, Pinto R, Savelyev PA, Yavitz A. The rate of return to the High/Scope Perry Preschool Program. *J Public Econ* 2010;**94**:114–28. https://doi.org/10.1016/j.jpubeco.2009.11.001

175. Aos S, Lieb R, Mayfield J, Miller M, Pennucci A. *Benefits and Costs of Prevention and Early Intervention Programs for Youth*. Olympia, WA: Washington State Institute for Public Policy; 2004, 04-07-3901.

176. Curtis L. *Unit Costs of Health and Social Care 2014*. Canterbury: Personal Social Services Research Unit, University of Kent; 2014.

177. Beecham J, Knapp M. Costing Psychiatric Interventions. In Thornicroft G, editor. *Measuring Mental Health Needs*. 2nd edn. London: Royal College of Psychiatrists; 2001.

178. Department of Health (DH). *NHS Reference Costs 2013–14*. London: DH; 2014.

179. Beecham J. *Unit Costs - Not Exactly Child's Play. A Guide to Estimating Unit Costs for Children's Social Care*. Dartington: DH, Dartington Social Research Unit and Personal Social Services Research Unit; 2000.

180. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall; 1993. https://doi.org/10.1007/978-1-4899-4541-9

181. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ* 1994;**3**:309–19. https://doi.org/10.1002/hec.4730030505

182. Richardson J, Joughin C. *Parent Training Programmes for the Management of Young Children with Conduct Disorders: Findings from Research*. 2002.

183. Department of Health (DH). *Achieving Better Access to Mental Health Services by 2020*. London: DH; 2014.

184. Beecham JK, Green J, Jacobs B, Dunn G. Cost variation in child and adolescent psychiatric inpatient treatment. *Eur Child Adolesc Psychiatry* 2009;**18**:535–42. http://dx.doi.org/10.1007/s00787-009-0008-9

185. Green H, McGinnity A, Meltzer H, Ford T, Goodman R. *Mental Health of Children and Young People in Great Britain 2004*. Basingstoke: Palgrave Macmillan; 2005. https://doi.org/10.1037/e557702010-001

186. Knapp M, Mangalore R. 'The trouble with QALYs . . .'. *Epidemiol Psichiatr Soc* 2007;**16**:289–93. https://doi.org/10.1017/S1121189X00002451

187. Knapp M, King D, Healey A, Thomas C. Economic outcomes in adulthood and their associations with antisocial conduct, attention deficit and anxiety problems in childhood. *J Ment Health Policy Econ* 2011;**14**:137–47.

188. Goldstein H. *Multilevel Statistical Models*. 4th edn. Oxford: Wiley; 2011.

189. Gardner F. Methodological issues in the direct observation of parent-child interaction: do observational findings reflect the natural behavior of participants? *Clin Child Fam Psychol Rev* 2000;**3**:185–98. https://doi.org/10.1023/A:1009503409699

190. Le Couteur A, Gardner F. Structured Interviews and Observations. In Rutter M, Bishop D, Pine D, Scott S, Stevenson J, Taylor E, *et al.*, editors. *Rutter's Child and Adolescent Psychiatry*. 5th edn. Oxford: Blackwell; 2008. https://doi.org/10.1002/9781444300895.ch19

191. van Aar J, Leijten P, de Castro BO, Overbeek G. Sustained, Fade-out or Sleeper Effects? A Multilevel Meta-Analysis of Parenting Interventions for Disruptive Child Behavior. *Clin Psychol Rev* 2016;**51**:153–63. https://doi.org/10.1016/j.cpr.2016.11.006

192. Montgomery P, Grant S, Hopewell S, Macdonald G, Moher D, Michie S, Mayo-Wilson E. Protocol for CONSORT-SPI: an extension for social and psychological interventions. *Implement Sci* 2013;**8**:99. http://dx.doi.org/10.1186/1748-5908-8-99

193. Scott S, O'Connor TG. An experimental test of differential susceptibility to parenting among emotionally-dysregulated children in a randomized controlled trial for oppositional behavior. *J Child Psychol Psychiatry* 2012;**53**:1184–93.

194. Forehand R, Lafko N, Parent J, Burt KB. Is parenting the mediator of change in behavioral parent training for externalizing problems of youth? *Clin Psychol Rev* 2014;**34**:608–19. http://dx.doi.org/10.1016/j.cpr.2014.10.001

195. Scott S. Conduct Disorders. In Thapar A, Pine D, Leckman J, Scott S, Snowling M, Taylor E, editors. *Rutter's Child and Adolescent's Psychiatry*. 6th edn. Oxford: Wiley Blackwell; 2015. https://doi.org/10.1002/9781118381953.ch65

196. Piotrowska PJ, Stride CB, Maughan B, Goodman R, McCaw L, Rowe R. Income gradients within child and adolescent antisocial behaviours. *Br J Psychiatry* 2015;**207**:385–91. http://dx.doi.org/10.1192/bjp.bp.113.143636

197. Scott S, Briskman J, O'Connor TG. Early prevention of antisocial personality: long-term follow-up of two randomized controlled trials comparing indicated and selective approaches. *Am J Psychiatry* 2014;**171**:649–57. http://dx.doi.org/10.1176/appi.ajp.2014.13050697

198. Department of Health (DH). *Future in Mind: Promoting, Protecting and Improving Our Children and Young People's Mental Health and Wellbeing*. London: DH; 2015.

199. Simkiss DE, Snooks HA, Stallard N, Kimani PK, Sewell B, Fitzsimmons D, *et al.* Effectiveness and cost-effectiveness of a universal parenting skills programme in deprived communities: multicentre randomised controlled trial. *BMJ Open* 2013;**3**:e002851. http://dx.doi.org/10.1136/bmjopen-2013-002851

200. Ng MY, Weisz JR. Annual Research Review: Building a science of personalized intervention for youth mental health. *J Child Psychol Psychiatry* 2016;**57**:216–36. https://doi.org/10.1111/jcpp.12470

201. National Institute for Health and Care Excellence (NICE). *Attention Deficit Hyperactivity Disorder*. Evidence Update 45. London: NICE; 2013.

202. Webster-Stratton. *Incredible Years® Parent Program Satisfaction Questionnaire BASIC Parent Program*. Seattle, WA: The Incredible Years; 2013.

203. Brestan EV, Jacobs JR, Rayfield AD, Eyberg SM. A consumer satisfaction measure for parent-child treatments and its relation to measures of child behavior change. *Behav Ther* 1999;**30**:17–30.

204. Gaheer S, Paull G. *Evaluation of Children's Centres in England (ECCE) Strand 6*. London: Department for Education; 2016.

205. Curtis L, Netten A. *Unit Costs of Health and Social Care 2004*. Canterbury: University of Kent, PSSRU; 2005.

206. Rutter J. *Childcare Costs Survey 2015*. London: Family and Childcare Trust; 2016.

Appendix 1 Data sharing agreement

14 January 2014 data sharing agreement

This agreement is drawn up between:

(name of participating department and university – source of trial data)

and

Department of Social Policy and Intervention, Oxford University

and

Institute of Psychiatry, King's College London

and

School of Psychology, Bangor University

and

Personal Social Services Research Unit, London School of Economics

This agreement commences in January 2014 and continues for the duration of the project and 1 year following the completion of the project – until October 2016. After this point there will be further discussions with all collaborators to determine if and how the data should be used in the future.

Principles of collaboration

The standard principles underlying this plan include respecting the original investigators' ownership of their data and the responsibility of all parties to ensure that all of the data are anonymised and kept confidential.

A pooled data set will be created to be used for the questions set out in the Public Health Research project proposal, evaluating the cost and effectiveness and moderators of the IY parenting intervention and related questions exploring parenting, child outcomes, service delivery, geographical factors or other family factors. All analyses and publications must contain aggregate data, no information will be included on individual participants and data will not be analysed separately at an individual centre or site level.

One investigator from each data site will be included as a coauthor on each of the two main papers from the pooled data set. We would expect that coauthors make a contribution to the writing and approval of the paper, in line with standard guidelines on authorship. All principal investigators will be acknowledged in all further papers.

Data ownership and use

The data from each trial which are being used for this study will be owned by the original data sites. The Pooling Study project team have the right to use these data for the purposes of this study, as part of the pooled data set.

Data being shared

Fully anonymised data from individual participants who consented for their data to be stored and used for research purposes, for the purpose of evaluating the effectiveness of the IY parenting intervention.

Specific conditions

The data must be fully anonymised and used solely for the purposes outlined above. It must not be shared with any organisation or individual, other than those named in this agreement.

Obligations

By signing this agreement, the parties are acknowledging guardianship of and an explicit agreement to maintain the confidentiality, security, safety and integrity of the data shared under the remit of this agreement regardless of whether they are confidential or sensitive.

Data storage and transfer

All stored data will be protected by user identifier and strong password.

All data will be transferred by the University of Oxford's web service 'Oxfile'. Data transferred via this web service are encrypted.

All data storage and transfer throughout the study will be in accordance with the University of Oxford's information security policy which can be found at: www.it.ox.ac.uk/infosec/ispolicy/

Breach of agreement

All users of the data will notify the originator of the data in case of any breach of the terms of this agreement.

At the appropriate time, we will provide Frances Gardner at the University of Oxford with access to these data sets to conduct analyses and the right to combine these data with other trial data to perform pooled data analysis. We will not unreasonably withhold access to the data sets.

Signatures

List of permitted users – January 2014
Professor Stephen Scott (psychiatrist, coinvestigator, King's College London)

Professor Frances Gardner (psychologist, coinvestigator, University of Oxford)

Dr Sabine Landau (statistician, coinvestigator, King's College London)

Professor Andrew Pickles (statistician, coinvestigator, King's College London)

Professor Jennifer Beecham (economist, coinvestigator, London School of Economics and the University of Kent)

Professor Judy Hutchings (psychologist, Bangor University)

Ms Angela Zhang (Doctor of Philosophy student working on the London data sets with Scott and Landau)

Dr Patty Leijten (research officer, for this data pooling study, University of Oxford)

Dr Joanna Mann (research assistant, for this data pooling study, University of Oxford)

Statistician working with Dr Landau, to be arranged after October 2014

Researcher working with Professor Beecham, to be arranged after October 2014

Appendix 2 Further details of the trials

List (or table) of included trials and references

Design features of the trials

Table 26 shows a list of the trials included in the pooled data set and their design features. Across the trials some offered the IY intervention only in the treated group, whereas others offered a literacy intervention alongside IY. Similarly, the control condition differs across trials, with some offering waiting list or treatment as usual and others offering minimal treatment. As this study is focused upon the IY training programme, participants randomised to a literacy-only intervention were excluded from the pooled data set, as well as participants randomised to the IY parent and IY child programme. *Table 1* shows the country where the trial was carried out. The column labelled *n* gives the total sample size from the trial that was used in the IY pooling data set. The column control group details the type of control condition and the column arms used details the arms included in the pooled sample. The trial may have included arms that have not been used in the pooled sample (e.g. reading intervention only or IY child programme), but these are excluded from *Table 1* if they were not included in the pooling study. 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 than those normally accessible to the patient during their daily life (in particular in NL-BS, in which mothers were recently released from incarceration; they may not have had 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, for example a telephone helpline. Most trials used a waiting list design, in which parents in the waiting list control condition were crossed over to the intervention after 6 months but these families were no longer observed after crossover. A full list of the trials included in the pooled sample and corresponding references is given in *Chapter 5*, *Methods*.

Trials included are:

1. NOR: a clinic-based treatment trial in Norway, with children with oppositional defiant and conduct problems (4–8 years).[113]
2. SWED: a clinic-based treatment trial in Sweden, with children with oppositional defiant and conduct problems (4–8 years).[114]
3. PORT: a clinic-based treatment trial, in Portugal, in a university-based clinic, with middle-class families. This trial took place with children with conduct problems (3–6 years).[115,116]
4. IRE: a community-based trial in Ireland, with disadvantaged communities and children with conduct problems (2–7 years).[48]
5. NL-BS: a community-based trial in the Netherlands, involving mothers who were recently released from incarceration. Children were not screened for conduct problems (aged 2–10 years).[118]
6. NL-SES: a community-based trial in the Netherlands, with families of mainly low SES and immigrant families. Children had conduct problems or relational problems (3–8 years).[96]
7. WL-SS: a community-based trial in Wales, in Sure Start areas, with disadvantaged families. Children were showing early signs of conduct problems (3–4 years old).[85]
8. WL-FS: a community-based trial of the IY toddler parenting programme in Wales, in Flying Start areas, with disadvantaged families. Children were not screened for conduct problems (12–36 months).[104]
9. BIRM: a community-based trial in England. Children were showing early signs of conduct problems (3–5 years).[105,119]
10. LON-SPO: a community-based trial in England, investigating IY and a literacy intervention in low-income primary schools, with children with conduct problems (1–6 years).[100]
11. LON-PAL: a community-based trial in England, investigating IY and a literacy intervention, in low-income primary schools, with children from disadvantaged families. Children were at high risk of social exclusion (2–7 years).[59]

TABLE 26 Further features of trial design

Trial number	Trial acronym	Duration of randomisation to first assessment (months)	Duration of end of intervention to second assessment (months)	Duration of end of intervention to third assessment (months)	Randomisation unit	Stratified randomisation	Stratifiers used in randomisation	Randomisation ratio by trial design	Variable randomisation ratio (i.e. whether or not ratio changed during trial)
1	NOR	1	0–2	N/A	Individual	Yes	Child age, child gender, child scored > 97th percentile on ECBI-I, site	1 : 1	Yes
2	SWED	1	0–2	N/A	Individual	Yes	Site	2 : 1	Yes
3	PORT	1	0–2	N/A	Individual	Yes	Child age, child gender	1 : 1	Yes
4	IRE	1	0–2	N/A	Individual	Yes	Site	2 : 1	No
5	NL-BS	1	0–2	4	Individual	No	None	2 : 1	Yes
6	NL-SES	1	0–2	N/A	Individual	No	None	2 : 1	No
7	WL-SS	1	0–2	N/A	Individual	Yes	Child age, child gender, site	2 : 1	No
8	WL-FS	1	0–2	N/A	Individual	Yes	Child age, child gender, site	2 : 1	No
9	BIRM	1	0–2	N/A	Individual	Yes	Child age, child gender, children's centre attachment	2 : 1	No
10	LON-SPO	1	0–2	N/A	Individual	Yes	School year (10 strata formed from eight schools over 3 years)	1 : 1	No
11	LON-PAL	1	0–2	N/A	Cluster – classrooms	No	None	1 : 1	Yes
12	LON-HCA	1	0–2	5–7	Individual	Yes	Recruitment cohort	1 : 1 : 1	Yes
13	OXF	1	0–2	N/A	Individual	No	None	1 : 1	Yes
14	LON-NHS	1	0–2	N/A	Cluster – time period	No	None	2 : 1	Yes

N/A, not applicable.
Including randomisation type, randomisation ratio by design and whether or not the randomisation ratio changed over the course of the trial. For some trials there was no second follow-up.

12. LON-HCA: a community-based trial in England, investigating IY and a literacy intervention alone, and IY plus literacy intervention, in low-income primary schools, with children with conduct problems (2–6 years).[101]
13. OXF: a community NGO-based trial in England, with children referred with conduct problems (2–9 years).[103]
14. LON-NHS: a clinic-based trial in England, with children referred for antisocial behaviour (3–8 years).[120]

Appendix 3 Harmonising of self-reported parenting practices

Positive parenting: use of praise, use of tangible rewards and monitoring

Across measures, items theoretically fitted three different constructs of positive parenting. Praise was defined as any verbal compliment in reaction to the child's behaviour. Tangible rewards were defined as any rewards for the child that are not verbal or physical, for example privileges, stickers on a chart, special food, a small toy or money. Monitoring was defined was parental supervision and knowledge of the child's whereabouts when the child is out of the parents' sight, including knowing the child's friends.

To assess positive parenting practices, four different instruments were used: PaPI (trials 1, 3, 6 and 10), APQ (trials 5 and 12), PS (trials 3, 7, 9 and 13) and interview version 1 (trials 10–12 and 14).

Several trials included multiple instruments: trial 3 had data on both the PS and the PaPI, although there were > 50% missing data on the PaPI. PS data were therefore used when available. Trial 10 had data on both the PaPI (selected items only) and the interview. PaPI data were used when available. Trial 12 had data on both the APQ (selected items only) and the interview. APQ data were used when available.

The most frequently used instrument was the PS. This instrument provides scores on a 7-point Likert scale. Scores from other instruments were therefore converted to a 7-point Likert scale. For the APQ, for example, scores are on 5-point scale. These were converted into a 7-point scale using $1 = 1$, $2 = 2.5$, $3 = 4$, $4 = 5.5$ and $5 = 7$.

Whenever possible, items selected were based on the original subscales of the instruments (*Table 27*). A detailed list of which items from which instrument were included is provided in the following sections. Internal consistency was sometimes low (lowest was $\alpha = 0.34$), often when there was a limited number of items. When more items were included internal consistency went up to 0.75 on the PaPI and 0.99 on the APQ.

TABLE 27 Overview of included positive parenting constructs and original instruments' subscales

Instrument	Construct	Original subscales
PS	Monitoring	Item was an extra item for the total score on the PS and was not part of any subscale
PaPI	Praise	Items are part of 'praise and incentives' subscale
	Tangible reward	Items are part of 'praise and incentives' subscale
	Monitoring	Items are part of the 'monitoring' subscale
APQ	Praise	Items are part of 'positive parenting' subscale
	Tangible reward	Items are part of 'positive parenting' subscale
	Monitoring	Items are part of 'poor supervision' subscale
Interview	Praise	N/A
	Tangible reward	N/A
	Monitoring	N/A

N/A, not applicable.

Items from the Parenting Practices Inventory (trials 1, 3, 6 and 10)
Praise (two items: 6b and 8a)

- In general, how often do you praise or compliment your child when your child behaves well or does a good job?
- Within the last 2 days how many times did you praise or compliment your child for anything he/she did well?

Tangible rewards (four items: 6d–f and 8b)

- How often do you buy something for him/her (such as special food, a small toy) or give him/her money for good behaviour when your child behaves well or does a good job?
- How often do you give points or stars on a chart when your child behaves well or does a good job (excluded in trial 10)?
- How often do you give him/her an extra privilege (such as cake, go to the movies, special activity for good behaviour) when your child behaves well or does a good job?
- Within the last 2 days how many times did you give him or her something extra, like a small gift, privileges, or a special activity with you, for something he/she did well?

Monitoring (five items: 12, 13, 14a–c; excluded in trial 10)

- About how many hours in the last 24 hours did your child spend at home without adult supervision?
- Within the last 2 days, about how many total hours was your child involved in activities outside your home without adult supervision?
- What percentage of the time do you know where your child is when he/she is away from your direct supervision?
- What percentage of the time do you know exactly what your child is doing when he/she is away from you?
- What percentage of your child's friends do you know well?

Items from the Alabama Parenting Questionnaire (trials 5 and 12)
Praise (four items: 2, 13, 16 and 27)

- You let your child know when he/she is doing a good job with something.
- You compliment your child when he/she does something well.
- You praise your child if he/she behaves well.
- You let your child know that you appreciate the child's help with chores (excluded in trial 12).

Tangible rewards (one item: 5)

- You reward or give something extra to your child for obeying you or behaving well (excluded in trial 12).

Monitoring (10 items: 6, 10, 17, 19, 21, 24, 28–30 and 32)

- Your child fails to leave a note, or let you know where he/she is going.
- Your child stays out in the evening past the time he/she is supposed to be home.
- Your child is out with friends you don't know.
- Your child goes out without a set time to be home.
- Your child is out after dark without an adult with him/her.
- You get so busy that you forget where your child is and what he/she is doing.
- You don't check that your child comes home at the time she/he was supposed to.
- You don't tell your child where you are going.
- Your child comes home from school more than an hour past the time you expect him/her.
- Your child is at home without adult supervision.

Item from the Parenting Scale (trials 3, 7, 9 and 13)
Praise (none)

Tangible rewards (none)

Monitoring (one item: 13)

- When my child is out of sight, I often don't know what my child is doing.

Items from parenting interview (trials 10 and 11)
Praise (one item)

- How many times did you praise your child for doing something you asked them or doing something well?

Tangible reward (one item)

- In the last week, how many times did you give a reward to your child for doing what you asked, such as sweets or crisps, extra television, more time playing football or a game he/she likes or a sticker?

Monitoring (none)

Items from parenting interview (trial 12)
Praise (one item)

- How many times per day do you praise (child)?

Tangible reward (one item)

- In the last week, how many times did you give a reward to your child for doing what you asked, such as sweets or crisps, extra television, more time playing football or a game he/she likes or a sticker?

Monitoring (none)

Items from parenting interview (trial 14)
Praise (none)

Tangible rewards (none)

Monitoring

Combined indoor and outdoor supervision. Score (ranging 0–3, recoded into 1–7) based on several items. Examples: How long would it be before you checked where he/she was, if he/she was playing outside? Do you always know where he/she is and with whom he/she is playing, by name?

Negative parenting practices (corporal punishment, harsh threatening, laxness and shouting)

Across measures, items theoretically fitted four different constructs of negative (i.e. harsh and/or inconsistent) parenting. Corporal punishment was defined as any physical punishment. Threatening was defined as threatening to punish the child (but not really punishing him/her). Laxness was defined as intending to punish the child but not actually following through/letting child get away with it. Shouting was defined as any voice raising, shouting, scolding, using of bad language, cursing or saying mean things.

To assess negative parenting practices, four different instruments were used: PaPI (trials 1, 3, 6 and 10), APQ (trials 5 and 12), PS (trials 3, 7, 9 and 13) and interview version 1 (trials 10–12 and 14). Please see *Chapter 2*, *Harmonisation of individual-level data*, for how data were harmonised on an item level.

Several trials included multiple instruments: trial 3 had data on both the PS and the PaPI, although > 50% missing data on the PaPI. PS data were therefore used when available. Trial 10 had data on both the PaPI (selected items only) and the interview. PaPI data were used when available. Trial 12 had data on both the APQ (selected items only) and the interview. APQ data were used when available.

The most frequently used instrument was the PS. This instrument provides scores on a 7-point Likert scale. Scores from other instruments were therefore converted to a 7-point Likert scale. For the APQ, for example, scores are on 5-point scale. These were converted into a 7-point scale using $1 = 1$, $2 = 2.5$, $3 = 4$, $4 = 5.5$ and $5 = 7$.

Whenever possible, items selected were based on the original subscales of the instruments (*Table 28*). A detailed list of which items from which instrument were included is provided in the following sections. Internal consistency was sometimes low (lowest was $\alpha = 0.41$), often when there was a limited number of items. When more items were included, internal consistency went up to 0.69 on the PS, 0.84 on the PaPI and 0.61 on the APQ.

We were able to compute correlations between the PS and the PaPI based on a small sample ($n = 44$) from one trial (trial 3). Correlations were small for pretest scores on threatening and laxness (0.30 and 0.34). However, all correlations on other time points and for all other constructs of negative parenting where more satisfactory, ranging from 0.53 to 0.87.

TABLE 28 Overview of included negative parenting constructs and original instruments' subscales

Instrument	Construct	Original subscales
PS	Corporal punishment	Item is from the 'hostility' subscale
	Threatening	Items are not part of any subscale
	Laxness	Items are from 'laxness'. Only items that matched items from other measures are included
	Shouting	Items are from the 'overreactiveness' and 'hostility' subscale. Some items are not part of any subscale
PaPI	Corporal punishment	Identical subscale
	Threatening	Items are part of 'harsh and inconsistent discipline' subscale
	Laxness	Items are part of 'harsh and inconsistent discipline' subscale
	Shouting	Items are part of 'harsh and inconsistent discipline' subscale
APQ	Corporal punishment	Identical subscale
	Threatening	Item is part of 'consistency subscale'
	Laxness	Items are part of 'consistency subscale'. Only items included that matched items from other measures
	Shouting	Items were not part of any subscale
Interview		N/A

N/A, not applicable.

Item from the Parenting Scale (trials 3, 7, 9 and 13)
Corporal punishment (one item: 18)

- I spank, grab, slap or hit my child most of the time.

Threatening (two items: 7 and 20r)

- I threaten to do things that I know I won't actually do.
- When I give a fair threat or warning, I always do what I said (reversed: I often don't carry it out).

Laxness (five items: 16, 19r, 24, 26r and 30r)

- When my child does something I don't like, I often let it go.
- When my child won't do what I ask, I take some other action (reversed: I often let it go or end up doing it myself).
- If my child misbehaves and then acts sorry, I let it go that time.
- When I say my child can't do something, I stick to what I said (reversed: I let my child do it anyway).
- If my child gets upset, I stick to what I said (reversed: I back down and give in).

Shouting (five items: 10r, 17r, 22, 25 and 28)

- I speak to my child calmly (reversed: I raise my voice or yell).
- Things build up and I do things I don't mean to (reversed: things don't get out of hand).
- I get so frustrated or angry that my child can see I'm upset.
- I almost always use bad language or curse.
- I insult my child, say mean things or call my child names most of the time.

Items from the Alabama Parenting Questionnaire (trials 5 and 12)
Corporal punishment (three items: 31, 33 and 36)

- You spank your child with your hand when he/she has done something wrong (in trial 12: you smack your child with your hand when he/she has done something wrong).
- You slap your child when he/she has done something wrong.
- You hit your child with a belt, switch or other object when he/she has done something wrong (in trial 12: you hit your child with a belt when he/she has done something wrong).

Threatening (one item: 3)

- You threaten to punish your child and then do not actually punish him/her.

Laxness (three items: 7, 21 and 24)

- Your child talks you out of being punished after he/she has done something wrong.
- You let your child out of a punishment early (such as lift restrictions earlier than you originally said).
- Your child is not punished when he/she has done something wrong (excluded in trial 12).

Shouting (two items: 37 and 38r; excluded in trial 12)

- You yell or scream at your child when he/she has done something wrong.
- You calmly explain to your child why his/her behaviour was wrong when he/she misbehaves (reversed).

Items from the Parenting Practices Inventory (trials 1, 3, 6 and 10)
Corporal punishment (six items: 1h, 1i, 2h, 2i, 3h and 3i)

- How often do you do each of the following things when your child misbehaves?
  Give your child a spanking (in trial 10: give your child a smack).
  Slap or hit your child (but not spanking) (in trial 10: slap or hit your child).
- If your child hit another child, how likely is it that you would discipline your child in the
  following ways?
  Give your child a spanking (in trial 10 give your child a smack).
  Slap or hit your child (but not spanking) (in trial 10: slap or hit your child).
- If your child refused to do what you wanted him/her to do, how likely is it that you would use each of
  the following discipline techniques?
  Give your child a spanking (in trial 10: give your child a smack).
  Slap or hit your child (but not spanking) (in trial 10: slap or hit your child).

Threatening (three items: 1d, 2d and 3d)

- How often do you do each of the following things when your child misbehaves?
  Threaten to punish him/her (but not really punish him/her).
- If your child hit another child, how likely is it that you would discipline your child in the following ways?
  Threaten to punish him/her (but not really punish him/her).
- If your child refused to do what you wanted him/her to do, how likely is it that you would use each of
  the following discipline techniques?
  Threaten to punish him/her (but not really punish him/her).

Laxness (three items: 5a, 5br and 5c)

- If you ask your child to do something and he/she doesn't do it, how often do you give up trying to get
  him/her to do it?
- If you warn your child that you will discipline him/her if he/she doesn't stop, how often do you actually
  discipline him/her if he/she keeps misbehaving (reversed)?
- How often does your child get away with things that you feel he/she should have been disciplined for?

Shouting (five items: 1b, 2b, 3b, 5e and 5f)

- How often do you do each of the following things when your child misbehaves?
  Raise your voice (scold or yell).
- If your child hit another child, how likely is it that you would discipline your child in the following ways?
  Raise your voice (scold or yell).
- If your child refused to do what you wanted him/her to do, how likely is it that you would use each of
  the following discipline techniques?
  Raise your voice (scold or yell).
- How often do you show anger when you discipline your child?
- How often do arguments with your child build up and you do or say things you don't mean to?

Items from parenting interview (trials 10–12 and 14)
Corporal punishment (one item)

- Thinking about last week, how many times did you give your child a tap or smack if he/she misbehaved?

Threatening (none)

Laxness (none)

Shouting (two items: excluded in trials 10 and 14)

- How many days a week do you find yourself raising your voice or shouting at your child or getting angry or cross at him/her?
- On the days that it does happen, how many times do you actually find yourself being critical or shouting at your child?

Appendix 4 Parental satisfaction with the intervention

It was not possible to analyse these data, owing to low sample size. However, we describe available data and potential methods for harmonising below.

The IY Satisfaction Questionnaire (PSQ9; Webster-Stratton[202]) was used to assess parental satisfaction with the IY intervention. Three different versions of the PSQ were used across the following trials: 2, 3, 5–11, 13 and 14. Trial 1 included two questions on parent satisfaction, but it was decided not to include these because the data were limited. For trials 10 and 11 it was not possible to match the item-level data and, therefore, these data were not included in the pooled data set. PSQ data were collected for trials 7 and 8, but it was not possible to match these data to the families' identifier numbers, so these data could not be included. Trial 13 used the Consumer Satisfaction Therapy Attitude Inventory;[203] however, only the total scores were available so it was not possible to harmonise these data at an item level. Data from this trial were also not included.

The following subscales were used: A, general satisfaction with the course; BU, usefulness of the teaching format; BD, difficulty of the teaching format; CU, usefulness of the content; CD, difficulty of the content; D, Evaluation of the group leader [a four-item subscale was created, rather than the three-item subscale as in the Webster-Stratton (2013)[202] version because many trials had split the first item into two ('leader's preparation' and 'leader's teaching'); E, satisfaction with the parent group.

A 'core items' mean score was also calculated for each of the above scales, which included items that were collected across the majority of trials. The following items were used as core items: subscale A – all items; subscale BD–BD1 (lecture information), BD2 (video-tapes), BD3 (group discussions), BD4 (role-plays), BD7 (practising skills at home); subscale BU–BU1 (lecture information), BU2 (video-tapes), BU3 (group discussions), BU4 (role-plays), BU7 (practising skills at home); subscale CU–CU1 (play), CU2 (attends), CU4 (rewards), CU6 (ignoring), CU7 (positive commands), CU13 (overall techniques); subscale CD–CU1 (play), CU2 (attends), CU4 (rewards), CU6 (ignoring), CU7 (positive commands), CU13 (overall techniques); subscale D – no core items were used as all trials which included the questions in this subscale used the same items; subscale E – no core items scale was used, as all items were included for trials that had included this subscale.

Appendix 5 Results data

TABLE 29 Descriptive data for full sample (data provided by trial arm in *Chapter 3*)

Variable	Combined	
	n	Mean (SD) (%)
Child gender (male)	1696	63.4
Child age (months)	1682	63.1 (17.8)
SES low income	1614	57.7
Low education	1696	38.6
SES lone parent	1606	35.4
SES teenage parent	1609	12
SES unemployed	1303	34.6
Ethnic minority	1651	29.8
ECBI-I total baseline	1622	137.9 (37.0)
ECBI-I total post test	1445	119.8 (36.2)
SDQ ADHD baseline	1532	5.9 (2.7)
SDQ ADHD post test	1171	5.6 (0.1)
SDQ emotional baseline	1340	3.4 (2.6)
SDQ emotional post test	1006	2.7 (0.1)
Monitoring baseline	1088	5.2 (1.7)
Monitoring post test	959	5.3 (1.6)
Tangible rewards baseline	625	3.3 (1.3)
Tangible rewards post test	544	3.5 (1.3)
Praise baseline	630	5.4 (1.2)
Praise post test	460	5.0 (1.2)
Corporal punishment baseline	1393	2.1 (1.4)
Corporal punishment post test	1038	2.1 (1.5)
Threatening baseline	999	3.6 (1.6)
Threatening post test	987	3.0 (1.5)
Laxness baseline	978	3.3 (1.3)
Laxness post test	945	3.2 (1.2)
Shouting baseline	967	3.1 (1.4)
Shouting post test	882	2.8 (1.3)
Parent depression BDI total baseline	1395	11.4 (10.5)
BDI total post test	1131	8.7 (9.0)
PSI-SF total baseline	542	91.1 (28.4)
PSI-SF total post test	502	81.3 (34.0)
PSOC scale total baseline	417	54.1 (7.6)
PSOC scale total post test	384	57.3 (13.8)

TABLE 30 Full results from moderator analyses

Moderator	Test of Between/within effects	Quadratic term (continuous variables)	Between/within trials	CC analysis Effect size	95% CI	p-value	MI analysis Effect size	95% CI	p-value
SES									
Low income[a]	$p = 0.286$			2.76	–4.10 to 9.62	0.43	1.91	–4.77 to 8.59	0.58
Low parental educational level	$p = 0.267$			3.12	–3.44 to 9.69	0.35	4.37	–2.17 to 10.90	0.49
Unemployed	$p = 0.669$			5.99	–1.84 to 13.81	0.13	4.88	–2.67 to 12.42	0.21
Teenage parent	$p = 0.051$		Between	–70.57	–149.46 to 8.31	0.08	–76.06	–166.10 to 13.98	0.10
			Within	8.47	–1.15 to 18.09	0.08	7.32	–2.24 to 16.87	0.13
Child demographic variables									
Child gender[a]	$p = 0.210$			–5.24	–11.60 to 1.12	0.11	–6.65	–13.03 to –0.27	0.04
Child age[a]	$p = 0.446$	$p = 0.885$		–0.20	–3.74 to 3.34	0.91	0.04	–0.14 to 0.22	0.65
Child problem severity									
Baseline ECBI-I[a]	$p = 0.004$	$p = 0.088$	Between	–16.37	–24.51 to –8.22	0.00	–18.29	–24.64 to –11.95	0.00
			Within	–3.13	–6.62 to 1.04	0.08	–4.30	–7.87 to –0.73	0.02
Baseline ADHD	$p = 0.576$	$p = 0.0219$	Linear	0.82	–2.63 to 4.27	0.07	0.74	––2.61 to 4.09	–0.07
			Quadratic	3.42	0.50 to 6.35		3.00	0.28 to 5.72	
Baseline emotional problems	$p = 0.2815$	$p = 0.3775$		–2.01	–5.74 to 1.73	0.29	–2.92	–6.73 to 0.89	0.13
Parental mental health									
Baseline parental depression[a]	$p = 0.3009$	$p = 0.3050$		–4.52	–8.00 to –1.04	0.01	–4.79	–8.43 to –1.14	0.01
Positive parenting									
Parenting: monitoring	$p = 0.3422$	$p = 0.3586$		1.89	–2.39 to 6.16	0.39	1.76	–2.51 to 6.03	0.42
Parenting: tangible rewards	$p = 0.8287$	$p = 0.3428$		–1.92	–7.36 to 3.53	0.49	–3.03	–7.68 to 1.61	0.20
Parenting: praise[a]	$p = 0.6807$	$p = 0.5529$		–0.63	–6.06 to 4.80	0.82	–3.67	–8.73 to 1.39	0.16

Moderator	Test of		Between/ within trials	CC analysis			MI analysis		
	Between/ within effects	Quadratic term (continuous variables)		Effect size	95% CI	*p*-value	Effect size	95% CI	*p*-value
Negative parenting									
Parenting: corporal punishment[a]	$p = 0.4240$	$p = 0.1765$		1.19	−2.25 to 4.63	0.50	0.38	−3.04 to 3.81	0.83
Parenting: threatening	$p = 0.1653$	$p = 0.4287$		0.47	−3.69 to 4.63	0.83	0.71	−3.48 to 4.91	0.74
Parenting: laxness	$p = 0.3904$	$p = 0.0891$	Linear	−5.77	−10.20 to −1.35	0.02	−4.25	−8.58 to 0.09	0.12
			Quadratic	2.70	−0.41 to 5.81		1.61	−1.40 to 4.61	
Parenting: shouting	$p = 0.8893$	$p = 0.2475$		−0.51	−4.82 to 3.80	0.82	−0.02	−4.06 to 4.01	0.99
Ethnicity									
Ethnic minority	$p = 0.042$		Between	18.44	0.91 to 35.98	0.04	19.54	1.04 to 38.05	0.04
			Within	−1.66	−10.31 to 6.99	0.71	−1.37	−9.81 to 7.08	0.75

a Denotes that the variable is the representative variable for that domain.
Moderator effect sizes by domain for both CC and MI analyses. The *p*-values for the test for separate between- and within-trial effects and for the addition of a quadratic term for continuous variables are displayed. If these are significant then both between- and within-trial effects or linear and quadratic terms are reported.

Appendix 6 Public and patient involvement

Parent involvement

Two focus groups were held during the project to seek parents' views on some preliminary results and their thoughts on how we could disseminate the results from the project. Both focus groups took place in the south of England: one in a children's centre and one in a NGO concerned with youth mental health.

Seven parents took part in these groups and provided feedback. Six parents who took part in the focus groups had taken part or were taking part in an IY course and one parent had taken part in a different parenting intervention. Included in the group were parents with low income and parents from an ethnic minority background.

The main points raised by parents are summarised below.

Severity of child's behaviour problems

Parents thought that IY may be more helpful for children with higher levels of behaviour problems, owing to them having more potential for change in their behaviour, but that there might be a level at which the course becomes less helpful.

A parent highlighted how the level of behaviour problems that a parent is experiencing can be very subjective and what one parent considers to be challenging behaviour may not be for a different parent.

Parents wondered if parents who were not having difficulty with their child's behaviour would go to the course; parents suggested that they may be less motivated.

Parents identified two points as important to the outcome of the intervention: parents' engagement with the course and the level of severity of the child's behaviour problems. Parents also identified that the group experience was very important.

Parental education

Parents did not think that the level of qualifications that a parent has could impact upon how helpful an IY group is for parents or children. Parents commented that the materials are accessible to all parents.

Symptoms of parental depression

All six parents said that IY could help children regardless of whether or not their parent is experiencing symptoms of depression. Some parents commented that parents who are experiencing symptoms of depression may benefit from the course by meeting other people and doing something for themselves.

A couple of parents said they have experienced depression and felt that the IY course would be more helpful for parents who are experiencing depression, as the skills would help parents know how to manage their child's behaviour when they feel low or are experiencing depression. One parent commented 'I suffer with depression, I say I do, but since I did this course my children are a lot happier and consequently I am happier . . . I feel like I have more time on my hands to enjoy being me and enjoy being a mum'. 'I felt like I wasn't doing enough but now I feel like I am and that it shown in their behaviour and their smiles'.

Another commented 'for me, being somebody who feels low sometimes, I felt it helped to have that stability in your reactions . . . and that can then help the children'.

Different ethnicities

Parents did not think there would be any difference in how helpful the course is for parents of different ethnicities.

Child age

Most parents thought that younger children may benefit more than older children, a parent commented 'starting to have it all set up for when they grow up, or as they are growing is best'. One parent said that it could be the case that a child may be able to learn more from the course when they are older.

Another commented that 'you see more of a difference if you start young, starting on the right path . . . it can be easier to change for younger children, if they are older they have more patterns of behaviour'.

However, one parent highlighted the fact that all parents in one of the focus groups had children who are in the middle age bracket, which could be why they themselves felt that the age of the child would not make a difference (parents also highlighted how this was evident on the graphs).

Other ideas to explore: in future research

A parent felt that it would be helpful to explore the type of service in which a group was delivered in relation to outcomes, owing to possible differences between services.

Reasons for attending the course

Parents described attending the course because of their child's challenging behaviour or they wanted to change their parenting skills, or both.

Parental experience of Incredible Years

All parents said that they found the IY programme helpful and that they found it useful to have the group environment in which to talk to other parents. Parents described IY as helping them to cope with being a parent.

Teenage parents

In one group parents thought it could be more helpful for parents and children if the parent was a teenager, as it could be the case that the parent would then have more to learn.

A parent highlighted how it may be that teenage parents would be more likely to attend if the group was specifically a group for teenage parents and the materials were adapted specifically for them.

In another group, parents thought that whether or not you were a teenager at the time of your child's birth would not have an effect on how helpful the intervention was for a family.

One parent commented on the fact that the parent has to be motivated to take part in a group, so it would depend on whether or not the teenage parent was motivated to participate: 'If they wanted to learn, absolutely, but if they didn't . . . they have to want to learn'.

Drawbacks to attending an Incredible Years course

Parents could not think of any drawbacks to attending a course other than the time it takes. In relation to time they emphasised the need to be able to fit the course in and make the commitment to it.

Parents highlighted how it is important 'to have everyone around the child on board', as otherwise the child would encounter other discipline strategies.

Dissemination of results

Parents thought that it was important to disseminate the results to parents via a number of services. Parents thought dissemination via NHS services was important, for example GPs or health visitors: 'In the NHS – let people know, that would be the most effective way of letting parents know.' Parents suggested that this could be via letters. Parents also said that informing parents via children's centres may be helpful, as we did for one of the focus groups.

Appendix 7 Harmonisation of economic variables

Sources of 'structural' differences between trials

Differences in Client Service Receipt Inventory coverage

All five studies used a variant of the CSRI[177] to record information on to service use alongside the trial. Differences in the level and variation in costs can arise from 'true' differences in service utilisation, but also from structural differences in the data collection instrument. The CSRIs for the included studies were scrutinised, and emerging differences thought to potentially impact costs guided the harmonisation effort. Note that CSRI data were available for only four trials, and not necessarily for both baseline and follow-up. Although all studies collected information on the number of service contacts, only two recorded the estimated average duration of contacts. Duration data are often a source of missingness in service use data, and this holds true in the current trials that attempted to collect these.

Although all trials explicitly asked for contacts with key primary care staff (GPs, general practice nurses), hospitals (inpatient, outpatient, A&E) and social workers, there was some variability in the services listed. Although not inherently problematic in a single trial, this may have influenced response patterns between trials, as participants were exposed to different prompts. Additional education support was covered in two trials. Note that this is in part related to the age of the study population.

Four trials provide information on the impact of the child's behaviour on parental employment, three have some data on parental use of services and two have additional data on the impact on parental time spent caring for the child.

Although all studies provided space for participants to record 'other' services beyond those listed on the CSRI, few data sets actually included information on the type of service, with most reporting only the number of contacts. Again, this question was a notable source of missing data when this variable was included.

The setting of service provision has implications for the unit cost. For example, overheads for a hospital-based health professional are higher than if that same contact had taken place in a community setting. Three trials recorded service contacts including their setting at baseline, whereas another provided data on whether or not contacts were 'usually' at home. At follow-up, this information was available for all but one trial and this was provided at baseline.

The period prior to randomisation covered was 6 months for three trials, and 12 months for two trials. The time between the baseline assessment and the follow-up data collection point was 6 months for three trials, 3 months for one trial and 12 months for one trial.

Another potential source of difference is whether questions on the CSRI were phrased to include all service contacts or only those relating to a particular problem ('marginal' perspective).

Assumptions made at point of data entry and missing data strategies

In some cases, the 'raw' data provided by the research teams allowed us insight into the assumptions made when interpreting information provided by participants on the CSRI at the point of data entry. This often related to the number of contacts that were entered. A common assumption was that one service contact took place when no contact number was recorded. Another issue encountered at the point of data entry was to translate a number given for a specific time period (e.g. 'once per week') into a number that matched the CSRI time frame (e.g. 'contacts per term').

In addition to these 'implicit' missing data strategies, for several trials it was noted that missing data appeared to have been dealt with more 'explicitly' and comprehensively. In at least one trial it is assumed that all missing values were replaced with zeros.

It is worth noting that it may not have been possible to recognise and identify all such strategies that may have been applied to data, and it is impossible to know whether or not these have been applied consistently within a particular trial, especially given that often more than one person is responsible for data entry.

Implications for harmonisation

The following strategies were employed to harmonise data, and to account for the structural differences between trials.

A dummy variable was created to identify trials from England compared with trials from other countries. Similarly, a variable indicating the (original) follow-up period was included in the analyses, whereas costs were adjusted to cover a 6-months period across all trials. Different missing data strategies (that may be common across several trials) cannot be identified but are captured to some extent by the fixed effect for 'trial' in our models.

To minimise the impact of structural differences between data collection instruments, the main analysis focuses on health and social care costs. We show results when at least two trials have provided information. Given that most trials did not report duration data, and duration data were often missing, we determined a 'typical' duration for each service contact (see *Appendix 8*). Standard unit costs were used for 'other' services, as details on the type of service used were more often than not unavailable or missing. As the setting of service provision was not available for all trials, and in practice it is not possible to distinguish unit costs for services received within a GP practice compared with a health clinic (for example), these data were summarised into contacts that take place at the health or social care professional's typical place of work compared with contacts that take place elsewhere and therefore likely include a travel element.

Appendix 8 Unit costs (2014 GBP) used in economic analyses

TABLE 31 Unit costs (2014 GBP) used in economic analyses

Service	Per hour (£)	Duration (£)	Unit cost (£)	Source
GP surgery/clinic			38	PSSRU 2014,[176] p. 195
GP home/other		23.4	76	PSSRU 2014,[176] p. 195
Nurse/GP nurse surgery/clinic	50.5	19.75	17	PSSRU 2014,[176] p. 187
Nurse other	50.5	31.75	27	PSSRU 2014,[176] p. 187
Health visitor surgery/clinic	65	24	26	PSSRU 2014,[176] p. 189
Health visitor other	65	50	54	PSSRU 2014,[176] p. 189
Hearing problems/audiologist	32	30	16	PSSRU 2014,[176] p. 179
Speech and language therapist surgery	32	45	24	PSSRU 2014,[176] p. 179
Speech and language other	32	57	30	PSSRU 2014,[176] p. 179
Other primary	32	30	16	PSSRU 2014,[176] p. 179
Physiotherapist surgery	32	45	24	PSSRU 2014,[176] p. 179
Physiotherapist other	32	57	30	PSSRU 2014,[176] p. 179
Community paediatrician			310	PSSRU 2014,[176] p. 85
Social worker	55	60	55	PSSRU 2014,[176] p. 207
Sessional worker	50	60	50	PSSRU 2014,[176] p. 212
Home help/home care worker	24	60	24	PSSRU 2014,[176] p. 210
CAMHS			236	*NHS Reference Costs 2013–14*[178]
Child guidance centre/psychiatric worker	14	90	22	Cost-effectiveness of children's centres in England[204]
Child development centre			310	PSSRU 2014,[176] p. 85
A&E/casualty			141	*NHS Reference Costs 2013–14*[178]
Ambulance			180	*NHS Reference Costs 2013–14*[178]
Outpatient appointment			189	PSSRU 2014,[176] p. 85
Inpatient stay			1095	*NHS Reference Costs 2013–14*[178]
Day hospital/day care centre			296	PSSRU 2014,[176] p. 85
Other (hospital)			189	PSSRU 2014,[176] p. 85
Home Start			18	PSSRU 2004,[205] uprated
Day care centre			22	*Childcare Cost Survey 2015*,[206] nursery
Drop-in centre			3	*Childcare Cost Survey 2011*,[206] after-school club
Counselling/advice services	50	60	50	PSSRU 2014,[176] counsellor primary care

TABLE 31 Unit costs (2014 GBP) used in economic analyses (*continued*)

Service	Per hour (£)	Duration (£)	Unit cost (£)	Source
Support group	13.8	60	14	PSSRU 2014[176]
Telephone help line			16	Own calculations
Web pages			0	
Voluntary agency – home based			18	As Home Start, PSSRU 2004,[205] uprated
Other			18	As home care worker, PSSRU 2014,[176] p. 210
Respite foster care		700 per week	700	PSSRU 2014,[176] p. 88
Children's home		2995 per week	2995	PSSRU 2014,[176] p. 86
Foster home		700 per week	700	PSSRU 2014,[176] p. 88

PSSRU, Personal Social Services Research Unit.

Appendix 9 Full service use table for economic analyses

TABLE 32 Full service use for economic analyses

Service	Baseline				Follow-up			
	Control (n = 236)		IY (n = 372)		Control (n = 236)		IY (n = 372)	
	Number using service	Per cent using service	Number using service	Per cent using service	Number using service	Per cent using service	Number using service	Per cent using service
Hospital								
A&E/casualty	40	17	69	19	29	12	58	16
Ambulance	3	1	7	2	4	2	2	1
Outpatient appointment	49	21	77	21	38	16	51	14
Inpatient stay	11	5	20	5	5	2	17	5
Day hospital	2	1	7	2	0	0	2	1
Other hospital	6	3	10	3	3	1	3	1
Community health care								
GP surgery/clinic	170	72	263	71	116	49	195	52
GP home/other	1	0	7	2	0	0	2	1
Nurse/GP nurse surgery	25	11	47	13	17	7	38	10
Nurse other	5	2	4	1	1	0	3	1
Health visitor surgery/clinic	33	14	64	17	12	5	14	4
Health visitor other	27	11	67	18	13	6	23	6
Hearing problems/audiologist	10	4	8	2	4	2	1	0
Speech and language therapist surgery	20	8	43	12	15	6	29	8
Speech and language therapist other	8	3	17	5	2	1	8	2
Other primary care	9	4	19	5	9	4	15	4
Physiotherapist surgery	1	0	8	2	3	1	4	1
Physiotherapist other	4	2	7	2	2	1	4	1
Community paediatrician	9	4	12	3	4	2	12	3
Homeopath	0	0	0	0	0	0	0	0

TABLE 32 Full service use for economic analyses (*continued*)

	Baseline				Follow-up			
	Control (*n* = 236)		IY (*n* = 372)		Control (*n* = 236)		IY (*n* = 372)	
Service	Number using service	Per cent using service	Number using service	Per cent using service	Number using service	Per cent using service	Number using service	Per cent using service
Specialist mental health								
CAMHS	3	1	8	2	4	2	2	1
Child guidance centre/psychiatric worker	10	4	16	4	4	2	14	4
Child development centre	7	3	9	2	5	2	5	1
Social work								
Social worker	11	5	30	8	12	5	20	5
Sessional worker	0	0	6	2	0	0	0	0
Home help/home care worker	1	0	0	0	1	0	0	0
Voluntary sector								
Home Start	2	1	5	1	2	1	2	1
Day care centre	3	1	6	2	1	0	4	1
Drop-in centre	1	0	3	1	2	1	1	0
Counselling/advice services	2	1	1	0	3	1	2	1
Support group	1	0	1	0	0	0	0	0
Telephone helpline	0	0	1	0	1	0	1	0
Web pages	0	0	1	0	0	0	0	0
Voluntary agency – home based	0	0	0	0	0	0	0	0
Other voluntary sector service	3	1	4	1	1	0	1	0
Accommodation								
Respite foster care	1	0	3	1	0	0	1	0
Children's home	0	0	0	0	0	0	0	0
Foster home	1	0	0	0	0	0	0	0

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library