

This is a post-peer-review, pre-copyedit version of an article published in Children and Youth Services Review. The final authenticated version is available online at:

<https://www.sciencedirect.com/science/article/abs/pii/S0190740920303170?via%3Dihub>

The effectiveness of a therapeutic parenting program for children aged 6-11 years with behavioral or emotional difficulties: results from a randomized controlled trial

Nick Axford (NIHR ARC South West Peninsula (PenARC), University of Plymouth, Plymouth, UK)*

Gretchen Bjornstad (University of Exeter Medical School, University of Exeter, Exeter, UK)

Justin Matthews (PenARC, University of Exeter, Exeter, UK)

Sarah Heilmann (Erasmus University Rotterdam, Rotterdam, Netherlands)

Anam Raja (University of Oxford, Oxford, UK)

Obioha C. Ukoumunne (PenARC, University of Exeter, Exeter, UK)

Vashti Berry (PenARC, University of Exeter, Exeter, UK)

Tom Wilkinson (Devon Partnership NHS Trust, Exeter, UK)

Luke Timmons (Stand Together, Washington DC, US)

Tim Hobbs (Dartington Service Design Lab, Dartington, UK)

Tim Eames (University of Exeter, Exeter, UK)

Angeliki Kallitsoglou (University of Roehampton, London, UK)

Sarah Blower (NIHR ARC Yorkshire and Humber, University of York, York, UK)

Georgina Warner (Uppsala University, Uppsala, Sweden)

* Corresponding author: University of Plymouth, N10, ITTC Building, Plymouth Science Park, PLYMOUTH PL6 8BX, UK, nick.axford@plymouth.ac.uk

Funding support: This research was funded by Big Lottery (Realising Ambition Programme Grants: Agreement 30118942). The funding organization had no role in the design and conduct of the study; collection, analysis or preparation of data; or preparation, review, or approval of the manuscript. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of Big Lottery.

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: Ethical approval was granted by the Warren House Group Research Ethics Committee on 20th October 2014 (ref. WHG2014-002). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained for all individual participants included in the study.

Human and animal studies: This article does not contain any studies with animals performed by any of the authors.

Trial Registration: Current Controlled Trials ISRCTN32083735. Retrospectively registered 28 October 2014.

Acknowledgements

We are grateful to Malachi Trust staff and all trial participants for making this study possible. The time of Nick Axford, Vashti Berry, Justin Matthews and Obioha C. Ukoumunne is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration South West Peninsula (PenARC). The time of Sarah Blower is supported by the NIHR Applied Research Collaboration Yorkshire and Humber. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Abstract

Objective: The aim of the study was to evaluate the implementation and effectiveness of a therapeutic parenting program that targets parents of children aged 6 to 11 years identified as having behavioral and emotional difficulties. The intervention comprises two parts, delivered sequentially: a 10-12-week group-based program for all parents, and one-to-one sessions for up to 12 weeks with selected parents from the group-based element.

Methods/Design: In a randomized controlled trial, 264 participants were allocated to the *Inspiring Futures* program (intervention) or services as usual (control) arms with follow-up assessments at 16 (post-group program) and 32 (post-one-to-one sessions) weeks. The primary outcome was the parent-rated Strengths and Difficulties Questionnaire (SDQ) Total Difficulties score at 32 weeks. Secondary outcomes included parent-rated SDQ subscales, parent coping strategies, empathy in parenting and parenting skills.

Results: All 264 participants were included in outcome analyses. There was no statistically significant effect on SDQ total difficulties (standardized mean difference: -0.07; 95% CI: -0.30 to 0.16; $p = 0.54$). There were no sub-group effects. Only 1 of 40 comparisons between the trial arms for secondary outcomes across both follow-ups was statistically significant at the 5% level. The mean number of group sessions attended by intervention arm participants was 6.1 (out of 10 to 12) and only 1 in 20 intervention arm participants received one-to-one support. Based on independent observation, mean adherence, quality and participant responsiveness scores indicated scope to improve fidelity.

Conclusions: The intervention is not more effective than services as usual at improving targeted outcomes. This may be related, in part, to implementation issues but arguably more to the inability of a non-behavioral intervention to improve caregiving adequately, particularly when it is not targeted at new parents who have experienced trauma or deprivation early in life or subsequently.

Keywords: Parenting, early intervention, group psychotherapy, child behavioral and emotional problems, randomized controlled trial

Highlights

- The intervention was not more effective than services as usual at improving outcomes
- This may relate, in part, to implementation issues; fidelity could be improved
- It may also reflect the relative lack of behavioral components in the program

1. Introduction

Left unchecked, behavioral difficulties in early childhood elevate children's risk for poor outcomes across multiple domains, including academic achievement, health, social relationships and offending (Bailey, Hill, Oesterle and Hawkins, 2009; Breslau, Lane, Sampson and Kessler, 2008; Calkins and Keane 2009; Fletcher, 2010; Nagin and Tremblay, 1999; Patel, Flisher, Hetrick and McGorry, 2007; Roza, Hofstra, van der Ende and Verhulst, 2003). Longitudinal research indicates that serious anti-social behavior can be predicted in childhood (Farrington and Welsh, 2007) and parenting, particularly poor parental monitoring, psychological control and negative aspects of support such as rejection and hostility, has been linked to delinquency (Hoeve et al., 2009). National surveys over the last 20 or so years reveal a slight increase in the prevalence of a mental disorder in 5 to 15 year-olds in England, rising from 9.7% in 1999 to 10.1% in 2004 and 11.2% in 2017 (Sadler et al., 2018).

Meanwhile, a cross-national comparison of 21 industrialized countries in 2007 ranked the UK lowest on two out of six dimensions of child well-being, namely behavior – including poor health behaviors, risk-taking and experience of violence – and family and peer relationships (Adamson, Bradshaw, Hoelscher and Richardson, 2007).

Partly in response to these challenges, the years 2005 to 2010 saw central government in the UK support the increased use of evidence-based parenting programs and provided financial incentives to local authorities, such as training grants, to adopt them (Lewis, 2011). Several of these were rolled out nationally in England, reaching over 6,000 parents in 43 local authorities, with significant improvements over time in aspects of parenting and child conduct problems (Lindsay and Strand 2013). There is a robust evidence base from efficacy trials for the effectiveness of these mostly group-based behavioral interventions for improving parenting practices, parent mental health and child emotional well-being and behavior (including for children with Attention Deficit Hyperactivity Disorder (ADHD))

(Zwi et al., 2011; Barlow et al., 2012; Furlong et al., 2012). The effects on disruptive behavior appear to increase in size with each level of prevention: universal (Cohen's $d=0.27$); selective ($d=0.33$); indicated ($d=0.65$); and treatment ($d=0.79$) (Leijten et al., 2019), while economic analyses support likely long-term savings to the public sector (e.g. Bonin et al., 2011; Gardner et al., 2017). However, the parenting programs with the strongest evidence of impact on child behavior, notably Incredible Years (Menting et al., 2013; Leijten et al., 2018a) and Triple P (Sanders et al., 2014), were developed in North America and Australia respectively, albeit with evidence from randomized controlled trials and quasi-experimental studies demonstrating their effectiveness in Europe, including the UK (Incredible Years only – Gardner et al., 2006; Hutchings et al., 2007; Scott et al., 2010; Morpeth et al., 2017) and Ireland (Incredible Years – McGilloway et al., 2012; Triple P – Doyle et al., 2018).

In 2012, the National Lottery Community Fund launched Realising Ambition, a five-year £25 million program that provided support to voluntary and community sector organizations across the UK to prevent young people aged 8 to 14 years from taking pathways into crime. It did this by replicating evidence-based programs in new geographical areas but also, critically, seeking to build evidence for the effectiveness of home-grown interventions. One of these was *Inspiring Futures*, an indicated parenting program for parents of children with behavioral and emotional difficulties. The program was developed and is delivered by Malachi Specialist Family Support Services ('Malachi'), a not-for-profit therapeutic family support organization working extensively in the UK Midlands for over 25 years. According to the program logic model, parents with traumatic early experiences (e.g. their own parents separating, witnessing domestic violence, the death of a significant person, not feeling safe physically or emotionally) are more likely to use maladaptive coping strategies and show less empathy in parenting, both of which contribute to their children displaying emotional and behavioral problems. *Inspiring Futures* aims to break these links by using a structured

process to increase parent awareness of how: (i) past experiences influence current behavior, (ii) maladaptive coping strategies affect parenting behavior and (iii) parenting behavior affects child behavior, and then encouraging the application of this learning and self-reflection. The program also uses child development education and solution-focused therapy (De Shazer et al., 2007) to help parents develop better parenting skills, such as praise and effective communication.

In contrast to the predominantly *behavioral* group-based parenting interventions referred to earlier, *Inspiring Futures* is primarily aligned with a type of parenting program that is *therapeutic* in orientation, focusing on enhancing the relationship between the parent and child rather than on helping parents develop skills to change the child's behavior. It does this by helping parents to mentalize – to envision mental states in themselves and others, in this instance their child, so that they can better reason about their own and other's behavior – in order to improve parent-child relationships and, in turn, caregiving practices (Slade, 2007; Coyne, 2013; Kalland et al., 2016; Luyten et al., 2017). Mentalization and reflective functioning are predictive of sensitive parenting, more secure child attachment and children's social skills and ability to regulate their emotions. However, parents' ability to mentalize and interact responsively with their children is shaped by their own experiences of being parented, which may not have been positive, and other forms of deprivation and trauma (Felitti et al., 1998; Murphy et al., 2014). Improving parents' ability to mentalize is therefore theorized to help prevent the intergenerational transmission of psychopathology. This type of intervention contrasts with behavioral parenting programs, which are rooted in social learning theory and emphasize the development and practice of specific skills, such as collaborative play, selective attention and praise to promote sociable behavior (Scott and Gardner, 2015). These are evident in *Inspiring Futures* but to a lesser extent.

The program was selected for trial based on it being an established and manualized intervention with preliminary qualitative evidence indicating that participating parents feel supported by Malachi workers and perceive their parenting skills and family relationships to have improved (Hickman, 2007). The objectives of the trial were to: (i) estimate the effect of the *Inspiring Futures* program on the behavior and emotional well-being of children with parent-reported emotional and behavioral difficulties; (ii) estimate the impact of *Inspiring Futures* on secondary outcomes, namely parent coping strategies, empathy in parenting and parenting skills (all potential mediators, albeit not analyzed as such in this study); and (iii) describe the extent to which *Inspiring Futures* is implemented with fidelity to the program design.

2. Methods

2.1 Design

A two-arm, individually randomized controlled, parallel group, superiority trial was conducted to evaluate the effectiveness of *Inspiring Futures* in improving behavioral and emotional outcomes in primary school children with elevated psychosocial difficulties. All aspects of the study design are described in detail in the protocol article (Reference removed for peer review) and summarized here.

2.2 Participants

The sample was drawn from primary schools in two trial sites in England: Birmingham and Somerset. The former is a large ethnically diverse city while the latter is a large rural county (in which two large towns were involved). Participants were the parents of children aged 6 to 11 years at the point of referral (school Years 2 to 6) who were referred by school staff to

Inspiring Futures. In order to be eligible, and in accordance with usual practice when delivering *Inspiring Futures*, children needed to display psychosocial difficulties in the *home* context (given the program's focus on parenting), identified by a parent-report in the 'borderline' or above range of the Strengths and Difficulties Questionnaire (SDQ) Total Difficulties score (i.e. ≥ 14) (Goodman, 1997). Malachi applied the following exclusion criteria using professional judgement: a parental mental health issue, substance abuse issue or significant self-esteem and/or confidence issue that would seriously affect their ability to participate in the group sessions; a family situation that does not allow the parent to engage fully in the process; other reasons that prevent the parent from participating in the group (not proficient in English, physical health issues, childcare obligations, work commitments); and parent is already receiving other therapeutic support.

2.3 Control arm

Children and parents assigned to the control arm were able to receive services as usual (with the exception of other Malachi services) because the aim was to examine the superiority of *Inspiring Futures* relative to a comparator that represents what children and their families normally receive. Malachi stated that usual care varies depending on school setting and area, although there may be other organizations that cover the whole city/area. The offer was considered likely to include support from a school pastoral team or voluntary organizations, and there are also parenting programs run by local children's centres. However, whereas other programs tend to focus on the practical aspects of parenting, *Inspiring Futures* concentrates on parents' own experience of being parented and subsequent attachment styles. Referrers were also signposted to standard universal children's services directories in both sites, which may be used to refer children to other services.

2.4 Intervention arm

Inspiring Futures comprises two parts, delivered sequentially: a group-based element for all parents, and one-to-one sessions with selected parents from the group-based element. Group facilitators (Malachi staff) are trained to or working towards at least Level 3¹ in counselling (or equivalent) and attend an intensive 3- to 4-day training session. During this training, they first participate in *Inspiring Futures*, delivered by existing facilitators, to ensure that they are familiar and comfortable with the content. They are then trained in program delivery.

Parents are invited to attend 10 to 12 (90-minute) weekly group sessions (each group ideally comprised of between 4 and 10 parents) at their child's school facilitated by two Malachi group facilitators. Group sessions run during a school term; the number depends on term length, but the same content is covered in all groups. The sessions first identify early traumatic experiences of the parents and aim to raise awareness of how these can influence current behavior. Maladaptive coping strategies, such as alcohol or substance misuse and self-harming behaviors, are identified and parents are made aware of how these can affect parenting behavior. A combination of child development education and psychodynamic, transactional and solution-focused therapy is used to improve empathy towards the child and, in turn, parenting skills. Specifically, participants come to understand more about the nature of children's needs, the barriers to them as parents meeting those needs (e.g. money problems, poor mental health, domestic violence, own experiences of being parented), and how children's feelings of fear, anxiety or rejection can manifest as aggression or violence. Malachi anticipates that improved parental empathy and parenting skills – the latter developed in part through encouraging greater reflection on one's actions but also facilitators

¹ Equivalent to an A-Level in the UK, which is the qualification typically received at age 18 years and provides a means of accessing university. It is pre-accreditation as a qualified counsellor.

teaching simple routines for dealing with challenging behavior and addressing conflict – will improve the child’s behavior and emotional well-being.

Topics covered in sessions include: children’s needs (e.g. for stimulation, affection, boundaries) and barriers to parents meeting them (see above); how one’s experience of being parented, and related attachment styles, affects one’s own relationships and parenting; how actions are driven by thoughts and feelings, and – in the light of that – how children’s aggression, violence or rejection may reflect fear, shame or anxiety; how adults’ coping mechanisms to meet or regulate their own needs can be healthy or problematic and, in the case of the latter, make it harder for parents to meet a child’s needs; the drivers of behavior and how one’s default responses can change through a more reflective or mindful approach; how the people one relies on, seeks advice from and looks to for support can affect parenting styles and decisions; how conflicts occur and can be resolved, and how one’s default role in conflicts relate to early experiences; and how reflection on one’s values, early experiences and other behavioral drivers can help with moving towards healthier responses to everyday situations and dilemmas.

Each group session follows the same structure. First, facilitators seek to raise awareness of the designated topic for the session. This often involves facilitators sharing teaching points, explaining psychological theories and leading group discussion. Second, perspective taking (i.e. understanding the difficulty of implementing the theory that has been delivered) is promoted in order to alleviate any guilt that parents might feel as a result of the learning that has taken place. Third, parents are encouraged to think about how the topic impacts on their life, and specifically how their childhood experiences affect their adult and parenting behavior. This can include storytelling, self-disclosure, practical activities or quiet self-reflection. Facilitators disclose relevant examples from their own life in order to normalize the experience and encourage sharing. Finally, most sessions conclude with a ‘homework’

task focused on implementing change over the next week; parents are asked to reflect on how well they managed the task at the next session (verbally with the group, or submitted in confidence in writing to the facilitator). These tasks include a mix of reflection (e.g. on the barriers that parents feel prevent them from meeting their child's needs, or how to change an important relationship) and action (e.g. using a new technique covered in the program to manage child conduct problems, or adopting a different role in conflicts, or practising mindful parenting).

The one-to-one component is delivered to a selection of parents who attended the groups and takes place during the term following group sessions. Malachi consider the following factors when determining eligibility for this further support: difficulty in applying the 'homework' tasks; parent difficulty in connecting past experiences to the present day; parent difficulty in recognizing the impact of their parenting behavior on the child; and self-disclosure of an issue that suggests the parent has significant unresolved emotional issues. The additional support runs initially for six weeks (one 60-minute session per week) and recaps sessions 2 to 7 of the group element, with the one-to-one delivery allowing deeper discussion with the parent. At week 6 a review involving the facilitator and parent determines whether further input is needed for the remainder of the school term (usually a further 4 to 6 weeks). The worker is consistent for all one-to-one sessions for any given parent, and ideally is the facilitator of the group that the parent attended.

Malachi supports facilitators with implementation via monthly supervision sessions in which the supervisor (a trained counsellor) reviews the self-report checklists completed in each group session and discusses with the facilitator how to overcome any challenges.

2.5 Participant timeline

Referrals in the study were made by school staff (e.g. a teacher or Special Educational Needs Coordinator) who knew the child well and had concerns about their behavior or emotional well-being. Malachi assessed participant eligibility using the parent SDQ Total Difficulties score at a parent information session in the child's school (or in a few cases by telephone for parents unable to attend). For those who were eligible, remaining baseline data were collected by a researcher in a home visit (usually two weeks after baseline SDQ). Randomization was designed to take place 10 weeks after referral. Follow-up data were collected from parents by a researcher using paper-based questionnaires in home visits 16 and 32 weeks after randomization (equivalent to the end of the parent group and one-to-one parts of the program respectively). The time between processing a referral and second follow-up was designed to be about 9 months. Data were collected from all participants who could be contacted and who consented to participate in data collection, regardless of their level of participation in the intervention. Implementation fidelity data were collected by Malachi supervisors and shared with the research team. Two sessions per course (ideally 4 and 8) were video-recorded by Malachi (with participants' consent) for coding by researchers. All data collection took place between October 2014 and June 2017.

2.6 Measures

Measures were selected based on their reliability, validity and – in most cases – previous use in similar trials. The study primarily measured levels of children's behavioral and emotional difficulties from the main parent's perspective using the Strengths and Difficulties Questionnaire (SDQ) (4 to 17 years), a 25-item questionnaire comprising 5 subscales (each with 5 items scored 0=Not true, 1=Somewhat true, 2=Certainly true) assessing conduct, hyperactivity, emotional difficulties, peer relations and pro-social behavior respectively (sample item: "Often fights with other children or bullies them"). Scores on each subscale

can range from 0 to 10, with higher scores indicating more problems, except for the prosocial subscale where higher scores indicate more prosocial behavior. The Total Difficulties Score, calculated by summing the 20 items in the first four subscales (range 0 to 40; higher scores indicating greater problems), was the primary outcome. The five subscale scores were all secondary outcomes. The parent SDQ also has a brief Impact supplement, which measures the extent to which the child's difficulties with emotions, concentration, behavior, or being able to get on with other people affect their home life, friendships, classroom learning or leisure activities (range 0 to 10; higher scores indicate greater impact). The SDQ has good psychometric properties (e.g. Cronbach's alpha (α) for internal consistency is 0.80 for parent ratings of Total Difficulties) and is regarded as suitable for identifying children with behavioral and emotional difficulties in clinical and community populations (Goodman, 2001; Stone, Otten, Engels, Vermulst and Janssens, 2010).

The Eyberg Child Behavior Inventory (ECBI) is a more sensitive 36-item parent-rated measure of behavior problems exhibited by children aged 2 to 16 years (Eyberg and Ross, 1978; Eyberg, 1980). It has two scales: Intensity ($\alpha = 0.94$) and Problem ($\alpha = 0.93$) (Burns and Patterson 2001). For the Intensity scale (sample item "Teases or provokes other children"), parents indicate the frequency of each of 36 behaviors on a 7-point scale (from 1=Never to 7=Always; total scoring range 36 to 252, with higher scores indicating a greater frequency of behavior problems). The Problem scale assesses whether parents consider the child's behavior to be a problem for themselves (0=No, 1=Yes; scoring range 0 to 36, with higher scores indicating that behaviors are more problematic).

The Ways of Coping Questionnaire (WCQ) is a 66-item self-report measure of coping skills with 50 'critical' items (used in this study) forming an 8-factor structure with acceptable to good internal consistency ($\alpha =$ between 0.61 to 0.79) (Folkman and Lazarus, 1988). The respondent is asked to think of the most stressful situation they have experienced

in the last week and rate how they coped with it using a 4-point scale (from 0=Does not apply or not used to 3=Used a great deal) per item (sample item: “I talked to someone to find out more about the situation”). The eight subscales range in size from 4 to 8 items (higher scores indicate greater use of that skill).

A 10-item subscale of the Adult-Adolescent Parenting Inventory (AAPI-2) (Bavolek and Keene, 2001) was used to measure parents’ empathy. Items (each rated from 1=Strongly agree to 5=Strongly disagree) assess parents’ understanding and recognition of children’s feelings and needs, parental nurturing skills, encouragement of positive growth, communication and management of stressful situations (sample item: “Children should keep their feelings to themselves”). The total score ranges from a possible 10 to 50 (higher scores indicate more negative parenting attitudes). The subscale has good internal consistency ($\alpha = 0.84$).

The Alabama Parenting Questionnaire (APQ) (Frick, 1991) measures parenting practices (sample item: “You praise your child if he/she behaves well”). Three subscales were examined separately in the study: positive involvement with children (10 items, range of possible scores for subscale 10 to 50); use of positive discipline techniques (6 items, range 6 to 30), and consistency in the use of such discipline (6 items, range 6 to 30). Each item refers to a parenting practice and respondents are required to indicate how often they typically use each practice on a 5-item scale (1=Never to 5=Always; higher scores indicate greater use). The included subscales demonstrate acceptable to good internal consistency ($\alpha = 0.67$ to 0.80) (Shelton, Frick and Wootton, 1996).

A short questionnaire was used to gather basic demographic information about the child and their family (adapted from one used in Hutchings et al., 2007), including date of birth, age, gender, ethnicity, special educational needs (SEN) status, education, members of household, relationship quality, family health and financial situation / socio-economic

situation (SES). A brief questionnaire based on the Client Service Receipt Inventory (CSRI) (Beecham, 1995; Chisholm et al., 2000) was used to record the receipt of targeted school services and additional services, detailing the typical length and number of contacts, and also whether any of the services received concerned the child's behavior.

Fidelity monitoring tools covered dose, quality of delivery, adherence to core program components and level of parent engagement. They included an attendance register, a self-report checklist for group facilitators, and an adapted version of the Parent Programme Implementation Checklist (PPIC), an observational tool which provides a global measure of adherence, quality and parent responsiveness (Bywater et al., 2019). The PPIC has 18 items but three items (concerning role play and video clips) that do not apply to *Inspiring Futures* were replaced with two more relevant items,² creating a 17-item version. It has three subscales, each of which is scored by summing the score for each item (from 1=Not at all to 5=Excellent). The numbers of items differs across subscales, resulting in different ranges of possible scores (higher is better): adherence (7 items, range 7 to 35); quality of delivery (8 items, range 8 to 40); and participant responsiveness (2 items, range 2 to 10). A total score is calculated by adding all subscale scores (range 17 to 85). In order to aid comparison with fidelity ratings in other studies, percentage scores can be calculated for each subscale and the total (by dividing the actual score by the maximum possible and multiplying by 100).

2.6 Sample size

The study sought to recruit 248 participants. This sample size is large enough to detect an effect size (Cohen's d) of 0.40 with 80% power at the 2-sided 5% level of significance, allowing for up to 20% loss-to-follow-up.

² These were to assess whether the facilitator (i) supported discussions by using self-disclosure sufficiently and appropriately and (ii) used sufficient and timely perspective statements.

2.7 Randomization

Recruitment and randomization took place between October 2014 and September 2016 and involved six cohorts, where each cohort represented eligible children whose parents were ready to start *Inspiring Futures* and could therefore be randomized. An online central computer-randomization service provided by Exeter Clinical Trials Unit facilitated concealment of the allocation sequence until assignment to trial arm. In the first two cohorts the first 25% of the total number of recruits at each site was allocated by simple randomization (1:1 ratio), and then minimization was used to minimize the imbalance between the intervention and control arms on age (<10 or ≥ 10 years) and gender (as per protocol). The approach was dynamic, with each case randomized as soon as baseline measures were completed. However, in one instance this resulted in a school having a non-viable intervention group of 2 parents despite 8 parents having been randomized. To avoid this happening again, from cohort three onwards (October 2015-) the randomization procedure described in the protocol article was revised. Individuals at each center (usually representing a single school) within a trial site were randomized (ratio 1:1) simultaneously in blocks, once recruitment was complete, by the trial coordinator using a system developed by the CTU. Blocks needed to include at least 8 participants to ensure that the minimum viable intervention group size (4) was reached. If fewer than 8 individuals at a school were recruited, they were grouped with participants from nearby schools in the study (if possible) in order to form a block of ≥ 8 parents. In the event, difficulties with recruitment meant that a minimum block size of 7 participants was permitted, with block sequences only used where four participants were randomly allocated to the intervention arm.

2.8 Blinding

Following randomization, the trial coordinator notified Malachi, the child's family and the referrer about the trial arm allocation. The participating family were therefore *not* blind to allocation. Data collectors were blind to trial arm allocation. The statisticians remained blind to allocation.

2.9 Analysis methods

Baseline and demographic characteristics were summarized. The comparison of the trial arms used an intention-to-treat framework with participants analyzed according to the trial arm to which they were randomized. The primary outcome was the parent-rated SDQ Total Difficulties score at 32 weeks post-randomization. All other outcomes were secondary. The trial arms were compared in crude (unadjusted) analyses presenting the mean difference for continuous outcomes and odds ratio for the single binary outcome (SDQ Total Difficulties ≥ 14). Linear and logistic mixed effects ('multilevel') models (for continuous outcomes and the binary outcome respectively) were used, and included random effects for parenting group (cluster) to allow for clustering by group in the intervention arm (Flight et al., 2016). Adjustments were made in these comparisons for the stratification factors (site, age and gender), ethnicity, SES, SEN, parent education, parent marital status and the baseline score on the outcome being analyzed. The adjusted analyses were considered the main analyses. Tests of interaction were used to examine whether the effect of the intervention on the primary outcome differs across categories based on age (<10 versus ≥ 10 years), gender, ethnicity and level of difficulties on the baseline SDQ Total Difficulties score (borderline [14 to 16] vs. abnormal [≥ 17]). An ancillary analysis using a complier average causal effect analysis (CACE) (Hewitt, Torgerson and Miles, 2006; Dunn and Bentall, 2007) was undertaken to quantify the intervention effect on the primary outcome among children whose parents attend all group sessions. The CACE analysis compares 'compliers' in the

intervention arm (those who ‘comply’ with the intervention offered, in this case attending all group sessions) with a comparable group in the control arm (those who would have complied had they – counterfactually – been offered the intervention). The findings presented were based on analyses of 20 multiply imputed datasets to handle missing data; the data were assumed to be missing at random. All outcome analyses were carried out using R software 3.5.0. (R Core Team, 2018).

Implementation fidelity was summarized using descriptive statistics. Attendance registers were used to record dose as the number of group and one-to-one sessions attended as well as the number of participants who attended group sessions. They were also used to quantify participant responsiveness by dividing their total applications scores by the number of sessions attended. Self-reported adherence from facilitator checklists was calculated as a percentage per session, taking into account scores for both facilitators. Adherence, quality and participant responsiveness from the PPIC were calculated as total and percentage scores per session.³ Analyses of fidelity were carried out using SPSS Version 25 software.

3. Results

3.1 Flow through the study

The CONSORT diagram (Moher et al., 2010) depicts the flow of referral, recruitment and retention in the trial (Figure 1). It shows that 327 children were assessed as being eligible for the intervention. Of these, 63 (19.3%) did not proceed to randomization because they actively withdrew, or could not be contacted for additional baseline data collection, or were subsequently found to be ineligible, or for other reasons. Active withdrawal mainly concerned parents’ inability to participate in the intervention (e.g. work/study, other

³ The possible range for percentage scores is 20% to 100%.

commitments, poor health, lack of childcare). The remaining 264 children, who came from 28 schools, were randomly allocated to the intervention (N=134) and control (N=130) arms. Within the intervention arm, 93 out of 122 participants (parents) for whom information is available (76.2%) received the intervention (meaning that they attended at least one group session). Of the 130 participants in the control arm, five (3.8%) received some of the intervention (Inspiring Futures). By 32 weeks post-randomization, 69 participants had dropped out of the study (withdrawn or unable to contact), meaning that overall attrition was 26.1%, with a slightly higher rate in the intervention arm (28.4% (38/134)) than in the control arm (23.8% (31/130)).

3.2 Participant characteristics and attrition

Table 1 shows baseline socio-demographic characteristics of all trial participants by trial arm, those lost to follow-up (withdrawn or unable to contact) and those remaining in the trial to the end.

Supplementary Table S1 shows baseline scores for the outcomes for all trial participants (by arm), those lost to follow-up (withdrawn or unable to contact) and those remaining in the trial to the end. The strong equivalence between trial arms on baseline assessments of outcomes (including subscales) was broadly retained after attrition. It is notable that mean baseline scores were in the clinical range for SDQ Total Difficulties (≥ 17) and ECBI 'Intensity' (≥ 131) and 'Problem' (≥ 15).

3.3 Effect on outcomes

Outcome analyses were conducted on all 264 participants as randomized. Table 2 shows the unadjusted and adjusted mean differences, and adjusted standardized mean differences (SMD), at 32 weeks post-randomization (results for 16 weeks post-randomization are in

Supplementary Table S2). There was no statistically significant effect on the primary outcome of parent-rated SDQ Total Difficulties score (adjusted SMD -0.07; 95% CI -0.3 to 0.16; $p = 0.54$). Apart from parent use of positive reappraisal as a coping skill at 16 weeks post-randomization (adjusted SMD 0.24; 95% CI: 0.01 to 0.46; $p = 0.04$), there were no statistically significant effects on child or parent secondary outcomes at either follow-up.

3.4 Sub-group analyses

Exploratory analyses found no statistically significant moderator effects for the primary outcome (Supplementary Table S3).

3.5 Other services received

There were no statistically significant differences between the trial arms for receipt of any school-based services or additional services, and no differences on whether additional services were received in relation to children's behavior (Supplementary Tables S4 and S5).

3.6 Missing data

Baseline variables were largely non-missing. The amount of missing data increased at follow-up (summarised in Figure 1). The primary outcome was 26.1% missing at 32 weeks post-randomization, due to loss of contact or withdrawal of participants.

3.7 Implementation fidelity

Ten group facilitators were involved in the study (2 in Somerset, 8 in Birmingham). Supplementary Table S6 shows indicators of implementation fidelity as reported by group facilitators. Of the 134 participants allocated to the intervention arm, data on group attendance from facilitators were available for 97 (72.4%). These data apply to 23 of the 31

groups that were run (it is not possible to say if the same pattern applied to the other eight groups). The mean number of group sessions attended per participant was 6.1, which translates into a mean attendance rate of 59.6% of available group sessions; this includes 26 participants (26.8%) who attended no sessions. The mean group size for participants attending groups was 3.3. According to facilitator report, mean self-reported adherence was 91.2%. Mean participant responsiveness, measured using application scores, was 3.0 (out of 5).

Thirty-nine of a possible 70 group sessions were video-recorded (18 for Session 4, 15 for Session 8, 1 for Session 6, 5 unknown). Reasons for non-recordings were participants refusing consent or technical difficulties with recording equipment. Of the 39 videos, 31 videos (44.3% of the total) were coded using the adapted PPIC (16 for session 4, 14 for session 8, 1 for session 6). Reasons for videos not being coded were videos going missing or technical difficulties meaning that they could not be viewed. The recordings were initially coded by two raters (one researcher, one Malachi) who worked independently and then agreed final scores through discussion. Subsequent recordings were coded by one rater only (researcher). Interrater reliability for the 14 double-coded videos showed a mean percentage agreement of 86%⁴ (the PPIC manual stipulates a minimum of 80% agreement between coders on the total score). A good level of agreement was also found between the subscale ratings for adherence, quality, and the total score, but agreement was poor for the participant responsiveness subscale (the Malachi rater scored higher on average than the researcher). Two-way mixed model ICCs with absolute agreement were used to assess agreement between the two raters. The single measures intraclass correlational coefficients (ICC) were as follows for each subscale: adherence ICC = 0.73, 95% confidence interval = 0.36 to 0.90

⁴ Calculated by taking the smaller of the two raters' total scores for a given session and dividing it by the larger score (then *100).

($p < 0.001$), quality ICC = 0.70, 95% confidence interval = 0.30 to 0.89 ($p = 0.002$), participant responsiveness ICC = 0.22, 95% confidence interval = -0.19 to 0.62 ($p = 0.161$), total score ICC = 0.70, 95% confidence interval = 0.29 to 0.90 ($p = 0.002$). Fidelity scores were calculated for the 14 double-coded videos and the 17 single-coded videos. The total mean fidelity score was 73%, with subscales ranging from 72% to 76% (Table 3).

Of the 132 participants for whom data on one-to-one sessions were available, eight (6.1%) took part in one or more sessions. The mean number of one-to-one sessions they received was 5.9 (median 6; range 2 to 12).

3.9 CACE analysis

The CACE analysis indicated no statistically significant effect (adjusted MD = -1.08; 95% CI -4.44 to 2.26; $p = 0.52$) on the primary outcome for the parents who attended all sessions (Supplementary Table S7).

4. Discussion

This study evaluated the effectiveness of an established UK parenting intervention that focuses primarily on helping parents reflect on how their past experiences affect their relationship and interactions with their children and less on skills development for behavior change. It goes some way towards addressing the knowledge gap regarding the effectiveness of parenting programs developed and practised in the UK. There was no statistically significant effect on the primary outcome, parent-rated SDQ Total Difficulties score at 32 weeks post-randomization, and only one statistically significant effect favoring the program for 40 secondary outcomes spread across 16- and 32-weeks post-randomization. There were no sub-group effects for the primary outcome and no consistent pattern of non-significant

trends favoring the intervention arm. The results may be generalized to similar populations in high-income countries.

Potential reasons for the null results should be explored. The first of these is implementation fidelity given the known association between fidelity and outcomes in prevention and early intervention (Durlak and DuPre, 2008), including for parenting programs (Axford et al., 2017). Based on independent observation, mean adherence, quality and participant responsiveness scores fell short of high fidelity measured using a threshold of 80%, although they considerably exceeded the 50% regarded by some as indicating low fidelity (Borrelli, 2011). Regarding dose, a quarter of participants in the *Inspiring Futures* arm did not attend any group session, and just over two-fifths of available group sessions were not attended. This may relate to the need for a small number of parents to attend a group in a neighboring school; while the notion might have seemed acceptable in theory when the parents concerned agreed to it, the reality of needing to travel further (potentially) or spending time with unfamiliar people in an unfamiliar setting may have proved off-putting. Another explanation for non-attendance could be an unwillingness to discuss sensitive topics such as early traumatic experiences. A range of mechanisms could be at work here: fears about the confidentiality of the information disclosed to the facilitator and group members (Koerting et al., 2013); failure to establish a quality therapeutic alliance, or a perception that the facilitator was judgemental or not empathetic (Weisenmuller and Hilton, 2020); or a belief that one's own early experiences were irrelevant to their child's behaviour problems (Ingoldsby, 2010). These may have been experienced differentially by parents depending on how they perceived their situation and that of others; for example, parents with early childhood trauma may have been reticent to disclose to parents lacking such experiences, equally the latter may have considered this focus unimportant. All that said, the levels of intervention drop-out are not unusual for parenting interventions (Whittaker and Cowley,

2012), and the CACE analysis found no effect for compliers (those parents attending all available sessions).

Mean group size was low, and in some sessions attendance fell to one (three groups) or two (11 groups) parents, meaning that technically in those instances the intervention ceased to be a 'group' (defined as three or more people plus one or more facilitators). An analysis of group effects in another parenting program indicates that it is group composition rather than size that affects outcomes (Byrne et al., 2013). However, in addition to making the intervention more costly per head, it is hard to imagine that small group size in Inspiring Futures did not undermine the intra-and inter-personal change processes instigated by group-based social interactions. These co-occur and interact dynamically, and include learning from one another and gaining self-insight, influencing one another's beliefs, motivations and behaviors, providing support (practical, emotional, informational) and being accountable for progress against goals set (Borek et al., 2019).

Only one in 20 intervention arm participants received any one-to-one support, which might seem significant given the high levels of child need at baseline and the fact that therapeutic parenting programs targeting high-risk groups often include – and in some cases solely comprise – this level of intensive support. Against that, a recent moderator analysis of individual participant data from 13 trials of the Incredible Years group-based intervention found that reductions in child conduct problems were greater among those with more severe conduct problems, that children of mothers with more depressive symptoms also benefited more (Leijten et al., 2020) and there were no differential effects by family disadvantage (indicated by poverty, lone parenthood, teenage parenthood household joblessness, or low education) (Gardner et al., 2019a). This suggests that it is less delivery format than intervention content that matters as regards impact on high-need groups.

While it cannot be concluded, therefore, that the fidelity data indicate flawed execution, there is clearly considerable scope for improving fidelity. A recent study investigating the effects of the various components of implementation integrity on the effectiveness of several parenting programs reported that implementation quality (adherence and quality of delivery) did *not* influence effects on parent and child outcomes (Giannotta et al., 2019). Rather, participant involvement was associated with improvements in parenting and child conduct, and parents' perceptions of their leaders as supportive and understanding were associated with parents' responsiveness and attendance. Applying these findings to *Inspiring Futures* would entail ensuring that facilitator training and supervision are fit for purpose, and that facilitators are a 'good fit' for the program with a strong personal belief in program content. In particular, facilitators need suitable characteristics and training to enable them to foster the change processes critical to intervention outcomes (Borek et al., 2019). Efforts to strengthen the fidelity of *Inspiring Futures* delivery will also require developing and applying a tailored, multifaceted and precisely defined implementation strategy (Proctor et al., 2013; Lewis et al., 2018). This should cover activities relating to different determinants of effective implementation, from the implementation process itself to the immediate organizational and wider policy context, and articulate who needs to do which activity, when, how much and for what purpose in order to address known barriers.

A second possible explanation for the lack of an effect concerns the families targeted by the intervention. With some exceptions (e.g. Adkins et al., 2018), therapeutic or relationship-enhancing parenting interventions have tended to focus on the parents of babies, toddlers and those of pre-school age (Lieberman and Van Horn, 2008; Barlow, Bennett, Midgley, Larkin and Yinghui, 2015; Suchman et al., 2017). It seems plausible that working with parents of older children – as in *Inspiring Futures* – is harder because parenting behavior and critical aspects of child development are less malleable. However, recent meta-analyses show that

parenting programs for older children are equally as effective for reducing disruptive behavior as those delivered earlier in childhood (Gardner et al., 2019b). A more likely explanation relates to the level of need of the children and parents. Participants in *Inspiring Futures* are selected based on the child's behavior and emotional well-being, not an assessment of parenting behaviors or known risk factors for relationship difficulties. While this is common for behavioral parenting programs, it points to a possible disconnect here between intervention strategy and participant profile. Specifically, families in which parent-child relationships were less distressed may not have needed relational enhancement, and therefore not found that program content met their needs. Similarly, some parents entering the program may not have experienced significant deprivation or trauma in early or later life or have had difficulties mentalizing or responding sensitively to their child's needs. Targeted therapeutic programs tend to select participants based on indicators such as risk of disorganized attachment owing to child maltreatment, parent mental health problems and parent history of domestic violence, crime, substance use or childhood trauma. This approach could be adopted by Malachi.

A third possible explanation for the null results is that the intervention did not adequately address the risk and protective factors associated with children's behavioral and emotional difficulties. There is no evidence of effects on parent outcomes, namely parenting skills, empathy and coping skills, which in turn – and based on the logic model – would help to explain the lack of effect on child outcomes. This should be viewed in the context of the wider evidence base on parenting programs. Evaluations of therapeutic interventions have found positive effects on parent reflective functioning, sensitive parenting, parent stress and, to a lesser extent, child attachment and externalizing problems (Luyten et al., 2017; Byrne et al. 2019). However, collectively the evidence is somewhat mixed and limited by the relative dearth of trials and a failure always to measure child outcomes. For example, Mercer (2015)

examined studies of one such program, concluding that it is highly plausible and acceptable (to parents) but “weakly supported” (p.382) by research evidence, and that this is comparable to the level of research support for similar programs. Although there are few head-to-head trials comparing behavioral with non-behavioral parenting interventions, evidence for the impact of the former on child conduct and – to a lesser extent – emotional problems is much stronger (Furlong et al., 2012; NICE, 2013; Mazzuchelli and Sanders, 2014; Scott and Gardner, 2015; Leijten et al., 2018a), and – particularly relevant given the target group of *Inspiring Futures* – covers interventions for school-aged as well as pre-school children. A recent meta-analysis showed that the parenting techniques associated with stronger parenting program effects on disruptive child behavior are positive reinforcement, praise and non-violent disciplining (e.g. natural/logical consequences), all of which are typically taught in behavioral programs, whereas relationship-building and parental self-management, both of which feature more strongly in therapeutic programs such as *Inspiring Futures*, are associated with stronger effects in treatment but weaker effects in prevention (Leijten et al., 2018b, 2019). In short, both preventive and clinical populations need behavior management skills, but only clinical populations need relationship-building as well (rather than instead). There is a good case, therefore, for a more personalized approach whereby a program such as *Inspiring Futures* has a much stronger behavioral focus for *all* families served and only adds the more therapeutic relationship-building elements – as part of an integrative approach – in those cases where children’s behavior problems are more severe and parent-child relationships more distressed.

There is a need for more robust impact evaluations of therapeutic parenting programs, notably large well-conducted trials with suitable measures (including for child outcomes). The relative lack of such studies to date calls into question the decision currently to offer a program such as *Inspiring Futures* as an early community-based intervention to parents of

children with behavior and emotional difficulties when evidence-based behavioral parent training programs may be more suitable, and particularly absent efforts to target based on parental risk factors. The apparent growing popularity of therapeutic parenting programs has been attributed to criticisms that behavioral interventions emphasize control, punishment and compliance, diminish a child's intrinsic motivation, fail to help parents recognize and deal with their own and their child's mental states, and ignore how relational histories influence child behavior (Coyne, 2013). However, Mazzuchelli and Sanders (2014) offer a strong rebuttal of each of these critiques, and note that many well-known behavioral parent training programs, such as Incredible Years and Triple P, include intervention components designed to promote secure attachment and positive parent-child relationships in addition to their behavioral components.

4.1 Strengths and limitations

The present study has several strengths. It was possible to recruit slightly more participants than the target sample size (264 participants rather than 248), meaning that despite higher-than-predicted attrition (26% cf. 20%) the required number of participants (196 compared with the hoped-for 198) remained. Other strengths include: the use of tried-and-tested measures; the use of two measures of child behavior – the parent-reported SDQ and the more sensitive ECBI; and the reasonably detailed data on other service use by children and families. There are also limitations. Some of these concern measures. It is sometimes considered not to be best methodological practice to use the same measure for the primary outcome and screening participants for eligibility. There were no measures of attachment style and parent reflective functioning, a focus of the intervention, and no observational measure of child behavior (owing to resources, and common in many pragmatic parenting program trials). Another set of limitations concern attrition and missing data. Attrition was

higher than in most (but not all) trials in a review of group-based parenting programs (Furlong et al., 2012), although the comparison may be misleading given that 32-week data collection point falls several months after the end of the intervention for most participants (only 6 intervention arm participants received one-to-one support after the group element, which finished before the 16-week post-randomization data collection point). In mitigation, there was no differential attrition and multiple imputation was used so that all study participants were included in the outcome analyses. Missing data for implementation fidelity from facilitator self-completion records (including participant attendance) and group session video-recordings have been mentioned; the latter could be addressed in part through providing better quality equipment and standardized training in its use, supported by written guidance, although the increased use of video-recording facilities on tablets and mobile phones in recent years arguably makes the use of specialist technology redundant. The number of parents offered individual support and therefore the proportion (though not number) who took it up are unknown. Other limitations include five participants in the control arm receiving some of the intervention, poor agreement between raters on the PPIC participant responsiveness scale, and a lack of long-term follow-up, although delayed effects are considered unlikely given the lack of observed effects.

5. Conclusions

Based on the findings of this trial, the *Inspiring Futures* therapeutic parenting intervention is not more effective than services as usual at improving targeted outcomes. This may be related in part to implementation issues, and certainly there is scope to improve fidelity by: enhancing facilitator recruitment, training and supervision; getting parents along to at least one group session, and retaining them thereafter; and improving group facilitator adherence and quality (which will hopefully increase participant responsiveness). However, unlike

evidence-based behavioral parent training programs, it is not clear that a more therapeutic parenting program such as *Inspiring Futures* is able to improve caregiving practices sufficiently to prevent or reduce child conduct or emotional difficulties, particularly when it is not targeted at parents who have experienced trauma or deprivation in childhood or later life. For primary school children with elevated behavior or emotional difficulties, therefore, and until there is more research demonstrating the effectiveness of non-behavioral programs, the present study suggests that it is preferable in a community-based (as opposed to treatment) setting to implement a proven behavioral parenting program with some attachment components as opposed to an attachment-orientated one with some behavioral elements.

References

- Adamson, P., Bradshaw, J., Hoelscher, P., & Richardson, D. (2007). *Child Poverty in Perspective: An Overview of Child Well-being in Rich Countries*. Florence, Italy: Unicef Innocenti Research.
- Adkins, T., Luyten, P., & Fonagy, P. (2018) Development and preliminary evaluation of Family Minds: a mentalization-based psychoeducation program for foster parents. *Journal of Child and Family Studies* 27 (8), 2519-2532.
- Axford, N., Bywater, T., Blower, S., Berry, V., Baker, V. & Morpeth, L. (2017) Critical factors in the implementation of parenting programmes: fidelity, adaptation and promoting quality. In Dixon, L., Perkins, D. F., Hamilton-Giachritsis, C. and Craig, L. A. (Eds) *The Wiley Handbook of What Works in Child Maltreatment: An Evidence-Based Approach to Assessment and Intervention in Child Protection*. Chichester, Wiley.
- Bailey, J., Hill, K., Oesterle, S., & Hawkins, J. (2009). Parenting practices and problem behavior across three generations: Monitoring, harsh discipline, and drug use in the intergenerational transmission of externalizing behavior. *Developmental Psychology*, 45(5), 1214-1226.
- Barlow, J., Bennett, C., Midgley, N., Larkin, S., & Yinghui, W. (2015). Parent-infant psychotherapy

- for improving parental and infant mental health. *Cochrane Database of Systematic Reviews*.
- Barlow, J., Smailagic, N., Bennett, C., Huband, N., Jones, H., & Coren, E. (2012). Individual and group based parenting programmes for improving psychosocial outcomes for teenage parents and their children (Review). *John Wiley & Sons, Ltd., The Cochrane Collaboration* (6).
- Bavolek, S., & Keene, R. (2001). *Adult-Adolescent Parenting Inventory AAPI-2: Administration and Development Handbook*. Park City, UT: Family Developments Resources.
- Beecham, J. (1995). *The Client Service Receipt Inventory (CSRI). Report prepared for the Concerted Action Research Program*. CEMH Working Paper 053. Brussels: European Commission.
- Bonin, E-M., Stevens, M., Beecham, J., Byford, S., & Parsonage, M. (2011). Costs and longer-term savings of parenting programmes for the prevention of persistent conduct disorder: a modelling study. *BMC Public Health*, 11(803).
- Borek, A. J., Abraham, C., Greaves, C. J., Gillison, F., Tarrant, M., Morgan-Trimmer, S., McCabe, R., & Smith, J. R. (2019). Identifying change processes in group-based health behaviour-change interventions: development of the mechanisms of action in group-based interventions (MAGI) framework. *Health Psychology Review*, 13(3), 227-247.
- Borrelli, B. (2011) The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, 71(s1), S52-S63.
- Breslau, J., Lane, M., Sampson, N., & Kessler, R. C. (2008). Mental disorders and subsequent educational attainment in a US national sample. *Journal of Psychiatric Research*, 42(9), 708-716.
- Burns, G. L., & Patterson, D. R. (2001). Normative Data on the Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory: Parent and Teacher Rating Scales of Disruptive Behavior Problems in Children and Adolescents. *Child & Family Behavior Therapy*, 23(1), 15-28.
- Byrne, G., Sled, M., Midgley, N., Fearon, P., Mein, C., Bateman, A. & Fonagy, P. (2019)

Lighthouse Parenting Programme: description and pilot evaluation of mentalization-based treatment to address child maltreatment. *Clinical Child Psychology and Psychiatry*, 24(4), 680-693.

Byrne, S., Salmela-Aro, K., Read, S., & Rodrigo, M. J. (2013). Individual and group effects in a

community-based implementation of a positive parenting program. *Research on Social Work Practice*, 23(1), 46-56.

Bywater, T., Gridley, N., Berry, V., Blower, S., & Tobin, K. (2019). The Parent Programme Implementation Checklist (PPIC): the development and testing of an objective measure of skills and fidelity for the delivery of parent programmes. *Child Care in Practice*, 25(3), 281-309.

Calkins, S. D., & Keane, S. P. (2009). Developmental origins of early antisocial behavior.

Development and Psychopathology, 21(4), 1095-1109.

Chisholm, D., Knapp, M. R. J., Knudsen, H. C., Amaddeo, F., Gaité, L., & van Wijngaarden, B.

(2000). Client Socio-Demographic and Service Receipt Inventory – European Version: development of an instrument for international research: EPSILON Study 5. *British Journal of Psychiatry*, 177(39), 28-33.

Coyne, J. (2013). Parenting from the outside-in: reflections on parent-training during a potential paradigm shift. *Australian Psychologist* 48, 379-387.

De Shazer, S., Dolan, Y., Korman, H., Trepper, T., McCollum, E., & Kim Berg, I. (2007). *More than Miracles: The State of the Art of Solution-Focused Brief Therapy*. Philadelphia: Haworth Press.

Doyle, O., Hegarty, O., & Owens, C. (2018) Population-based system of parenting support to reduce the prevalence of child social, emotional, and behavioural problems: difference-in-differences study. *Prevention Science* 19(6), 772-781.

- Dunn, G., & Bentall, R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine*, 26(26), 4719-4745.
- Eyberg, S. M., & Ross, A. W. (1978). Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child Psychology*, 7(2), 113-116.
doi:10.1080/15374417809532835
- Farrington, D., & Welsh, B. (2007). *Saving children from a life of crime: Early risk factors and effective interventions*. Oxford: Oxford University Press.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . . Marks, J. S. (1998). Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*, 14(4), 245-258.
- Fletcher, J. M. (2010). Adolescent depression and educational attainment: results using sibling fixed effects. *Health Economics*, 19(7), 855-871.
- Flight, L., Allison, A., Dimairo, M., Lee, E., Mandefield, L. and Walters, S. J. (2016) Recommendations for the analysis of individually randomised controlled trials with clustering in one arm – a case of continuous outcomes. *BMC Medical Research Methodology* 16(1):165
- Folkman, S., & Lazarus, R. (1988). *Ways of Coping Questionnaire Manual*. Palo Alto, California: Consulting Psychologists Press.
- Frick, P. (1991). *The Alabama Parenting Questionnaire*. Unpublished rating scale: University of Alabama.
- Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S., & Donnelly, M. (2012). Behavioral and cognitive-behavioral group based parenting programs for early-onset conduct problems in children aged 3 to 12 years (Review). *John Wiley & Sons, Ltd., The Cochrane*

Collaboration (2).

- Gardner, F., Burton, J., & Klimes, I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry*, 47(11), 1123-1132.
- Gardner, F., Leijten, P., Mann, J., Landau, S., Harris, V., Beecham, J., et al. (2017). 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 Research*, 5(10).
- Gardner, F., Leijten, P., Harris, V., Mann, J., Hutchings, J., Beecham, J., Bobin-E-M., Berry, V., McGilloway, S., Gaspar, M., João Seabra-Santos, M., & Orobio de Castro, B. (2019a). Equity effects of parenting interventions for child conduct problems: a pan-European individual participant data meta-analysis. *Lancet Psychiatry*, 6(6), 518-527.
- Gardner, F., Leijten, P., Melendez-Torres, G. J., Landau, S., Harris, V., Mann, J., et al. (2019b). The earlier the better? Individual participant data and traditional meta-analysis of age effects of parenting interventions. *Child Development*, 90(1), 7-19.
- Giannotta, F., Özdemir, M. & Stattin, H. (2019). The implementation integrity of parenting programs: which aspects are most important? *Child Youth Care Forum* 48, 917-933.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581-586.
- Goodman, R. (2001). Psychometric Properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.
- Hewitt, C. E., Torgerson, D. J., & Miles, J. N. V. (2006). Is there another way to take account of noncompliance in randomized controlled trials? *Canadian Medical Association Journal*, 175(4), 347-347.
- Hickman G. (2007). *Parents' understanding of the impact of the Malachi Community trust*

intervention on their parenting behaviour: an interpretive phenomenological analysis; Masters dissertation.

- Hoeve, M., Dubas, J. S., Eichelsheim, V. I., van der Laan, P. H., Smeenk, W., & Gerris, J. R. M. (2009). The Relationship Between Parenting and Delinquency: A Meta-analysis. *Journal of Abnormal Child Psychology*, 37(6), 749-775.
- Hutchings, J., Bywater, T., Daley, D., Gardner, F., Whitaker, C., Jones, K., . . . Edwards, R. T. (2007). Parenting intervention in Sure Start services for children at risk of developing conduct disorder: pragmatic randomised controlled trial. *BMJ*, 334(7595), 678.
- Ingoldsby, E. M. (2010). Review of interventions to improve family engagement and retention in parent and child mental health programs. *Journal of Child and Family Studies*, 19(5), 629-645.
- Kalland, M., Fagerlund, A., von Koskull, M., & Pajulo, M. (2016). Families First: the development of a new mentalization-based group intervention for first-time parents to promote child development and family health. *Primary Care Research & Development*, 17(1), 3-17.
- Koerting, J., Smith, E., Knowles, M. M., Latter, S., Elsey H., McCann, D. C., Thompson, M., & Sonuga-Barke, E. J. (2013). Barriers to and facilitators of, parenting programmes for childhood behaviour problems: a qualitative synthesis of studies of parents' and professionals' perceptions. *European Child & Adolescent Psychiatry*, 22(11), 653-670.
- Leijten, P., Gardner, F., Landau, S., Harris, V., Mann, J., Hutchings, J., Beecham, J., Bonin, E-M., & Scott, S. (2018a). Research review: harnessing the power of individual participant data in a meta-analysis of the benefits and harms of the Incredible Years parenting program. *Journal of Child Psychology and Psychiatry*, 59(2), 99-109.
- Leijten, P., Gardner, F., Melendez-Torres, G. J., van Aar, J., Schulz, S., & Overbeek, G. (2018b). Are relationship enhancement and behavior management “the golden

- couple” for disruptive child behavior? Two meta-analyses. *Child Development*, 89(6), 1970-1982.
- Leijten, P., Gardner, F., Melendez-Torres, G. J., Van Aar, J., Hutchings, J., Schultz, S., ... & Overbeek, G. (2019). What to teach parents to reduce disruptive child behavior: two meta-analyses of parenting program components. *Journal of the American Academy of Child and Adolescent Psychiatry*, 58(2), 180-190.
- Leijten, P., Scott, S., Landau, S., Harris, V., Mann, J., Hutchings, J., Beecham, J., & Gardner, F. (2020) Individual participant data meta-analysis: impact of conduct problem severity, comorbid Attention-deficit/Hyperactivity Disorder and emotional problems, and maternal depression on parenting program effects. *Journal of the American Academy of Child & Adolescent Psychiatry*. doi: 10.1016/j.jaac.2020.01.023.
- Lewis, C. C., Scott, K., & Marriott, B. R. (2018). A methodology for generating a tailored implementation blueprint: an exemplar from a youth residential setting. *Implementation Science*, 13, 68-81.
- Lewis, J. (2011). Parenting programmes in England: policy development and implementation issues, 2005-2010. *Journal of Social Welfare and Family Law*, 33(2), 107-121.
- Lieberman, A., & Van Horn, P. (2008). *Psychotherapy with Infants and Young Children: Repairing the Effects of Stress and Trauma on Early Attachment*. New York, NY: Guilford Press.
- Lindsay, G., & Strand, S. (2013). Evaluation of the national roll-out of parenting programmes across England: the parenting early intervention programme (PEIP). *BMC Public Health*, 13:972.
- Luyten, P., Nijssens, L., Fonagy, P., & Mayes, L. C. (2017). Parental reflective functioning: theory, research, and clinical applications. *The Psychoanalytic Study of the Child*, 70(1), 174-199.
- Mazzuchelli, T. G., & Sanders, M. R. (2014) Parenting from the outside-in: a paradigm shift in parent training? *Behaviour Change*, 31(2), 102-109.
- McGilloway, S., Mhaille, G. N., Bywater, T., Furlong, M., Leckey, Y., Kelly, P., Comiskey, C., &

- Donnelly, M. (2012). A parenting intervention for childhood behavioural problems: a randomized controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology, 80*(1), 116-127.
- Menting, A. T. A., de Castro, B. O., & Matthys, W. (2013). Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behaviour: a meta-analytic review. *Clinical Psychology Review, 33*(8), 901-913.
- Mercer, J. (2015) Examining Circle of Security™: a review of research and theory. *Research on Social Work Practice, 25*(3), 382-392.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., . . . Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ, 340*:c869.
- Morpeth, L., Blower, S., Tobin, K., Taylor, R. S., Bywater, T., Edwards, R. T., Axford, N., Lehtonen, M., Jones, C., & Berry, V. (2017). The effectiveness of the Incredible Years pre-school parenting programme in the UK: a pragmatic randomised controlled trial. *Child Care in Practice, 23*(2), 141-161.
- Murphy, A., Steele, M., Dube, S. R., Bate, J., Bonuck, K., Meissner, P., . . . Steele, H. (2014). Adverse Childhood Experiences (ACEs) Questionnaire and Adult Attachment Interview (AAI): Implications for parent child relationships. *Child Abuse & Neglect, 38*(2), 224-233.
- Nagin, D., & Tremblay, R. E. (1999). Trajectories of Boys' Physical Aggression, Opposition, and Hyperactivity on the Path to Physically Violent and Nonviolent Juvenile Delinquency. *Child Development, 70*(5), 1181-1196.
- NICE (National Institute for Health and Care Excellence) (2013) *Recognition, Intervention and Management of Antisocial Behaviour and Conduct Disorders in Children and Young People*. London: National Institute for Health and Care Excellence.
- Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: a global

- public-health challenge. *The Lancet*, 369(9569), 1302-1313.
- Proctor, E. K., Powell, B. J., & McMillen, J. C. (2013). Implementation strategies: recommendations for specifying and reporting. *Implementation Science*, 8, 139-150.
- R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org>
- Roza, S. J., Hofstra, M. B., van der Ende, J., & Verhulst, F. C. (2003). Stable Prediction of Mood and Anxiety Disorders Based on Behavioral and Emotional Problems in Childhood: A 14-Year Follow-Up During Childhood, Adolescence, and Young Adulthood. *American Journal of Psychiatry*, 160(12), 2116-2121.
- Sadler, K., Vizard, T., Ford, T., Marcheselli, F., Pearce, N., Mandalia, D., et al. (2018). *Mental Health of Children and Young People in England, 2017: Summary of Key Findings*. London: NHS Digital.
- Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple-P Positive Parenting Program: a systematic review and meta-analysis of a multi-level system of parenting support. *Clinical Psychology Review*, 34(4), 337-357.
- Scott, S., Sylva, K., Doolan, M., Price, J., Jacobs, B., Crook, C., & Landau, S. (2010). Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: the SPOKES project. *Journal of Child Psychology and Psychiatry*, 51(1), 48-57.
- Scott, S. and Gardner, F. (2015). Parenting programs. In: Thapar, A., Pine, D. S., Leckman, J. F., Scott, S., Snowling, M. J. and Taylor, E. (Eds) *Rutter's Child and Adolescent Psychiatry (Sixth Edition)*. Chichester: Wiley.
- Shelton, K. K., Frick, P. J., & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, 25(3), 317-329.
- Slade, A. (2007). Reflective parenting programs: theory and development. *Psychoanalytic Inquiry* 26(4). 640-657.

- Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review. *Clinical Child and Family Psychology Review*, 13(3), 254-274.
- Weisenmuller, C., & Hilton, D. (2020). Barriers to access, implementation, and utilization of parenting interventions: considerations for research and clinical applications. *American Psychologist*, doi: 10.1037/amp0000613.
- Whittaker, K. A., & Cowley, S. (2012). An effective programme is not enough: a review of factors associated with poor attendance and engagement with parenting support programmes. *Children & Society*, 26(2), 138-149.
- Zwi, M., Jones, H., Thorgaard, C., York, A., & Dennis, J. A. (2011) Parent training interventions for Attention Deficit Hyperactivity Disorder (ADHD) in children aged 5 to 18 years. *Cochrane Database of Systematic Reviews*, Issue 12, Article No: CD003018.

Author statement

Nick Axford: Conceptualization, Methodology, Formal analysis, Writing – Original Draft, Writing – Review & Editing, Supervision **Gretchen Bjornstad:** Methodology, Formal analysis, Writing – Original Draft, Visualization, Writing – Review & Editing, Supervision **Justin Matthews:** Methodology, Software, Formal analysis, Visualization, Writing – Original Draft, Writing – Review & Editing **Sarah Heilmann:** Methodology, Investigation, Supervision, Project administration **Anam Raja:** Investigation, Supervision, Project administration **Obioha C. Ukoumunne:** Methodology, Software, Formal analysis, Visualization, Supervision, Writing – Original Draft, Writing – Review & Editing **Vashti Berry:** Conceptualization, Methodology, Writing – Review & Editing **Tom Wilkinson:**

Methodology, Investigation, Supervision, Project administration **Luke Timmons:**

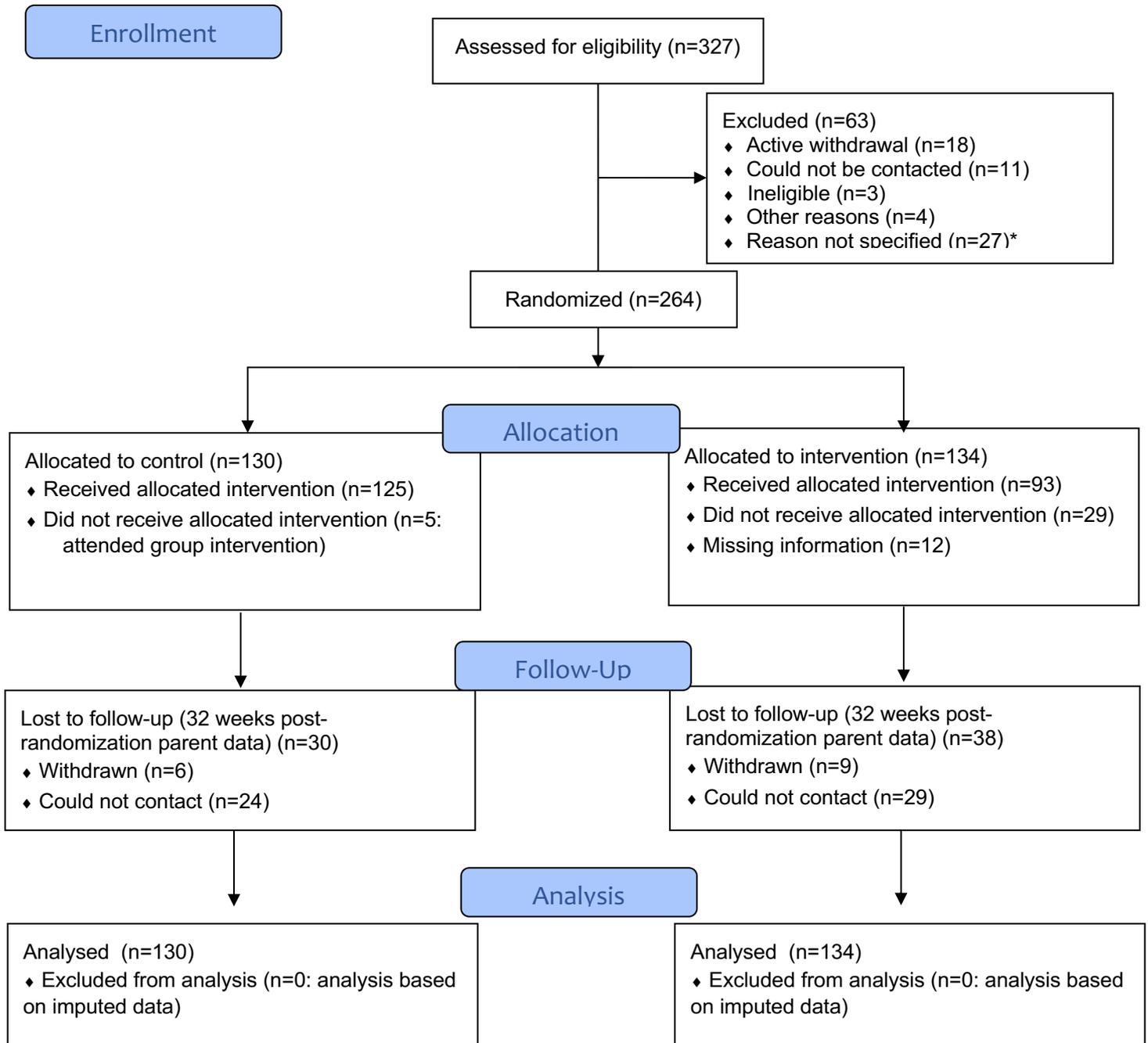
Investigation, Supervision, Project administration **Tim Hobbs:** Conceptualization, Funding

acquisition **Tim Eames:** Methodology, Software **Angeliki Kallitsoglou:** Methodology,

Supervision **Sarah Blower:** Conceptualization, Methodology, Writing – Review & Editing

Georgina Warner: Methodology, Writing – Review & Editing, Supervision

Figure 1: CONSORT 2010 flow diagram



*Includes 24 participants who were not baselined, suggesting that the reason for drop-out was inability to contact.

Table 1: Baseline demographic characteristics by trial arm status of all participants in the trial, those lost to follow-up 32 weeks post-randomization (withdrawn or unable to contact), and those remaining in the trial to the end. Values are percentages (numbers).

Baseline variable	All participants		Participants lost to follow-up		Remaining participants	
	Intervention (N=122 to 134)	Control (N=114 to 130)	Intervention (N=39 to 44)	Control (N=23 to 30)	Intervention (N=82 to 90)	Control (N=91 to 100)
Child age						
<9 years	59.7 (80)	52.3 (68)	59.1 (26)	36.7 (11)	60.0 (54)	57.0 (57)
≥9 years	40.3 (54)	47.7 (62)	40.9 (18)	63.3 (19)	40.0 (36)	43.0 (43)
Gender						
Male	54.9 (73)	70.1 (89)	41.9 (18)	67.9 (19)	61.1 (55)	70.7 (70)
Female	45.1 (60)	29.9 (38)	58.1 (25)	32.1 (9)	38.9 (35)	29.3 (29)
Ethnicity						
White	85.8 (115)	79.4 (100)	88.6 (39)	85.7 (24)	84.4 (76)	77.6 (76)
Asian/Asian British	7.5 (10)	11.1 (14)	11.4 (5)	7.1 (2)	5.6 (5)	12.2 (12)
Mixed/multiple ethnic groups	4.5 (6)	4.8 (6)	0 (0)	3.6 (1)	6.7 (6)	5.1 (5)
Black/African/Caribbean/Other black	2.2 (3)	3.2 (4)	0 (0)	0 (0)	3.3 (3)	4.0 (4)
Other	0 (0)	1.6 (2)	0 (0)	3.6 (1)	0 (0)	1.0 (1)
Parent marital status						
Married or living together	56.4 (75)	46.4 (58)	61.4 (27)	33.3 (9)	53.9 (48)	50.0 (49)
Lone parent	43.6 (58)	53.6 (67)	38.6 (17)	66.7 (18)	46.1 (41)	50.0 (49)
SEN status						
Child receiving SEN provision at school (parent report)	21.6 (29)	31.5 (40)	20.5 (9)	43.3 (13)	22.2 (20)	27.8 (27)

Child not reported by parent to be receiving SEN provision at school	78.4 (105)	69.2 (87)	79.5 (35)	56.7 (17)	77.8 (70)	72.2 (70)
Site						
Birmingham	75.4 (101)	77.7 (101)	79.5 (35)	86.7 (26)	73.3 (66)	75.0 (75)
Somerset	25.6 (33)	22.3 (29)	20.5 (9)	13.3 (4)	26.7 (24)	25.0 (25)
Housing type						
Owned	22.6 (30)	22.1 (27)	18.2 (8)	15.4 (4)	24.7 (22)	24.0 (23)
Other	77.4 (103)	77.9 (95)	81.8 (36)	84.6 (22)	75.3 (67)	76.0 (73)
Housing quality						
Good	68.0 (83)	59.7 (68)	55.0 (22)	52.2 (12)	74.4 (61)	61.5 (56)
Acceptable	28.7 (35)	31.6 (36)	42.5 (17)	34.8 (8)	22.0 (18)	30.8 (28)
Substandard	3.3 (4)	8.7 (10)	2.5 (1)	13.0 (3)	3.7 (3)	7.7 (7)
Income (weekly, exc. housing costs)						
≤ £150	17.5 (22)	18.1 (21)	20.5 (8)	12.0 (3)	16.1 (14)	19.8 (18)
> £150	82.5 (104)	81.9 (95)	79.5 (31)	88.0 (22)	83.9 (73)	80.2 (73)
Socio-economic status [how hard it is to live on household income right now]						
Not at all / somewhat / difficult	85.8 (115)	87.3 (110)	90.9 (40)	82.1 (23)	83.3 (75)	88.8 (87)
Very difficult / extremely difficult	14.2 (19)	12.7 (16)	9.1 (4)	17.9 (5)	16.7 (15)	11.2 (11)

Table 2: Summary and comparison of outcomes between the intervention and control arms at 32 weeks post-randomization

		Intervention Mean (sd)	Control Mean (sd)	Unadjusted MD/ OR	Adjusted[‡] MD / OR (95% CI)	p	Adjusted SMD[‡]
PSDQ	TD[#]	18.4 (7.4)	19.6 (6.7)	-0.7	-0.6 (-2.6 to 1.4)	0.54	-0.07 (-0.30 to 0.16)
AAPI	Empathy	41.2 (6.9)	40.6 (6.3)	0.9	1.1 (-0.8 to 2.9)	0.25	0.14 (-0.10 to 0.39)
APQ	Inconsistent discipline	15.2 (5.1)	15.9 (4.3)	-0.9	-0.3 (-1.6 to 0.9)	0.62	-0.05 (-0.27 to 0.16)
	Involvement	40.6 (4.4)	40.3 (5.8)	0.0	0.1 (-1.2 to 1.5)	0.87	0.02 (-0.20 to 0.23)
	Positive parenting	27.1 (2.9)	27.0 (2.6)	0.1	-0.3 (-0.9 to 0.3)	0.35	-0.09 (-0.27 to 0.10)
ECBI	Intensity	132.8 (47.3)	140.0 (40.3)	-6.2	-3.4 (-12.3 to 5.5)	0.45	-0.07 (-0.27 to 0.12)
	Problem	14.1 (10.7)	13.6 (9.7)	-0.2	0.5 (-2.1 to 3.0)	0.71	0.04 (-0.18 to 0.26)
PSDQ	Conduct	3.8 (2.4)	4.1 (2.3)	-0.1	-0.2 (-0.8 to 0.4)	0.50	-0.08 (-0.31 to 0.15)
	Emotional	4.3 (2.6)	4.8 (2.7)	-0.3	-0.3 (-1.1 to 0.5)	0.46	-0.09 (-0.33 to 0.15)
	Hyperactivity	6.6 (2.7)	6.8 (2.5)	-0.1	0.1 (-0.7 to 0.8)	0.85	0.02 (-0.21 to 0.26)
	Impact	3.4 (2.9)	3.3 (2.9)	-0.1	0.3 (-0.5 to 1.0)	0.49	0.08 (-0.15 to 0.31)
	Peer	3.7 (2.3)	3.9 (2.3)	-0.1	-0.2 (-0.9 to 0.6)	0.69	-0.04 (-0.27 to 0.18)
	Prosocial	7.0 (2.4)	6.8 (2.0)	0.0	0.2 (-0.4 to 0.8)	0.53	0.07 (-0.16 to 0.30)

	TD \geq 14 [^]	-	-	0.8	0.9 (0.4 to 1.8)	0.70	-
WCQ	Accepting responsibility	4.2 (2.6)	4.9 (3.0)	-0.9	-0.6 (-1.4 to 0.2)	0.15	-0.18 (-0.42 to 0.07)
	Confrontive coping	6.6 (3.7)	6.7 (3.3)	-0.3	-0.3 (-1.3 to 0.8)	0.62	-0.06 (-0.31 to 0.19)
	Distancing	7.4 (3.8)	7.2 (3.6)	0.1	0.5 (-0.6 to 1.6)	0.38	0.11 (-0.13 to 0.34)
	Escape avoidance	7.7 (5.1)	8.2 (5.5)	-0.6	-0.2 (-1.8 to 1.3)	0.77	-0.03 (-0.27 to 0.20)
	Planful problem solving	9.0 (3.9)	3.0 (4.4)	-0.0	-0.3 (-1.4 to 0.9)	0.68	-0.05 (-0.28 to 0.18)
	Positive reappraisal	6.9 (4.9)	6.1 (4.3)	0.9	1.0 (-0.3 to 2.4)	0.14	0.18 (-0.06 to 0.41)
	Self controlling	9.3 (4.8)	10.1 (4.2)	-0.9	-0.9 (-2.5 to 0.6)	0.23	-0.15 (-0.40 to 0.10)
	Social support	8.4 (4.8)	8.6 (4.6)	-0.3	0.1 (-1.2 to 1.4)	0.90	0.02 (-0.22 to 0.25)

Abbreviations: PSDQ: parent-completed strengths and difficulties; AAPI: AAPI-2 Adult-Adolescent Parenting Inventory; APQ: APQ Alabama Parenting Questionnaire; ECBI: Eyberg Child Behavior Inventory; WCQ: Ways of Coping Questionnaire; TD: total difficulties score; CI: confidence interval; MD: mean difference; OR: odds ratio; SMD: standardized mean difference

primary outcome

£ adjusted for: age group (<9 / \geq 9), site (Birmingham / Somerset), gender, binary ethnicity (White / non-White), SEN, SES (<4 / \geq 4), marital status (married or living together / lone parent), parents' education (left school at or before age 16 / further education).

[^] threshold analysis (expressed as odds ratio)

Table 3: Total and subscale scores for the Parent Program Implementation Checklist (PPIC)

	Adherence	Quality	Parent responsiveness	Total score
Mean score (SD)	25.3 (7.5)	29.3 (7.5)	7.6 (2.7)	62.1 (16.0)
Mean threshold (SD)	0.72 (0.21)	0.73 (0.19)	0.76 (0.27)	0.73 (0.19)

Table S1: Baseline outcomes for all participants in the trial, those lost to follow-up at 32 weeks post-randomization (withdrawn or unable to contact), and those remaining in the trial to the end. Values are mean and standard deviation.

Baseline variable	All participants		Participants lost to follow-up		Remaining participants	
	Intervention [#]	Control [£]	Intervention	Control	Intervention	Control
<i>Parent-rated Strengths and Difficulties Questionnaire</i>						
Conduct problems	5.2 (2.7)	5.1 (2.3)	4.75 (2.7)	5.50 (2.1)	5.49 (2.6)	5.01 (2.3)
Emotional problems	5.7 (2.5)	5.6 (2.8)	5.45 (2.3)	6.27 (2.5)	5.87 (2.6)	5.42 (2.9)
Hyperactivity	7.4 (2.3)	7.5 (2.0)	6.84 (2.5)	7.63 (1.7)	7.64 (2.1)	7.48 (2.1)
Peer problems	4.8 (2.4)	4.5 (2.5)	4.09 (2.2)	4.83 (2.7)	5.20 (2.5)	4.45 (2.4)
Prosocial behaviour	5.2 (2.0)	5.4 (1.9)	5.59 (1.9)	5.67 (1.9)	4.99 (2.1)	5.32 (1.9)
Impact	4.2 (3.1)	4.4 (3.2)	3.79 (3.1)	4.92 (3.5)	4.40 (3.1)	4.23 (3.1)
Total difficulties	23.2 (6.5)	22.8 (6.4)	21.14 (6.0)	24.23 (7.7)	24.21 (6.5)	22.37 (6.2)
<i>Eyberg Child Behaviour Inventory</i>						
Intensity	150.8 (43.2)	153.8 (40.8)	146.88 (38.2)	156.26 (37.2)	152.72 (45.5)	153.17 (41.9)
Problem	18.1 (9.5)	18.4 (9.0)	16.42 (9.5)	19.16 (7.8)	18.90 (9.4)	18.24 (9.3)
<i>WCQ</i>						
Accepting responsibility	4.7 (3.2)	5.4 (2.9)	4.50 (2.7)	5.00 (3.3)	4.76 (3.4)	5.55 (2.8)
Confrontive coping	7.3 (3.5)	7.7 (3.4)	6.58 (3.0)	7.64 (4.0)	7.71 (3.6)	7.68 (3.2)
Distancing	7.3 (3.8)	8.2 (4.0)	7.43 (3.6)	8.07 (4.2)	7.19 (3.9)	8.21 (3.9)
Escape avoidance	8.9 (5.7)	9.2 (5.4)	8.00 (6.2)	9.46 (5.4)	9.32 (5.4)	9.07 (5.4)
Planful problem solving	9.4 (3.5)	9.0 (4.0)	9.30 (3.5)	8.70 (4.7)	9.47 (3.6)	9.02 (3.8)
Positive reappraisal	6.9 (4.9)	7.4 (4.9)	6.68 (4.7)	7.07 (5.6)	7.00 (5.1)	7.56 (4.6)
Self controlling	10.4 (4.1)	10.6 (4.5)	9.27 (3.9)	9.64 (5.1)	10.99 (4.0)	10.86 (4.4)
Social support	8.8 (4.5)	9.0 (4.5)	8.02 (4.4)	7.25 (4.6)	9.20 (4.6)	9.53 (4.4)
<i>AAPI</i>						

Empathy	39.9 (5.5)	39.4 (6.5)	40.33 (5.0)	38.82 (7.9)	39.71 (5.7)	39.58 (6.1)
<i>APQ</i>						
Inconsistent discipline	16.3 (4.7)	17.0 (4.3)	16.42 (4.9)	16.78 (4.8)	16.22 (4.6)	17.05 (4.2)
Involvement	39.5 (5.4)	39.7 (5.2)	39.05 (5.7)	39.50 (4.1)	39.79 (5.2)	39.70 (5.5)
Positive parenting	27.4 (2.5)	26.7 (2.9)	27.23 (2.7)	26.54 (3.0)	27.44 (2.5)	26.75 (2.8)

WCQ Ways of Coping Questionnaire
AAPI Adult-Adolescent Parenting Inventory
APQ Alabama Parenting Questionnaire

n=126 to 134

£ n=123 to 130

Table S2: Comparison of intervention and control (with multiple imputation) at 16 weeks post-randomization

		Intervention Mean (sd)	Control Mean (sd)	Unadjusted estimate	Adjusted estimate[‡] (95% CI)	p	Adjusted SMD[‡]
AAPI	Empathy	41.3 (6.5)	40.7 (5.9)	0.4	0.8 (-1.0 to 2.5)	0.39	0.10 (-0.14 to 0.34)
APQ	Inconsistent discipline	15.4 (4.9)	16.2 (4.4)	-0.9	-0.4 (-1.6 to 0.9)	0.56	-0.06 (-0.28 to 0.15)
	Involvement	41.0 (4.6)	40.5 (5.0)	0.6	0.6 (-0.5 to 1.7)	0.32	0.10 (-0.10 to 0.30)
	Positive parenting	27.1 (2.7)	26.6 (3.0)	0.5	0.1 (-0.6 to 0.8)	0.75	0.03 (-0.17 to 0.24)
ECBI	Intensity	141.5 (44.8)	140.9 (43.0)	-3.0	0.0 (-8.9 to 9.0)	0.99	0.00 (-0.19 to 0.19)
	Problem	15.8 (10.0)	15.4 (10.2)	0.0	0.6 (-2.0 to 3.2)	0.65	0.05 (-0.17 to 0.28)
PSDQ	Conduct	4.3 (2.6)	4.2 (2.3)	0.0	-0.0 (-0.7 to 0.7)	0.94	-0.01 (-0.24 to 0.22)
	Emotional	4.2 (2.5)	4.5 (2.9)	-0.2	-0.2 (-1.0 to 0.6)	0.58	-0.06 (-0.30 to 0.17)
	Hyperactivity	6.6 (2.6)	7.0 (2.3)	-0.3	-0.1 (-0.8 to 0.6)	0.74	-0.04 (-0.27 to 0.19)
	Impact	3.3 (2.8)	3.2 (2.9)	0.1	0.3 (-0.4 to 1.0)	0.41	0.09 (-0.13 to 0.32)
	Peer	4.2 (2.3)	4.0 (2.5)	0.3	0.2 (-0.5 to 1.0)	0.54	0.07 (-0.17 to 0.31)
	Prosocial	6.5 (2.4)	6.8 (1.9)	-0.1	0.0 (-0.6 to 0.6)	0.99	0.00 (-0.23 to 0.23)
	TD	19.3 (7.3)	19.8 (6.9)	-0.2	-0.2 (-2.2 to 1.9)	0.88	-0.02 (-0.25 to 0.21)

WCQ	Accepting responsibility	4.2 (2.7)	5.3 (3.1)	-1.0	-0.6 (-1.4 to 0.3)	0.22	-0.15 (-0.39 to 0.09)
	Confrontive coping	7.3 (3.5)	7.4 (3.2)	-0.1	-0.0 (-1.0 to 1.0)	0.99	0.00 (-0.25 to 0.24)
	Distancing	7.4 (3.8)	7.3 (3.4)	0.1	0.4 (-0.6 to 1.4)	0.45	0.09 (-0.14 to 0.32)
	Escape avoidance	7.4 (5.3)	8.2 (5.3)	-0.8	-0.5 (-2.1 to 1.1)	0.53	-0.07 (-0.30 to 0.16)
	Planful problem solving	9.7 (4.2)	8.9 (4.0)	0.9	0.6 (-0.4 to 1.7)	0.24	0.13 (-0.09 to 0.35)
	Positive reappraisal	7.6 (5.0)	6.5 (4.3)	1.2	1.3 (0.1 to 2.5)	0.04	0.24 (0.01 to 0.46)
	Self controlling	10.6 (4.4)	9.7 (4.0)	1.1	1.0 (-0.1 to 2.2)	0.08	0.19 (-0.02 to 0.41)
	Social support	8.9 (4.6)	8.2 (4.7)	0.6	0.6 (-0.7 to 1.9)	0.37	0.11 (-0.13 to 0.35)

Abbreviations: PSDQ: parent-completed strengths and difficulties; AAPI: AAPI-2 Adult-Adolescent Parenting Inventory; APQ: APQ Alabama Parenting Questionnaire; ECBI: Eyberg Child Behavior Inventory; WCQ: Ways of Coping Questionnaire; TD: total difficulties score; CI: confidence interval; SMD: standardized mean difference

primary outcome

£ adjusted for: age group (<9 / ≥9), site (Birmingham / Somerset), gender, binary ethnicity (White / non-White), SEN, SES (<4 / ≥4), marital status (married or living together / lone parent), parents' education (left school at or before age 16 / further education).

Table S3: Analyses of moderators of primary outcome (PSDQ Total Difficulties at 32 weeks post-randomization)

Moderator	Subgroup	Control mean (SD)*	Intervention mean (SD)*	Adjusted mean difference (95% CI)#	p-value for interaction#
Age	< 9 years	18.1 (6.6)	18.4 (7.1)	-0.30 (-2.87 to 2.26)	0.71
	≥ 9 years	21.6 (6.2)	18.5 (8.0)	-1.04 (-3.96 to 1.88)	
Gender	Male	20.0 (6.5)	19.6 (6.2)	-0.15 (-5.99 to 5.69)	0.87
	Female	18.8 (7.1)	16.6 (8.8)	-0.50 (-2.83 to 1.83)	
PSDQ baseline	< 17	13.9 (5.2)	11.4 (6.6)	-3.31 (-10.43 to 3.81)	0.44
	≥ 17	21.2 (6.2)	19.6 (6.9)	-3.20 (-9.98 to 3.59)	
Ethnicity	White	20.7 (6.7)	18.7 (6.9)	-0.62 (-2.72 to 1.48)	0.98
	non-White	16.1 (5.5)	16.5 (9.7)	-0.69 (-5.02 to 4.87)	

* complete case analysis summary

imputed data analysis, with adjustments for: age, gender, borough, ethnicity, SEN, SES, marital status, parents' education, baseline PSDQ Total Difficulties

Table S4: Use of school services in the last 6 months by children and families in the intervention and control arms (parent report at 32 weeks post-randomization)*

Service received	Intervention (%)&	Control (%)#
Extra parent consultation with head teacher	31.7	27.8
Extra parent consultation with class teacher	47.0	46.2
School nurse	10.3	12.9
Educational social worker	13.0	10.7
School doctor	1.3	2.4
Other school service	50.0	38.9
One-to-one help	52.3	49.0
Small group work	53.9	55.1
Special teaching	18.0	26.6
Other help at school	14.0	15.8
Special Educational Needs statement issued at school	18.2	18.2
Psychological assessment at school	16.5	12.6
Attended a special school	7.0	2.1

* None of the differences between the conditions are statistically significant at $p < 0.05$

& Base figure (n) ranges from 77 to 89

Base figure (n) ranges from 72 to 99

Table S5: Use of other additional services in the last 6 months by children and families in the intervention and control arms (parent report at 32 weeks post-randomization)*

Service received	Intervention (%) ^{&}	Control (%) [#]
GP	16.1	17.2
CAMHS	6.9	2.1
Hospital	2.3	4.2
Other health services	18.4	20.0
Social work	4.6	3.2
Educational support	3.4	3.2
Family or parent support	5.7	9.5
Other services	4.6	7.4
Any of above services used as a results of child's behavior [§]	65.1	56.7
Number of services received [§]		
0	64.4	58.9
1	17.2	23.2
2	6.9	9.5
3	8.0	4.2
4	3.4	4.2
Use of any additional service [^]	35.6	41.1

* None of the differences between the conditions are statistically significant at $p < 0.05$

[&] Base figure (n) = 87 unless otherwise stated

[#] Base figure (n) = 93 to 95 unless otherwise stated

[§] Base figures (n) = 87 (intervention) and 95 (control). Maximum possible number of services is 4.

[§] Base figure (n) = 63 (intervention) and 67 (control)

[^] Base figure (n) = 87 (intervention) and 95 (control)

Table S6: Implementation fidelity according to facilitator self-report[§]

Fidelity measure	Mean (standard deviation) unless otherwise stated
<i>Dose</i>	
Group sessions [§] :	
Number of sessions attended (all participants)	6.1 (4.4)
Available sessions attended (all participants)	59.6%
0 sessions	26.8%
≥8 sessions	54.6%
Number of sessions attended (if attended ≥1)	8.3 (2.7)
Available sessions attended (if attended ≥1)	80.9%
Mean group size	3.3 (1.0)
One-to-one sessions*:	
Participants who attended any session	6.1%
Number of sessions attended (all participants)	0.4
Number of sessions attended (if attended ≥1)	5.9
<i>Adherence</i>	
Facilitator-reported adherence score [#]	91.2% (23.8)
<i>Participant responsiveness ('application')</i>	
Total score (out of 5)	3.0 (1.3) ^{&}
≤ 1	15.9%
≥ 4	23.8%

[§] Based on facilitator self-report (available for 23 of 31 groups, n=97 participants)

* Based on data on n=132 participants

[#] Calculated as follows: (Facilitator 1 total + Facilitator 2 total / total possible score)*100

[&] Based on 63 participants for whom information is available

Table S7: Treatment effect estimates by ITT or CACE. The CACE estimate is the estimated average effect of treatment on the primary outcome for individuals who would have attended all parenting group meetings had they been offered.

	Control arm mean (sd)	Intervention arm mean(sd)	Adjusted mean difference (95% confidence interval)#	p-value
Intention to treat (ITT)	19.6 (6.7)	18.4 (7.4)	-0.62 (-2.62 to 1.38)	0.54
Complier Average Causal Effect (CACE)			-1.08 (-4.44 to 2.26)	0.52

#Adjustments made for baseline variables anticipated to affect both outcome and participation:

Baseline PSDQ Total Difficulties score, age category, gender, site, ethnicity (white/non-white), SES, SEN, Marital status, parent education