



Enhancing Social-Emotional Outcomes in Early Years (E-SEE): Randomized Pilot Study of Incredible Years Infant and Toddler Programs

Sarah L. Blower¹ · Vashti L. Berry² · Matthew C. Bursnall³ · Judith Cohen^{3,8} · Nicole Gridley^{1,9} · Amanda Loban³ · Laura Mandefield^{3,10} · Amanda J. Mason-Jones¹ · Sinéad McGilloway⁴ · Kirsty L. McKendrick³ · Siobhan B. Mitchell² · Kate E. Pickett¹ · Gerry A. Richardson⁵ · M. Dawn Teare³ · Louise C. Tracey⁶ · Simon M. Walker⁵ · Karen A. Whittaker⁷ · Jessica Wright³ · Tracey J. Bywater¹

Received: 21 May 2020 / Accepted: 25 May 2021 / Published online: 12 June 2021
© The Author(s) 2021

Abstract

Social emotional development in infancy is a predictor of outcomes in later life, yet there is little evidence of effectiveness for parenting interventions designed to enhance social emotional wellbeing in infancy. An 18-month two-arm randomized controlled pilot trial evaluated the feasibility of a definitive trial of Incredible Years (IY) Infant and Toddler parent programs delivered in a proportionate universal model, called Enhancing Social-Emotional Health and Wellbeing in the Early Years (E-SEE) Steps. Intervention families received an IY Babies book (universal dose), followed by the IY Infant and/or the Toddler group-based programs, based on parent depression (PHQ-9) and/or child social emotional development (ASQ:SE-2) scores. Control parents received services as usual. Parents from two English local authorities with a child eight-weeks-old or younger participated, and were block randomized using a web-based system. Primary endpoints for the study were feasibility parameters relating to recruitment, retention, intervention fidelity and appropriateness of measures. 205 participants were randomized (152:53, intervention:control). Our target was 288 parents. Trial retention rate was higher than expected, with a completion rate of 88% (n = 181, 137:44) at follow-up 3; equating to 94% of 192 expected participants. Intervention uptake was lower than expected. Fidelity of delivery was acceptable and measures were deemed appropriate. A definitive trial is feasible with design amendments to include: introduction of a child screener for intervention eligibility; enhanced intervention material; revised sample size and random allocation ratio. Our internal pilot became an external pilot due to these changes.

Keywords Proportionate universalism · Social emotional · Parent program · Incredible years · Infant

Highlights

- This is the first pilot of a proportionate universal delivery of the IY parent program.
- Trial retention was high at 88% at final follow-up (18 months post baseline).
- Emerging findings suggest a definitive trial of E-SEE STEPS is warranted.
- Definitive trials should consider methods to enhance intervention uptake.
- Trials should consider intervention compliance with relevant guidelines.

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR the Department of Health and Social Care, or the ARCs.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s10826-021-01991-7>.

✉ Tracey J. Bywater
Tracey.Bywater@york.ac.uk

Extended author information available on the last page of the article

One in eight children and young people (aged 2–19 years) in England experience a mental health disorder (Sadler et al., 2018). There are strong continuities between childhood and adult mental health difficulties (Winsper et al., 2020). The cost of these difficulties to individuals, families, communities and wider society is large (Romeo et al., 2006). Early intervention and prevention is therefore of significant public health importance (Jones et al., 2015).

Onset of mental health issues in very young children is difficult to detect, therefore consideration of risk factors to adaptive infant development is key when planning (preventative) interventions. Parents experiencing socio-economic disadvantage (e.g., low income or education) are more likely to experience poor mental health associated with higher stress levels and financial and time constraints (Kalil et al., 2020). Depressed parents are more likely to experience reduced emotional and cognitive capacity for sensitive and responsive parenting, which can impact on the quality of parent-infant bonds (Galbally and Lewis, 2017), and greater use of punitive discipline practices (Kalil et al., 2020). These risk factors are established predictors of social, emotional and behavioral difficulties and disorders in children (Bender et al., 2007).

Parenting programs are effective at promoting children's development through changing parent behavior (Barlow et al., 2010). Enhancing parenting early in a child's life is likely to be the most effective and cost-effective point to intervene (Ryan et al., 2017). Cochrane reviews have highlighted the effectiveness of *group-based* parent programs to promote child and parent well-being in children aged 3 years and older (e.g., Furlong et al., 2012). There is a, however, a dearth of evidence for the effectiveness of parenting interventions targeted specifically at parents of infants under 3 years (Barlow et al., 2010).

In the UK, there is significant policy interest in *proportionate universal* parenting interventions (Department of Health, 2009; Leadsom et al., 2013). Proportionate universalism refers to the universal provision of services, with a scale and intensity proportionate to levels of disadvantage or need (Carey et al., 2015). For example, a proportionate universal parenting intervention might involve offering a low intensity intervention to all parents regardless of their circumstances and needs, alongside offering additional interventions with increased intensity targeted at subgroups of parents with increasing needs or levels of disadvantage. Theoretically, by increasing the health and well-being of all families the gap between the poorest and wealthiest is reduced (Benach et al., 2013). Proportionate universal interventions can be referred to as “tiered”, “stepped care”, “adaptive interventions” or “dynamic treatment regimes” (Candlish et al., 2019).

Despite the significant policy interest, there is an evidence gap for parenting programs for 0–2 year olds, and more specifically for proportionate universal interventions (Hurt et al., 2018). The current study was funded in response to a commissioned call by one of the UK's largest funders of health research (the National Institute for Health Research) to attempt to address this evidence gap. We developed a proportionate universal model for parents of 0–2-year-olds called Enhancing Social-Emotional Health and Well-being in the Early Years (E-SEE) Steps,

comprising Incredible Years® (IY) materials and programs (Webster-Stratton, 2011), and planned a randomized controlled trial evaluation (RCT).

The IY programs are manualized, collaborative (non-didactic) parent education and training interventions informed by social learning theory and designed to enhance the social and emotional well-being of children aged 0–12 years. IY is effective and cost-effective when delivered to parents of children aged 3+ years as demonstrated in independent trials across several countries/contexts (Furlong et al., 2012). Examples include community settings in disadvantaged urban areas of Ireland (McGilloway et al., 2012), urban and rural areas of North and Mid-Wales (Hutchings et al., 2013), disadvantaged inner-city in England (Morpeh et al., 2017), as well as multi-cultural applications in low and middle income countries (Pidano and Allen, 2015). IY has been delivered preventatively, and as a targeted intervention (Leijten et al., 2019), making it an ideal candidate for a universal proportionate delivery model. Analysis of data pooled from several European IY trials suggested a large moderating effect for depression and that IY was more beneficial for children where parents were more depressed, and resulted in positive child outcomes regardless of socioeconomic status (Gardner et al., 2017). Meta-analyses suggested IY may be beneficial to younger children and their parents (Menting et al., 2013). The strength of IY's UK and international evidence base from independent trials for children 3+ years was a primary factor in its selection over alternatives.

E-SEE STEPS comprises the Infant (IY-I) book and the program, and the Toddler (IY-T) program, for 0–1 and 1–3-year-olds respectively. Both programs build on the strategies and content of the IY (3-years +) programs. Evidence for IY-I and IY-T, although promising, is currently lacking. A small randomized non-targeted study (N = 80) of IY-I in Wales, UK, showed that control mothers were significantly less sensitive during play with their baby (e.g., less likely to respond in a positive manner to their child's positive vocal/physical actions or to help an infant label, identify and understand their emotions) at the 6 months follow-up (Jones et al., 2016). A second small trial of IY-I delivered universally in Denmark, found differential outcomes for the lowest and highest functioning families suggesting that IY-I should be targeted (Pontoppidan et al., 2016). Two IY-T trials have been conducted. A small community-based trial in Wales, UK, targeted families living in disadvantaged ‘Flying Start’ geographical areas (Hutchings et al., 2013), rather than on the basis of individual need, which led to a lower risk sample. The study reports modest short-term benefits for IY-T in parental mental well-being and parental praise compared with the control arm. A US trial evaluated IY-T delivered in ‘non-community’ primary care settings

(pediatric practices) and documented improvements in parenting practices and child disruptive behaviors (Perrin et al., 2014).

Although small scale trials of IY-I and IY-T have been conducted, there has been no independent RCT evaluation of IY-I and IY-T delivered in a proportionate universal format. The Medical Research Council (MRC) framework for evaluating complex interventions suggests a pilot study should be conducted to identify potential challenges to be addressed before an expensive full-scale definitive trial is conducted (Craig et al., 2013). This is particularly relevant for E-SEE Steps, given the complexities of the intervention and the observation that RCTs of group-based parenting programs have previously been undermined by challenges in recruiting and retaining sufficient participants, ensuring viable group sizes and maintaining intervention fidelity (Simkiss et al., 2010; Axford et al., 2012; Axford et al., 2020). The emergence of proportionate universal interventions and RCTs also raises challenges in statistical design and analysis that fall outside current evaluation guidelines (see Candlish et al., 2019 for a summary), providing a strong case for analyzing pilot data before finalizing a protocol for a definitive trial (Craig et al., 2013). The current study was designed as an internal pilot, however, several amendments were necessary to ensure viability of the definitive trial phase. As a result our pilot is now defined and presented throughout as an external pilot.

Aims and Objectives

The study's main aim was to establish if a definitive RCT of E-SEE STEPS is feasible. The primary objective was to assess the feasibility of trial processes and intervention delivery against progression criteria for a definitive trial. Progression criteria (see Table 1) were set in relation to: (a) recruitment; (b) retention; (c) intervention delivery; (d) intervention acceptability; and (e) intervention fidelity. The secondary objective was to gather data to inform a statistical power and sample size calculation for a main trial to answer the primary research question, i.e., "to what extent does the proportionate delivery model of IY (and each dose level) enhance child social emotional well-being at 20 months of age and adult well-being compared to services as usual?" Pilot data was required to estimate: (a) the variability in the primary outcome (ASQ:SE-2) of each arm at follow-up timepoints; (b) the correlation between the primary clinical outcome at different timepoints; (c) the pooled standard deviation (SD) of the primary outcome at follow-up timepoints; (d) the average group size attending IY-I and IY-T; (e) the prevalence of mild to severe parent depression at all timepoints; and (f) the prevalence of ASQ:SE-2 monitoring levels at all timepoints.

Methods

Participants and Sampling Procedures

Trial inclusion criteria were: the participant had the main parental responsibility for a child aged ≤ 8 weeks at initial engagement; was willing to participate in the research; was willing to be randomized and, if allocated to intervention, able and willing to receive IY services offered; and was fully competent to give consent. Exclusion criteria applied if the child had obvious or diagnosed organic or developmental difficulties or the parent was enrolled on another group parent program at recruitment.

Parents (defined as primary caregivers who have the main parenting responsibility for the index child, including biological parents, step-parents, foster parents and legal guardians) were identified and referred into the study by health visitors and family support workers (self-referral was also possible). Researchers assessed potential participants for eligibility via the referral form and a follow-up telephone call to the parent. Eligible and willing parents were visited by a researcher to discuss the study in more detail, and obtain written, informed, consent. Consenting parents could invite a co-parent (an adult who shared parenting responsibilities for the child at least 3 days or nights a week) into the trial. Each family received shopping vouchers following completion of each data collection timepoint: £10 at Baseline and increasing by an additional £5 at each timepoint. Families also received a £10 voucher if they informed us of a change of address.

Sample Size, Power and Precision

The target sample size for the internal pilot was 288 parents (144 in each site) at baseline. This represented one third of the target sample size (and equivalent to the first two sites) for the full definitive trial; it is not based on a specific sample size calculation as would be expected for an external pilot. The sample size was calculated in the presence of several unknown statistical design parameters—the intra-class correlation coefficient (ICC), the correlation between measurements made on the same participant over time and the standard deviation (SD) of the primary outcome measure (Ages and Stages Questionnaire Social and Emotional supplement 2nd edition; ASQ:SE-2). We applied conservative estimates of the SD ($SD = 25$) (Squires et al., 2015) and correlation between the repeated measures and the ICC ($ICC = 0.05$ and correlation = 0.6) based on prior studies of group-based interventions (Adams et al., 2004).

Research Design

The study was designed as an 18-month pragmatic two-armed pilot RCT evaluation of E-SEE STEPS. Participants

in two English Local Authorities (LAs, one in the North and one in the South) were randomly allocated to intervention or control (services as usual, SAU). The study at four time-points: baseline (BL) when the index child was approximately 2 months old; follow-up 1 (FU1) when the child was 4 months; follow-up 2 (FU2) when the child was 11 months; and follow-up 3 (FU3) when the child was 20 months.

Procedure

Random assignment method

Participants were block randomized using a web-based system with allocation ratio 3:1 (intervention: control) and random blocks of size 4 and 8 and stratified by site, sex of primary carer, sex of child and primary parent's baseline Patient Health Questionnaire (PHQ-9) score. The final sequence was generated by the web-based system, allowing it to be concealed from all study team members. Randomization occurred after baseline data collection to reduce initial attrition. Data collectors were masked. Participants, group leaders, research staff (involved in recruitment, initial assessments, fidelity assessment) and data managers were not masked. The trial statistician was unmasked when conducting the final analysis.

Intervention

E-SEE STEPS is a proportionate universal intervention comprising three levels of the IY program (see <http://incredibleyears.com/> for program specifics and logic models). E-SEE STEPS' three levels are described below:

1. **IY-B (universal)**: 'A guide and journal of your Baby's first year' is a book that discusses how to promote and understand a baby's physical, social, emotional and language development. It includes safety alerts, developmental principles, and a section to record progress.
2. **IY-I (targeted)**: includes content on how to help babies feel loved, safe, and secure as well as how to encourage babies' physical and language development. It comprises 8–10 two-hour weekly sessions delivered by two trained leaders to 8–10 parents and their babies. Group leaders use video clips of real-life situations and provide opportunities for group discussions and role-play.
3. **IY-T (targeted)**: parents learn how to help their toddlers feel loved and secure and how to scaffold language, social, and emotional development. They learn how to establish clear and predictable routines,

and handle separations and reunions. The program comprises 12 weekly sessions, in two-hour blocks by two trained leaders and was designed for groups of 10–14 parents. Unlike IY-I, parents attended without their toddlers.

All intervention parents received the Incredible Babies Book (IY-B) as the universal level of E-SEE STEPS. IY-B was posted to all intervention families following randomization to read and use at home. IY-I was offered to parents who met eligibility criteria as assessed at FU1. IY-I groups were delivered between FU1 and FU2. IY-T was offered to parents who met eligibility criteria at FU2. IY-T groups were delivered between FU2 and FU3. It did not matter whether participants had previously met criteria for IY-I, all parents in the intervention arm were eligible for IY-T if they met the criteria at FU2. Four possible 'doses' of IY existed; some families would receive IY-B only, others IY-B and IY-I, others IY-B and IY-T, and others the full dose of IY-B, IY-I and IY-T.

Criteria for being invited to either IY-I or IY-T was originally: parent score ≥ 5 on the Patient Health Questionnaire (PHQ-9, indicating mild to severe depression). The criteria for eligibility were revised mid-way through the pilot, (see the results section for more detail), the PHQ-9 threshold was reduced to ≥ 4 and scores on ASQ:SE2 in the monitoring zone or higher was added as "and/or" to the PHQ-9 criteria.

IY-I and IY-T were delivered in local community settings (e.g., Children's Centers) by a combination of one health professional (e.g., health visitor) and one local authority staff member (e.g., family support worker) co-leading each group. Group leaders attended separate 3-day training sessions for IY-I and IY-T delivered by accredited UK-based IY trainers and received fortnightly telephone supervision from IY accredited mentors. Implementation partners/organizations were advised to deliver a 'dry run' practice of an IY-I or IY-T group after the training and prior to delivering research groups. "Service design" meetings facilitated by the research team with key decision makers in each site confirmed the specifics of the delivery model for each site. Based on information from population level data on the prevalence of maternal depression (sourced from <https://fingertips.phe.org.uk/profile-group/child-health>), four IY-I groups and two groups of IY-T per site were planned.

Control treatment

Control condition parents/co-parents received SAU which largely consisted of "stay and play groups" but other parenting programs were periodically available (in Site 1 the

Solihull Approach, and a locally developed program; Site 2 offered HENRY and a form of Triple P).

Data collection

Data collection took place in participant homes unless parents requested an alternative venue. Data collectors held postgraduate qualifications or had equivalent experience of working with families and were trained in local safeguarding practices, good clinical practice and administering measures.

Measures

Child primary outcome

The ASQ:SE-2 (Squires et al., 2015) is a 36-item parent-report based tool for screening children's social and emotional development with different versions covering nine specific developmental ages. It covers 6 key social and emotional development areas: self-regulation, compliance, adaptive functioning, autonomy, affect, social-communication, and interaction with people. Total scores are transferred onto a simple score-grid, which include cut off scores indicating "low or no risk" (indicating development is on schedule), 'monitor' (indicating a child may have a problem) and 'refer' (indicating the child needs a referral). The reliability and validity of the ASQ:SE-2 has been investigated with 14,074 diverse children across the age intervals and their families. Test-retest reliability is 89%; internal consistency is 84%; sensitivity is 81%, and specificity is 84% (Squires et al., 2015).

Parent primary outcome

Depression was measured using the PHQ-9 (Kroenke et al., 2001). Respondents are required to provide answers to 9 items based on the way they have been feeling over the last two weeks. Items are summed into a total score and there are thresholds indicating the potential severity of depression symptoms: minimal depression (0–4); mild depression (5–9); moderate depression (10–14); moderately severe depression (15–19); severe depression (20–27). The PHQ-9 has established good diagnostic and construct validity, the internal reliability is also reported to be excellent with Cronbach alphas ranging between 0.86 and 0.89 and test retest reliability reported at 0.84 (Kroenke et al., 2001).

A bespoke parent report form captured information on age, ethnicity, religion, income, marital status, parent/co-parent education, and quality of relationships between parents. Secondary outcome measures were implemented but not used in the assessment of progression criteria or sample size calculation (details can be found in our published main trial protocol:

<https://www.york.ac.uk/healthsciences/research/public-health/projects/e-see-trial>). Data on adverse events and minor protocol non-compliances were also collected and reported in Appendix A.

Measures for Assessing Progression Criteria

Intervention delivery

Book receipt was confirmed through tracking postage. The number of groups delivered in each site was gathered through project monitoring data. Group size and parent attendance was assessed via parent contact logs completed by group leaders.

Intervention acceptability

Parent attendance data gathered via contact logs (as above). Parent satisfaction was assessed using standard IY parent satisfaction questionnaires completed after each session, and at the end of each program.

Intervention fidelity

Group leaders' adherence to core components was measured using standard, weekly-completed, self-rated IY checklists that correspond with the components set out in the respective intervention manuals (e.g., number of vignettes shown). IY checklists include a 'Yes' or 'No' section regarding specified actions, for example "Did I....Write the agenda on the board". Quality of delivery was assessed via the researcher-rated Parent Program Implementation Checklist (PPIC), which comprises indices for three constructs; adherence, quality of delivery, and participant responsiveness (see Bywater et al., 2019). A random subset of group sessions for each program was observed and coded by two PPIC-trained researchers; week three and week eight for IY-I, and week two and week nine for IY-T.

Research Governance and Patient and Public Involvement

The trial was overseen by a Trial Management Committee (TMG) and an independent Trial Steering Committee (TSC) and Data Monitoring and Ethics Committee (DMEC). A Parent Advisory Committee (PAC) was set up to: input into the development of information/consent forms and other literature to enhance inclusivity; assist measure selection; provide lay member attendance at the TSC; and assist in training data collectors. Members of the PAC were 16 mothers and 1 father (of children aged 0–3) representing both pilot sites (site 1 n = 4; site 2 n = 13).

Analytical Procedures

Primary endpoints

Numbers of participants identified, approached, recruited, retained, and those dropping out (with reasons) are presented. Descriptive statistics are provided to summarize data relating to intervention delivery, acceptability and fidelity. Free text questions on parent satisfaction questionnaires were subject to content analysis.

Assessment of progression criteria

The assessment of progression criteria was informed by the primary endpoint analyses and performed by the TMG. Each criterion was assessed to determine if had been achieved, partially achieved or not achieved. Some criteria are multidimensional, where evidence for one dimension (e.g., number of IY-T groups) meets the stated target but another dimension does not (e.g., number of IY-I groups) an assessment of ‘partially achieved’ is made. The outcomes of this assessment were scrutinized by both the E-SEE TSC and DMEC and ultimately the research funder (NIHR) when deciding whether to fund a definitive trial.

Secondary endpoints

Missing data was not imputed for clinical outcomes. Where participants missed one follow up point but returned data at a subsequent point they were included in the analyses ($n=5$). Analyses were confined to key parameters estimated to examine whether data supports a definitive trial. For clinical outcomes, descriptive statistics are presented for continuous outcomes at FU1, FU2, and FU3, mean differences between groups along with 95% confidence intervals (CIs) are also presented. The 95% CIs were calculated using a multiple linear regression model adjusting for site, stratifying factors (sex of parent and child and initial level of depression) and the baseline values of the outcome by including as covariates. No hypothesis testing was conducted. The prevalence of mild-severe depression, the standard deviation of the outcome measure and correlation between consecutive measures of ASQ:SE-2 are presented. Estimates are accompanied by appropriate 95% confidence intervals (CI).

Results

Primary Endpoints

Key findings, summarizing the detail below, relating to each of the progression criteria are presented in Table 1.

Recruitment

We recruited 205 parents (and 30 co-parents) between November 2015 and May 2016 (see Fig. 1). The baseline recruitment target was not met, however high retention meant we achieved the expected sample size at FU3. Of the 205 participants, 152 participants were randomly allocated to the intervention arm and 53 to the control arm. The characteristics of participants in each arm were balanced (see Appendix B for participant characteristics). The progression criterion for recruitment was not met.

Retention

From the 205 participants, 181 (88%) of participants completed the study; 24 were lost to follow-up, 15 intervention and 9 control (12% of the sample). The main reason was ‘no response to contact from the study team’, see Fig. 1 for more details. Participant retention was uniformly high; from 96% at FU1 to 88% at FU3. The progression criterion relating to retention was met.

Intervention Delivery

Book receipt and ability to deliver required number of groups

All 152 participants randomized to the intervention arm received IY-B. Both sites identified and trained sufficient numbers of group leaders to deliver the planned number of groups (18 in site 1; 10 in site 2). This included training additional back up-leaders in each site in case of staff turnover or sickness. Appropriate venues were secured for each group; all were delivered in children’s centers (now called family hubs). Group leaders accessed fortnightly supervision with their accredited IY coach mentor. None of the leaders implemented the recommended ‘dry run’, however group leaders in both sites met prior to study groups commencing to plan and prepare for delivery, and in site 2 they also practiced delivering content together.

Site 1 delivered two IY-I groups and two IY-T groups. Site 2 delivered one IY-I group and two IY-T groups. Eligibility rates were higher for IY-T compared to IY-I (see Table 1), partially accounted for by an adjustment in the eligibility thresholds mid-way through the pilot and addition of a second screener (ASQ:SE-2) however acceptance of a place by eligible participants was lower for IY-T.

Group size and attendance

For IY-I, 18 parents and 3 co-parents accepted a place on one of three groups (two in site 1 and one group in site 2). Of these, 13 participants attended, with group sizes of 5 parents

Table 1 Overview of progression criteria assessment and design implications for a definitive RCT of E-SEE STEPS

Progression criteria	Endpoints/measures	Assessment of criterion	Implications for design of definitive RCT
(A) Recruitment Recruit and randomize at least 144 participants at each site (288 total) – to achieve 192 at final follow up.	Sample size (recruited and randomized) achieved at baseline.	Not achieved: N = 205 Representing 71% of the target baseline sample of 288. Although recruitment was below target, high retention mitigated this (at FU3 N = 181, 94% of original FU3 target of 192).	Embed recruitment targets in site level agreements. Create awareness raising materials to promote the study. Utilize QUINTET approach. Investigate a single research question “Do the scores of children in the IY arm, on average, stay below those scores for children in SAU over the three follow-up measures?” to reduce required sample size and increase statistical power.
(B) Retention Maximum 12% loss at each data collection (follow-up time point (equivalent to 32% overall loss).	Sample size at the three follow-up data collection timepoints (FU1, FU2, FU3).	Achieved: 181 participants were retained by FU3 (12% overall loss). Average 4% loss at each data collection point. This is much lower than the 12% loss at each timepoint and 32% overall that was anticipated.	Replicate piloted strategies for retaining parents; change of address procedures, tokens branded with contact details, PAC input to data collector training, consistent data collector across timepoints, appointment by letter procedures, re-attempt contact with those lost to follow-up at next timepoint. Consider SWAT.
(C) Intervention delivery Books received by parents and ability for sites to deliver required number of groups (to include identification and training of group leaders and suitable venues). IY-I planned: 4 groups per site. IY-T planned: 2 groups per site. Intervention groups consist of a viable number (min 8 parents invited, 5 parents in attendance >50% of sessions)	Book receipt monitored via track and trace postage. Program monitoring data collated by the study team on IY training attendance, group venue and number of groups delivered in each site). Parent contact logs completed by group leaders.	Partially achieved: All parents in the intervention arm received IY-B. Sufficient practitioners trained and attended supervision in each site. Suitable venues used (children’s centers). Viable group size’ was not achieved in all groups - impacted by trial recruitment issues and staff capacity to run groups at times convenient for parents. IY-I: 2 groups ran in site 1 (n = 6). 2 groups started in site 2 which were combined to 1 group (n = 3); IY-T: 2 groups site 1 (n = 2, n = 2), 2 groups site 2 (n = 3, n = 7). Spaces offered to 8 or more in most groups. IY-I site 1: 15 eligible, 6 accepted and attended. IY-I site 2: 12 eligible, 4 accepted and attended. IY-T site 1: 24 eligible, 8 accepted, 4 attended. IY-T site 2: 31 eligible, 17 accepted, 10 attended	Replicate delivery procedures. Ensure 10+ practitioners trained in IY-I and IY-T per site. Allow flexibility in group delivery (HVs only or LA only or combination). Ensure referring practitioners offer a study application to all parents of children aged ≤ 8 weeks. Supplement PHQ-9 (scores ≥ 5) with ASQ:SE-2 scores when determining eligibility for IY-I and IY-T. Increase randomization ratio to 5:1. Emphasize need for creche, transport, flexible timing and wrap-around support in service design processes. Utilize matched non-research participants to supplement groups to ensure minimum group sizes.
(D) Intervention acceptability IY retention levels to reach 70% at IY-I and IY-T end. Parent satisfaction (supplements data on retention levels).	Parent contact logs completed by group leaders. Satisfaction forms completed weekly, and at the end of the program.	Achieved: IY-I: Average weekly attendance as % total of those attending at least one session = 73% IY-T: Average weekly attendance as % total of those attending at least one session = 87% Parent satisfaction forms indicate high level of parent satisfaction	Efforts should be focused on increasing uptake of groups among eligible parents – see above. Suggestions from parents included increasing group sizes, ensure venues are accessible/provide transport, timing groups to start when children are younger. Additional feedback from parents e.g., increasing the number and duration of IY sessions need further exploration.
(E) Intervention fidelity Adherence and quality of delivery assessment of 80% in each local authority across co-leaders.	Group leader weekly self-report checklist (adherence) and PPIC completed by study team (quality of delivery)	Achieved: Overall adherence to key components ranged from 80% to 96%. PPIC tool shows acceptable threshold of 80% was exceeded for quality of delivery for IY-I and IY-T. Overall PPIC scores for fidelity were good for IY-I (73%) and IY-T (78%).	Retain independent measure of intervention fidelity alongside standard self-report checklists. Provide accredited IY training and fortnightly supervision to group leaders and promote delivery ‘dry runs’ through service design processes.

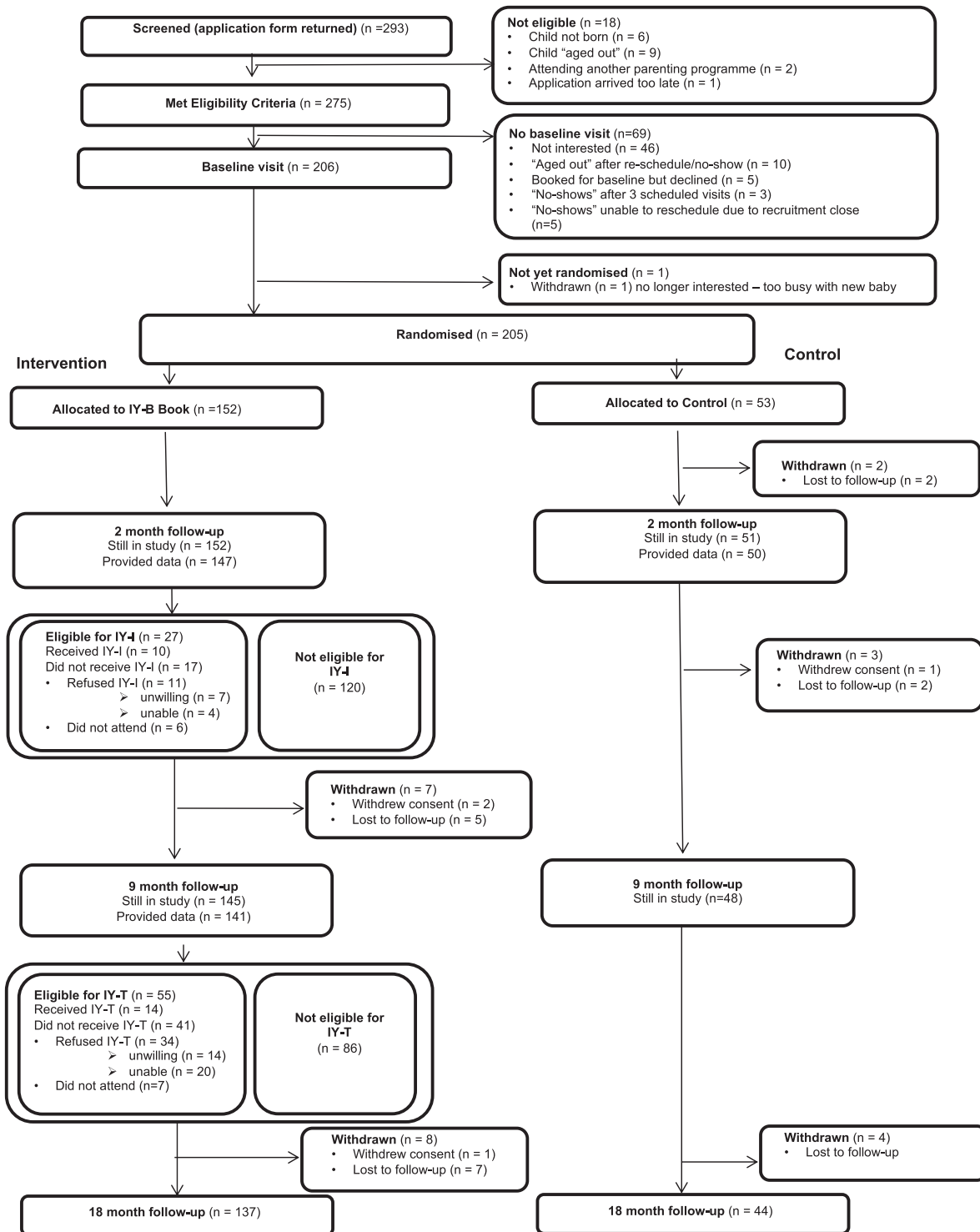


Fig. 1 CONSORT flowchart of participants

and 3 parents for site 1, and 5 parents for site 2. One IY-I group (in site 1) delivered across 10 weeks, all other IY-I groups ran for 9 weeks of delivery (small group size enabled parents and leaders to move through the program content quicker – this is in line with IY developer guidance). For IY-

T, the 14 people who consented were formed into two groups per site. Group sizes were 2 parents in each of the two groups in site 1, and 3 parents and 7 parents in site 2. Site 1 delivered their IY-T groups over 11 weeks (due to small group size), site 2 delivered both over 12 weeks.

Table 2 Eligibility, acceptance and parent participation for IY-I and IY-T

Program	IY-I	IY-T
Responses at timepoint	147 at FU1	141 at FU2
Eligible (% of respondents)	27 (18%)	55 (39%)
Accepted (% of eligible)	18 (67%)	14 (25%)
Attended at least 1 session (% of accepted)	10 (56%)	14 (100%)
Number attending at least.... sessions		
1	13 (100%)	14 (100%)
2	10 (77%)	14 (100%)
3	10 (77%)	14 (100%)
5	9 (69%)	14 (100%)
8	6 (46%)	13 (93%)
10	3 (34%)	11(79%)
12	*	3(21%)
Number of sessions that occurred	32	46
Median (IQR)	3.00 (2,4)	2.50 (2,4)
Min, Max	1, 5	1, 7
Number including co-parents who attended one or more sessions	13 (62%)	14 (100%)
Median sessions attended (IQR)	7.00 (3, 9)	10.50 (10, 11)
Min, Max	1, 10	5, 12

These results present a mixed picture in relation to the progression criteria for intervention delivery. Whilst group leaders were successfully identified and trained, venues secured, and both IY-B and IY-T were successfully delivered, the number of IY-I groups was lower than expected as was the group size in some cases. Consequently, the criterion for intervention delivery was assessed as only partially achieved.

Intervention acceptability

Group retention levels All those who accepted a place on IY-T attended at least one session (100%), whereas 44% of participants accepting an IY-I place did not attend at least one session (see Table 2). The intervention attrition rate was comparable for IY-I and IY-T up to 5 sessions, but was subsequently higher for IY-I than for IY-T. The number of people in each session was higher for IY-T. The criterion of 70% retention at group end was met for both IY-I (73% of those who attended at least one session retained) and IY-T (87% of those who attended at least one session retained).

Parent satisfaction A total of 219 weekly evaluation forms were completed by research parents across both programs and sites. Out of 76 weekly attendees, 68 evaluation forms (89.5% response rate) were completed for the IY-I program, while out of 143 weekly attendees, 119 responses (83.2%

response rate) were completed for IY-T groups. Responses for all groups across all weeks, and at the end of the program were generally positive and indicated a high level of parent satisfaction for program content and delivery (See Appendix C). The evaluation form also allowed parents to add what they found most useful from the session. Parents from IY-I highlighted playing, routine and observing your child's reaction when playing. IY-T parents noted ignoring bad behavior, praising good behavior and the benefits of routine. Parents also found it reassuring that others were struggling with the same things, valued hearing new stories and experiences and bouncing ideas off one another. When asked about group leader ability, parents emphasized the flexibility of the group leaders, space for discussion and the provision of positive information and feedback. Issues for development highlighted included the video vignettes being paused too frequently, that they were Americanized, outdated and mainly show "ideal world" scenarios. Parents said that sites should also invite more parents to the groups, start the course earlier in the baby's life, make the location of the groups more accessible and increase the length of course and sessions. The satisfaction data provides supplementary information on intervention acceptability, when considered alongside group retention levels the criterion on intervention acceptability was met.

Fidelity of intervention

Adherence All IY-I and IY-T groups met the progression criterion threshold of 80% self-reported adherence on the number of vignettes shown (IY-I: Site 1—95%, Site 2—91%; IY-T: Site 1 $m=6$, Site 2 $m=5$) and "did I" statements (IY-I: Site 1 80%, Site 2 IY-I; IY-T: Site 1—83%, IY-T 80%).

Quality of delivery The notional threshold of 80% was met or exceeded for delivery quality as measured via the PPIC, with the group leaders' observed adherence to facilitation processes scoring the lowest for all three IY-I groups. Group leaders scored higher for all three PPIC constructs (adherence, quality of delivery, and participant responsiveness) for IY-T, compared to IY-I (see Table 3).

Summary assessment of Progression Criteria

We successfully, partially or fully, achieved 5 of our 6 progression criteria (see Table 1). The recruitment criterion was not met but the impact on study feasibility was mitigated by high levels of participant retention. The assessment of progression criteria was submitted to the funder (NIHR) on completion of the pilot, and approved progression to the definitive main trial was given. Opportunities to optimize main trial design and processes were identified based on

Table 3 Mean scores for IY infant and IY toddler on the PPIC measure of implementation fidelity (max score of 5)

Program	Site	Session number	Adherence	Quality	Parent responsiveness	
IY-I	Site 2	3	3.38 (67.6%)	4.14 (82.8%)	3.67 (73.4%)	
		8	3.00 (60.0%)	4.14 (82.8%)	3.67 (73.4%)	
	Site 1A	3	3.88 (77.6%)	4.14 (82.8%)	4.67 (93.4%)	
		8	3.25 (65.0%)	4.43 (88.6%)	3.67 (73.4%)	
	Site 1B	3	2.75 (55.0%)	3.43 (68.6%)	3.00 (60.0%)	
		8	2.75 (55.0%)	4.43 (88.6%)	5.00 (100%)	
	IYT	Site 2 Group A	2	4.00 (80.0%)	4.14 (82.9%)	4.33 (86.7%)
			9	3.63 (72.5%)	4.57 (91.4%)	3.67 (73.3%)
Site 2 Group B		2	2.88 (57.5%)	4.00 (80.0%)	4.00 (80.0%)	
		9	3.00 (60.0%)	4.29 (85.7%)	3.67 (73.3%)	
Site 1 Group A		2	4.00 (80.0%)	4.43 (88.6%)	4.67 (93.3%)	
		9	3.88 (77.5%)	4.29 (85.7%)	5.00 (100.0%)	
Site 1 Group B		2	3.50 (70.0%)	4.14 (82.9%)	5.00 (100.0%)	
		9	2.29 (45.7%)	4.00 (80.0%)	3.67 (73.3%)	

Table 4 Primary outcome/screener main differences between arms

Time	Treatment			Control			Adjusted ¹ Mean diff [95% CI]	Mean diff [95% CI]
	n (%)	Mean	SD	n (%)	Mean	SD		
ASQ:SE-2								
BL	152 (100%)	22.35	12.74	53 (100%)	18.32	11.69	3.12 [−1.51, 7.74]	4.03 [0.13, 7.93]
FU1 (2-MO)	70 (46%)	16.17	11.9	19 (36%)	11.32	7.97		4.86 [−0.82, 10.54]
FU1 (6-MO)	146 (96%)	22.38	16.28	50 (94%)	17.5	12.83		4.88 [−0.09, 9.85]
FU2	141 (93%)	27.77	17.68	47 (89%)	26.06	17.03		1.70 [−4.09, 7.49]
FU3	137 (90%)	24.31	19.25	43 (81%)	25.93	20.48		−1.62 [−8.32, 5.08]
PHQ-9 reported by parent								
BL	152 (100%)	2.7	2.84	53 (100%)	2.53	2.89	0.6 [0.48, 0.72]	0.17 [−0.72, 1.06]
FU1	147 (97%)	2.23	3.16	50 (94%)	1.64	2.26		0.59 [−0.36, 1.54]
FU2	141 (93%)	2.42	3.08	48 (91%)	1.81	2.12		0.61 [−0.33, 1.55]
FU3	137 (90%)	2.31	3.19	44 (83%)	2.2	2.64		0.10 [−0.94, 1.14]

ASQ:SE-2 Ages and Stages Social Emotional 2nd edition, PHQ-9 Patient Health Questionnaire-9, BL Baseline, FU1 follow-up 1, FU2 follow-up 2, FU3 follow-up 3, 2-MO ASQ:SE-2 for 2 month olds, 6-MO ASQ:SE-2 for 6 months old

^aCalculated using multiple linear regression with covariates adjusting for site and stratifying factors (sex of child, initial level of child’s social and emotional well-being using ASQ:SE-2 and initial level of parent depression measured using PHQ-9)

challenges, lessons learnt and remedial actions in the pilot, as described in the discussion section (and summarized in Table 1).

Secondary endpoints

Outcomes and estimations

This pilot study was not powered for effectiveness analysis, but observations based on consideration of the estimates of SD are reported. After considering the reported estimates of standard deviation, no systematic difference between arms for the child primary outcome at any of the time points (see Table 4) are apparent, although scores tended to be higher

(i.e., possibly worse) in the intervention arm at all time-points except FU3. However, the estimated mean difference between arms after adjusting for initial levels of parent depression and child social and emotional well-being is lower than the unadjusted mean difference.

Prevalence of monitoring levels for child and parent primary outcomes

Scores above a defined cut-off value on the ASQ:SE-2 indicate a recommendation to “review behaviors of concern and monitor”. The number of children at monitoring level was generally lower at follow up than at baseline in both arms (see Table 5). Scores of ≥5 on the PHQ-9 suggest

Table 5 Prevalence of ASQ:SE-2 and PHQ-9 (Parent) monitoring levels

Time	Intervention		Control		Overall	
	Number responding	Eligibility criteria met	Number responding	Eligibility criteria met	Number responding	Eligibility criteria met
ASQ:SE-2						
BL	152	64 (42%)	53	18 (34%)	205	82 (40%)
FU1	70	18 (26%)	19	3 (16%)	89	21 (24%)
FU2	146	42 (29%)	50	9 (18%)	196	51 (26%)
FU3	141	30 (21%)	48	14 (29%)	189	44 (23%)
PHQ-9 \geq 5						
BL	152	53 (35%)	53	14 (26%)	205	67 (33%)
FU1	147	49 (33%)	50	12 (24%)	197	61 (31%)
FU2	141	59 (42%)	48	16 (33%)	189	75 (40%)
FU3	137	58 (42%)	44	20 (45%)	181	78 (43%)
PHQ-9 \geq 4						
BL	152	65 (43%)	53	17 (32%)	205	82 (40%)
FU1	147	60 (41%)	50	16 (32%)	197	76 (39%)
FU2	141	74 (52%)	48	19 (40%)	189	93 (49%)
FU3	137	67 (49%)	44	24 (55%)	181	91 (50%)

Eligibility criteria for ASQ:SE-2 score at monitoring levels or above; PHQ-9 mild to severe depression

ASQ:SE-2 Ages and Stages Social Emotional 2nd edition, PHQ-9 Patient Health Questionnaire-9, BL Baseline, FU1 follow-up 1, FU2 follow-up 2, FU3 follow-up 3

parents may be experiencing mild to severe depression. In contrast to the child primary outcome, the number of parents in the intervention and control arms with mild to severe depression was slightly higher at FU2 and FU3 than at baseline (see Table 5). The lower than expected proportion of parents with depression symptoms in the sample reduced the number of parents eligible for the IY-I and IY-T program at FU1 and FU2. Because of this we reduced the threshold for eligibility to IY-I and IY-T from 5 to 4 for the pilot, this was designed as a temporary strategy to ensure that there were enough eligible participants to enable the piloting of group delivery processes. Moreover, we added ASQ:SE-2 as a screener because we noticed that some parents who scored their child in the monitoring zone of the ASQ:SE-2 screener (child primary outcome) did not score themselves as depressed on the PHQ-9 screener.

Discussion

The results suggested that a proportionate universal approach could be successfully delivered with a suite of IY programs. The progression criteria were largely achieved.

Recruitment

The study did not achieve the target sample size at baseline. There were initial barriers around participant identification and recruitment, for example, participant

identification began the week that commissioning for health visiting (HV) in England moved from the NHS to Local Authorities. Although senior site management were fully committed to the trial, health visitors faced a lot of uncertainty and had limited capacity which may have led to reduced identification and recruitment. Documentation with regard to working partnerships, expectations and site requirements should be clear and detailed in a definitive trial. Awareness raising and training resources for health visitors and other staff involved in identifying potential participants, in an electronic format that can be easily shared and accessed by professionals working in a range of contexts, might also improve recruitment rates. One site had a late start, due to delays in securing finance to cover health visitor time to train in and deliver the IY programs. This impacted on time available for participant identification and recruitment, however a simplified process for seeking intervention delivery costs and approvals from the UK's Health Research Authority (HRA) for trials conducted in health and social care settings should alleviate this issue for a future definitive UK trial of E-SEE STEPS.

The challenges of recruiting to RCTs are well documented, with one recent review suggesting that 56% of publicly funded RCTs in the UK fail to achieve their target sample size (Walters et al., 2017). The Qualitative Research Integrated within Trials (QuinteT) Recruitment Intervention (Donovan et al., 2016) provides a methodology for understanding the process of recruitment to RCTs, identifying difficulties and

producing a plan to address those difficulties, all in real-time. Our pilot study suggests that this approach should be considered in the design of a definitive trial.

Retention

Despite achieving only 205 (71%) of the target sample of 288, the high retention rate of 181 (88%) participants at final follow-up approximated the expected number at trial end ($n = 192$, 94%), as 32% attrition was estimated, but only 12% occurred. This level of retention is encouraging, particularly when considering that families were recruited during a major life event (soon after birth of a child), the inclusion of four timepoints and final follow-up was approximately 18 months after baseline. The range of strategies we employed to support retention should be replicated in a definitive trial. The PAC reviewed the measures and materials and participated in training our data collectors in data collection procedures (using role-play scenarios) during study set-up. This ensured that measures were ‘parent friendly’ and acceptable to parents of young children, and that data collectors were confident in administering them. Data collectors left ‘change of address’ forms and freepost envelopes along with E-SEE branded trolley tokens and pens (that had our contact telephone number on) with participants at each data collection visit. Parents who notified us of a change of address received a £10 shopping voucher as a token of appreciation. When booking data collection visits we contacted parents by phone to discuss and arrange appointments and they received a letter in the mail confirming the dates and times of the appointment and the name of the visiting data collector. We had teams of data collectors local to each site, and wherever possible we ensured the researcher who visited the family was consistent across each timepoint. When we could not reach a parent by phone, we sent an ‘appointment by letter’ notifying the parent that a named data collector would visit them at a specific date and time or they could arrange a more convenient time (if the letter was unsuccessful twice we marked the participant ‘lost to follow-up’). Those who we were unable to contact and marked lost to follow-up were contacted again at the next follow-up timepoint. We are unable to pinpoint specifically which, if any, of these strategies resulted in the successful retention of parents, and therefore recommend future studies explore the differential effectiveness of such strategies (for example using ‘Studies within a trial’ (SWAT) approaches, Clarke et al., 2015).

Intervention Delivery

We initially envisaged Health Visitors co-leading all intervention groups with LA staff, however with the move of

health visiting commissioning arrangements to LAs, changes in service structures and budgets, and the fact that services are delivered differently across sites we propose that delivery agents should be specific to local context needs, that is, both HVs and other LA staff could be IY group leaders, or just LA staff such as children center staff, or just health visiting staff. This flexibility reflects variation in real-world service set-up.

The planned number of IY-T groups was delivered in the two sites, however we delivered fewer IY-I groups than planned, due to a lower than expected percentage of eligible parents at FU1. Implementation partners in both sites also struggled to sufficiently engage and retain eligible parents to meet the target of at least five parents attending at least 50% of sessions in each group, although eight or more parents were invited per group. Reasons why uptake of the invitation and initial attendance may have been low are outlined below.

The number of parents meeting the designated threshold on the depression screener (PHQ-9) was much lower than expected. This could have been because not all families were informed of the trial, or that families with threshold scores chose not to hear more about the trial from the researchers, or that the population depression data we used to estimate prevalence of maternal depression was inaccurate. The low number of parents with depressive symptoms meant fewer eligible parents for the IY-I and IY-T programs at FU1 and FU2. We therefore reduced the threshold temporarily from a PHQ-9 score of 5 to 4 for the pilot, and added the ASQ:SE-2 as a screener as it was the primary outcome plus we noticed that some parents who scored their child in the ASQ:SE-2 monitoring zone did not score themselves as depressed on the PHQ-9 screener. Since the IY programs were chosen to enhance child social emotional well-being, and the demonstrated relationship between children’s early difficulties and parental mental health (Quist et al., 2019), using two screeners is acceptable. Both screeners should be used in a definitive trial.

Timing of groups and availability of wrap-around support may have been a factor for low uptake. One site had staff capacity issues that meant that the IY-T program could only be run on limited weekdays, with no evening sessions and no crèche facility (a crèche is supervised childcare offered to parents while parents undertake some activity, usually on the same premises). Inflexible delivery times made it difficult for working parents, or those with childcare challenges to attend a group. Flexible group delivery times, to suit parents, should be included in site level agreements in a definitive trial. In addition, some groups were delivered during school holidays to meet study timeframes, which reduced parents’ ability to attend (e.g., those with older children). The timings for IY-I and IY-T delivery in a definitive trial should be mapped to ensure they can be delivered in term time to retain inclusivity.

To ensure future viable groups (minimum of 5 parents) we recommend inviting up to 12 parents per group, working with sites to implement effective engagement and retention strategies, and allowing ‘non-research’ participants to be invited to low-uptake groups to promote collaborative learning. Non-research participants would need to have similar levels of need to eligible research parents and have children a similar age, while no research data would be collected from them, their presence enables group dynamics to operate (a key element of the IY-I and IY-T theory of change). Additional support within the model to boost engagement of parents in the intervention at the outset should be considered. For example, Fosco et al. (2014) included motivational interviewing techniques in the Family Check-Up model and found a significant relationship with intervention response.

Intervention Acceptability

Group retention levels for participants who attended at least one of the weekly sessions were high; 73% for IY-I and 87% for IY-T. This suggests that participants found the intervention acceptable and is consistent with reported high levels of satisfaction in weekly and end of program questionnaires. These findings are comparable to other IY studies (e.g., Morpeth et al., 2017; Jones et al., 2016). However, participants did highlight potential areas of improvement for the programs that could be considered in a future trial (see Table 1).

We encountered unanticipated challenges in relation to practitioner perspectives on the acceptability of IY. The IY programs were developed in the US and, during the pilot, sites raised concerns about risk to their UNICEF Baby Friendly Initiative (BFI) accreditation as the IY-B book and IY-I program contained some information around feeding and sleeping that did not conform to BFI recommendations (although it conformed to US and World Health Organization guidance). This was initially managed by inserting a one-page addendum to the book informing parents that the book was from the US and signposting them to UK guidance. Local BFI Infant Feeding Coordinators, HVs, IY trainers and members of our PAC inputted to the addendum. The research team were transparent about the BFI concerns and worked closely with the IY developer, IY trainers, local service staff and local BFI staff while the developer revised the book and published an updated version in 2017 which is now used worldwide. These challenges highlight the limitations of printed books, as they can quickly become outdated as guidance inevitably changes. E-books or online materials can be more easily updated (although there are associated risks in relation to digital exclusion).

Intervention Fidelity

Notional thresholds of 80% fidelity were largely achieved in relation to different aspects of intervention fidelity

(particularly in relation to quality of delivery and participant responsiveness). There were some discrepancies between self-reported adherence and objective coding of videotaped sessions by researchers. Subjective, desirability biases, can influence the outcomes of self-report measures of fidelity (Bywater et al., 2019) highlighting the importance of incorporating independent measures of fidelity in trials. Overall IY-I and IY-T were being delivered as intended and the training and supervision model for E-SEE STEPS is sufficient. However, “dry run” delivery practice for newly trained group leaders is recommended to promote intervention fidelity within a definitive trial.

Research Questions and Analyses for Definitive Trial

The E-SEE trial was originally designed to deliver four primary analyses examining the evidence for the overall effect of the intervention at FU3, and the effect of the individual dose levels of the intervention (i.e., the IY book, IY-I and IY-T). However, the evaluation of the component level effects was always going to be compromised due to the lack of randomization at each subsequent proportionate stage. While a series of two further randomized stages would have enabled us to evaluate the individual components, it would have made the trial prohibitively large and hence power to detect an overall effect of the intervention was prioritized. We therefore, propose an alternative design consisting of a single research question “Do the scores of children in the IY arm, on average, stay below those scores for children in SAU over the three follow-up measures?” The benefit of asking a single research question is that power is increased and there is no need for Bonferroni adjustment. The disadvantage is that we are unable to examine the potential effectiveness of each of the three levels, and each of the four doses of intervention. Arguably this approach is more appropriate, given that the intention is to implement and scale E-SEE STEPS as a whole proportionate universal intervention, rather than individual components or programs.

Sample Size Calculations for Definitive Trial

We will use all the follow-up measurements in a single mixed effects model repeated measures analysis informed by the results generated by the secondary endpoints from the pilot. Rather than simply expecting scores to be lower at FU3 adjusted for baseline, we now expect the average scores to be lower at each follow-up in IY compared to SAU. We assume that correlation with baseline and FU3 will be the same as observed in the pilot for baseline and FU2 (0.26). As was justified in the original design we set the clinically important difference at FU3 to be 5 units of the ASQ:SE-2. We expect this effect to be consistently seen over the three follow-up

points (original analyses assumed 5.5 points difference at FU2, so we are allowing for a slightly lower average effect overall). The sample size for the full definitive trial is calculated bearing in mind we need to ensure that there are sufficient numbers of eligible and willing to attend parents for the IY-I and IY-T groups to be viable. Groups need to have at least 5 parents attending. So our calculations are driven to deliver an expected total of 48 parents who will attend the IY-T sessions. We further assume that the IY-I and IY-T groups may lead to treatment induced clustering of outcomes and hence apply a design effect inflation factor to account for this clustering in one arm. These two requirements lead to an unbalanced allocation ratio. Assuming an average effect of 5 units below SAU, the SD at FU3 is 18 and a design effect of 1.25 for the IY arm with a two-sided 5% significance level and 90% power we would require retention at FU3 of 441 in SAU and 92 in IY. Allowing for the overall attrition of 12% we would require 606 to be randomized with an allocation ratio of 4.8:1 for the definitive trial.

Practical Implications of the Study for Clinical Practice, Training or Policy

The study demonstrates that LA and HV services can work together to deliver IY-I and IY-T in a proportionate universal model. When delivering parent programs services should consider using both a parent and child screener to ensure families get the support they need, and enhanced engagement activities may be required in order to improve the uptake of offer to IY. The findings imply that any trial of an intervention to promote social and emotional development of infants needs to pay careful attention to whether intervention materials comply with BFI guidance as BFI accredited services may be wary about delivering interventions which may threaten their accreditation status.

Limitations

The sample were not as depressed as population data suggested, which led to lower eligibility numbers for the IY groups. The E-SEE STEPS model is a proportionate universal approach and a representative sample is key to test the model. We are exploring the issues of representative samples within trials and generalizability (Gridley et al., submitted) using this pilot data and data from a similar study (Leckey et al., 2019). Low take up of IY-I and IY-T is a limitation that needs to be address in a future definitive trial.

Conclusion

Progression to a definitive trial was acceptable with the design amends outlined above, and with a recalculated

sample size. The published main trial protocol (Bywater et al., 2018) was informed by the results and lessons learned during the pilot.

Trial registration number

ISRCTN11079129.

Data and/or Code Availability

Requests for participant-level quantitative data and statistical codes to be made to the corresponding author and each request to be considered by members of the original TMG, including the chief investigator and members of Sheffield CTRU. Data sharing principles as described by Tudur et al. (2015) will be applied. The data will not contain direct identifiers, we will minimize indirect identifiers and remove free text data to reduce the risk of identification.

Acknowledgements We express sincere thanks to; all families in our patient and public involvement activities during the preparation of the application and families who participated in our study Parent Advisory Committee or pilot trial; our Trial Steering Committee and Data Monitoring and Ethics Committee; our site collaborators and data collectors; Kelly Blockley for process evaluation support, Kath Kiernan for her study support and expertise, Carolyn Webster-Stratton (Incredible Years (IY) developer) and IY mentors who provided training and supervision (Kate Rhee, Sue Evans, Linda Bennett and Jeannie Gordon); Kiera Solaiman for support with formatting, our funders Public Health England who provided excess treatment costs and the NIHR for funding the study, and our sponsor, the University of York. T.B., V.B., S.B. and G.R. are supported by the National Institute for Health Research Applied Research Collaboration, T.B., S.B., G.R. by Yorkshire and Humber, and V.B. by South West Peninsula.

Author Contributions All authors substantially contributed to the conception or design of the work, and/or the acquisition, analysis, or interpretation of data for the work, in addition to drafting the work or revising it critically for important intellectual content. All authors gave final approval of this version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This study was funded by the National Institute for Health Research's Public Health Research Program (NIHR PHR 13/93/10).

Compliance with Ethical Standards

Conflict of Interest All authors declare no competing interests, except the CI, T.B., who is a Trustee for Children's Early Intervention Trust (CEIT). Early Intervention Wales Training (EIWT) is owned by CEIT and offers training courses, including Incredible Years (I.Y.). Trustees do not benefit financially from trainings or CEIT/EIWT activities. ESEE study trainings were arranged via the IY developer in the U.S.

Ethics Approval Approved by University of York (sponsor) Education Ethics Committee (ref: FC15/03, 10 August 2015); UK NHSREC 5 (ref: 15/WA/0178, 22 May 2015). The pilot protocol is V6 (08.05.2017).

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bywater, T., Berry, V., Blower, S. L., Cohen, J., Gridley, N., Kiernan, K., Mandefield, L., Mason-Jones, A., McGilloway, S., Mckendrick, K., Picket, K., Richardson, G., Teare, D.T., Tracey, L., Walker, S., Whittaker, K. & Wright, J. (2018). Enhancing Social-Emotional Health and Wellbeing in the Early Years (E-SEE): a study protocol of a community-based randomised controlled trial with process and economic evaluations of the incredible years infant and toddler parenting programmes, delivered in a proportionate universal model. *BMJ open*, *8*(12), e026906.
- 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.
- Gridley, N., Hickey, G., Bywater T., McGilloway, S., Blower, S., Whittaker, K. & Berry, V. (2021). The Challenges of Assessing Sample Representativeness within Community-Based Evaluations of Parenting Programmes delivered in England and Ireland [Manuscript submitted for publication]. Carnegie School of Education, Leeds Beckett University.
- Adams, G., Gulliford, M. C., Ukoumunne, O. C., Eldridge, S., Chinn, S., & Campbell, M. J. (2004). Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *Journal of Clinical Epidemiology*, *57*(8), 785–794.
- Axford, N., Lehtonen, M., Kaoukji, D., Tobin, K., & Berry, V. (2012). Engaging parents in parenting programs: lessons from research and practice. *Children and Youth Services Review*, *34*(10), 2061–2071.
- Axford, N., Bjornstad, G., Matthews, J., Heilmann, S., Raja, A., Ukoumunne, O. C., & Warner, G. (2020). The effectiveness of a therapeutic parenting program for children aged 6–11 years with behavioral or emotional difficulties: results from a randomized controlled trial. *Children and Youth Services Review*, *117*, 105245.
- Barlow, J., Smailagic, N., Ferriter, M., Bennett, C., & Jones, H. (2010). Group-based parent-training programmes for improving emotional and behavioural adjustment in children from birth to three years old. *Cochrane Database of Systematic Reviews*, (3), CD003680.
- Benach, J., Malmusi, D., Yasui, Y., & Martínez, J. M. (2013). A new typology of policies to tackle health inequalities and scenarios of impact based on Rose's population approach. *Journal of Epidemiol Community Health*, *67*(3), 286–291.
- Bender, H. L., Allen, J. P., McElhaney, K. B., Antonishak, J., Moore, C. M., Kelly, H. O. B., & Davis, S. M. (2007). Use of harsh physical discipline and developmental outcomes in adolescence. *Development and Psychopathology*, *19*(1), 227.
- Candlish, J., Teare, M. D., Cohen, J., & Bywater, T. (2019). Statistical design and analysis in trials of proportionate interventions: a systematic review. *Trials*, *20*(1), 151.
- Carey, G., Crammond, B., & De Leeuw, E. (2015). Towards health equity: a framework for the application of proportionate universalism. *International journal for Equity in Health*, *14*(1), 81.
- Clarke, M., Savage, G., Maguire, L., & McAneney, H. (2015). The SWAT (study within a trial) programme; embedding trials to improve the methodological design and conduct of future research. *Trials*, *16*(2), 1–1.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2013). Developing and evaluating complex interventions: the new Medical Research Council guidance. *International Journal of Nursing Studies*, *50*(5), 587–592.
- Department of Health. (2009). *Healthy child programme*. London: Department of Health. Accessed 1/5/2020 <https://www.gov.uk/government/publications/healthy-child-programme-pregnancy-and-the-first-5-years-of-life>.
- Donovan, J. L., Rooshenas, L., Jepson, M., Elliott, D., Wade, J., Avery, K., Mills, N., Wilson, C., Paramasivan, S., & Blazeby, J. M. (2016). Optimising recruitment and informed consent in randomised controlled trials: the development and implementation of the Quintet Recruitment Intervention (QRI). *Trials*, *17*(1), 283. <https://doi.org/10.1186/s13063-016-1391-4>.
- Fosco, G. M., Van Ryzin, M., Stormshak, E. A., & Dishion, T. J. (2014). Putting theory to the test: examining family context, caregiver motivation, and conflict in the Family Check-Up model. *Development and Psychopathology*, *26*(2), 305.
- Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., & Smith, S. M., & Donnelly, M. (2012). Behavioural and cognitive-behavioural group-based parenting programme for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database of Systematic Reviews*, 2. <https://doi.org/10.1002/14651858.cd008225.pub2>
- Galbally, M., & Lewis, A. J. (2017). Depression and parenting: the need for improved intervention models. *Current Opinion in Psychology*, *15*, 61–65.
- Gardner, F., Leijten, P., Mann, J., Landau, S., Harris, V., Beecham, J., ... & Scott, S. (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).
- Hurt, L., Paranjothy, S., Lucas, P.J., Watson, D., Mann, M., Griffiths, L.J., Ginja, S., Paljarvi, T., ... & Lingam, R., (2018). Interventions that enhance health services for parents and infants to improve child development and social and emotional well-being in high-income countries: a systematic review. *BMJ Open*, *8* (2). <https://doi.org/10.1136/bmjopen-2016-014899>
- Hutchings, J., Griffith, N., Bywater, T., Williams, M. E., & Baker-Henningham, H. (2013). Targeted vs universal provision of support in high-risk communities: comparison of characteristics

- in two populations recruited to parenting interventions. *Journal of Children's Services*, 8(3), 169–182.
- Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, 105(11), 2283–2290.
- Jones, C. H., Erjavec, M., Viktor, S., & Hutchings, J. (2016). Outcomes of a comparison study into a group-based infant parenting programme. *Journal of Child and Family Studies*, 25(11), 3309–3321. <https://doi.org/10.1007/s10826-016-0489-3>.
- Kalil, A., & Ryan, R. (2020). Parenting practices and socioeconomic gaps in childhood outcomes. *The Future of Children*, 30(2020), 29–54.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.
- Leadsom, A., Field, F., Burstow, P., & Lucas, C. (2013). The 1001 critical days: the importance of the conception to age two period. A Cross-Party Manifesto. Accessed 1 May 2020. <https://www.nw.cscsenate.nhs.uk/files/8614/7325/1138/1001cdmanifesto.pdf>
- Leckey, Y., Hickey, G., Stokes, A., & McGilloway, S. (2019). Parent and facilitator experiences of an intensive parent and infant programme delivered in routine community settings. *Primary Health Care Research & Development*, 20, 1–12.
- Leijten, P., Gardner, F., Melendez-Torres, G. J., Van Aar, J., Hutchings, J., Schulz, S., Knerr, W., & Overbeek, G. (2019). Meta-analyses: Key parenting program components for disruptive child behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58(2), 180–190.
- Menting, A. T., de Castro, B. O., & Matthys, W. (2013). Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behavior: a meta-analytic review. *Clinical Psychology Review*, 33(8), 901–913.
- McGilloway, S., Mhaille, G. N., Bywater, T., Furlong, M., Leckey, Y., Kelly, P., Comiskey, C., & Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: a randomized controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology*, 80(1), 116.
- 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 United Kingdom: a pragmatic randomised controlled trial. *Child Care in Practice*, 23(2), 141–161.
- Perrin, E. C., Sheldrick, R. C., McMenamy, J. M., Henson, B. S., & Carter, A. S. (2014). Improving parenting skills for families of young children in pediatric settings: a randomized clinical trial. *JAMA Pediatrics*, 168(1), 16–24.
- Pidano, A. E., & Allen, A. R. (2015). The Incredible Years series: a review of the independent research base. *Journal of Child and Family Studies*, 24(7), 1898–1916.
- Pontoppidan, M., Klest, S. K., & Sandoy, T. M. (2016). The incredible years parents and babies program: a pilot randomized controlled trial. *PLoS One*, 11(12).
- Quist, M., Kaciroti, N., Poehlmann-Tynan, J., Weeks, H. M., Asta, K., Singh, P., & Shah, P. E. (2019). Interactive effects of infant gestational age and infant fussiness on the risk of maternal depressive symptoms in a nationally representative sample. *Academic Pediatrics*, 19(8), 917–924.
- Romeo, R., Knapp, M., & Scott, S. (2006). Economic cost of severe antisocial behaviour in children-and who pays it. *The British Journal of Psychiatry*, 188(6), 547–553.
- Ryan, R., O'Farrelly, C., & Ramchandani, P. (2017). Parenting and child mental health. *London Journal of Primary Care*, 9(6), 86–94.
- Sadler, K., Vizard, T., Ford, T., Marchesell, F., Pearce, N., Mandalia, D., Davis, J., Brodie, E., Forbes, N., Goodman, A. & Goodman, R., 2018. Mental health of children and young people in England, 2017.
- Simkiss, D. E., Snooks, H. A., Stallard, N., Davies, S., Thomas, M. A., Anthony, B., & Stewart-Brown, S. (2010). Measuring the impact and costs of a universal group based parenting programme: protocol and implementation of a trial. *BMC Public Health*, 10(1), 1–11.
- Squires, J., Bricker, D., & Twombly, E. (2015). *ASQ:SE-2 User's guide*. 2nd ed. Baltimore: Paul Brookes Publishing Company.
- Tudur Smith, C., Hopkins, C., Sydes, M.r., Woolfall, K., Clarke, M., Murray, G., & Williamson, P. (2015) How should individual participant data (IPD) from publicly funded clinical trials be shared?. *BMC Medicine*, 13(1), 1–7.
- Walters, S. J., dos Anjos Henriques-Cadby, I. B., Bortolami, O., Flight, L., Hind, D., Jacques, R. M., Knox, C., Nadin, B., Rothwell, J., Surtees, M., & Julious, S. A. (2017). Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. *BMJ Open*, 7(3), e015276.
- Webster-Stratton, C. (2011). *The Incredible Years: parents, teachers, and children's training series: program content, methods, research and dissemination 1980–2011*. Seattle: Incredible Years.
- Winsper, C., Bilgin, A. & Wolke, D. (2020). Associations between infant and toddler regulatory problems, childhood co-developing internalising and externalising trajectories, and adolescent depression, psychotic and borderline personality disorder symptoms. *Journal of Child Psychology and Psychiatry*, 61(2), 182–194.

Affiliations

Sarah L. Blower¹ · Vashti L. Berry² · Matthew C. Bursnall³ · Judith Cohen^{3,8,8} · Nicole Gridley^{1,9,9} · Amanda Loban³ · Laura Mandefield^{3,10,10} · Amanda J. Mason-Jones¹ · Sinéad McGilloway⁴ · Kirsty L. McKendrick³ · Siobhan B. Mitchell² · Kate E. Pickett¹ · Gerry A. Richardson⁵ · M. Dawn Teare³ · Louise C. Tracey⁶ · Simon M. Walker⁵ · Karen A. Whittaker⁷ · Jessica Wright³ · Tracey J. Bywater¹

¹ Department of Health Sciences, University of York, Seeborn Rowntree Building, Heslington YO10 5DD, UK

² College of Medicine and Health, South Cloisters, University of Exeter, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK

³ Sheffield Clinical Trials Research Unit, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

⁴ Centre for Mental Health and Community Research John Hume Building, Maynooth University, Maynooth, Co Kildare W23 F2H6 Ireland, UK

- ⁵ Centre for Health Economics, University of York, Heslington YO10 5DD, UK
Hull HU6 7RX, UK
- ⁶ Department of Education, University of York, Heslington, York YO10 5DD, UK
- ⁷ School of Nursing, University of Central Lancashire, Preston PR1 2HE Lancashire, UK
- ⁸ Present address: Hull Health Trials Unit, Hull York Medical School, University of Hull, Allam Medical Building,
- ⁹ Present address: Carnegie School of Education, Headingley Campus, Leeds Beckett University, Leeds LS6 3QS, UK
- ¹⁰ Present address: Department of Health Sciences, Seebohm Rowntree Building, University of York, Heslington YO10 5DD, UK