

ON ESTIMATING THE FISCAL BENEFITS OF EARLY INTERVENTION

Leon Feinstein,* Haroon Chowdry* and Kirsten Asmussen**

In this paper we explain some of the difficulties of providing forecasts of the financial benefits of early intervention programmes, focussing on those delivered during the early childhood period. We highlight the diversity of early intervention, and the complexity and multiplicity of outcomes. We summarise recent work at the Early Intervention Foundation to assess the evidence on the impacts of early intervention, recognising the diversity of approaches to delivery and the importance of innovation and local practice as well as of rigorous approaches to evaluating causal effects. We also describe new ways of assessing accurately the local fiscal costs of late intervention and consider the implications of this for addressing the well-established barriers to investment in prevention. Our analysis brings to the fore gaps in the evidence from which even the most rigorous ‘gold-standard’ research is not immune. These limitations prevent the production of an accurate and realistic cost-benefit ratio or net present value for the majority of programmes as delivered in practice. We suggest some paths towards a firmer foundation of evidence and a better alignment of evidence and policy.

Keywords: noncognitive skills; early childhood; child development; early intervention; programme evaluation; randomised controlled trials; cost-benefit analysis

JEL Classifications: H43, H51, H52, H53, H72, H75, I14, I24, I31, I38, J13, J18, J24

1. Introduction

The early childhood period, from conception to age 5, is recognised as an important period for children’s development (see e.g. Phillips and Shonkoff, 2000; Shonkoff *et al.*, 2009, 2012; Heckman, 2008). Children’s relationships and interactions with parents and caregivers are a key mechanism through which skills, behaviours, and other developmental milestones are achieved (Piaget, 2013; Asmussen *et al.*, 2016). It is also widely established that events, developments and shocks occurring in the first five years of a child’s life can be expected to have effects that last into adulthood (Almond and Currie, 2011). Thus, the broad case that early life matters to outcomes is well established.¹ Moreover, well-regarded direct (i.e. retrospective) cost-benefit analyses are available for a small number of highly stylised programmes, many of which were implemented in small, rigorously controlled and evaluated experiments in the past (e.g. Eckenrode *et al.*, 2010; Heckman *et al.*, 2010; Kitzman *et al.*, 1997, Campbell *et al.*, 2012). From this evidence we learn that

some forms of intervention, with sufficient specificity and intensity, can potentially make a substantive difference to child development in certain contexts. As a result, policies to promote improved parenting and home environments during this period have become of significant interest to local and national policymakers (Allen, 2011; National Audit Office, 2013), as well as the research and scientific communities. This includes activities known as early intervention, for babies, children and adolescents. We define early intervention as targeted activity (in response to emerging signals of additional need) which addresses problems before they become entrenched, thereby preventing costly outcomes to individuals, society and the state.

The Early Intervention Foundation (EIF) was set up in England by the coalition government of 2010–15, to draw together the evidence on this sort of intervention and support local authorities and other agencies in translating

*Office of the Children’s Commissioner. E-mails: haroon.chowdry@childrenscommissioner.gsi.gov.uk; Leon.FEINSTEIN@childrenscommissioner.gsi.gov.uk. **Early Intervention Foundation. We are grateful to funders, trustees and birth parents of the Early Intervention Foundation for establishing the Centre and steering the research agenda and to Carey Oppenheim, Jack Martin, Daniel Acquah and Donna Molloy for multiple comments and discussions about the work. We owe a particular debt to the many programme developers, practitioners and academics whose work we have reviewed and whose insights have taken the field forward and to the children and families who have participated in the research. We are grateful to Peter Fitzsimons, Lara Doubell, Georgia Vourloumi and Ilenia Piergallini for valuable research assistance. All errors and oversights remain our own.

this into action. The EIF was established as a ‘What Works Centre’, one of a growing number of independent organisations seeking to review and synthesise evidence on the effectiveness of policy and practice interventions and to share this with policymakers and practitioners in accessible ways. Other What Works Centres have tended to focus on specific domains of practice. Perhaps the best known example is the Education Endowment Foundation (EEF) which has centred on reviewing and developing the evidence on the cost and effectiveness of teaching practice in the classroom. The EEF’s Teaching and Learning Toolkit² assesses different teaching and learning methods, using ratings on a 1–5 scale to represent cost of delivery, quality of evidence and scale of impact. This approach is underpinned by three key features and assumptions: i) an emphasis on evidence that can establish causality through the use of high quality, randomised control trial designs that minimise the biases that result from observational approaches with weak, irrelevant or missing comparison groups or poor measurement; ii) a focus on the single outcome of progress in literacy or numeracy scores, expressed in a single currency of months of learning; iii) the assessment of cost in terms of staff time or other inputs.

This model, rooted in recognition of economic principles of cost effectiveness and a concern for valid causal inferences, provides an important benchmark for other What Works Centres such as on Local Economic Growth³ or Crime Reduction.⁴ This has stimulated a lively debate about the validity of RCTs as opposed to other forms of evaluation such as realist approaches or process evaluation (Cartwright, 2011; Bonell *et al.*, 2012; Marchal *et al.*, 2013; Pawson, 2013). Methodologists and practitioners have been inspired to engage in new ways in a fascinating example of the messy frontier between technical issues of methodology and the realities of practice.

This paper describes some of the challenges in applying this model to the field of early intervention, and sets out findings from two substantial studies the EIF has conducted that make small steps towards resolving these conundrums. Section 2 describes some of the issues and difficulties in applying the EEF model to early intervention, with a focus on the challenge of considering the fiscal benefits. Section 3 summarises recent EIF evidence on currently delivered early childhood programmes in the UK, and their levels of evidence of effectiveness. Section 4 reports on a new way of measuring the fiscal costs of late intervention. Section 5 concludes with a discussion of the challenges of integrating findings from sections 3 and 4 and makes a small number of recommendations for future research and practice in this field.

Carefully targeted early intervention and prevention activities are widely recognised to provide benefits for children and their families and wider society and government (Allen, 2011; Department of Health, 2013). However, local agencies often rightly seek evidence in the form of cost-benefit analysis demonstrating credible financial or social returns fit to their populations, location and investments. As will be seen, the current state of the evidence is some way from having this breadth and specificity of information. Therefore, we do not present here a model or assessment of long-term impact; instead, we seek to explain why the demand for simple cost-benefit numbers cannot yet be met meaningfully. We offer new ways of thinking about (a) what is now known about the effectiveness of available programmes that are currently widely commissioned, and (b) what is known about the fiscal costs of the ‘late intervention’ problems which those programmes seek to prevent. We hope in this way to mitigate against the kind of “magical thinking” (Brooks-Gunn, 2003) that can sometimes become embedded in policy debate.

2. Challenges in modelling and measuring the impact of early intervention

A first key challenge in estimating the impact of early intervention is the diversity of what is meant by early intervention and of approaches to delivering it. Early intervention as defined by the EIF is not just policy for the early years, but as action that is preventive, corresponding to the notion of secondary intervention in public health (e.g. Durlak and Wells, 1998), supportive of a reduction in harm where there is evidence that risk is above that for a population norm. The NAO landscape review (2013, *op. cit.*) emphasises the importance of this form of policy and practice, that is situated between universal services and the wide array of statutory and acute services that are necessarily involved when public harms are so intense for individuals, their families or communities that the state must intervene through, for example, child protection, criminal justice or mental health services. Early intervention can usefully be thought of as activity that exists between universal services such as GPs and schools, and these more statutory responses and ‘late’ intervention.⁵ Early intervention is a wide terrain that ranges from early years activities (such as targeted health visiting in low income communities) to support for adolescents (such as early identification and support for mental health difficulties). Hence early intervention refers to a wide array of different sorts of activity, in different settings, for children of different ages, at different levels of need, with need and risk defined in different ways, by different practitioner groups and professionals. Not

only is early intervention not one thing, the effectiveness of any form of early intervention will depend on how well it is targeted to meet need and how it fits with other parts of the local portfolio of supports and interventions.

A second issue we would highlight is external validity and the transferability of evidence from the best studies to the realities of local action. As stated above there is evidence from rigorous trials (such as Eckenrode *et al.*, 2010) of very specific forms of early intervention but this often describes activities that are very different from many local forms of early intervention. The issue of external validity has been highlighted recently by the attempt to transfer the success of the US programme Nurse Family Partnership to the UK as the Family Nurse Partnership (FNP). This form of evidence based policymaking has seen the attempt to replicate the success of a rigorously evaluated US programme in the UK, by adapting what has been shown to work in the US with a carefully managed roll out across multiple parts of England and a high quality randomised controlled trial (Robling *et al.*, 2016), following an earlier pilot study (Barnes *et al.*, 2009). Although this has been a careful and valuable trial, the early findings at 24 months have been disappointing for advocates of this translation approach.

We referred above to the diversity implicit in the term early intervention. There is also considerable variation in how each type of early intervention is delivered. The recently emerging field of implementation science has emphasised the difficulty for narrow definitions of evidence based policy that for human services interventions in part 'the practitioner is the intervention'.⁶ In practice, the child and family are also part of the intervention. This makes each local implementation subject to complex, individual level heterogeneity. The role of science and evidence in this approach is much more to support the quality of practice than to develop rigorous evaluations of gold standard products that can then be easily 'rolled out'.

As described in Section 3 the EIF has taken a different approach that has recognised more the diversity of existing practice and sought to develop more evidence based approaches by supporting existing local practice to improve its evidence, rather than by importing wholesale the most evidence based approaches.

The third issue that we highlight here is the multifinality and long-term nature of the impacts of early intervention (see e.g. Cicchetti and Rogosch, 1996), i.e. that human development allows a wide diversity of possible pathways from variable starting conditions to variable outcomes,

which makes specific impacts hard to predict accurately across multiple domains of outcome.

There are no large sample studies that monitor the impact of programmes through from the early years to later life in ways that can capture all of the hypothesised benefits in different domains of outcome. This adds to the difficulty of construction of a single currency, although in principle cost-benefit analysis is a way to evaluate various types of benefit in comparable forms – such as quality adjusted life years (QALYs), social value and hedonic pricing. In the case of early intervention, the quality of estimate of benefits depends on the degree to which short-run impacts can be linked to subsequent outcomes upon which an economic valuation can be placed.

Important work of this sort has been conducted elsewhere. For example, the Washington Institute for Public Policy (WSIPP)⁷ has combined meta-analyses of programme effects with other longitudinal research on the links between childhood circumstances and future adolescent and adult outcomes. In this way WSIPP has been able to produce hypothetical estimates of long-run economic and fiscal benefits – including net present values, cost-benefit ratios, and benefits by public sector agency – for a range of programmes.⁸ This form of appraisal has expanded the pool of programmes for which cost-benefit estimates are available. However, it focuses on a select group of branded, well-evidenced programmes, using evidence and modelling that is more relevant to US context⁹ and makes considerable assumptions about the relationships between early impact and long-term development. The lag between immediate impact and fiscal savings introduces unavoidable uncertainty over impacts that increase with the length of the lag. Moreover, ultimate impacts may be diffuse and hard to predict because of multifinality.

For the current crop of provision, especially early intervention and early years programmes delivered in the UK, there remains a lack of information on the expected potential returns on investment, in large part because there has not been enough modelling of local population trajectories in the absence of early intervention, linked to estimates of impact that can address the issues of long lags and multifinality. This serves as a barrier to the funding and commissioning of programmes, since local and national commissioners of services often demand credible, rigorous evidence of future budgetary savings before committing to invest in early intervention. Commissioners and policymakers are interested not just in the general scientific evidence that early intervention

might work in theory; they also demand evidence that it will work in practice when they commission it for their populations. Moreover, where cost-benefit analysis is based only on projections from the best-evidenced programmes, it is likely to be an overestimate of what may be achieved from more standard practice. A lack of evidence relevant to the majority of early intervention practice – as delivered ‘on the ground’ – therefore inhibits opportunities to improve outcomes and life chances for children.

3. Effective support for parent-child interactions in the UK

The Early Intervention Foundation was set up in part to begin to address some of these gaps in evidence. Because of the diversity of types of early intervention referred to above, a series of reviews have been established which consider the evidence for different forms of early intervention. The review we summarise here (Asmussen *et al.*, 2016) provides the most extensive summary of what is currently known about the effectiveness of programmes that aim to improve developmental outcomes of children up to age 5 through improved parent-child interactions.¹⁰ The definition of developmental outcomes included parent-child attachment, behaviour or self-regulation, and cognitive/language development. We reiterate that this particular theme – parent-child interactions in the early years – is just one type of early intervention. Other types of early intervention currently under EIF review include activities to support the relationship between parents, to support social and emotional skills in adolescence, and to support vulnerable groups such as children in gangs. In this paper, we focus on what was learnt about scale and nature of impact and provide a brief description of methods and findings, emphasising uncertainty and the need for innovation as well as the importance of rigour in evaluation of scale of impact.

A full description of the sample frame, methods of identification, methods of assessment and findings is provided on the EIF website.¹¹ This includes an up to date assessment through the EIF Guidebook of those programmes with the best evidence of impact.

The review assessed the research and evidence underpinning 75 programmes available in the UK. For each programme, the EIF assessment process generated a rating summarising the strength of evidence for impact on a child outcome, based on detailed and transparent standards of evidence.¹² The higher levels of this scale are based on the Maryland Scientific Methods Scale (Sherman *et al.*, 1998) and Flay *et al.*

Table 1. Evidence ratings used in the review

Level	Meaning
4	Evidence from at least two high-quality, well-conducted RCT/QED studies showing positive impact on a child outcome.
3	Evidence from one high-quality, well-conducted RCT/QED study showing positive impact on a child outcome.
2	Evidence from a weaker design (e.g./ pre-post or unmatched comparison) positive difference or change in a child outcome.
NL2	No direct evidence of impact.
NE	Evidence from a high-quality, well-conducted RCT/QED study showing no impact on children or parents. ^(a)

Note: (a) In principle, programmes may also have harmful effects or side effects and would receive a rating of ‘NE’. In this review, however, no such programmes were found.

(2005). The standards also include more formative types of evidence in order to recognise a wider range of activities than those with large sample trials. The ratings summarise the confidence with which one can state that a programme has demonstrably (and causally) led to an improvement in a child outcome; this combines a consideration of the quality of the research design with a binary judgement about whether the balance of the findings is consistent with a positive and statistically significant benefit. There are five levels to this EIF scale: Level 3 on this scale is equivalent to what is commonly considered as efficacy (Rush, 2009), or proof of concept under supportive circumstances; Level 4 equates to effectiveness at scale, with replicated impacts under multiple trials including adaptations. Both of these levels involve evidence where a causal link between exposure to an intervention and subsequent outcomes would be accepted as credible, subject to good implementation of the study and accurate data collection. By contrast, Level 2 would not count as causal evidence of impact in standard evidence-based policymaking frameworks such as the UK government’s Magenta Book (HM Treasury, 2011), but would be recognised as important steps in developing a sound understanding of the mechanism by which impacts might be achieved and how to measure them. The rating ‘NL2’ was used for programmes that do not have evidence from which even preliminary assessments about impact upon a child outcome could be made. Such programmes may only have examined effects on a parent outcome, or may have had qualitative or process evaluations. Programmes assessed as ‘NE’ or ‘No effect’ have evidence from a high quality study with null findings in relation to child developmental outcomes.

It must be emphasised that the ratings reflected the evidence from the balance of studies conducted on each programme and that the most rigorous evidence was always privileged in the rating system. It must also be emphasised in line with the discussion of Section 1 that the intention was not to determine a set of proven programmes that should then be rolled out nationally, but to provide local decision makers with findings about a range of programmes that provided detailed information about how much evidence was available for each, together with guidance on how to interpret it in the light of local conditions, assets and needs.

It is an important feature of this approach that it assesses the evidence that a given programme has achieved in empirical studies a hypothesised set of impacts, rather than asking whether the programme is based on activities that have a sound scientific basis, or that have been shown to have impact when delivered in other forms by different organisations. We address the question of whether specific programmes have demonstrated effectiveness, not whether they are rooted in evidence, and not the general science of what might work without regard to implementation and delivery. We also took the approach of seeking to assess in a formative as well as summative sense, providing information on how programmes might develop their evidence as well as on what is currently well evidenced. This is important in a field for which much activity is voluntary or developed locally and participants may be hostile or unsupportive to the parachuting in of well-established programmes with higher levels of evidence.

The detailed findings on individual programmes are best reviewed on the EIF website where the information can be kept up to date as the evidence changes and programmes change. We focus here on the general findings of the review considered in terms of aggregate findings across the 75 programmes of this type. Briefly, we found that there were a good number of well-evidenced programmes that if carefully commissioned are likely to be effective. We also found many promising approaches that are based on firm scientific principles but have not yet been proven. The evidence from this review was strongest for programmes that target based on early signals of risk in child development, although universal programmes or programmes that target on the basis of demographic factors can in principle be effective. Evidence was also strongest for programmes that aimed to promote early behavioural self-regulation rather than attachment or cognitive development. This does not necessarily mean that attachment or cognitive development programmes are ineffective; more and better evaluation is required.

Table 2. Evidence-based parenting support programmes (as of January 2016)

Programme Name	Primary outcome domain
1. Child First Programme	Attachment
2. Child-Parent Psychotherapy	
3. Family Foundations	
4. Family Nurse Partnership	
5. Infant-Parent Psychotherapy	
1. Empowering Parenting and Empowering Communities (EPEC)	Behaviour
2. Family Check-Up (FCU)	
3. Helping the Noncompliant Child	
4. Hitkashrut	
5. Incredible Years Preschool	
6. ParentCorps	
7. The New Forest Parenting Programme	
8. Triple P Discussion Groups	
9. Triple P Group	
10. Triple P Standard	
1. Let's Play in Tandem	Cognitive
2. Raising Early Achievement in Literacy	

Nonetheless, evidence-based approaches to improving behaviour are better prepared for replication and taking to scale.

A total of 17 programmes were assessed with an evidence rating of '3' or '4'. We refer to these programmes as 'evidence-based', which for the purposes of this paper is a shorthand label for having demonstrated statistically significant effectiveness in at least one rigorously controlled study.¹³ The 17 evidence-based programmes are listed in table 2 above together with their primary outcome domain.

The review also assessed the implementation costs for these programmes, providing a rating which indicated a broad estimate of implementation cost per child treated. This cost assessment is not the same as the actual market price that a commissioner would pay for an intervention (although both may be correlated, and the latter may influence the former). The actual market price will in practice need to be negotiated between provider and commissioner, taking into account commercial policies, trading conditions and local resources and circumstances. Typically, therefore, information on market prices or unit costs is either unavailable or commercially sensitive, so we did not have access to it in a robust, consistent and comparable way. We therefore developed a new approach that would provide us with comparable data for each programme. We surveyed programme providers and

developers to obtain information on their programme's implementation inputs and activities, including costs but also other features of delivery:

- Training fees
- Training time for each practitioner
- Whether booster training is required
- Costs of programme material (initial and ongoing)
- Programme delivery hours for each practitioner involved
- Qualification level of each practitioner involved
- Whether internal and external supervision are required
- Qualifications of internal and external supervisors (if applicable)
- Whether a licence is required
- Typical size of intervention group.

Using statistical and econometric methods, we then weighted and combined this information into a single scale (see Annex 3 of Asmussen *et al.*, 2016 for further detail). We then discretised the scale into the following estimated unit cost bins:

- Low (< £100 per child)
- Medium-Low (£100–£499 per child)
- Medium (£500–£999 per child)
- Medium-High (£1,000–£1,999 per child)
- High (≥ £2,000 per child).

These cost ratings permit one programme to be compared to another in broad terms on the basis of the total amount of resources that one would normally expect are required to deliver it, normalised according to the number of children that would receive the treatment simultaneously. It should be noted that our methodology is an attempt to arrive at a high-level cost for a number of programmes, in a relatively scalable and parsimonious manner. Other approaches exist which produce a more detailed estimate for an individual programme or a small group of related programmes, e.g. Charles *et al.* (2013).

The distribution of implementation cost ratings for the evidence-based programmes is shown in figure 1. By and large these are inexpensive programmes: we estimate that many could be delivered at a cost of less than £500 per child. However, this set also includes two attachment programmes with a significantly higher estimated implementation cost of over £2,000 per child.

For each of the review's evidence-based programmes, data were also collected on scale of impact and effect size from its most rigorous studies. Across the 17 programmes above, we identified 53 such studies containing 762

Figure 1. Distribution of estimated programme implementation costs (per child)

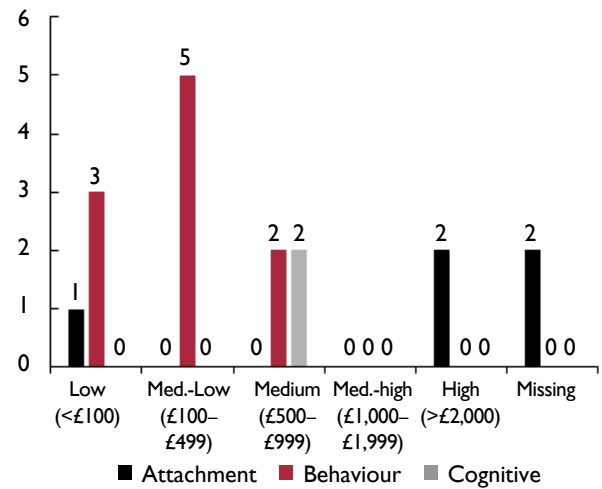


Table 3. Effect size in evidence-based programmes

Primary outcome	N	Mean	Min.	Max.	Std. Dev.
Attachment	8	0.32	0.20	0.63	0.14
Behaviour	67	0.70	0.21	2.20	0.38
Cognitive	7	0.43	0.18	0.82	0.22
Grand Total	82	0.64	0.18	2.20	0.38

Notes: See Appendix table A2 for further details on specific programmes.

effects, of which 304 related to relevant child outcomes – attachment, behaviour, or cognitive development.¹⁴ Of these, 155 were genuine effect size parameters; in the other cases, an effect size was not actually reported in the original study (the most common alternative was a p-value). We focus our attention on the cases where an actual Cohen's D or equivalent parameter was reported, and in particular the 82 effect sizes that were statistically significant. These are shown in table 3. Effects are coded such that positive numbers represent improvements in a child outcome, and higher numbers represent larger improvements. For full details on the effect sizes for each of the 17 programmes, see table A2 in the Appendix.

Overall, effect sizes appear to be larger for the evidence-based programmes that target child behaviour, and smaller for evidence-based programmes that target attachment – although the differences would not be statistically significant.

Cost-benefit analysis of early intervention programmes requires knowledge of, or assumptions about, potential

Table 4. Long-term effect sizes for evidence-based programmes

	N	Mean	Min.	Max.	Std. Dev.
Attachment	8	0.32	0.20	0.63	0.14
Post-test	6	0.32	0.20	0.63	0.16
Follow-up (≤ 3 yrs)	2	0.33	0.25	0.41	0.11
Behaviour	67	0.70	0.21	2.20	0.38
Post-test	44	0.80	0.24	2.20	0.41
Follow-up (≤ 3 yrs)	14	0.48	0.21	0.87	0.23
Follow-up (3 yrs +)	9	0.56	0.25	0.91	0.25
Cognitive	7	0.43	0.18	0.82	0.22
Post-test	5	0.35	0.18	0.61	0.17
Follow-up (≤ 3 yrs)	1	0.82	0.82	0.82	N/A
Follow-up (3 yrs +)	1	0.42	0.42	0.42	N/A
Grand Total	82	0.64	0.18	2.20	0.38

long-term effects, in order to make statements about economic and fiscal implications. Table 4 investigates this by breaking down the 82 effect sizes according to the point at which effect sizes were measured. As can be seen, the majority of these effects were measured at post-test, i.e. at the end of the intervention, with only a small number of long-term follow-up measures of impact. This starkly illustrates a key challenge for carrying out cost-benefit analysis of early years parenting programmes – despite our focus on the most rigorous available evidence. The outcomes of these programmes are not generally measured over sufficient duration to permit direct, non-speculative cost-benefit analysis. Furthermore, in the case of behaviour programmes, there appears to be a downward trend whereby effect sizes measured in long-term follow-ups are smaller than effect sizes measured at post-test.¹⁵

When only short-term and immediate shifts in skills, development or human capital are observed, auxiliary methods linking aspects of early and later development are required in order to make statements about potential longer-term outcomes. As highlighted in Section 2, we note that other evidence bodies engaged in the practice of cost-benefit analysis have combined data on programme effects with other research and data on longitudinal trajectories between child outcomes and adult outcomes. There is a rich literature on quantitative approaches for estimating the link between early and later measures of skill or development. These vary from structural models of the dynamics of skills and human capital (Cunha and Heckman, 2008; Cunha *et al.*, 2010) and of the effects of early-life shocks (Almond *et al.*, 2017), to more reduced-form approaches involving regressions of adolescent or adult outcomes on early childhood conditions (Smith *et al.*, 2010; Goodman *et al.*, 2011, 2015).

Of course, there are a number of challenges that make this difficult to execute in ways that are not open to criticism. First, programme effects may dissipate or fade out over time – see table 4 – in ways that can be difficult to pin down without direct observation. Hence, forward projections based on short-term or immediate effect sizes risk overestimating the actual long-term effect. Second, even though similar constructs may be measured in evaluation trials and longitudinal cohort studies such that they can be linked, they may be structurally different in important ways. Programme effects from an RCT are not the same as naturally occurring cross-sectional differences in outcomes. Individual heterogeneity, both observed and unobserved, may be responsible for much of the latter; this may mean that long-term benefits may not materialise in the manner hypothesised through the application of *ex ante* arithmetic. As an example of this, we note evidence that programmes might improve parenting without a corresponding longer-term improvement in child outcomes (e.g. Kalinauskiene *et al.*, 2009), even though the correlation between parenting and child outcomes is well-established in observational studies (Utting, 2007; Dearden *et al.*, 2011; Bono *et al.*, 2016).

Given these considerable modelling challenges, we are not able to – and do not attempt to – make cost-benefit statements in relation to the programmes considered here. In the next section, we consider evidence from the other end of the spectrum on known or accepted costs of adolescent and adult outcomes but we do not attempt to link the two formally.

4. Evidence on financial and fiscal costs

An increasingly popular way of evaluating the financial case for early intervention is to explore the financial costs of failing to intervene. For example, in the case of behaviour programmes, the costs of conduct disorder provide information and evidence on the case for intervening earlier. Important work has been done in this area by Scott *et al.* (2001), Friedli and Parsonage (2007), Sainsbury Centre for Mental Health (2009), Bonin *et al.* (2011), and most recently NICE (2013). However there is considerable variation among these sources in the types of cost, types of benefit, types of analysis and time periods considered. Our work relates more directly to, and builds on, existing evidence on the unit costs of public services relevant to targeted interventions, most notably the Unit Cost Database (produced by New Economy)¹⁶ and the Unit Costs of Health and Social Care (produced by the Personal Social Services Research Unit).¹⁷

EIF has recently added to this evidence with estimates of the annual cost of ‘late intervention’. We define ‘late

Table 5. Components of the fiscal cost of late intervention

Issue	Information used to calculate estimated fiscal cost
Crime and anti-social behaviour	<ul style="list-style-type: none"> • Reported cases of domestic violence and abuse • Reported anti-social behaviour incidents • Young people in the Youth Justice System (YJS)
School absence and exclusion	<ul style="list-style-type: none"> • Number of persistent absentees • Number of permanent school exclusions • Annual spending on Pupil Referral Units
Children's social care	<ul style="list-style-type: none"> • Annual spending on Looked After Children • Number of Child Protection Plans • Number of Children in Need
Child injuries and mental health problems	<ul style="list-style-type: none"> • Children admitted to hospital due to injuries • Children admitted to hospital due to mental health • Children admitted to hospital due to self-harm
Youth substance misuse	<ul style="list-style-type: none"> • Young people admitted to hospital due to substance misuse • Children using specialist substance misuse treatment services • Children admitted to hospital due to alcohol
Youth economic inactivity	<ul style="list-style-type: none"> 16-17 year-olds who are NEET 18-24 year-olds who are NEET

Source: Chowdry and Fitzsimons (2016a).

intervention' as the acute or statutory services and benefits – such as social care, hospital, police and criminal justice services – that may be required if problems are not addressed. These services, while necessary, are costly to individuals, society and state, and tend to be aimed more at management or containment of problems rather than remedying them. The associated costs and harms may, therefore, extend into the future as problems occur repeatedly, although that is not the focus of our estimates. A fuller description of what is captured and included within the EIF estimate of the fiscal costs of late intervention is shown in table 5.

An important feature of this estimated cost of late intervention is that it is a current, short-term annual, fiscal cost – the fiscal cost this year – rather than a projected cost cumulated over years or decades. Also, as it relates only to known costs to taxpayers, it is not an estimate of social costs (e.g. costs to others from the experience of crime) or economic costs (such as in lost output or lost productivity). Although limited in this way, expressing the cost of late intervention as a short-run fiscal cost offers a number of benefits. First, it is more comparable to the current costs of investing in early intervention and prevention. Second, it is subject to less uncertainty than estimated lifetime costs or costs over a generation, important though those may be (e.g. Action for Children,

2013). Third, our focus on the current annual cost of late intervention sits more easily within budgetary or political cycles relevant to decision-makers – unlike cumulative estimates over a longer period. Focussing on *current annual* government spending on late intervention for children and young people – while they are still children and young people – enables a clearer discussion of the size and scope of resources currently expended which might become available to use in other ways if early intervention is employed in effective and appropriate ways. It also focusses attention on where the costs of late intervention currently fall between different parts of government, in ways that provide greater impetus for policy response.

The EIF estimate of late intervention complements other estimates of the annual costs of specific problems: youth crime and youth unemployment have both been estimated to cost over £1 billion a year (The Prince's Trust, 2010), and the cost of dealing with child behavioural disorders is estimated at £1.6 billion a year.¹⁸ In relation to healthcare, National Audit Office (2013) estimated that the NHS spent nearly £10 billion in 2011–12 on the costs of obesity, alcohol misuse and smoking-related illness. In comparison to these estimates, the EIF late intervention estimate is a more 'global' summary of costs in that it aggregates across all key issues rather than focussing on one. The EIF estimate is also a 'bottom-up' estimate, rooted in actual data on children and young people and the services they use, within each local area and for the country as a whole.

Our general approach for arriving at the annual fiscal cost of late intervention is to take total demand or caseloads for each form of intervention shown in table 5 – obtained from published statistics – and combine that with an estimated 'unit cost' of each case that provision. This has the advantage of being directly linked to what we know about demand for late intervention services for children and young people. We have made use of existing unit cost data¹⁹ that are widely accepted, including by HM Treasury, as a basis for financial analysis and cost-benefit analysis. Information on unit costs tends to be available at a national level only, even though the true unit cost may vary from one local area to another. Therefore, where it would lead to more robust results, we have also used published data on actual local authority spending on particular acute services.

In summarising this work we are mindful of its key limitations. First, judgements have been made about which items to include in the analysis. Table 5 is not a list of every activity that counts as late intervention, and the items in the table do not add up to the true

Table 6. Late intervention costs in England and Wales

Cost item	Caseload	Estimated unit cost	Total cost (£m)
Reported cases of domestic violence and abuse	943,628		5,230
Reported anti-social behaviour incidents	1,925,952	£364	701
Young people in the Youth Justice System (YJS)	37,946	£9,031	342
Number of persistent absentees	256,632	£1,886	484
Permanent school exclusions	5,029	£2,545	442
Annual spending on Looked After Children	75,150	£70,645	5,309
Number of Child Protection Plans	52,624	£5,583	294
Number of Children in Need	355,328	£1,627	578
Children admitted to hospital due to injuries	114,475	£1,319	158
Children admitted to A&E due to injuries	1,982,660	£80	151
Children admitted to hospital due to mental health	10,240	£42,236	433
Children admitted to hospital due to self-harm	14,386	£2,241	32
Young people admitted to hospital due to substance misuse	19,130	£404	8
Children using specialist substance misuse treatment services	24,159	£17,007	411
Children admitted to hospital due to alcohol	13,736	£1,770	24
16–17 year olds who are NEET	42,940	£630	27
18–24 year olds who are NEET	752,889	£3,507	2,640

Source: Figures taken from Chowdry and Fitzsimons (2016b).

Note: All amounts in 2016/17 prices.

amount of total spending on late intervention. The table is better interpreted as setting out some key types of late intervention for which national and local data on caseloads are available, along with information on total spend or unit costs. For this reason our estimate is necessarily an underestimate of the true current annual fiscal cost of late intervention. Second, late intervention relates to *services* not *wellbeing*. The data we have used will reflect reporting and recording of incidents, local and national decisions about the availability, referral and resourcing of acute services, as well as the broader fiscal climate. While such data are useful for understanding the fiscal implications of policy decisions, and for informing discussions about spending, these data do not capture the underlying *outcomes* and *wellbeing* of children and families.

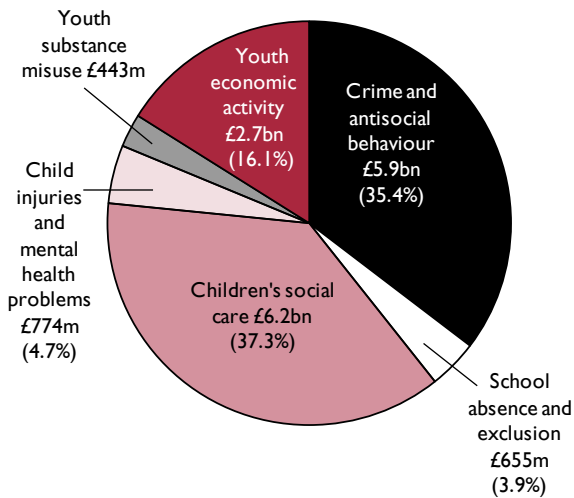
We estimate that, annually, £17 billion per year is spent by the public sector in England and Wales on the costs of providing late intervention for children and young people (Chowdry and Fitzsimons, 2016a). Further details behind the total overall cost are shown in table 6. The single largest components of this are the fiscal costs of responding to incidents of domestic violence and abuse, and spending on looked after children.

Figure 2 presents a visual summary in terms of the social issues that drive costs. Children's social care – including spending on children in need, child protection and looked after children – represents the issue that constitutes the largest portion of the £17 billion total. The second largest issue relates to crime and anti-social behaviour.

The data can also be considered in terms of who pays, as shown in figure 3. Nearly 40 per cent of the cost of late intervention, as currently defined, is borne by local government. Importantly, however, the majority is borne elsewhere. Given that most early intervention programmes and provision are currently funded and commissioned by local authority services, this sheds light on another major economic and financial challenge: even if benefits from early intervention do materialise, they are likely to accrue to other organisations and agencies – for example, police and health services, or the national exchequer. This externality reduces the incentives for a single agency to invest in early intervention and prevention. However, this reflects institutional shortcomings and failures of governance rather than of early intervention itself. We believe that analysis such as figure 3 has the potential to drive the practice and improvement of joint investment and commissioning arrangements, ensuring that investment and benefit are distributed across all relevant public bodies.

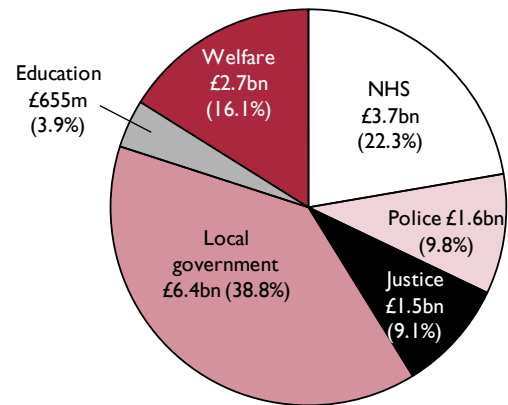
Finally, because the cost of late intervention is a bottom-up estimate computed by aggregating local data on late intervention caseloads and spending, late intervention costs can also be computed for any local authority in England. The analyses shown in table 6, figure 2 and figure 3 can all be replicated for any English local authority. For reference local authority levels of late intervention cost, normalised on a per-capita basis, range from a minimum of £164 to a maximum of £531. The 25th percentile is £248; the 75th percentile £337.

Figure 2. Fiscal cost of late intervention by social issue



Source: Chowdry and Fitzsimons (2016a).

Figure 3. Fiscal cost of late intervention by budget area



Source: Chowdry and Fitzsimons (2016a).

We stop short of implying that the fiscal amounts presented above represent achievable fiscal savings as a result of early intervention. First, some amount of late intervention will be inevitable and unavoidable – the cost cannot be brought down to zero, and we do not have evidence to prove that it is far from a natural limit. Second, unit costs are usually defined as the ratio between total budgetary spend and total caseloads, hence they may not be equal to the marginal saving that would result from incremental reductions in demand for such services. Third, in reality, the services mentioned in table 4 may be *de facto* rationed, in the sense that there may be thresholds or waiting lists. Even if early intervention programmes successfully improve outcomes, the result may be simply freeing up of capacity within these services and the reallocation of places, rather than a financial saving from reduced total service use. Even if total service use can be reduced, these services may be governed by fixed budgets in the short run. If so, service expenditure will not be directly linearly related to caseloads and savings will not be proportional to the change in caseloads. To be clear, this is a limitation of the structure and organisation of services, not of early intervention. Nevertheless, this caveat should be borne in mind when assessing the capability of early intervention to deliver financial savings, and so we are careful not to treat the costs of late intervention interchangeably as the benefits of early intervention.

While our estimates are not an attempt at cost-benefit analysis, they do provide important data and evidence

which could support it. Further analysis to get closer to cost-benefit estimates, building on this work, would have to model the relationship between the presence or prevalence of child problems and the resulting likelihood of use of, or volume of demand for the late intervention services considered here. Specific cost items, such as anti-social behaviour, youth justice, persistent absenteeism and school exclusions, may be the most fertile ground for savings as a result of improvements to child behaviour, for example.

5. Conclusion and discussion

We summarise our headline findings and messages in the following key points:

1. The scientific basis for early intervention is strong. The importance of promoting improved early childhood environments and child development is strongly supported by research across a range of disciplines. The rationale for policies that take a preventative approach is also sound.
2. Early intervention is a broad and diverse concept. It can vary by target group, type of risk, level of need, type of intervention and type of outcome. The Early Intervention Foundation has conducted substantial reviews of precisely defined types of early intervention; specific provision to achieve specific outcomes for specific target groups. As an example, we focus in this paper on programmes that aim to promote early child development via improved parent-child interactions (up to a child age of 5 years).

3. Cost-benefit analysis for commissioners should reflect what is available, not just what has the best evidence. We reviewed 75 such programmes available in the UK, and found that 1 in 4 have shown a statistically significant improvement in child attachment, behaviour or cognitive development in a well-conducted and rigorous trial with a good counterfactual. Most of these programmes can be delivered at a low or moderate cost per child. Other programmes have less rigorous evidence, but may still be of interest to commissioners.

4. Establishing that a programme is effective is important, but not sufficient, for cost-benefit analysis. Even among programmes with the most rigorous evidence, data on scale and longevity of impact are relatively weak. We find a plethora of different outcome measures, limited reporting of effect sizes, and a lack of long-term follow-up.

5. Future research should focus on the future economic and fiscal value that results from initial programme effectiveness. We support this approach by presenting data on the immediate fiscal costs of late intervention for children and young people. Further evaluation of programmes is required to improve what is known about their impacts on long-run outcomes in adolescence and adulthood, and on demand for late intervention services.

6. Cost-benefit analysis should be applicable to the contexts in which commissioners operate. Commissioners seek to understand how evaluation findings will translate into benefits for their local population and their local system (including workforces, referral pathways and other provision). Commissioners also seek to understand how benefits are distributed across agencies, as well as the extent to which these benefits might lead to tangible savings.

Discussion

While it is important for commissioners and policy-makers to have clear information on effectiveness and value for money to guide decision-making, there are important aspects of the reality of implementation of early intervention that should not be ignored when decisions are made. For these reasons the Early Intervention Foundation in its early work has sought to map out the evidence for early intervention programmes recognising the diversity of early intervention activities, the importance of innovation, and the crucial role of implementation. There is a danger that the recognition of wider complexity and uncertainty falsely undermines confidence in the well-established science of early intervention and of evaluation. However, we think the rewards of addressing gaps in evidence makes this risk worthwhile.

The work summarised here assessed programmes that aimed to improve child development by supporting the quality of parent-child interactions in early childhood. We have established that there are programmes available in the UK which have demonstrated the capability to positively shift child outcomes in rigorously controlled environments. There are also many other approaches, not mentioned in this paper, that may be promising but do not yet have rigorous evidence of impact. We find that programmes that focus on behaviour tend to have the strongest evidence of effectiveness. In our view, further innovation, testing and learning is required. Attachment represents another important area of innovation that has much potential, but where few programmes have yet demonstrated that the underpinning science can be translated into effective practice. Importantly, there is a paucity of evidence on long-run effectiveness several years from the point at which interventions are delivered.

We note that the evidence we summarise here is largely academic in nature. More work is required to understand how to translate this evidence and best apply it in the context of a particular population and system. The effectiveness of programmes in the real world will depend on features of the broader system, including implementation, workforces, assessments, referrals and monitoring. This requires an integration of implementation science with economic thinking and other sciences, and new ways of understanding the messy frontier between practice, policy and evidence.

We hope that some gaps in knowledge will be addressed over time by improvements to the evidence base, with more studies over longer duration that make a clearer connection between ‘what works’ and ‘for whom’. This requires more evaluation, not just with the necessary standards of rigour for causality to be ascertained, but also with the types of data (e.g. linked administrative records) that facilitate long-run tracking of economic and fiscal outcomes in an affordable manner. While we have emphasised the importance of linking to fiscal costs for any discussion of benefits, it is important that estimates of potential benefits of early intervention are aware of the challenges involved in quantifying and ultimately achieving cashable fiscal benefits – tangible, in-year financial savings. Too often it is assumed that if children benefit then savings will result; much more sophisticated analysis is required.

It is also important to recognise that a diversity of outcomes generally implies a diversity of beneficiaries. We present here an attempt to move towards bridging the gap, showing the degree to which the costs of late

intervention fall across different public agencies. In so doing, we hope to offer further stimulus to those agencies to come together in developing approaches that better match services to populations and reduce the collective costs of failure. Devolution is one such opportunity, but despite some promising pilots, as in Greater Manchester, there is not substantive evidence yet that the opportunity to reduce the mismatch of agencies, budgets and populations has been grasped.

Returning to the title of this paper, we have shown that the current state of the evidence in relation to these programmes is not yet mature and established enough to enable rigorous, place-specific cost-benefit analysis on a wide scale. Even the ‘gold-star’ evidence containing rigorous causal estimates of effectiveness, while a useful and important point of departure, provides an insufficient basis for estimation and modelling. In order to make a firm and credible connection to either of these, much more evaluation, modelling and analysis is required; we hope this stimulates further work by government, researchers and the EIF over the coming years.

NOTES

- 1 www.eif.org.uk/wp-content/uploads/2015/08/Annex-The-evidence-on-the-benefits-of-El-web.pptm.
- 2 <https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit>.
- 3 <http://www.whatworksgrowth.org/>.
- 4 <http://whatworks.college.police.uk/Pages/default.aspx>.
- 5 We describe below in section 4 the costs involved in delivery of these forms of late intervention.
- 6 Fixsen *et al.*, 2009.
- 7 <http://www.wsipp.wa.gov/>.
- 8 See <http://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf> for more details on the WSIPP cost-benefit analysis methodology.
- 9 Dartington Social Research Unit (DSRU) has adapted this evidence and analysis for the UK context through its Investing in Children project.
- 10 See Asmussen *et al.* (*op. cit.*) for more information about this review and its findings.
- 11 <http://www.eif.org.uk/publication/foundations-for-life-what-works-to-support-parent-child-interaction-in-the-early-years/>
- 12 <http://www.eif.org.uk/eif-evidence-standards/>.
- 13 It is also required that a programme does not have a null finding – no effect or a negative effect – in another study of equal or greater rigour (such that estimates of scale of impact have internal validity).
- 14 The other 458 effects related to parent outcomes or features of the family environment. For further information on the specific child measures in our data, see table A1 in the Appendix.
- 15 We acknowledge, however, that the use of potentially different measures at different time points could confound this comparison.
- 16 <http://www.neweconomymanchester.com/our-work/research-evaluation-cost-benefit-analysis/cost-benefit-analysis/unit-cost-database>.

17 <http://www.pssru.ac.uk/project-pages/unit-costs/>.

18 Department of Health (2013) *op. cit.*

19 We particularly wish to acknowledge use of the New Economy Unit Cost Database (<http://www.neweconomymanchester.com/our-work/research-evaluation-cost-benefit-analysis/cost-benefit-analysis/unit-cost-database>).

REFERENCES

- Action for Children (2013), *The Red Book 2013: Children under Pressure*.
- Allen, G. (2011), *Early Intervention: The Next Steps, An Independent Report to Her Majesty's Government by Graham Allen MP*, The Stationery Office.
- Almond, D. and Currie, J. (2011), ‘Human capital before age five’, in Ashenfelter, O. and Card, D. (eds), *The Handbook of Labor Economics*, Vol. 4, Chapter 15, Amsterdam: Elsevier, pp. 1315–486.
- Almond, D., Currie, J. and Duque, V. (2017), ‘Childhood circumstances and adult outcomes: act II’, NBER Working Paper 23017.
- Asmussen, K., Feinstein, L., Martin, J. and Chowdry, H. (2016), *Foundations for Life: What Works to Support Parent Child Interaction in the Early Years*, London: Early Intervention Foundation.
- Barnes, J., Ball, M., Meadows, P., Belsky, J. (2009), *Nurse-Family Partnership Programme. Second Year Pilot Sites Implementation in England. The Infancy Period*, Research Report DCSF-RR166.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T. and Moore, L. (2012), ‘Realist randomised controlled trials: a new approach to evaluating complex public health interventions’, *Social Science and Medicine*, 75(12), pp. 2299–306.
- Bonin, E.M., Stevens, M., Beecham, J., Byford, S. and Parsonage, M. (2011), ‘Parenting interventions for the prevention of persistent conduct disorders’, in Knapp, M.J., McDaid, D. and Parsonage, M. (eds), *Mental Health Promotion and Mental Illness Prevention: The Economic Case*, 6 – 8, London: Department of Health.
- Bono, E.D., Francesconi, M., Kelly, Y. and Sacker, A. (2016), ‘Early maternal time investment and early child outcomes’, *Economic Journal*, 126, F96–F135.
- Brooks-Gunn, J. (2003), ‘Do you believe in magic? What we can expect from early childhood intervention programs’, *Social Policy Report*, 17(10), pp. 3–14.
- Campbell, F.A., Pungello, E.P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B.H., Barbarin, O.A., Sparling, J.J. and Ramey, C.T. (2012), ‘Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up’, *Developmental Psychology*, 48(4), pp.1033–43.
- Cartwright, N. (2011), ‘A philosopher’s view of the long road from RCTs to effectiveness’, *The Lancet*, 377(9775), pp. 1400–1.
- Charles, J. M., Edwards, R.T., Bywater, T. and Hutchings, J. (2013), ‘Micro-costing in public health economics: steps towards a standardized framework, using the Incredible Years Toddler Parenting Program as a worked example’, *Prevention Science*, 14(4), pp. 377–89.
- Chowdry, H. and Fitzsimons, P. (2016a), *The Cost of Late Intervention: EIF analysis 2016*, London: Early Intervention Foundation.
- (2016b), *The Cost of Late Intervention: EIF analysis 2016 – Technical Report*, London: Early Intervention Foundation.
- Cicchetti, D. and Rogosch, F. (1996), ‘Equifinality and multifinality in developmental psychopathology’, *Development and Psychopathology*, 8(4), pp. 597–600.
- Cunha, F. and Heckman, J.J. (2008), ‘Formulating, identifying and estimating the technology of cognitive and noncognitive skill

- formation', *Journal of Human Resources*, Fall, 43(4), pp. 738–82.
- Cunha, F., Heckman, J.J. and Schennach, S.M. (2010), 'Estimating the technology of cognitive and noncognitive skill formation', *Econometrica*, 78, pp. 883–931.
- Dearden, L., Sibietta, L. and Sylva, K. (2011), 'The socio-economic gradient in early child outcomes: evidence from the Millennium Cohort Study', *Longitudinal and Life Course Studies*, 2(1), pp. 19–40.
- Department of Health (2013), *Chief Medical Officer's Annual Report 2012: Our Children Deserve Better: Prevention Pays*, 24 October.
- Durlak, J.A. and Wells, A.M. (1998), 'Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents', *American Journal of Community Psychology*, 26, pp. 775–802.
- Eckenrode, J., Campa, M., Luckey, D., Henderson, C., Cole, R., Kitzman, H., Anson, E., Sidora-Arcoleo, K., Powers, J. and Olds, D. (2010), 'Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial', *Archives of Pediatric and Adolescent Medicine*, 164, pp. 9–15.
- Fixsen, D., Blase, K., Naoom, S., Wallace, F. (2009), 'Core implementation components', *Research on Social Work Practice*, 19, pp. 531.
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., Mościcki, E.K., Schinke, S., Valentine, J.C. and Ji, P. (2005), 'Standards of evidence: criteria for efficacy, effectiveness and dissemination', *Prevention Science*, 6(3), pp. 151–75.
- Friedli, L. and Parsonage, M. (2007), *Mental Health Promotion: Building an Economic Case*, Northern Ireland Association for Mental Health.
- Goodman, A., Joshi, H., Nasim, B. and Tyler, C. (2015), *Social and Emotional Skills in Childhood and their Long-Term Effects on Adult Life*, London: Early Intervention Foundation.
- Goodman, A., Joyce, R. and Smith, J.P. (2011), 'The long shadow cast by childhood physical and mental problems on adult life', *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), pp. 6032–7.
- Heckman, J.J. (2008), 'Schools, skills and synapses', *Economic Inquiry*, 46(3), pp. 289–324.
- Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A. and Yavitz, A. (2010), 'The rate of return to the HighScope Perry Preschool Program', *Journal of Public Economics*, 94(1), pp. 114–28.
- Hiscock, H., Bayer, J.K., Price, A., Ukoumunne, O.C., Rogers, S. and Wake, M. (2008), 'Universal parenting programme to prevent early childhood behavioural problems: cluster randomised trial', *BMJ: British Medical Journal*, 336(7639), pp. 318–21.
- HM Treasury (2011), *The Green Book: Appraisal and Evaluation in Central Government*, London: The Stationery Office.
- Kalinauskienė, L., Cekuoliene, D., VanIjzendoorn, M.H., Bakermans-Kranenburg, M.J., Juffer, F., and Kusakovskaja, I. (2009), 'Supporting insensitive mothers: the Vilnius randomized control trial of video-feedback intervention to promote maternal sensitivity and infant attachment security', *Child: Care, Health and Development*, 35, pp. 613–23.
- Kitzman, H., Olds, D.L., Henderson, C.R., Hanks, C., Cole, R., Tatelbaum, R., McConnochie, K.M., Sidora, K., Luckey, D.W., Shaver, D., Englehardt, K., James, D. and Barnard, K. (1997), 'Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing', *Journal of the American Medical Association*, 278(8), pp. 644–52.
- Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G. and Pawson, R. (2013), 'Realist RCTs of complex interventions – an oxymoron', *Social Science and Medicine*, 94, pp. 124–8.
- National Audit Office (2013), *Early Action: Landscape Review*, London: The Stationery Office.
- NICE (2013), *Antisocial Behaviour and Conduct Disorders in Children and Young People – Costing Report: Implementing NICE Guidance*, National Institute for Health and Clinical Excellence.
- Pawson, R. (2013), *The Science of Evaluation: A Realist Manifesto*, London: Sage.
- Phillips, D.A. and Shonkoff, J.P. (eds). (2000), *From Neurons to Neighborhoods: The Science of Early Childhood Development*, National Academies Press.
- Piaget, J. (2013), *The Construction of Reality in the Child (Vol. 82)*, Abingdon: Routledge.
- Robling, M., Bekkers, M., Bell, K., Butler, C., Cannings-John, R., Channon, S., Martin, B., Gregory, J., Hood, K., Kemp, A., Kenkre, J., Montgomery, A., Moody, G., Owen-Jones, E., Pickett, K., Richardson, G., Roberts, Z., Ronaldson, S., Sanders, S., Stamuli, E. and Torgerson, D. (2016), 'Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial', *The Lancet*, 387(10014), pp. 146–55.
- Rush, J. (2009), 'The role of efficacy and effectiveness trials', *World Psychiatry*, February, 8(1), pp. 34–5.
- Sainsbury Centre for Mental Health (2009), *The Chance of a Lifetime: Preventing Early Conduct Problems and Reducing Crime*, London: Sainsbury Centre for Mental Health.
- Scott, S., Knapp, M., Henderson, J. and Maughan, B. (2001), 'Financial cost of social exclusion: follow up study of antisocial children into adulthood', *British Medical Journal*, 323, pp. 1–5.
- Sherman, L.W., Gottfredson, D.C., MacKenzie, D.L., Eck, J., Reuter, P. and Bushway, S.D. (1998), *National Institute of Justice: Research in Brief, July 1998*, U.S. Department of Justice.
- Shonkoff, J.P., Boyce, W.T. and McEwen, B.S. (2009), 'Neuroscience, molecular biology, and the childhood roots of health disparities: building a new framework for health promotion and disease prevention', *JAMA*, 301(21), pp. 2252–9.
- Shonkoff, J.P., Garner, A.S., Siegel, B.S., Dobbins, M.I., Earls, M.F., McGuinn, L. and Committee on Early Childhood, Adoption, and Dependent Care (2012), 'The lifelong effects of early childhood adversity and toxic stress', *Pediatrics*, 129(1), e232–e246.
- Smith, J.P., Monica, S. and Smith, G.C. (2010), 'Long-term economic costs of psychological problems during childhood', *Social Science and Medicine (1982)*, 71(1), pp. 110–15.
- The Prince's Trust (2010), *The Cost of Exclusion: Counting the Cost of Youth Disadvantage in the UK*, London: The Prince's Trust.
- Utting, D. (2007), *Parenting and the Different Ways it Can Affect Children's Lives: Research Evidence*, York: Joseph Rowntree Foundation.

APPENDIX

Table A1. Effect sizes by primary outcome and measure

	N	Mean	Min.	Max.	Std. Dev.
Attachment	8	0.32	0.20	0.63	0.14
Attachment Classification	1	0.25	0.25	0.25	N/A
Child Sleep Questionnaire	2	0.25	0.23	0.27	0.03
Family involvement in CPS	1	0.41	0.41	0.41	N/A
Infant Behaviour Questionnaire	2	0.28	0.21	0.35	0.10
Infant Behaviour Questionnaire	1	0.20	0.20	0.20	N/A
Traumatic Stress Disorder	1	0.63	0.63	0.63	N/A
Behaviour	67	0.70	0.21	2.20	0.38
ADHD Rating Scale-IV	6	1.56	1.07	2.20	0.39
ADHD Symptoms	1	0.69	0.69	0.69	N/A
Antisocial Process Screening Device	1	0.70	0.70	0.70	N/A
Arrested	1	0.31	0.31	0.31	N/A
Callous-unemotional traits	1	0.85	0.85	0.85	N/A
Child and Adolescent Psychiatric Assessment	2	0.90	0.89	0.91	0.02
Child Behaviour Checklist	11	0.41	0.21	0.67	0.17
Child Behavioural Questionnaire	1	0.47	0.47	0.47	N/A
Child Behaviours	1	0.46	0.46	0.46	N/A
Concerns about my child measure	1	0.85	0.85	0.85	N/A
Conners Abbreviated Parent-Teacher rating scale	1	0.61	0.61	0.61	N/A
Conners Rating Scale-Revised	6	1.07	0.89	1.24	0.14
Convicted	1	0.48	0.48	0.48	N/A
Eyberg Child Behaviour Inventory	14	0.61	0.23	0.89	0.19
Head Start Competence Scale	1	0.43	0.43	0.43	N/A
Infant-Toddler Social and Emotional Assessment	1	0.85	0.85	0.85	N/A
Internalising disorders	1	0.25	0.25	0.25	N/A
Kendall Self Control rating scale	1	0.38	0.38	0.38	N/A
New York Rating Scale	3	0.55	0.37	0.69	0.16
Observed child negative behaviour	1	0.78	0.78	0.78	N/A
PACS	1	0.43	0.43	0.43	N/A
Parent account of child symptoms	2	0.60	0.31	0.89	0.41
Parent Daily Report	1	0.75	0.75	0.75	N/A
Parent defined problems	1	0.92	0.92	0.92	N/A
Parental Account of Childhood Symptoms	1	0.87	0.87	0.87	N/A
Strengths and Difficulties Questionnaire	3	0.43	0.37	0.51	0.07
Teacher Report Form	1	0.26	0.26	0.26	N/A
Used cigarettes, alcohol or marijuana in past 30 days.	1	0.65	0.65	0.65	N/A
Cognitive	7	0.43	0.18	0.82	0.22
Academic Performance Rating	1	0.25	0.25	0.25	N/A
Infant-Toddler Developmental Assessment	2	0.71	0.61	0.82	0.15
Kaufman Test of Educational Achievement Brief Form	1	0.18	0.18	0.18	N/A
Letter recognition	1	0.30	0.30	0.30	N/A
Sheffield Early Literacy Development Profile	1	0.41	0.41	0.41	N/A
Weschler Objective Reading Dimensions Score	1	0.42	0.42	0.42	N/A
Grand Total	82	0.64	0.18	2.20	0.38

Notes: Effect sizes are limited to statistically significant Cohen's D or equivalent parameters on a measure of child attachment, behaviour or cognitive development.

Table A2. Effect sizes by primary outcome and programme

	N	Mean	Min.	Max.	Std. Dev.
Attachment	8	0.32	0.20	0.63	0.14
Child FIRST	1	0.41	0.41	0.41	N/A
Child Parent Psychotherapy	1	0.63	0.63	0.63	N/A
Family Foundations	5	0.25	0.20	0.35	0.06
Infant-Parent Psychotherapy	1	0.25	0.25	0.25	N/A
Behaviour	67	0.70	0.21	2.20	0.38
Child FIRST	1	0.85	0.85	0.85	N/A
Child Parent Psychotherapy	4	0.49	0.24	0.67	0.20
Empowering Parenting and Empowering Communities (EPEC)	3	0.60	0.37	0.85	0.24
Family Check-Up (FCU)	6	0.24	0.21	0.30	0.03
Family Foundations	3	0.48	0.43	0.55	0.06
Family Nurse Partnership (FNP)	4	0.42	0.25	0.65	0.18
Helping the Noncompliant Child	8	1.24	0.37	2.20	0.57
Hitkashrut	4	0.68	0.47	0.85	0.17
Incredible Years Preschool	21	0.64	0.31	0.92	0.20
The New Forest Parenting Programme	10	0.95	0.43	1.66	0.35
Triple P Discussion Groups	1	0.86	0.86	0.86	N/A
Triple P Group	2	0.59	0.43	0.74	0.22
Cognitive	7	0.43	0.18	0.82	0.22
Child FIRST	2	0.71	0.61	0.82	0.15
Incredible Years Preschool	1	0.42	0.42	0.42	N/A
ParentCorps	2	0.22	0.18	0.25	0.05
REAL	2	0.36	0.30	0.41	0.08
Grand Total	82	0.64	0.18	2.20	0.38

Notes: Effect sizes are limited to statistically significant Cohen's D or equivalent parameters on a measure of child attachment, behaviour or cognitive development.