Does improving parenting practices in childhood lead to happier adults?

A shallow exploration into the cost-effectiveness of early childhood psychosocial stimulation

Joel McGuire, Ben Stewart, Samuel Dupret, Ryan Dwyer, Michael Plant

November, 2024

Happier Lives Institute

Contents

Summary

Early childhood is a critical period for development. But many parents in low and middle income countries (LMICs) are less likely to play or engage with their kids enough, and some use violence to punish their children. These parenting practices can impair cognitive, emotional, and physical growth, leading to lower lifetime wellbeing.

Psychosocial stimulation interventions improve parenting practices by teaching parents to engage with their children in more enriching ways, such as storytelling, reading, and playing, and to avoid maltreatment. These interventions can be delivered by lay health workers in home visits or to groups of caregivers, and they are relatively cheap to provide in LMICs. These interventions have clear short term benefits to development, and potentially much longer effects on wellbeing.

In this **shallow report,** we spent about 90 hours exploring the cost-effectiveness of these interventions on the subjective wellbeing of children in LMICs. To our knowledge, this is the first wellbeing cost-effectiveness analysis of an intervention to improve parenting.

This forms part of our broader work to assess the cost-effectiveness of interventions and charities based on their impact on subjective wellbeing, measured in terms of wellbeing-adjusted life years (WELLBYs). One WELLBY is equivalent to a 1-point increase on a 0-10 wellbeing scale for one person over one year. We focus on subjective wellbeing because we believe it best captures what ultimately matters, wellbeing. By using wellbeing as a common outcome, it allows to make apples-to-apples comparisons between very different interventions.

The best funding opportunity we found is the Reach Up intervention delivered by the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b). This is a home visiting and group parenting programme in which health workers make regular home visits to parents of young children (often ages 0 to 3) to help train parents to engage in stimulating daily activities with their children.

There's considerable evidence of the effect of parenting interventions on short and medium term development outcomes like cognition (20+ RCTs). However, the long-term evidence on wellbeing is relatively weak. We estimate the impact of parenting interventions on long-term wellbeing using two sources of evidence, to which we give equal weight in our analysis (i.e. 50% to each):

- home visitation programmes in LMICs (2 RCTs, n = 426)
- psychosocial stimulation interventions delivered through preschool programmes in high income countries (causal studies = 4, n = 2,502)

Based on this evidence, we **speculatively** estimate that home visiting parenting programmes have an effect of 0.23 standard deviations (SDs)[1] on depression that lasts 32 years. The naive total individual benefit is 14.5 WELLBYs with a spillover effect of 7 WELLBYs for a total effect

[1] See our methods website page for general definitions about what these units are.

of ~22 WELLBYs. However, we discount this by 77% (primarily for replicability concerns) to arrive at **our final effect estimate of 5 WELLBYs.**

We estimate that it costs **$98** per child treated by the icddr,b, leading to a cost-effectiveness of **50 WELLBYs per $1k (WBp1k)**. For context, this is around **7 times** more cost-effective than GiveDirectly[2] for which we estimated to be 7.55 WBp1k (i.e., $132 per WELLBY; McGuire et al., 2022a). We compare this opportunity to other funding opportunities on our website here.

We expect the icddr,b has room to reduce costs if they scale further, which would improve the cost-effectiveness of their programme. The icddr,b has indicated they could absorb millions more in funding to scale the programme across Bangladesh over the next decade.

Funding the Icddr,b's Reach Up programme seems like a promising, albeit speculative funding opportunity for donors interested in high impact opportunities for increasing wellbeing.

Acknowledgements: We would like to recognise in these footnotes the contributions of authors[3] and staff from the charities we have evaluated[4].

0. Outline

In **Section 1** we introduce and motivate the issue psychosocial stimulation is trying to solve, define key terms, explain how these interventions work and why it's plausible they have long term benefits.

In **Section 2** we discuss our non-systematic literature review of psychosocial stimulation interventions' long term wellbeing effects.

In **Section 3** we evaluate the impact of psychosocial interventions, with an emphasis on those delivered to parents, by lay practitioners, in LMICs.

In **Section 4** we consider the promise of some organisations delivering psychosocial stimulation interventions, discuss possible funding opportunities, and perform a back of the envelope calculation for a psychosocial stimulation programme based in Bangladesh.

In **Section 5** we discuss the quality of the evidence we use in our analysis.

In **Section 6** we conclude and provide our overall views on how promising further research is and how cost-effective funding opportunities appear.

---

[2] GiveDirectly is an NGO which provides cash transfers to very poor households. We take cash transfers as a useful benchmark because they are a straightforward, plausibly cost-effective intervention with a solid evidence base. (For more detailed and updated charity comparisons, see our charity evaluations page.)

[3] Joel McGuire contributed to the conceptualization, investigation, analysis, data curation, and writing of the project. Ben Stewart, contributed to the analysis, and writing of the project. Samuel Dupret contributed to the analysis and data curation of this project. Ryan Dwyer and Michael Plant contributed to the supervision, and writing of the project.

[4] We thank Jena Hamadani for information about icddr,b's implementation of Reach Up in Bangladesh.

1. Why intervene in early childhood?

Parents shepherd children through the most sensitive period of life. Protection from harm and access to stimulation at a young age can have long-lasting consequences for the child's wellbeing for two reasons. Compared to adults:

1. Children can experience permanent benefits for longer.
2. A child may experience larger or more persistent benefits from the same intervention because it puts them on a more favourable developmental trajectory.

This latter point stems from the idea that there are potentially sensitive periods in human development, where an event at one time of life exerts a greater influence than it would have had at another time[5] (Bornstein, 1989, p. 179; Chapter 1 by Troller-Renfree & Fox). While applied to development in general, it also applies to mental health (and presumably mental wellbeing; Räikkönen et al., 2012).

Interventions in childhood could represent unique opportunities to improve wellbeing over time. Notably, if early harm is avoided, reparatory interventions in adulthood might not be needed, making these potentially cost-effective opportunities.

We focus on these interventions in LMICs, where a safe and fulfilling childhood is less guaranteed and it's typically much cheaper to intervene.

1.1 Parenting psychosocial interventions

Parent-based psychosocial stimulation interventions, which I refer to in this report simply as "parenting interventions", are educational programmes primarily aimed at encouraging parents to play with their children more and in more enriching ways. In addition to playing with their child, parenting interventions often also attempt to teach appropriate discipline, avoid maltreatment, and improve parental mental health (Jeong et al., 2021). Parenting interventions can also provide basic nutritional advice or teach healthy behaviours like hand washing, and when possible connect families to relevant health or educational services.

But are parenting interventions really needed? I think so for a few reasons. First, in low income countries, parents don't report playing with their children in a very enriching manner. Cuartas et al. (2020), using a dataset covering 62 LMICs (n = 205,150), estimated that 17.4% of mothers (4.8% for fathers) in low income countries engage in a majority (4 or more) of the following six activities with their children (ages 3 to 4) in the last three days:

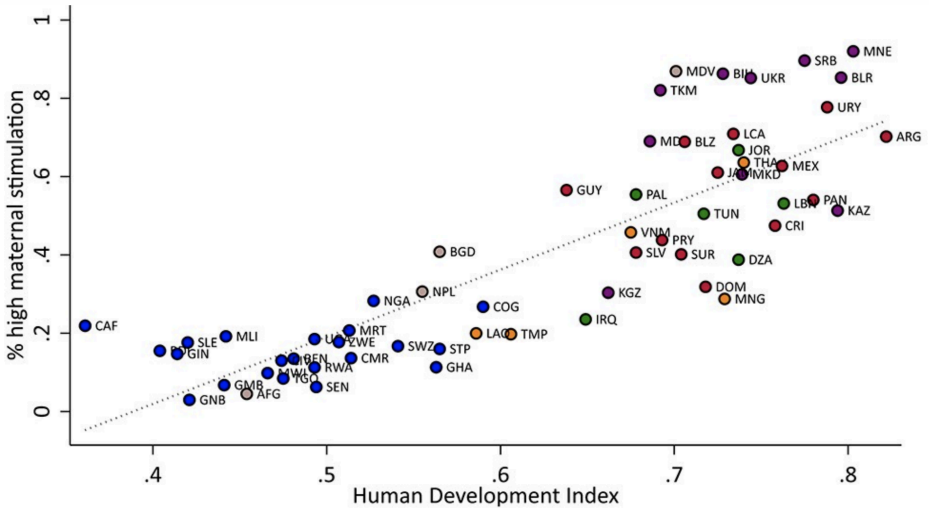
1. reading books or looking at picture books
2. telling stories
3. singing songs or lullabies
4. taking the child outside the home
5. playing with the child

[5] Whilst specific periods for specific phenomena need to be defined, the literature often refers to the first 1000 days of life, middle childhood, and adolescence (Bundy et al., 2018).

6. naming, counting, or drawing things for or with the child.

This might seem a high standard, but countries with a higher Human Development Index (HDI), but without high incomes, report much higher responses to the majority of those questions. This is illustrated in Figure 1 below (Figure 2 from Cuartas et al.).

Figure 1: Relationship between high maternal stimulation and HDI



As Cuartas et al. (2020) go on to argue, these activities are related to the development of a child's language, cognitive, and socioemotional abilities (Jeong et al. 2017; Hamdani et al. 2010; Sun et al. 2016). And it's through these types of activities that parenting interventions, particularly those labelled "childhood psychosocial stimulation interventions" aim to improve development (Jeong et al., 2018).

To put the point more starkly, people aren't automatically all equally good parents[6]. UNICEF reports that globally, 60% of children between the ages of two and four are violently punished by their caregivers (2017). There have been over 230 million girls and women (around 6% of women) who've undergone genital mutilation (2024). Note that this report doesn't explicitly tackle the topic of violence in the home, but we think it's important enough that we plan to address it in a separate report.

We think that the psychosocial stimulation programmes primarily work through two mechanisms: education related to the benefits of certain practices, how to perform those practices and regular encouragement to maintain them.

1.2 What do these interventions look like in practice?

To give a more concrete sense of the content of these interventions, we share a description of what one of the most studied and widely deployed parenting interventions, "Reach Up", looks

[6] A related explanation is the idea that poverty saps individuals' cognitive bandwidth (Schilbach et al., 2016), and this leads parents to systematically neglect long term investments (Haushofer and Fehr, 2014), including in their children (Lichand et al., 2022; Burlacu et al., 2023). Parenting interventions plausibly jump start good practices that are then reinforced with regular reminders and so are relatively easier to maintain despite the continued pressures of poverty.

like in practice. The intervention is described in the Reach Up sample curriculum, training manual, and several papers studying the intervention (Grantham- McGregor, 2016; Smith et al. 2018). This is also the intervention delivered in the best funding opportunity we find and evaluate later in this report.

The programme employs community health workers to make home visits to parents of young children (often ages 0 to 3) on a weekly or fortnightly basis. These visits may continue for a period of several months to several years. The content of the visit is quite simple. As Walker et al. (2016) describe it:

> "[After greetings and checking in,] the home visitor then engages the mother and child in a play session. The visitor introduces new activities through an interactive approach: observing what the child does, demonstrating and describing the activity to mother and child, helping the child with the activity, encouraging mother and child to practise, giving positive feedback[7], and celebrating success. The visit ends with a review of activities to continue during the week and encouragement to continue the activities and to try to include them in daily routines."

2. Literature and evidence overview

For this review, we used a non-systematic search strategy, which we explain next.

We used Elicit and Google Scholar to look for reviews of parenting interventions and childhood interventions containing evidence of the long term causal effects on any outcome. When we came across a review, we searched through the description of each article included in the review to see if it mentions affective mental health[8] (MHa) or subjective wellbeing (SWB) outcomes. We did not look through every individual article but used summary information. It's possible that we missed some studies this way. A more systematic search in the future could be valuable.

We then, using Google Scholar, searched for articles that cited these reviews, and looked through the titles to see whether they clearly included MHa or SWB outcomes[9]. If there were no reviews, we searched for individual RCTs or natural experiments and followed a similar search strategy. We paid particular attention to papers that were highly cited. When we found a study that held MHa or SWB as a primary outcome, we repeated this process and looked through the studies it cited, and the studies that cited that paper as well using Google Scholar.

As part of an earlier and unpublished project, we found literature indicating that parenting interventions had some long-term causal evidence of a wellbeing effect in LMICs. Given this

[7] The emphasis on positive feedback seems warranted. A recent RCT by Justino et al. (2022) found that a test of messages meant to persuade individuals about parental investments increased maternal time investment by 0.2 SDs when a positive feedback message was added.

[8] By this we refer to mental health issues that relate solely to mood and are captured by measures of depression, distress, anxiety, and potentially post traumatic stress disorder.

[9] We don't think there's a huge chance of missing hidden outcomes with SWB because there have been separate reviews which look at long term outcomes, which are a prerequisite for SWB (e.g., Jeong et al. 2021). And there are only a handful of long-term follow-ups, all of which we have searched for wellbeing outcomes.

starting point, we then looked for further evidence from parenting interventions' long term effects in HICs and the short term effects in LMICs.

Again, the evidence we reference here is very likely not to be exhaustive. There are probably more relevant studies out there and their inclusion could potentially shift our conclusions.
We spent 10-15 hours searching for studies. We stopped when we felt confident we had relatively exhausted the literature we could find with a non-systematic search[10].

2.1 How much evidence did we find on the topic in general?

We think the overall *wellbeing* evidence for parenting interventions is weak. This is mostly because the evidence specifically about the long-run wellbeing effects of parenting interventions is very slim (2 RCTs, n = 426). And the overall effect when meta-analysed is not statistically significant. Note that we won't have short term wellbeing information because people assume (probably correctly) that young children can't reliably quantify their feelings.

We wouldn't say that the evidence is "very weak" because there appears to be robust evidence on plausible mechanisms for the interventions to improve wellbeing through improving factors like cognition. There is also a slightly more robust literature studying the wellbeing effects of similar programmes in HICs. These other sources of evidence are discussed below (and previewed in Table 1). We split the evidence we found into the A) short run effects in HICs and LMICs on non-wellbeing outcomes like cognition, and B) the long-run effects in HICs and LMICs.

Table 1: Evidence of parenting interventions

Evidence	Relevance	Description
Long run wellbeing effect LMICs	High	Early childhood parenting: 2 RCTs, n = 426
Long run wellbeing effect HICs	Medium	Pre-school programmes = 4, n = 2,502
Short run development effect in LMICs	Medium	Reach Up: 18 RCTs, n = 7,463; Early childhood parenting: 41 RCTs
Short run development effect in HICs	Low	Early childhood parenting: 61 RCTs

There is considerably less wellbeing evidence of parenting interventions compared to the best evidenced interventions we've reviewed: cash transfers (45 causal studies, n = 116,999), psychotherapy (84 RCTs, n = 25,363), anti-malaria bed nets (23 RCTs, n = 275,000) and even deworming (1 RCT, n = 5,259). But we feel like we are in a slightly favourable evidential position relative to other more exploratory reports we have performed, such as on lead exposure, immigration, or pain (none had directly relevant RCTs).

[10] We think we are reaching diminishing marginal returns when we've searched for an hour or two and found no promising papers that study or reference papers that directly study the causal impacts of an intervention we are searching for.

3. Wellbeing effects of psychosocial stimulation

To estimate the effect of early childhood psychosocial stimulation programmes on long-term wellbeing, we aggregated evidence from LMICs and HICs using subjective weights (50/50). The evidence from LMICs is more relevant to the funding opportunities we identified, but there's more evidence from HICs. We also discuss the degree to which similar psychosocial stimulation interventions improve non-wellbeing outcomes as a way to sanity check our long-term wellbeing results.

3.1 Psychosocial stimulation's long effects on children in LMICs

3.1.1 The Jamaica RCT and the Reach Up programme

The long-run MHa effects on adolescents or adults who were exposed to parenting interventions as children are captured in two RCTs. Both of these interventions are home visiting programmes where a lay practitioner visits the parents and encourages them to play with their children in an enriching manner.

The first RCT, which I refer to as "the Jamaica study", was initially studied in Grantham-McGregor et al. (1991). There were three long-term follow-ups reported by Walker et al. (2021; 2011; 2006). These studies are detailed in Table 2 below. Notably, their sample size is quite small.

With any long-term follow-up, there will be attrition, and that's the case with this trial. A sizable amount of this appears to be due to death. Eleven individuals died by the 29 year follow-up, constituting about a third of all attrition. Other attriters were just lost or refused to participate. Attrition was balanced, with control and treatment arms differing by only a few individuals across follow-ups, limiting the likelihood the results are systematically biassed due to dropout. Given the very long time frames involved, this seems like an impressive follow-up rate, and we don't feel particularly concerned about attrition in this instance. However, in a more in-depth report, it may be worth considering how accounting for attrition could increase the uncertainty in these effects.

Table 2: Jamaica parenting studies

Study	Years after intervention ended	Sample size	MHa (depression) measure
Grantham-McGregor et al., (1991)	0	127 (129 baseline)	NA
Walker et al., (2006)	14	103	SMFQ-SR
Walker et al., (2011)	19	105	SMFQ-SR
Walker et al., (2021)	29	95	CESD-20

The intervention content is extremely similar to the "Reach Up" programme described above. Indeed, the content of the "Reach Up" programme was adapted from the one delivered in the

original RCT with few alterations. What changes they made appear to mostly involve optional adaptations to decrease costs by delivering fewer sessions or to groups instead of individuals.

The programme involved 50 to 100 hour-long house visits (depending on the child's age) for 2 years to families with very young children (ages 1 to 2) who had stunted growth living in poor neighbourhoods of Kingston, Jamaica. These children were recruited in a "house to house survey". The fact these children are stunted potentially raises generalisability concerns, which we haven't had time to investigate here. The control group was assigned no intervention, which is often the more realistic alternative to the services an organisation delivers in practice, so this is often our preferred type of control group.

The Jamaica study finds large and statistically significant beneficial long term effects of the parenting intervention on depression (0.47, 0.42 and 0.58 SDs, 14, 19 and 29 years after). These imply a very large total effect of the programme[11]. They also found persistent effects on other outcomes like income and cognition.

Despite its initial small sample size (n = 127), its long-term effects have inspired a coterie of adaptations in at least 10 countries, including becoming government policy in Peru.

3.1.2 Are these long-term effects plausible?

The best sanity check we have is to see how well the non-wellbeing effects replicate since there are other RCTs that study the short term non-wellbeing effects[12]. The reasoning is that if the non-wellbeing outcomes replicate, then this should increase the plausibility that the long-term wellbeing outcomes will also replicate.

Jervis et al. (2023) meta-analyses the causal evidence (which it rates as moderate quality on GRADE[13]) of the effects of the Reach Up parenting programme on developmental outcomes. Again, the Reach Up programme was based on the programme from the Jamaica RCT. Using this evidence, we can compare the non-wellbeing outcomes of the Reach-Up programme in general, to the original trial (Grantham-McGregor et al., 1991).

While Jervis et al. included the original trial, it only plays a small role in the average effects given its low power. In Table 3 below, we compare the effects of the original trial to the meta-analytic effects of Reach Up across the non-wellbeing[14] outcomes relevant to a child's development that they share.

[11] Note that we're unsure exactly whether the effects, insomuch as they replicate, stem from the specific curriculum or just from the company the regular visits provided.
[12] We can't observe the short term wellbeing effects on children, because the recipients are often very young (under 3).
[13] This is also the system for assessing the quality of evidence that HLI uses. This article provides a brief overview.
[14] Jervis et al. (2023) also reports that the typical effect of Reach Up on "Home stimulation", which is measured as the response to sets of questions asking about stimulating resources and parenting behaviours based on the FCI or HOME set of indicators. The effects are similar (0.39 SDs for original trial, 0.37 for average across RCTs).

Table 3: developmental effects in SDs (95% CIs) reported in Reach Up RCTs

Study	Programme	Length months	Sample	Lang and Cognitive	Fine Motor	Gross Motor
Grantham- McGregor et al. (1991)	Jamaica Home Visiting	24	1 RCT, n = 127	0.50 (0.15, 0.84)	0.67 (0.34, 1.00)	0.34 (0.01, 0.67)
Jervis et al. (2023)	Reach Up (LMICs)	6 to 12	18 RCTs, n = 7,463	0.39 (0.24, 0.51)	0.29 (0.24, 0.51)	0.05 (-0.05, 0.31)

The average replicated effect is 45% of the original. The replicated effects on these developmental outcomes[15] make us think it is likelier that the long term MHa effects will replicate to some degree. If they do, we think it'll probably be a smaller magnitude. Exactly how much the long-term effects truly replicate in practical settings depends on a few factors:

- The degree to which the Jervis et al. (2023) RCTs will generalise to practise.
- The degree to which those studies suffer from publication bias or other factors that inflate the effects reported.
- What we expect the mechanism for long-term impacts to be, and how much of it flows through:
- the replicated channels (language, cognition, and fine motor skills)
- versus other channels that didn't replicate (e.g., gross motor skills)
- or aren't observed (e.g., early in life mental health).

We try to make some adjustments (discussed in Section 3.5) to deal with these concerns, but these are necessarily speculative.

Note that the effect size of the non-wellbeing outcomes found in Jervis et al. (2023) are similar in magnitude to those reported for early childhood parenting interventions more broadly (Jeong et al., 2021; RCTs = 102).

3.1.3 The South Africa RCT and Parenting for Lifelong Health

We refer to the second study of a parenting intervention in a LMIC with long-term causal evidence as the "South Africa study". This RCT includes results from St. Claire et al. (2019), which follows Cooper et al. (2009) 13 years after the original trial. This intervention was briefer, totalling 16 hour-long home visits, focused on younger children (newborns, 0 to 6 months) and was deployed by lay counsellors instead of community health workers. The aim was described as "to encourage the mother in sensitive, responsive interactions with her infant." which doesn't sound very different in content than the "Reach up" programme, but perhaps with a slightly more clinical approach. In addition, both the treatment and control groups received a visit from a community health worker who "*assessed the physical and medical progress of mothers and infants, and encouragement [sic] of mothers to take their infant to the local clinic to be weighed, have their physical health assessed, and be immunised.*" This control condition seems more potent than the "nothing" comparator in the Jamaica study.

[15] See World Bank (2023) for a discussion of what constitutes developmental outcomes.

Similar to the Jamaican RCT, there's an organisation (Global Parenting Initiative) dedicated to spreading and scaling this programme, which refers to the South Africa programme as "Parenting for Lifelong Health for infants". However, the organisation doesn't appear directly involved in delivery and we don't find any organisations that are – a topic we discuss further in Appendix B.

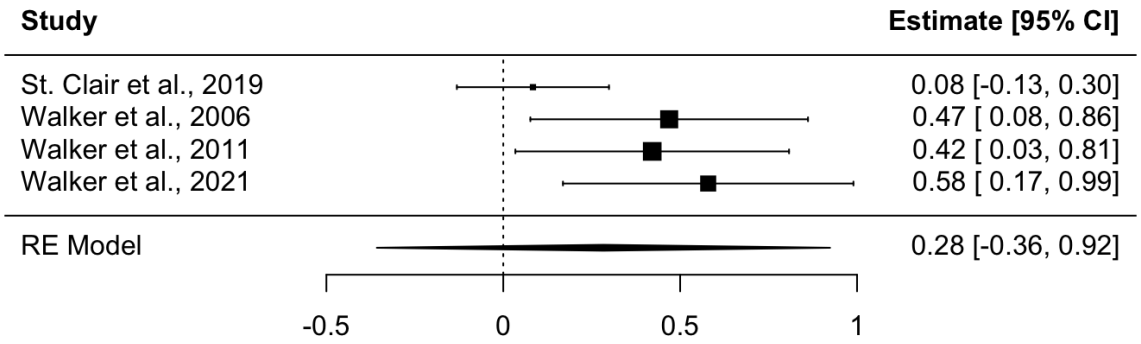
St. Claire et al. (2019) found a small positive (0.08 SD, n = 333) but n.s. effect on self-reported distress after 12.5 years for the individual who was exposed as a child. However, Tomlinson et al. (2021), also a 12 year follow-up to the original study (Cooper et al.; 2009), found a persistent positive effect on *maternal* depressive symptoms (SMD of 0.24). Notably, maternal mental health has been repeatedly evidenced to be a good predictor of children's mental health (Luoma et al., 2001; Nadeem et al., 2018; Coles & Cage, 2022). This makes St. Claire et al.'s small, null result for the child's wellbeing somewhat surprising.

3.1.4 Meta-analytic summary of results

We combine the results across these studies using a meta-analysis, which in simple terms is an average weighted by the precision of each study's results – with the precision often being highly related to the sample size of the study. More technically, we use meta-analytic models that allow us to leverage more complex data structures such as having multiple time points per study without giving a study with more data points disproportionately more weight. For more discussion of how we use meta-analyses see Section 2.2 of our psychotherapy report for the most up to date discussion (McGuire et al., 2024b).

The meta-analytic effect combining results of the Jamaica and South Africa studies is 0.28 (95% CI: -0.36, 0.92) SDs after an average of around 20 years and is not statistically significant ($p > 0.10$). The forest plot is shown below in Figure 2. The null overall effect is a result of the null finding from the better powered St. Clair et al. and the clear difference between the two studies (which increases the error of random effects models through the inter-study variance term).

Figure 2: Forest plot of LR effects on young children in SDs



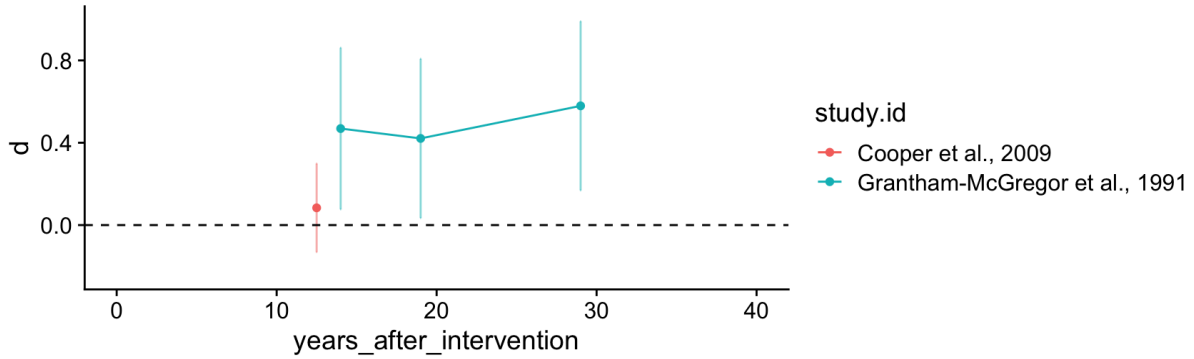
Note: The effect is in SDs (Cohen's d), calculated using a multi-level meta-analysis. Effect sizes were standardised so a positive effect represents a beneficial effect or improvement in wellbeing.

Note that Jeong et al. (2021) find similar differences in effect sizes between the Jamaica and the South Africa study for cognitive development outcomes. They suggest this difference is due to the substantial differences in "dosage" between the two interventions. Recall that the Jamaica

study involved 50-100 home visits in total over two years, while the South Africa study totalled only 16 home visits across the first five months post-delivery[16].

In Figure 3 below we plot the effect sizes of the interventions against how long after the intervention they were observed. While the effects are very imprecisely measured we think it suggests the possibility of long-term and persistent effects of psychosocial stimulation in early childhood. Based on the flat trajectory of the Jaimaca studies we assume the effects are constant over time when we estimate the total effect.

Figure 3: long term effects of parenting programmes in LMICs over time



3.1.5 Potential further evidence on the long-run effects in LMICs

Evidence is scarce here, so it's worth considering how we could expand the evidence we use in our analysis. We list some possibilities for further research in order of promise:

- There are many Reach Up replications listed in Jervis et al. (2023). Most of them don't yet have long-term follow-ups. It seems worth considering whether to encourage or invest in more long-term follow-ups, particularly if they look into the long-term effects on SWB or MH outcomes.
- Replicate and expand Yang (2021), to include the effect of childhood education on later life-satisfaction or mental health[17].
- A systematic review of the literature to see if there is more SWB evidence of related interventions in LMICs.

3.2 Psychosocial stimulation's long term effects in HICs

Because of the limited evidence for the long term wellbeing effects of psychosocial stimulation interventions in LMICs, we tried to collect broader evidence that could inform our views. We considered some other sources of evidence but decided to focus on reviewing the long term effects of childhood psychosocial stimulation programmes in HICs. While we think this evidence

[16] If we crudely adjust for the St. Clair et al. results for dosage by taking the average number of session in Jamaica to be 75, then we can say that the Jamaica study has 75/16 = 4.7x more sessions than those in the SA study, therefore we would expect that with an equal number of sessions the SA effect might increase to 0.08 * 4.7 = 0.38, which is much closer to the Jamaica effect sizes.

[17] Yang (2021) uses a propensity score matching approach, which is considered a weaker form of causal identification because it requires matching on all potential confounders (Craig et al. 2017; Kim and Steiner, 2016).

is much less relevant, given that there's much more of it, we end up weighing the smaller effects found in these studies the same amount as the effects found in evidence from LMICs.

The other sources of evidence we considered were: A. the long term effects of being adopted earlier and B. the near term effects of parenting programmes on *older* children and adolescents in LMICs. However, we think the other sources of evidence differ too much in the intervention content, population or context to be relevant. For conciseness, we have moved the discussion of these less relevant sources of evidence to Appendix A.

Across all of these evidence sources, we **only performed a very shallow and quick review.** That is, we were less thorough than we were on the LMIC evidence. We certainly missed some studies, and it's plausible their inclusion could change our results.

Preschool psychosocial programmes for young (but not very young) children

The evidence we found from psychosocial programs carried out in HICs is not from parenting programmes. That is, they don't involve home visits or primarily provide instructions for parents on how to interact with their children in a more psychologically enriching manner.

These psychosocial stimulation programmes are primarily educational programmes for preschoolers. The population studied includes older children than the programmes discussed in the previous section (ages 0 to 6 compared to ages 0 to 3). Given the older sample, the content is more sophisticated and aimed more directly at preparing children for school. There are also varying levels of parental involvement depending on the exact programme. However, we think they share a core mechanism: creating a more stimulating psychological environment for young children.

Due to the shallowness of the review, we think we probably have a biassed (and optimistic) sample of published papers. This is because we mostly included the studies we could easily find, which may tend to report larger effects. Our concern about this is a reason why we are comfortable with the large discount we apply (described in Section 3.5).

The studies, which we summarise below in Table 4, use a variety of techniques to estimate the causal effects of early childhood education on wellbeing later in life. We briefly explain the content of each programme and the design of its study next.

The Child-Parent Center programme (CPC) involves both a child[18] and parent[19] based educational component. Studies of the CPC, while reasonably well powered, attempt to estimate

[18] "The CPC program provides educational and family-support services between the ages of 3 and 9 years (preschool to second or third grade). Within a structure of comprehensive services similar to Head Start, the intervention emphasises the acquisition of basic skills in language arts and maths through relatively structured but diverse learning experiences that include teacher-directed whole- class instruction, small-group and individualised activities, and field trips." Reynolds et al. (2007).

[19] "An intensive parent program that includes parenting education, volunteering in the classroom, attending school events and field trips, furthering educational attainment, and home visitation; and health and nutrition services, including screening and diagnostic services, speech therapy, meal services, and referrals by program nurses. Parents are expected to participate in the program up to half of a day per week through a variety of supported activities." Reynolds et al. (2007).

the effects based on a matching estimator where a control group was statistically produced to resemble the treatment group based on observable characteristics. This is conventionally considered one of the weaker forms of causal estimation as it strongly depends on the richness of the observable characteristics collected in the data.

Table 4: Overview of psychosocial stimulation studies in HICs

Programme	Study	Effect size (sds)	Population and design	Sample (children)	Follow up
Child-Parent Center (CPC)	Reynolds et al. 2007 ; Mondi et al. 2020	0.20, 0.32	Ages 3-4 in 1980s Chicago from low income, mostly black families. Control made of matched preschools with a normal curriculum.	1,142 - 1,368	19 years
Education + family participation	Varshney et al. 2022:	-0.027		1,098	33 years
Head Start	Carneiro and Ginja (2014)	0.2	Ages 3-5 in 1980s USA from low incomes. Design was regression discontinuity based on income.	1,053	12.5 years
Education + family participation	Lacey (2023)	0.1		2,084	25 years
Carolina Abecedarian Project	McLaughlin et al., (2007)	0.4	Ages 0-5 in from low SES status*. Design was an RCT.	101	15 years
BEEP	Palfrey et al. (2005)	0.18	Ages 0-5 in 1970s Boston from low income. Control group was made of matched children.	120	25 years

*"All families had to qualify on the basis of a High-Risk Index (Ramey & Smith, 1977) based on sociodemographic characteristics." McLaughlin et al., (2007)

The Brookline Early Education Project (BEEP)[20] also used a matching estimator but on a much smaller sample (n = 120).

The Head Start program[21], a federal programme aimed at directly providing early childhood education and stimulation alongside education for parents, has two studies of long term effects of exposure (not necessarily participation). The studies use a regression discontinuity design based on income eligibility and provide, in our view, the most reliable results.

The Carolina Abecedarian Project, which also provided educational childcare like the other programmes[22] was studied using a classic RCT design, but on a small sample size (n = 101).

[20] Palfrey et al. (2005) describes the BEEP as "The program's essential goal was to ensure that children in the project would enter kindergarten healthy and ready to learn. The project's interventions were based on a growing knowledge base indicating that children's developmental trajectories depended to a large ex- tent on the nurturance and support they received in the first few years of their lives.26 The major inter- vention was a multifaceted health and education program designed to engage the children's families as their first and best teachers."

[21] Carneiro and Ginja (2014) states that Head Start "provides comprehensive educa- tion, health, nutrition, and parent involvement services to around 900,000 low-income children 0 to 5 years old (of which 90 percent were 3–5 years old in 20095 and until 1994 the program served children ages 3–5) and their families."

[22] "Treatment was an intensive child- based program delivered in a full-time child-care setting from infancy until kindergarten entry." (McLaughlin et al., 2007)

There are also plausibly relevant studies we were unable to include because we lacked sufficient information to calculate an effect size from these studies. With more time we would contact the authors and try to include these and other studies to form a more representative sample of early childhood psychosocial stimulation programmes in HICs. We discuss these studies in a footnote[23].

3.2.1 Meta-analytic effect of psychosocial interventions in the USA

The combined effect (0.17 SDs) displayed in Figure 4, is around half the size of the long-term effect found in LMICs (0.28 SDs). However, the effect is more precisely estimated and is statistically significant ($p < 0.05$). The average follow-up is about the same (20 years) as the average in LMICs.

Overall, we think these studies usefully update us towards believing that psychosocial stimulation programmes can have long term effects on wellbeing.

Figure 4: Forest plot of long-term effects on young children in HICs

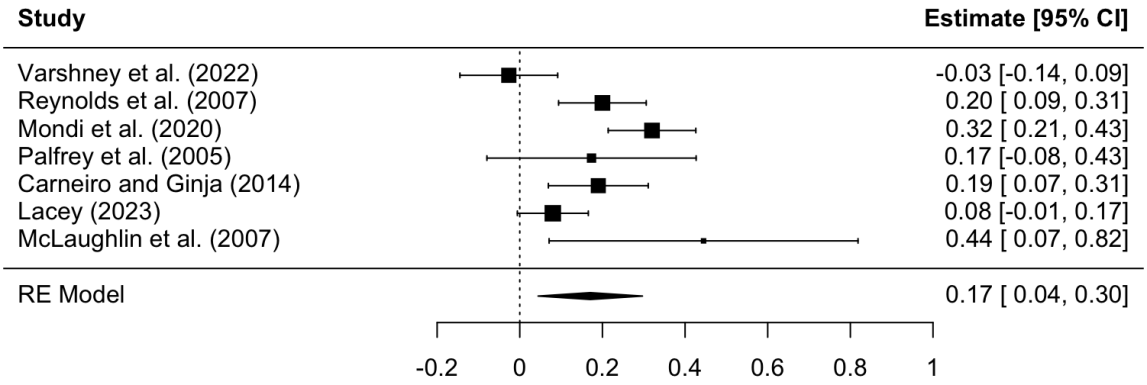
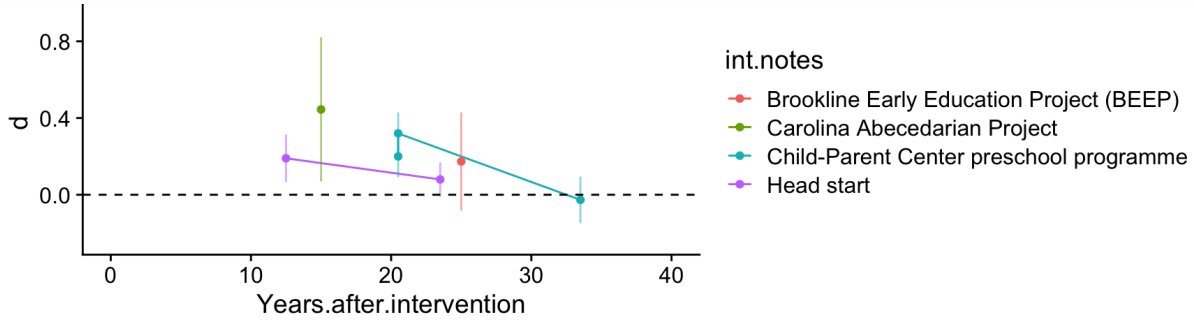


Figure 5: Plot of long-term effects on young children in HICs



[23] The excluded studies were Dunkel et al. (2021) for the Carolina Abecedarian Project, which had a longer term follow-up, but reported insufficient information to extract an effect size. Similarly with Kagitcibasi et al. (2009) which studied the 19-year follow-up of an early childhood enrichment program in Turkey. We also excluded Ginja (2021) which studied Sure Start, an early childhood programme in England, because its mental health outcome (a significant decrease in the likelihood of a mental health hospital admission in adolescence) was not a self-reported outcome.

3.2.2 Gathering further evidence from HICs

We think it'd be valuable to perform a more systematic review and meta-analysis of studies with an analysis of publication bias to combat what we think is likely (a) a selection bias in our search process and (b) publication bias. Both of these biases likely inflate the positive effects we found.

3.3 Forming a view on the long-term effects of early childhood psychosocial stimulation interventions

Here, we attempt to combine the broader evidence from HICs with the more direct evidence from LMICs based on subjective weights. We assigned 50% of the weight to the long-run evidence from LMICs, and 50% to the long-run evidence from HICs. We summarise the effects based on these two sources of evidence in Table 5 below.

We assigned as much weight to the long-run evidence from HICs as from LMICs because the HIC evidence, while less relevant, studies interventions with similar content, time of follow-ups and importantly much better power. This was, of course, the result of a rather quick analysis that we could improve on with time.

The result of this synthesis is an estimated effect of 0.23 SDs (at an average follow-up of 20.5 years). We assume this effect lasts around 32 years. This estimated duration comes from the trend suggested in the moderator analysis of the HIC data (see Figure 4 above). Note the trajectory is n.s. and increasing for the long-run LMIC data, which would imply a larger effect if we put any weight on this other trend (so using the trend from HICs is more conservative). A duration of 32 years also seems plausible given that we see positive effects at more than 30 years in some studies.

Table 5: Long-run (LR) effects of psychosocial interventions on mental wellbeing

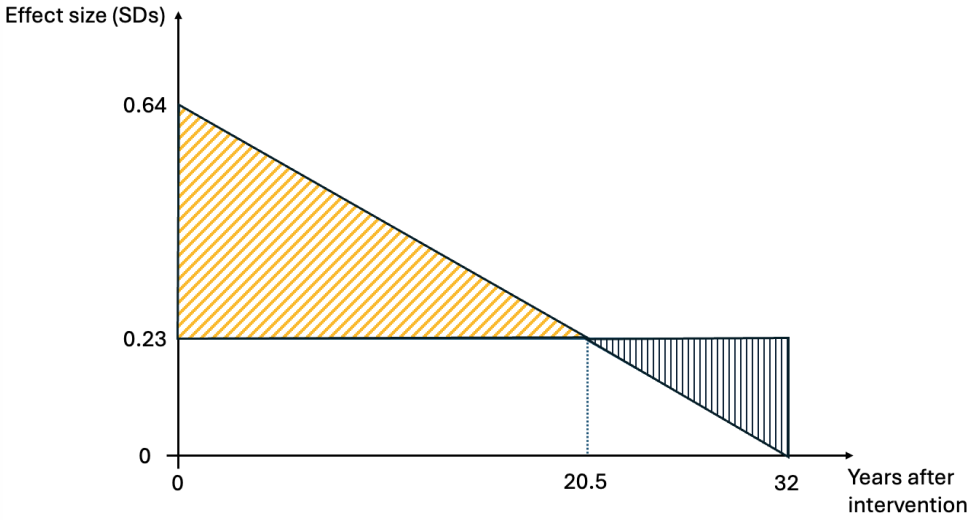
Evidence	Description	Follow up range	LR effect in SDs	Weight or credence
LR effect LMICs	RCTs = 2, n = 426	12 to 30 years	0.28	0.5
LR effect HICs	Interventions = 4, n = 2,502	10 to 35 years	0.17	0.5
Weighted average			0.23	

We assume a flat trajectory. In other words, the initial effect is the same as the follow-up effect. While this may imply the effects are lifelong, we don't accept this implication because we are cautious of taking the data too seriously (i.e. extrapolating beyond the data we have). A flat trajectory is also a conservative choice because if there is decay in the effects, as we normally assume, then the effects earlier in life (and thus the total effect) would be much larger. The studies we build our estimated effect size from have an average follow-up time of 20.5 years i.e., at the tail end of the decay process. Diagrammatically, if we imagine there would have been linear

decay at the rate suggested by the HIC data, then as we show in Figure 6 below, by assuming a flat trajectory we lose the orange area and gain the (smaller) blue area.

Finally, as per our general approach, we convert SD-years (e.g., a 0.2 SD effect lasting 10 years is 0.2 * 10 = 2 SD-years) to WELLBYs (e.g., points on a 0 to 10 wellbeing scale) based on a conversion factor of 2 – where the 2 represents the global average SD of life-satisfaction (see Dupret et al., 2024 for more details).

Figure 6: diagram illustrating why a flat trajectory is in fact conservative



A 0.23 SD effect lasting for 32 years (with a 2x SD-year to WELLBY conversion rate) implies a total individual effect of 0.23 * 32 * 2.0 = **14.54 WELLBYs**. This is huge, but smaller than the 20.48 WELLBYs we would estimate if we assumed the effects decayed at the rate suggested by the HIC data.

3.4 Household effects

There is some evidence of the effect of psychosocial stimulation interventions on household members other than the child recipient. Jervis et al. (2023, k = 18) who meta-analysed Reach Up RCTs found they led to, on average, a non-significant reduction in maternal depression (–0.09 SDs, 95% CI: –0.19 to 0.01). It's unclear what's the average follow-up as it's unreported. It's also unclear if we should call these "spillover effects" in line with our typical characterization of household effects (e.g., McGuire et al., 2022) – since the mothers are often involved in these programmes. However, we maintain the label of "spillover" to be consistent in terminology and structure with our other reports.

Jeong et al. (2021) meta-analysed the effects of parenting interventions in the first three years of life more broadly across HICs and LMICs[24] (RCTs = 102). They found a similar magnitude effect of -0.13 SDs on maternal depression symptoms (RCTs = 12, 95% CI: -0.27, 0.01). We are

[24] An earlier meta-analysis, Jeong et al. (2018), found that stimulation and parenting interventions in LMICs had a less precise effect on maternal depressive symptoms (k = 9; SMD = −0.10; 95% CI, −0.23 to 0.03). Shah et al. (2022) meta-analysing healthcare based parenting interventions found a null effect on parental depression (no access, SMD and k unclear).

also aware of at least 7 recent RCTs that haven't been included in either meta-analysis, which suggests further work updating these with new studies could be useful[25].

After a shallow search, we only found one long-term study of the wellbeing effects of a psychosocial intervention on parents. This study is Tomlinson et al. (2021), which we already mentioned in Section 3.1.3. They found, in a 12-year follow-up to a parenting intervention in South Africa (Cooper et al., 2009), a persistent positive effect on maternal depressive symptoms (A reduction of -0.24 SDs). This is while the programme had a n.s. effect of -0.08 SDs at the same follow-up on children.

We also found an RCT that captures the (presumably short term) mental wellbeing effects on siblings. Jeong et al. (2023) find that (older) siblings of children who received a parenting and nutrition intervention have lower internalising symptoms and this represents a fairly large standardised effect (0.3 SDs). However, it's unclear how much of this effect is attributable to the parenting or nutrition component and the timeframe is unclear since we are unable to access the article.

Together, we think that this evidence is sufficient to assume that parenting interventions will have at least the same 16% spillover for their household we use in our psychotherapy report (McGuire et al., 2024b). A 16% spillover is equivalent to an effect of 0.04 SDs. This assumption of 16% thus seems conservative given the evidence we've discussed suggests larger effects than 0.04 SDs. Note that even if we assumed fathers (who are often a smaller share of households than mothers) had no spillover effect, the average spillover effect would still be larger if we took the data seriously. Here, we prefer to use the more conservative 16% figure. This is in line with our more general preference for taking a conservative tact with more speculative analyses (such as this one) since we think that effect sizes tend to shrink with further scrutiny. We assume this spillover benefit will last for as long as the effect on the recipient persists.

The average household size of Bangladesh (which is where the programme we evaluate in a later section is based) is 3.98[26]. Based on this, including household effects would add a (3.98-1) * 16% * 14.54 WELLBYs = 6.94 WELLBY effect for non-recipient household members, for a total household effect of 14.54 (recipient) + 6.94 (other household members) = **21.48 WELLBYs in total.**

[25] There are more studies not included in previously mentioned meta-analyses that also study the effects of a parenting intervention on maternal depression (Bliznashka et al., 2020; Baumgartner et al., 2021; Gureje et al., 2022; Islam et al., 2022; Tofail et al., 2023; Carneiro et al., 2019, Brathwaite et al., 2023). These happen to mostly find statistically significant benefits. In the future we could collect and synthesise these studies with any further new evidence we find (e.g., in Kaminski et al., 2022). Some of these interventions specifically target parental mental health with, for example, a psychoeducation component (e.g., Islam et al., 2022) – these components are notably not a part of the interventions we're focusing on here though.

[26] This is calculated by estimating the trend in household size over time using the UNPD data, and then extrapolating out to 2024.

3.5 Total effect

In the previous section, we arrived at the naive estimate for the total wellbeing effect of psychosocial stimulation interventions. However, we do not take these results at face value. To make these results more realistics, we apply several adjustments and discounts (explained after Table 6).

First, we apply the standard 51% replicability discount to these findings[27]. On top of that add a 45% discount for a Reach-Up specific replicability discount. This Reach Up specific discount comes from comparing the effect sizes from the meta-analysis, Jervis et al. (2023) of Reach Up RCTs compared to original Reach Up RCT, Grantham-McGregor et al. (1991) – a comparison we made in Section 3.1.2.

Table 6: Total effect of parenting interventions.

Parameter	Value	Description
Individual effect (SDs)	0.23	Average of HIC and LMIC LR effects.
Duration	32.30	In years (from HIC LR meta-regression)
Mental health to WELLBY conversion	2.00	SD-year to WELLBY conversion
Total individual effect in WELLBYs	14.54	Calculation
Household spillover effect	16%	From our psychotherapy report
Household member effect (WELLBYs)	2.3	Calculation
No. non-recipient house members	2.98	Prediction based on UNPD data
Total household member effect (WELLBYs)	6.94	Calculation
Total household effect (WELLBYs)	21.48	Calculation
Reach-Up specific replicability adjustment	0.45	From comparing Jervis et al. (2023) with Grantham-McGregor et al. (1991)
General replicability adjustment	0.51	Weighted average of the replication effect across multi-lab replications.
Total adjustment	0.23	Calculation
Total adjusted household effect (WELLBYs)	4.93	Calculation
Expected cost to treat one person	$98.34	
WELLBYs per $1k spent on parenting	50.12	Calculation
GiveDirectly WELLBYs per $1k	8.17	From our cash transfer paper
Parenting CE (in multiples of GiveDirectly)	6.14	Calculation

While we think these discounts may overlap considerably and potentially double count, we think they also account for our general scepticism at the size and persistence of these effects. Note that this sort of scepticism can be rationalised in terms of Bayesian epistemology. Our 'prior' belief is that such interventions - indeed most interventions - won't have large, long-lasting effects. We

[27] The adjustment is calculated as a weighted average of the proportion of the size of effect sizes as replicated in replication studies in the broader social science literature: based on the results from Camerer (2018, n = 21), Open Science Collaboration (2015, n = 94) and the Multi-Lab studies (1,2,3,4; n = 77), as reported in Nosek et al. (2022). However, it seems plausible that this is too harsh given that a higher than average proportion of Reach Up replications reported in Jervis et al., (2023) had pre-analysis plans (11/18).

then update this prior based on the evidence. In this case, we are doing this process in reverse: we look at the evidence, then apply a discount based on a sceptical prior.

Putting both of these discounts together results in an overall 0.51*0.45*21.5 = **4.93 WELLBY effect.** We summarise the total effect calculation in Table 6 above.

4. Programmes, organisations, and costs

There appear to be many large, multi-national organisations deploying parenting psychosocial stimulation programmes, but we struggled to identify any small to medium scale charities delivering thes. We tend to prefer small to medium organisations for reasons we elaborate on below. Some examples of organisations that appear involved in parenting and early childhood development that have a budget of around $1 billion or more:

- Children's Investment Fund Foundation
- International Rescue Committee (IRC)
- Catholic Relief Services
- Plan International
- Save the Children
- The World Bank
- World Vision
- UNICEF
- BRAC

However, we think evaluating funding opportunities in very large charities is difficult. In order to keep the report a reasonable length we address these concerns in a separate short document. However, below we paraphrase our key takeaways, which we explain in more depth in the memo. When we fund a large organisation, we think in many cases we're faced with concerns about:

- Fungibility - Are we just freeing organisations to spend more on their ineffective projects?
- Transparency - In large organisations it's often very difficult to find basic details necessary for an evaluation like the cost it takes to deliver a programme, how much money that programme can usefully absorb, or if there is a way to donate to that programme directly.
- Diseconomies of scale - The cost per person treated doesn't always shrink (economies of scale), it can sometimes grow – especially with organisations that span a wide variety of unrelated activities.

For these reasons, we attempted to find a smaller organisation where our funds go further. However, the only funding opportunity we have confirmed is Reach Up in Bangladesh delivered by the icddr,b – which had an annual budget of $78 million in 2022[28]. We only spent around 10

[28] The icddr,b reported $78 million in expenses in 2022, of which $68 million was restricted (icddr,b, 2022). While they have a lot of large institutional backers (Gates, USAID, UNDG, CDC, etc). Fungibility may be less of an issue due to the large share of restricted funding.

hours searching for charities so with more time we would likely consult experts in the field, and do a deeper search for charities.

We came across some other possibilities that seem plausibly cost-effective that we haven't yet confirmed (no response from charities). We detail these in Appendix B.

4.1 Reach Up in Bangladesh by the icddr,b

Reach Up is the programme based on the Jamaica RCT discussed above. As we have argued, it is the best (but still very weakly) evidenced programme for the long term wellbeing effects of an early childhood parenting intervention. The eponymous Reach Up organisation appears focused on coordinating research and supporting training others in the "Reach Up" curriculum and does not appear involved in active delivery. While it may be worth it to review their work through the lens of "advocacy", given that advocacy interventions are almost always much more speculative, we focused here on opportunities for direct delivery.

The only organisation that we have confirmed is deploying Reach Up and is open to further funding is the icddr,b[29]. As of June 2024, they've said that using the Reach-Up programme, "...more than 14,000 caregivers of children aged 6-36 months have so far been trained [unclear since when]." (icddr,b, 2024).

Cost

Hossain et al. (2024) attempt to estimate the cost of the program once scaled by extrapolating from the micro-costing performed in two RCTs in Bangladesh: one of a group format (Mehrin et al. 2021, n = 715) and home visit programme (Hossain et al. 2022, n = 599). They estimate that the costs of the programme in the RCTs were $156 for the group intervention and $136 for the home visiting programme, but that at scale (416,000 children[30]) the costs would decline to $44 and $47 per child.

These costs are within the range of costs reported in RCTs of Reach Up collected by Bowden et al. (2024). Bowden et al. found that the lowest cost for Reach Up was group delivery in India at $38 per mother-child pair.

We also think that these costs at scale-up are plausible, and probably on the conservative side. This is because it seems like Reach Up, which involves training a lay person to discuss a topic, should have similar economies of scale as psychotherapy. For reference, StrongMinds' costs dropped by nearly an order of magnitude from their start in 2015 (approximately RCT scale: n = 1,289, cost per person = $419), to 2023 (arguably "at scale" with ~200k reached in 2023, cost per person = $41).

[29] The Reach-Up website claims that the intervention is only clearly being scaled in Bangladesh in collaboration with the IRC and the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b). However, the IRC doesn't mention anything like Reach-Up in their 2022 or 2023 report on activities in Bangladesh. For the IRC's Bangladesh office, they report only a small share of donations coming from private donors (8% of total).

[30] The authors explain that this comes from a previous estimate of the capacity of the 13,000 existing community clinics in Bangladesh.

Back of the envelope cost-effectiveness estimate

To estimate the cost for icddr,b to deliver Reach Up at their present scale (treating ~14,000) we interpolate between the estimated cost at scale ($46 when treating 416,000 people) and the RCT cost per person treated ($156 and $136 for treated populations of 715 and 599[31]). To perform this interpolation we assume a logarithmic relationship between scale and cost so that every doubling in treatments reduces the cost the same. We show the calculations in Table 7.

Using this approach we estimate that the icddr,b delivers Reach Up for $98.34 at their current scale (around 14,000 treated per year). Given the effect of 4.93 WELLBYs per treatment estimated for the programme generally (after a 77% discount), this would lead to an estimated 4.93 * (1000/98.34) = **50 WELLBYs per $1,000 donated (WBp1k).**

Table 7: interpolation of costs between RCT and predicted cost at scale

Cost exercise for Reach up in Bangladesh	Children treated	Cost ($)
RCT cost	657	146.00
	1,345	134.83
	2,755	123.67
	5,642	112.50
	11,553	101.33
	23,658	90.17
	48,445	79.00
	99,205	67.83
	203,148	56.67
Estimate of cost at scale	416,000	45.50
Therefore at 14,000...	14,000	98.34

We think there are several interesting features of this opportunity that potentially make it more cost-effective than estimated:

1. We could place more weight on replications that occur in South Asia. The effects reported in the Bangladesh trials of Reach Up on cognition, language, and motor skills are higher than the Jamaica RCT effects measured at the follow-up[32].
2. The costs may be conservative (but we aren't sure). StrongMinds started with much higher costs and at a smaller scale reached smaller costs. Granted, Reach Up is much more intensive than StrongMinds so costs may stay higher if the variable costs relating to deliverer recruitment, training, salaries and management make up a large share of total costs.
3. The icddr,b appears like it's trying to shift the delivery of the programme to the Bangladeshi government. As with any attempt to change policy, especially one that as

[31] We take a simple average of the cost and number treated between the RCTs to set the bottom of our scale (i.e., $146 when treating 657 people).

[32] The effects for cognition are 0.85 and 1.3 SDs for the group and pair interventions; 0.69 and 1.1 for language and 0.52 and 1.2 for motor (Hossain et al., 2024). Compare to the effects of the original trial shown in Table 1.

Hossain et al. (2024) estimates, could consume 0.5% of their health budget, this seems like a tall order. But given that the icddr,b appears to be one of the country's most well respected research organisations, we would probably give this effort a higher chance of success than the base rate (which we haven't established).

Of course, weighing against these factors is the massive uncertainty about the long-term effects, which are crucial for the cost-effectiveness estimate. We discuss our key uncertainties, and how they influence our overall view in the next section.

Funding gap

This programme seems potentially cost effective and worth funding. Dr. Jena Hamadani, a scientist working in the child health department at the icddr,b told us that the programme can absorb potentially millions more in funding. However, she was unable to provide a clear sense of how much funding the project could reasonably absorb. Based on our conversation though, we would guess they could usefully absorb something around $5 million in the next three years[33].

The icddr,b, is a large organisation, so we are trying to establish a way to donate to this project directly and will update this report accordingly. In the meantime, interested parties should email donate@icddrb.org about restricting donations to the Reach Up programme.

5. Evidence quality and depth

We characterise the **evidence quality** as **low or weak**, and thus the analysis that's based on it as speculative.

We provide a shallow assessment of the quality of evidence used in this analysis based on the widely incorporated GRADE (Grading of Recommendations, Assessment, Development and Evaluation) framework, which we explain on our website. Alternatively, see this article for a brief overview.

The highest quality of evidence, in GRADE, is characterised by good study designs (e.g., RCTs), low risk of bias in the studies, precisely measured effects within studies, low variation in effects between studies, high relevance to the real world context, and low publication bias. We didn't do an in depth GRADE analysis, but we will go through our impressions of what we'd expect an analysis to say about each of these factors.

Study design: The evidence we draw on relies primarily on studies with a plausible causal identification strategy (RCTs and natural experiments). This typically indicates high quality.

[33] This comes from a discussion where they mentioned they could roll out to 600 clinics for $1 million USD which would mean $1667 per clinic or around $26 million to roll out across the country (~15,000 remaining clinics), or $2.6 million every year for 10 years. I assumed they'd otherwise receive $1 million in funding each year for a $1.6 gap in the next 3 years.

Risk of bias: The way risk of bias works is that your evidence as a whole is only low risk of bias if the weakest evidence you have is low risk of bias. In our case, we think that at least the most important study we use the: the original Reach Up RCT, would be characterised as high risk of bias due to the attrition in its long-term follow-ups..

Imprecision: Our most important pieces of evidence (the follow-ups to the Jamaica study) have very imprecisely measured effect sizes and low power, making it likelier that their results are due to chance.

Inconsistency: Many of the estimates within and across evidence sources vary substantially, which should undermine our belief in their credibility.

Indirectness: It's unclear how direct the evidence we use is. Some of it is about the Reach Up programme, the programme we are considering funding. But most of that evidence is not about wellbeing or long run outcomes. But around half of the evidence we use is from non Reach Up programmes to support child psychosocial development in high-income countries. Hence, we're using rather indirect evidence on most accounts. And, of course, the most direct evidence we do have – is not very good.

Publication bias: We did not assess publication bias in this analysis. It's also not possible when we only have a small number of studies (less than 10, which is the case here). In the absence of evidence against publication bias, we tend to assume the worst.

For the reasons we outlined above, particularly the high degree of indirectness, we rate the evidence quality as **low or weak**.

We also rate the depth of work that went into creating this estimate as **low**. By this we mean that we believe we have only A) reviewed some of the relevant available evidence on the topic, and B) we have completed only some (10-60%) of the analyses we think are useful. Another way of expressing this is we view this report as shallow. For example, we put around ~80 hours into this report. Our most in-depth reports might have absorbed 5 to 10 times as much time.

6. Conclusion

Here we try and summarise our views on how cost-effective we think it is to improve wellbeing through early life psychosocial interventions, whether there are actionable funding opportunities, and how valuable further research would be.

6.1 The cost-effectiveness of supporting direct delivery

We think there is enough evidence to suggest that the Reach-Up parenting intervention as delivered by the icddr,b may result in around 50 WELLBYs per $1,000 donated (WBp1k). We think it is a reasonable but very speculative bet to assist with their scale up in Bangladesh. We don't think that the current evidence justifies large amounts of funding, but we think that some

amount of funding would likely reveal more information about the cost-effectiveness of this space – which we think would be quite valuable.

The effect per treated child was estimated as an average between the data from two RCTs (one in South Africa, one in Jamaica) of parenting interventions in LMICs and four interventions (studied in seven papers) in HICs. The cost was then derived by interpolating between the predicted cost at scale, and the cost during two small scale RCTs. Of course, this is a very speculative estimate, based on relatively flimsy evidence. We tried to adjust for this with an overall discount of 77%, which was itself a product of guesswork.

We think of the speculativeness of this cost-effectiveness as only slightly less than those in our other cause area reports (lead exposure, immigration, or pain), and significantly more than for our analysis of psychotherapy, cash transfers, and anti-malaria bed nets.

Aside from the icddr,b, this research raised the question of how we should treat funding very large NGOs in general. We address this in a separate short document. We are open to updating our beliefs on this topic, but currently, we are sceptical that we can have as much impact funding very large NGOs that deliver a variety of vaguely related projects, as funding smaller impact-oriented NGOs. Our primary concern is we are unable to evaluate all the programmes carried out by large NGOs. We are also concerned that any small grant ($250k or less) made will be fungible due to their pools of unrestricted funding. If they redirect the money they would have spent on an effective project to other programmes we haven't reviewed then we cannot be confident of a grant's impact.

For this reason, we are interested in identifying smaller parenting NGOs, or NGOs for whom parenting is their only intervention. Due to time constraints and SEO on Google, we were only able to identify very large charities with multiple programs at this time.

6.2 The value of supporting further research

Given our high level of uncertainty in our current estimate, we think there are plausibly quite valuable research opportunities here which would improve our confidence considerably. The opportunity we think is most promising would be to fund the inclusion of SWB measures in any planned follow-ups to Reach Up RCTs (such as those included in Jervis et al., 2023). We haven't looked through these RCTs to prioritise which ones would be best to follow up, but this would be our next step.

While not a Reach Up programme, we previously (in 2023) emailed the authors of Islam et al. (2022) who studied a combined mental health and psychosocial stimulation programme delivered in Bangladesh for quite cheap ($45 per child). We discuss this study in Appendix B since it isn't related to an actionable funding opportunity. The authors welcomed the possibility of funding a follow-up to their RCT.

Appendix A: Orphans and parenting adolescents

The effects of being an orphan

One of the stories about parenting interventions is that they increase the supply and quality of parenting care a child receives. One rather extreme way to bound the effects of an intervention that presumably slightly improves parenting is to see what the effect is on being deprived of having a parent. Some papers use the often random nature of parental loss as a way to study the impact of this common tragedy.

Humphreys et al. (2015) found the effects of being assigned to foster care instead of remaining in an orphanage did not affect internalising symptoms (n = 136) by age 12. But when they compared these children to those who had never been in an institution, unsurprisingly, the internalising symptoms were much higher. Sonuga-Bark et al. (2017), comparing the effects of orphans who spent longer or a shorter period in orphanages in Romania found that the adult outcomes of those who spent more than six months were very bad for just about every outcome.

Askeland et al. (2017) found in a meta-analysis of eleven studies with 17,919 adoptees as adolescents and 1,090,289 non-adopted peers, that international adoptees had more mental health problems in general compared to their non-adopted peers (0.16 SDs, 95% CI: 0.03 to 0.28).

However, most of the evidence we found of orphanhood is related to human capital accumulation.

Beegle et al. (2010) found that adults who were orphaned in Tanzania were shorter by 2 cm and lacked a year less of school. Cas et al. (2014) found that children whose parents died in the 2004 Indian Ocean tsunami had about a year and a half less schooling five years after the event. Garcia-Brazales also finds (2023) that orphanhood results in about a year less of schooling in India.

There appear to be even more studies about the effects of the loss of a single parent. Such as Dupraz and Ferrara (2023) who found that sons of men killed in the American Civil War were far less likely to work in well-paid occupations. Or that having a parent killed in a mine accident meant that sons earned 15% less over their life in the USA in the 20th century (Goldstein, 2021).

There's more evidence along all of these veins that we could draw on. But reflecting on the evidence we have collected so far, we don't think this sharpens our views of how good marginally increasing the quality of parenting is. We already believe that having a parent is important. So it isn't really surprising that losing parents is bad.

If there was much more robust evidence of the long-term wellbeing effects of orphanhood then we could use that as a sanity check on the long-term wellbeing effects of parenting programmes. For instance, if parenting programmes had a larger positive effect than orphanhood's effect would be negative, that's a sure sign of implausibility. We would also be sceptical if parenting

interventions were 50% or even 25% as good as orphanhood was bad. But based on our search so far we don't think we have the appropriate evidence. This is potentially something to look into with more time.

The effects of parenting interventions on adolescents in the short run

To further broaden the evidence we considered, and further build a prior for the effects of parenting interventions, we also estimated the short term (no apparent long-term evidence) causal effects of parenting interventions on older children and adolescent wellbeing in LMICs. The thinking is that if parenting interventions have a long term effect on very young children, it seems plausible that they should at least have a short term benefit on older children.

However, we collected this evidence before thinking more about why these parenting programmes probably work. On reflection, it seems very unclear to us that these interventions which often focus on the parent-adolescent relationship, should have much of any relationship to the effect of increasing psychosocial stimulation for very young children. These two interventions seem to probably work through pretty distinct causal mechanisms.

For the evidence, we used data from Pederson et al. (2019), a systematic review of parenting interventions (primarily with short run follow-ups) effects on the mental health of their participants. We enriched the existing information the study provided with the information necessary to perform a meta-analysis. We also excluded studies which didn't include measures of depression, anxiety, or distress. Since Pederson et al. was published a few years ago, there are probably additional studies we could collect and include, but we didn't assess that this would be worth our time as there already appeared plenty of evidence available in Pederson et al. (2019).

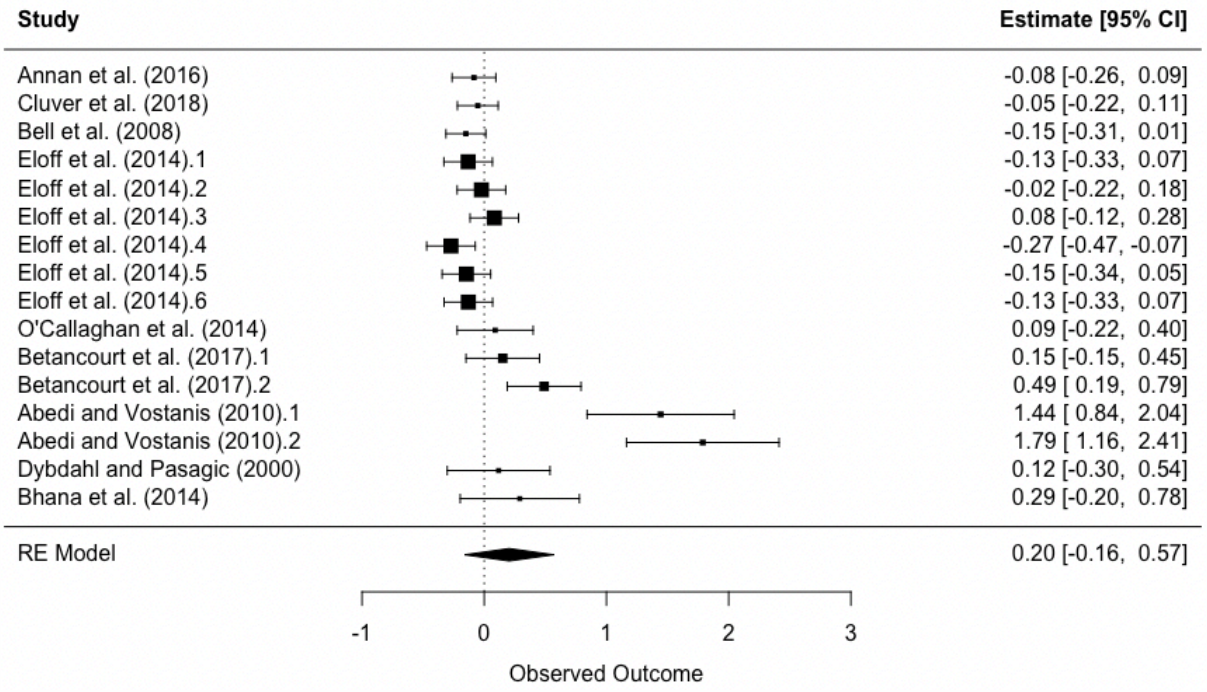
When we performed a meta-analysis of the results, the results came back non-significant, 0.20 SDs (95% CI: -0.16, 0.57; see Figure 5), with Abedi and Vostanis (2010) as a clear outlier[34]. When we remove this study's outcomes, the effect size shrinks dramatically to around 0.01 SDs (95% CI: -0.13, 0.15), which implies to us that there's probably a null effect of parenting interventions on older children and adolescents.

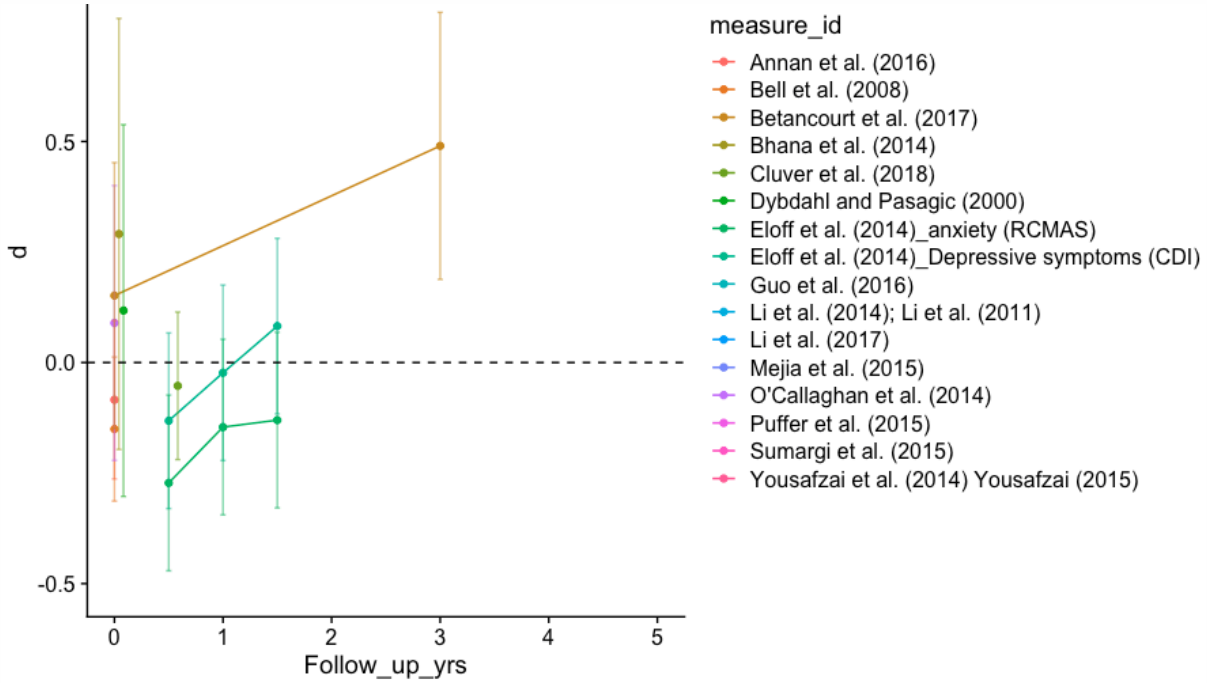
Why are these effects so small? One idea, which doesn't seem particularly good, is that the typical participants in the studies from the Pederson et al. review were young teenagers. The psychotherapy literature base makes us think that perhaps psychological interventions work less well on adolescents (From Metapsy: 0.39 SD effect for adolescents compared to 0.72 for adults).

Insomuch as this evidence informs the LR effect of parenting interventions on younger children (which we don't think it should), then we would interpret this evidence to imply a null LR effect.

Figure A1: Forest plot of SR effects of parenting interventions on MHa.

[34] This study is looking at the effect of therapy for parents of children with OCD. It seems plausible this could be why the effect is so different.





Further evidence

- Wait et al. (2022) review child maltreatment programme RCTs that include MH outcomes, but do not meta-analyse them.
- Bosqui et al. (2024) is a systematic review of parenting interventions' effects on the mental health of children and adolescents. It likely contains additional studies of parenting that can be extracted. But none are particularly well powered, RCTs, and contain self-reported child or adolescent MH or SWB outcomes. So we did not investigate further due to time constraints.

Appendix B: Other potential opportunities

The issue with parenting programmes isn't a lack of organisations delivering a plausibly efficient intervention – it's more a scarcity of clear funding opportunities with estimable costs. We discuss a few in this section. Some of these cases illustrate the difficulty in actually identifying a discrete parenting programme. In other cases there appear to be programmes, with a plausible funding opportunity, but we haven't heard back from them yet (after 2+ weeks). We didn't have a clear methodology for searching through organisations, so it's plausible we have missed a smaller organisation doing good work here. Using Google appears to direct nearly all search traffic to mega charities which provide little if any details about their programmes. With more time we would probably ask experts, and other actors in the field for recommendations.

Parenting for Lifelong Health

Parenting for lifelong health, the programme which was evaluated in the South African studies previously described and despite being featured by UNICEF, WHO, and a claim of 400,000 people reached through it or that it's in 12 countries we are unable to find an organisation claiming to actively deploy it. The organisations that mention it such as the Peace Culture Foundation in Thailand or Clowns without Borders, also appear to be referring to research programmes not instances of implementation. It's possible it's still primarily a research programme. Regardless, we haven't yet found any instances of deployment.

Save the Children, Bangladesh Parenting Programme

In GiveWell's review of psychosocial stimulation programmes, they cited one programme by Save the Children operating in Bangladesh as having an unusually low cost of $7 per person treated. While we don't have any data on this programme (and all links to the original study are broken), its very low cost means that even if it has a fraction of the long-term benefits we think Reach Up has, it could still be very cost-effective. However, Save the Children's only programme in Bangladesh that seems related is the Shishuder Jonno, which is a mega programme that uses the same deployment system (community health workers) to deliver nutritional, early child development and stimulation, education, and child rights and protections interventions across childhood. So there's no clear programme that focuses solely on childhood psychosocial stimulation.

Save the Children's "First Steps" Programme in Rwanda

Dusabe et al. (2023) provide an overview of Save the Children's scale-up of a parenting programme delivered over the radio in Rwanda. The first RCT of this programme, Abimpaye et al. (2019), showed promising results. In it caregivers attended 17 weekly education sessions delivered by radio followed by lay facilitated discussions (n =485) or the same but with three to four home visits and more resources (n = 486), both compared to a control that received nothing (n = 479). They found significant and similar development outcomes (communication, problem solving, social skills, gross and fine motor). Eye-balling it, the radio-only arm's effects seemed 80% the size of the full intervention.

A two year follow-up found that developmental outcomes were maintained (Justino et al. 2020), but only for the "full treatment" arm (0.2 SD effects for all development outcomes except fine motor, which were similar in magnitude but n.s.). The content was A/B tested within the original RCT context, to consider including positive feedback, which was found to increase maternal time investment by 0.2 SDs. The framing of the positive feedback was based on the previous RCT results, so as Dusabe et al. (2023) write "When parents heard that
their simple actions such as singing and talking were leading to improved outcomes for children, they did those actions even more". So this content was incorporated into the radio programme (Justino et al. 2023). After this, they scaled the programme up to the whole of Rwanda.

From reading the "what we do" page on Save the Children's Rwanda website, it seems like three years of scaling the programme was funded by Grand Challenges Canada for $660,865, which seems like it'd be incredibly cheap. If the First Steps programme had 1% of the impact we estimate for home visiting programmes (4.3 WELLBYs), then it'd take the radio programme affecting 310,712 children to be 20 WBp1k. Note that Rwanda has around 5,000,000 individuals below the age of 18, which naively implies around 900k below the age of three (assuming uniform distribution across ages).

However, if we had to guess, we would assume it's more than 1% as effective and more like 15% as effective. In that case, if the programme reached 100,000 children during those three years it'd have a cost-effectiveness of 23 WBp1k, and if it reached 200,000 children the cost-effectiveness would be 45 WBp1k. Both numbers seem reasonable.

Looking more deeply into this programme made us search the literature for any evidence on the mass media delivery of parenting programmes. Overall, we came away thinking that mass media delivery may actually be trickier and this means that the details of an organisation attempting to scale these interventions likely matter more than normal. We discuss the evidence in Appendix B. We found one potential organisation working to scale mass media parenting interventions, but haven't heard back from them.

Mass media delivery?

The example of First Steps in Rwanda raises the question of whether the parenting interventions we've looked at could be adapted to a cheaper medium such as radio, text message, or digital format. We think this is worth considering, but there isn't much causal evidence of such programmes in LMICs.

Solis-Cordero et al. (2022), a systematic review of remotely delivered parenting programmes reports mixed results from 4 RCTs, only 1 of them in an LMIC, and this RCT was Abimpaye et al. (2019) of the First Steps radio programme – which was only partially delivered remotely in the RCT settings.

Searching for papers published since then, we have found a few demonstrating positive results.

Most relevant to previously discussed programmes, Smith et al. (2023) study the effects of Reach-Up delivered remotely, and they report effects of 0.34 SDs on "behaviours that support child development" (n = 246), which we assume are weaker results than the in-person which report "harder" outcomes like cognitive benefits.

Lelys et al. (2023), in a large app and SMS based RCT (n = 1,113) found:

> "a 0.12 SD reduction in violence against children. Treatment children also experience fewer emotional problems (0.17 SD). When we return nine months later, we also find reductions in caregiver depression (0.12 SD), anxiety (0.16 SD), and parental stress (0.16 SD) for treatment caregivers."

The intervention is a digital adaptation of the Irie Homes Toolbox, a programme to reduce violence targeting parents of 2-to-6-year-old children (the in-person programme was studied by Francis and Baker-Henningham (2020)).

The authors report that it costs $62 per person treated. Note, if we assume the self-reported wellbeing benefits are shared across a family of four, and the effects dissipate after 9 months, then this alone would indicate a cost-effectiveness of 15 WBp1k. We emailed the authors and related organisations but have received no response.

However, it's worth noting that there are a few studies involving scaling parental interventions with technology that report negative and unexpected effects. These suggest that the details of the implementation may be quite important. We are not sure what to make of the fact that we have seen no similar study for a much larger body of evidence around in-person parenting programmes. We take it as weak evidence that SMS based interventions are more fraught for some reason. Maybe there are delicate cultural subtleties involved that a deliverer from the population is well attuned to navigate, but misunderstandings are more likely over a medium of text where there's less feedback regarding whether the recipient understands the gist of the message.

Amaral et al. (2022) studied a digital parenting intervention meant to improve parenting practices and the mental health of the caregiver in El Salvador. They found that the programme "increased stress and anxiety and lowered caregiver-child interactions among male caregivers. In contrast, among female caregivers, we did not detect changes in mental health and observed a decrease in the use of physical violence against children." Since this was driven by men with partners in situations of greater economic deprivation, one plausible story is that the intervention increased the salience of the differences in parenting done by mothers, and this had a discouraging effect.

Barrera et al. (2020) report the results of a text message experiment about parenting practices in Nicaragua. When households received messages, these had no effect on cognition, but when local opinion leaders (teachers, health workers, village heads) received texts "parental investments declined and children's outcomes deteriorated." The authors speculate that this is likely due to the information in the messages contradicting the opinion leader's views, leading them to

advocate against the advice or that it crowds out the opinion leader's advice which could be more tailor-made to the situation.

While educating parents about positive practices through cheaper remote mediums seems promising, and there are three RCTs supporting its plausibility, there are also two cases of RCT failures which provide some reason for caution. The takeaway is that while we should be careful about generalising from evidence to intervention in general, this seems like a situation where we should want very relevant evidence demonstrating success before making a recommendation.

BRAC programme from Islam et al. (2022) in Bangladesh

Islam et al. (2022) studies a large, pre-registered lay-delivered combined psychoeducation and parenting intervention (The home-based Humanitarian Play Lab or hHPL). The programme was delivered to young mothers (average age = 26 years) and their children (average age = 14 months) who were Rohingya refugees in Bangladesh. The intervention consisted of 44 weekly hour long sessions covering common psychoeducation topics as well as "highlighting the importance of childcare and psychosocial stimulation through playing for healthy development. [...] mothers also engaged in various play activities with their children and other participating mothers." (p. 10). The last 20 sessions were delivered over the phone due to COVID.

They found effects on the mothers' self-reported happiness (0.12 SDs) a month after the year-long intervention, and similar effects on depression (0.14 SDs, larger for those who were depressed 0.29 SDs). The mothers also report a 0.12 SD reduction in children's depression symptoms[35]. On developmental measures, the effects seem smaller than the Reach-Up programme (which were around 0.4 SD): "Improved communications skills of children by 0.23 SD ($p < 0.01$), gross-motor skills by 0.18 SD ($p < 0.01$), problem-solving skills by 0.18 SD ($p < 0.01$), and social skills by 0.13 SD ($p < 0.10$)." But the RCT was also at a larger scale, pre-registered, and interrupted by COVID, so we think it's far likelier these results reflect the results at scale.

Islam et al. (2022) also report "The intervention cost about $45 per dyad and is currently being scaled up in refugee camps in Bangladesh, where about 17 thousand mother-child pairs have already benefited from it."

This programme seems potentially on par with the cost-effectiveness of other programmes considered. However, as usual with large multinational NGOs, we can't find further details on the programme. At the time of writing this report, we also haven't heard back after inquiring to BRAC.

[35] The questions ask about how often they "cry for no reason", "are always sad / down / annoyed", "never gets excited" or "cannot concentrate". (p. B2).