RESEARCH

Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial

OPEN ACCESS

Orazio P Attanasio *Jeremy Bentham chair of economics*[1], Camila Fernández *senior survey researcher*[2], Emla O A Fitzsimons *professor of economics*[3], Sally M Grantham-McGregor *emerita professor of international child health*[4], Costas Meghir *Douglas A Warner III professor of economics*[5], Marta Rubio-Codina *senior research economist*[6]

[1]Department of Economics, University College London, London, UK; [2]Mathematica Policy Research, Princeton, NJ, USA; [3]Institute of Education, London WC1H 0AL, UK; [4]UCL Institute of Child Health, London, UK; [5]Department of Economics, Yale University, New Haven, CT, USA; [6]Centre for the evaluation of development policy, Institute for Fiscal Studies, London

Abstract

Objective To assess the effectiveness of an integrated early child development intervention, combining stimulation and micronutrient supplementation and delivered on a large scale in Colombia, for children's development, growth, and hemoglobin levels.

Design Cluster randomized controlled trial, using a 2×2 factorial design, with municipalities assigned to one of four groups: psychosocial stimulation, micronutrient supplementation, combined intervention, or control.

Setting 96 municipalities in Colombia, located across eight of its 32 departments.

Participants 1420 children aged 12-24 months and their primary carers.

Intervention Psychosocial stimulation (weekly home visits with play demonstrations), micronutrient sprinkles given daily, and both combined. All delivered by female community leaders for 18 months.

Main outcome measures Cognitive, receptive and expressive language, and fine and gross motor scores on the Bayley scales of infant development-III; height, weight, and hemoglobin levels measured at the baseline and end of intervention.

Results Stimulation improved cognitive scores (adjusted for age, sex, testers, and baseline levels of outcomes) by 0.26 of a standard deviation (P=0.002). Stimulation also increased receptive language by 0.22 of a standard deviation (P=0.032). Micronutrient supplementation had no significant effect on any outcome and there was no interaction between the interventions. No intervention affected height, weight, or hemoglobin levels.

Conclusions Using the infrastructure of a national welfare program we implemented the integrated early child development intervention on a large scale and showed its potential for improving children's cognitive development. We found no effect of supplementation on developmental or health outcomes. Moreover, supplementation did not interact with stimulation. The implementation model for delivering stimulation suggests that it may serve as a promising blueprint for future policy on early childhood development.

Trial registration Current Controlled trials ISRCTN18991160.

Introduction

An estimated 219 million children aged less than 5 years in developing countries are failing to reach their developmental potential.[1] Poverty and the associated poor nutrition and lack of psychosocial stimulation have been identified as major risk factors.[2] However, there is robust evidence that well targeted and well designed interventions have positive impacts on the development of vulnerable children and some evidence that benefits are sustainable.[3-6] Translating this evidence into effective policy requires information on how to scale up early childhood development programs that may reduce some of the developmental risks associated with poverty. Indeed, there are increasing demands for integrated early childhood development programs at scale that include nutrition and psychosocial

Correspondence to: E O A Fitzsimons e.fitzsimons@ioe.ac.uk

stimulation components.[7] A crucial challenge in the scale-up process is to achieve cost effectiveness of early childhood development services while maintaining their quality and attaining similar impacts on child development to those identified in efficacy studies.

One promising approach to scaling up early childhood development programs is linking them to existing social welfare systems that have established administrative capacity and local community networks.[4] This paper contributes to filling the evidence gap on scalable interventions with findings from a large scale cluster randomized controlled trial of an integrated early childhood development program, which delivered psychosocial stimulation and micronutrient supplementation to socioeconomically vulnerable families who were beneficiaries of the Familias en Acción conditional cash transfer program in Colombia. Familias en Acción is the largest national welfare program in the country; it began in 2002 and targets the poorest 20% of households ("beneficiaries"). Beneficiaries receive cash payments if they comply with two requirements: regular health check-ups for children aged less than 6 years and regular school attendance for children age more than 5 years. The effects of the program on schooling and health have been documented,[8 9] but little is known about its effects on child development outcomes.

Beneficiaries periodically elect a representative from among themselves, known as the madre líder (mother leader), who acts as a liaison between approximately 50 beneficiary families and local program officials, as and when problems arise. These mother leaders are influential and well connected in their communities. They are not paid for this role, but having been elected by fellow beneficiaries the position is considered prestigious. We tapped into this mother leaders network to deliver an early childhood development program at scale.

The effect of interventions that combine psychosocial stimulation with nutrition are not well understood. We only found four efficacy trials that examined this[10-13]; one found a synergistic interaction between stimulation and zinc supplementation on child development,[12] whereas the others found either independent effects[10 11] or no effects[13] from the addition of macronutrient supplementation to stimulation. We found no published studies at scale that examined the effects of the nutrition and stimulation components individually and combined, except a study in progress.[14]

Children living in the lowest income households in Colombia often accumulate substantial developmental delays. For instance, a recent study[15] found that significant differences in development (cognition and language) between children from different socioeconomic status emerge as early as 12 months. This gap widens with age, and by age 3 the differences between middle class children and those from lower socioeconomic classes are as large as 1 standard deviation of a standardized score.

In Colombia the incidence of early childhood anemia is high. The 2010 survey Encuesta Nacional de la Situación Nutricional (ENSIN) estimates a rate of 31% among children from the lowest socioeconomic category. However, the proportion attributable (at least in part) to iron deficiency is 19% for these children,[16] which is low compared with the 63% to 34% found in many other countries.[17] Iron deficiency may affect children's development.[2] Other micronutrient deficiencies are likely to coexist with iron deficiency and the use of multiple micronutrients is recommended.[18] Multiple micronutrients have been shown to benefit children's motor development, but evidence on mental development is less consistent.[2] Micronutrient deficiency may contribute to stunting[17]; the

prevalence of stunting in Colombia is around 15% among children aged 12-24 months.[16]

In our study we therefore included a multiple micronutrient arm and a stimulation arm to compare the independent effects of each and their combined effects on children's cognitive, language, and motor development, and their nutritional status. We hypothesized that stimulation would benefit cognitive, receptive and expressive language, and fine motor development; and that micronutrient supplementation would benefit gross motor development as well as height, weight, and hemoglobin levels. We also hypothesized that secondary outcomes such as maternal depression and household investment in the quality of the home environment would change in a way that reinforced the impact of the program on the primary outcomes.

Methods

Study design

We conducted a cluster randomized controlled trial in 96 towns in Colombia using a 2×2 factorial design with arms of psychosocial stimulation alone, micronutrient supplementation alone, both combined, and control. The intervention was rolled out across a period of four months February-May 2010, and phased out over a similar period from September-December 2011 (see supplementary appendix part I). Alongside the phase-in and phase-out period, we conducted the baseline and follow-up survey immediately before and after each phase periods. Each survey also took place over a period of four months.

Sampling frame and participants

To identify the sample, we first selected eight departments located in three geographical regions proximate to Bogotá: Cundinamarca, Boyacá, and Santander (oriental region); Antioquia, Risaralda, and Caldas (coffee zone region); and Huila and Tolima (central region). Within each of these three regions, we identified 32 municipalities (clusters) in which Familias en Acción had been in operation since its inception in 2002, and the population ranged from 2000 to 42 000 inhabitants. This latter requirement was to ensure that we had sufficient children in our sample while retaining a focus on relatively small towns. The municipalities were similar in terms of their cultures and customs. This was in order to be able to design one curriculum—and associated materials such as pictures and books—identifiable to all, which was particularly important given the ethnic and geographical diversity of Colombia.

In late 2009 we identified all mother leaders in each of the municipalities in our sample and randomly selected three of them per municipality. Each mother leader was elected by and represented approximately 50 households. We conducted door to door house listing among these beneficiary households to identify families with children in the target age range (12-24 months at enrolment). Among these families we randomly selected five to be enrolled in the study. We targeted this age range for two reasons. Firstly, we deemed it important that by the end of the intervention the children would be of an age where they would naturally be progressing on to community preschools or childcare centers. Secondly, although lacking rigorous evidence on this, we also suspected that it would be more difficult to change development in the first year of life, based on findings from Jamaica.[19]

Sample size

We computed the sample size to detect an effect of 0.33 of a standard deviation of a Bayley scale on infant development[20] for either the stimulation only group or the micronutrient supplementation only, against the control group. Similar interventions attained this effect size in efficacy studies in Bangladesh.[21] This sample provided 80% power and 5% significance level, allowing for an attrition rate of 10% with 24 villages per intervention and assuming an intracluster correlation of 0.09 in the outcome. This level of intracluster correlation had been estimated from a sample of rural Mexican children who were part of the evaluation of a conditional cash transfer program. It turned out to be conservative as we had an intracluster correlation of 0.04, conditional on observables, in our baseline sample.

Randomisation

Within each region we used computer generated codes to randomly assign eight clusters each to psychosocial stimulation, micronutrient supplementation, both combined, and control. To do this we used version 9.0 of the data analysis and statistical software STATA (StataCorp, College Station, TX).

Blinding

It was not possible to blind study participants to their allocation to stimulation and, for ethical reasons, we did not use a placebo for micronutrients. However, testers and interviewers were blind to the treatment status of participants.

Interventions

Our intervention lasted 18 months, which was the maximum period our budget allowed while ensuring that by the end of the intervention, children would reach an age when they would naturally "graduate" into the existing community care services. Previous studies found sustainable benefits from similar interventions lasting from nine months to three years.[3 22-24]

The psychosocial stimulation program was based on the Jamaican home visiting model, which demonstrated positive short term and long term effects.[3 6 11 24] We adapted the curriculum and materials to the Colombian sociocultural context. The curriculum was delivered by home visitors, who were hired by us on a part time basis.

Home visitors made weekly home visits where they demonstrated play activities using low cost or homemade toys, picture books, and form boards. These materials were left in the homes for the week after the visit and were changed weekly. The aims of the visits were to improve the quality of maternal-child interactions and to assist mothers to participate in developmentally appropriate learning activities, many centered on daily routines.

The micronutrient supplementation consisted of Sprinkles (Hexagon Nutrition, Mumbai, India)—encapsulated micronutrients in powder form—developed to treat childhood anemia.[25] Their efficacy, bioavailability, and safety have been shown in diverse settings.[25-30] Each single dose sachet contains 12.5 mg iron, 5 mg zinc, vitamin A 300 μg retinol equivalents, 160 μg folic acid, and 30 mg vitamin C. Sachets, which include a pictorial representation of use, were delivered to households every two weeks. At the beginning of the study, home visitors also provided mothers with a monitoring booklet with use and storage instructions and forms for recording intake, which were checked by home visitors regularly. Households were provided with enough sachets for all children aged less than 6 years to prevent reduction of dose from sharing and for ethical reasons.

Strategy for large scale implementation: program staff, training, and monitoring

Home visitors were selected from among the mother leaders. In municipalities assigned to receive stimulation, 63% of selected mother leaders took on the role of home visitor. Of the remaining 37%, most declined owing to other work commitments and a few were deemed unsuitable as a result of their literacy levels being too low to deliver the curriculum (as determined from a short reading comprehension test designed and administered by us). In these cases, while retaining the beneficiary children to ensure comparability between treatment and control samples, the mother leader helped us identify a replacement from the community. We used this referral procedure because it would be easy to bring to scale.

Six mentors with an undergraduate degree in psychology or social work—or fieldwork experience with families and children—were selected to train and supervise the home visitors. Mentors underwent six weeks' pre-service training on the home visiting curriculum and protocols, training and supervision skills, and conducting home visits, including supervised practice. Each mentor trained and supervised 24 home visitors from eight municipalities. The home visitors' pre-service training lasted two weeks, with an additional week of training one to two months after the program began. The mentors visited communities once every 7-10 weeks to monitor the fidelity of the implementation, provide support, and sustain home visitors' motivation. During this visit they also distributed one page bulletins to home visitors, with reminders of best practices in home visiting. In addition, mentors sent short text messages to home visitors every month, which reinforced key advice such as "listen to the mother and praise her." Home visitors were also encouraged to call mentors (at our expense) for advice where necessary.

Outcomes and measurements

Primary outcomes

We used the Bayley scales of infant and toddler development, third edition (Bayley-III). The Bayley-III subscales were translated into Spanish, back translated to English to ensure accuracy, and piloted by testers. Following standard procedures, we administered five subscales of the Bayley-III[20]: cognition, receptive language, expressive language, fine motor, and gross motor. Children were assessed in local community centers with their mothers present. Testers held degrees in psychology and had six weeks' training, including practice sessions with children of the target age groups. Inter-rater reliability (intracluster correlation) was above 0.9 on each subscale. In a concurrent study,[15] test-retest reliability on the Spanish version of the Bayley-III was 0.95 to 0.98 (intracluster correlation, n=20, median time between tests 8 (range 2-11) days).

In the analysis, we use raw scores controlling for sex and age (second order polynomial) rather than using the composite scores, which are standardized with a representative sample of US children (reference population) and may not be appropriate in Colombia. Moreover, using raw scores allowed us to examine each subscale separately, and in particular to keep fine and gross motor scores separate (since we did not hypothesize that stimulation would affect gross motor scores). In part II of the supplementary appendix, however, we report results on the composite scores.

Hemoglobin level was assessed using Hemocue Hb 201 microcuvette machines (distributed by Vital Tecnologia, Colombia). Weight was measured using a scale (SECA 872) accurate to 0.1 kg, and length (height at post-intervention) was measured using wooden length boards (ShorrBoards; Proveedor Shorr Productions, MD) accurate to 1 mm. Stunting was computed following World Health Organization standards and software (WHO Anthro version 3.2.2).

Secondary outcomes

We consider as secondary outcomes those that could mediate the effect of the intervention on primary outcomes: they included maternal depression, measured using the Spanish translation of the Center for Epidemiologic Studies short depression scale (CES-D 10),[31-33] and household investment in the quality of the home environment. This was measured using the number of varieties of play materials in the home that the child often played with and the number of play activities the child engaged in with an adult over the three days before the interview, from UNICEF's family care indicator.[34] We previously reported that stimulation increased substantially the varieties of play materials and play activities (0.53 SD and 0.54 SD, both P<0.01) as measured by the family care indicator. These results are shown in part II of the supplementary appendix and elsewhere.[35] Our final outcome measure was children's consumption of iron-rich food, based on maternal reports. In particular, respondents were asked how many times in the previous week the target child had eaten certain items.

Socioeconomic variables

We surveyed the mother or head of household in the home to obtain information on socioeconomic characteristics. These included age, education level, and employment status of all household members, as well as household assets and expenditures.

All measures were collected at baseline and follow-up. The primary caregivers of all children provided written informed consent before enrolment.

Statistical analyses

We considered two sets of primary outcomes: the first included the four Bayley subscales (cognition, expressive language, receptive language, and fine motor) hypothesized to be impacted by both interventions; the second referred to physical outcomes (height, weight, hemoglobin levels, and the Bayley gross motor subscale) hypothesized to be affected by the nutritional intervention.

Using multiple regression analysis, we examined the treatment effect of stimulation and supplementation, and their interactions on our primary outcomes (intention to treat analyses). All statistical inferences controlled for cluster effects at the municipality level and for the stratification (by region) in the randomization procedure. We report the coefficient of the effect of stimulation, supplementation, and of the interaction between the two interventions. To obtain these effects we regressed the outcome on an indicator that takes the value of 1 if any stimulation (alone or in combination), an indicator that takes the value of 1 if any supplementation (alone or in combination), and an indicator that takes the value of 1 if both stimulation and supplementation. We controlled for tester effects (two for each region), baseline level of the outcomes, sex, and a second order polynomial in age, which was chosen considering the fit of the data.

We report means and 95% confidence intervals for the treatment effects, with the confidence intervals computed by bootstrap methods. The P values reported refer to one tailed tests for the impacts of the two interventions (reflecting the presumption that these interventions could not harm the children) and two tailed tests for the interaction term (reflecting the possibility of negative interactions between the two interventions). We computed these using the Romano and Wolf[36 37] step-down procedure on each of the two sets of primary outcomes we considered, to take into account that we were testing multiple hypotheses. We considered a P value of less than 0.05 to be statistically significant for the main effect. We used STATA version 12.1 for statistical analyses and the Matlab algorithm written by D Wunderli of the University of Zurich to control for multiple testing by the Romano and Wolf step-down procedure.

We also carried out two additional checks. Firstly, we aggregated all outcomes listed in table 3, after standardizing them, and estimated the effect on this aggregate, which we refer to as the aggregate index. Secondly, we conducted factor analysis on all outcomes in table 3 and estimated effects on this factor index. In both cases the coefficients were already expressed in effect sizes (standard deviations). Moreover, there was no need to adjust P values since there was only one aggregated outcome.

Results

At baseline we administered the Bayley-III to 1420 children and the survey to 1429 households (figure⇓). We excluded from analyses two children who scored more than 3 SDs below the mean on the Bayley-III cognitive subscale at baseline (possible disability) and three children who scored more than 4 SDs below the mean on the same outcome at follow-up (extreme observations). The attrition rate between baseline and follow-up for the Bayley-III sample was 10.7% (n=151) across treatment arms: 36 (11%) of the children from the stimulation arm were not measured at follow-up, 46 (15%) from the supplementation arm, 39 (12%) from the combined arm, and 30 (9%) from the control arm. The difference in loss among the groups was not statistically significant. However, we also computed the impacts using weighted regressions to take into account sample loss. The results of this exercise, which are not different from those we report here, are in supplementary tables A1 and A2 (part II).

Compliance was high and the average number of home visits made was 63 (81% of those scheduled). An average of 396 sprinkles (73% of those scheduled) was recorded as having been given by the mothers.

Baseline characteristics

Table 1⇓ shows the characteristics of children who remained in the study at follow-up, their mothers, and their households. No differences were apparent between the groups, with the exception of the supplementation group, which had lower levels of stunting and lower proportions of single mothers but a higher proportion of divorced mothers than the control group. We controlled for these initial measures, and the results remained similar (see supplementary tables A3 and A4 in part II).

The age of children in the control group was an average of 18.27 (SD 4.02) months. The prevalence of anemia was 46% and of stunting was 16%. The mean age of the mothers was 27.63 (SD 6.96) years and on average they had 7.70 (SD 3.51) years of education, which corresponds to having completed more than primary school.

The children's Bayley-III scores were well balanced across groups, as were height, weight, and hemoglobin levels (table 2⇓).

Impacts

Table 3⇓ shows that the estimated treatment effects on child development as measured using the four indicators we hypothesized could have been affected by the stimulation intervention: the Bayley-III cognitive, receptive language, expressive language, and fine motor subscales. For each impact we report its estimate, bootstrapped 95% confidence interval, and effect size. The reported P values, however, take into account that we tested 12 separate hypotheses (four outcomes for each of three treatments) using the Romano and Wolf procedure.

The effect of stimulation on Bayley-III cognition was 0.260 SDs (P=0.002, table 3). We did not detect any significant effects of supplementation on cognition and no significant interaction with stimulation (P>0.5 in both cases).

Stimulation benefited receptive language by 0.218 SDs (P=0.032, table 3). There were no significant effects from supplementation and no significant interactions between treatments on any outcome.

None of the treatments had a significant effect on expressive language or fine motor skills, with most of the P values being larger than 0.5. Some of the coefficients would be marginally significant when considered in isolation (such as that of supplementation on fine motor skill, with a non-reported individual P=0.085). However, when controlling for multiple hypotheses testing, none of the estimated impacts were significantly different from zero. Results for the aggregate index and factor index were qualitatively similar, even if the effect sizes were smaller, as expected. Similarly, the effects were robust to the use of composite scores (see supplementary table A5).

Table 4⇓ reports effects of the interventions on the set of primary outcomes that we hypothesized would be impacted by micronutrient supplementation: hemoglobin levels, height, weight, and gross motor skills. None of the effects on these outcomes was estimated to be significantly different from zero.

Discussion

In a cluster randomized controlled trial at scale, we found that stimulation of children aged 12-24 months and delivered through home visits resulted in a significant improvement of 0.26 SDs in cognitive scores and 0.22 SDs in receptive language. We previously reported that stimulation also improved the amount of stimulation (play activities and play materials) being provided by parents in the home.[35] In contrast, micronutrient supplementation had no significant effect on the Bayley-III scores and there was no significant interaction between the interventions. None of the treatments had a significant effect on maternal depression or on children's consumption of iron-rich food. Most importantly, the study showed that stimulation could be delivered effectively at scale as part of an existing welfare program.

Innovation

Earlier efficacy studies[2] have established that stimulation can substantially improve the cognitive development of infants in low income and middle income countries. Compared to these studies, we innovated in two ways. Firstly, we mimicked the workings of a scaled up program by using the infrastructure of a national welfare program to access the communities and identify local home visitors and families, and by implementing the program across 96 municipalities, spread across a large scale—eight of the 32 departments of Colombia, covering an area three times the size of England. Secondly, we looked at the effects of micronutrient supplementation and stimulation interventions separately and combined, in the context of a scale-up intervention.

Though the trial was not fully at scale, it was designed to be scalable. The intervention was spread over 96 towns in three regions and used mother leaders who are readily available nationally. While considerable expertise was involved in the adaptation of the curriculum, and some of the authors took part in the training of the mentors, the profile of the mentors and mother leaders was chosen so that the program could be scaled up at the national level in Colombia, or indeed in other countries.

Relevance

By emphasizing parenting in the intervention and by showing that parental behavior changed and home stimulation increased, there is hope that the benefits will be sustainable and will translate into higher schooling, economic benefits, and wellbeing in the longer term. Studies in economics have shown a direct relation between cognition as children and earnings in later life.[38 39] Recent studies have shown that early cognitive gains from a similar intervention, on which ours is based, were sustained to adulthood[3 24] and have translated into educational attainment and earnings.[6]

An important question is whether the size of the effects of the stimulation arm on cognition is relevant from a developmental and economic standpoint. An effect of 0.26 SDs is usually considered small to moderate. To examine the clinical importance of this finding, we used data on 203 young adults drawn from the national longitudinal survey of youth 1979 (NLSY79), a survey of the biological children of women in the NLSY79 obtained from the US Bureau of Labor Statistics. For this sample we have both an early childhood developmental test as well as an ability score for the mother. From this we have estimated that an increase in the Peabody picture vocabulary test scores at age 4 years of 0.25 of a standard deviation is associated with a 7.5% increase in income at age 30 (P=0.057), once we control for maternal ability scores. This gain in income is equivalent to the gain from one extra year of schooling. Therefore if sustained the cognitive improvement is likely to be important in the longer term. Also in relation to clinical significance, a separate study[15] of a sample representative of low income and middle income households in Bogota, initiated after the beginning of this trial, showed that the average difference in children's cognition (Bayley-III) between families in the top and bottom fourths of wealth, was 0.812 SDs at ages 31-42 months. Average cognitive development among those in the bottom fourth in the Bogota sample is similar to that of our control group. It is therefore reasonable to conclude that our stimulation arm fills almost a third of the gap in cognition between children living in an extremely disadvantaged environment and children living in middle class families in Bogota. Hence if sustained the improvement we recorded makes the longer run cognitive development of these children closer to those living in families that are substantially better off.

Recent studies (such as[40]) have stressed that the high levels of income inequality in Colombia, and more generally Latin America, are strictly linked to inequality in education achievement, which in turn is related to the gaps in early development that our intervention has the potential to fill.

Limitations and possible improvements

This trial was the first attempt to take a stimulation intervention to scale, and subsequent to this experience we could almost certainly increase the impact in future trials. Possible improvements include hiring home visitors full time instead of part time, thus requiring fewer overall, and employing mentors who live near the intervention sites. The former improvement would imply having home visitors more dedicated to their job and would allow better selection from the same pool of local women as well as more personalized attention and training by their mentors; the mentors would allow more frequent and effective supervision and feedback to the home visitors, while at the same time reducing travel time and costs.

There were no significant benefits from supplementation. Although it might be surprising that micronutrient supplementation did not improve hemoglobin levels, given that mothers reported that they gave sprinkles regularly, there are several explanations. Firstly, it is possible that the reports about compliance were not accurate. Secondly, it is possible that the high levels of anemia observed at baseline were not linked to micronutrient deficiency but to other factors such as chronic infections and hemoglobinopathies. Limitations to the study are that we had no measure of iron status and did not use a placebo for micronutrients.

Future research

Important research questions remain to be addressed. Firstly, follow-up studies of our sample are needed to establish the sustainability of the impacts and the factors that may contribute to sustaining cognitive gains. They will also establish whether the effects of our intervention enable children to gain more from future opportunities and interventions, such as performance in preschools and schools. Secondly, more information is needed on the most effective age to enroll a child and what duration is necessary to achieve sustained benefits. Thirdly, we need to develop and test an approach for fitting such early childhood development programs into the continuum of services from early childhood to school age, and possibly beyond. In this context, other models for delivering early childhood development programs such as mothers' groups could also be explored.

Finally, we need to understand how programs interact with parental behavior and affect other inputs into child development. Our results suggest that the effects of the stimulation program may be sustainable in the long run because the program induces positive change in parental behavior. Understanding the ways in which programs influence the larger community are also of interest, as good parenting practices may be disseminated more widely in communities in the longer term.

Conclusions

This study offers important new evidence that child stimulation programs delivered through the local community can be delivered effectively on a large scale and at a low cost. The program cost about $500 per year per child, which could be further reduced at scale, and which is considerably lower than the $1300 per year per child the Colombian government has budgeted for some of its flagship programs (such as the construction of large centers targeted at children aged between 6 and 60 months) that are an integral part of its early childhood policy.

1   Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B. Developmental potential in the first 5 years for children in developing countries. *Lancet* 2007;369:60-70.
2   Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, et al. Inequality in early childhood: risk and protective factors for early child development. *Lancet* 2011;378:1325-38.
3   Walker SP, Chang SM, Vera-Hernandez CM, Grantham-McGregor S. Early childhood stimulation benefits adult competence and reduces violent behaviour. *Pediatrics* 2011;127:849-57.
4   Engle P, Fernald L, Alderman H, Behrman J, O'Gara C, Yousafzai A, et al. Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *Lancet* 2011;378:1339-59.
5   Hoddinott J, Maluccio JA, Behrman JR, Flores R, Martorell R. Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. *Lancet* 2008;371:411-6.
6   Gertler PJ, Heckman JJ, Zanolini A, Pinto R, Zanolini A, Vermeersch C, et al. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science* 2014;344:998-1001.
7   Grantham-McGregor SM, Fernald LCH, Kagawa RMC, Walker S. Effects of integrated child development and nutrition interventions on child development and nutritional status. *Ann N Y Acad Sci* 2014;1308:11-32.
8   Attanasio O, Fitzsimons E, Gomez A, Rodriguez MI, Meghir C, Mesnard A. Children's education and work in the presence of a conditional cash transfer program in rural Colombia. *Econ Dev Cult Change* 2010;58:181-210.
9   Attanasio O, Vera-Hernandez M, Gomez LC, Rojas AG. Child health in rural Colombia: determinants and policy interventions.. *Econ Human Biol* 2004;2:411-38.
10   Waber DP, Vuori-Christiansen L, Ortiz N, Clement JR, Christiansen NE, Mora JO, et al. Nutritional supplementation, maternal education, and cognitive development of infants at risk of malnutrition. *Am J Clin Nutr* 1981;34:807-13.
11   Grantham-McGregor S, Powell C, Walker S, Himes J. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. *Lancet* 1991;338:1-5.
12   Gardner JM, Powell CA, Baker-Henningham H, Walker SP, Cole TJ, Grantham-McGregor SM. Zinc supplementation and psychosocial stimulation: effects on the development of undernourished Jamaican children. *Am J Clin Nutr* 2005;82:399-405.
13   Nahar B, Hossain MI, Hamadani JD, Ahmed T, Huda SN, Grantham-McGregor SM, et al. Effects of a community-based approach of food and psychosocial stimulation on growth and development of severely malnourished children in Bangladesh: a randomised trial. *Eur J Clin Nutr* 2012;66:701-9.
14   Yousafzai AK, Aboud F. Review of implementation processes for integrated nutrition and psychosocial stimulation interventions. *Ann N Y Acad Sci* 2014;1308:33-45.
15   Rubio-Codina M, Attanasio O, Meghir C, Varela N, Grantham-McGregor S. The socio-economic gradient of child development: cross-sectional evidence from children 6-42 months in Bogota. J Human Resources 2014 (in press).
16   Instituto Colombiano de Bienestar Familiar. Encuesta nacional de la situación nutricional (ENSIN) en Colombia 2010. Bogota; 2011.
17   Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet* 2013;382:427-51.
18   Allen LH, Peerson JM, Olney DK. Provision of multiple rather than two or fewer micronutrients more effectively improves growth and other outcomes in micronutrient-deficient children and adults. *J Nutr* 2009;139:1022-30.
19   Powell C, Grantham-McGregor SM. The effect of home visiting of varying frequency on child development. *Pediatrics* 1989;84:157-64.
20   Bayley N. Bayley scales of infant and toddler development. 3rd ed. Pearson Education, 2006.

What is already known on this topic

An extensive and growing literature has shown that environmental factors during the first few years of life have long lasting consequences on a variety of developmental outcomes, and thus that experiences in these formative years are key for subsequent development

Several interventions targeted at vulnerable children in developed and developing countries, based on stimulation (and, at times, nutrition) have been shown to be effective in obtaining sustainable improvement in developmental outcomes

These interventions, however, have been conducted on a relatively small scale and could be classified as efficacy studies

What this study adds

This study evaluated an integrated early childhood intervention combining stimulation and micronutrient supplementation, targeted at young children living in the poorest fifth of households in Colombia, in a 2×2 design, using a large cluster randomized controlled trial

The intervention was designed to be scalable in terms of both costs and implementation, by using local community based human resources and the infrastructure of an existing national welfare program

The stimulation arm of the intervention had significant impacts on cognition and receptive language, while supplementation had no effects and there were no interactions between stimulation and supplementation

We have shown that successful interventions can be implemented with limited resources on a large scale and have identified aspects of the implementation that could be improved

21 Hamadani JD, Huda SN, Khatun F, Grantham-McGregor SM. Psychosocial stimulation improves the development of undernourished children in rural Bangladesh. *J Nutr* 2006;136:2645-52.

22 Grantham-McGregor SM, Powell C, Walker S, Chang S, Fletcher P. The long term follow-up of severely malnourished children who participated in an intervention programme. *Child Dev* 1994;65:428-39.

23 Walker SP, Chang SM, Younger N, Grantham-McGregor SM. The effect of psychosocial stimulation on cognition and receptive language in a cohort of term, low-birthweight Jamaican children. *Dev Med Child Neurol* 2010;52:e148-54.

24 Walker SP, Chang SM, Powell C, Simonoff E, Grantham-McGregor S. Effects of psychosocial stimulation and dietary supplementation in early childhood on psychosocial functioning in late adolescence: follow-up of randomised controlled trial. *BMJ* 2006;333:472.

25 Zlotkin SH, Schauer C, Christofides A, Sharieff W, Tondeur MC, Hyder SMZ. Micronutrient sprinkles to control childhood anaemia. *PLoS Med* 2005;2:e1.

26 Zlotkin SH, Arthur P, Antwi KY, Yeung G. Treatment of anemia with microencapsulated ferrous fumarate plus ascorbic acid supplied as 'sprinkles' added to complementary (weaning) foods. *Am J Clin Nutr* 2001;74:791-5.

27 Schauer C, Zlotkin SH. "Home-fortification" with micronutrient sprinkles—a new approach for the prevention and treatment of nutritional anemias. *Paediatr Child Health* 2003;8:87-90.

28 Tondeur MC, Schauer CS, Christofides AI, Asante KP, Newton S, Serfass RE, et al. Determination of iron absorption from intrinsically labeled microencapsulated ferrous fumarate (sprinkles) in infants with different iron and hematologic status by using a dual-stable-isotope method. *Am J Clin Nutr* 2004;80:1436-44.

29 Giovannini M, Sala D, Usuelli M, Livio L, Francescato G, Braga M, et al. Double-blind, placebo-controlled trial comparing effects of supplementation with two different combinations of micronutrients delivered as sprinkles on growth, anaemia, and iron deficiency in Cambodian infants. *J Pediatr Gastroenterol Nutr* 2006;42:306-12.

30 Sharieff W, Yin SA, Wu M, Yang Q, Schauer C, Tomlinson G, et al. Short-term daily or weekly administration of micronutrient sprinkles has high compliance and does not cause iron overload in Chinese schoolchildren: a cluster-randomized trial, *Public Health Nutr* 2006;9:336-44.

31 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385-401.

32 Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med* 1994;10:77-84.

33 González V, Stewart A, Ritter P, Lorig K. Translation and validation of arthritis outcome measures into Spanish. *Arthritis Rheum* 1995;38:1429-46.

34 Frongillo EA, Sywulka SM, Kariger P. UNICEF psychosocial care indicators project. final report to UNICEF 2003. Cornell University.

35 Attanasio O, Fernández C, Fitzsimons E, Grantham-McGregor S, Meghir C, Rubio-Codina M. Enriching the home environment of low-income families in Colombia: a strategy to promote child development at scale. Early Childhood Matters. Van Leer Foundation, 2013.

36 Romano JP, Wolf M. Stepwise multiple testing as formalized data snooping. *Econometrica* 2005;73:1237-82.

37 Romano JP, Wolf M. Exact and approximate stepdown methods for multiple hypothesis testing. *J Am Stat Assoc* 2005;100:94-108.

38 Cunha F, Heckman J. Estimating the technology of cognitive and noncognitive skill formation: the linear case. In: Molenaar PCM, Lerner RM, Newell KM, eds. Handbook of developmental systems theory and methodology. Guilford Press, 2013:221-69.

39 Cunha F, Heckman J, Schennach S. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 2010;78:883-93.

40 Azevedo VMR, Bouillon C. Intergenerational social mobility in Latin America: a review of existing evidence. *Econ Anal Rev* 2010;25:7-42.

Tables

Table 1| Baseline characteristics of participant children, their mothers, and their households by randomization status. Values are percentages unless stated otherwise

Characteristics	Control (n=318)	Stimulation (n=318)	Supplementation (n=308)	Both interventions (n=319)
Children:				
Mean (SD) age (months)	18.27 (4.02)	18.07 (3.76)	17.96 (3.60)	18.01 (3.73)
Boys	50	47	54	51
Premature	19	14	18	12
Mean (SD) birth weight (g)	3222.48 (554.20)	3266.94 (476.39)	3244.75 (499.36)	3247.15 (514.63)
Stunted: Z score height for age <–2 SD	16	14	10	14
Anemic	46	47	46	45
First born	42	36	42	36
Mothers:				
Mean (SD) age (years)	27.63 (6.96)	28.34 (6.95)	27.50 (6.23)	27.92 (6.55)
Mean (SD) education (years)	7.70 (3.51)	7.21 (3.41)	7.41 (3.53)	7.48 (3.43)
Married	69	70	70	66
Divorced	9	11	17	13
Single	23	18	13	21
Mean (SD) depression score*	9.43 (5.54)	8.38 (5.60)	9.51 (5.47)	8.82 (5.24)
Household:				
Mean (SD) household size	5.22 (2.20)	5.38 (2.31)	5.23 (2.15)	5.22 (2.17)
Mean (SD) crowding (No of rooms over household size)	0.60 (0.30)	0.57 (0.29)	0.59 (0.27)	0.62 (0.32)
Home ownership	34	39	40	36
Mean (SD) wealth index†	–0.08 (0.92)	0.04 (0.98)	0.07 (1.06)	0.03 (1.04)
Home environment:				
Mean (SD) No of varieties of play materials‡	3.34 (1.59)	3.41 (1.51)	3.19 (1.61)	3.10 (1.47)
Mean (SD) No of varieties of play activities over past 3 days§	3.69 (1.76)	3.70 (1.71)	3.71 (1.65)	3.62 (1.67)

*Center for Epidemiologic Studies short depression scale (CES-D 10). Scores range from 1 to 30; with a score greater than 10 being considered depressed using the reference population norms.

†First principal component of household asset and characteristics: dirt floor, solid walls, crowding index, home ownership, sewage, and ownership of car, computer, blender, fridge, washing machine, and cellphone.

‡Toys that make or play music; toys or objects meant for stacking, constructing or building; things for drawing, writing, coloring, and painting; toys for moving around; toys to play pretend games; picture books and drawing books for children; and toys for learning shapes and colors.

§Reading books or looking at picture books; telling stories to child; singing songs with child; taking child outside home place or going for a walk; playing with child with toys; spending time with child scribbling, drawing, or coloring; and spending time with child naming things or counting.

Table 2| Outcomes of participant children at baseline and end of intervention by randomization status

Outcomes	Mean (SD)			
	Control (n=318)	Stimulation (n=318)	Supplementation (n=308)	Both interventions (n=319)
Baseline				
Bayley-III raw scores:				
Cognition	52.21 (7.72)	51.76 (7.28)	51.78 (7.37)	51.83 (7.41)
Receptive language	20.57 (4.99)	20.51 (5.00)	20.20 (5.23)	19.97 (4.77)
Expressive language	20.48 (6.30)	20.57 (6.69)	19.96 (6.18)	19.97 (6.14)
Fine motor	34.85 (3.96)	34.64 (4.01)	34.29 (4.07)	34.08 (3.89)
Gross motor	50.52 (6.91)	51.11 (7.19)	50.54 (6.34)	50.23 (6.90)
Hemoglobin (g/L)	114.64 (13.05)	112.91 (12.86)	113.71 (12.77)	113.81 (13.05)
Height (cm)	78.90 (4.66)	79.07 (4.40)	79.48 (4.37)	78.96 (4.51)
Weight in (kg)	10.35 (1.43)	10.42 (1.48)	10.43 (1.35)	10.39 (1.49)
End of intervention				
Bayley-III raw scores:				
Cognition	71.68 (4.38)	72.74* (4.31)	71.63 (4.26)	72.43 (4.27)
Receptive language	34.06 (3.55)	34.81* (3.56)	34.06 (3.73)	34.48 (3.43)
Expressive language	35.97 (5.43)	36.32 (5.70)	36.13 (5.85)	36.10 (5.81)
Fine motor	50.38 (4.65)	50.77 (4.45)	50.52 (4.56)	50.59 (4.50)
Gross motor	63.31 (2.79)	63.30 (2.77)	63.19 (2.99)	62.97 (3.08)
Hemoglobin (g/L)	121.18 (12.03)	121.55 (11.76)	120.54 (12.03)	122.49 (11.33)
Height (cm)	92.79 (4.20)	93.11 (4.39)	93.24 (4.37)	92.82 (4.37)
Weight (kg)	14.04 (1.90)	14.07 (1.80)	14.19 (1.80)	14.07 (1.90)

*P<0.05 for difference with respect to control group; P values for difference in means adjusted for clustering standard errors at municipality level.

Table 3| Estimated effects of intervention on cognition, language, and fine motor scores, and on aggregate scores in participants (n=1263)

	Stimulation			Supplementation			Stimulation×supplementation		
	β (95% CI)	P value*	D†	β (95% CI)	P value*	D†	β (95% CI)	P value*	D†
Bayley-III scores‡:									
Cognition	1.139 (0.538 to 1.776)	0.002	0.260	0.196 (−0.294 to 0.676)	>0.50	0.045	−0.352 (−1.254 to 0.424)	>0.50	−0.080
Receptive language	0.776 (0.270 to 1.332)	0.032	0.218	0.128 (−0.458 to 0.706)	>0.50	0.036	−0.330 (−1.138 to 0.384)	>0.50	−0.093
Expressive language	0.455 (−0.286 to 1.250)	>0.50	0.084	0.403 (−0.580 to 1.269)	>0.50	0.074	−0.375 (−1.741 to 0.803)	>0.50	−0.060
Fine motor§	0.567 (−0.060 to 1.247)	0.34	0.122	0.455 (−0.184 to 1.102)	>0.50	0.098	−0.492 (−1.444 to 0.308)	>0.50	−0.106
Indexes:									
Aggregate index (Z score)¶	0.182 (0.050 to 0.314)	0.007	0.182	0.091 (−0.038 to 0.219)	0.165	0.091	−0.116 (−0.303 to 0.072)	0.223	−0.116
Factor index (Z score)**	0.217 (0.075 to 0.359)	0.003	0.217	0.075 (−0.061 to 0.210)	0.276	0.075	−0.091 (−0.286 to 0.104)	0.357	−0.091

Each row represents a separate regression. βs are estimated intervention effects (stimulation, supplementation) and interaction (stimulation×supplementation), controlling for child's sex; age and age squared in months at time of measurement; level of cognition, receptive language, expressive language and fine motor development at baseline; and tester dummies. Standard errors (not shown) are computed by bootstrap methods (5000 replications) and adjusted for clustering at municipality level and stratification in randomization.

*P values are computed using Romano Wolf step-down procedure to take into account multiple hypotheses testing (one tailed tests for impact of interventions, two tailed test for interaction term). In this table 12 hypotheses are considered, corresponding to first four rows.

†D=(β/SD controls), where SD controls is standard deviation for control group within estimation sample.

‡Expressed in raw scores.

§Fine motor score missing for one child.

¶Index that adds up standardized Bayley-III cognitive, receptive language, expressive language, and fine motor scores. Standard P values. Coefficients already expressed in standard deviations.

**Standardized factor index constructed by conducting factor analysis on Bayley-III cognitive, receptive language, expressive language, and fine motor raw scores. Standard P values. Coefficients already expressed in standard deviations.

Table 4| Estimated effects of intervention on children's hemoglobin levels, height, weight, and gross motor scores

| Variables	No	Stimulation				Supplementation				Stimulation×supplementation		
		β (95% CI)	P value*	D†		β (95% CI)	P value*	D†		β (95% CI)	P value*	D†
Hemoglobin (g/L)	1208	0.178 (−0.036 to 0.392)	>0.50	0.148		0.048 (−0.181 to 0.264)	>0.50	0.039		−0.014 (−0.317 to 0.294)	>0.50	−0.012
Height (cm)	1231	0.095 (−0.328 to 0.518)	>0.50	0.022		−0.177 (−0.635 to 0.295)	>0.50	−0.042		0.084 (−0.564 to 0.748)	>0.50	0.020
Weight (kg)	1233	−0.023 (−0.164 to 0.118)	>0.50	−0.012		−0.012 (−0.175 to 0.158)	>0.50	−0.007		0.021 (−0.197 to 0.248)	>0.50	0.011
Bayley-III gross motor score‡	1236	−0.070 (−0.379 to 0.241)	>0.50	−0.025		−0.097 (−0.427 to 0.244)	>0.50	−0.034		−0.064 (−0.592 to 0.442)	>0.50	−0.023

Each row represents a separate regression. βs are estimated intervention effects (stimulation, supplementation) and interaction (stimulation×supplementation), controlling for child's sex; age and age squared in months at time of measurement; baseline level of hemoglobin, height, weight, and gross motor development; and tester dummies. Hemoglobin regression additionally controls for altitude and regional dummies. Standard errors (not shown) are computed by bootstrap methods (5000 replications) and adjusted for clustering at municipality level and stratification in randomization.

*P values are computed using the Romano Wolf stepdown procedure to take into account multiple hypotheses testing (one tailed tests for impacts of interventions, two tailed test for interaction term). In this table 12 hypotheses are considered.

†D=(β/SD controls), where SD controls is standard deviation for control group within estimation sample.

‡Expressed in raw scores.

Figure
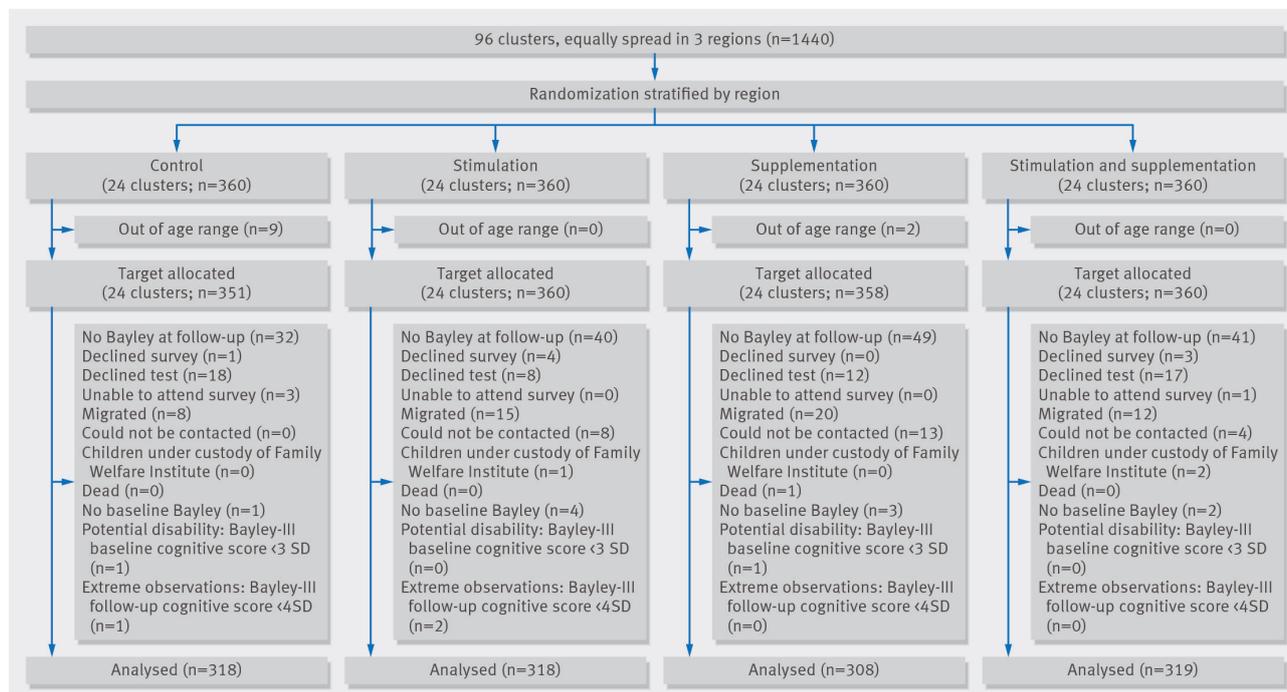
Flow of participants through study