Effects of an early stimulation curriculum and a nutritional intervention in a national public parenting program in Colombia on child development and growth: a cluster-randomized controlled trial

Orazio Attanasio
Raquel Bernal
Helen Baker-Henningham
Costas Meghir
Marta Rubio-Codina

First version: 27/08/2016
This version: 19/04/2017

Abstract

In this paper, we evaluate the effects of the implementation of a structured early stimulation curriculum and a nutritional intervention through public parenting support services for vulnerable families living in rural areas of Colombia (known as FAMI), on children's development and parental behaviors. We use a clustered randomized controlled trial that assigns 87 municipalities to treatment and control, to evaluate the effects of these interventions on children growth and development. 1,460 children younger than 1 year of age were assessed at baseline. The interventions were also complemented with training, supervision and coaching of FAMI program facilitators. We assessed program effects on children's nutritional status by anthropometric measures, cognitive, receptive and expressive language, and fine and gross motor using the Bayley scales of infant development-III and socio-emotional development based on the Ages and Stages Questionnaire for the socioemotional domain. The interventions had positive and significant effects on Bayley-III cognitive scale (0.15 SD), receptive language (0.11 SD), expressive language (0.14 SD) and gross motor development (0.14 SD). We also report a reduction in the risk of stunting of -0.13 SDs. We do not find any effects on socio-emotional development. We report positive and statistically significant effects on the number of toy materials at home (0.36 SD), the number of varieties of play materials (0.28 SD), and the number of varieties of play activities with adults at home over the past three days (0.17 SD). This trial is registered with Current Controlled Trials ISRCTN93757590.

1. Introduction

In this study we evaluate the effects of an intervention aimed at *improving the quality* of public parenting support services for vulnerable families living in rural areas of Colombia, on children's development and parental behaviors. The Family, Women and Infancy Program (FAMI for its acronym in Spanish) aims at improving pre and postnatal services for pregnant women and their newborn children since 1991 when it was first established. Starting in 2014 we designed, piloted and evaluated the implementation of a structured curriculum to improve child development (including cognitive, verbal and motor skills), a nutritional supplement plus psychoeducation around feeding and nutrition, and training and supervision of program facilitators in the FAMI program using a cluster (town) randomized controlled trial (RCT) in a sample of 87 towns in 3 of Colombia's 32 departments. The program was delivered for an average of 10.4 months. We evaluate the effects of the intervention on children's nutritional status, cognitive, receptive and expressive language, and fine and gross motor. We also study the effects on mothers' parenting skills, parental knowledge and perceptions, parental self-efficacy, mental health, and the home environment.

The family is probably the most important institution in an individual's process of skill formation. Families are in charge of making some of the most crucial decisions in children's lives since conception and up until they reach young adulthood. These decisions range from prenatal care of the pregnant mother to the choice of schools, and other critical decisions regarding the quality and quantity of time that parents devoted to their children, early childhood education choices, financial investments devoted to promoting child development, the choice of discipline strategies and a healthy home environment. Furthermore, early investments by families leverage and facilitate further learning, making it crucial for the consolidation of all human functions.

However, optimal investment decisions of families in their children depend upon household time and financial constraints, the knowledge and perceptions that parents have about the process of formation of children's skills, and the extent to which they care about their children and the future. Bernal and Fernández (2013) report that poor mothers in Colombia respond correctly only about 40% of items in the Knowledge on Infant Development Inventory –KIDI- (MacPhee, 1981) regarding developmental milestones, parental practices, health and child safety. Nores and Bernal (2014) report that 40% of parents of children younger than 6 years of age in low-SES households think that children will develop eventually without any effort from their parents, and 55% think that all children will eventually develop the same skills regardless of their circumstances. 36% report that playing is not crucial for appropriate early development.

In addition, Bernal (2016) reports that parental practices in Latin-America are far from ideal. For example, the median of exclusive breastfeeding is well below 4 months in most countries of the region. Only 10% of children in low-SES households consume balanced diets (all food groups at least

once a day) in a subset of Latin-American countries where such information is available (*circa* 2010) and 58% of parents of children 2-14 years of age report using physical or verbal reprimands often with their children (*circa* 2011). On average, 44% of low-SES mothers of children younger than 5 report at least 4 weekly activities with their children aimed at promoting development (e.g. reading, drawing, singing, etc.) while 22% of fathers do. And only about 34% of low-SES households with children younger than 5 report owning books for children, among a variety of other information documented in the report.

Thus, the intervention of governments through a variety of public policies such as conditional cash transfers, infrastructure, and family-friendly policies can contribute to the alleviation of some of these constraints which would, in turn, promote children's early development. Parenting programs are one of such initiatives which aim at improving parental knowledge, perceptions, parental self-efficacy and the home's learning environment. Parenting programs have been shown to have large impacts on parental behaviors and children's development in certain settings (Aboud and Yousafzai, 2015). However, most of the results on parenting programs are derived from small and fully controlled trials (Berlinsky and Schady, 2015), and little is known about how to develop the capacity to deliver high quality parenting services at scale, particularly in rurally disperse areas.

In Colombia, there are 4.3 million children younger than 6 years. Of these, 2.5 million are socio-economically vulnerable. Enrollment in early childhood programs ranged from 20% to 40% for most of the period since the late 1980s, and was one of the highest in the Latin-American region (Bernal and Camacho, 2011). The FAMI program was until 2014 the only program that supported socioeconomically vulnerable families with children younger than 2 in terms of parental practices, health and nutrition. By 2014, approximately 250 thousand children of about 1.2 million that are eligible participated in the program.

This study is relevant to developing countries, not only in Latin-America, for various reasons. First, it adds to the evidence on the importance of parenting programs on children's development. In particular, the best known programs with rigorous evaluations are based on individual home visits (Grantham-McGregor et al., 1997; Olds et al., 1986a, 1986b, 1994) while the program we work with is mostly based on group sessions. The results are important because groups can be more cost-effective than individual visits. Second, it seeks to understand the impact of scalable quality improvements in a public parenting program that works at scale serving 250 thousand children younger than 2 years of age and their families in socioeconomic vulnerability nationwide. Third, it contributes to the understanding of how to best serve disperse rural populations in a cost-effective way by providing itinerant training and coaching to front line workers. And finally, it allows us to assess the barriers, challenges and opportunities of improving quality of parenting programs at scale by leveraging on local low-skilled human resources and low cost enhancements.

Our results indicate that the interventions had positive and significant effects on the Bayley-III cognitive scale (0.15 SD), receptive language (0.11 SD), expressive language (0.14 SD) and gross

motor development (0.14 SD). We also report a reduction in the risk of stunting of -0.13 SDs (P=0.037). We do not find any effects on socio-emotional development. We also report positive and statistically significant effects on the number of toy materials at home (0.36 SD), the number of varieties of play materials (0.28 SD), and the number of varieties of play activities with adults at home over the past three days (0.17 SD).

2. Methods

2.1 Background and intervention

Description of the FAMI program

The FAMI program aims at supporting families in issues related to pregnancy, childbirth, breastfeeding, nutrition, health, childrearing and parental practices more generally. The program targets pregnant women, breastfeeding women and parents of children younger than two years of age in socioeconomic vulnerability.[1] The program is delivered through both, group sessions of about two hours and home visits of one hour. In particular, two group sessions per month are delivered for pregnant women, two sessions per month for breastfeeding women, and one weekly group session for parents of children between 0 and 24 months of age. Families receive one monthly 1-hour home visit to reinforce topics covered during group meetings. The program also delivers a nutritional supplement that corresponds to 30% of (monthly) recommended nutritional intake. The size of each FAMI unit varies from 12 to 15 beneficiaries. Of these, approximately 75% are parents of children between 0 and 24 months of age and 25% are pregnant women.

The program is facilitated by women in the community with no particular early childhood training at the moment of their hiring. They are just required a high school degree. There is no national guideline for a specific curriculum to be used in early childhood services, including parenting services provided by the FAMI program. The Board for Early Childhood has emphasized the principle of curricular freedom, and national standards are intentionally broad. Program providers are expected to adapt the learning standards to their own programs (Comisión Intersectorial de Primera Infancia, 2012). The average cost of program delivery is 310 US dollars (USD)[2] per child per year (Bernal, 2013).

Starting in 2011, the national government decided to prioritize early childhood investment in Colombia as a result of increased interest and evidence about the importance of pre-primary education programs for growth, development, and equity. In that year, the government launched the national early childhood strategy *De Cero a Siempre* - DCAS (From Zero and Forever). The strategy

[1] The program is targeted by using Colombia's proxy means test designed to identify potential beneficiaries of social policies. It is based on a household survey that collects information on socioeconomic vulnerability to construct an index known as SISBEN.

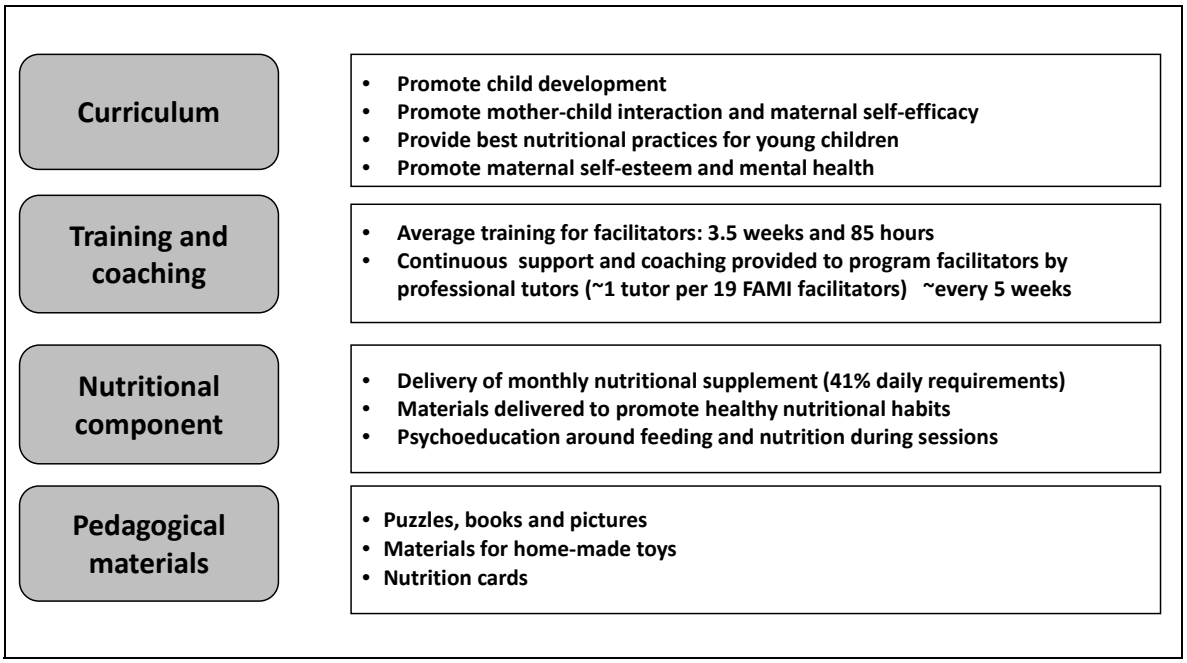
[2] At an average exchange rate of COLP 2,800 per dollar during 2015-2016.

aimed at delivering high-quality and comprehensive early childhood services for 1.2 million disadvantaged children with a budget close to 800 million USD per year over four years (Bernal and Camacho, 2014). As part of the government's interest in improving early childhood services nationwide, this research team worked in association with the Family Welfare Agency and Fundación Éxito in the design, pilot and evaluation a quality enhancement of the only parenting service available in 2014 targeted at socioeconomically vulnerable families, the FAMI program.

Description of the interventions

The components of this intervention, aimed at improving the quality of the FAMI parenting program, are described in Figure 1. The program consists of four complimentary elements: a structured early stimulation curriculum to improve child development, a nutritional supplement that is larger and of better quality than the one that had been typically received by FAMI participants, itinerant training, supervision and coaching for FAMI program facilitators, and pedagogical materials to facilitate the delivery of the curriculum.

Figure 1. Components of the intervention to enhance the quality of the FAMI program



Source: Own elaboration.

The new curriculum is based on the Jamaican home visitation program now known as *Reach Up* (Grantham-McGregor et al., 1997) but required extensive adjustment because the FAMI program is mostly delivered through group meetings.[3] The early stimulation curriculum aims at strengthening the child development component of the FAMI program (in particular, language, cognitive, motor,

[3] A pilot was conducted during the second semester of 2013 in 5 towns in Antioquia to assess feasibility of the new curriculum and delivery through group meetings.

social-emotional development) and strengthening the nutrition component by providing psychoeducation around feeding and nutrition. It also aims at improving mothers' knowledge, practices and enjoyment of bringing up her child and improving mothers' self-esteem and mental health. The curriculum includes 24 home visiting sessions and 20 group meeting sessions.

Most of the program content was delivered through the group sessions as they were held on a weekly basis. However, the home visits provided the opportunity to reinforce the material covered and to personalize to specific needs of the family as necessary. Group meetings were delivered for pregnant and breastfeeding women with children up to 6 months, mothers with children from 6 to 11 months and mothers with children aged 1-2 years. The curriculum involves materials to be used during home visits and group meetings, including age-appropriate books, puzzles, pictures, construction blocks and nutrition cards. The cost of these materials was of about USD 27[4] per child per year. The program had a particular emphasis in teaching how to elaborate home-made toys with recyclable materials that could be used to practice the activities proposed by the curriculum with children at home.[5]

In addition to the set of activities and materials, the qualification of the FAMI program also included a training and coaching component to support and maintain the quality of home visits and group meetings. In particular, a team of nine professional tutors (trained and supervised by the research team) provided initial pre-service training and then continued to provide in-service training and support to FAMI program facilitators throughout the implementation of the program. Training was provided sequentially by town. All FAMI mothers were trained simultaneously in a given town for an average of 3.5 weeks and 85 hours.[6] More specifically, towns with less than 5 FAMI units received 75 hours of training in 3 weeks, towns with 6 to 9 FAMI units were trained for 100-125 hours in 5-6 weeks and towns with more than 10 FAMI units received training during 150-175 hours offered during 6-7 weeks. The training protocol included face to face instruction (10%), design and preparation of home-made toys (20%), practice in pairs by role playing (55%), and practice with children under supervision (10%) and individual work (10%). The one-time cost of pre-service training per FAMI mother was USD 113 or USD 11 per child. [7]

After training was finalized, tutors coached FAMI facilitators continuously throughout the duration of the intervention. The observations and feedback sessions took place approximately every 6 weeks. In 44% of treated FAMIs 4 to 5 weeks elapsed between supervision visits, in other 42% treated FAMIs 6 to 7 weeks elapsed and the remainder 12% received a supervision visit every 8 weeks or more. Each tutor was in charge of 5 towns and 19 FAMIs, on average. During the supervision visit, the tutor would meet with the FAMI facilitator to work jointly in the preparation of a group session

[4] Computed at the average exchange rate 2015-2016.

[5] In Appendix 1 we present a more detailed description of the curriculum.

[6] This was done in two stages: an initial stage of 2 weeks and a second stage of 1.5 on average two months later.

[7] Both computed at the average exchange rate 2015-2016 (2,800 COP/USD).

and a home visit, observe a group session and one home visit, provide feedback about her performance after the observations, and finally, the tutor would gather all FAMI facilitators in one town to have a discussion about advantages and difficulties of the program. The cost of coaching was around USD 82 per month per FAMI or approximately USD 8 per child per month.

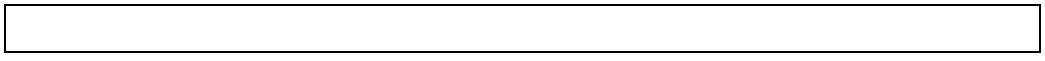
Finally, the intervention also included a nutritional component. It was comprised of the delivery of a monthly nutritional supplement to FAMI participants. The nutritional supplement corresponds to 35% of daily calorie intake requirements and 54% of daily protein intake requirements for pregnant women, breastfeeding mothers and children younger than 2 years of age (for 30 days). The cost of the package and delivery was USD 26 per month. We include more details about the specific composition of the nutritional package in Appendix 2.

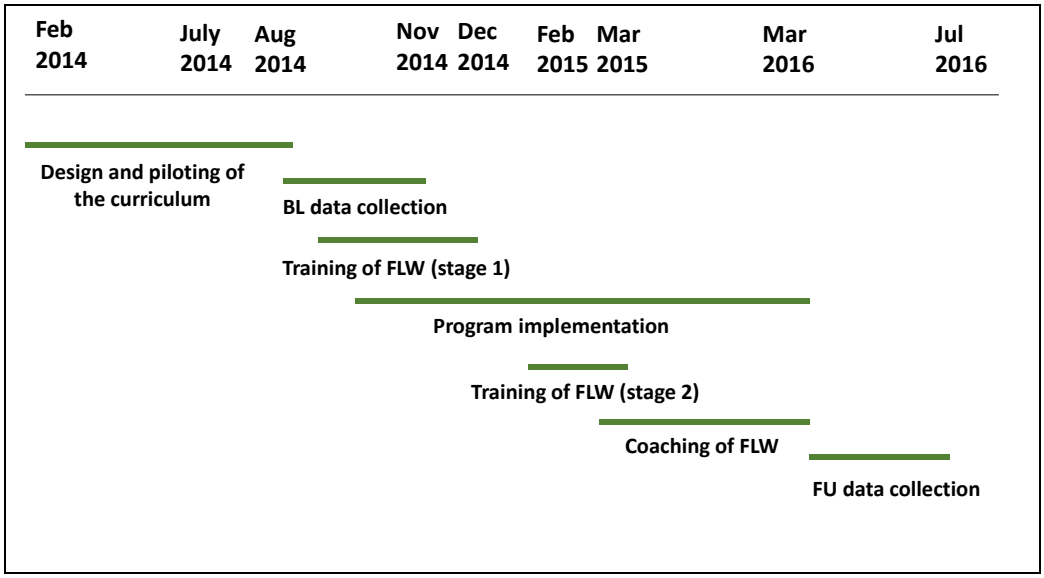
We expect these improvements in curricular content and pre-service and in-service training for program facilitators to improve the ability of FAMI facilitators to deliver more effective sessions and follow-up participants in a more comprehensive way, work more specifically in helping mothers improve their ability to stimulate their children and enrich their home environments to improve the opportunities for child development. This would in turn, improve children's development. Similarly, the delivery of the nutritional supplement and including nutritional content in the curriculum aims at improving parental practices regarding food handling, feeding, breastfeeding, health and topics relevant to the child's physical development. Thus, the hypothesis that we test is whether these enhancements to the FAMI program had positive effects on children's development and mothers' parenting skills, parental knowledge, parental self-efficacy, and the home environment (see Appendix 3).

2.2 Research Design

We evaluate the impact of the implementation of a structured curriculum in the FAMI program plus pre-service and in-service training for FAMI facilitators using a cluster-randomized effectiveness trial in Colombia. We measure the effects of the intervention on children's nutritional status, cognitive and socio-emotional development. The main mechanisms through which the intervention was expected to have an effect on children's development include: more structured guidelines for session delivery with specific content aimed at promoting children's development and parental abilities, the availability of pedagogical materials to be used with mothers and children during group sessions and home visits, the availability of a nutritional supplement that increased the amount and quality of food for program participants, professional training of front line workers, and coaching throughout the implementation period which provides continuous feedback on ways to improve the quality of the intervention.

Figure 2. Study's timeline





Source: Own elaboration. FLW: Front-line workers.

We conducted a cluster randomized controlled trial in 87 towns in Colombia involving 340 FAMI units using a two-arm design with treatment and control group. The intervention was rolled out across a period of four months September-December 2014, and finalized in March 2016. Alongside the phase-in period, we conducted the baseline survey immediately before, and between April-July 2016 we conducted follow-up survey right after the end of the program. Each survey took place over a period of four months. The average duration of the program was 45 weeks (10.4 months) with a minimum of 34 weeks and a maximum of 58 weeks. Previous studies found sustainable benefits from similar interventions lasting from nine months to three years (Walker, Chang, Vera-Hernández and Grantham-McGregor, 2011; Walker, Chang, Powell, Simonoff and Grantham-McGregor, 2006; Attanasio, et al., 2014). See Figure 2 for details about the study's timeline.

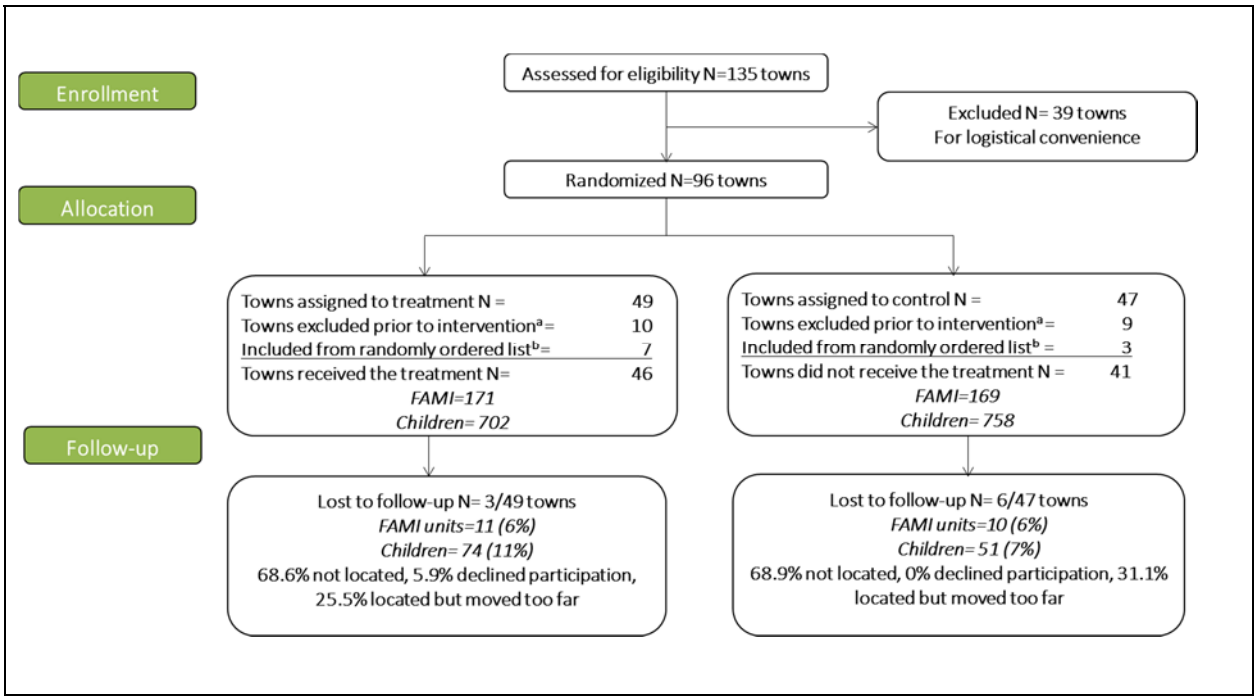
2.3. Sampling frame and sample size

The universe of towns considered for this study was defined as those with less than 40.000 inhabitants (to encompass semi-urban and rural towns), with at least two FAMI units and with no more than one MF unit (*Modalidad Familiar*)[8] located in three departments in central Colombia (Cundinamarca, Boyacá, and Santander). We restricted the study sample to these regions because they had a large number of towns that complied with the eligibility criteria mentioned above, had a

[8] The MF is a public parenting program that was introduced during the first half of 2014. MF is similar to FAMI in that it serves beneficiaries through one monthly home visit and one weekly group meeting but: (1) it serves children 0 to 5 years of age under the same scheme while FAMI units only serve children 0 to 2, (2) it has infrastructure for group meetings while FAMI uses other community spaces or the FAMI mother's own home, (3) serves, on average, 45 beneficiaries as compared to 15 in FAMI units, (4) is led by a professional and an assistant (sometimes a former FAMI mother) as compared to a single person which is not required to have a college degree in FAMI units, (5) offers a nutritional component that is at least 5 times bigger than in FAMI and (6) has access to a group of professionals including psychologists and health professionals who support MF activities.

larger number of FAMI units per town, on average, than other departments and are also much less geographically disperse than other departments[9]. In addition, the municipalities in these regions are similar in terms of their cultures and customs. This would make the use of a single curriculum —and associated materials such as pictures and books—more feasible, which was particularly important given the ethnic and geographical diversity of Colombia. These screening criteria yield 135 eligible towns. In Figure 3 we show the study's flow chart for sample selection. We then selected a sample of 96 out of this universe of 135 eligible towns by choosing all towns with no MF units, choosing towns with only one MF and selecting towns whose geographic location favored the construction of clusters to be assigned to program tutors for training and supervision.

Figure 3. The study's flow chart



Source: Consort Flow Chart. Own Elaboration

[a] Once in the field for data collection, we realized some towns did not have any FAMI units as they had made the transition to other models (MF)

[b] Towns not originally assigned to the sample were randomly ranked and used as replacements. However we did not have enough in all randomization strata.

We randomly assigned our sample of 96 towns into treatment and control. We defined the following three stratification variables for our random assignment procedure: presence of MF (0 or 1), department (Boyacá, Cundinamarca, Tolima, Boyacá) and population size (0-10.000 inhabitants, 10.000-40.000 inhabitants). Through this procedure, we assigned 49 towns in the treatment group

[9] These features characterize at least 75% of semi-urban areas in Colombia so the results are, in principle, still generalizable.

and 47 in the control group.[10] We further assigned the remaining 39 towns of our universe (=135-96) to randomly ordered waiting lists within department, in case we needed to substitute towns initially selected to make part of the study.

Power calculations for the trial assumed program effects of 0.25 of a standard deviation (SD) relative to the control on the Bayley scale on infant development, 4 FAMI units per town and 4 children per FAMI. We used an inter-class correlation between towns of 0.04 (in the Bayley-III scale) that is reported by Attanasio et al. (2014) for a similar study in Colombia that was implemented in towns with similar characteristics and had a very similar design.[11] This sample provided 95% power at 5% significance level, allowing for an attrition rate of 10%.

However, once our team reached the regions, we realized that 19 towns originally selected had completely made a transition from FAMI to MF so we did not find any FAMI units in the field.[12] In these cases, we replaced towns that had completely migrated out of FAMI in a given department, with the first town in the randomly ordered list in the same department. The replacement town was assigned to the same group (treatment or control) as the town being replaced. By following this procedure, we could successfully replace 10 of the 19 towns that no longer ran the FAMI program (see Figure 3). That yields a final sample of 87 towns, with 46 in the treatment group and 41 in the control group. In particular, 171 FAMI units received the treatment while 170 FAMI units remained as control. Each town had an average of four FAMI units in the final sample. In Figure 4 we show the final distribution of treatment and control towns in the sample of 87 municipalities.

In each FAMI unit, we sampled all participant children between 0 and 12 months of age at baseline to make part of the study sample for a total of *N*=1,460 children. By doing this we wanted to maximize sample size by selecting an age range as large as possible, at the same time that we maximized the potential time of exposure to the program by including the younger children who would be eligible for the program for a longer time. Each cluster (town) had an average of 17 children. In sum, 171 FAMI units and 702 children younger than 12 months of age at baseline in 46 towns received the treatment and 169 FAMI units and 758 children in 41 towns remained as control throughout the trial. The actual number of children per FAMI varied from one to 11 with a mean of 4.3 and SD of 1.9. The number of FAMIs per town in our final sample ranged from one to 13 with a mean of 3.9 and SD 2.3 (see Appendix 4). The observed inter-class correlation between towns in the

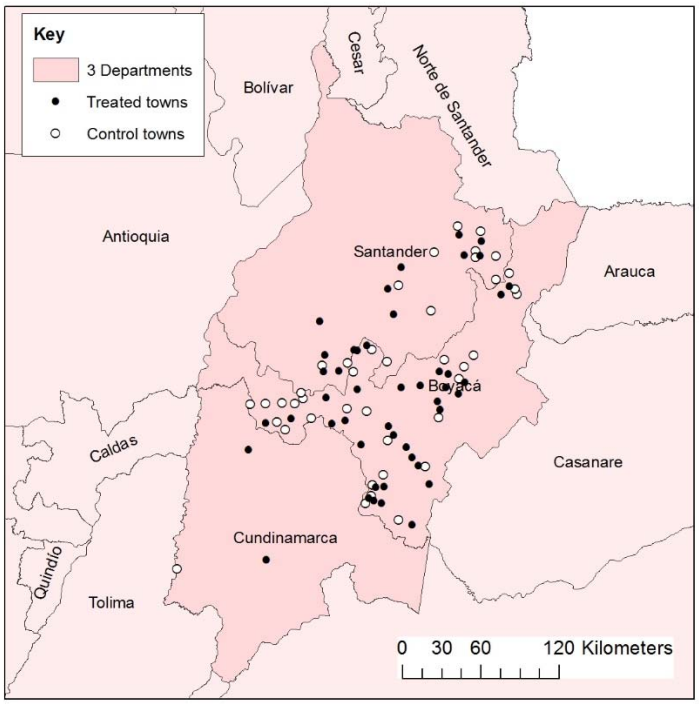
[10] Within each strata a random number from a standard uniform distribution (0,1) was drawn for each municipality, municipalities with odd numbers were assigned to treatment and municipalities with even numbers were assigned to control. In cases were a strata had an even number of municipalities, this procedure was followed up to an even number and the municipality with the highest random draw was assigned to treatment or control with 50-50 probability.

[11] No reliable data was available for FAMI units that could be used to calculate the inter-cluster correlation at the beginning of this trial. Thus, we used an inter-class correlation *between FAMIs* of 0.0135 in height for age which is the inter-class correlation between home-based nurseries of 12 children each unit for children 0 to 5 years of age (*Hogares Comunitarios* program) with similar care providers and located in similar towns (Bernal and Fernández, 2013).

[12] The transition from FAMI to MF also made part of the qualification of early childhood services promoted by the DCAS strategy. In practice, one program was replaced by the other in some of our study towns and other parts of the country.

Bayley scale was 0.15 and 0.16 between FAMIs. Assuming program effects of 0.25 of a standard deviation (SD) relative to the control on the Bayley scale on infant development and an attrition rate of 10%, the updated power calculation indicates that our final sample provided 70% power at 5% significance level.

Figure 4. Geographic location of the final sample



Note: Treated towns depicted in black and control towns depicted in white.

2.3 Data collection procedures

Informed consent[13] and parent questionnaires on household's socio-demographic characteristics and children's socio-emotional outcomes were collected at the child's household. When this was not possible, the assessment took place during the FAMI group meeting.[14] Background information on FAMI facilitators was collected directly from them. Data on child's attendance status at follow-up were collected by parental report. Data on FAMI facilitators was collected directly from them. Apart from their sociodemographic characteristics, we also included verbal ability measured by the Peabody Picture Vocabulary test and personality traits by the Big Five instrument.

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

[13] The Universidad de los Andes ethics committee (No. 287/2014) and the University College London research ethics committee (No 2168/011) approved this study.
[14] The FAMI program group meeting generally takes place at the FAMI mother's own home or a community hall.

target age groups. Inter-rater reliability (intracluster correlation) was above 0.9 on each subscale of the Bayley-III. When it was not possible to assess children in the town's urban center, they were assessed at their own homes trying to guarantee similar testing conditions.

2.4 Measures

We measure different developmental domains using several tests that have been used extensively in evaluations of early care or education and/or have been recommended for developing countries (Fernald et al., 2009). All instruments used had been previously adapted to Spanish. The primary outcomes included in this study are the following.

Nutrition. In line with similar international studies (Walker et al., 2004; Fernald, Gertler, and Neufeld, 2008), we collected information on height, weight, and body mass index (BMI) following World Health Organization (WHO) standards (WHO Multicentre Growth Reference Study Group, 2006, 2007) for all children in our sample, both at *baseline and follow-up*. Based on these measures, we constructed a variety of nutritional indicators depending on the child's age and based on WHO (2006, 2007) standards.

Food Insecurity. We measure food insecurity by using the ELCSA scale -Latinamerican Scale for the Measurement of Food Insecurity- which was validated in Colombia (ELCSA Scientific Committee, 2012) at *baseline and follow-up*. Total scores allow classification of households in four different levels of food insecurity: secure, mild insecurity, moderate insecurity and severe insecurity (Álvarez et. al, 2008).

Motor development. We used the World Health Organization (WHO) Milestones to measure gross motor development of children younger than 12 months of age at *baseline*. The scale measures six gross motor milestones of children younger between 6 and 24 months of age. The fieldworker carefully trained determines whether the child can achieve each of the six gross motor milestones assessed. In particular, the items assessed include sitting without support, standing with assistance, hands & knees crawling, walking with assistance, standing alone and walking alone. The windows of achievement based on the WHO milestones represent normal variation in ages of milestone achievement among healthy children. They are recommended for descriptive comparisons among populations and to signal the need for appropriate screening when individual children appear to be late in achieving the milestones (WHO Multicentre Growth Reference Study Group, 2006).

Cognitive development. We used the Bayley scales of infant and toddler development, third edition (Bayley-III) at *follow-up*. 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[15]: cognition, receptive language, expressive language, fine motor, and gross motor. Children were assessed in local community centers with their mothers present.

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.

Socio-emotional development. We used the ASQ for the socio-emotional domain (ASQ:SE) (Squires, Bricker, and Twombly, 2009a) for all children at *follow-up*. The ASQ:SE is a parent-completed assessment system for children ages 6–60 months completely through culturally sensitive questionnaires focusing on socio-emotional development and the identification of children at risk of social-emotional difficulties. It includes self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interactions with others. The ASQ:SE shows high levels of consistency, reliability, validity, and specificity (Squires et al., 2002; Squires, Bricker, and Twombly, 2009b), and has been used for early development assessments in low and middle low income countries (Handal et al., 2007; Heo, Squires and Yovanoff, 2007). We report age-standardized scores. It is important to note that both, ASQ and ASQ:SE are instruments that screen for developmental risk and are traditionally not the best measures for assessing change in children's learning and behavior.

Secondary outcomes

We consider as secondary outcomes those that could mediate the effect of the intervention on primary outcomes. First we assess 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 (Frongillo, Sywulka and Kariger, 2003).

We also include the Self-efficacy in the Caregiver Role Test – Modified (Pedersen et al., 1989; Porter and Hsu, 2003) which evaluates feelings mothers could have while taking care of their children and the DUKE-UNC Functional Social Support Questionnaire 11 (Broadhead, et. al, 1988). This scale measures an individual's perception of the amount and type of personal social support he or she receives. These instruments are included to measure maternal self-esteem and motivation. We also use an adaptation of the domain III of the UNICEF Care Indicator Questions to inquire about discipline strategies at home (Frongillo, Sywulka and Kariger, 2003). Similarly, we use a 10-item scale to measure maternal knowledge about child development which uses some of the questions available

[15] Bayley N. Bayley scales of infant and toddler development. 3rd ed. Pearson Education, 2006.

in the Knowledge of Infant Development Inventory –KIDI- (MacPhee, 1981), and was piloted in relevant populations in Colombia. Finally, we use the 10-item version of the Center for Epidemiologic Studies Depression Scale – CESD-10 (Anderson et al., 1994).

Household demographic and socio-economic status data were collected at both baseline and follow-up. We collected information about characteristics of the dwelling such as type of property, type of floors, roofs, and walls, number and type of bathrooms, access to public utilities, availability of durable goods, characteristics of members of the household such as educational attainment, employment status and wages, ethnicity, participation in social programs, and household income by source. The socio-economic status score was computed as the principal component of a set of dichotomous variables that describe characteristics of the household, ownership of durable goods, and access to public utilities. A lower socio-economic status score is denoted by a negative factor and a higher status by a positive factor (Vyas and Kumaranayake, 2006).

2.5 Data Analysis Strategy

There are three main hypotheses in this study. The first one assesses whether the implementation of a structured curriculum in the FAMI parenting program in addition to training for program facilitators, supervision and coaching had positive impacts on child development, nutrition, and morbidity. The second evaluates whether the intervention had positive average effects on parental practices and the learning environment for children in their household. Finally, the third hypothesis evaluates whether the qualification of FAMI has increasing effects in the duration of actual exposure to the treatment. *ITT* was randomly assigned given the cluster-randomized effectiveness trial design of the study. Thus, we estimate *ITT* effects on children's development with the following specification:

$$Y_{isl,1} = \beta_o + \beta_1 T_{sl} + \gamma Y_{isl,o} + X'_{isl,0}\delta + D_{isl,0}\theta + F_{isl,0}\sigma + S_{isl,0}\tau + Z_{isl,1}\rho + \varepsilon_{isl,1} \qquad (1)$$

where $Y_{isl,1}$ is the outcome of interest for child *i* in FAMI center *s* in municipality *l* at follow-up (*t*=1); T_{1sl} is a dummy equal to 1 if the FAMI center *s* in municipality *l* receives the treatment; and $Y_{isl,o}$ is the baseline (*t*=0) level of the outcome of interest (or level of the corresponding aggregate construct in the case that the same measure was not administered at baseline and follow-up) for child *i* in FAMI center *s* in municipality *l* at follow-up. For child developmental outcomes (with the exception of nutritional status) we do not have the same outcome at baseline and follow-up since the tests could not be administered given children's ages at baseline. For these outcomes, we use a composite score constructed using age-standardized birth weight and weight-for-age and height-for-age Z-scores. The purpose of this approach is to maximize efficiency. $X'_{isl,0}$ is a set of basic child and

household characteristics, which are also added to improve efficiency (minimize residual variance)[16] and control for the imbalance in some baseline characteristics observed between groups at baseline (child's father present in the household, household wealth quintile, maternal PPVT score and maternal conscientiousness from the Big Five personality test); $D_{isl,0}$ are a set of department fixed effects, $F_{isl,0}$ are a set of dummies indicating the presence or not of the alternative parenting program in town (*modalidad familiar)* and $S_{isl,0}$ are a set of municipality population size dummy variables indicating above and below 10 thousand inhabitants (all included due to our stratified randomization procedure)[17] and , and $Z_{isl,1}$ are a complete set of tester or interviewer dummies. $\varepsilon_{isl,1}$ is the random error term, clustered at the municipal level *l* (the unit of randomization).

In principle, given that the treatment was randomly assigned, we could have estimated equation (1) by OLS. However, as we report later, we find statistically significant differences between individuals found at follow-up and individuals not found at follow-up. For this reason, we estimate equation (1) by maximum likelihood implementing a correction for self-selection into the follow-up sample.[18] β_1 is the estimated average impact of the treatment on outcome $Y_{isl,1}$ (intent-to-treat estimate). We also use equation (1) to estimate the impact of the treatment on parental skills and the home environment. In this case $Y_{isl,1}$ corresponds to the intermediate outcomes that were described in section 2.5.

We report means, 95% confidence intervals, standard error and statistical significance at 1%, 5% and 10%. The p- values reported refer to one-tailed tests for the impacts on children's cognitive and socio emotional development (reflecting the presumption that this intervention could not harm these developmental outcomes) and two tailed tests for the impacts of the intervention on child's nutritional status. This choice reflects the possibility that the nutritional supplement might have increased overweight or obesity if used inappropriately.

For some of the developmental domains analyzed in this study, we have more than one outcome measure with which to explore treatment effects. To deal with multiple hypothesis testing we group our outcome measures into domains or "constructs" using factor analysis or computing total scores and estimate equation (1) using the resulting factor index as the relevant dependent variable. This procedure is based on the idea that items within a domain are measuring an underlying common "construct" (or factor).[19]

---

[16] Child's age and age squared, gender, birth order, household size, number of siblings, maternal schooling attainment and teenage mother.

[17] Item non-response in baseline covariates is not correlated with treatment status. Thus, we imputed missing covariate values with the average of the non-missing observations and accounted for this imputation with a dummy variable in equation (1).

[18] We use the distance between the households and the town's city hall as the exclusion restriction in the selection model.

19 The standard errors of versions of regression (1) that are not ran on aggregate factors of a single developmental dimension, but are ran on different subscales or measures of the same construct have been adjusted for multiple hypotheses testing using the Romano-Wolf (2005) (RW) step-down procedure.

Finally, we also present results for the impacts of the intervention depending on the duration of actual program exposure. In particular, we look into the effects of the program as the number of program contacts (either group sessions or home visits) increases. However, the duration of exposure to the program is not random because certain parents assigned to the treatment group decide to participate in more or less program activities than other parents for reasons that might be, in turn, correlated with the child's developmental outcomes and other household characteristics. In this case, we estimate the impact of duration of exposure to the program by a two-stage least squares procedure (instrumental variables approach). In the first stage we predict the duration of program exposure explained by *ITT*, and in the second stage we estimate the main effects by using the predicted duration of exposure that resulted from the first stage. Random assignment is clearly exogenous to child and household characteristics and it also significantly predicts duration of exposure to the intervention.[20]

3. Results

At baseline, we assessed 1,460 children and surveyed their households and 340 FAMIs. The child attrition rate between baseline and follow-up was 8.6% (n=125) across treatment arms: 74 (10.5%) of the children from the treatment arm were not measured at follow-up and 51 (6.7%) from the control arm. The difference in loss among the groups was statistically significant at 5% confidence level. In Appendix 5 we show the attrition analysis. Thus, we estimated a selection model (into the follow-up sample) to correct the impacts by using a maximum likelihood estimation procedure.[21] Children lost to follow-up were older, less likely to have a resident father at home, and more likely to have mothers with lower Peabody Picture Vocabulary Test (PPVT) score. In Appendix 6 we show the selection model used in the maximum likelihood correction.

We excluded from the analyses 12 children who scored more than 3 SD below the mean on the Bayley-III cognitive subscale (possible disability) and 15 children who were 6 SDs below the mean and 6 SD above the mean of height-for-age, 42 children who were 5 SD below the mean and 5 SD above the mean of weight-for-length, 25 children who were 6 SD below the mean and 5 SD above the mean of weight-for-age, and 40 children who were 5 SD below the mean and 5 SD above the mean of Body mass Index-for-age (extreme observations).

The qualification of FAMI units was in place between last quarter of 2014 and the end of March 2016. Excluding vacation periods (Christmas), the program effectively operated for an average of 10 to 11 months. This means that a family could have attended a total 44 weekly group sessions (if age-eligible throughout the period) and received 11 monthly home visits. In terms of effective participation in the program in FAMI units in the treatment group, we know that 74% of all children

20 La F-test for statistical significance of the instrument is above 25.
21 In this selection correction model we use the distance from the household to the urban center's town hall as the exclusion restriction.

assessed at baseline participated in at least one FAMI activity during the intervention.[22] This implies that sometime between baseline data collection and the end of pre-service training of FAMI facilitators, 26% of participants dropped out of the FAMI program.

In Appendix 7 we show the distribution of children in the intervention group by the duration of exposure to the program. Among children who attended at least one group session, close to 25% attended 1-13 group sessions, 25% attended 14-22 sessions, 25% attended 23-32 sessions and the remainder 25% attended more than 32 sessions. On average, children who participated in at least one session, attended 19.7 (SD=10), i.e., approximately half of the proposed sessions. In terms of home visits, we know that 72% of treated children assessed at baseline are registered in home visit attendance lists at least once. Among these, 25% received 1 to 5 home visits, 25% received 6-7, 30% received 8 or 9, and the remainder 20% received more than 9 home visits throughout the duration of the program. On average, children that received at least one home visit, received 7.8 visits (SD=3).

Based on a satisfaction survey collected from all 171 FAMI facilitators in the treatment group, we know that 98% of them found the program extremely useful (65%) or useful (33%). In addition, 57% report to have found the curriculum very different from their regular practice and 25% different. In particular, the issues that seem to be different with respect to how they had typically worked are: (i) practicing play activities with mothers and their children, (ii) practicing language activities with babies, (iii) making homemade toys with mothers, (iv) encouraging parents to play with their children at home, and (v) listening to parents about their achievements at home. Basically all of them (99%) reported that they would continue to use the proposed curriculum after the end of the project.

3.1 Descriptive Statistics

Table 1 shows baseline characteristics of children, their mothers, and their households and in Table 2 we show baseline and follow-up final outcomes by randomization status (*ITT*). Children are, on average, 5.6 months of age, about 30% do not live with their fathers, and households in the sample have, on average, two children. Maternal average schooling is 8.6 years, and 23% of mothers are teenagers. In terms of nutritional status, we report an average weight-for-age of 0.2 SD and height-for-age of -0.1 SD below the mean of the reference population. Child development in our sample is comparable with that of children in the Colombian Longitudinal Survey (ELCA, 2013). Chronic malnutrition is about 11% in our sample compared to 14% for the lowest urban SES in Colombia, 12% in selected rural regions and 7.6% nationwide by ELCA results. Overweight is 19% in our sample compared to 16% in the lowest urban SES and 12.4% in selected rural areas (ELCA, 2013).

We find statistically significant differences between study groups in several sociodemographic characteristics and in a few developmental outcomes at baseline. In particular, more fathers live in the child's household in the control group (75% vs. 70%), the wealth index is

22 This information is available from attendance lists that FAMI facilitators collected as part of the supervision protocol of this intervention. Thus, the data is only available for participants in the treatment group.

higher in the treatment group and maternal verbal scores by PPVT are higher in the treatment group (0.16 SD vs. -0.14 SD) as well as maternal conscientiousness (0.15 SD vs. -0.15 SD).

Table 1. Baseline characteristics of children and their households by randomization status

	ITT=1		ITT=0		
	Mean	SD	Mean	SD	
Sociodemographic characteristics					
Child's age in months	5.76	(3.44)	5.54	(3.29)	
Boys (%)	0.52	(0.50)	0.51	(0.50)	
First Born (%)	0.47	(0.50)	0.45	(0.50)	
Child's birth weight (gr)	3,190	(572)	3,156	(500)	
Low birth weight (%)	0.07	(0.26)	0.07	(0.26)	
Mother's age (number of years)	26.16	(6.95)	26.48	(6.81)	
Mother's education (years)	8.85	(3.42)	8.40	(3.31)	
Father present (%)	0.70	(0.46)	0.75	(0.43)	**
Mother present (%)	1.00	(0.06)	1.00	(0.07)	
Mother single (%)	0.24	(0.43)	0.21	(0.41)	
Mother divorced (%)	0.01	(0.11)	0.03	(0.16)	*
Teenage mothers (%)	0.25	(0.44)	0.21	(0.41)	*
Household wealth index	0.06	(0.96)	-0.06	(1.04)	**
Home ownership	0.37	(0.48)	0.39	(0.49)	
SISBEN survey	0.71	(0.45)	0.70	(0.46)	
Household size	4.08	(1.47)	4.10	(1.43)	
No. of observations	702		758		
Caregiver Characteristics					
Peabody Picture Vocabulary Test (raw score)	22.33	(8.53)	19.76	(8.8)	**
Personality Traits (raw score)					
Extroversion	3.33	(0.54)	3.35	(0.54)	
Agreeableness	3.88	(0.57)	3.86	(0.61)	
Conscientiousness	3.92	(0.60)	3.73	(0.60)	***
Emotional Stability	2.73	(0.61)	2.77	(0.56)	
Openness	3.80	(0.60)	3.78	(0.57)	
No. of observations	702		758		
Intermediate Outcomes					
FCI No. of adult books, magazines and newspapers	2.57	(3.08)	2.76	(3.17)	
FCI No. of toy sources	1.35	(0.94)	1.38	(0.97)	
FCI No. of varieties of play materials	1.42	(1.41)	1.42	(1.45)	
FCI No. of varieties of play activities over past 3 days	2.53	(1.56)	2.53	(1.57)	
FCI No. of parental care activities over past 3 days	4.75	(1.04)	4.72	(1.06)	
Exclusive breastfeeding at least 6 months (%)	0.72	(0.45)	0.73	(0.44)	
Social support DUKE UNC-11 total (raw score)	41.22	(8.26)	42.21	(8.89)	
Parental self-efficacy scale (raw score)	26.47	(5.60)	26.45	(4.76)	
Mothers with depression symptoms (%)	0.15	(0.36)	0.13	(0.33)	
Use of verbal or physical abuse in the household (%)	0.02	(0.15)	0.01	(0.11)	
No. of observations	702		758		

Note: ***p<0.01; **p<0.05; *p<0.1

Standard errors clustered by town in parenthesis

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

	ITT=1		ITT=0		
	Mean	SD	Mean	SD	
Baseline					
Nutrition					
Height (cm)	64.98	(6.85)	63.99	(7.05)	*
Weight (kg)	7.38	(1.97)	7.34	(2.02)	
Weight for age Z-score	0.25	(1.40)	0.27	(1.42)	
Length for age Z-score	-0.02	(1.70)	-0.21	(1.74)	
Weight for length Z-score	0.37	(1.59)	0.55	(1.65)	
Global Malnutrition (Underweight)	0.07	(0.25)	0.05	(0.22)	
Acute Malnutrition (Wasting)	0.06	(0.24)	0.07	(0.25)	
Chronic Malnutrition (Stunting)	0.10	(0.30)	0.14	(0.35)	
Overweight by BMI	0.10	(0.29)	0.09	(0.28)	**
Food insecurity (mild, moderate or severe)	0.50	(0.50)	0.42	(0.49)	
Food insecurity (moderate or severe)	0.04	(0.21)	0.03	(0.18)	
No. of observations	702		758		
Motor Development					
No. of WHO motor milestones achieved	2.70	(1.661)	2.78	(1.870)	
No. of observations	344		361		
End of the Intervention					
Nutrition					
Height (cm)	84.34	(4.50)	84.30	(4.59)	
Weight (kg)	12.05	(1.69)	11.94	(1.66)	
Weight for age Z-score	0.02	(1.09)	-0.06	(1.03)	
Length for age Z-score	-0.82	(1.15)	-0.89	(1.18)	
Global Malnutrition (Underweight)	0.02	(0.14)	0.04	(0.19)	
Acute Malnutrition (Wasting)	0.01	(0.09)	0.02	(0.15)	
Chronic Malnutrition (Stunting)	0.16	(0.37)	0.16	(0.37)	
Overweight by BMI	0.26	(0.44)	0.22	(0.42)	
Food insecurity (mild, moderate or severe)	0.49	(0.50)	0.65	(0.48)	**
Food insecurity (moderate or severe)	0.05	(0.22)	0.07	(0.25)	
No. of observations	628		707		
Socio-emotional Development					
ASQ:SE Raw Scores Total	44.30	(26.02)	48.44	(28.69)	
No. of observations	628		707		
Bayley-III Raw Scores					
Cognition	59.37	(7.95)	59.47	(7.53)	
Receptive language	24.34	(5.25)	24.67	(5.06)	
Expressive language	25.78	(5.76)	25.47	(5.49)	
Gross motor	54.58	(4.74)	54.37	(4.44)	
Fine motor	38.18	(4.18)	38.06	(3.80)	
No. of observations	628		707		

Note: ***p<0.01; **p<0.05; *p<0.1

Standard errors clustered by town in parenthesis

In terms of developmental outcomes, we report only a difference in height in favor of the treatment group and a statistically significant difference in food insecurity in favor of the control group (65% vs. 49%). In our estimating equation, we controlled for these initial measures.

In Table 3 we present basic characteristics of program facilitators by evaluation group. FAMI facilitators are, on average, 41 years of age, have completed 13 years of education,[23] they have almost 12 years of work experience in the FAMI program and over 75% report to have an early childhood education certificate. They have an average of 2.5 children of their own. Both groups of women are for most part similar. However, we find statistically significant differences between study groups. In particular, program facilitators in the control group report more hours devoted to FAMI planning activities (6.8 hours vs. 4.9 hours) and FAMI facilitators in the treatment group have higher language scores by PPVT than those in the control group (0.16 SD vs. -0.16 SD).

Table 3. Baseline characteristics of FAMI program facilitators by randomization status

	ITT=1		*ITT=0*		
Variables	Mean	SD	Mean	SD	
Age	41.8	(10.04)	41.4	(10.36)	
Education (years)	13.3	(1.66)	13.0	(1.96)	
Work experience (years)	11.7	(7.96)	11.9	(8.48)	
Certified in early childhood	0.75	(0.43)	0.77	(0.42)	
Single, divorced or widowed	0.24	(0.43)	0.21	(0.41)	
Number of children	2.7	(1.35)	2.5	(1.50)	
MC's household size	3.9	(1.48)	3.9	(1.43)	
Number of children (0-12 months old) attending	4.8	(2.06)	5.1	(2.29)	
Number of pregnant women attending	1.8	(1.34)	1.9	(1.45)	
Number of group sessions held last month	5.4	(4.50)	5.1	(3.39)	
Number of home visits held last month	12.1	(6.66)	13.5	(7.12)	
Hours devoted to FAMI planning activities (hours)	4.9	(3.02)	6.8	(6.92)	**
Peabody Picture Vocabulary Test (Z-score)	0.163	(1.03)	-0.165	(0.94)	*
Knowledge about ECD (Raw Score: correct)	7.29	(1.72)	7.11	(1.39)	
Depressive Symptoms by CESD10	0.16	(0.37)	0.12	(0.32)	
Personality Traits					
Extroversion (Z-score)	0.056	(0.98)	-0.057	(1.02)	
Agreeableness (Z-score)	-0.052	(0.96)	0.052	(1.03)	
Conscientiousness (Z-score)	-0.054	(1.04)	0.055	(0.96)	
Emotional Stability (Z-score)	-0.010	(1.06)	0.010	(0.94)	
Openness (Z-score)	0.026	(1.01)	-0.026	(0.99)	
No. of observations	171		169		

Note: ***p<0.01; **p<0.05; *p<0.1

Standard errors clustered by town in parenthesis

[23] A high school degree corresponds to 11 years of schooling; 12 years of schooling correspond to vocational education; 13 years of schooling correspond to a 2-year t*echnical* or *technological* college degree and 15 years would correspond to a full professional undergraduate degree.

3.2 Impact Results

The first hypothesis of this study evaluates whether the quality enhancement of FAMI units had positive average effects on children's development with respect to FAMI units without the quality enhancement program. In particular, we report average effects by intent to treat, regardless of whether children actually attended the program or how many times they attended. In each table, we present the estimated effect in column 2 along with the 95% confidence interval. In the third column we present the p-value corresponding to the one or two-tailed hypothesis test depending on the outcome, and the fourth column shows the RW-adjusted P-value. Finally, in the last column we present the estimated effect in terms of standard deviations (SD) with respect to the control group.

3.2.1. Cognitive and motor development

Table 4 shows the estimated treatment effects on child cognitive and motor development using the indicators we hypothesized could have been affected by the intervention: the Bayley-III cognitive, receptive language, expressive language, and gross and fine motor subscales. In the last row we show the age-standardized factor that aggregates all the subscales. The effect of the program on Bayley-III cognitive subscale was 0.15 SDs, on receptive language was 0.11 SDs, on expressive language was 0.14 SDs and on gross motor was 0.14 SDs. There were no significant effects of the program on fine motor development. The overall effect of the program on total Bayley-III scores was 0.14 SDs which is statistically significant at 5% confidence level.

Table 4. Results on children's cognitive development

VARIABLE: Bayley-III	Beta (95% CI)	P-value	RW P-value	D
Cognitive	0.154 (0.008,0.299)	0.020**	0.057*	0.154
Receptive language	0.115 (-0.010,0.240)	0.036**	0.057*	0.115
Expressive language	0.138 (-0.005,0.281)	0.029**	0.057*	0.138
Gross motor	0.141 (-0.021,0.302)	0.044**	0.087*	0.141
Fine motor	0.046 (-0.094,0.187)	0.260	0.257	0.046
Total Bayley-III	0.136 (-0.013,0.285)	0.037**	--	0.136

Note: ***p<0.01; **p<0.05; *p<0.1; standard errors clustered by town in parenthesis
Confidence interval at 95% confidence level in parentheses, p-values reported for one-tailed t-tests
Estimated by maximum likelihood correcting for selection into the follow-up sample
Bayley-III scores have been non-parametrically age-standardized
D=(ß/SD controls), where SD is standard deviation for control group within the estimation sample
P-values in the column (4) are corrected by RW (2005). We consider 3 hypotheses for cognitive development and 2 hypotheses for motor development. Total Bayley-III score excluded.

3.2.2. Nutrition

Table 5 reports effects of the intervention on the set of primary outcomes that we hypothesized would be impacted by nutritional supplementation and psychoeducation around feeding and nutrition: household food insecurity and children's nutritional outcomes. The effect of the program on the risk of stunting was of -0.13 SDs. None of the effects on other nutritional outcomes or on ELCSA food insecurity was estimated to be significantly different from zero.

Table 5. Results on children's nutritional status

VARIABLE	Beta (95% CI)	P Value	RW P-Value	D
Risk of underweight	0.006 (-0.038,0.050)	0.795	0.521	0.017
Stunting	-0.020 (-0.086,0.045)	0.540	0.521	-0.056
Risk of stunting	-0.058 (-0.113,-0.004)	0.037**	0.075*	-0.126
Risk of overweight	-0.004 (-0.057,0.049)	0.882	0.521	-0.010
Overweight	-0.023 (-0.062,0.016)	0.248	0.234	-0.086
ELCSA Food insecurity	-0.033 (-0.111,0.045)	0.411	--	-0.068

Note: ***p<0.01; **p<0.05; *p<0.1; standard errors clustered by town in parenthesis
Confidence interval at 95% confidence level in parentheses, p-values reported for one-tailed t-tests
Estimated by maximum likelihood correcting for selection into the follow-up sample
D=(ß/SD controls), where SD is standard deviation for control group within the estimation sample
P-values in the column (4) corrected by RW (2005).We consider 5 hypotheses and food insecurity is excluded.

3.2.3. Socio emotional development

Finally, in Table 6 we present the effects of the intervention on socio-emotional development using the ASQ:SE (total and subscale scores). We do not have ASQ:SE scores at baseline so we construct a composite score using age-standardized birth weight and weight-for-age and height-for-age Z-scores to use as a baseline control. None of the effects on these outcomes was estimated to be significantly different from zero with the only exception of a positive effect on the interaction subscale of 0.13 SD. However, this effect is not statistically significant after the RW adjustment.

Table 6. Results on children's socio-emotional development

VARIABLE	Beta (95% CI)	P Value	RW P Value	D
ASQ:SE total score	0.046 (-0.070,0.161)	0.780	--	0.046
Self-regulation	0.017 (-0.102,0.135)	0.608	0.934	0.017
Compliance	0.015 (-0.109,0.140)	0.595	0.934	0.015
Communication	0.066 (-0.094,0.226)	0.791	0.934	0.066
Adaptive functioning	-0.044 (-0.192,0.105)	0.283	0.725	-0.044
Autonomy	-0.083 (-0.239,0.073)	0.149	0.529	-0.083
Affect	0.022 (-0.080,0.125)	0.666	0.934	0.022
Interaction	-0.131 (-0.272,0.010)	0.034**	0.195	-0.131

Note: ***p<0.01; **p<0.05; *p<0.1; standard errors clustered by town in parenthesis
Confidence interval at 95% confidence level in parentheses, p-values reported for one-tailed t-tests
Estimated by maximum likelihood correcting for selection into the follow-up sample
D=(ß/SD controls), where SD is standard deviation for control group within the estimation sample
P-values in the column (4) corrected by RW (2005).We consider 7 hypotheses and exclude total ASQ:SE score.

3.3. Heterogeneous effects

In this section, we investigate whether the effects of the intervention on children's development varies depending on the characteristics of children and their families.[24] In particular, we assess whether the effects vary by the child's gender, household wealth, maternal education and nutritional status of the child at baseline. We only show results for Bayley-III are these are the most robust effects. However, we assessed heterogeneous effects for socio emotional development and nutritional status, but did not find any differences worth mentioning.

For this analysis, we divide the sample in two groups below and above a certain threshold. In the case of household wealth, we divide the sample of households in the first three wealth quintiles from the households in the top two quintiles. In terms of maternal education, we divide the sample of children with mothers that had completed a high school degree or more from children of mothers with lower educational attainment. Finally, we divide the sample of children below and above the median value of weight for age at baseline (median=0.258).

In Table 7 we report heterogeneous effects of the program on cognitive development. The effects are reported in SD with respect to the control group within estimation sample and RW adjusted P-values are reported in parenthesis. The effect of the program is stronger for children in poorer households, in particular, 0.25 SD on the cognitive subscale, 0.21 SD on receptive language,

[24] These moderating effects are assessed using children's and households' baseline characteristics.

and 0.24 SD on expressive language. There is no statistically significant difference of program effects by household wealth on gross or fine motor development. The effect on the total Bayley-III score is 0.24 SD for children in the three poorest quintiles (compared to 0.13 SD in the total sample), and the effect is not statistically significant for children in the top quintiles. This result is not surprising as there is a vast literature that highlights that the impacts of early childhood interventions are stronger for children at risk.

The results indicate that there is no discernible difference in program impacts by maternal education except in the case of gross motor development which seems to be larger for more educated mothers, 0.3 SD. The results in Table 7 also suggest that the positive effects of the program do not vary significantly by the child's gender. However, there is a higher effect for boys on gross motor development (0.3 SD) and total Bayley-III scores (0.18 SD for boys vs. 0.9 SD for girls). Finally, we report in Table 7 that program impacts were higher in most cases for children with better nutritional status at baseline (by weight-for-age). For example, the effect on receptive language for children with higher weight-for-age at baseline was 0.18 SD, 0.28 SD on expressive language, and 0.25 SD on gross motor development. For total Bayley-III scores the effect for the subsample of children with better weight-for-age at baseline is 0.3 SD (compared with 0.13 in the complete sample). This result suggests a significant complementarity between a child's nutritional status and her learning ability.

Table 7. Heterogeneous program impacts on cognitive development

Outcome variables	High wealth Q4 & Q5	Low wealth Q1-Q3	Mother >= complete secondary education	Mother incomplete secondary or lower	Boy	Girl	BL weight-for-age above mean	BL weight-for-age below mean
	Cognitive and Language Development: Bayley Scales of Infant Development III (BSID III)							
Cognitive	-0.030	0.258	0.144	0.144	0.161	0.063	0.152	0.184
	(0.784)	(0.020)**	(0.206)	(0.206)	(0.231)	(0.255)	(0.175)	(0.051)*
Receptive Language	-0.057	0.209	0.103	0.127	0.152	0.107	0.182	0.076
	(0.784)	(0.020)**	(0.206)	(0.206)	(0.231)	(0.250)	(0.038)**	(0.264)
Expressive Language	0.012	0.245	0.132	0.137	0.173	0.100	0.281	0.053
	(0.692)	(0.019)**	(0.206)	(0.206)	(0.157)	(0.255)	(0.006)***	(0.327)
Gross motor	0.190	0.129	0.302	-0.010	0.216	0.109	0.251	0.038
	(0.149)	(0.189)	(0.011)**	(0.709)	(0.099)*	(0.366)	(0.044)**	(0.521)
Fine motor	0.047	0.021	0.108	-0.007	0.081	0.070	0.159	-0.075
	(0.477)	(0.477)	(0.324)	(0.709)	(0.366)	(0.366)	(0.110)	(0.809)
BSID III Total	0.016	0.243	0.182	0.075	0.200	0.079	0.300	0.015
	(0.442)	(0.005)***	(0.027)**	(0.267)	(0.079)*	(0.206)	(0.005)***	(0.447)

Note: ***p<0.01; **p<0.05; *p<0.1

Effect reported as D=(ß/SD controls), where SD is standard deviation for control group within the estimation sample

RW-adjusted P-values in parentheses. The multiple hypotheses considered are: three cognitive development measures x 2 heterogeneous characteristics (e.g., high/low wealth); and 2 motor development measures x 2 heterogeneous characteristics.

Estimated by maximum likelihood correcting for selection into the follow-up sample

3.4. Effects on intermediate outcomes

In this section we examine the effects of the intervention on intermediate outcomes of the children's households, which we hypothesized could have been affected by the intervention: parental involvement and the home's learning environment, the use of violent discipline, maternal self-efficacy and social support (see Appendix 3).

In Table 8 we present the effects of the program on secondary outcomes using the intermediate outcomes described in detail in section 2.4. The effect of the program on the number of adult books, magazines and newspapers are home was of 0.39 SD, on the number of toy sources at home was of 0.36 SDs, 0.28 SDs on the number of varieties of play materials at home and 0.17 SDs on the number of varieties of play activities with mother, father or any other adult care provider at home. Using the principal component resulting from factor analysis on all FCIs items, we report an effect of the program of 0.37 SDs. We also report an effect of 0.11 SDs on social support that is not statistically significant after the RW adjustment.

Table 8. Results on household and parental intermediate outcomes

VARIABLE	Beta (95% IC)	P Valor	RW P Valor	D
No. of adult books, magazines and newspapers	1.507 (0.808,2.206)	0.000***	0.001***	0.396
No. of Toy Sources	0.229 (0.138,0.319)	0.000***	0.001***	0.359
No. of varieties of play materials	0.495 (0.281,0.709)	0.000***	0.001***	0.281
No. of varieties of care activities over past 3 days	0.018 (-0.033,0.069)	0.499	0.484	0.042
No. of varieties of play activities over past 3 days	0.427 (0.086,0.767)	0.014**	0.034**	0.170
FCI Home Environment Quality (PCA)	0.378 (0.248,0.508)	0.000***	0.000***	0.378
Parental Knowledge (Raw Score)	0.025 (-0.502,0.551)	0.927	0.999	0.007
Uses Violent Discipline	-0.027 (-0.081,0.027)	0.331	0.864	-0.081
Social Support DUKE UNC-11 Scale	0.945 (-0.136,2.026)	0.087*	0.464	0.110
Self-Efficacy Score Above Median	-0.041 (-0.089,0.006)	0.090*	0.464	-0.096
Presence of Depressive Symptoms (CESD-10)	0.003 (-0.052,0.057)	0.919	0.999	0.007

Note: ***p<0.01; **p<0.05; *p<0.1. Standard errors clustered by town in parenthesis
Estimated by maximum likelihood correcting for selection into the follow-up sample
D=(ß/SD controls), where SD is standard deviation for control group within the estimation sample
P-values in the column (4) corrected by RW (2005).We consider 5 hypotheses for the household environment (FCI principal component excluded) and 5 hypotheses for all the other outcomes.

We also report negative program effects on maternal self-efficacy and depression symptoms. However, neither of these is statistically significant after RW adjustment. This might be related to the fact that as parents understand better the process of early skill formation, they become more aware of their actions and perhaps, more critical about their own role as models. We do not report statistically significant effects on the use of violent discipline at home.

3.5. Effects on child development by duration of exposure to the program

In this section we assess whether the impacts of the program on child development change as the duration of program exposure varies. In particular, we present two different sets of results. First, we compare children in the treatment group that participated in at least one of the activities of the program with children in the control group. This information is available from attendance registries in the treatment group gathered by FAMI program facilitators as part of the program's supervision protocol. Second, we compare children that participated in at least 21 activities offered by the program (the median number of activities, including group sessions and home visits) with children in the control group. We present results on cognitive and motor development measured by Bayley-III given this results are the most robust.

Duration of program exposure is not random because some parents assigned to the treatment group could have chosen to participate in more or less program activities for reasons that could be, in turn correlated with children's development such as socioeconomic status, maternal motivation or her knowledge about early child development. Thus, we estimate the impact of the duration of exposure to the program using a two-staged least squares procedure and use random assignment as the instrumental variable to predict duration of exposure.[25] Random assignment to treatment is clearly exogenous and it effectively predicts children's duration of exposure to the program.[26]

In Table 9 we present the results for this first case which compares children in the treatment group who participated in at least one program activity with children in the control group. As reported earlier, 78% of children in the treatment group assessed at follow-up were registered at least once in program attendance lists (both, groups and home visits). This means that 22% of children in the treatment group (141 children) that were found at follow-up did not participate in any activity while the program was operating.

[25] La F-test for significance of the instrumental variable is always above 25.
[26] The results of the first stage are presented in Appendix 9.

Table 9. Impacts on Bayley-III for children who participate in at least one program activity

VARIABLE	Beta (95% IC)	P Value	RW P-Value	D
Cognitive	0.159 (-0.041,0.360)	0.060*	0.063*	0.159
Receptive language	0.198 (0.012,0.385)	0.019**	0.050**	0.198
Expressive language	0.224 (-0.003,0.451)	0.027**	0.051*	0.224
Gross motor	0.216 (0.006,0.425)	0.022**	0.046**	0.216
Fine motor	0.126 (-0.079,0.330)	0.114	0.120	0.126
Total Bayley-III score	0.144 (-0.061,0.350)	0.084*	--	0.144

Note: ***p<0.01; **p<0.05; *p<0.1. Standard errors clustered by town in parenthesis
Estimated by 2SLS using random assignment as an instrumental variable for duration of program exposure
D=(ß/SD controls), where SD is standard deviation for control group within the estimation sample
The number of children found at follow-up that attended at least one program activity in the treatment group is 487.
P-values in the fourth column are corrected by RW (2005). We consider 3 hypotheses for cognitive development and 2 hypotheses for motor development. Total Bayley-III score excluded.
Selection into the follow-up sample is corrected by including the inverse mills ratio resulting from the selection model in the estimating equation.

The results indicate that the effects of the program on cognitive and motor development are higher than average effects estimated on the entire sample for children that participated in at least one program activity. In particular, the effect on the cognitive subscale increases to 0.16 SD, on receptive language is 0.2 SD, expressive language 0.22 SD and gross motor development 0.21 SD. The effect on total Bayley-III scores is 0.14 SD. In other words, the effects increase slightly more than 50% with respect to the impacts on the entire sample.

Similarly, in Table 10 we present effects that result from comparing children in the treatment group who attended at least 21 activities with the control group. In particular, 353 children in the treatment group that were found at follow-up out of 628 attended at least 21 program activities. The results show that program effects increase as the number of program contacts increases. In this case, the effect on the cognitive scale is of 0.20 SD, receptive language 0.24, expressive language is 0.28 SD and gross motor is 0.27 SD. The effect on total Bayley-III scorer is 0.18 SD. In other words, the effects are basically twice as big with respect to the estimates on the entire sample.

In Appendix 10 we present basic differences between children that participated in at least 21 program activities in the treatment group and children with fewer contacts. Children in the latter group are older so they could have lost age-eligibility sooner. These children have younger mothers (25.4 years vs. 26.7 years), with higher educational attainment (9.1 vs. 8.5 years) and lower probability of living with their fathers (66% vs. 71%). On the other hand, children with lower duration of program exposure reside in households with better quality of the learning environment by FCI at baseline. For example, the number of personal care routines is 4.8 during the last three days

compared with 4.6 in the group of children with higher duration of program exposure. Parents of children in the former group also report to have used violent discipline more often (4% vs. 1%). Their initial development as measured by the WHO gross motor milestones indicates that they were better (2.87/6 milestones vs. 2.52/6 milestones achieved). Finally, among the reasons reported by parents for leaving the FAMI program, children with fewer program contacts report with higher probability that they dropped out because they moved (20% vs. 5%), and children with higher program exposure report more often that they dropped out because they reached the maximum age for eligibility (63% vs. 38%). In sum, children with a fewer program contacts were better off at baseline in terms of development and household characteristics and the main reason for leaving the program is not age-eligibility.

Table 10. Impacts on Bayley-III for children who participate in at least 21 program activities

VARIABLE	Beta (95% CI)	P Value	RW P-Value	D
Cognitive	0.200 (-0.047,0.447)	0.056*	0.061*	0.200
Receptive language	0.245 (0.019,0.470)	0.017**	0.053*	0.245
Expressive language	0.279 (-0.001,0.558)	0.025**	0.053*	0.279
Gross motor	0.270 (0.010,0.529)	0.021**	0.046**	0.270
Fine motor	0.157 (-0.097,0.412)	0.113	0.116	0.157
Total Bayley-III score	0.182 (-0.073,0.436)	0.081*	--	0.182

Note: ***p<0.01; **p<0.05; *p<0.1. Standard errors clustered by town in parenthesis
Estimated by 2SLS using random assignment as an instrumental variable for duration of program exposure
D=(ß/SD controls), where SD is standard deviation for control group within the estimation sample
The number of children found at follow-up that attended at least one program activity in the treatment group is 353.
P-values in the fourth column are corrected by RW (2005). We consider 3 hypotheses for cognitive development and 2 hypotheses for motor development. Total Bayley-III score excluded.
Selection into the follow-up sample is corrected by including the inverse mills ratio resulting from the selection model in the estimating equation

In both, Tables 9 and 10, we could still be facing a downward bias of the estimates, because children could have left the program long before we collected follow-up assessments. This would introduce an attenuation bias because mothers could have lost motivation and stopped practicing program activities with their children if a long period elapsed between their last program participation and follow-up data collection. In Appendix 11 we show information about the number of months elapsed between a child's last attendance registry in the program and the date in which he or she was assessed at follow-up. The average duration of this period is 5.2 months (SD=3.22), and close to 46% of children in the treatment group ceased program participation 5 months or more prior to follow-up data collection.

4. Discussion and conclusions

In a cluster randomized controlled trial at scale, we found that the implementation of a structured curriculum in public parenting program services in rural areas plus pre-service and in-service training for program facilitators and a nutritional intervention, resulted in a significant improvement of 0.15 SD in children's cognitive scores, 0.11 SD in receptive language, 0.14 SD in expressive language and 0.14 SD in gross motor development. We also report a reduction in the risk of stunting of -0.13 SDs. We do not find any effects on social-emotional development. We also report positive and statistically significant effects on the number of toy materials at home (0.36 SD), the number of varieties of play materials (0.28 SD), and the number of varieties of play activities with adults at home over the past three days (0.17 SD). Most importantly, the study showed that this quality enhancement can be effectively implemented at scale as part of an existing public early childhood program in rural areas.

This study offers important new evidence that quality low-cost enhancements of already existing programs at scale which leverage on local low-skilled human resources can be effective. The program cost about $320 per year per child ($28 in pedagogical materials, $82 in supervision and $212 in additional nutritional supplementation) plus $12 one-time per child cost in FAMI pre-service training. The cost of the unenhanced FAMI program is about $310 per child per year. That means that the quality enhancement excluding the nutritional supplement corresponds to approximately 40% of the total cost. The cost of the enhanced program excluding the nutritional supplement would be 1.7 monthly minimum wages per child per year or 2.5 monthly minimum wages per year including the nutritional supplement. The cost per child per year in center-based care in Colombia is approximately $1,100 or 4.4 monthly minimum wages per child per year.

Innovation

Earlier efficacy studies (Walker et al., 2011) have established that stimulation can substantially improve the cognitive development of infants in low income and middle income countries. Compared to these studies, we contribute in the following ways. First, we mimicked the workings of a scaled up program by using the infrastructure of a national parenting program to deliver a structured early stimulation curriculum and pre-service and in-service training for program facilitators, and by implementing the program across 46 municipalities, spread across a large scale in three departments of Colombia, covering an area half the size of the state of New York and a third of the size of Ecuador. Second, it contributes to the understanding of how to deliver effective parenting programs through group meetings instead of individual home visits. Most of the well-known effective parenting programs with rigorous evaluations are based on individual home visits (Grantham-McGregor et al., 1997; Olds et al., 1986a, 1986b and 1994) while the program we work with is mostly based on group sessions. The results are important because groups can be more cost-effective than individual visits.

Third, this study contributes to the understanding of how to best serve rurally disperse populations in a cost-effective way by providing itinerant training and coaching to front line workers. And finally, it allows us to assess the barriers, challenges and opportunities of improving quality of parenting programs at scale by leveraging on local low-skilled human resources and low cost enhancements. Though the trial was not fully at scale, it was designed to be scalable. The intervention was spread over 46 towns in three regions and used early childhood program facilitators who are readily available nationally. The results presented in this study, indicate that the curriculum in addition to pre-service and in-service training provided to these front line workers was effective in attaining positive effects on children's development so that the program could be scaled up at the national level.

Relevance

This study also shows that parental behavior changed and home stimulation increased, which suggests that the benefits might be sustainable and might 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 (Cunha and Heckman, 2013; Cunha, Heckman and Schennach, 2010).

The size of the effects is small relative to impacts of comparable early childhood interventions in the Latin-American region. Nores and Barnett (2010) report effect sizes of about 0.25 SD for continuous outcomes in verbal ability and/or cognition for early education programs. However, it is worth noting that there are three key features of this particular study that makes us believe that the estimates reported here are lower bounds of the potential of the intervention. First, the control group had access to the basic program without the quality enhancement - unlike similar pilots in which the control group does not receive anything. Second, the children and households in our sample were not as vulnerable at baseline in terms of sociodemographic characteristics or initial developmental outcomes as expected. It is well known that early childhood interventions are particularly effective on vulnerable populations, especially if the intervention is a marginal enhancement and not a stand-alone intervention. Finally, and most importantly, given that we worked with an already existing public parenting program, we did not have full control of all the elements that we might expect to have an impact on children. In fact, we faced various challenges such as very high turnover rates of children and lower duration of exposure to the program than we expected. In particular, treated children who participated in at least one group session, attended an average of 19 group sessions out of a total 45 sessions (average duration) that they could have attended. Together, these issues imply that the intervention could have had a higher impact if the FAMI program had better targeting and higher compliance rates.

Limitations

First, we face the typical limitations of a randomized controlled trial, and in particular, issues related to external validity. The sample of municipalities is located in three departments in central Colombia (Cundinamarca, Boyacá, and Santander). We restricted the study sample to these regions because they had a large number of towns that complied with the eligibility criteria for the study, had a larger number of FAMI units per town than other departments and are also much less geographically disperse than other departments. We also selected this sample because the municipalities in these regions are similar in terms of their cultures and customs. This would make the use of a single curriculum —and associated materials such as pictures and books—more feasible, which was particularly important given the ethnic and geographical diversity of Colombia. This implies that it is not clear that our curriculum (as it is currently designed) would be suitable for other regions of Colombia, specifically, for diverse communities such as those with afro-Colombians and indigenous populations.

Second, our randomized evaluation design was threatened by the fact that towns in the originally selected sample were transited out of the FAMI program and had to be removed and replaced. Although we replaced these towns from a randomly ordered list, we do find statistically significant differences between study groups in some dimensions at baseline, which might imply that this replacement procedure was not ideal. We have controlled for these differences at baseline in our main estimation to ameliorate the possibility of confounding factors.

Third, the characteristics of our sample at baseline indicate that this population was not as vulnerable as it was anticipated. In fact, we report that children and their families are better off in a variety of dimensions with respect to other comparable studies in the country. It is well known that early childhood interventions are particularly effective on vulnerable populations, especially if the intervention is a marginal enhancement rather than a stand-alone program. For this reason, we expect our results to be lower bounds of the potential effects of our program if targeting could be improved.

Finally, it is possible that developmental outcomes might have been collected with significant measurement error, particularly those that that are reported by parents such as Ages and Stages for socio-emotional development, and that these specific measures lack enough sensitivity▪

Acknowledgment

This study was funded by Grand Challenges Canada and *Fundación Éxito*. We thank the *Instituto Colombiano de Bienestar Familiar* (ICBF), ICBF program supervisors in Cundinamarca, Santander and Boyacá and program coordinators at *Fundación Éxito* for their cooperation in facilitating the intervention; the FAMI program facilitators, children and families who willingly participated in this study; all the study staff including our nine tutors and María Lucía Gómez our field manager; all the research staff including Santiago Lacouture, Alejandro Sánchez and Sara Ramírez; the data collection team from IQuartil; and the experts from Grand Challenges Canada who gave their time to provide technical advice during the course of this study.

References

Aboud, F.E. and A.K., Yousafzai (2015) Global Health and Development in Early Childhood. Annual Review of Psychology 66 (January):433-57.

Andersen E. M., Malmgren J. A., Carter W. B., Patrick D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). American Journal of Preventive Medicine, 10, 77-84.

Álvarez M, Estrada A, Melgar-Quiñonez H, Fonseca Z. (2008). Adaptación y validación interna y externa de la Escala Latinoamericana y Caribeña para la medición de seguridad alimentaria en el hogar - ELCSA- Colombia: Componente adaptación lingüística de ELCSA. Medellín: ICBF, PMA, FAO, DANE, UDEA.

Andrew, A., Attanasio, O., Bernal, R., Cardona, L., Krutikova, S., Martínez, D., Medina, C., Peña, X., Rubio-Codina, M., and Vera-Hernández, M. (2016). Evaluation of centers of infant development: an early years intervention in Colombia. Unpublished manuscript, Institute for Fiscal Studies.

Attanasio, O. P., Fernández, C., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M. (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early childhood development program in Colombia: Clustered randomized controlled trial. British Medical Journal, 349, g5785.

Belinski, S. and N. Schady (2015) The early years: child well-being and the role of public policy. Inter-American Development Bank. Palgrave Macmillan.

Bernal, R. (2016) Todo empieza en casa: el papel de la familia. En: Eds. Berniell, L. and De la Mata, L. Más habilidades para el trabajo y la vida: los aportes de la familia, la escuela, el entorno y el mundo laboral. Reporte anual Banco de Desarrollo de América Latina. ISSN: 980-6810-01-5.

Bernal, R. (2015). The impact of a vocational education program for childcare providers on children's well-being. Economics of Education Review, 48, 165–183

Bernal, R. (2013) The Cost of Early Childhood Policy in Colombia. Unpublished manuscript Universidad de los Andes.

Bernal, R., O. Attanasio, X. Peña, M. Vera-Hernández (2016) The Effects of the Transition from Home-based Community Nurseries to Childcare Centers on Children in Colombia. Unpublished Manuscript Universidad de los Andes.

Bernal, R. and Camacho, A. (2014). Early childhood policy in the context of equity and social mobility in Colombia. In A. Montenegro and M. Meléndez (Eds.), *Equidad y Movilidad Social: Diagnósticos y Propuestas para la Transformación de la Sociedad Colombiana*. Editorial Uniandes.

Bernal, R. and Camacho, A. (2011). The importance of early childhood programs in Colombia. Imprenta Nacional de Colombia.

Bernal, R. and Fernández, C. (2013). Subsidized childcare and child development in Colombia: Effects of Hogares Comunitarios de Bienestar as a function of timing and length of exposure. Social Science and Medicine, 97, 241–249.

Bernal, R. and Quintero, C. (2014). Characterization of children younger than five in Colombia based on the Longitudinal Colombia Survey ELCA. Manuscript Universidad de los Andes.

Broadhead, W., Gehlbach, S., De Gruy, F., & Kaplan, B. (1988). The Duke-UNC Functional Social Support Questionnaire: Measurement of Social Support in Family Medicine Patients. Medical Care, 26(7), 709-723.

Comisión Intersectorial de la Primera Infancia. (2012). Estrategia de Cero a Siempre: Presentación. Mayo 3, 2015. Consejería Presidencial para la Primera Infancia.

Cunha F, Heckman J. (2013) 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, 221-69.

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

ELCSA Scientific Committee (2012). Escala Latinoamericana y Caribeña de Seguridad Alimentaria (ELCSA). Manual de uso y aplicación. Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). ISBN 978-92-5-307354-2.

Fernald, L. C. H., Gertler, P. J., and Neufeld, L. M. (2008). Role of cash in conditional cash transfer programmes for child health, growth, and development: An analysis of Mexico's Oportunidades. *Lancet*, 371(9615), 828–837.

Fernald, L. C. H., Kariger, P., Engle, P., and Raikes, A. (2009). *Examining Early Child Development in Low-Income Countries.* Washington DC: World Bank.

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

Handal, A. J., Lozoff, B., Breilh, J., and Harlow, S. D. (2007). Sociodemographic and nutritional correlates of neurobehavioral development: A study of young children in a rural region of Ecuador. Revista Panamericana de Salud Pública, 21, 292–300.

Heo, K. H., Squires, J., and Yovanoff, P. (2007). Cross-cultural adaptation of a pre-school screening instrument: Comparison of Korean and US populations. Journal of Intellectual Disability Research, 52(3), 195–206.

Grantham-McGregor, S. M., Walker, S. P., Chang, S. M., & Powell, C. A. (1997). Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children. American Journal of Clinical Nutrition 66, 247-253.

McKay, H., Sinisterra, L., McKay, A., Gomez, H., & Lloreda, P. (1978). Improving cognitive ability in chronically deprived children. Science, 200, 270-278.

McPhee, D. (1981). Knowledge of infant development inventory. Chapel Hill, NC: University of North Carolina.

Nores, M., R. Bernal and S. Barnett (2016) Center-Based Care for Infants and Toddlers: The aeioTU Randomized Trial. Unpublished manuscript National Institute for Early Education Research.

Nores, M. and R. Bernal (2014) The AeioTu Early Childhood Longitudinal Study Effects of ECD intervention on family and community. Progress report National Institute for Early Education Research and Universidad de los Andes.

Nores, M. and Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under)Investing in the very young. Economics of Education Review, 29, 271–328.

Olds, D., Henderson, C., Tatelbaum, R., & Chamberlin, R. (1986a). Improving the delivery of prenatal care and outcomes of pregnancy: a randomized trial of nurse home visitation. Pediatrics, 77, 16-28.

Olds, D., Henderson, C., Chamberlin, R., & Tatelbaum, R. (1986b). Preventing child abuse and neglect: a randomized trial of nurse home visitation. Pediatrics, 78, 65-78.

Olds, D., Henderson, C., & Kitzman, H. (1994). Does prenatal and infancy nurse home visitation have enduring effects on qualities of parental caregiving and child health at 25 to 50 months of life?. *Pediatrics, 93*(1):89-98

Pedersen, F. A., Bryan, Y., Huffman, L., & Del Carmen, R. (1989, April). Constructions of self and offspring in the pregnancy and early infancy periods. Paper presented at the Meetings of the Society for Research in Child Development, Kansas City, Missouri.

Porter, C. L., & Hsu, H. (2003). First-time mothers' perceptions of efficacy during the transition to motherhood: Links to infant temperament. Journal of Family Psychology*, 17*, 54-64.

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

Romano JP, Wolf M. 2005. Stepwise Multiple Testing as Formalized Data Snooping. Econometrica; 73(4):1237-1282

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 (2015) J Human Resources Spring vol. 50 no. 2 464-483.

Squires, J. and Bricker, D. (2009). Ages & Stages Questionnaires, Third Edition (ASQ-3). Baltimore, MD: Paul H. Brookes Publishing.

Squires, J., Bricker, D., and Twombly, E. (2009a). Technical Report on ASQ:SE. Baltimore, MD: Paul H. Brookes Publishing.

Squires, J., Bricker, D., and Twombly, E. (2009b). Ages & Stages Questionnaires: A parent-completed child monitoring system. Baltimore, MD: Paul H. Brooks Publishing.

Squires, J., Bricker, D., Heo, K., and Twombly, E. (2002). Identification of social-emotional problems in young children using a parent-completed screening measure. Early Childhood Research Quarterly, 16(4), 405–419.

Super, C. M., Herrera, G., and Mora, J. O. (1990). Long-term effects of food supplementation and psychosocial intervention on the physical growth of Colombian infants at risk of malnutrition. Child Development, 61, 29–49.

UNICEF (2010). Child Disciplinary Practices at Home: Evidence from a Range of Low- and Middle-Income Countries, New York.

Vargas-Baron, E. (2006). Impuestos sobre la Nómina Destinados al Desarrollo Infantil: Lecciones aportadas por Colombia. UNESCO Notes on the Importance of Early Childhood Policies, p. 35.

Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, et al. (2011) Inequality in early childhood: risk and protective factors for early child development. Lancet2;378:1325-38.

Walker SP, Chang SM, Vera-Hernandez CM, Grantham-McGregor S. (2011) Early childhood stimulation benefits adult competence and reduces violent behaviour. Pediatrics 127:849-57.

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

Walker, S., Chang, S., Powell, C. A., and Grantham-McGregor, S. M. (2004). Psychosocial intervention improves the development of term low-birth-weight infants. *American Society for Nutritional Sciences*, 134(6), 1417–1423.

WHO Multicentre Growth Study Group (2006) WHO Motor Development Study: Windows of Achievement for Six Gross Motor Development Milestones. Acta Pediatrica Suppl 450: 86-95.

WHO Multicentre Growth Reference Study Group (2006). WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development, p. 312. Geneva: World Health Organization.

WHO Multicentre Growth Reference Study Group (2007). WHO Child Growth Standards: Head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: Methods and development. Geneva: World Health Organization

Yoshikawa, H., Ponguta, L., Nieto, A., van Ravens, J., Portilla, X., Rebello Brito, P., and Leyva, D. (2014). Evaluating mechanisms for governance, finance and sustainability of Colombia's comprehensive early childhood development policy De Cero a Siempre. Unpublished manuscript, New York University.

Supplemental Materials.

Appendix 1.

Detailed Description of the Interventions

The design of the new curriculum aims at strengthening the child development component of the existing FAMI program in order to improve the development of the child (in particular, language, cognitive, motor, social-emotional development) and strengthening the nutrition component of the existing FAMI program by providing psychoeducation around feeding and nutrition and a nutritional supplement. It also aims at improving mothers' knowledge, practices and enjoyment of bringing up her child and improving mothers' self-esteem and mental health. Given that the program is delivered through group meetings and home visits, this intervention includes two complementary curricula. In both cases, the components, actions and activities used to promote better maternal child rearing practices are the same. These include making the mother the agent of change and empower her to improve her child's development, demonstrating the use of age-appropriate play materials and activities and providing opportunities to practice age appropriate activities and provide supportive feedback. It also offers the opportunity to set up a toy and child library to use at home, prove opportunities for mothers to make low-cost toys and practice using them in ways that promote specific child development goals and opportunities for social support, sharing of experiences, and group problem-solving. The program also aims at increasing mothers' motivation to improve her child's development by helping her to understand how her actions can make a difference to her child, improving mother's self-esteem with praise, support and encouragement, training in sensitive and responsive parenting and appropriate behavior management and encouraging positive mother-child interactions and preventing child maltreatment.

Most of the program content was delivered through the group visits as they were held on a weekly basis. However, the home visits provided the opportunity to reinforce the material covered and to personalize to specific needs of the family as necessary. During home visits, FAMI mothers gave and received feedback about the group sessions and reinforced mothers' participation in all program activities, reinforced the key messages learnt in the group meetings, made sure that the activities that the mother and child were bringing home from the group sessions were developmentally appropriate and introduced additional activities targeted to the child's age and developmental level if necessary. FAMI mothers also discussed with mothers how to integrate the program activities discussed and practiced in the group sessions into every day family activities, identified materials in the home that could be used to promote child development and helped families to use their home environment to promote child development. Mothers were encouraged appropriate mother-child interaction with a strong focus on promoting children's language development in a variety of ways and engaged in problem solving around attendance at group

sessions, the ability to do activities at home in addition to any individual concerns or issues that the families could face. Home visits were, on average, one hour long and took place once a month with each participant family.

The group curriculum aimed at providing opportunities to share parenting experiences in a group setting, provide opportunities to discuss and practice effective child rearing skills and positive interactions with children, and demonstrate and practice the use of age-appropriate play materials and language activities. Group meetings also provided the opportunity to discuss how these activities help in children's development, and show mothers how to make simple toys so that each family could set up a toy library for home use.

Mothers were asked to attend one group meeting according to the age of their children. Specifically, group meetings were delivered for pregnant and lactating women with children up to 6 months, mothers with children from 6 to 11 months and mothers with children aged 1-2 years. We expected mothers of children from 6 to 24 months of age to attend four meetings per month and pregnant and lactating with children up to 6 months to attend three meetings per month. However, in practice, this did not always occur and the curriculum had been designed so that it could be delivered to groups with children over the entire age range. Each group session is structured in six different moments: arrival and free play; feedback from the previous group session (10 minutes); song (5 minutes); demonstration and practice of age-appropriate play activity and language activity for the week with material that will be taken home (30 minutes); discussion around a parenting theme or activity (15 minutes); review of the session to ensure that mothers understand the activities and commitment to practice with children at home (10 minutes), and finally, they share a snack. If mothers cannot attend their group according to their children's age, then play and language activities are divided into age bands (birth-5 months, 6-11 months and 1-2 years). The themes for discussion during the group meetings include issues such as the importance of spending time playing with the child, praising the child, talking to the child, things to do at bath time or mealtimes, learning to trust, understanding the child's feelings, teaching the baby about her environment, and child behavior.

The curriculum involves materials to be used during home visits and in the group meeting. The curriculum includes discussion topics or key parenting messages, a selection of age appropriate activities to promote child development using simple play materials (e.g. home-made toys, materials in the home, puzzles) and activities to promote children's language development (using games, books, pictures and a using everyday activities to encourage mothers to talk more with their child). The curriculum also includes a set of nutrition cards relevant to their child's age that are discussed with the mother during each home visit. The complete kit of materials has a cost of USD 27[27] per child per year.

[27] Computed at the average exchange rate 2015-2016.

In addition to the set of activities and materials, the qualification of the FAMI program also included a coaching component (in-service training) to support and maintain the quality of home visits and group meetings. Shifting away from a supervision model, the new approach consists of a team of facilitators who provided the initial pre-service training and then continued to provide in-service training and support during the implementation period. Professional tutors trained and supervised by the research team were in charge of training FAMI mothers. Training was provided sequentially by town. All FAMI mothers were trained simultaneously in a given town for an average of 3.5 weeks and 85 hours.[28] Towns with a larger number of FAMI units spent up to 170 hours of training in cases with more than 10 FAMI units per town. On average, the number of training hours per FAMI mother was 19 hours (total number of hours per town divided by FAMI units in that town). Training included instructional time (10%), toy making workshops (20%), practice in pairs (55%), practice with children under supervision (10%) and individual work (5%). The one-time cost of pre-service training per FAMI mother was of about USD 113 or USD 11 per child. [29]

After training was finalized, tutors coached FAMI mothers continuously throughout the duration of the intervention. Tutors observed one group session and one home visit in each supervision round which took place approximately every 5 weeks. Each tutor was in charge of 5 towns and 19 FAMI mothers, on average. They also facilitated a group meeting of FAMI mothers in each town to discuss and share positive experiences and challenges and engage in problem-solving. The facilitators were supervised by an intervention supervisor who conducted visits with each tutor every 2 months. The cost of coaching was around USD 82 per month per FAMI mother or approximately USD 8 per child per month.

In addition to the introduction of the early stimulation curriculum, the intervention also includes a nutritional component. It is comprised of the delivery of a monthly nutritional supplement to FAMI participants, and psychoeducation around feeding and nutrition during group meetings and home visits. The nutritional supplement corresponds to 35% of daily calorie intake requirements for pregnant women, breastfeeding mothers and children younger than 2 years of age (for 30 days). The cost of the package is USD 26 per month including shipping costs. It contains tuna, sardines, canola oil, and whole milk with iron supplement, beans and lentils. In terms of educational contents, we developed a cooking book that takes into account the socioeconomic characteristics of households in our sample, brochures used to handle and classify foods and 19 nutrition cards that are discussed the mother during each home visit. Mothers receive a nutrition card relevant to their child's age at these monthly home visits. The topics covered include things like breastfeeding, bottle-feeding, breastmilk extraction and storage, weaning, hygiene, finger foods, menu ideas, mealtimes, and chatting while feeding.

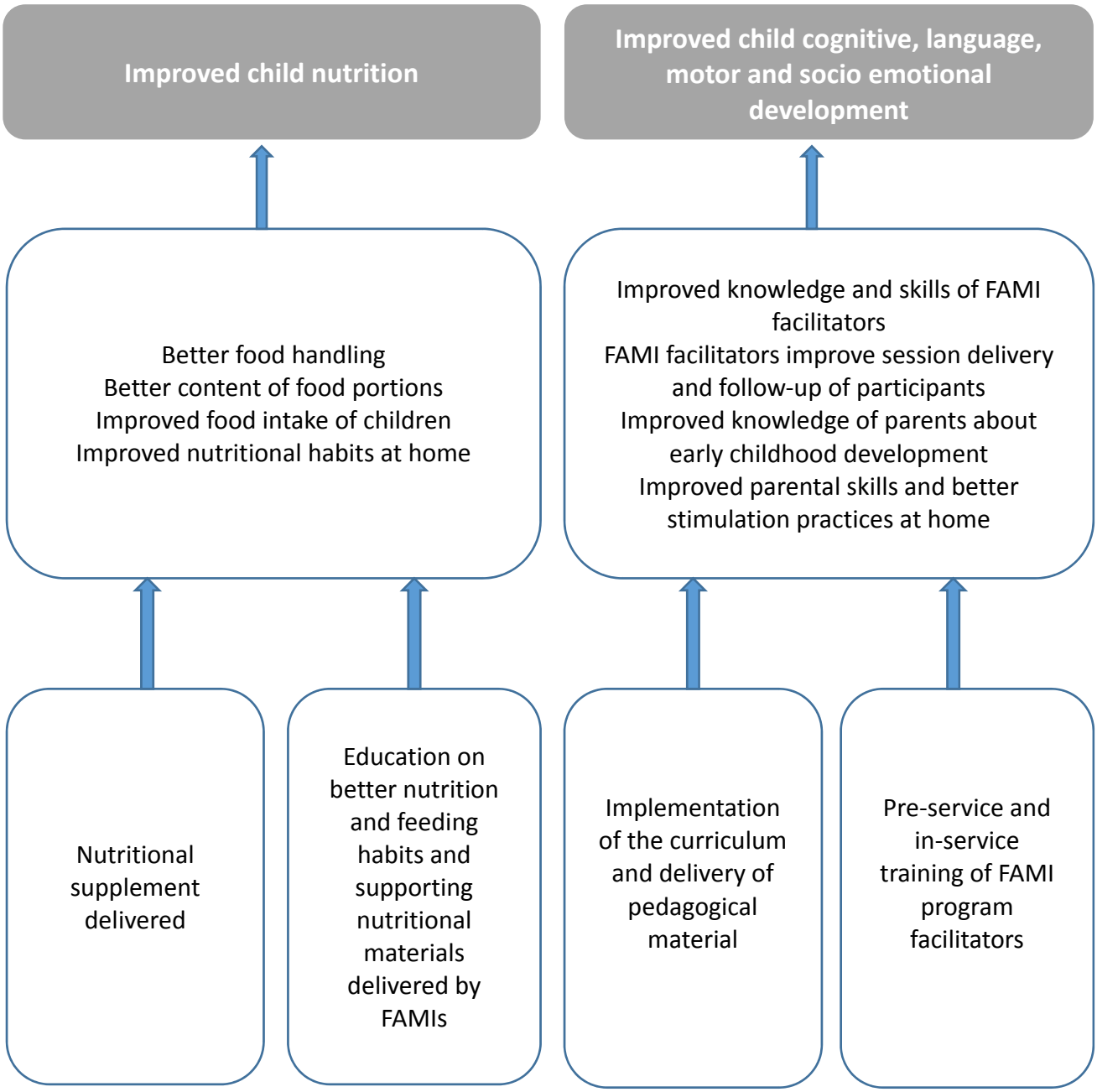
[28] This was done in two stages: an initial stage of 2 weeks and a second stage of 1.5 on average two months later.
[29] Both computed at the average exchange rate 2015-2016 ($2,800 COP/USD).

Appendix 2.

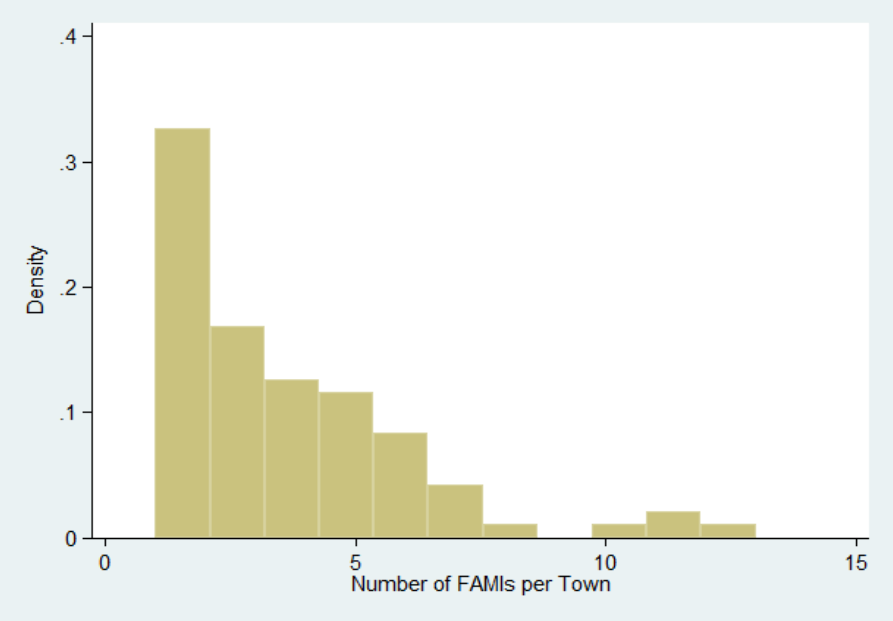
Content of the nutritional supplement of the program.

	Nutritional content prior to program					Additional nutritional content					Total nutritional content				
Individual	Calories	Protein	Calcium	Iron	Folic acid	Calories	Protein	Calcium	Iron	Folic acid	Calories	Protein	Calcium	Iron	Folic acid
Pregnant woman	16,1%	23,1%	33,8%	22,2%	56,6%	19,0%	31,1%	24,8%	11,0%	36,9%	35,1%	54,2%	58,5%	33,2%	93,4%
Father	3,2%	4,1%	0,9%	7,1%	7,8%	8,0%	18,2%	20,1%	22,3%	42,4%	11,2%	22,2%	21,0%	29,4%	50,2%
Child	7,6%	12,6%	1,5%	11,0%	17,2%	19,1%	56,2%	32,2%	34,7%	94,2%	26,6%	68,8%	33,7%	45,7%	88,0%

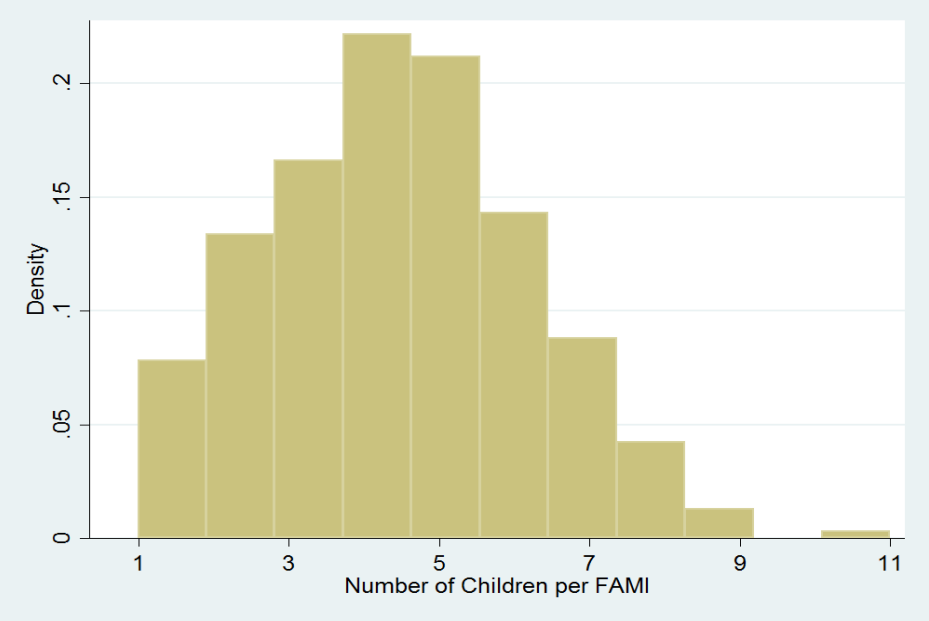
Appendix 3. Theory of change



Appendix 4. Distribution of clusters



Mean=4.3; SD=1.9



Mean=3.9; SD=2.3

Appendix 5. Attrition analysis

VARIABLES	(1) Attrition	(2) Attrition
ITT	0.0381*	0.0427**
	(0.0212)	(0.0210)
Age in months		0.0167**
		(0.0067)
Age sq.		-0.0010*
		(0.0006)
Gender (Male)		-0.0177
		(0.0127)
Number of previous pregnancies		-0.0059
		(0.0147)
Number of Siblings		-0.0105
		(0.0149)
Wealth Quintile = 2		-0.0037
		(0.0272)
Wealth Quintile = 3		-0.0208
		(0.0283)
Wealth Quintile = 4		-0.0247
		(0.0280)
Wealth Quintile = 5		-0.0447*
		(0.0256)
Maternal years of educaction		-0.0009
		(0.0026)
Father present		-0.0466**
		(0.0201)
Household size		-0.0066
		(0.0071)
Maternal verbal ability (PPVT)		-0.0018*
		(0.0011)
Constant	0.0673***	0.1665***
	(0.0115)	(0.0551)
Observations	1460	1460
R-squared	0.0046	0.0336
F - Stat	3.237	2.682
Prob > F	0.0755	0.00266

Standard errors clustered by town in parenthesis
*** p<0.01, ** p<0.05, * p<0.1

Appendix 6. Selection model into the follow-up sample

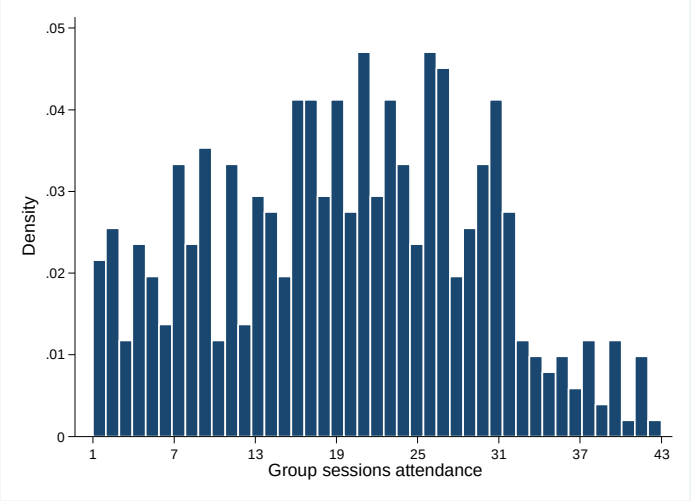
VARIABLES	Found at FU
ITT	-0.2994**
	(0.1339)
Age in months	-0.1268**
	(0.0589)
Age sq.	0.0081*
	(0.0049)
Gender (Male)	0.1143
	(0.0877)
Birth Order	0.1109**
	(0.0451)
Wealth Quintile = 2	-0.0517
	(0.1600)
Wealth Quintile = 3	-0.0059
	(0.1739)
Wealth Quintile = 4	-0.0235
	(0.1834)
Wealth Quintile = 5	0.1502
	(0.1748)
Maternal Education = 2, Complete or incomplete secondary	-0.1006
	(0.1236)
Maternal Education = 3, Higher than secondary	0.0892
	(0.2029)
Father present	0.3168***
	(0.1093)
Household size	0.0576
	(0.0490)
Maternal verbal ability (PPVT)	0.0158**
	(0.0078)
Conscientiousness (Z) [Mother]	-0.0906
	(0.0600)
Time to Town Hall (Hours)	-0.1351***
	(0.0477)
Constant	1.1920***
	(0.3475)
Observations	1460
Wald Test	76.45
Prob > chi2	0.000

Standard errors clustered by town in parenthesis

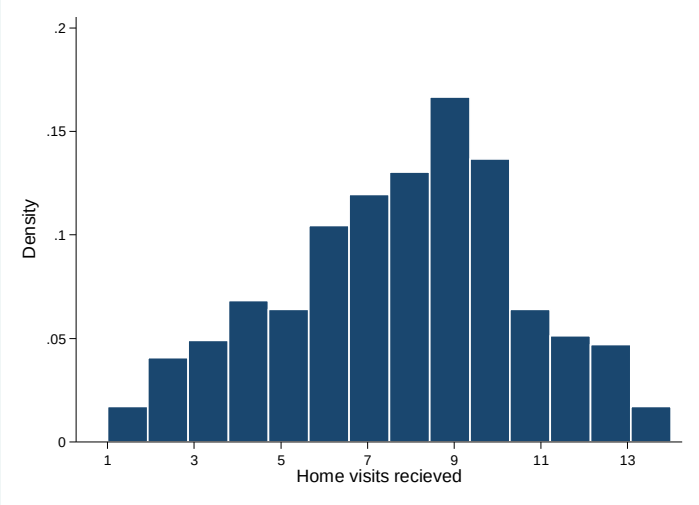
Probit Model

*** p<0.01, ** p<0.05, * p<0.1

Appendix 7. Effective program participation

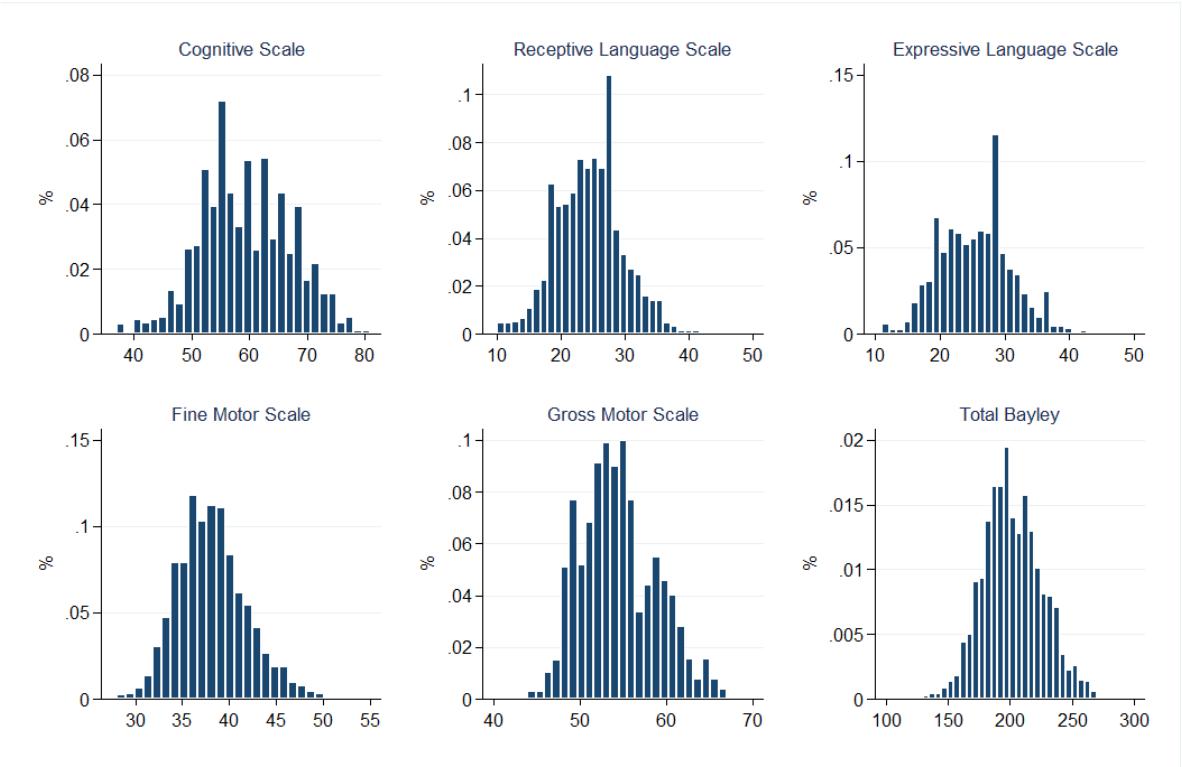


Source: Program attendance registry (recorded by FAMI facilitators)
Nota: Subsample of children registered at least once in group session attendance lists (74% of treated children found at follow-up)

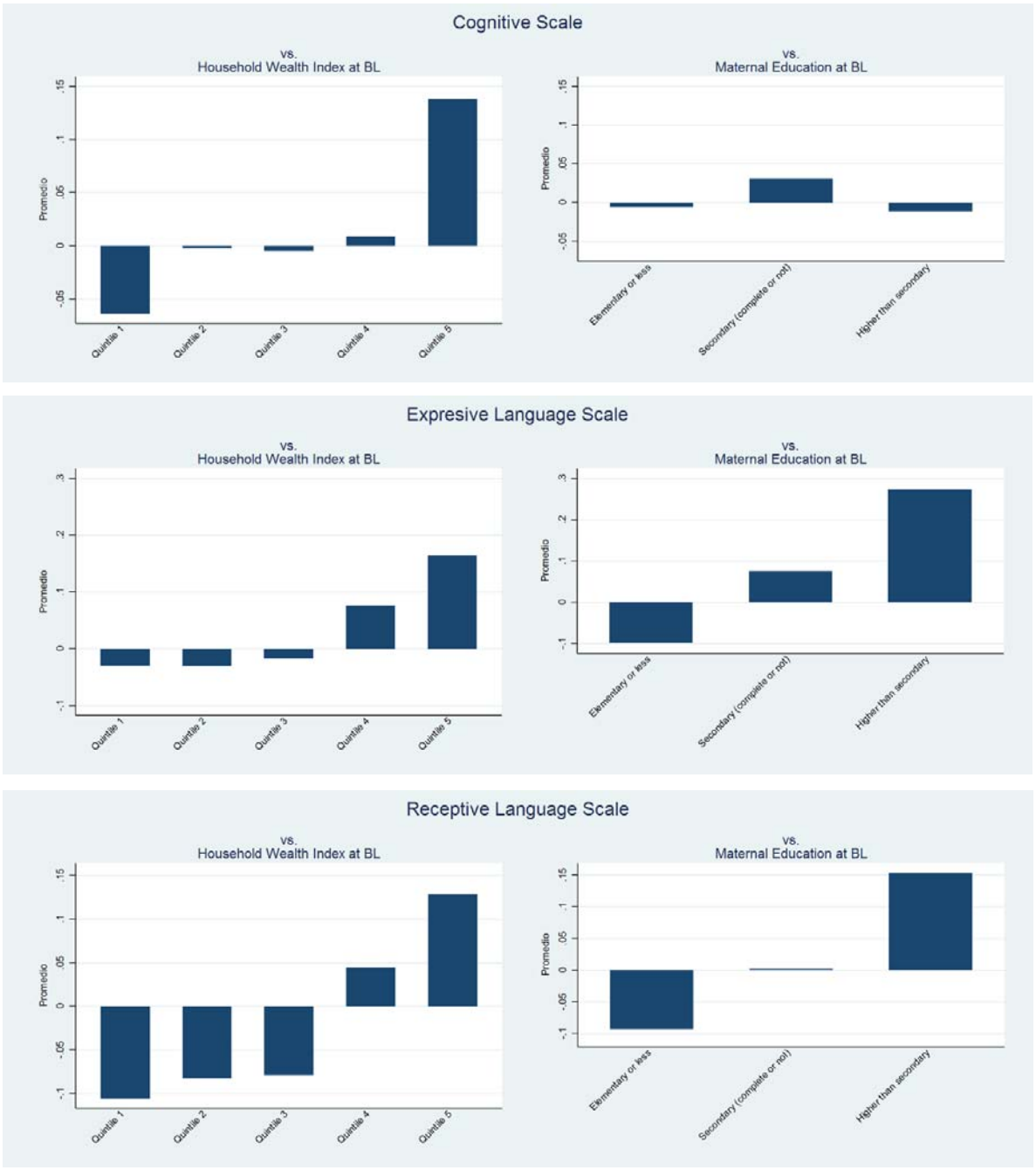


Source: Program attendance registry (recorded by FAMI facilitators)
Nota: Subsample of children registered at least once in home visit attendance lists (72% of treated children found at follow-up)

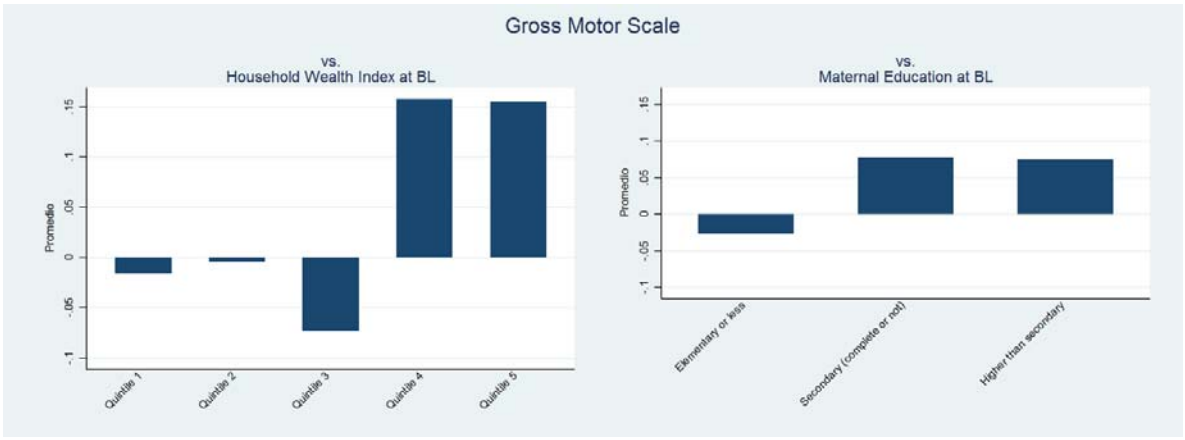
Appendix 8. Distribution of Bayley-III and correlations with children's sociodemographic characteristics

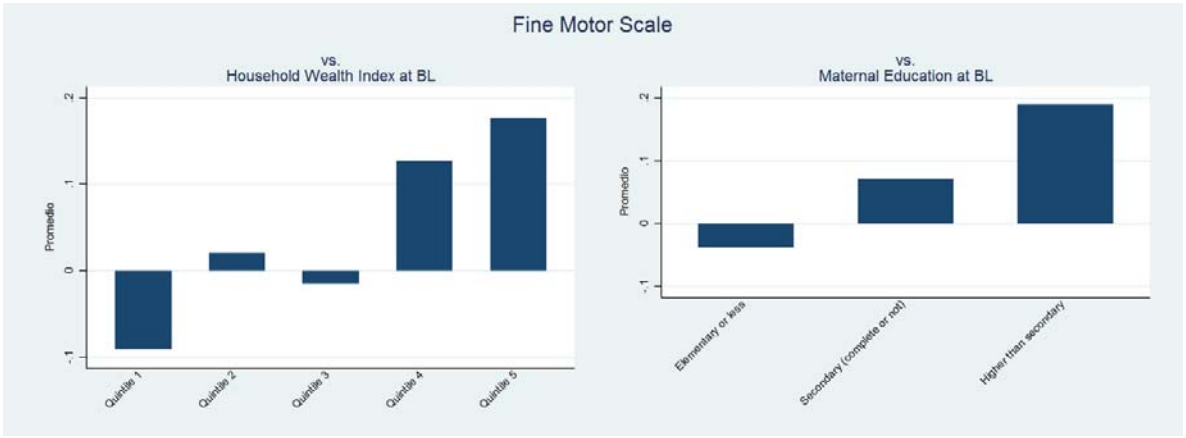


Appendix 8 (continuation)



Appendix 8 (continuation)





Appendix 9. First-stage results for program impacts by duration of exposure to the program

VARIABLES	At least one program activity	22 or more activities
ITT	0.7694***	0.6174***
	(0.0225)	(0.0376)
Age in months	-0.0156**	0.0192**
	(0.0070)	(0.0073)
Age sq.	0.0007	-0.0021***
	(0.0006)	(0.0006)
Gender (Male)	0.0049	0.0077
	(0.0153)	(0.0209)
Birth Order	-0.0209	-0.0018
	(0.0173)	(0.0232)
No. Of Siblings	0.0142	-0.0061
	(0.0153)	(0.0241)
Wealth Quintile = 2	-0.0234	-0.0161
	(0.0242)	(0.0290)
Wealth Quintile = 3	-0.0179	-0.0176
	(0.0235)	(0.0269)
Wealth Quintile = 4	0.0097	-0.0107
	(0.0281)	(0.0346)
Wealth Quintile = 5	-0.0413	-0.0518
	(0.0276)	(0.0317)
Maternal Education = 2, Complete or incomplete secondary	-0.0348*	-0.0402
	(0.0208)	(0.0254)
Maternal Education = 3, Higher than secondary	-0.0403	-0.0453
	(0.0394)	(0.0515)
Father present	0.0547**	0.0338
	(0.0220)	(0.0294)
Household size	0.0031	-0.0012
	(0.0074)	(0.0092)
Maternal verbal ability (PPVT)	0	-0.0018
	(0.0012)	(0.0017)
Teenage Mother	-0.0375	-0.0411
	(0.0270)	(0.0296)
BL composite score	0.0009	-0.0049
	(0.0033)	(0.0041)
Stratification variables	X	X
Department fixed effects	X	X
Tester fixed effects	X	X
Constant	0.1042	0.0473
	(0.0775)	(0.1208)
Observations	1213	1213
R-squared	0.68	0.4498
F-test	1069.52	24.6
Prob>F	0	0

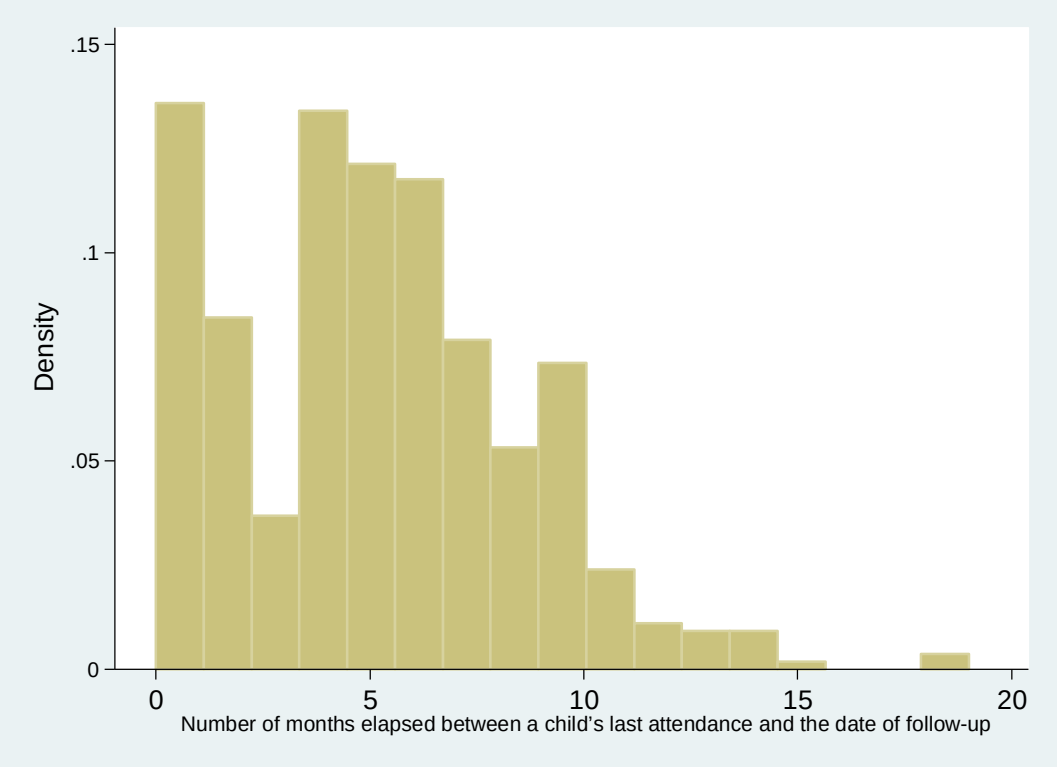
Estimated by OLS; Standard errors clustered by town in parenthesis
*** p<0.01, ** p<0.05, * p<0.1

Appendix 10. Differences between children with different durations of exposure to the program

	22 or more activities		*Less than 22 activities*		
	Mean	SD	Mean	SD	
Sociodemographic characteristics					
Chilld's age in months	5.43	(3.31)	6.11	(3.55)	***
Chilld's age in months at FU	24.03	(3.42)	24.85	(3.74)	***
Boy (%)	0.53	(0.50)	0.503	(0.50)	
Child's birthweight (gr)	3197	-585	3170	-509	
Birth order	1.06	(1.34)	1.01	(1.22)	
First Born (%)	0.434	(0.50)	0.5	(0.50)	
Maternal years of schooling	8.58	(3.48)	9.14	(3.34)	*
Father present (%)	0.74	(0.44)	0.66	(0.48)	**
Number of siblings	1.03	(1.26)	1.02	(1.36)	
Mother married / cohabiting	0.71	(0.45)	0.71	(0.45)	
Mother is Single	0.22	(0.41)	0.27	(0.45)	
Mother is Divorced	0.01	(0.12)	0.01	(0.09)	
Teenage mothers (%)	0.23	(0.42)	0.28	(0.45)	
Mother's age (years)	26.79	(7.00)	25.48	(6.58)	**
Household wealth index [a]	0.02	(0.96)	0.10	(0.96)	
No. of observations	366		336		
Intermediate Outcomes					
FCI No. of adult books, magazines and newspapers	2.54	(3.07)	2.60	(3.10)	
FCI No. of toy sources	1.33	(0.94)	1.38	(0.93)	
FCI No. of varieties of play materials	1.33	(1.37)	1.52	(1.44)	
FCI No. of varieties of play activities over past 3 days	2.45	(1.51)	2.62	(1.60)	
FCI No. of parental care activities over past 3 days	4.65	(1.13)	4.86	(0.91)	***
FCI home environment (PCA) b	-0.08	(0.98)	0.07	(0.98)	*
Social support DUKE UNC-11 total (raw score)	40.75	(8.39)	41.74	(8.10)	
High maternal self-efficacy (% above median)	0.39	(0.49)	0.41	(0.49)	
Mothers with depression symptoms (%)	0.14	(0.35)	0.17	(0.37)	
Use of verbal or physical abuse in the household (%)	0.01	(0.10)	0.04	(0.19)	*
No. of observations	366		336		
Nutrition					
Weight-for-age z-score (n1=351, n0=311)	0.18	(1.42)	0.33	(1.36)	
Length/height-for-age z-score (n1=345, n0=309)	0.00	(1.66)	-0.04	(1.74)	
BMI-for-age z-score (n1=338, n0=299)	0.33	(1.64)	0.39	(1.62)	
Weight-for-length/height z-score (n1=331, n0=295)	0.41	(1.56)	0.32	(1.62)	
Food-Insecure (Any, %)	0.47	(0.50)	0.54	(0.50)	
No. of observations	366		336		
Reasons for dropout					
Enrolled in another parenting or early childhood program (%)	0.31	(0.46)	0.41	(0.49)	
Child got sick (%)	0.01	(0.11)	0.01	(0.08)	
Found attendance useless (%)	0.00	(0.00)	0.01	(0.08)	
Moved away (%)	0.05	(0.22)	0.20	(0.40)	***
Lost eligibility by age (%)	0.63	(0.48)	0.38	(0.49)	***
No. of observations	173		177		

Standard errors clustered by town in parenthesis
*** p<0.01, ** p<0.05, * p<0.1

Appendix 11. Months elapsed between the last program activity attended and the date in which the child was assessed at follow-up



Source: Program attendance registry and follow up data.

Note: Subsample of children found at follow up and registered at least once in home visit attendance lists (N=487).