

Evaluation of the Happy Child Program: a randomized study in 30 Brazilian municipalities

Iná S. Santos (<https://orcid.org/0000-0003-1258-9249>)¹

Tiago N. Munhoz (<https://orcid.org/0000-0003-1281-9542>)¹

Raquel Siqueira Barcelos (<https://orcid.org/0000-0001-6580-0509>)¹

Cauane Blumenberg (<https://orcid.org/0000-0002-4580-3849>)¹

Caroline Cardozo Bortolotto (<https://orcid.org/0000-0003-3318-7900>)¹

Alicia Matijasevich (<https://orcid.org/0000-0003-0060-1589>)²

Cristiane Salum (<https://orcid.org/0000-0003-0191-371X>)³

Hernane Guimarães dos Santos Júnior (<https://orcid.org/0000-0002-9998-2141>)⁴

Letícia Marques (<https://orcid.org/0000-0001-5963-2166>)⁵

Luciano Correia (<https://orcid.org/0000-0001-8948-8660>)⁶

Marta Rovey de Souza (<https://orcid.org/0000-0001-6910-843X>)⁷

Pedro Israel Cabral de Lira (<https://orcid.org/0000-0002-1534-1620>)⁸

Vitor Pereira (<https://orcid.org/0000-0002-6487-7704>)⁹

Cesar G. Victora (<https://orcid.org/0000-0002-2465-2180>)¹

¹ Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas. R. Marechal Deodoro 1160, 3º piso, Centro. 96020-220 Pelotas RS Brasil. inasantos.epi@gmail.com

² Departamento de Medicina Preventiva, Faculdade de Medicina, Universidade de São Paulo. São Paulo SP Brasil.

³ Centro de Matemática, Computação e Cognição, Universidade Federal do ABC. Santo André SP Brasil.

⁴ Instituto de Saúde Coletiva, Universidade Federal do Oeste do Pará. Santarém PA Brasil.

⁵ Instituto de Humanidades Artes de Ciências, Universidade Federal da Bahia. Salvador BA Brasil.

⁶ Departamento de Saúde Comunitária, Universidade Federal do Ceará. Fortaleza CE Brasil.

⁷ Departamento de Saúde Coletiva, Universidade Federal de Goiás. Goiânia GO Brasil.

⁸ Departamento de Nutrição do Centro de Ciências da Saúde, Universidade Federal de Pernambuco. Recife PE Brasil.

⁹ Escola Nacional de Administração Pública. Brasília DF Brasil.

Abstract *The Happy Child Program (Programa Criança Feliz - PCF, in Portuguese) reaches 1.4 million Brazilian children under three years of age with home visits aimed at promoting neuro-psychomotor development. Based on a conceptual model, PCF implementation and impact were evaluated in a randomized study in 30 municipalities. A total of 3,242 children were allocated to the intervention (IG) or control (CG) group, 80.0% of whom were prospectively followed up from late 2018 to late 2021. Development was assessed by the Ages and Stages Questionnaire (ASQ3). During the three-year study period, visits were replaced by virtual contacts for an average of 12 months due to COVID-19. At the endline survey, intent-to-treat analyses showed mean scores of 203.3 in the IG and 201.3 in the CG. Additional analyses using instrumental variables and propensity scores matching also showed no effect, since the number of contacts with the program was not associated with ASQ3 scores. No impact was observed on stimulation, responsive interactions or psychological attributes of children. The implementation study revealed low coverage in the IG, contamination of the CG, deficiencies in management and low quality of visits in many municipalities. The study did not demonstrate an impact of PCF implemented under routine conditions, but provides elements for its improvement.*

Key words *Child development, Program evaluation, Impact*

Introduction

Home visiting programs are considered effective strategies for promoting child development. Systematic reviews of interventions performed in low- and middle-income countries show positive effects on the cognitive, language, psychosocial, and motor domains^{1,2}. Meta-analyses of randomized trials conducted in high-, middle-, and low-income countries also found positive effects on child development, in addition to improvements in mediating outcomes, such as parental knowledge, parenting practices, and interaction with the child³⁻⁵. It is important to note that most of these experimental studies were carried out in small and selected samples, with the interventions being implemented by the researchers themselves.

Several randomized intervention studies were carried out in Latin America in recent years. In Colombia, in a study in 96 municipalities with 1,420 children, the group who received home visits for 18 months improved their adjusted cognitive score by 0.26 standard deviations (SD) ($p=0.002$) and 0.22 SD their adjusted language score by 0.22 SD ($p=0.032$) in relation to children who were not visited⁶. In Peru, the proportion of children scoring above average for age in all domains was 12-23 percentage points higher in the group that received home visits⁷. In Mexico, the integration of two large-scale programs (a cash transfer program and an early childhood education program) resulted in higher cognitive, language and memory indices, equivalent to 0.26-0.29 SD, compared to the group that received only financial support⁸. A separate analysis of the same program showed positive effects of the combined interventions on mediating outcomes – parent-child play and children's book reading – that accounted for 32% of the effect on development⁹.

In Brazil, a review of the literature identified a gap in studies evaluating the impact of home visiting programs implemented as public policies focused on child development¹⁰. Regarding interventions led by universities, a cohort study with 156 children conducted in the interior of Pernambuco relied upon home visits and group activities carried out by occupational therapists¹¹. There were significant differences between the intervention and control groups of 9.4 points in cognitive development and 8.2 points in psychosocial development ($p<0.001$ in each case). A recent randomized controlled study trial with 900 families in the city of São Paulo evaluated

the impact of an adapted version of the Reach Up program¹², an evidence-based intervention¹². Intent-to-treat analysis showed no effects on children's development, but per-protocol analysis, restricted to mother-child dyads who completed at least 10 visits, showed an improvement in children's development of 0.22 (95%CI 0.01-0.43) SD¹³. A quasi-experimental study of the impact of the *Primeira Infância Melhor* (PIM) Program in Pelotas-RS, with 601 children enrolled in the program and an individually matched control group, found no effect of the program on development at age four¹⁴. However, children whose mothers were visited since pregnancy showed 60% lower prevalence of developmental delay compared to the control group ($p=0.02$), and also lower prevalence than children for whom visits started after birth (prevalence ratio=0.40; 95%CI 0.18-0.89).

In March 2016, Law 13,257 (known as the *Marco Legal da Primeira Infância*) was enacted. The law established that interventions to promote integral development in early childhood would become priority strategies supported by public social policies. In October 2016, at the initiative of the then Ministry of Social and Agrarian Development (currently the Ministry of Citizenship - MCid), Brazil became the first country in the world to launch a national home visiting program aimed at promoting the development of socially-vulnerable children – the Happy Child Program (PCF)¹⁵. Eligible children were those whose families were enrolled in the *Bolsa Família* (currently *Auxílio Brasil*), a cash transfer program that covered 40 million Brazilians¹⁶. With gradual implementation, PCF was reaching approximately 1.4 million children under three years of age as of June 2022.

PCF is implemented in a decentralized and integrated manner involving three administrative levels: the federal, state and municipal governments¹⁷. Implementation takes place at municipal level, relying upon the existing networks of Social Assistance Reference Centers (CRAS) and being integrated with the delivery of other social protection public policies for vulnerable families. After signing of the PCF Acceptance and Commitment Term, municipal authorities received funds from the federal government and became responsible for implementation with a substantial degree of autonomy.

Specific objectives of the PCF include guiding and supporting pregnant women and their families in preparation for the birth of the child; collaborating in the exercise of parenting;

strengthening the roles of families in the care, protection and education of children; and favoring the strengthening of affective and community bonds. Additional objectives include promoting actions aimed at integral development in early childhood, encouraging the development of recreational activities with the involvement of other family members; promoting and monitoring child development; and facilitating access to other public services¹⁵. The cornerstone of PCF were weekly visits to the children and their families, starting during pregnancy and continuing until the child reaches 36 months of age.

At the time PCF was launched, the Ministry commissioned researchers from the Postgraduate Program in Epidemiology at the Federal University of Pelotas (UFPEL) to carry out an independent impact evaluation. The research team proposed a randomized controlled trial in several parts of the country.

The impact model proposed by the research team postulated a necessary chain of events for PCF to impact on child development¹⁸. The first link in the chain would be the PCF implementation in the municipalities and the second, the selection and appropriate training of home visitors in the promotion of child stimulation and development. The third element would consist in carrying out weekly home visits to the families selected for the intervention group. The fourth link was the proper application of the protocols defined by the program, in terms of the contents of the visits and in establishing productive relationships between visitors and families. With this, the fifth element in the chain would be achieved, namely the increase in the mother's or caregiver's knowledge about child stimulation – which would then promote the child's interaction and stimulation (sixth link). Finally, it was expected that children in the intervention group would score higher in development tests than their peers living under similar conditions who were not beneficiaries of the program.

During the three years of the evaluation, each of the links described above was evaluated using different indicators and information collection tools. Given that the links in the impact model are interconnected in a logical and sequential order, deficiencies in any of the levels would compromise the subsequent links^{19,20}. In order to properly assess the proposed impact model, the evaluation study included two main components: implementation studies at municipal level, and household interviews for testing the children on an annual basis during three years.

Impact and implementation evaluations can help policymakers and managers improve and strengthen early childhood programs¹⁰. In addition, the study provided scientific evidence on the profile of the population served by *Bolsa Família* and PCF, as well as on factors associated with child development. A previous publication based on this evaluation found that vaccination coverage in the intervention group was higher than in the control group when the children were on average 18.9 months²¹. High prevalence (26.5%) of depressive symptoms among mothers participating in the study was demonstrated²², and that – in addition to maternal depression – the factors associated with lower child development scores included low maternal education and their perception of not having received support during pregnancy²³. Finally, the research also aims to fill the gap in the literature of studies on the impact of home visiting programs to stimulate child development linked to public policies.

Methods

Design and eligibility

The longitudinal evaluation had two main components. The first (impact study) was an individual-level randomized trial. Children under one year of age who were eligible for the PCF were enrolled in 2018-19 and followed up for three years.

The second component (implementation study) relied upon observational mixed-methods designs. The evaluation team visited the municipalities at the ends of 2019 and 2021, and collected data by telephone in late 2020, as actual visits at that time were not possible due to the COVID-19 pandemic.

Selection of municipalities

Thirty municipalities were selected jointly with the Ministry staff, covering four of the five subnational regions in the country. The Southern region (which has only three states) was left out because there were few municipalities with enough eligible children, and because the existence of programs that were similar to PCF that might contaminate the study design. One state was selected in each of the four regions, except in the Northeast where three states were chosen because this is the region with the largest clientele of the cash transfer program.

Figure 1 shows a map with the selected states and municipalities.

Within each of the six states, three to six municipalities were selected based on the following criteria: local government willingness to join the program and the evaluation study; 10,000 or more inhabitants; number of eligible children at least four times greater than the capacity to deliver the program; and ability to enroll at least 80 children. The requirement for at least four eligible children for each vacancy in the program was essential due to ethical reasons as described below.

Sample size calculations

Based on the literature review, it was hypothesized that participation in PCF would lead to an average increase over time of 0.20 SD in results of development tests relative to the comparison group. Given that full compliance was unlikely, a difference of 0.15 SD was used in the sample size calculations. Additional parameters were set at 0.05 for the alpha error (two-tailed), study power of 90% (beta error of 10%), and 20% of losses to follow-up, resulting in a desired national sample size of 2,880 children at baseline. This would be achieved with an average of 96 children per municipality, that is, 48 children in each group.

Randomization

The cash transfer program database at the Ministry was used to generate lists of potentially eligible children in each municipality. From late 2018 to mid-2019, the research teams visited each municipality and mapped the geographic areas where PCF would be implemented. In most municipalities, rural areas and regions with poor security conditions were excluded by the city social work staff. The eligibility lists included families with children aged under 12 months or pregnant women likely to deliver a child in 2018. Using the list of eligible families, the research team carried out the baseline study and generated a listing of children whose families agreed to participate in the randomized study and signed the informed consent form. Because 26 municipalities had already started the PCF, children who had already been visited once or more times were not eligible for inclusion in the evaluation.

As soon as the baseline study was completed in a municipality, an anonymized listing of eligible children was sent electronically to the UFPEL where randomization was carried out immediately. Each child was identified by their date of birth and their Social Identification Number, a unique number allocated by the Ministry to all beneficiaries of social welfare. At UFPEL, all eli-



Figure 1. States and municipalities included in the study.

Source: Authors.

gible children in each municipality were ordered according to their dates of birth, and blocks of 20 children in ten pairs were created. Using a sequence of ten coin tosses, children in each pair were allocated to group A or to group B, resulting in the following sequence within each group of 20 children: BABAABABABABBABABA. Randomization within each block of 20 children was only carried out once and used for the 30 municipalities. As the lists of eligible children arrived from a municipality, the last digit of the first prize of the most recent weekly drawing of the national lottery was used to allocate group A (in case the last digit was an odd number) or group B (for even numbers) to receive the home visits. All randomization sessions were transmitted live by the internet using Zoom, and remain available in the YouTube platform²⁴. The listings were saved immediately as pdf files and sent to the Ministry which forwarded them to the 30 municipalities, specifying which children should receive the visits (the intervention group) and which ones should not be visited (the comparison group).

Ethical aspects

The selected municipalities had at least four times the number of PCF-eligible children relative to the number of potential vacancies, thus ensuring that random selection would not leave out children who may have benefited. The Project was approved by the Research Ethics Committee at UFPEL, which is affiliated with the National Research Ethics Committee of the Ministry of Health under the protocol number CONEP 2.148.689. The study protocol was deposited in the Brazilian Registry of Clinical Trials (protocol REBEC RBR-4x7dny).

The study protocol ensured the confidentiality of personal information and written informed consent was obtained from a parent or caregiver in the baseline interview. All databases are anonymized. All study data are publicly available for use by other research teams.

Evaluation study: data collection

The original design of the evaluation study included four visits to all children: the baseline study in late 2018 (referred to as T0 or T zero), the first-year follow-up study (T1) in late 2019, the second follow-up (T2) in late 2020, and the third and last follow-up (T3) in late 2021. The pandemic led to the cancellation of the T2 visits, and the design was adapted to include five telephone in-

terviews with the families in June and September 2020, and in January, April, and July 2021.

For the T0, T1 and T3 interviews, each state team included an experienced principal investigator, one or two supervisors and 6-10 field workers. All supervisors and field workers had at least a secondary school diploma. Face-to-face training of the interviewers took place in Brasilia before the T0 and T1 phases; web-based training was carried out before the T3 phase due to the pandemic. In each state, supervisors and field workers carried out the interviews in each municipality during 1-2 weeks and if necessary, returned on a later occasion to complete additional interviews.

Quality control measures in all phases included standardization and pilot testing of the questionnaires, preparation of detailed interviewer manuals, re-interviews of 10% of all families using a shortened questionnaire by the state supervisors, and visits to selected municipalities in all states by members of the central team (UFPEL) during the field work phases. Interviewers were not informed about the intervention or comparison status of each child, although this was possible to discover as the questionnaires included variables related to participation in PCF.

Instruments

The three visits included the application of questionnaires to the children's mothers or caregivers (from now on referred to as "mothers") and testing the development and mental health of children. The main questionnaire incorporated widely-used questions from the UNICEF *Multiple Indicator Cluster Survey* (MICS)²⁵, and from questionnaires used in the Pelotas Birth Cohort Studies (<http://www.epidemiio-ufpel.org.br/site/content/studies/>).

Additional questions were adapted from the PCF training manuals. The topics included in the questionnaires covered sociodemographic information on the family and parents, childcare and stimulation, discipline, feeding, and use of health services. Questions on participation in PCF, including age at enrollment, frequency of visits and maternal recall of advice by the home visitors, were collected for both intervention and comparison groups. Maternal depression was assessed using the Brazilian version of the Edinburgh Postnatal Depression Scale²⁶⁻²⁸. The complete questionnaires are available at the link <http://www.epidemiio-ufpel.org.br/site/content/pesquisas/index.php?estudo=5>.

Cognitive and sensorial stimulation was assessed using a score derived from MICS survey questions (UNICEF) and on the recommendations from the PCF visitor's training handbook²⁹. According to the child's age, mothers were asked about recommended interactions such as telling stories, reading books, singings, teaching names of body parts and of foods, drawing and helping in the house. For children aged 30-36 months, the maximum score was 38 points, and for those aged over 36 months, 30 points. For this last group, the score was multiplied by 38/30 to harmonize with the younger group before the regression analyses.

Child discipline practices were assessed using subsets of two validated scales, the Parent-Child Conflict Tactics Scales³⁰ and the Juvenile Victimization Questionnaire³¹.

Child development was assessed in the T0, T1 and T3 visits using the Ages and Stages Questionnaire (ASQ3)³². ASQ3 includes 30 items from five domains: cognition (problem solving), communication, fine and gross motor coordination, and personal-social. Each domain includes six questions answered by the mother regarding developmental milestones, with three possible replies: not yet (0 point), sometimes (5 points) and yes (10 points). The items investigated are age specific; for example, the T3 questionnaire included the modules for average ages 36, 42 and 48 months. For each age stratum, the maximum score including the 30 questions equals to 300 points.

The standard ASQ3 solely relies on respondent's report regarding abilities of the child. This version was used in all three phases of the study. To increase the objectivity of the results, the research team derived an additional ASQ3 version using exactly the same milestones as the standard version but based on observation of task completion by the interviewers. Each interviewer carried objects such as toys, child clothes, personal hygiene items, paper and pencils, and children were requested to complete each ASQ3 task while being observed. When the child was unable to complete a task, the respondent was asked and her/his information was recorded. This led to two distinct scores per child, the standard (or reported) score and the observed score, and both were analyzed separately. Pearson's correlation coefficient between the two scores was equal to 0.858 ($p < 0.001$). In the data analysis, children with some mental disability (according to the respondent) and those with scores below 30 were excluded because the distribution was bimodal,

with a first peak between 0 and 30, indicating children who failed to respond to nearly all questions. In total, there were 16 exclusions for the reported test (8 each for the intervention and control groups) and 292 for the observed test (145 for the intervention group and 147 for the control group). The difference in numbers is related to the fact that the observed test was more demanding than the reported version. Among children who were excluded, only three from the intervention group and one from the control group had received a visit from the PCF in the four weeks prior to the interview.

The child psychological attributes were assessed by means of the Strengths and Difficulties Questionnaire (SDQ) – which was part of the T3 questionnaire. This screening instrument was developed in the United Kingdom and later validated in Brazil^{33,34}. It covers five scales with five items each: emotional symptoms, conduct problems, hyperactivity-inattention, peer relationship problems, and prosocial behaviors. Each item was scored between 0 and 2, and three categories (generated by the sum of the results of all scales, except the sociability scale - prosocial behaviors) were defined, indicating an increasing prevalence of symptoms: 0-13, 14-16 and 17 or more points. For regression analysis, the continuous variable from 0-40 points was used.

Parent-child responsive caregiving was assessed in the T1 and T3 rounds for approximately 10% of the mother-child dyads who were filmed during five minutes while carrying out two simple tasks. After receiving a printed image and several building blocks, the mother worked with the child in reproducing the two images with the blocks. Each video was separately assessed and scored by two psychologists at UFPEL, who were unaware of the child's intervention or comparison status. The test was recently validated in Canada and Brazil^{35,36}.

Telephone interviews

To minimize the impact of the pandemic on the evaluation design, the T2 visit was replaced by five telephone interviews with the families during 2020 and 2021. Using the telephone numbers collected during the T0 and T1 visits, attempts were made to contact all families. The phone interviews were aimed at updating address and contact information and assessing participation in PCF by home visits or other types of interactions with the PCF visitors, such as social media or phone calls.

Implementation study

The implementation study was carried out in parallel with the annual household visits, except during the T2 phase when it was conducted by telephone. Supervisors of the state evaluation teams carried out data collection. Each round included key-informant interviews with the state and municipal PCF coordinators, who were also requested to complete standardized forms with information on the program. Coordinators of social work centers in each city were also interviewed; these were in charge of program implementation in their catchment areas. Three sets of activities were carried out with the home visitors in an anonymous base: a knowledge test, a self-applied questionnaire on their views regarding positive and negative aspects of the program, and a focus group discussion. Within each municipality, an average of ten home PCF visits were observed by the supervisor from the evaluation team using a checklist of the expected contents to be delivered according to the age of the child. Lastly, after completion of the T3 round of interviews, a single web-based focus group session was carried out by the central study team including the six state evaluation supervisors, in which these discussed their impressions on the strengths and weaknesses of the PCF.

The Ministry database on number of visits or contacts

The PCF monitoring database included the number of visits or virtual contacts (from now on referred to as “contacts”) with each child from January 2019 to December 2021. This information was matched to the study database using the children’s Social Identification Numbers.

Data processing and analyses

Data from the household and telephone interviews, as well as from the implementation study, were entered in the field using a tablet programmed with REDCap software³⁷, and downloaded to UFPEL in the same day. Range, consistency and logical checks were carried out initially in the field using the software, and double-checked at UFPEL when any remaining inconsistencies were sent back to the state teams. Stata 17.0 (StataCorp®, College Station, TX, USA) was used for data cleaning, recoding, and statistical analyses.

Intent-to-treat (ITT) analyses constituted the primary strategy. Children who were initially allocated to the intervention and comparison groups remained in the same group regardless of compliance with the original randomization scheme.

Two sets of additional analyses were carried out. Using propensity score matching (PSM), 445 children with 30 or more contacts according to the PCF monitoring database were matched to children without any contacts. Matching variables included age, municipality, baseline ASQ3 scores, wealth quintiles, maternal variables (schooling, skin color, work outside the home and depression) and the number of children in the household.

Finally, analyses were performed using instrumental variable (IV) methodology to estimate the additional impact of each program visit on child development. Because the selection of children for the intervention group was at random and was associated with an increase in the average number of contacts, the treatment assignment can be used as an instrumental variable for the number of visits received. Assuming the absence of spillover effects and that the allocation of the child to the intervention group does not cause a reduction in the number of contacts (monotonicity hypothesis), the IV estimates can be interpreted as the mean effect of each visit on the outcomes observed among children whose number of visits increased when they were randomly assigned to the intervention group. This approach is known as Local Average Treatment Effects or LATE.

Using two-stages linear regression the at random treatment assignment was used to predict the number of contacts, and then the predicted values were used to estimate the impact of each additional contact on the outcomes, after adjusting for age and municipality. Instrumental variable estimates were initially calculated using the number of contacts obtained from the PCF monitoring database and, in the sensitivity analyses, using the number of contacts calculated from the interviews and phone calls of the study.

The study sample was described using absolute and relative frequencies for categorical variables and using means and standard deviations for continuous variables. Sample characteristics of intervention and comparison groups were compared using chi-squared tests for proportions and Student t tests to compare means. Crude and adjusted multilevel linear regression models were used in ITT analyses.

Results

Impact study

The baseline study or T0 enrolled 3,242 children aged 0-11 months with an average of 108 per municipality. Of these, 1,623 were randomized to the intervention and 1,619 to the comparison group. The average age of the children was 7.3 months (SD=2.9). Table 1 shows sociodemographic and health care characteristics of the study sample. The mother was the primary caregiver for over 90% of the children. About 87% of the families had running water inside the house, and 78% of the households had a flush toilet. A fifth of the dwellings had more than three people per room used for sleeping. Just over a third of the mothers did not live with the child's father, their average age was 26 to 27 years, and approximately three quarters classified themselves as having brown skin color. About 10% of mothers had attended school for four years or less. Virtually all mothers underwent prenatal care, with an average number of eight consultations. Virtually all deliveries were in hospital, with about half by cesarean section. There was no difference between mothers regarding the presence of depressive symptoms. About 9% of the children were preterm and 5% weighted less than 2,500 g at weight.

Randomization resulted in two highly comparable groups (Table 1). Of 50 indicators studied, only one showed a statistically significant difference among the two groups: 2.2% of the intervention group children were left alone for one hour or more in the week before the interview, compared to 1.1% in the comparison group ($p=0.01$). Detailed results of the T0 phase on baseline comparability are available in Table 1 and in a previous publication²⁴.

The ASQ3 test was applied during the baseline study. As shown in Table 2, the mean scores were equal to 248.1 in both groups. Detailed information on these results is available in a previous publication²³. In a reanalysis, including the scores of 292 children excluded because they had an observed ASQ3 lower than 30, the means of the two groups (intervention and control) decreased but remained practically the same and not significant.

The first year follow up (T1) took place from September 2019 to January 2020. A total of 3,008 children (92.8% of those enrolled during T0) were located, including 1,482 from the intervention and 1,526 from the comparison group. The aver-

age age was 18.6 months (SD=3.5). According to maternal report (Table 3) only 31.3% of the children in the intervention group had been visited in the previous month and 24.1% of the mothers recalled any pieces of advice provided by the PCF visitors. There was evidence of contamination of the comparison group, with 12.9% of the children reportedly enrolled in PCF and 7.3% having been visited in the previous month. These preliminary results were fed back to the Ministry in February 2020. As shown in Table 2, the mean ASQ3 score was 2.4 points higher in the intervention than in the comparison group, but the difference was not statistically significant ($p=0.145$). This difference is equivalent to about 0.05 SD. Ignoring random allocation and using propensity score matching, an additional analysis compared all 565 children who had been visited in the month prior to the interview with a similar group of 565 children not enrolled in the program. On this occasion, the visited group had a slightly higher mean score than the non-visited group (233.8 points *versus* 228.5 points; $p=0.043$).

Table 3 also shows results from the telephone interviews on PCF participation during the pandemic, when home visits were markedly reduced and virtual contacts were introduced. These included web-based calls, messages, and videos as well as phone calls from the visitors to the families. Around 30% of the families of children in the intervention group and around 10% of those in the comparison groups reportedly received such contacts.

With the suspension of the T2 household interviews due to the pandemic, the third and last visit (T3) took place from October 2021 to January 2022. A total of 2,594 interviews were carried out, with an overall follow-up rate of 80.0% (79.9% in the intervention and 80.1% in the comparison group). Average ages of the children were 42.7 months (SD=3.5) in the comparison and 42.8 months (SD=3.4) in the intervention group ($p=0.790$).

Follow-up rates according to baseline characteristics are shown in Table 4. Although follow-up varied significantly according to the age of the children (higher for young children), maternal skin color (higher for brown mothers), wealth (higher for poorer households), maternal schooling (higher for those with nine or more years) and maternal depression (lower for depressed mothers), at least 69% of the children in all groups were reached. These variables were included as potential confounders in the ASQ3 regression models. More importantly, there was no

Table 1. Comparability between intervention and control groups at baseline, according to selected indicators.

	Com- parison (%)	Inter- vention (%)	P
Characteristics of the household			
Type of building			0.945
Bricks/building	93.2	93.1	
Others	6.8	6.9	
Piped water			0.176
No	8.0	7.2	
Yes, inside the house	87.3	86.7	
Yes, in the backyard	4.8	6.1	
Toilet			0.768
Absent	2.0	2.0	
Flush toilet	79.4	78.3	
Toilet without flush or latrine	18.7	19.6	
Number children <7 years (excluding index child)			0.266
0	24.4	25.5	
1	57.1	58.6	
≥2	18.5	15.9	
Household crowding (>3 persons/bedroom)	22.0	23.1	0.482
Respondent of the interview			0.642
Mother	98.6	98.8	
Other	1.4	1.2	
Socio-economic position and characteristics of the father			
Socioeconomic position of the household			0.754
Quintil mais pobre	20.6	19.5	
Poorest quintile	20.2	19.8	
Second quintile	19.8	20.2	
Third quintile	19.1	21.0	
Fourth quintile	20.3	19.6	
Father's age in years (mean and SD)	30.5 (8.6)	30.7 (8.8)	0.476
Father's skin color			0.206
White	24.2	23.9	
Black	8.7	10.6	
Brown	65.2	64.3	
Other	1.8	1.3	
Father's schooling in years (mean and SD)	8.4 (3.6)	8.5 (3.8)	0.621
Father is currently working			0.601
No	26.7	25.3	
Every day	49.7	51.3	
Some days	23.5	23.4	
Family enrolled in Bolsa Família Program	98.0	98.1	0.898

it continues

Table 1. Comparability between intervention and control groups at baseline, according to selected indicators.

	Com- parison (%)	Inter- vention (%)	P
Maternal characteristics			
Mother's age in years (mean and SD)	26.7 (6.8)	26.9 (6.6)	0.466
Mother's skin color			0.257
White	15.5	13.6	
Black	10.1	10.0	
Brown	71.8	73.0	
Other	2.6	3.4	
Mother's schooling in years			0.717
0-4	8.8	9.7	
5-8	30.9	30.6	
≥ 9	60.3	59.7	
Mother's schooling in years (mean and SD)	9.3 (3.2)	9.3 (3.2)	0.829
Mother lives with husband or partner	63.5	62.6	0.585
Mother works outside the home			0.599
No	89.9	90.7	
Every day	4.2	3.5	
Some days	5.9	5.8	
Maternal smoking	7.9	7.8	0.844
Depressive symptoms (Edinburgh ≥10)	25.4	27.6	0.171
Depressive symptoms (Edinburgh ≥13)	14.7	14.4	0.880
Characteristics of the child's gestation			
Parity (excluding index child)			0.057
0	32.7	28.6	
1	30.9	33.3	
2	20.7	22.8	
≥3	15.7	15.3	
Antenatal care attendance	98.9	99.0	0.734
Antenatal visits (mean and SD)	7.8 (3.0)	7.9 (3.0)	0.085
Delivery in a hospital	99.7	99.5	0.580
Vaginal delivery	53.1	52.9	0.972
Characteristics of the newborns			
Female sex	50.2	50.1	0.972
Preterm (<37 weeks of gestation)	8.3	8.6	0.843
Birth weight in grams (mean and SD)	3.271 (509)	3.266 (513)	0.575
Low birth weight (<2500 grams)	6.0	6.1	0.941
5-minute Apgar score < 7	1.1	1.2	0.699

it continues

Table 1. Comparability between intervention and control groups at baseline, according to selected indicators.

	Com- parison (%)	Inter- vention (%)	P
Characteristics of the children			
Age groups (months)			0.990
<3	9.2	9.0	
3 a 4	12.1	12.0	
5 a 6	21.0	21.5	
7 a 8	25.2	25.0	
9 a 10	23.5	23.0	
11 a 12	8.9	9.5	
Currently breastfed	75.1	76.6	0.345
Childcare, stimulation-interaction and development promotion			
Main caretaker			0.559
Mother	93.0	93.4	
Father	0.4	0.4	
Grandmother/ grandfather	3.0	3.3	
Mother and father	2.7	1.9	
Other	1.0	1.0	
Daycare center attendance	1.4	1.3	1.000
Child was left alone for ≥1 hour at the last week	1.1	2.2	0.012
Another child (≤10 years) took care of the child for ≥1 hour at the last week	0.8	0.7	0.838
Child has a book or comic book	8.7	10.2	0.183
Maternal/caretaker stimulation-interaction score (0-5 months) (mean and SD)	12.8 (2.3)	12.8 (2.3)	0.749
Maternal/caretaker stimulation-interaction score (6-12 months) (mean and SD)	14.9 (2.6)	14.8 (2.7)	0.926
Development-promoting activities score (last 3 days) (mean and SD)	3.4 (1.3)	3.4 (1.3)	0.950
Use of inadequate disciplinary methods (last month) (mean and SD)	1.1 (0.8)	1.1 (0.8)	0.560
Agrees with the statement: “to educate and raise a child well, the child does not need to be beaten”	95.1	94.3	0.308

Maternal/caretaker stimulation-interaction score (0-5 months): maximum 16 points. Maternal/caretaker stimulation-interaction score (6-12 months): maximum 20 points. Development-promoter activities score (last 3 days): maximum 6 points. Use of inadequate disciplinary methods (last month): maximum 7 points.

Source: Authors.

statistical evidence of differential follow-up between the intervention and comparison groups within any category of the above-mentioned variables.

We report on six main outcomes: stimulation-interaction scores; child discipline; responsive interactions (based on videos); psychological attributes; and observed and reported ASQ3 test results. First, we present means or proportions of the outcome indicators using ITT analyses without any adjustments (Table 5). The mean stimulation-interactions scores were very similar for the intervention and comparison groups, for both age ranges studied. Mean scores were about 80% of the maximum values of the scale. The responsive caregiving scores derived from the videos recorded for a subsample were relatively low, with average values of 2.1 out of a maximum of 5.0 in both groups. Inadequate disciplinary methods were reported slightly more often in the comparison group (mean of 3.5 methods out of 12) than in the intervention group (3.4 methods). The distribution of psychological attributes scores was similar in both groups.

As shown in Table 2, observed and reported ASQ3 scores were similar in the intervention and comparison groups, with mean values around 202 out of a possible 300 points. As expected, reported scores were higher than observed scores.

The above-presented unadjusted results were complemented by regression-based ITT and LATE analyses (Table 6) with adjustment for child age and municipality. The beta coefficients show the differences in outcomes between the intervention and comparison groups in the ITT analyses, while the betas in LATE correspond to the change in outcome associated with each additional PCF contact.

In the adjusted analyses, stimulation/interaction scores for the two age groups (Table 6) were expressed on the same scale (see Methods) and then pooled. Neither ITT nor LATE analyses showed any evidence of differences between the intervention and comparison groups. Likewise, no differences were observed in the ITT and LATE analyses for responsive caregiving, disciplinary methods, or psychological attributes.

The regression analyses of the primary study outcomes – the two ASQ3 scores – were also adjusted for the baseline (T0) value of ASQ3 and for the above-described variables that were associated with the T3 follow-up rates. Neither ITT nor LATE analyses showed statistically significant differences (Table 6). Analyses using propensity score matching were carried out to confirm

Table 2. Unadjusted ITT analyses of ASQ3 results at baseline, first- and third-year follow-up studies*.

	Comparison			Intervention			P
	N	Mean	SD	N	Mean	SD	
ASQ-3 (reported)							
T0	1,505	248.1	40.2	1,521	248.1	37.7	0.953
T1	1,518	226.2	44.5	1,479	228.6	44.2	0.145
T3	1,245	219.8	50.6	1,253	221.3	49.4	0.453
ASQ-3 (observed)							
T3	1,105	201.3	58.2	1,108	203.3	57.1	0.427

*Children with ASQ3 scores below 30 points and those who reportedly presented learning disabilities were excluded; this was more common with the observed than with the reported scale.

Source: Authors.

Table 3. Percent of children who received PCF visits or contacts during the month preceding the interviews.

Percent of children	Household interviews (T1)		Telephone interviews										Household interviews (T3)	
	September 2019-January 2021		June 2020		September 2020		January 2021		April 2021		July 2021		September 2019-January 2021	
	GC	GI	GC	GI	GC	GI	GC	GI	GC	GI	GC	GI	GC	GI
PCF visits	7.3	31.3	1.0	3.8	6.9	28.2	4.7	17.8	4.2	11.4	2.8	9.0	0.8	6.1
PCF virtual contacts	0.0	0.0	5.7	24.8	14.0	31.9	15.4	25.3	15.7	32.4	13.4	27.7	17.3	35.5
Visits and/or contacts	7.3	31.3	5.8	25.0	14.1	32.6	15.4	26.1	16.4	35.6	13.9	29.4	16.8	35.4
Number of children	1,482	1,526	1,022	1,022	897	891	894	905	770	765	820	856	1,300	1,294

Source: Authors.

these results by comparing 445 pairs of children with 30 or more contacts during 2019-2021 according to the PCF database and closely matched children without any contacts. In these analyses, the beta estimate for observed ASQ3 was equal to -2.44 (95%CI -9.62;4.74, p=0.506), that is, the mean score of children in the group with 30 or more contacts was slightly but non-significantly lower than for children without any contacts.

To explore the effects of face-to-face visits during 2019, the two groups of 565 children (those visited in the last month and who had never been visited before the T1 interview) were reassessed at the T3 stage. On that occasion, the mean scores for the observed ASQ3 test were 202.7 and 201.2 (p=0.712). Therefore, there was

no evidence of a lasting effect of in-person visits in 2019 on performance at the end of 2021.

We also carried out separate analyses for each of the five domains of the ASQ3 scores (Tables 7 and 8), but found no significant differences between the intervention and comparison groups.

The lack of impact on child development led us to explore the correlation between the observed ASQ3 score and the number of contacts from 2019-2021 according to the PCF database. Figure 2, in which each child is represented by a dot regardless of his/her allocation group shows that there was no association (Pearson correlation coefficient of -0.0133; p=0.436). Because the PCF database does not discriminate face-to-face visits from virtual contacts, we repeated

Table 4. Follow-up rates at T3, according to baseline characteristics (T0).

Group	Comparison	Intervention	Total	P level (Intervention x Comparison)
Child's sex				
Male	79.1%	78.6%	78.8%	0.798
Female	81.1%	81.2%	81.2%	0.927
p intra-groups	0.320	0.180	0.099	
Child's age (T0)				
<5 months	85.3%	83.8%	84.5%	0.603
5-6 months	77.7%	77.9%	77.8%	0.959
7-8 months	79.0%	80.0%	79.5%	0.717
9-10 months	78.0%	77.7%	77.9%	0.915
11-12 months	82.1%	81.2%	81.6%	0.841
p intra-groups	0.068	0.245	0.008	
Mother's skin color				
White	76.5%	81.0%	78.6%	0.239
Black	75.2%	69.4%	72.3%	0.247
Brown	82.2%	81.2%	81.7%	0.549
p intra-groups	0.024	0.002	0.000	
Wealth quintiles				
Poorest quintile	80.4%	79.9%	80.2%	0.874
Second quintile	85.3%	82.7%	84.0%	0.378
Third quintile	80.0%	83.4%	81.7%	0.259
Fourth quintile	77.5%	76.7%	77.1%	0.791
Wealthiest quintile	77.3%	77.2%	77.2%	0.979
p intra-groups	0.076	0.098	0.006	
Mother's schooling (years)				
0-4	78.8%	78.1%	78.4%	0.886
5-8	73.9%	78.5%	76.2%	0.096
≥9	83.1%	80.0%	81.6%	0.085
p intra-groups	0.000	0.753	0.004	
Mother's depressive symptoms				
No	81.5%	80.4%	81.0%	0.504
Yes	75.9%	79.0%	77.5%	0.287
p intra-groups	0.016	0.526	0.032	

Source: Authors.

Table 5. Unadjusted ITT analyses of key outcomes in the comparison and intervention groups.

	Maximum score	Comparison		Intervention		P
		Mean (SD)	(SD)	Mean (SD)	(SD)	
Stimulation-interaction score (30-36 mo)	38	28.4	(4.4)	27.7	(5.0)	0.541
Stimulation-interaction score (>36 mo)	30	24.8	(3.9)	25.0	(4.1)	0.349
Responsive caregiving score	5	2.1	(0.7)	2.1	(0.8)	0.974
Inadequate disciplinary methods	8	1.6	(1.5)	1.5	(1.4)	0.066
Psychological problems		%		%		
About average		57.9		60.2		0.358
Borderline		15.5		13.7		
At risk		26.6		26.1		

Source: Authors.

Table 6. ITT regressions and LATE analyses, adjusted for age and municipality.

	MeI	Beta	95%CI	P
Stimulation-interaction score				
ITT		0.092	-0.215 0.399	0.558
LATE		0.004	-0.010 -0.018	0.562
Responsive caregiving score				
ITT		-0.008	-0.180 0.160	0.926
LATE		-0.00086	-0.0084 0.0066	0.822
Inadequate disciplinary methods score				
ITT		-0.060	-0.247 0.126	0.524
LATE		-0.004	-0.009 0.001	0.101
Psychological problems score				
ITT		-0.417	-0.950 0.112	0.122
LATE		-0.019	-0.040 0.010	0.133
Observed ASQ3 score				
ITT*		2.62	-2.16 7.41	0.282
LATE		0.08	-0.14 0.30	0.485
Reported ASQ3 score				
ITT*		1.87	-2.04 5.78	0.348
LATE		0.08	-0.10 0.26	0.373

*Adjusted for baseline ASQ3 and covariates.

Source: Authors.

these analyses using information on 758 children whose mothers were interviewed at T1, T3 and in all five rounds of telephone interviews. Pearson correlation coefficients with the observed ASQ3 scores were equal to 0.028 for face-to-face visits and 0.005 for virtual contacts, confirming the absence of associations.

Sensitivity analyses included restricting the sample to municipalities with strong PCF implementation according to the implementation study, subgroup analyses examining impact according to socioeconomic, maternal and child characteristics, and LATE analyses using results from the telephone interviews for estimating the instrumental variable. None of these analyses showed significant impacts of the program. Detailed results are available elsewhere (<http://www.epidemiio-ufpel.org.br/site/content/pesquisas/index.php?study=518>).

Implementation study

The findings from the implementation study are available at the UFPEL website (<http://www.epidemiio-ufpel.org.br/site/content/pesquisas/index.php?study=518>). Due to the pandemic, home visits were stopped in all municipalities,

mostly from March 2020 to around June 2021. During the 36 months of the evaluation (January 2019 to December 2021), the number of calendar months with ongoing home visits ranged from 5 to 35 in the 30 municipalities, with a median of 24.

The PCF monitoring database was used to assess the frequency of contacts and compliance with the original randomization. Figure 3 shows two circles for each municipality, the light blue circle being the average number of visits per child in the control group, and the dark blue circle the average for the intervention group. The number of visits was small in most municipalities, with means of 6.6 contacts in the comparison group and 28.2 in the intervention group. The median child in the study reached 36 months of age in June 2021, and at this time should be discharged from PCF. Therefore, the average number of monthly contacts should equal 30, and with weekly contacts 120. It is also observed that some municipalities disregarded the random allocation scheme, since the number of contacts was very similar in the intervention and control groups.

The upper part of Table 9 shows selected findings from the key informant interviews and reviews in the six states and 28 municipalities, giv-

Table 7. Mean scores by age (excluding 96 children with deficiency and 317 children with total score <30 points) in the five domains of reported ASQ-3 test in intervention and comparison groups.

	Comparison			Intervention			P
	N	Mean	(SD)	N	Mean	(SD)	
Reported ASQ-3	(36 months)						
Total (0-300)	203	223.3	(49.7)	191	218.6	(52.1)	0.367
Communication (0-60)	203	43.1	(13.8)	191	42.6	(13.6)	0.727
Broad motor coordination (0-60)	203	52.4	(11.6)	191	51.0	(11.1)	0.242
Fine motor coordination (0-60)	203	43.8	(16.0)	191	42.4	(16.5)	0.398
Problem solving (0-60)	203	39.4	(15.8)	191	39.2	(15.8)	0.916
Personal-social (0-60)	203	44.6	(12.7)	191	43.4	(13.9)	0.351
Reported ASQ-3	(42 months)						
Total (0-300)	672	219.2	(51.3)	687	222.7	(46.7)	0.190
Communication (0-60)	672	42.0	(15.0)	687	42.1	(14.3)	0.890
Broad motor coordination (0-60)	672	51.0	(11.0)	687	51.4	(11.1)	0.511
Fine motor coordination (0-60)	672	38.1	(15.2)	687	39.1	(14.2)	0.233
Problem solving (0-60)	672	39.1	(16.0)	687	40.6	(15.4)	0.077
Personal-social (0-60)	672	49.1	(12.5)	687	49.6	(11.0)	0.405
Reported ASQ-3	(48 months)						
Total (0-300)	370	218.8	(50.0)	375	219.9	(52.8)	0.763
Communication (0-60)	370	46.4	(15.2)	375	47.3	(15.6)	0.432
Broad motor coordination (0-60)	370	51.5	(11.4)	375	52.5	(10.3)	0.187
Fine motor coordination (0-60)	370	34.9	(17.0)	375	34.2	(17.2)	0.611
Problem solving (0-60)	370	38.0	(15.0)	375	38.5	(15.5)	0.659
Personal-social (0-60)	370	48.1	(10.7)	375	47.4	(11.4)	0.418

Source: Authors.

Table 8. Mean scores by age (excluding 96 children with deficiency and 317 children with total score <30 points) in the five domains of observed ASQ-3 test in intervention and comparison groups.

	Comparison			Intervention			P
	N	Mean	(SD)	N	Mean	(SD)	
Observed ASQ-3	(36 months)						
Total (0-300)	178	205.1	(52.6)	170	197.4	(62.0)	0.208
Communication (0-60)	178	40.1	(14.4)	170	39.4	(15.6)	0.664
Broad motor coordination (0-60)	178	49.8	(13.9)	170	48.2	(14.1)	0.288
Fine motor coordination (0-60)	178	44.0	(15.7)	170	41.2	(17.8)	0.112
Problem solving (0-60)	178	36.0	(17.0)	170	34.9	(17.1)	0.538
Personal-social (0-60)	178	35.1	(14.8)	170	33.6	(16.1)	0.377
Observed ASQ-3	(42 months)						
Total (0-300)	596	204.7	(59.1)	614	207.0	(54.6)	0.494
Communication (0-60)	596	38.5	(16.5)	614	38.3	(15.7)	0.871
Broad motor coordination (0-60)	596	47.6	(13.8)	614	48.3	(13.5)	0.369
Fine motor coordination (0-60)	596	37.2	(15.8)	614	38.0	(15.0)	0.397
Problem solving (0-60)	596	37.6	(16.6)	614	38.3	(16.1)	0.431
Personal-social (0-60)	596	43.8	(14.7)	614	44.0	(14.1)	0.812
Observed ASQ-3	(48 months)						
Total (0-300)	331	193.2	(58.8)	324	199.4	(58.8)	0.177
Communication (0-60)	331	41.1	(17.3)	324	42.6	(17.6)	0.289
Broad motor coordination (0-60)	331	49.3	(13.8)	324	50.8	(11.8)	0.137
Fine motor coordination (0-60)	331	31.8	(17.8)	324	32.0	(18.0)	0.890
Problem solving (0-60)	331	32.4	(15.4)	324	34.2	(16.0)	0.147
Personal-social (0-60)	331	38.6	(15.3)	324	39.9	(16.3)	0.294

Source: Authors.

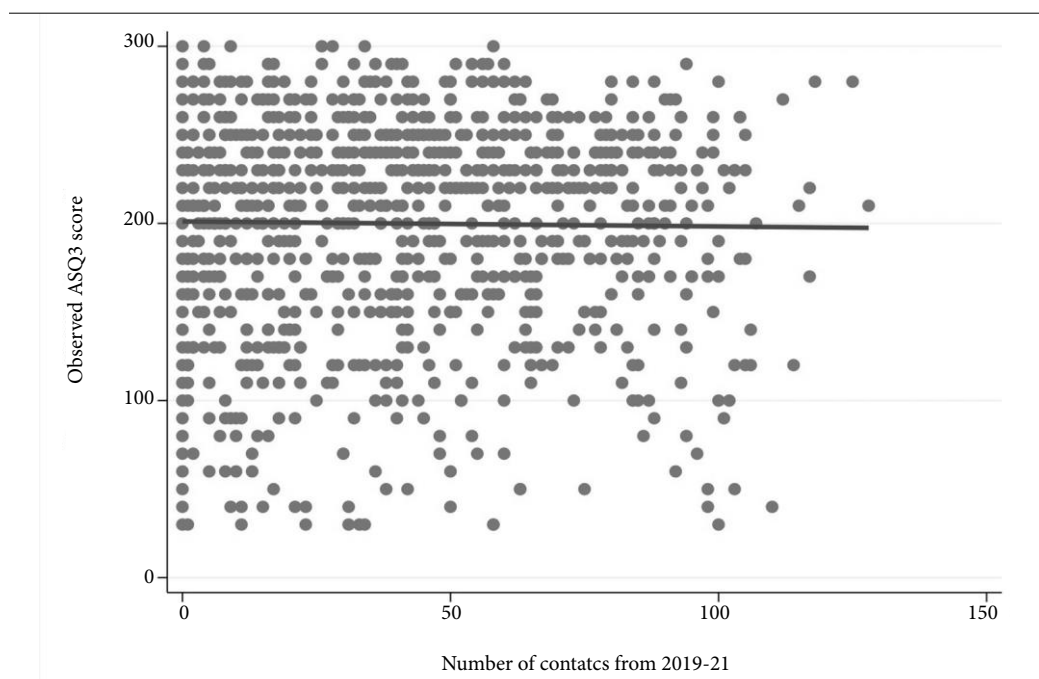


Figure 2. Scatter diagram of observed ASQ3 scores in the T3 study and the number of contacts from 2019-2021 according to the PCF database.

Source: Authors.

en that PCF activities had stopped in two of the 30 original municipalities. State PCF coordinators believed that only 30% of the municipalities had strong implementation. Staff turnover was an important issue – about half of the municipal coordinators were the same as one year earlier, and more than half of the municipalities failing to keep most of their home visitors since the launch of the program. In 60% of the municipalities, visitors were selected through a competitive process as opposed to personal indications. Four or more methods for contacting families during the pandemic were used in 43% of the municipalities, and only 20% had three or more vehicles for the program. The median number of families per visitor was 25, less than the number of 30 families recommended by the national program³⁸.

In each municipality, home visitors took a knowledge test and completed scales from 0 to 10 regarding their satisfaction with the program. In total, 363 visitors were assessed. These scores were averaged by municipality and Table 9 shows the mean values of municipal averages. A simple

knowledge test was applied to visitors, resulting in a mean score of 5.9 out of 10. Average scores by municipality ranged from 4.4 to 7.6.

The mean monthly salary in the 30 municipalities was equal to R\$ 1,103, or about US\$ 200, ranging from R\$ 400 to R\$ 1,450. Temporary contracts without fringe benefits were the prevailing type (67%) of arrangement. On average, home visitors gave a score of 5.3 out of 10 to their satisfaction with salaries. Regarding satisfaction with the number of hours worked, the average score was 7.3. Inadequate transportation arrangements for the visits were also frequently reported, with a mean score of 5.2.

In order to assess program fidelity, 182 home visits were observed by the study supervisors using a checklist of the PCF-recommended visit contents according to the child's age. The percentage of items addressed by the visitor during each visit relative to the expected items was calculated. On average, only 9.8% (SD=7.4) of recommended items were addressed, with municipal averages ranging from 1.0% to 71.0%. It was

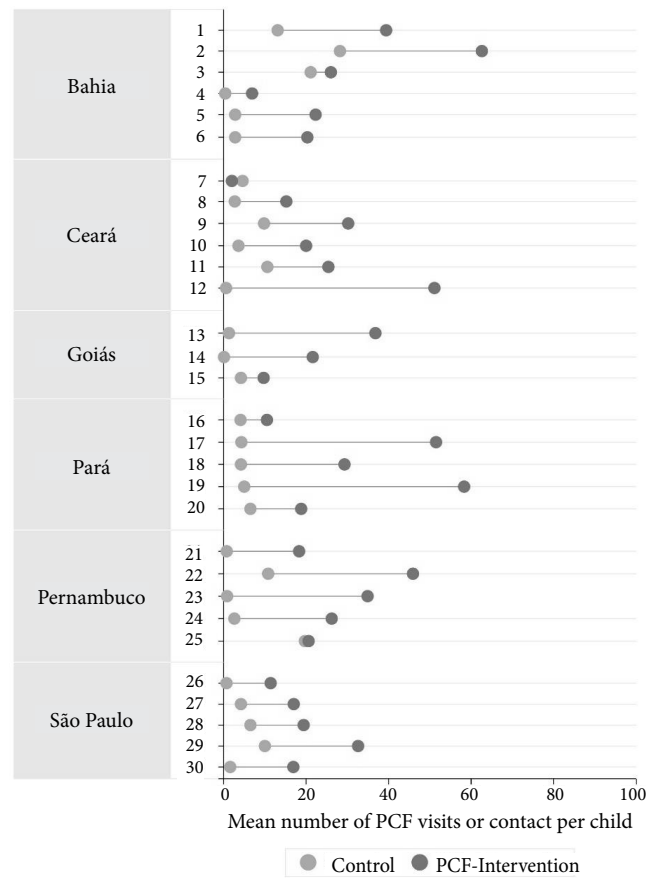


Figure 3. Mean number of PCF visits or contacts with the study children from January 2019 to December 2020 according to allocation group. Each line corresponds to one municipality, grouped by state.

Source: Authors.

not possible to carry out observations in eight municipalities, where visits were not taking place at the time of the research.

Twenty-eight focus groups were carried out (as two municipalities had no visitors at the time of the implementation study), with three to eight visitors per group. Most groups reported that the PCF was impacted by the pandemic due to full loss of contact (or remote contacts only) with families, lack of access by families to the internet or cell phones, and loss of the previously existing bonds with families and children. In most municipalities, the visitors expressed their dissatisfaction with transportation, wages and their satisfaction with the workload. Other positive points mentioned by the visitors included the good receptivity and trust by most families,

the strengthening of bonds between mother and child, the contributions of the PCF to child development, and the confidence built between visitors and children, mothers, and other family members.

Finally, a focus group was organized using Zoom software with all six state survey supervisors who carried out the implementation studies from 2019 to 2021. The topics discussed included characteristics of the home visits, training and working conditions of visitors, turnover, and management of the program. Supervisors reported low fidelity to program content during home visits; their observation of many visits without pre-defined contents and without taking into account the age of the children; completely untrained visitors who operated in some municipalities, with many children to cover (median of 26,

Table 9. Key results from the 2021 implementation study in the 30 municipalities.

Criteria	Proportion or mean
State PCF coordinator considered that the municipality had strong implementation	30%
Municipality coordinator was the same as in 2020	47%
Most home visitors were the same since the launch of the program	44%
Visitors were selected through a competitive process	60%
Four or more methods for contacting families were used during the pandemic (phone, voice/ audio message, text message, video call, other)	43%
Three or more vehicles were available for home visiting	20%
Home visitors' mean scores (scales from 0 to 10)	
Knowledge test scores	5.9
Satisfaction with salaries	5.3
Satisfaction with the number of hours worked,	7.3
Satisfaction with transportation arrangements	5.2

Source: Authors.

ranging from 11 to 37) and receiving low salaries; and high turnover among municipal coordinators and visitors (mainly due to municipal elections, low remuneration and inadequate employment contracts), thus hampering the program's continuity. In addition, supervisors reported coordination and management problems, starting with the lack of oversight of the PCF by state and municipal managements. Several problems were mentioned regarding the role of the CRAS, which in many municipalities were overloaded and dissociated from the PCF, without interacting sufficiently with the visitors. Some municipalities did not have PCF focal points within the CRAS. In few municipalities did municipal PCF supervisors regularly meet the home visitors in the CRAS. Other reported problems included the lack of data records available for program management and the lack of linkages between the PCF and *Bolsa Família*, which makes it difficult for families to understand and participate in both programs. Finally, an extremely worrying aspect was the total lack of supervision of the work of the visitors: in none of the municipalities did the supervisors accompany home visits periodically to provide support and guidance to the visitors.

Discussion

This section includes a discussion of the study design, summaries of the main results of the impact and implementation studies, a comparison with the existing literature and recommendations for public policies.

The study employed a rigorous design for assessing the impact of public health programs, because randomization resulted in two highly comparable groups of children at baseline (T0) and ensured that any trends in contextual variables would affect both groups. Because the numbers of children potentially eligible for the program far exceeded the numbers of vacancies in the selected municipalities, random allocation was ethically justifiable. The sample size was adequate to measure an impact compatible with that observed in the literature, with more than 3,000 children enrolled in 30 municipalities from six states. Follow-up rates were adequate (80% after three years) and similar across subgroups of children and families, indicating low risk of bias due to loss to follow-up. The main analyses were performed on an "intention-to-treat" basis, as originally planned, and were supplemented by additional approaches.

Other strengths include the longitudinal nature of the design and the use of multiple outcomes based on an *a priori*-defined impact model. These outcomes included the assessment of parent-child interactions, discipline, psychological attributes and child development, the latter using two versions of the ASQ3 test. In addition to face-to-face household interviews, the assessments relied on videotaped interactions between mothers and children, interviews with key informants, knowledge testing for PCF visitors and structured observation of home visits.

The COVID-19 pandemic resulted in changes to some aspects of the original design, including the cancellation of the second-year follow-up

(T2) interviews and the adoption of telephone interviews to monitor compliance during 2020 and 2021. In addition, the Ministry allowed municipalities to rely upon virtual channels to keep in touch with families and deliver counselling messages. The original effectiveness designed also had to be adopted. The findings of inadequate program implementation at the end of 2019 (T1) were communicated to the Ministry in early 2020. At that time, the research team considered that it would be unethical to fail to provide feedback to the Ministry. Consequently, during the years 2020 and 2021, the Ministry undertook a series of measures to strengthen implementation in the 30 municipalities. This affected the original study design that entailed a “pure” effectiveness evaluation under routine implementation conditions. Within the *continuum* from efficacy and effectiveness, the reinforcement of activities in the 30 municipalities – but not in the rest of the country – moved the study design towards the efficacy end of the scale³⁹.

Possible limitations of the study include the fact that the PCF home visits were planned to be interrupted at 36 months of age and – partially due to the pandemic – the median age of children at the T3 interview was 42 months, when many of them were no longer being visited. Nevertheless, the program was expected to have a lasting impact⁴⁰ that would persist after the visits ended. Another limitation is that structured observations of home visits could only be carried out in 22 of the 30 municipalities, mainly because visits were not taking place in eight municipalities.

The impact model¹⁸ built by the researchers in the planning phase of the study on the basis of the PCF guidelines^{15,29} postulated a chain of events necessary for the program to help improve child development. The model (described above in the Introduction) guided the selection of indicators and evaluation tools aimed at documenting the full process from the beginning of program implementation until its postulated impact on child development and on a broad reduction of vulnerability^{15,29}.

The combination of process and impact measurement strategies is essential for assessing the performance of any program^{19,20}. During the three-year evaluation, all links in the impact model were evaluated through various data collection strategies. Given that the model describes a causal chain in which all links are interconnected in a logical and sequential order, inadequate performance on any one of them was expected to compromise subsequent links.

The present impact results were robust and consistent across outcomes and analytical strategies. Intent-to-treat analyses failed to show any significant differences between intervention and control groups in terms of stimulation, interaction, disciplinary measures, psychological attributes or child development scores. Consistent with the above results, there was no association between the number of visits or contacts and the outcomes studied.

Possible reasons for the lack of impact were investigated through the implementation studies for which detailed results of which are available in additional documents (<http://www.epidemiologia-ufrpe.org.br/site/content/pesquisas/index.php?study=518>) The main explanations include: low coverage and low frequency of visits in the intervention group; lack of adherence to the randomization scheme at municipal level, resulting in contamination of the control group; interruption of face-to-face visits due to the pandemic; and implementation challenges. These will be discussed below.

Analyses of follow-up visits and telephone interviews, complemented by results from the PCF monitoring database, revealed important challenges in reaching high coverage in the intervention group. According to the database, the average number of visits over the three years was 28.2, ranging from 2.0 to 62.6 – that is, well below the expected number of weekly (or even monthly) contacts. Contamination was also evident in several municipalities, which apparently disregarded the randomization listings sent by the Ministry in 2019 and reinforced in 2020 after the feedback provided by the study. This finding may highlight communication challenges between the federal and municipal levels despite repeated attempts by the research team to raise awareness of implementation problems that became evident in the first year of the evaluation. Another possibility that cannot be ruled out is the use of criteria other than those recommended by the program, such as partisan or political criteria.

In a program where face-to-face interactions with caregivers and children are essential, the necessary suspension of home visits during the pandemic led to a major disruption. Although the Ministry has developed counseling materials for virtual contacts with families and most municipalities relied on such contacts to a greater or lesser extent, these are poor substitutes for home visits. Nevertheless, municipal managers reported that home visits were carried out on average during 24 of the 36 months of the study and

thus some degree of impact could be expected. It should also be noted that at the end of the first year of the program – before the pandemic, when only home visits were carried out – the mean ASQ3 scores in the intervention group were only 2.4 points or 0.05 SD higher in the intervention group than in the control group ($p=0.145$).

Low coverage in the intervention group, contamination of the control group and the pandemic may partially explain the lack of impact in the intention-to-treat analyses. However, LATE analyses and propensity score matching also showed no evidence of impact among the treated, that is, even among children who actually received the intervention. The complete lack of association between the number of contacts and the ASQ3 scores (Figure 2) was an unexpected finding that confirms the weakness of implementation. This was confirmed by results from the implementation study showing poor program management at municipal level, high turnover, inadequate selection and training of visitors, low wages and inadequate employment contracts, lack of field supervision and low fidelity to program contents during home visits.

In relation to similar studies in Latin America, positive results were observed for interventions applied on a much smaller scale than PCF. This was the case in studies from Mexico⁸, Peru⁷, Colombia⁶, and Brazil¹¹, all of which had intensive supervision and quality control of the visits. Two other Brazilian studies evaluated interventions similar to PCF. A randomized efficacy study in São Paulo¹³ showed no impact according to intent-to-treat analyses, but an improvement was detected in children visited more frequently. The non-randomized, observational evaluation of the *Primeira Infância Melhor* program in Pelotas¹⁴ did not show an impact on the full group of children studied, but only among those whose mothers were recruited during pregnancy. These last two studies may be affected by residual confounding that is inherent to observational analyses, as both the greater number of visits and early recruitment may be associated with family characteristics that would favor the child's development.

Our findings agree with the international literature, which shows that efficacy assessments – conducted in small-scale, strongly-implemented studies, under the direct monitoring of researchers – generally reveal a positive impact of home visits on child development. On the other hand, evaluations of the effectiveness of large-scale programs under routine implementation tend to have little or no impact⁴¹⁻⁴³.

It is also important to consider that, in its original design, the PCF was an extremely ambitious program, with weekly visits to millions of children in over 3,000 municipalities during which multiple contents had to be worked with families. In fact, this is the largest program ever implemented in any country relying upon home visits to promote child development. As discussed, the pandemic affected the characteristics of the program for several months, which together with the implementation challenges already present in the pre-pandemic period – particularly with regard to the quality of visits – jointly explain the lack of impact.

The difficulties faced by the PCF as a decentralized program, in which most actions take place at the municipal level with federal funding, are also observed in several Brazilian programs for health care, education and social assistance⁴⁴.

The primary objective of program evaluations is to provide input for policy and management improvement. Impact evaluations must answer the main question: is the program achieving its proposed objectives? If not, the process evaluation sub-study should suggest how to improve the program. Detailed results of the impact and implementation components are publicly available and have been communicated to the Ministry on numerous occasions to assist with program improvement.

From a broader point of view, the PCF has succeeded in placing early childhood, and specifically the promotion of neuro-psycho-motor development, at the highest level of the national citizenship agenda. The Ministry had the foresight and courage to fund a prospective randomized evaluation, initiated at an early stage of implementation, which provided inputs for program improvement. Few, if any, programs in Brazil or other low- and middle-income countries have been able to benefit from rigorous effectiveness assessments such as the current one. Even without a measurable impact, the program needs to be valued for its pioneering approach to implementation and evaluation. As demonstrated in a previous study, the PCF intervention group had a higher vaccination coverage than the control group²¹. In addition, the program may have impacts in other areas to be investigated, such as the identification and reporting of domestic violence and other multiple vulnerabilities, as well as strengthening intersectoriality and the connection of families with public services. The results show that several measures are needed to strengthen the implementation of the program,

so that future governments expand can continue to invest in early childhood as a key national priority.

Collaborations

IS Santos, TN Munhoz and CG Victora participated in the design of the research project. CG Victora, IS Santos, TN Munhoz, RS Barcelos, CC Bortolotto and C Blumenberg were responsible for the central coordination of the study. A Matijasevich, C Salum, HG Santos Júnior, L Marques, L Correia, MR Souza and PIC Lira were responsible for conducting the research in their respective states. C Blumenberg, TN Munhoz, RS Barcelos, CC Bortolotto and V Pereira performed the statistical analyses. All authors contributed to the interpretation of data. CG Victora and IS Santos wrote the first version of the article and all authors participated in the critical review of its intellectual content. All authors have approved the final version to be published and are responsible for all aspects of the work in ensuring the accuracy and integrity of any part of the work.

Acknowledgements

This article was carried out with data from the study “Impact Assessment of the Happy Child Program”, conducted by the Postgraduate Program in Epidemiology of the Federal University of Pelotas, in partnership with the Instituto Humanidades Artes e Ciências Professor Milton Santos (Bahia), Universidade Federal do ABC Paulista, Universidade Federal do Ceará, Universidade Federal de Goiás, Universidade Federal do Oeste do Pará, Universidade Federal de Pernambuco, and Universidade de São Paulo.

Funding

The study was funded by the Ministério da Cidadania, Fundação Maria Cecília Souto Vidigal, Itaú Social and the World Bank, with support from the United Nations Development Program.

Conflict of interests

In addition to the potential conflicts of interest related to the funding described above, Marina Chicaro and Elisa Altafim work at the Maria Cecília Souto Vidigal Foundation, a partner entity of the Ministry of Citizenship in actions aimed at early childhood.

References

1. Aboud F, Yousafzai A. Global health and development in early childhood. *Annu Rev Psychol* 2015; 66:433-457.
2. Britto P, Pyle S, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Perez-Escamilla R, Rao N, Ip P, Fernald LCH, MacMillan H, Hanson M, Wachs TD, Yao H, Yoshikawa H, Cerezo A, Leckman JF, Bhutta ZA; Early Childhood Development Interventions Review Group, for the Lancet Early Childhood Development Series Steering Committee. Nurturing care: promoting early childhood development. *Lancet* 2017; 389(10064):91-102.
3. Jeong J, Franchett E, Oliveira CR, Rehmani K, Yousafzai A. Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLoS Med* 2021; 18(5):e1003602.
4. Prado E, Larson L, Cox K, Bettencourt K, Kubes J, Shankar A. Do effects of early life interventions on linear growth correspond to effects on neurobehavioural development? A systematic review and meta-analysis. *Lancet Glob Health* 2019; 7(10):e1398-e1413.
5. Henwood T, Channon S, Penny H, Robling M, Waters C. Do home visiting programmes improve children's language development? A systematic review. *Int J Nurs Stud* 2020; 109:103610.
6. Attanasio O, Fernandez C, Fitzsimons E, Grantham-McGregor S, Meghir C, Rubio-Codina M. Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial *BMJ* 2014; 349:g5785.
7. Hartinger S, Lanata C, Hattendorf J, Wolf J, Gil AI, Obando MO, Noblega M, Verastegui H, Mäusezahl D. Impact of a child stimulation intervention on early child development in rural Peru: a cluster randomised trial using a reciprocal control design. *J Epidemiol Community Health* 2017; 71:217-224.
8. Fernald L, Kagawa R, Knauer H, Schnaas L, Guerra A, Neufeld L. Promoting child development through group-based parent support within a cash transfer program: Experimental effects on children's outcomes. *Dev Psychol* 2017; 53(2):222-236.
9. Knauer H, Kagawa R, Garcia-Guerra A, Schnaas L, Neufeld L, Fernald L. Pathways to improved development for children living in poverty: A randomized effectiveness trial in rural Mexico *Int J Behav Dev* 2016; 40(6):492-99.
10. Fundação Maria Cecília Souto Vidigal (FMCSV). *Visita domiciliar como estratégia de promoção do desenvolvimento e da parentalidade na primeira infância*. São Paulo: FMCSV; 2018.
11. Eickmann SH, Lima AC, Guerra MQ, Lima MC, Lira PI, Huttly SR, Ashworth A. Improved cognitive and motor development in a community-based intervention of psychosocial stimulation in northeast Brazil. *Dev Med Child Neurol* 2003; 45(8):536-541.
12. Grantham-McGregor S, Powell C, Walker S, Himes J. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican study. *Lancet* 1991; 338(8758):1-5.
13. Brentani A, Walker S, Chang-Lopez S, Grisi S, Powell C, Fink G. A home visit-based early childhood stimulation programme in Brazil—a randomized controlled trial. *Health Policy Plan* 2021; 36(3):288-297.
14. Silva EV, Hartwig F, Barros F, Murray J. Effectiveness of a large-scale home visiting programme (PIM) on early child development in Brazil: quasi-experimental study nested in a birth cohort. *BMJ Global Health* 2022; 7:e007116.
15. Brasil. Ministério do Desenvolvimento Social e Agrário (MDS). *A intersectoralidade na visita domiciliar*. Brasília: MDS; 2017.
16. Brasil. Lei nº 10.836, de 9 de janeiro de 2004. Cria o Bolsa Família. *Diário Oficial da União* 2004; 1 dez.
17. Brasil, Ministério da Cidadania. Portaria MC nº 664, de 2 de setembro de 2021. Consolida os atos normativos que regulamentam o Programa Criança Feliz/Primeira Infância no Sistema Único de Assistência Social - SUAS. *Diário Oficial da União* 2021; 6 set.
18. Bryce J, Victora C, Boerma T, Peters D, Black R. Evaluating the scale-up for maternal and child survival: a common framework. *Int Health* 2011; 3(3):139-146.
19. Habicht J, Victora C, Vaughan J. Evaluation designs for adequacy, plausibility and probability of Public Health programme performance and impact. *Int J Epidemiol* 1999; 28:10-18.
20. Donabedian A. The quality of care: how can it be assessed? *JAMA* 1988; 270:1743-1748.
21. Barcelos RS, Santos IS, Munhoz TN, Blumenberg C, Bortolotto CC, Matijasevich A, Salum C, Santos Júnior HGD, Santos LMD, Correia L, Souza MR, Lira PIC, Altafim E, Macana EC, Victora CG. Vaccination coverage in children up to 2 years old, receiving financial support from the Family Income Transfer Program, Brazil. *Epidemiol Serv Saude* 2021; 30(3):e2020983.
22. Santos IS, Munhoz TN, Blumenberg C, Barcelos R, Bortolotto CC, Matijasevich A, Santos Júnior HG, Santos LM, Correia LL, Souza MR, Lira PIC, Altafim E, Marino E, Macana EC, Silva RS, Ohana EF, Fontes MTA, Victora CG. Post-partum depression: a cross-sectional study of women enrolled in a conditional cash transfer program in 30 Brazilian cities. *J Affect Disord* 2021; 281:510-516.
23. Munhoz TN, Santos IS, Blumenberg C, Barcelos RS, Bortolotto CC, Matijasevich A, Santos Júnior HG, Santos LMD, Correia LL, Souza MR, Lira PIC, Altafim ERP, Macana EC, Victora CG. Fatores associados ao desenvolvimento infantil em crianças brasileiras: linha de base da avaliação do impacto do Programa Criança Feliz. *Cad Saude Publica* 2022; 38(2):e00316920.

24. Brasil. Ministério da Cidadania (MC). Secretaria de Avaliação e Gestão da Informação. *Estudo de Linha de Base da Avaliação de Impacto do Programa Criança Feliz. Cadernos de Estudos nº 35*. Brasília: MC; 2020.
25. United Nations Children's Fund (UNICEF). *Multiple indicator cluster surveys (MICS)* [Internet]. [acessado 2021 jan 20]. Disponível em: <https://mics.unicef.org/>.
26. Cox JGC, Murray UD, Jones P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. *J Affect Disord* 1996; 39(3):185-189.
27. Santos IS, Matijasevich A, Tavares BF, Barros AJ, Botelho IP, Lapolli C, Magalhães PV, Barbosa AP, Barros FC. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in a sample of mothers from the 2004 Pelotas Birth Cohort Study. *Cad Saude Publica* 2007; 23(11): 2577-2588.
28. Matijasevich A, Munhoz TN, Tavares BF, Barbosa AP, Silva DM, Abitante MS, Dall'Agnol TA, Santos IS. Validation of the Edinburgh Postnatal Depression Scale (EPDS) for screening of major depressive episode among adults from the general population. *BMC Psychiatry* 2014; 14:284.
29. Brasil. Ministério da Cidadania (MC). *Manual do visitador: um olhar sobre a visita domiciliar* [Internet]. 2021 [acessado 2022 jan 20]. Disponível em: <https://www.gov.br/cidadania/pt-br/acoes-e-programas/crianca-feliz/publicacoes-1/MANUALDOVISITADORVERSOFINAL.pdf>.
30. Reichenheim ME, Moraes CL. Adaptação transcultural do instrumento Parent-Child Conflict Tactics Scales (CTSPC) utilizado para identificar a violência contra a criança. *Cad Saude Publica* 2003; 19:1701-1712.
31. Tremblay RE, Nagin DS, Séguin JR, Zoccolillo M, Zelazo PD, Boivin M, Pérusse D, Japel C. Physical Aggression During Early Childhood: Trajectories and Predictors. *Pediatrics* 2004; 114(1):e43-e50.
32. Squires J, Bricker D. *Ages and Stages Questionnaires Third Edition (ASQ3)*. Baltimore: Paul H. Brookes Publishing Co.; 2009.
33. Goodman R. The strengths and difficulties questionnaire: a research note. *J Child Psychol Psychiatry* 1997; 38:581-586.
34. Fleitlich B, Cortázar P, Goodman R. Questionário de capacidades e dificuldades (SDQ). *Infanto Rev Neuropsiq Inf Adol* 2000; 8(1):44-50.
35. Prime H, Browne D, Akbari E, Wade M, Madigan S, Jenkins JM. The development of a measure of maternal cognitive sensitivity appropriate for use in primary care health settings. *J Child Psychol Psychiatry* 2015; 56:488-495.
36. Schneider A, Rodrigues M, Falenchuk O, Munhoz TN, Barros AJD, Murray J, Domingues MR, Jenkins JM. Cross-Cultural Adaptation and Validation of the Brazilian Portuguese Version of an Observational Measure for Parent-Child Responsive Caregiving. *Int J Environ Res Public Health* 2021; 18(3):1246.
37. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42(2):377-381.
38. Brasil. Ministério de Desenvolvimento Social (MDS). *Criança Feliz: guia para visita domiciliar*. 2ª ed. Brasília: MDS; 2017.
39. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004; 94(3):400-405.
40. Heckman J. Skill Formation and the Economics of Investing in Disadvantaged Children. *Science* 2006; 312(5782):1900-1902.
41. Engle PL, Fernald LC, Alderman H, Behrman J, O'Garra C, Yousafzai A, Mello MC, Hidrobo M, Ulkuer N, Ertem I, Iltus S; Global Child Development Steering Group. Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *Lancet* 2011; 378(9799):1339-1353.
42. Victora C, Barros F, Assunção M, Restrepo-Méndez M, Matijasevich A, Martorell R. Scaling up maternal nutrition programs to improve birth outcomes: A review of implementation issues. *Food Nutr Bull* 2012; 33(2 Supl.):S6-S26.
43. List J, Suskind D, Supplee L. *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It*. 1ª ed. Oxfordshire: Routledge; 2021.
44. Arretche M. Políticas Sociais no Brasil: descentralização em um Estado federativo. *Rev Bras Cien Soc* 1999; 14(40):111-141.

Article submitted 22/08/2022

Approved 24/08/2022

Final version submitted 26/08/2022

Chief editors: Romeu Gomes, Antônio Augusto Moura da Silva

