FUNDAÇÃO GETULIO VARGAS

ESCOLA DE ECONOMIA DE SÃO PAULO

GABRIEL WEBER COSTA

INTERVENTIONS ON HUMAN CAPITAL FORMATION AMONG VULNERABLE POPULATIONS: EXPERIMENTAL EVIDENCE FROM TWO LARGE-SCALE PROGRAMS IN BRAZIL

São Paulo

2022

GABRIEL WEBER COSTA

INTERVENTIONS ON HUMAN CAPITAL FORMATION AMONG VULNERABLE POPULATIONS: EXPERIMENTAL EVIDENCE FROM TWO LARGE-SCALE PROGRAMS IN BRAZIL

Tese de doutorado apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas, como requisito para a obtenção de título de Doutor em Economia.

Orientador: Prof. Dr. André Portela Souza

São Paulo

2022

Costa, Gabriel Weber.

Interventions on human capital formation among vulnerable populations: experimental evidence from two large-scale programs in Brazil / Gabriel Weber Costa. - 2022.

118 f.

Orientador: André Portela Fernandes de Souza.

Tese (doutorado CDE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Avaliação de pesquisa (Programas de ação social). 2. Capital humano. 3. Populações vulneráveis - Brasil. 4. Integração social. I. Souza, André Portela Fernandes de. II. Tese (doutorado) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 304(81)

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

GABRIEL WEBER COSTA

INTERVENTIONS ON HUMAN CAPITAL FORMATION AMONG VULNERABLE POPULATIONS: EXPERIMENTAL EVIDENCE FROM TWO LARGE-SCALE PROGRAMS IN BRAZIL

Tese de doutorado apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas, como requisito para a obtenção de título de Doutor em Economia.

Campo de Conhecimento:
Microeconomia Aplicada

Data de aprovação:
07/07/2022

Banca examinadora:

Prof. Dr. André Portela Souza (Orientador)
FGV EESP

Prof. Dr. Bruno Ferman
FGV EESP

Prof. Dr. Nicolás Ajzenman
FGV EESP

Prof. Dr. Naércio Menezes Filho
Insper

Prof. Dr. Pedro Carneiro
UCL

AGRADECIMENTOS

Aos tantos professores que contribuíram de formas diversas à minha formação durante esses anos e que me apresentaram à excelência no ensino e pesquisa.

Ao meu orientador, André Portela, por toda a paciência e por me ensinar a enxergar os caminhos.

Aos colegas de mestrado e doutorado. Espero encontrar ambientes leves e colaborativos como esse em todos os lugares.

Aos colegas do Clear, sem os quais este trabalho não seria possível. Em especial à Lycia Lima, fundamental nessa trajetória e referência em tudo que faço.

À minha família, sempre presente para brindar aos sucessos e para enfrentar os momentos mais difíceis.

Aos meus amigos, que são também família. Não seria nada sem vocês.

À minha esposa, Mariane Furtado, que estava lá para me escutar quando decidi embarcar nessa jornada, e que seguirá comigo em todas as jornadas daqui pra frente.

À minha irmã, Suélem Costa, pelo exemplo a ser seguido.

À minha mãe, Angela Weber, por sempre acreditar em mim e por me dar o que há de mais precioso.

Ao meu pai, Aldemir Costa, meu super-herói, que agora comemora comigo de outro lugar.

Agradeço pelo apoio da Fundação Getulio Vargas, do Banco Interamericano de Desenvolvimento e da Fundação Maria Cecília Souto Vidigal na execução da avaliação experimental do Programa Primeira Infância Melhor.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

ABSTRACT

Essay 1 - Home Visiting, Child Development and Parenting: Experimental Evidence from the Primeira Infância Melhor Program

We study a large-scale parenting program in the south of Brazil whose goal is to promote the full development of children during the early childhood period. We explore random assignment of eligible families into a treatment and a control group to evaluate program effects on measures of child development, parenting, caregiver mental health and family access to services. Our results suggest that, under regular implementation, the program was effective in improving motor skills of treated children. We also find evidence of improvements in parenting skills and family interactions consistent with program design. Children with caregivers of lower educational backgrounds seem to benefit the most from the intervention.

Key-words: Child development, early childhood, parenting, RCT.

Essay 2 - Can Human Capital Investments for At-Risk Youth Affect Welfare Dependency? Experimental Evidence from the Protejo Program

In this paper, we present evidence on the effects of a comprehensive training program that encompasses general, technical and soft skills, as well as activities related to social participation and citizenship building, under a curriculum of 800 hours of activities. Targeted at youth at-risk, vacancies were assigned randomly to a cohort applying to the program in Rio de Janeiro, allowing for comparability between treatment and control groups. We use administrative data to assess program impacts on formal employment and welfare receipt, and explore primary data related to family formation outcomes. Our results suggest that male beneficiaries experience persistent gains in formal employment, while women have an increased chance of receiving welfare benefits up to ten years after the program. Women with lower educational levels and with no children at the baseline are most affected, the same subgroups that face an increase in fertility after two years following the end of the intervention.

Key-words: Human capital, vocational training, social inclusion, RCT.

RESUMO

Ensaio 1 - Visitação Domiciliar, Desenvolvimento Infantil e Parentalidade: Evidências Experimentais do Programa Primeira Infância Melhor

Neste artigo, é estudado um programa de parentalidade em larga escala no sul do Brasil cujo objetivo é promover o pleno desenvolvimento das crianças durante a primeira infância. A atribuição aleatória de famílias elegíveis em grupos de tratamento e controle é utilizada para avaliar os efeitos do programa em medidas de desenvolvimento infantil, parentalidade, saúde mental do cuidador e acesso da família a serviços. Os resultados sugerem que, sob condições normais de implementação, o programa foi eficaz em melhorar as habilidades motoras das crianças tratadas. Também são encontradas evidências de melhorias nas habilidades parentais e nas interações familiares, consistentes com o design do programa. Crianças com cuidadores de menor escolaridade parecem se beneficiar mais da intervenção.

Palavras-chave: Desenvolvimento infantil, primeira infância, parentalidade, avaliação experimental.

Ensaio 2 - Investimentos em Capital Humano para Jovens em Risco Podem Afetar a Dependência de Programas Sociais? Evidências Experimentais do Programa Protejo

Neste artigo, são apresentadas evidências sobre os efeitos de um programa de treinamento que engloba competências gerais, específicas e socioemocionais, bem como atividades relacionadas à participação social e cidadania, sob um currículo de 800 horas de atividades. Focalizando jovens em situação de risco, as vagas para uma coorte de inscritos no Rio de Janeiro foram atribuídas aleatoriamente, permitindo a comparabilidade entre os grupos de tratamento e controle. São utilizados dados administrativos para avaliar os impactos do programa no emprego formal e na participação em programas sociais, e dados primários para observar variáveis relacionadas à formação de família. Os resultados sugerem que os beneficiários do sexo masculino experimentam ganhos persistentes no emprego formal, enquanto as mulheres têm uma chance maior de serem beneficiárias de programas sociais até dez anos após o programa. Mulheres com menor escolaridade e sem filhos no período pré-tratamento são as mais afetadas, os mesmos subgrupos para os quais é observado um aumento da fecundidade após dois anos do fim da intervenção.

Palavras-chave: Capital humano, educação vocacional, inclusão social, avaliação experimental.

List of Figures

List of Tables

TABLE OF CONTENTS

1 Home Visiting, Child Development and Parenting: Experimental Evidence from the Primeira Infância Melhor Program[1]

1.1 Introduction

The early childhood period, commonly defined to span the gestational period until age 6, has been largely documented to be a defining moment in the life cycle of an individual towards the full development of their skills and capabilities. Such an importance seems to be intrinsically related to the process of neurodevelopment, which starts in the uterus and continues through childhood, adolescence and youth (CURRIE; ALMOND, 2011).

In turn, the development achieved at early ages serves as a basis for future skill acquisition, in a process of self-productivity and complementarity by which skills achieved at some point in life increase subsequent development of abilities and increase productivity of human capital investments (CUNHA; HECKMAN, 2007; CUNHA et al., 2010). As such, early development brings huge potential benefits throughout childhood and even adulthood, as several studies have documented. Conversely, gaps in development early in life pose a seemingly permanent threat, since they are hard to compensate by later human capital investments and lead inequalities visible in the first years of life to remain throughout the life cycle (SHONKOFF et al., 2000; CUNHA et al., 2006; PAXSON; SCHADY, 2007).

High-quality home visiting programs are promising policy alternatives to promote the full development of young children living in contexts of socioeconomic vulnerability, providing families with the necessary skills to achieve this goal. Influential studies have analyzed programs of this kind aimed at promoting early childhood development. Experimental evidence from the well-known Reach Up and Learn, for example, implemented in Jamaica and later in countries such as Colombia and India, have shown that the home visiting program targeting vulnerable stunted children of up to two years of age was able to generate persistent improvements in cognitive abilities, better performance at school and an increased income level up to three decades after the intervention (GRANTHAM-MCGREGOR et al., 1997; GERTLER et al., 2014; GERTLER et al., 2021; WALKER et al., 2022). Preschool initiatives such as the Perry Preschool Program, Abecedarian and Head Start have also been shown to lead to cognitive gains and increased school performance, as well as benefiting healthy behaviors and health outcomes in adulthood (CURRIE; THOMAS, 1995; GARCES et al., 2002; HECKMAN et al., 2010; HECKMAN et al., 2013; CONTI et al., 2016).

[1] This study was approved by the Comissão Nacional de Ética em Pesquisa (CONEP/Plataforma Brasil) on June 19, 2018 under approval number 2.747.117, and by the Comitê de Conformidade Ética em Pesquisas Envolvendo Seres Humanos (CEPH/FGV) on February 9, 2018, under number 01/2018. The trial was registered in the American Economic Association's registry for randomized control trials under ID AEARCTR-0006799 (https://www.socialscienceregistry.org/trials/6799).

Despite the well documented short and long-term impacts that high-quality small-scale programs such as the home visiting intervention in Jamaica are able to produce, the evidence base is much scarcer when considering large-scale initiatives, particularly for children in low and middle-income countries. Nevertheless, this evidence is central for the policy debate in these regions, and there are challenges in various domains when scaling up an intervention, e.g. related to planning, organizational design, content adaptation and assuring quality and sustainability (RADNER et al., 2018).

An emerging body of literature has recently begun to address this gap by conducting randomized trials in various developing countries. For instance, Araujo et al. (2021) study a large-scale home visiting program implemented in rural Peru and estimate small gains in problem solving and communication skills (0.10 and 0.11 standard deviations (σ), respectively) after two years. Studies of large-scale initiatives in countries such as China (SYLVIA et al., 2021), Colombia (ATTANASIO et al., 2014), India (ANDREW et al., 2020), and Bangladesh (BOS et al., 2021) estimate increased cognition among treated children, with related gains in language skills (ATTANASIO et al., 2014; BOS et al., 2021), the home stimulation environment (ATTANASIO et al., 2014; ANDREW et al., 2020), and parenting skills (SYLVIA et al., 2021), but with some evidence of fade out after two years in the Colombian case (ANDREW et al., 2018).

The present study aims to evaluate the 'Primeira Infância Melhor' Program (Better Early Childhood, PIM), an ongoing large-scale home visiting program targeting socioeconomically disadvantaged families to promote the health, cognitive and socioemotional development of young children during their early years of development in the Brazilian State of Rio Grande do Sul. PIM is a State Government program that exists since 2003. It was a pioneering program when it was implemented, and it has served as a model for similar state and federal programs since.

The ultimate goal of PIM is to promote the full development of young children from disadvantaged families by fostering an environment conducive to good parenting. In order to achieve such goals, families are assisted through personalized home visits from the program's group of trained professionals, during which parents are counseled regarding parenting practices that stimulate their children, as well as through small group community meetings where such practices are discussed. These activities aim to strengthen the quality of parenting and home environment among vulnerable populations. We take advantage of program expansions to new communities within already-served municipalities to randomize eligible children into a treatment group and a waiting list. Such comparable groups allow us to robustly assess program impacts on outcomes related to child development, parenting practices, caregiver mental health and family access to services.

Few studies have attempted to evaluate the impacts of PIM. Ribeiro et al. (2018) use a difference-in-differences approach with municipal data and find that PIM led to small reductions

in infant mortality by external causes when the program had been in place for more than 7 seven years in a given municipality. Junior et al. (2022) also employ difference-in-differences estimation with school-level identification of students potentially targeted by PIM in benefited municipalities, and find reductions in reported school violence, such as verbal or physical abuse, attacks or threats and robberies or thefts. Silva et al. (2022) use a propensity score matching algorithm to create a control group with individual-level data from a birth cohort in one municipality benefited by PIM and find no evidence of impacts on child development, except for a subgroup of families that received the intervention since the gestational period.

Our contribution is thus twofold. We add to the the growing body of literature documenting the effects of large-scale home visiting programs aimed towards child development in developing countries, and we do so by analyzing an understudied intervention with almost two decades of history that has directly impacted hundreds of thousands of families in southern Brazil since its start.

We analyze a wide range of outcomes related to the main goals of the program. Our primary data sources include measures of child development and health, parenting skills and practices, including methods used to discipline children and family interactions, an assessment of the mental health of caregivers and variables related to access to public services. Furthermore, in spite of concerns with attrition rates, our identification strategy stems from individual-level randomizations, allowing us to follow a much more credibly comparable control group, compared to previous assessments of PIM.

Our results suggest that, under regular conditions of implementation, PIM is effective in promoting child development. We find a statistically significant effect of 0.18 standard deviations on a measure of global development, which seems to be concentrated as an effect on fine motor skills (with gains of 0.18-0.22σ). Girls seem to benefit mostly in terms of fine motor skills, while boys show increased gross motor skills. Further heterogeneity analyses suggest that all impacts are concentrated among children with caregivers of lower educational attainment. Concerning parental practices, we estimate that PIM improves a measure of parental skills at a magnitude of 0.18 standard deviations, and that the program leads to a reduction of 21.9% in the use of physical punishments to discipline children in the treatment group. We also find improvements in the home environment, with an increase of 17.5% in the weekly hours spent by caregivers with treated children and an increase of 13.2% in the likelihood of family members interacting with the child everyday to sing and teach songs.

The remainder of this paper is organized as follows. Section 2 presents the main components of the program and details our experimental design. Section 3 describes the evaluation sample, our data sources and tests for differential attrition and balancedness across treatment arms. Section 4 outlines our estimation strategy. Finally, we present our results in section 5 and conclude in section 6.

1.2 Institutional Context and Study Design

1.2.1 The 'Primeira Infância Melhor' Program

In 2003 the Brazilian State of Rio Grande do Sul established the home visiting program called "Better Early Childhood" Program (PIM). The aim of PIM is to promote the early development of children from families in socioeconomically vulnerable contexts. The program combines weekly home visiting and community meetings with the objective of strengthening parental skills and ultimately fostering the full development of children – e.g., through playful stimulation exercises to teach shapes and colors. PIM has become a public policy by State Law nº 12.544/2006 and is now the reference for a national public policy, the program "Criança Feliz".

PIM has three overarching official goals: (i) promoting the full development of children during the early childhood period, in domains of motor, cognitive, socioeconomial and communication/language skills; (ii) strengthening parental interactions, family bonds and family protagonism; (iii) guaranteeing access to the public network of services via intersectoral integration. Weekly home visits by trained professionals last for about 60 minutes, during which visitors counsel families to foster the development of their children through activities specially crafted to promote learning by stimulating the creative, physical and emotional capacities of the child.

Specifically, a given home visit can be divided into three moments (VERCH, 2017). First, the home visitor is instructed to actively listen and check on the the family, discussing that week's progress and proposing a new set of activities to be performed. All activities are directed to the child caregiver to perform with the child, as a means to foster family bonds and commitment of the family with the proposed activities. The second moment then comprises the actual execution of play activities by the family and child, as facilitated by the visitor. Finally, family and visitor execute a quick informal assessment of the activity performed to identify progress and challenges, clarify questions and reinforce the importance of the activity proposed. Visitors can also identify risk factors within beneficiary families and guide them to seek adequate health or social assistance care, acting both as preventive and early-treatment mechanisms and as an entry point into the network of government services.

There are three modalities through which PIM operates: (i) children aged 0 to 3 receive weekly visits; (ii) children aged 4 to 6 receive regular visits that can happen weekly, biweekly or monthly; and (iii) expectant mothers participate in biweekly community meetings organized by program visitors, during which they are counseled regarding proper ante-natal care. For the purposes of our study, we randomize eligible families from groups (i) and (iii), during which we are less likely to have children leaving the program early to enter childcare facilities.[2]

[2] Expectant mothers in the evaluation sample are usually far enough into pregnancy that children randomized into

The financing of the PIM is carried out with funds from the state government and the municipalities that join the program. Up to the end of 2021, the government of the state of Rio Grande do Sul secured funds for each municipality based on its number of active visitors and their contractual working hours. Municipalities then complement the funds received to cover for additional expenses. Since its creation, the program has benefited over 200,000 families, including over 250,000 children and more than 60,000 expectant mothers. As of June 2022, PIM is present in 209 municipalities in the state, employing almost 1,300 visitors and caring for over 30,000 families.

Program adoption in a given municipality takes place through a mutual agreement between state and municipal governments. PIM is continuously offered by the state government to all municipalities in the state, which adhere to the program voluntarily. The municipalities provide local support for the development of PIM activities by hiring and training visitors and allocating infrastructure within the communities that receive the intervention (selected by local management). The state provides technical assistance and oversees program implementation. In 2019, the state of Rio Grande do Sul allocated a budget of R$ 16.8 million (around US$ 4.35 million at the time) to support program in the following year – this amount does not include local budgets.

Although the program has some degree of centralization in the State government, the selection of beneficiaries takes place at the community level, made by home visitors and their immediate local supervisors (usually experienced visitors who have been put in charge of a team of their colleagues). Under usual program implementation, visitors working in a given community select new families on a continuous base as long as they have not filled their quota (around 14-20 families per visitor, depending on the type of contract), taking into consideration several socioeconomic criteria[3]. Visitors are selected by municipalities and undergo training offered by the local program management, and each visitor can have a 20, 30 or 40 weekly hours contract.

1.2.1.1 Implementation during the covid-19 pandemic

The intended implementation of PIM through weekly home visits was severely affected starting at the onset of the covid-19 pandemic in early 2020. State-level program officers inquired implementing municipalities between late April and early May 2020 about the status of home visits.[4] From a total of 195 municipalities, 126 reported having stopped all home visits, and a

the treatment group are exposed to the same treatment regime of those in group (i).

[3] These criteria include the following: family income per capita; eligibility for cash transfer programs; housing conditions; household density; presence of women with high-risk pregnancy; history of infant morbidity and mortality; children with more than one hospitalization during the first year of life; mothers diagnosed with postpartum depression; families with caregivers, pregnant women and/or children with alcohol or drug-related problems; family member in specialized care and/or psychiatric treatment; illiteracy or low maternal or primary caregiver education; children not enrolled in the formal education network; domestic violence; family member in detention (VERCH, 2017).

[4] Their results are available at: https://www.pim.saude.rs.gov.br/site/pesquisa-1-pim-em-tempos-de-coronavirus/.

further 46 reported having partially suspended visitation, which is equivalent to over 90% of implementing municipalities altogether.

Among affected municipalities, 92% reported contacting families remotely, with resources varying across the state: 19% sent written guidance to families, 26% used recorded audios, 19% used recorded videos, 30% used telephone calls, and 6% contacted families by videoconference. Almost 3/4 of municipalities reported not being able to reach all previously-served families, with 35% percent of them only reaching less than half of intended families by remote means. The main challenges in contacting families were lack of internet access or mobile phones among vulnerable populations.

The earliest official state-level guidance on visitations during the pandemic was issued in April 2020[5] and mainly recommended keeping in touch with families by diversifying means of communication, including telephone calls or WhatsApp messaging, which seems to have been followed by municipalities. Official guidance was updated in April 2021[6], corroborating the possibility of remote or hybrid care but leaving decisions about which modality to follow up to each municipality.

Our survey data also illustrates the degree to which program implementation was affected during the pandemic scenario. As described later, our interviews took place between August 2021 and February 2022. Among interviewed families who reported having received any type of PIM care in 2020 or 2021 (143 families), close to 20% were only contacted remotely and approximately 40% went through a hybrid modality. Also, while 75% of families reported receiving weekly visits in 2019, this number went down to 60% in 2020 and 52% in 2021, with equivalent increases in biweekly or monthly visitation. Additionally, we also observed a decrease in the duration of visits, with only 55% of families reporting visits of 30 to 45 minutes, compared to 79% in 2019.

Overall, the data indicate that the covid-19 pandemic severely interfered with the intended implementation of PIM, both in the extensive margin – with fewer families receiving the intervention – and the intensive margin – with a shift to remote or hybrid care and a decrease in treatment intensity in terms of frequency and duration.

1.2.2 Experimental design

To conduct this study, we worked with program management at the state level to contact willing municipalities and invite them to participate in the evaluation. Because these were places where PIM was already in full operation, their participation in the study was conditional on a planned expansion to a new neighborhood, where a randomization would then take place to select beneficiary families.

[5] Available at: https://www.pim.saude.rs.gov.br/site/wp-content/uploads/2020/04/Mail-PIM-Covid-19-2-web.pdf.
[6] Available at: https://www.pim.saude.rs.gov.br/site/segunda-nota-de-orientacao-do-pim-pcf-em-tempos-de-novo-coronavirus-e-atualizada/.

Negotiations to implement this design began in 2017 at the state level and continued from 2018 until 2021 with the within-municipality lotteries. To enable the randomization procedure in each new community in participating municipalities, teams of home visitors were allocated to identify potential beneficiary families, following their usual selection criteria. We then performed a randomization to select families whenever these was an excess of demand, i.e., whenever the size of the list of potential beneficiaries exceeded local supply capacity.

For each list of eligible families, we randomized families into two groups: (i) those who would be offered to receive the intervention (the size of this group was determined by local program capacity), and (ii) a waiting list, to allow for replacements in the offer-group after refusals from the first draw. The formation of these two groups via lottery ensures their comparability over time. We hereafter refer to these groups as treatment and control groups, respectively.

Whenever there are enough eligible individuals in a given community/municipality, we further stratify the sample by child age group[7]. We generally aimed to have strata sizes of at least 30 individuals, varying according to the local program capacity and population size of eligible families. We use the terms "family" and "child" to refer to a unit in our sample interchangeably because only one child per family was allowed to be part of the study. The selected child was always the youngest of all siblings or chosen at random in case of twins. PIM visits are always individualized to each child, and families with multiple children receive more than one weekly visit to accommodate activities for all. We monitored compliance at the family level, meaning that children in our sample randomly allocated to the control group had all of their siblings marked as ineligible for visits, to avoid contamination.

1.3 Sample and Data

1.3.1 Sample

Our total sample for this study is composed of 2,419 children, divided into 61 strata in 15 municipalities, as detailed in Table 1. Porto Alegre (the capital of Rio Grande do Sul and its biggest city) appears twice because we conducted randomizations in four different communities, two at each date.

Although the percentage of treated children in each municipality varies from 42.9% to 72.3%, the overall treatment group size lies just below the target 50% mark, for a total of 1,191 children in the offer-group. Strata sizes also present some variation, with noticeable outliers in Canoas and the first draw of Porto Alegre (with above-mean averages of 134.6 and 62.8, respectively), and Viamão, the second draw of Porto Alegre and Palmeira das Missões, all with averages below 20. In spite of these, the average stratum in our sample is composed of roughly

[7] The age groups we considered for stratification were based on age groups for the Ages and Stages Questionnaires (ASQ-3).

Table 1 – Sample structure

Municipality	Lottery date (month/year)	Total sample size	N. treated	% treated	Number of strata	Average strata size
Porto Alegre	03/2018	628	273	43.5	10	62.8
Serafina Corrêa	04/2018	59	31	52.5	2	29.5
Viamão	09/2018	151	92	60.9	10	15.1
Canoas	09/2018	673	289	42.9	5	134.6
Porto Alegre	10/2018	148	74	50	10	14.8
Palmeira das Missões	10/2018	55	28	50.9	5	11
São Borja	11/2018	204	132	64.7	5	40.8
Rio Grande	09/2019	56	34	60.7	1	56
Alvorada	10/2019	78	45	57.7	2	39
Pelotas	01/2020	92	40	43.5	3	30.7
Caxias do Sul	02/2020	58	30	51.7	2	29
São Paulo das Missões	03/2020	32	15	46.9	1	32
Uruguaiana	03/2020	64	34	53.1	2	32
Santa Maria	01/2021	47	34	72.3	1	47
Cachoeirinha	03/2021	29	20	69	1	29
Ibirubá	05/2021	45	20	44.4	1	45
Total		2419	1191	49.2	61	39.7

Note – Each row represents a round of lottery draws. The total sample size is equal to the total number of children in both treatment arms for each lottery draw. N and % treated refer to the absolute and relative number of children randomly allocated into the treatment group, respectively. Number of strata is the total number of randomization blocks for each lottery date, and the average strata size is equal to the number of children in that row divided by the number of strata.

40 children.

1.3.2 Primary Data

1.3.2.1 First survey (2018)

Our first round of data collection took place in December 2018. Our goal was to interview the 1,659 children from our first five randomizations in Porto Alegre, Serafina Corrêa, Viamão and Canoas (listed under the first five lines of Table 1). Out of these, 311 were left out of interviewing efforts because program management informed us they had failed to be contacted to receive an offer to participate, meaning they had been registered wrongfully and should not have been included in the randomization from the start. This left us with 1,348 children to be interviewed, out of which we successfully contacted 502, an interview rate of 37.2% (or 30.3%, if we consider the 311 that were excluded).

Door-to-door data collection was conducted within communities by a group of interviewers specifically hired for this purpose by a subcontracted firm specialized in surveys. The interviewers

were trained by our research team and assisted by community leaders and PIM visitors.

Importantly, this first round was initially planned to serve as a baseline data collection. However, logistical issues and delays in signing contracts with the data collection firm meant that several months had passed since the lottery date for part of our sample. Table 2 summarizes the average time in months between lottery and interview dates.

Table 2 – First round of data collection

Municipality	Lottery date	Average time in months since lottery	Standard deviation	Interviewed sample
Porto Alegre	03/2018	9.04	0.15	113
Serafina Corrêa	04/2018	8.49	0.14	30
Viamão	09/2018	3.39	0.16	60
Canoas	09/2018	3.3	0.16	251
Porto Alegre	10/2018	2.42	0.16	48
Total		4.83	2.60	502

Note – Each row represents a round of lottery draws. The average time in months since the lottery date and the corresponding standard deviation are computed at the 2018 interview date for each child. Interviewed sample refers to the total number of interviews completed among the children in each lottery draw.

For our earliest randomization, in Porto Alegre, nine months had passed at the time of the interview since the lottery took place. The lowest average was also at Porto Alegre (the only municipality with two rounds of randomizations), with just under two and a half months between lottery and interview dates. Stardard deviations of the time passed are mostly small because all interviews happened within an interval of three weeks in December 2018. We leverage this unplanned delay and use these interviews to measure outcomes instead of using them as a baseline.

The questionnaire comprised three sets of questions: (i) a socioeconomic form to obtain data on family composition, income, health, relations and habits; and (ii) a measure of child development; and (iii) a measure of violent child discipline methods. Specifically, we measure outcomes using the variables detailed below.

Child development was measured based on the "Ages and Stages Questionnaire" (ASQ-3 Brasil), originally developed by Bricker et al. (1999) and adapted to the Brazilian context by Filgueiras et al. (2013), which is administered on the primary caregiver of the child. ASQ-3 provides measures on five domains of development: communication skills, gross motor skills, fine motor skills, problem solving, and personal-social ability.

In addition to child development scores, we consider four variables related to child health. Regarding hospitalizations, we obtained information from the interviewed caregivers on whether

children were ever hospitalized and, if so, how many hospitalizations they went through. We also collected information on child weight and height/length, as recalled by the caregiver, but only among children reported to have been weighed or measured recently.

The discipline methods employed by family members on the child were assessed with the Child Discipline module from the "Multiple Indicator Cluster Surveys" (MICS6), developed by UNICEF. For this round of data collection, we administered a short version of this module, allowing us to assess whether the family employs methods of psychological aggression, physical punishment and severe physical punishment.

Regarding the home environment and family-child interactions, we obtained information on the weekly time spent by the caregiver with the child (in hours), whether the child has children books available at the household, and how frequently a family member engages in the following activities with the child (adapted from MICS6/UNICEF): played with child, told stories to child, took child outside, sang/taught songs to child.

Finally, we consider four variables related to family access to public services. These are binary variables on whether the interviewed caregiver was able to provide information on which social assistance center and health unit the family attends, whether the child is weighted monthly and whether the child attends childcare.

1.3.2.2 Second survey (2021)

The second round of door-to-door data collection was planned to happen in early 2020, just after one year of the first round of interviews. However, because of the covid-19 pandemic, we had to postpone these efforts. The second round of interviews was conducted via telephone calls and started in August 2021. This questionnaire included six sets of questions: (i) a socioeconomic form to obtain data on family composition, income, access to services and other socioeconomic characteristics; (ii) measures of child development; (iii) measures of child discipline methods; (iv) measures of parenting skills; and (v) measures of mental health of the caregiver. Specifically, we consider the variables detailed below to measure outcomes.

Similarly to the 2018 interviews, child development was measured based on the "Ages and Stages Questionnaire" (ASQ-3 Brasil)[8], and child discipline questions followed the "Multiple Indicators Cluster Survey" (MICS6), this time including questions on non-violent discipline methods as well as violent ones.

Parental practices were measured based on the 18-item parenting scale in "Parenting and Family Adjustment Scales" (PAFAS), originally by Sanders et al. (2014) and adapted to Brazil by

[8] We had to slightly adapt the questionnaire to telephone-based interviewing, e.g. describing simple figures and shapes instead of presenting them. Questions that demanded explanations considered too intricate for telephone communication were dropped, and we adjusted the scores accordingly.

Santana (2018). PAFAS provides scores on parental consistency, coercive parenting, positive encouragement and quality of parent-child relationship, as well as a global parenting score.

Mental health of the caregiver was measured using the 21-item version of the "Depression, Anxiety and Stress Scales" (DASS-21) by Lovibond and Lovibond (1996), adapted to Brazil by Vignola and Tucci (2014) , which provides separate scores of depression, anxiety and stress of the interviewee. Based on the scores, symptoms for each dimension are then categorized as one of the following: normal, mild, moderate, severe or extremely severe.

Similarly to the 2018 survey, we also obtained information on family access to services. The interviewed caregiver had to respond whether anyone in the household had to use a public service in the four weeks preceding the interview (e.g. go to a health unit or social assistance center) and whether the child was currently attending childcare.

Table 3 presents the average time in months elapsed between lottery and interview dates. This time, standard deviations are higher than in the first round of interviews, since we took several months to complete the second round, with the latest interviews happening in early February 2022. The implications of this variation in timing are explored in section 1.5.

Table 3 – Second round of data collection

Municipality	Lottery date	Average time in months since lottery	Standard deviation	Interviewed sample
Porto Alegre	03/2018	43.75	1.62	89
Serafina Corrêa	04/2018	43.14	1.96	9
Viamão	09/2018	38.24	1.84	23
Canoas	09/2018	37.59	1.79	162
Porto Alegre	10/2018	37.02	1.31	19
Palmeira das Missões	10/2018	38.27	2.12	9
São Borja	11/2018	35.12	1.85	36
Rio Grande	09/2019	25.11	1.81	15
Alvorada	10/2019	24.11	2.16	23
Pelotas	01/2020	21.37	2.23	31
Caxias do Sul	02/2020	21.78	2.22	12
São Paulo das Missões	03/2020	20.7	1.48	8
Uruguaiana	03/2020	19.03	1.96	21
Santa Maria	01/2021	10.27	2.64	7
Cachoeirinha	03/2021	8.33	3.11	6
Ibirubá	05/2021	4.7	2.13	8
Total		33.77	9.68	478

Note – Each row represents a round of lottery draws. The average time in months since the lottery date and the corresponding standard deviation are computed at the 2021 interview date for each child. Interviewed sample refers to the total number of interviews completed among the children in each lottery draw.

1.3.3 Administrative Records

We have access to two sources of program administrative records, which we use to measure compliance with treatment assignment. The first is an online monitoring system maintained by PIM, in which visitors from each municipality are tasked with keeping an updated record of all children they are responsible for visiting. This system is usually accessible only by program staff, but was made available to us for research purposes. In addition to basic identification, it includes records of a mandatory quarterly development assessment.

The development assessment is conducted and filled by the family visitor, and contains information on basic child development milestones (e.g. whether they are able to speak full sentences and use pronouns), basic health information (e.g. use of prescribed medication, diagnosed diseases, whether the child's weight and height are considered appropriate for their age), a short assessment of family involvement in promoting development (e.g. whether they show affection, stimulate child initiative and respect the child's rhythm of learning), as well as a non-standardized written review of the child's progress since the previous assessment.

Most of this information is not of use for this study, both because it only covers children who have actually received the intervention and because questions are often incomplete. However, because visitors are in part monitored based on the existence of quarterly assessment data for their families and thus have an incentive to register said assessments into the system, we focus on the existence of assessments to identify whether children in our evaluation sample are part of PIM. Of the 2,419 children who participated in our lottery, we identify 472 who have a profile in the monitoring system, 465 of which have at least one quarterly assessment filled.

Our second source of administrative data comes from written records of home visits. In every municipality, visitors are required to fill a printed form with the dates of all visits conducted each month. This brief form includes only the names of the child and their mother, and the date of the visit. We contacted every municipality participating in the study to gain access to such printed forms and then we had the information manually filled into a database. A total of 370 children in our sample have records of at least one home visit. This number is smaller than the 465 children who had a quarterly assessment filled in the system likely because the online system is more closely monitored by program management. Of the 370 with records of home visits, 29 did not have any completed quarterly assessments, while 124 of those with assessment dates did not have any records of home visits.

1.3.4 Attrition

In our first round of data collection in 2018, we set out to interview 1,659 individuals, but only successfully obtained information from 502, or 30.3%. In 2021, we interviewed 19.8% of our complete registry list of 2,419, equivalent to 478 children surveyed.

One concern with attrition rates is that they might not be random between treatment and control groups, which would possibly lead to selection bias, e.g. if being treated causes a higher probability of response. To assess the presence of differential attrition, we estimate the following:

$$A_{is} = \alpha + \tau W_{is} + \mu_s + \epsilon_{is} \tag{1.1}$$

where A_{is} indicates whether individual i in strata s has attrited (i.e., has not responded to the survey); W_i is an indicator variable that assumes a value equal to one if the individual was selected via lottery to receive an offer to participate (and zero otherwise); μ_s is the fixed-effect of strata s; and ϵ_{is} is a random error term. We estimate this regression separately for the 2018 and 2021 surveys, and the results are presented in Table 4.

Table 4 – Differential attrition analysis

	2018		2021	
	Full sample	Restricted sample	Full sample	Restricted sample
	(1)	(2)	(3)	(4)
Lottery dummy	0.0694	0.0408	-0.0473	-0.0215
	(0.022)	(0.025)	(0.016)	(0.018)
	[0.002]	[0.109]	[0.004]	[0.242]
Control mean	0.6689	0.6784	0.8241	0.814
Clustered SE p-value	0.06	0.12	0.073	0.067
Fisher p-value	0.003	0.1205	0.004	0.267
N. obs.	1659	1256	2419	1902
N. strata	37	31	61	50

Note – Regressions of an attrition indicator on the lottery dummy. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Unfortunately, the estimates indicate the presence of differential attrition in both rounds of data collection. In 2018 (shown in column 1 of Table 4), it seems that being randomly allocated to the treatment group leads to higher attrition at a proportion of 10.4% relative to the control mean, which is unusual in survey settings similar to ours. This can be explained by the 311 individuals who were left out of interviewing efforts, as detailed beforehand, since 292 of them were initially drawn to the treatment group, thus relatively reducing our chances of interviewing treated individuals. Dropping these 311 from the regressions yields the opposite and less unexpected result: being selected into treatment leads to lower attrition rates (-13.3% relative to the control mean, statistically significant at 1%, results not shown). For the 2021 survey (column 3 of Table

4), estimates also indicate that individuals randomly selected into the treatment group are less likely to be attriters at a smaller proportion of 5.7% relative to the control mean.

We deal with the presence of differential attrition with two main strategies. First, we restrict our analyses to subsamples containing only strata with less evidence of differential attrition. To do this, we estimate equation 1.1 separately for each strata. Then, we drop every strata presenting evidence of differential attrition (i.e., for which the p-value associated with the lottery dummy is small enough[9]). We refer to these as the 2018 and 2021 restricted samples, respectively. Reestimating equation 1.1 for the restricted samples yields results no longer statistically significant at the 10% level, although point estimates are still not negligible, as presented in columns 2 and 4 of Table 4.

Finally, we also estimate bounds for the treatment effect in 2021 using measured effort to interview to truncate our sample, as proposed by Behaghel et al. (2015) and explained in further detail later. This procedure allows us to correct for potential non-response bias by forcing treatment and control groups to be comparable in terms of interview response rates and estimate bounds similar to those proposed by Lee (2009)[10].

1.3.5 Balancedness

As explained above, we proceed with four different samples in our analyses. The full samples of 2018 and 2021 refer to all the data we were able to collect in that respective round of surveys. The restricted samples refer to the subset of strata for which there is less evidence of differential attrition.

We assess balancedness of observable characteristics in each sample by estimating the following:

$$X_{is} = \alpha + \tau W_{is} + \mu_s + \epsilon_{is} \tag{1.2}$$

where X_{is} is an observed variable of individual i in strata s; W_i is the treatment initial offer indicator; μ_s is the fixed-effect of strata s; and ϵ_{is} is a random error term.

The nature of a randomized treatment assignment should mean that treatment and control groups are comparable, i.e. that observable (and also non-observable) characteristics are on average statistically similar between both groups. However, since we are limited by survey response rates (with high and differential attrition rates, as previously discussed), this could not always be the case.

[9] To achieve satisfactory differential attrition (i.e., with p-values higher than 0.1) in the restricted samples, we had to set the p-value threshold to 10% for the 2021 sample but to 15% for the 2018 sample.

[10] We also estimate Lee bounds, but these are highly uninformative because of our high attrition rates and thus are not reported.

Table 5 shows that most variables are balanced between treatment and control groups when considering the 2018 samples. Few exceptions with p-values smaller than 10% include the mean number of caregivers with incomplete primary education, proportion of female children, and number of prenatal appointments, as well as family income for the full sample only. However, testing for joint significance with an F test brings p-values equal to 0.43 and 0.26, suggesting an overall balancedness of these variables.

Table 5 – Balancedness tests, 2018 samples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Female caregiver	502	0.95	0.95	0.86	385	0.95	0.95	0.74
Caregiver with incomplete primary education	502	0.24	0.35	0.03	385	0.24	0.35	0.05
Caregiver with complete primary until incomplete secondary education	502	0.41	0.35	0.45	385	0.41	0.35	0.31
Caregiver with at least complete secondary education	502	0.35	0.30	0.20	385	0.35	0.30	0.43
Female child	502	0.51	0.43	0.05	385	0.52	0.42	0.04
Child age in months	502	22.61	23.41	0.94	385	21.58	20.54	0.48
Child lives with both parents	502	0.64	0.59	0.69	385	0.62	0.59	0.83
Child lives with mother only	502	0.29	0.31	0.91	385	0.30	0.29	0.40
Pregnancy was planned	497	0.30	0.33	0.56	380	0.32	0.32	0.74
Mother had at most 4 prenatal appointments	502	0.09	0.08	0.63	385	0.09	0.07	0.28
Mother had 5-7 prenatal appointments	502	0.20	0.29	0.03	385	0.19	0.29	0.04
Mother had 8 or more prenatal appointments	502	0.61	0.50	0.02	385	0.62	0.51	0.05
Number of prenatal appointments missing	502	0.10	0.14	0.28	385	0.10	0.14	0.30
Birth weight <2500	502	0.07	0.09	0.25	385	0.07	0.10	0.22
Birth weight 2500-2999	502	0.20	0.21	0.69	385	0.21	0.18	0.71
Birth weight 3000-3999	502	0.63	0.58	0.28	385	0.62	0.59	0.49
Birth weight >4000	502	0.07	0.09	0.57	385	0.07	0.10	0.29
Birth weight missing	502	0.03	0.02	0.55	385	0.04	0.02	0.42
Number of people living in the house	502	4.54	4.50	0.80	385	4.53	4.46	0.74
Number of rooms in the house	499	4.75	4.56	0.53	382	4.77	4.57	0.71
House has electricity	502	1.00	0.99	0.93	385	1.00	0.99	0.17
House has piped water	502	0.99	0.97	0.33	385	0.99	0.97	0.31
House is connected to sewage network	502	0.75	0.81	0.13	385	0.76	0.82	0.20
Family consumes untreated water	502	0.50	0.51	0.39	385	0.50	0.48	0.38
Family owns a computer	502	0.27	0.22	0.49	385	0.28	0.23	0.59
Family owns a tablet	502	0.18	0.15	0.59	385	0.19	0.15	0.60
Family has access to mobile internet	499	0.70	0.68	0.78	382	0.69	0.67	0.62
Family has access to dial-up internet	501	0.10	0.10	0.99	384	0.11	0.09	0.75
Family has access to broadband internet	500	0.43	0.42	0.76	383	0.44	0.40	0.91
Family monthly income up to 1 MW	502	0.42	0.46	0.69	385	0.40	0.46	0.42
Family monthly income 1-2 MW	502	0.31	0.27	0.37	385	0.32	0.26	0.22
Family monthly income 2-3 MW	502	0.12	0.08	0.32	385	0.12	0.07	0.32
Family monthly income 3 or more MW	502	0.05	0.07	0.29	385	0.07	0.07	0.66
Family monthly income missing	502	0.10	0.11	0.31	385	0.10	0.14	0.22
Someone in the household benefits from cash transfer programs	496	0.60	0.69	0.23	379	0.60	0.67	0.86
Joint F test				0.43				0.26

Note – Balancedness tests for the 2018 samples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

Similarly, most characteristics measured in 2021 are also balanced between treatment and control groups, as shown in Table 6. In this case, however, we reject the null hypothesis of the joint F test at the 2% level when using the restricted sample for 2021. Since this unbalancedness in observables seems to be driven by differences in income[11], we proceed by including family

[11] The p-value of the joint F test is equal to 0.1049 if we exclude monthly income dummies from the regression of the lottery dummy on the full set of variables.

income dummies in all our regressions for this particular sample.

With these results in mind, we are presented with an apparent trade-off between attrition and balancedness for the 2021 sample. Our full sample is one with evidence of differential attrition, which might bias our treatment effect estimates, but with balanced observed characteristics between treatment and control groups. On the other hand, restricting which strata enter the analysis yields a subsample with similar attrition rates in both groups, but with otherwise unbalanced observables (thus indicating limited comparability, which we attempt to solve by controlling for income). This apparent trade-off suggests that the profile of attriters differs according to their assigned treatment, on average, and influences the interpretation attributable to our results.

Table 6 – Balancedness tests, 2021 samples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Caregiver lives with the child	478	0.97	0.98	0.88	373	0.97	0.97	0.94
Female caregiver	478	0.97	0.98	0.43	373	0.96	0.98	0.21
Caregiver with incomplete primary education	476	0.21	0.21	0.71	371	0.20	0.23	0.28
Caregiver with complete primary until incomplete secondary education	476	0.27	0.29	0.84	371	0.26	0.28	0.82
Caregiver with at least complete secondary education	476	0.52	0.49	0.62	371	0.54	0.49	0.28
Female child	478	0.50	0.45	0.14	373	0.51	0.46	0.39
Child age in months	478	49.12	52.67	0.28	373	52.91	53.19	0.11
Child lives with both parents	477	0.63	0.58	0.97	372	0.61	0.60	0.68
Child lives with mother only	477	0.35	0.38	0.60	372	0.38	0.35	0.27
Number of people living in the house	478	4.18	4.22	0.71	373	4.28	4.14	0.28
Number of rooms in the house	478	5.18	5.22	0.49	373	5.25	5.22	0.78
House has electricity	478	1.00	1.00	0.33	373	1.00	0.99	0.33
House has piped water	478	0.97	0.94	0.42	373	0.98	0.95	0.12
House is connected to sewage network	478	0.66	0.70	0.45	373	0.66	0.70	0.61
Family owns a computer	478	0.35	0.36	0.91	373	0.38	0.37	0.36
Family owns a tablet	478	0.19	0.17	0.25	373	0.23	0.15	0.07
Family has access to mobile internet	477	0.88	0.87	0.73	372	0.88	0.85	0.34
Family has access to dial-up internet	471	0.23	0.27	0.43	369	0.22	0.27	0.41
Family has access to broadband internet	475	0.74	0.73	0.70	371	0.75	0.72	0.28
Someone in the household was infected with covid	476	0.31	0.30	0.58	372	0.26	0.30	0.31
Someone in the house lost their job during the pandemic	478	0.40	0.44	0.77	373	0.42	0.42	0.96
Caretaker had to stop taking care of child because of the pandemic	477	0.08	0.11	0.67	373	0.10	0.11	0.86
Family monthly income up to 1 MW	478	0.32	0.41	0.15	373	0.28	0.39	0.03
Family monthly income 1-2 MW	478	0.35	0.33	0.52	373	0.35	0.33	0.55
Family monthly income 2-3 MW	478	0.13	0.13	0.79	373	0.15	0.14	0.75
Family monthly income 3 or more MW	478	0.16	0.11	0.30	373	0.17	0.11	0.20
Family monthly income missing	478	0.03	0.02	0.38	373	0.04	0.02	0.29
Someone in the household benefits from cash transfer programs	477	0.48	0.54	0.88	372	0.46	0.51	0.66
Someone receives Auxílio Emergencial	478	0.59	0.57	0.77	373	0.57	0.58	0.74
Joint F test				0.92				0.02

Note – Balancedness tests for the 2021 samples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

1.4 Empirical Strategy

To estimate the treatment effects of PIM on our outcomes of interest, we focus on the Intention-to-Treat parameter, recovered from the following regression:

$$y_{is} = \alpha + \tau_{\text{ITT}} W_{is} + \mu_s + \epsilon_{is} \tag{1.3}$$

where y_{is} is the result of interest of individual i in strata s; W_i is an indicator variable that assumes a value equal to one if the individual was selected via lottery to receive an offer to participate (and zero otherwise); μ_s is the fixed-effect of strata s; and ϵ_{is} is a random error term. The intention-to-treat parameter captures the effect of being offered the chance to participate in the program, represented by $\tau_{\texttt{ITT}}$ in equation 1.3.

We set W_{is} as the "initial offer" indicator, which is equal to 1 only for individuals randomized into the treatment group from the start. The initial offer estimator is a consistent estimator for the ITT parameter (CHAISEMARTIN; BEHAGHEL, 2020). We focus on the initial offer, as opposed to the ever-offer estimator proposed by Chaisemartin and Behaghel (2020), for two main reasons: (i) only 4.5% of our initial-offer control group was reported by local program management to have eventually received an offer, or 56 children; and (ii) even among those who were never reported to have been offered the treatment, there was non-compliance, with 4.4% still having benefited from the program.

In all cases, our main strategy for testing the statistical significance of our results consists of estimating heteroskedasticity-robust standard errors, following the individual assignment in our lottery design as described by Abadie et al. (2017). In addition, we complement this strategy with two approaches: (i) clustering standard-errors at the municipality level, to account for the possibility of stochastic shocks affecting potential outcomes within municipalities and thus draw inference on an effect "netted out" of such shocks, as discussed and proposed by Deeb and Chaisemartin (2019); and (ii) performing a randomization inference procedure to obtain Fisher's p-value using the ITT estimate as a test statistic, which can be interpreted as the likelihood of observing that results under the sharp null hypothesis of no effect for all observations in the data, as detailed e.g. by Imbens and Rubin (2015).

We also estimate the effect on compliance with treatment assignment using a similar specification to above, but with an indicator of participating in the program as the dependent variable. This is equivalent to the first stage in the estimation of the Local Average Treatment Effect, via the analogous two-stage least squares approach, in which the lottery assignment acts as an exogenous instrument for program participation (as seen e.g. in Imbens and Angrist (1994)):

$$\begin{aligned} T_{is} &= \gamma + \tau_{\texttt{FS}} W_{is} + \mu_s + \xi_{is} \\ y_{is} &= \alpha + \tau_{\texttt{LATE}} \widehat{T}_{is} + \mu_s + \epsilon_{is} \end{aligned} \tag{1.4}$$

where T_{is} is a dummy variable equal to 1 if individual i in strata s ever benefited from PIM. With administrative records, we measure participation in the program as one of the following: (i) the child was reported to have received at least one visit from PIM since the lottery, or (ii) at least one quarterly assessment of the child was filled in the program monitoring system since the lottery. The τ_{FS} parameter then captures to which extent our randomizations were able to affect program

take-up.

The LATE parameter $\tau_{\texttt{LATE}}$ captures the average treatment effect for the subpopulation of compliers, i.e., those in our sample who were induced to participate in PIM because of the lottery assignment to the treatment group. For the estimation of the LATE, we focus on a first stage using the existence of quarterly assessment records as our primary measure of program participation. Since LATE results are largely similar to ITT estimates, but scaled up by the proportion of compliers, we leave these to the appendix.

Finally, we investigate the possibility of heterogeneity in our results on child development scores by stratifying our samples based on two relevant observable characteristics, despite escaping from our randomized design. First, we consider the gender of the child, splitting the sample into male and female children. Then, we also consider the educational level of the child caregiver, splitting the sample in two groups, one with caregivers who at most entered secondary education but did not complete it ("low" education) and the other with caregivers with at least complete secondary education ("high" education). In these subgroup analyses, we first assess balancedness of observable characteristics and control for unbalanced variables whenever applicable.

1.5 Results

1.5.1 Program participation

While being randomly selected into the treatment group was supposed to guarantee a slot for the child as a beneficiary of PIM, accepting the offer of being treated was not mandatory for families. Also, despite our best efforts to communicate with local program implementation teams, it is not certain whether the protocol of offering a slot to all lottery winners, and only progressing down the waiting lists upon rejections from initial offers, was strictly followed.

We estimate to which extent the lottery was successful in increasing take-up among the treatment group using our two main measures of program participation, as shown in Table 7. Panel A uses the existence of any record of home visits as an indication of program participation, while Panel B focuses on the existence of quarterly evaluation assessments of each child.

In all cases, there is strong evidence of increased participation among the treatment group, compared to the control group, with statistically significant estimates at the 1% level. Using any of our measures, there seems to be some degree of contamination, with 4% to 7.7% of the control group benefiting from the program at some point. We estimate an increase of up to 41.9 percentage points in take-up caused by the randomized assignment, for the restricted 2018 sample (column 2).

Because these measures of program participation come from administrative data, we are

Table 7 – Intention-to-Treat estimates on program participation, interviewed sample

	2018		2021	
	Full sample	Restricted sample	Full sample	Restricted sample
	(1)	(2)	(3)	(4)
Panel A: Has any record of home visit				
Lottery dummy	0.3405	0.3327	0.3196	0.2857
	(0.038)	(0.042)	(0.038)	(0.040)
	[0.000]	[0.000]	[0.000]	[0.000]
Control mean	0.0405	0.0459	0.0561	0.0608
Clustered SE p-value	0.013	0.012	0.000	0.001
Fisher p-value	0.000	0.000	0.000	0.000
Panel B: Has any record of quarterly assessment				
Lottery dummy	0.3983	0.4185	0.3489	0.3374
	(0.041)	(0.045)	(0.039)	(0.041)
	[0.000]	[0.000]	[0.000]	[0.000]
Control mean	0.0709	0.0734	0.0748	0.0773
Clustered SE p-value	0.007	0.003	0.000	0.001
Fisher p-value	0.000	0.000	0.000	0.000
N. obs.	500	383	471	366
N. strata	35	29	48	37

Note – Regressions of treatment compliance indicators on the lottery dummy. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

also able to estimate results for the complete registry of 1,659 children in 2018 and 2,419 children in 2021. These results are presented in Table A.1 in Appendix A and yield similar conclusions of increased take-up among lottery winners, even though point estimates are slightly smaller, as is program participation among children in the control group.

1.5.2 Results on main outcomes

With evidence of increased take-up among lottery winners, we now turn our attention to a wide range of outcomes related to the main objectives of the Better Early Childhood Program, using the variables described in section 1.3.2. Throughout this subsection, we present Intention-to-Treat estimates following the specification in equation 1.3, and also refer to the Local Average Treatment Effects estimates presented in Appendix B, following the two stage approach outlined in equation 1.4.

1.5.2.1 Child development and health

We start by examining program impacts on five domains of child development, including communication skills, gross and fine motor skills, problem solving and personal-social ability, as well as a global development score. We present results using standardized scores for 2018 and 2021 in Table 8.[12]

We estimate that winning the lottery has a statistically significant impact of 0.18σ (units of control group standard deviation) on overall development when considering our full 2018 sample (column 1, Panel A), although p-values increase slightly out of the 10% significance range for the restricted sample (column 1, Panel B). On the other hand, results for both the full and restricted 2021 samples are not statistically significant and point estimates are much smaller, although still positive.

Further inspection of individual domains of child development, in columns 2-6 of Table 8, indicate that the impacts of PIM on development seem to be concentrated as impacts on motor skills. Most point estimates are positive in all considered domains, but those related to fine motor skills (column 4) show statistical significance for both 2018 samples (Panels A and B), with an estimated effect of 0.17σ in the restricted sample. Estimates on gross motor skills are also statistically significant for the full 2018 sample (column 3, Panel A), but lose their significance when restricting the strata (Panel B).

Instrumental variable estimates of the LATE parameter in Table B.1 tell a similar story. The estimated LATE of PIM on the overall development score is 0.34σ when considering the 2018 restricted sample (column 1, Panel B), which again seems to be concentrated as an effect on fine motors skills (0.42σ), while none of the estimates for 2021 are statistically different from zero.

Interestingly, despite the small number of municipalities in this study, clustering standard errors does not change our conclusions of impacts on child development in 2018 (for both the ITT and the LATE analyses). As discussed by Deeb and Chaisemartin (2019), this could be interpreted as an indication of external validity of these results to stochastic shocks at the municipality level, for this specific population of interviewed families, and is likely the result of higher variability of the outcomes within rather than across municipalities.

The fact that we find significant results on child development in 2018 but not in 2021 is intriguing. One possible explanation would be a fade-out of any benefits created by PIM in the short-term. An alternative interpretation of these results, which we believe to be more likely, is that implementation issues caused by the covid-19 pandemic, as outlined in section 1.2.1, diluted

[12] We have also tested fitting graded response models to each ASQ domain and using the predicted latent variables instead of the original standardized ASQ scores. Results are roughly identical and thus we proceed with the more traditional approach.

Table 8 – Intention-to-Treat estimates on child development

	Overall development score (1)	Communication skills (2)	Gross motor skills (3)	Fine motor skills (4)	Problem solving (5)	Personal and social ability (6)
Panel A: Full sample 2018						
Lottery dummy	0.1833	0.0688	0.1567	0.2176	0.1002	0.0697
	(0.088)	(0.099)	(0.080)	(0.090)	(0.090)	(0.093)
	[0.039]	[0.488]	[0.050]	[0.016]	[0.266]	[0.454]
Clustered SE p-value	0.013	0.384	0.107	0.03	0.363	0.433
Fisher p-value	0.0345	0.472	0.0505	0.0135	0.2595	0.452
N. obs.	491	493	492	493	492	493
Panel B: Restricted sample 2018						
Lottery dummy	0.1446	0.0285	0.1276	0.1761	0.1043	0.0651
	(0.098)	(0.110)	(0.088)	(0.100)	(0.099)	(0.106)
	[0.141]	[0.795]	[0.146]	[0.080]	[0.295]	[0.540]
Clustered SE p-value	0.104	0.615	0.356	0.193	0.264	0.436
Fisher p-value	0.1535	0.795	0.1685	0.0885	0.309	0.5335
N. obs.	377	379	378	379	378	379
Panel C: Full sample 2021						
Lottery dummy	0.0608	0.0182	-0.0091	0.0906	0.0402	0.0396
	(0.097)	(0.100)	(0.096)	(0.093)	(0.096)	(0.094)
	[0.529]	[0.855]	[0.924]	[0.332]	[0.677]	[0.673]
Clustered SE p-value	0.492	0.834	0.912	0.372	0.458	0.69
Fisher p-value	0.498	0.8395	0.936	0.323	0.653	0.6375
N. obs.	465	470	469	467	470	471
Panel D: Restricted sample 2021						
Lottery dummy	0.0705	0.0458	0.0412	0.0839	0.0479	0.0221
	(0.104)	(0.104)	(0.106)	(0.100)	(0.107)	(0.102)
	[0.496]	[0.660]	[0.697]	[0.402]	[0.654]	[0.828]
Clustered SE p-value	0.455	0.585	0.687	0.475	0.435	0.816
Fisher p-value	0.4835	0.6465	0.701	0.398	0.67	0.828
N. obs.	360	365	364	362	365	366

Note – Regressions of child development ASQ-3 scores (reported in units of control-group standard deviations) on the lottery dummy. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

the potential benefits of PIM, particularly ones related to motor skills, which could be more severely affected by the lack of an at-home intervention. Then, a lack of statistical power could mean we are not able to precisely estimate smaller gains in child development.

We further investigate the effects of PIM on child development domains by assessing the possibility of heterogeneity in our results. We reestimate the ITT effect on child development for four subgroups. Although the experimental design of this study was not conceived to address such subgroup analyses, doing so could be beneficial to gain insight on possible mechanisms through which PIM might act. We split our samples based on two characteristics: child gender (female or male) and caregiver educational level (low or high, where a "low" education means not having completed high school). Because some of these subsamples are unbalanced in terms of observable characteristics (balancedness tests available in Appendices E and F), we control for variables that seem to be driving any unbalances.

First, splitting our sample by gender allows us to more clearly visualize gains in child development that differ significantly between girls and boys. For female children, in Table 9, we estimate increased fine motor skills in 2018 of around 0.23-0.29σ, significant at the 5% level for the restricted sample (column 4), which is consistent with our previous results. Interestingly, in this case we also observe a similar effect of PIM on fine motor skills in 2021 when using the restricted sample, with a point estimate of 0.27σ. For male children, on the other hand, we estimate statistically significant gains in gross motor skills (column 3, Table 10) or around 0.32-0.35σ in 2018 only, but with negative point estimates 2021 that are imprecisely estimated. Overall these results could suggest that girls thrive under PIM exposition in terms of their fine motor skills development, even if faced with possible implementation issues such as those led by the pandemic after 2020. As for boys, they seem to benefit the most in terms of gross motor skills development, but only under regular program delivery conditions.

Estimating heterogeneous effects in terms of caregiver education also yield interesting results. Children with low education caregivers seem to benefit the most from the treatment, with significant increases in both gross and fine motor skills in 2018 (columns 3 and 4, Table 11) of 0.19-0.22σ and 0.24-0.29σ respectively, and an equivalent gain in the global development score (column 1). Estimates for children with high education caregivers (Table 12) are considerably more unstable, going from positive to negative after restricting the sample in some cases, and all imprecisely estimated. While pinpointing that PIM does not generate benefits for this group is tricky, both because of possible concerns with power and because the experiment was not designed around this specific question, benefiting only children from the most vulnerable contexts would be aligned with PIM goals and could be an unforeseen result of intentional program design.

Finally, we estimate Intention-to-Treat effects on four variables related to child health, as can be seen in Table 13. Columns 1 and 2 use measures of hospitalization. Column 1 refers to an indicator of the child ever having been hospitalized, as reported by the caregiver, while column 2

Table 9 – Intention-to-Treat estimates on child development, female children

	Overall development score	Communication skills	Gross motor skills	Fine motor skills	Problem solving	Personal and social ability
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	0.1051	0.1356	0.0149	0.2355	-0.0378	-0.0428
	(0.133)	(0.150)	(0.126)	(0.132)	(0.131)	(0.148)
	[0.432]	[0.366]	[0.906]	[0.075]	[0.773]	[0.773]
Clustered SE p-value	0.17	0.317	0.895	0.002	0.76	0.212
Fisher p-value	0.448	0.341	0.917	0.0955	0.7885	0.773
N. obs.	220	221	221	221	220	221
Panel B: Restricted sample 2018						
Lottery dummy	0.1283	0.1294	0.0553	0.2887	-0.2066	0.1042
	(0.146)	(0.163)	(0.150)	(0.144)	(0.166)	(0.182)
	[0.382]	[0.428]	[0.713]	[0.046]	[0.215]	[0.569]
Clustered SE p-value	0.46	0.352	0.809	0.06	0.165	0.272
Fisher p-value	0.426	0.431	0.736	0.0545	0.2155	0.559
N. obs.	165	166	166	166	165	166
Panel C: Full sample 2021						
Lottery dummy	0.0714	-0.0112	0.0866	0.1683	0.0918	-0.0948
	(0.136)	(0.146)	(0.142)	(0.134)	(0.153)	(0.132)
	[0.600]	[0.939]	[0.541]	[0.212]	[0.550]	[0.475]
Clustered SE p-value	0.623	0.902	0.661	0.269	0.347	0.239
Fisher p-value	0.6105	0.9385	0.546	0.234	0.556	0.5135
N. obs.	213	215	214	214	214	215
Panel D: Restricted sample 2021						
Lottery dummy	0.1935	0.075	0.1357	0.2745	0.2496	-0.0262
	(0.153)	(0.156)	(0.155)	(0.156)	(0.172)	(0.145)
	[0.207]	[0.631]	[0.384]	[0.081]	[0.149]	[0.857]
Clustered SE p-value	0.31	0.372	0.552	0.234	0.066	0.76
Fisher p-value	0.2135	0.6315	0.394	0.0685	0.1025	0.8575
N. obs.	168	170	169	169	169	170

Note – Regressions of child development ASQ-3 scores (reported in units of control-group standard deviations) on the lottery dummy, for the subsample of female children. Regressions for the 2018 restricted sample include caregiver education, prenatal care and family income dummies. Regressions for the 2021 full sample include family income dummies. Regressions for the 2021 restricted sample include dummies for whether the caregiver lives with the child, whether a member of the household was infected with covid-19 and family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

Table 10 – Intention-to-Treat estimates on child development, male children

	Overall development score	Communication skills	Gross motor skills	Fine motor skills	Problem solving	Personal and social ability
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	0.2681	0.0361	0.3501	0.22	0.1609	0.1712
	(0.124)	(0.148)	(0.096)	(0.132)	(0.141)	(0.129)
	[0.032]	[0.808]	[0.000]	[0.098]	[0.255]	[0.187]
Clustered SE p-value	0.002	0.65	0.011	0.089	0.347	0.196
Fisher p-value	0.04	0.8115	0.002	0.1075	0.253	0.2195
N. obs.	257	258	257	258	258	258
Panel B: Restricted sample 2018						
Lottery dummy	0.1682	-0.0801	0.3185	0.0841	0.1886	0.0944
	(0.137)	(0.173)	(0.109)	(0.152)	(0.152)	(0.148)
	[0.222]	[0.643]	[0.004]	[0.581]	[0.217]	[0.526]
Clustered SE p-value	0.039	0.162	0.036	0.619	0.113	0.336
Fisher p-value	0.237	0.6235	0.0145	0.594	0.213	0.5465
N. obs.	199	200	199	200	200	200
Panel C: Full sample 2021						
Lottery dummy	0.0124	-0.0094	-0.1374	0.0633	0.0056	0.0512
	(0.138)	(0.134)	(0.141)	(0.145)	(0.135)	(0.132)
	[0.928]	[0.944]	[0.331]	[0.662]	[0.967]	[0.698]
Clustered SE p-value	0.933	0.954	0.311	0.563	0.958	0.724
Fisher p-value	0.9165	0.9485	0.3315	0.6685	0.97	0.6925
N. obs.	233	236	236	234	237	237
Panel D: Restricted sample 2021						
Lottery dummy	0.0304	0.0405	-0.1198	0.0657	-0.0128	0.0592
	(0.148)	(0.133)	(0.150)	(0.157)	(0.158)	(0.140)
	[0.837]	[0.760]	[0.427]	[0.677]	[0.935]	[0.674]
Clustered SE p-value	0.813	0.759	0.402	0.575	0.914	0.65
Fisher p-value	0.837	0.771	0.42	0.652	0.9275	0.662
N. obs.	176	179	179	177	180	180

Note – Regressions of child development ASQ-3 scores (reported in units of control-group standard deviations) on the lottery dummy, for the subsample of male children. Regressions for the 2018 restricted sample include dummies for whether the child lives with both parents or the mother and family income dummies. Regressions for the 2021 restricted sample include a dummy for whether the caregiver is female. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

Table 11 – Intention-to-Treat estimates on child development, low education caregivers

	Overall development score	Communication skills	Gross motor skills	Fine motor skills	Problem solving	Personal and social ability
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	0.2118	0.0296	0.1885	0.2424	0.1246	0.1236
	(0.107)	(0.117)	(0.092)	(0.116)	(0.115)	(0.113)
	[0.048]	[0.801]	[0.041]	[0.037]	[0.277]	[0.276]
Clustered SE p-value	0.018	0.498	0.013	0.058	0.255	0.275
Fisher p-value	0.0545	0.811	0.076	0.038	0.2825	0.2755
N. obs.	332	334	333	334	333	334
Panel B: Restricted sample 2018						
Lottery dummy	0.2405	-0.0364	0.2231	0.2921	0.1956	0.1396
	(0.114)	(0.130)	(0.104)	(0.121)	(0.127)	(0.120)
	[0.036]	[0.780]	[0.033]	[0.016]	[0.126]	[0.247]
Clustered SE p-value	0.007	0.433	0.015	0.055	0.118	0.169
Fisher p-value	0.0405	0.786	0.038	0.0235	0.107	0.258
N. obs.	254	256	255	256	255	256
Panel C: Full sample 2021						
Lottery dummy	0.0401	-0.0727	-0.0138	0.116	0.0578	0.026
	(0.171)	(0.163)	(0.160)	(0.156)	(0.166)	(0.151)
	[0.815]	[0.656]	[0.931]	[0.459]	[0.728]	[0.864]
Clustered SE p-value	0.8	0.69	0.953	0.223	0.49	0.882
Fisher p-value	0.7995	0.6305	0.917	0.435	0.7105	0.8485
N. obs.	220	223	221	222	222	223
Panel D: Restricted sample 2021						
Lottery dummy	-0.0051	-0.1014	0.0125	0.0605	0.0679	-0.0359
	(0.179)	(0.167)	(0.177)	(0.173)	(0.192)	(0.161)
	[0.977]	[0.544]	[0.944]	[0.727]	[0.724]	[0.824]
Clustered SE p-value	0.972	0.569	0.961	0.606	0.576	0.817
Fisher p-value	0.9795	0.5465	0.9405	0.7025	0.6875	0.821
N. obs.	165	168	166	167	167	168

Note – Regressions of child development ASQ-3 scores (reported in units of control-group standard deviations) on the lottery dummy, for the subsample of children with low education caregivers (i.e. at most incomplete secondary education). Regressions for the 2018 restricted sample include prenatal care dummies. Regressions for the 2021 full sample include a dummy for child gender and a variable for the number of rooms in the household. Regressions for the 2021 restricted sample include dummies for child gender, whether the child lives with both parents or the mother, whether someone in the household was infected with covid-19, whether the caregiver had to stop taking care of child because of the covid-19, whether the family has access to mobile internet connection, and variables for child age and the number of rooms in the household. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

Table 12 – Intention-to-Treat estimates on child development, high education caregivers

	Overall development score	Communication skills	Gross motor skills	Fine motor skills	Problem solving	Personal and social ability
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	0.0735	0.2006	0.0944	0.1285	-0.0485	-0.1524
	(0.175)	(0.173)	(0.180)	(0.155)	(0.182)	(0.185)
	[0.675]	[0.248]	[0.600]	[0.409]	[0.790]	[0.412]
Clustered SE p-value	0.565	0.438	0.268	0.326	0.758	0.354
Fisher p-value	0.693	0.233	0.619	0.439	0.774	0.4255
N. obs.	151	151	151	151	151	151
Panel B: Restricted sample 2018						
Lottery dummy	-0.0755	0.1923	-0.0047	-0.0357	-0.2499	-0.1525
	(0.194)	(0.187)	(0.198)	(0.182)	(0.209)	(0.219)
	[0.699]	[0.308]	[0.981]	[0.845]	[0.235]	[0.488]
Clustered SE p-value	0.638	0.349	0.978	0.836	0.14	0.374
Fisher p-value	0.7165	0.333	0.9825	0.853	0.243	0.5125
N. obs.	115	115	115	115	115	115
Panel C: Full sample 2021						
Lottery dummy	0.0979	0.1067	0.042	0.0677	0.1419	0.0278
	(0.133)	(0.150)	(0.135)	(0.139)	(0.133)	(0.133)
	[0.463]	[0.478]	[0.756]	[0.626]	[0.286]	[0.834]
Clustered SE p-value	0.383	0.313	0.834	0.611	0.148	0.811
Fisher p-value	0.4225	0.4245	0.762	0.5945	0.251	0.849
N. obs.	232	233	234	232	234	234
Panel D: Restricted sample 2021						
Lottery dummy	0.0993	0.1704	0.0078	0.0894	0.1295	-0.0097
	(0.137)	(0.152)	(0.140)	(0.153)	(0.151)	(0.136)
	[0.469]	[0.265]	[0.956]	[0.560]	[0.391]	[0.944]
Clustered SE p-value	0.491	0.103	0.974	0.603	0.168	0.934
Fisher p-value	0.437	0.2075	0.9625	0.52	0.3265	0.9415
N. obs.	181	182	183	181	183	183

Note – Regressions of child development ASQ-3 scores (reported in units of control-group standard deviations) on the lottery dummy, for the subsample of children with high education caregivers (i.e. at least complete secondary education). Regressions for the 2018 restricted sample include a dummy for child gender. Regressions for the 2021 restricted sample include a dummy for whether the family owns a tablet. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

Table 13 – Intention-to-Treat estimates on child health

	Child ever hospitalized (1)	Number of times hospitalized (2)	Weight-for-age (3)	Height-for-age (4)
Panel A: Full sample 2018				
Lottery dummy	-0.0134	-0.0229	0.0723	-0.8865
	(0.044)	(0.086)	(0.376)	(0.645)
	[0.761]	[0.791]	[0.848]	[0.172]
Control mean	0.3277	0.4966	-0.0304	-0.9767
Clustered SE p-value	0.746	0.812	0.728	0.077
Fisher p-value	0.728	0.795	0.832	0.1665
N. obs.	500	500	225	144
Panel B: Restricted sample 2018				
Lottery dummy	-0.034	-0.0244	0.0022	-0.8258
	(0.049)	(0.100)	(0.377)	(0.590)
	[0.487]	[0.809]	[0.995]	[0.165]
Control mean	0.3257	0.4908	0.0624	-0.59
Clustered SE p-value	0.481	0.857	0.994	0.085
Fisher p-value	0.4795	0.8215	0.9945	0.24
N. obs.	383	383	180	117

Note – Regressions of child health variables on the lottery dummy. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

considers the number of reported hospitalizations. Columns 3 and 4 use reported weight and height to calculate weight-for-age and height-for-age Z-scores, following growth standards by the World Health Organization (WHO, 2006). These measures of anthropometric growth were only answered by caregivers who had previously reported their child was being regularly weighed and measured. Still, since we find no effect of winning the lottery on the probability of being regularly measured (as reported later in this section), we proceed with these analyses under the caveat of smaller sample sizes.

None of our Intention-to-Treat estimates on health outcomes appear to be statistically significant. Clustering standard errors at the municipality level does yield p-values smaller than 10% for the estimated negative effect on height-for-age scores (column 4) but since neither heteroskedasticity-robust standard errors nor randomization inference indicate significant results,

Table 14 – Intention-to-Treat estimates on parenting practices

	Overall parenting score	Parental consistency	Coercive parenting	Positive encouragement	Quality of parent-child relationship
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample 2021					
Lottery dummy	-0.1634	-0.1179	-0.091	-0.094	-0.0698
	(0.101)	(0.103)	(0.110)	(0.105)	(0.095)
	[0.107]	[0.254]	[0.410]	[0.370]	[0.461]
Clustered SE p-value	0.091	0.184	0.474	0.192	0.33
Fisher p-value	0.09	0.2315	0.3935	0.3285	0.4475
N. obs.	459	463	467	469	470
Panel B: Restricted sample 2021					
Lottery dummy	-0.1819	-0.1599	-0.1314	-0.0736	-0.0348
	(0.108)	(0.113)	(0.120)	(0.112)	(0.107)
	[0.092]	[0.157]	[0.273]	[0.513]	[0.745]
Clustered SE p-value	0.116	0.182	0.404	0.337	0.542
Fisher p-value	0.103	0.145	0.2785	0.512	0.7585
N. obs.	357	359	362	365	366

Note – Regressions of parenting PAFAS scores (reported in units of control-group standard deviations) on the lottery dummy. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

the discussion of external validity makes less sense in this case, which seems to be the result of within-municipality variability. LATE estimates for child health outcomes are similar to the ITT analyses, in that there are no significant results (Table B.2).

1.5.2.2 Parenting

We now focus on outcomes related to parenting skills and practices, as well as selected variables related to the home environment and family relations. We begin by estimating Intention-to-Treat effects on four domains of parenting, including parental consistency, coercive parenting, positive encouragement and a measure of the quality of parent-child relationship, as well as a global parenting score. These data are available for our 2021 samples. All of these are standardized scores, with lower values indicating better parenting.

As can be seen in Table 14, all of our estimated coefficients are consistent with improve-

ments in parenting practices caused by the treatment, even though they are generally not precisely estimated. We find marginally significant results only for the overall parenting score (column 1), with p-values fluctuating around the 10% threshold. Still, we interpret these as suggestive of a positive effect on parenting, at a magnitude of around 0.16-0.18σ on the global parenting score.

Next, we turn our attention to several measures of the methods used by members of the household to discipline the child. Both in 2018 and 2021, we observe four binary measures of violent discipline: psychological aggression, physical punishment, severe physical punishment and any violent discipline (which combines the other three measures). For 2021 only, we also collected information on the use of non-violent discipline methods and the attitudes of the caregiver toward physical punishment. Table 15 presents Intention-to-Treat estimates on all of these outcomes, considering our four samples.

Our most prevalent result when considering the use of different child discipline methods is that of a reduction in the use of physical punishment in the offer-group for 2021, with a p-value equal to 6.5% when considering full sample, although it becomes slightly bigger than 10% for our restricted sample (column 3, Panels C and D). We estimate a reduction in physical punishment to discipline children in 2021 of 8.1-9.1 percentage points, or 21.9-23.6% compared to the control mean. Our estimate for the 2018 restricted sample is similar in magnitude, although smaller percentage-wise compared to a higher control mean (12.7%), and also marginally significant. The equivalent LATE estimates in Table B.4 indicate an effect of around 67% in the 2021 full sample.

Similarly, we find a decrease in severe physical punishment for the 2021 sample, which is marginally significant after restricting the strata (column 4, Panels C and D). We also find a marginally significant increase in the use of psychological aggression for the 2018 full sample, but this result is not sustained after dropping strata with higher differential attrition, neither is it present for 2021 (column 2 of Table 15).

Our final outcomes related to parenting practices concern the home environment and habits of family members when interacting with the child. We observe, in 2018, the self-reported number of weekly hours the caregiver spends with the child, whether the child has children books available at home, and whether someone in the family does the following activities with the child everyday: play, tell stories, take outside and sing/teach songs. Table 16 presents our Intention-to-Treat estimates on these six variables.

The results indicate an increase of around 10.4-12.1 in the number hours the caregiver spends with the child per week, significant at the 1% level for both the full and the restricted samples (column 1). Using the latter as reference, this corresponds to a 17.5% increase relative to the control mean of 69.1 weekly hours. We also observe an increase in the likelihood of having children books available at home, at the order of 7.6 percentage points for the restricted sample, or 14.5% relative to the control mean, although these are only marginally significant (column 2).

Table 15 – Intention-to-Treat estimates on child discipline methods

	Any violent discipline	Psychological aggression	Physical punishment	Severe physical punishment	Attitudes to physical punishment	Any non-violent discipline
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	0.0248	0.0697	-0.0252	0.0146		
	(0.037)	(0.042)	(0.043)	(0.016)		
	[0.505]	[0.093]	[0.558]	[0.367]		
Control mean	0.75	0.6385	0.598	0.0203		
Clustered SE p-value	0.538	0.092	0.363	0.201		
Fisher p-value	0.5115	0.0865	0.567	0.3485		
N. obs.	500	500	500	500		
Panel B: Restricted sample 2018						
Lottery dummy	0.0036	0.0511	-0.0744	0.0237		
	(0.045)	(0.049)	(0.048)	(0.020)		
	[0.937]	[0.297]	[0.123]	[0.234]		
Control mean	0.711	0.5963	0.5872	0.0183		
Clustered SE p-value	0.95	0.438	0.183	0.149		
Fisher p-value	0.9315	0.296	0.123	0.1795		
N. obs.	383	383	383	383		
Panel C: Full sample 2021						
Lottery dummy	-0.0499	-0.0095	-0.0906	-0.019	0.0198	-0.0084
	(0.052)	(0.052)	(0.049)	(0.011)	(0.037)	(0.022)
	[0.340]	[0.854]	[0.065]	[0.071]	[0.590]	[0.706]
Control mean	0.5421	0.4019	0.3832	0.0234	0.1542	0.9626
Clustered SE p-value	0.349	0.856	0.046	0.041	0.492	0.656
Fisher p-value	0.3155	0.8495	0.058	0.092	0.585	0.671
N. obs.	471	471	471	471	468	471
Panel D: Restricted sample 2021						
Lottery dummy	-0.0518	0.0024	-0.0811	-0.0196	0.0248	0.0029
	(0.056)	(0.056)	(0.053)	(0.012)	(0.039)	(0.019)
	[0.358]	[0.966]	[0.123]	[0.114]	[0.524]	[0.881]
Control mean	0.547	0.3978	0.3702	0.0276	0.1326	0.9669
Clustered SE p-value	0.394	0.97	0.106	0.149	0.381	0.88
Fisher p-value	0.3465	0.9655	0.13	0.1585	0.526	0.8835
N. obs.	366	366	366	366	364	366

Note – Regressions of MICS/UNICEF child discipline indicators on the lottery dummy. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity- robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table 16 – Intention-to-Treat estimates on family interactions

	Hours caretaker spends with child per week	Has children books	Plays with child	Tells stories to child	Takes child outside	Sings/teaches songs to child
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	10.3783	0.0693	-0.0153	0.0018	-0.0422	0.0863
	(3.851)	(0.044)	(0.026)	(0.037)	(0.036)	(0.043)
	[0.007]	[0.116]	[0.560]	[0.962]	[0.240]	[0.047]
Control mean	68.184	0.5236	0.922	0.1952	0.1993	0.6407
Clustered SE p-value	0.035	0.029	0.442	0.944	0.282	0.006
Fisher p-value	0.008	0.1005	0.542	0.964	0.235	0.044
N. obs.	489	500	499	494	500	497
Panel B: Restricted sample 2018						
Lottery dummy	12.1083	0.0764	0.0058	-0.0042	-0.0416	0.086
	(4.293)	(0.048)	(0.027)	(0.042)	(0.040)	(0.048)
	[0.005]	[0.116]	[0.827]	[0.920]	[0.302]	[0.071]
Control mean	69.0986	0.5275	0.9263	0.1963	0.211	0.6498
Clustered SE p-value	0.02	0.051	0.788	0.895	0.283	0.037
Fisher p-value	0.0045	0.1245	0.8365	0.924	0.3015	0.076
N. obs.	375	383	382	377	383	380

Note – Regressions of family relations indicators on the lottery dummy. Column 2 refers to a variable indicating whether the child has children books available at home. Columns 3-6 refer to a measure of whether the family reports doing the activities everyday with the child. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Columns 3-6 in Table 16 refer to habits of family members when interacting with the child. Among these, we document a significant impact of PIM on the likelihood of singing or teaching songs to the child everyday (column 6). Our estimates suggest an increase of 13.2% relative to the control mean.

1.5.2.3 Mental health of caregiver

To evaluate program impacts on the mental health of the caregiver, we collected information in 2021 on three dimensions of self-reported mental health: stress, anxiety and depression. For each of these, we use a standardized score, as well as dummies indicating whether the caregiver was categorized to have normal, mild, moderate, severe of extremely severe symptoms. Tables 17, 18 and 19 present the Intention-to-Treat estimates on stress, depression and anxiety, respectively.

Results for stress and depression suggest that PIM did not affect caregivers' mental

Table 17 – Intention-to-Treat estimates on caregiver stress

	Stress score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2021						
Lottery dummy	0.0854	-0.0604	0.0232	0.0319	-0.008	0.0133
	(0.113)	(0.053)	(0.033)	(0.039)	(0.026)	(0.027)
	[0.451]	[0.253]	[0.486]	[0.417]	[0.760]	[0.627]
Control mean		0.6075	0.0935	0.1682	0.0701	0.0607
Clustered SE p-value	0.301	0.143	0.422	0.24	0.678	0.543
Fisher p-value	0.3955	0.221	0.432	0.4075	0.7485	0.6025
N. obs.	469	469	469	469	469	469
Panel B: Restricted sample 2021						
Lottery dummy	0.079	-0.0562	0.0171	0.04	-0.0143	0.0135
	(0.122)	(0.056)	(0.034)	(0.042)	(0.027)	(0.030)
	[0.517]	[0.319]	[0.618]	[0.340]	[0.602]	[0.658]
Control mean	0.372	0.6022	0.0939	0.1657	0.0718	0.0663
Clustered SE p-value	0.5	0.227	0.605	0.221	0.495	0.624
Fisher p-value	364	0.311	0.64	0.352	0.6205	0.6595
N. obs.	364	364	364	364	364	364

Note – Regressions of DASS-21 stress scores (column 1, reported in units of control-group standard deviations) and stress categories (columns 2-6) on the lottery dummy. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

health on these dimensions, with high p-values for most outcomes. The exception is again when clustering standard errors at the municipality-level, which would indicate an increase in mild depression for caregivers in the offer-group; however, we fail to reject the null hypothesis both under heteroskedasticity-robust standard errors and randomization inference, especially for the restricted sample with p-values close to 30% (column 3, Table 18).

As for program effects on anxiety levels, despite a lack of evidence indicating a causal change on the anxiety score, we find a decrease in the occurrence of "normal" anxiety and a corresponding increase in "moderate" anxiety, both for the full and restricted samples (columns 2 and 4, Table 19). This surprising effect corresponds to a 13.4% decrease in normal anxiety and a 66.9% increase in moderate anxiety relative to control means in the restricted sample, or, equivalently, an approximate 10 to 11 percentage points shift from normal to moderate anxiety.

Table 18 – Intention-to-Treat estimates on caregiver depression

	Depression score	Normal	Mild	Moderate	Extremely severe
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample 2021					
Lottery dummy	0.0043	0.001	0.0354	-0.0067	0.0055
	(0.104)	(0.037)	(0.023)	(0.022)	(0.013)
	[0.967]	[0.979]	[0.123]	[0.758]	[0.666]
Control mean		0.8585	0.0377	0.0472	0.0142
Clustered SE p-value	0.94	0.946	0.074	0.65	0.507
Fisher p-value	0.9655	0.975	0.1045	0.735	0.6765
N. obs.	469	469	469	469	469
Panel B: Restricted sample 2021					
Lottery dummy	0.0133	0.004	0.0268	-0.0061	0.0017
	(0.113)	(0.039)	(0.024)	(0.022)	(0.015)
	[0.906]	[0.918]	[0.269]	[0.782]	[0.905]
Control mean		0.8611	0.0444	0.0389	0.0167
Clustered SE p-value	0.847	0.813	0.065	0.704	0.819
Fisher p-value	0.9055	0.9135	0.3015	0.788	0.8785
N. obs.	365	365	365	365	365

Note – Regressions of DASS-21 depression scores (column 1, reported in units of control-group standard deviations) and depression categories (columns 2-6) on the lottery dummy. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table 19 – Intention-to-Treat estimates on caregiver anxiety

	Anxiety score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2021						
Lottery dummy	0.1113	-0.0906	0.0177	0.0858	0.0131	-0.026
	(0.105)	(0.049)	(0.023)	(0.034)	(0.020)	(0.030)
	[0.287]	[0.064]	[0.446]	[0.013]	[0.509]	[0.381]
Control mean		0.7465	0.0469	0.0751	0.0376	0.0939
Clustered SE p-value	0.132	0.02	0.244	0.002	0.453	0.096
Fisher p-value	0.261	0.0445	0.412	0.007	0.5155	0.353
N. obs.	469	469	469	469	469	469
Panel B: Restricted sample 2021						
Lottery dummy	0.1568	-0.1023	0.0077	0.1113	0.0059	-0.0225
	(0.113)	(0.051)	(0.025)	(0.035)	(0.020)	(0.032)
	[0.167]	[0.045]	[0.756]	[0.001]	[0.768]	[0.484]
Control mean		0.7611	0.0444	0.0667	0.0333	0.0944
Clustered SE p-value	0.132	0.041	0.655	0	0.752	0.3
Fisher p-value	0.1495	0.047	0.745	0.0045	0.772	0.4775
N. obs.	364	364	364	364	364	364

Note – Regressions of DASS-21 anxiety scores (column 1, reported in units of control-group standard deviations) and anxiety categories (columns 2-6) on the lottery dummy. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

1.5.2.4 Access to services

Finally, we assess whether PIM had an impact on broadly-defined access to public services. To carry out this analysis, we observe whether interviewees in 2018 were able to provide information on which social assistance center and health unit they attend, whether the child is weighted monthly and whether the child attends childcare. In 2021, we asked caregivers whether anyone in the child's family used a public service during the four weeks preceding the interview (such as going to a health unit or social assistance center), and once again whether the child attends childcare. Results are presented in Table 20.

Our measures of access to services present no change under the treatment relative to the control group in 2018, as shown in columns 1-4 of Table 20. In 2021, after restricting the sample (Panel B), we find suggestive evidence of increased self-reported use of public services in the

Table 20 – Intention-to-Treat estimates on access to services

	2018				2021	
	Informed social assistance center	Informed health unit	Child weighted monthly	Child attends childcare	Used a public service (past 4 weeks)	Child attends childcare
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
Lottery dummy	-0.0086	-0.0017	0.0056	-0.0185	0.0703	-0.0229
	(0.037)	(0.017)	(0.046)	(0.036)	(0.050)	(0.046)
	[0.818]	[0.921]	[0.902]	[0.611]	[0.156]	[0.621]
Control mean	0.7568	0.9595	0.4865	0.2399	0.5888	0.4299
Clustered SE p-value	0.626	0.946	0.897	0.694	0.4	0.476
Fisher p-value	0.822	0.9235	0.899	0.5975	0.1525	0.6255
N. obs.	500	500	500	500	471	471
Panel B: Restricted sample						
Lottery dummy	-0.0211	-0.0105	0.0242	-0.0218	0.0899	-0.0443
	(0.041)	(0.019)	(0.051)	(0.037)	(0.054)	(0.050)
	[0.612]	[0.578]	[0.634]	[0.561]	[0.098]	[0.371]
Control mean	0.7752	0.9725	0.5	0.1881	0.558	0.4586
Clustered SE p-value	0.402	0.651	0.699	0.555	0.245	0.124
Fisher p-value	0.626	0.577	0.6325	0.547	0.0965	0.3935
N. obs.	383	383	383	383	366	366

Note – Regressions of indicators related to access to public services on the lottery dummy. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

four weeks prior to data collection, with an estimated effect of almost 9 percentage points, or 16.1%, although p-values are quite close to the 10% mark both using heteroskedasticity-robust standard errors and performing randomization inference. This result could be interpreted e.g. as a direct effect of being a beneficiary of PIM on knowledge of the local health system and health professionals, which could matter as a means to affect access to these services especially during the covid-19 pandemic.

1.5.3 Concerns with the timing of data collection

Two sources of variation affect the amount of time that has passed between the date of a lottery draw and the date of an interview, for a given child in our sample. First, as detailed in Table 1, we effectively performed 16 rounds of randomizations in 15 municipalities, for a total of 61 strata. Each of the 16 lottery draws took place at a different date between March 2018 and May 2021. This means that while approximately 9 months had passed since the lottery for our earliest-adopters at the moment of data collection in 2018, children in later municipalities had been through the randomization only 3 months prior to the survey.

While these differences of several months do mean that our main estimates are an average of intention-to-treat effects under different stages of child development, this source of variation is fixed for each strata, and thus we are able to appropriately average these differences out in our analyses because of strata fixed effects.

The second source of variation is then related to delayed interviews within a given strata. For 2018, as discussed in section 1.3.2, door-to-door data collection efforts took place during an interval of three weeks in December. This small window of time ensures that this source of variation is negligible, as evidenced by small standard deviations in Table 2.

However, for telephone-based interviews beginning in August 2021, six months had passed between the first and last completed interviews, which ended in February 2022. In a given week of data collection, the order of interview attempts was randomized. However, because attempts continued for several months, a potential problem arises if late interviewees differ significantly from early respondents, and differentially so in terms of treatment and control groups. Late respondents might differ from early respondents e.g. if they are harder to reach because they are more likely to work outside of the house, or if reluctance to respond signals a worse a home environment or is otherwise related to potential outcomes. Regressing a variable counting the time passed between lottery and interview dates on the lottery dummy and strata fixed effects yields a positive and statistically significant estimate for the 2021 sample, but not for 2018, as presented in Panel A of Table 21.

This results indicates that individuals in the offer-group are more likely to respond to the interview later in the data collection window, compared to the control. Another way of looking at these estimates is as in the framework of Behaghel et al. (2015). If we model response behavior as a function of treatment assignment and a latent (omitted) variable related to "reluctance" to respond to the survey, our positive and significant estimate for the 2021 sample suggests that the treatment is inducing a group of so-called "marginal respondents" to actually respond, who would otherwise (in the absence of the treatment) not have been interviewed.

Additionally, we investigate the profile of late respondents by regressing time on a set of caregiver education and family income dummies, as well as strata fixed effects. These estimates are presented in Panel B of Table 21. For the 2018 sample, we find statistically significant positive partial correlations between higher levels of caregiver education and responding to the interview later in the data collection process, as well as a negative and significant estimate for the 2-3 minimum-wage family income dummy, compared to the base level of income. As for 2021, the only statistically significant coefficient is that of the highest level of family income. Put together, these results could indicate that individuals who respond to the survey in later months have higher levels of education and family income.

Despite the possibility of endogeneity, we attempt to verify whether this delay in interviews

Table 21 – Regressions of time elapsed since lottery, by treatment arm and socioeconomic characteristics

	Full sample 2018 (1)	Full sample 2021 (2)
Panel A. Regression of time on the lottery dummy		
Lottery dummy	-0.006	0.6831
	[0.674]	[0.000]
Control mean	4.9319	35.3217
N. obs.	500	471
Panel B. Regression of time on caregiver education and family income dummies		
Caregiver education (omitted = incomplete primary education)		
Completed primary until incomplete secondary education	0.0577	0.1862
	[0.001]	[0.502]
Completed secondary education or more	0.0750	0.3585
	[0.000]	[0.148]
Family income (omitted = less than 1 MW)		
1-2 MW	-0.0118	0.2280
	[0.498]	[0.283]
2-3 MW	-0.0750	0.2224
	[0.005]	[0.457]
3 or more MW	0.0223	1.1034
	[0.479]	[0.001]
Family monthly income missing	0.0113	0.8967
	[0.646]	[0.149]
N. obs.	500	469

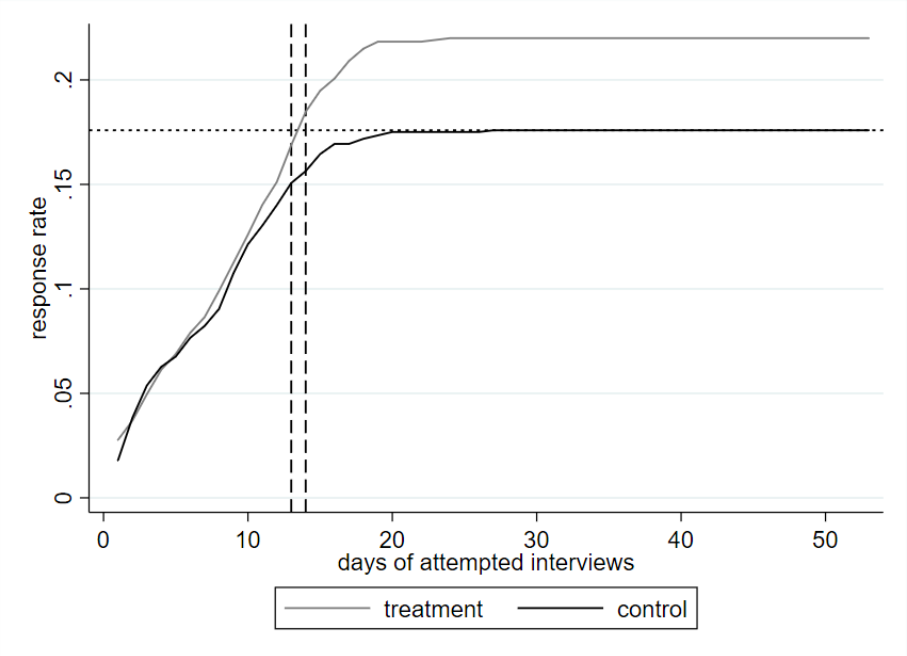
Note – Regressions of time, in months, between lottery and interview dates on the lottery assignment dummy (Panel A) and caregiver education and family income dummies (Panel B). All regressions include strata fixed effects. Robust SE p-value in brackets. MW = minimum wage.

within a given strata influences our results by including time as an independent variable in ITT regressions. Estimates are presented in Tables C.1 to C.9 in Appendix C. As expected, results for 2018 are largely unchanged, since most of the variation in time is already controlled for by the strata fixed effects. As for 2021, although there is more variation in relation to the base specification, with most coefficients experiencing a slight decrease in magnitude while a few decrease by several folds (e.g. estimates for the ITT effect on overall child development decreases by more than 60%, as shown in column 1 of Table C.1) and others even increase (e.g. the estimated coefficient for the ITT on coercive parenting, positive encouragement and quality of parent-child relationship in columns 3-5 of Table C.3), the general conclusions drawn from these estimates are still mostly the same as before.

Finally, we return to the framework proposed by Behaghel et al. (2015) to account for

response behavior and compute bounds on program treatment effects. We use the number of different days in which we called to attempt interviewing each individual in our 2021 sample to measure reluctance to respond. Individuals who only responded after a certain number of attempts are then dropped from the analysis in order to equalize response rates between treatment and control groups. This procedure can be visualized in Figure 1.

Figure 1 – Survey response rate by interview effort



Note – This figure reports 2021 survey response rates by interview effort in terms of days of attempted interview calls to each individual in our sample, for both treatment and control groups. The horizontal dashed line represents maximum response rate for the control group. Vertical lines cross the treatment group curve at the nearest integer values of interview attempt days below and above the point that equalizes response rates between both groups.

Bounds estimates for treatment effects on our main outcomes are presented in Tables D.1 to D.7, in Appendix D. We use each of our main outcomes both in its original form (Panel A) and as the residual after regressing on strata fixed effects (Panel B). In near all cases, the estimated interval agrees qualitatively with Intention-to-Treat estimates presented previously, even when it does not include the ITT point estimate. Unsurprisingly, the majority of lower and upper bounds are not precisely estimated, which was to be expected since most of our ITT estimates are also not statistically significant. One blatant exception is the estimated effect on caregiver moderate anxiety, which we find to be increased in the offer-group even when estimating bounds (column 4, Table D.6). In general, the fact that the estimated bounds seem to agree with ITT estimates allows us to proceed with our main strategy confident that concerns with the timing of data collection do not pose a serious threat to our results.

1.6 Conclusion

In this paper, we studied a large-scale parenting intervention targeting children from socioeconomically vulnerable contexts in their first stages of development. By randomizing assignment to receive a participation offer among eligible families, we were able to form a plausibly-comparable group of untreated children to serve as a control group in our analyses.

Early childhood is known to be a series of critical and sensitive periods to development, during which returns to human capital investments peak and after which it is often hard to mitigate skills acquisition gaps. Several studies have documented huge and prolonged potential of interventions during early childhood to not only increase cognition and socioemotional skills, but also to affect people's lives through adulthood and beyond.

An open question is how scalable are such types of interventions. With thousands of workers and potentially tens or hundreds of thousands of targeted families, assuring quality and homogeneous implementation of a program becomes increasingly challenging.

Our study finds significant program impacts on child development at a magnitude of 18% of a standard deviation. Analyzing different dimensions of development shows that these development gains come mainly from increased fine motor skills among children in the treatment group. We also investigate several measures related to parental skills and practices, and find significant evidence of PIM increasing parental skills among caregivers of treated children. Complementarily, we also find a reduced practice of disciplining children with physical punishments and evidence suggestive of better home environments, with caregivers of children in the treatment group reporting spending more time watching the children per week and an estimated increase in social interactions between children and their families in the treatment group.

Importantly, gains in development are mainly found in the short-term, between 3 and 9 months after lottery draws, while we do not find program effects in the second follow-up, comprising generally longer exposition times. Although this evidence is consistent with fade out, such as evidenced in Colombia (ANDREW et al., 2018), it is likely the result of decreased dosage of implementation caused by the pandemic after the start of 2020. The fact that we observe gains in parenting in spite of the pandemic might indicate that we are not able to detect existing but smaller increases in child development in 2021. Our evidence of increasing child development mainly through motor skills gains also differs from similar recent programs in developing countries, whose main effects seem to be on domains of cognition/problem solving and language development. To the extent that motor skill development might be more dependent on in-person stimulation, this could also explain the lack of effects after a more widespread adoption of remote or hybrid modalities of care.

We contribute to a growing, though still incipient, body of literature concerned with larger-scale early childhood programs in developing countries, much of whose evidence comes from randomized trials. This ongoing effort to document what works in programs aimed at young children is far from definitive. Policy-making in the last few decades has started to assimilate the importance of investing early in life, guided by evidence on the huge and lasting effects brought by early development. The role of research is now to inform policy on how to best invest at scale and specially with quality. Only then will children from all backgrounds be allowed to truly flourish, benefiting society as whole.

APPENDIX A. Intention-to-Treat estimates on program participation for the complete registry data

Table A.1 – Intention-to-Treat estimates on program participation - full registry data

	2018		2021	
	Full sample	Restricted sample	Full sample	Restricted sample
	(1)	(2)	(3)	(4)
Panel A: Has any record of home visit				
Lottery dummy	0.1651	0.1787	0.2351	0.2164
	(0.015)	(0.018)	(0.014)	(0.015)
	[0.000]	[0.000]	[0.000]	[0.000]
Control mean	0.0233	0.0249	0.0318	0.0325
Clustered SE p-value	0.052	0.046	0.000	0.003
Fisher p-value	0.000	0.000	0.000	0.000
Panel B: Has any record of quarterly assessment				
Lottery dummy	0.1981	0.2216	0.2737	0.2687
	(0.017)	(0.020)	(0.015)	(0.016)
	[0.000]	[0.000]	[0.000]	[0.000]
Control mean	0.0433	0.0439	0.0505	0.0457
Clustered SE p-value	0.017	0.018	0.000	0.001
Fisher p-value	0.000	0.000	0.000	0.000
N. obs.	1659	1256	2419	1902
N. strata	37	31	61	50

Note – Regressions of treatment compliance indicators on the lottery dummy. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

APPENDIX B. Local Average Treatment Effect estimates

Table B.1 – Local Average Treatment Effect estimates on child development

	Overall development score	Communication skills	Gross motor skills	Fine motor skills	Problem solving	Personal and social ability
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
PIM beneficiary	0.456	0.1728	0.3889	0.546	0.252	0.1749
	(0.225)	(0.251)	(0.200)	(0.230)	(0.229)	(0.235)
	[0.044]	[0.491]	[0.052]	[0.018]	[0.271]	[0.458]
Clustered SE p-value	0.005	0.433	0.091	0.003	0.375	0.397
N. obs.	491	493	492	493	492	493
Panel B: Restricted sample 2018						
PIM beneficiary	0.3415	0.0682	0.3008	0.4206	0.2495	0.1554
	(0.233)	(0.262)	(0.207)	(0.241)	(0.240)	(0.255)
	[0.144]	[0.795]	[0.146]	[0.082]	[0.298]	[0.543]
Clustered SE p-value	0.059	0.629	0.339	0.135	0.256	0.412
N. obs.	377	379	378	379	378	379
Panel C: Full sample 2021						
Lottery dummy	0.1737	0.052	-0.0263	0.2567	0.1151	0.1134
	(0.280)	(0.285)	(0.276)	(0.270)	(0.279)	(0.271)
	[0.536]	[0.855]	[0.924]	[0.343]	[0.680]	[0.675]
Clustered SE p-value	0.475	0.836	0.913	0.322	0.479	0.683
N. obs.	465	470	469	467	470	471
Panel D: Restricted sample 2021						
Lottery dummy	0.1997	0.1301	0.1178	0.2365	0.1363	0.063
	(0.298)	(0.296)	(0.304)	(0.288)	(0.307)	(0.290)
	[0.504]	[0.661]	[0.699]	[0.411]	[0.658]	[0.828]
Clustered SE p-value	0.435	0.6	0.678	0.432	0.459	0.812
N. obs.	360	365	364	362	365	366

Note – Local Average Treatment Effect of PIM on child development ASQ-3 scores (reported in units of control-group standard deviations). Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level.

Table B.2 – Local Average Treatment Effect estimates on child health

	Child ever hospitalized	Number of times hospitalized	Weight-for-age	Height-for-age
	(1)	(2)	(3)	(4)
Panel A: Full sample 2018				
PIM beneficiary	-0.0336	-0.0575	0.1742	-1.8127
	(0.110)	(0.217)	(0.908)	(1.330)
	[0.761]	[0.791]	[0.848]	[0.175]
Control mean	0.3277	0.4966	-0.0304	-0.9767
Clustered SE p-value	0.754	0.817	0.732	0.065
N. obs.	500	500	225	144
Panel B: Restricted sample 2018				
PIM beneficiary	-0.0813	-0.0582	0.005	-1.7862
	(0.117)	(0.240)	(0.857)	(1.297)
	[0.488]	[0.809]	[0.995]	[0.172]
Control mean	0.3257	0.4908	0.0624	-0.59
Clustered SE p-value	0.495	0.859	0.994	0.091
N. obs.	383	383	180	117

Note – Local Average Treatment Effect of PIM on child health variables. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table B.3 – Local Average Treatment Effect estimates on parenting practices

	Overall parenting score	Parental consistency	Coercive parenting	Positive encouragement	Quality of parent-child relationship
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample 2021					
PIM beneficiary	-0.4524	-0.3354	-0.2567	-0.2669	-0.1984
	(0.287)	(0.296)	(0.315)	(0.299)	(0.270)
	[0.116]	[0.258]	[0.416]	[0.372]	[0.462]
Clustered SE p-value	0.101	0.137	0.507	0.157	0.318
N. obs.	459	463	467	469	470
Panel B: Restricted sample 2021					
PIM beneficiary	-0.5095	-0.4517	-0.3679	-0.2096	-0.0991
	(0.310)	(0.322)	(0.341)	(0.322)	(0.305)
	[0.102]	[0.161]	[0.281]	[0.515]	[0.746]
Clustered SE p-value	0.128	0.121	0.449	0.305	0.533
N. obs.	357	359	362	365	366

Note – Local Average Treatment Effect of PIM on parenting PAFAS scores (reported in units of control-group standard deviations). Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level.

Table B.4 – Local Average Treatment Effect estimates on child discipline methods

	Any violent discipline	Psychological aggression	Physical punishment	Severe physical punishment	Attitudes to physical punishment	Any non-violent discipline
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
PIM beneficiary	0.0623	0.1751	-0.0632	0.0367		
	(0.093)	(0.106)	(0.108)	(0.041)		
	[0.506]	[0.099]	[0.559]	[0.372]		
Control mean	0.75	0.6385	0.598	0.0203		
Clustered SE p-value	0.573	0.169	0.307	0.215		
N. obs.	500	500	500	500		
Panel B: Restricted sample 2018						
PIM beneficiary	0.0085	0.1222	-0.1778	0.0565		
	(0.108)	(0.118)	(0.117)	(0.048)		
	[0.937]	[0.302]	[0.130]	[0.241]		
Control mean	0.711	0.5963	0.5872	0.0183		
Clustered SE p-value	0.95	0.477	0.126	0.147		
N. obs.	383	383	383	383		
Panel C: Full sample 2021						
PIM beneficiary	-0.1431	-0.0272	-0.2596	-0.0545	0.057	-0.0242
	(0.152)	(0.148)	(0.144)	(0.031)	(0.107)	(0.064)
	[0.347]	[0.854]	[0.072]	[0.079]	[0.593]	[0.707]
Control mean	0.5421	0.4019	0.3832	0.0234	0.1542	0.9626
Clustered SE p-value	0.38	0.855	0.112	0.036	0.488	0.668
N. obs.	471	471	471	471	468	471
Panel D: Restricted sample 2021						
PIM beneficiary	-0.1474	0.0068	-0.231	-0.056	0.0711	0.0082
	(0.162)	(0.158)	(0.152)	(0.036)	(0.113)	(0.055)
	[0.364]	[0.966]	[0.129]	[0.123]	[0.529]	[0.881]
Control mean	0.547	0.3978	0.3702	0.0276	0.1326	0.9669
Clustered SE p-value	0.426	0.97	0.192	0.124	0.374	0.878
N. obs.	366	366	366	366	364	366

Note – Local Average Treatment Effect of PIM on MICS/UNICEF child discipline indicators. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table B.5 – Local Average Treatment Effect estimates on family interactions

	Hours caretaker spends with child per week	Has children books	Plays with child	Tells stories to child	Takes child outside	Sings/teaches songs to child
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
PIM beneficiary	25.9472	0.174	-0.0384	0.0045	-0.106	0.2167
	(9.697)	(0.114)	(0.066)	(0.094)	(0.090)	(0.112)
	[0.008]	[0.127]	[0.560]	[0.962]	[0.241]	[0.054]
Control mean	68.184	0.5236	0.922	0.1952	0.1993	0.6407
Clustered SE p-value	0.046	0.024	0.392	0.944	0.352	0.004
N. obs.	489	500	499	494	500	497
Panel B: Restricted sample 2018						
PIM beneficiary	28.7717	0.1824	0.0139	-0.0101	-0.0995	0.2055
	(10.298)	(0.119)	(0.063)	(0.100)	(0.097)	(0.117)
	[0.005]	[0.125]	[0.826]	[0.919]	[0.304]	[0.079]
Control mean	69.0986	0.5275	0.9263	0.1963	0.211	0.6498
Clustered SE p-value	0.018	0.02	0.788	0.895	0.337	0.056
N. obs.	375	383	382	377	383	380

Note – Local Average Treatment Effect of PIM on family relations indicators. Column 2 refers to a variable indicating whether the child has children books available at home. Columns 3-6 refer to a measure of whether the family reports doing the activities everyday with the child. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table B.6 – Local Average Treatment Effect estimates on caregiver stress

	Stress score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2021						
PIM beneficiary	0.244	-0.1725	0.0662	0.091	-0.0228	0.0381
	(0.323)	(0.151)	(0.096)	(0.112)	(0.075)	(0.078)
	[0.450]	[0.254]	[0.489]	[0.418]	[0.761]	[0.627]
Control mean		0.6075	0.0935	0.1682	0.0701	0.0607
Clustered SE p-value	0.279	0.108	0.369	0.271	0.689	0.556
N. obs.	469	469	469	469	469	469
Panel B: Restricted sample 2021						
PIM beneficiary	0.2255	-0.1605	0.0487	0.114	-0.0407	0.0384
	(0.346)	(0.160)	(0.098)	(0.120)	(0.079)	(0.087)
	[0.514]	[0.318]	[0.619]	[0.343]	[0.605]	[0.658]
Control mean		0.6022	0.0939	0.1657	0.0718	0.0663
Clustered SE p-value	0.358	0.197	0.577	0.256	0.516	0.632
N. obs.	364	364	364	364	364	364

Note – Local Average Treatment Effect of PIM on DASS-21 stress scores (column 1, reported in units of control-group standard deviations) and stress categories (columns 2-6). Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table B.7 – Local Average Treatment Effect estimates on caregiver depression

	Depression score	Normal	Mild	Moderate	Extremely severe
	(1)	(2)	(3)	(4)	(6)
Panel A: Full sample 2021					
PIM beneficiary	0.0124	0.0028	0.1013	-0.0192	0.0159
	(0.296)	(0.105)	(0.065)	(0.062)	(0.037)
	[0.967]	[0.979]	[0.121]	[0.759]	[0.665]
Control mean		0.8585	0.0377	0.0472	0.0142
Clustered SE p-value	0.94	0.946	0.038	0.634	0.505
N. obs.	469	469	469	469	469
Panel B: Restricted sample 2021					
PIM beneficiary	0.0379	0.0113	0.0761	-0.0173	0.0049
	(0.320)	(0.110)	(0.069)	(0.063)	(0.041)
	[0.906]	[0.919]	[0.267]	[0.783]	[0.905]
Control mean		0.8611	0.0444	0.0389	0.0167
Clustered SE p-value	0.849	0.809	0.032	0.69	0.818
N. obs.	365	365	365	365	365

Note – Local Average Treatment Effect of PIM on DASS-21 depression scores (column 1, reported in units of control-group standard deviations) and depression categories (columns 2-6). Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table B.8 – Local Average Treatment Effect estimates on caregiver anxiety

	Anxiety score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2021						
PIM beneficiary	0.3195	-0.2601	0.0508	0.2464	0.0377	-0.0748
	(0.300)	(0.141)	(0.067)	(0.102)	(0.058)	(0.086)
	[0.288]	[0.067]	[0.446]	[0.016]	[0.512]	[0.387]
Control mean		0.7465	0.0469	0.0751	0.0376	0.0939
Clustered SE p-value	0.141	0.028	0.3	0.007	0.417	0.103
N. obs.	469	469	469	469	469	469
Panel B: Restricted sample 2021						
PIM beneficiary	0.448	-0.2922	0.0219	0.3179	0.0167	-0.0644
	(0.323)	(0.146)	(0.070)	(0.104)	(0.057)	(0.093)
	[0.167]	[0.046]	[0.756]	[0.002]	[0.768]	[0.489]
Control mean		0.7611	0.0444	0.0667	0.0333	0.0944
Clustered SE p-value	0.131	0.046	0.673	0.001	0.741	0.303
N. obs.	364	364	364	364	364	364

Note – Local Average Treatment Effect of PIM on DASS-21 anxiety scores (column 1, reported in units of control-group standard deviations) and anxiety categories (columns 2-6). Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table B.9 – Local Average Treatment Effect estimates on access to services

	2018				2021	
	Informed social assistance center	Informed health unit	Child weighted monthly	Child attends childcare	Used a public service (past 4 weeks)	Child attends childcare
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
PIM beneficiary	-0.0216	-0.0043	0.0141	-0.0465	0.2014	-0.0656
	(0.094)	(0.044)	(0.115)	(0.090)	(0.142)	(0.132)
	[0.818]	[0.921]	[0.902]	[0.606]	[0.157]	[0.619]
Control mean	0.7568	0.9595	0.4865	0.2399	0.5888	0.4299
Clustered SE p-value	0.602	0.945	0.899	0.71	0.427	0.49
N. obs.	500	500	500	500	471	471
Panel B: Restricted sample						
PIM beneficiary	-0.0503	-0.0251	0.0577	-0.052	0.2561	-0.1263
	(0.099)	(0.045)	(0.121)	(0.088)	(0.156)	(0.141)
	[0.612]	[0.578]	[0.635]	[0.555]	[0.101]	[0.372]
Control mean	0.7752	0.9725	0.5	0.1881	0.558	0.4586
Clustered SE p-value	0.36	0.635	0.708	0.576	0.294	0.139
N. obs.	383	383	383	383	366	366

Note – Local Average Treatment Effect of PIM on indicators related to access to public services. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

APPENDIX C. Intention-to-treat estimates controlling for time elapsed since lottery

Table C.1 – ITT estimates on child development, controlling for time

	Overall development score	Communication skills	Gross motor skills	Fine motor skills	Problem solving	Personal and social ability
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	0.1814	0.0668	0.1552	0.2164	0.0997	0.0686
	(0.089)	(0.099)	(0.080)	(0.090)	(0.090)	(0.093)
	[0.041]	[0.502]	[0.052]	[0.016]	[0.270]	[0.461]
Clustered SE p-value	0.015	0.409	0.103	0.031	0.372	0.438
Fisher p-value	0.0365	0.483	0.0545	0.0135	0.2615	0.4625
N. obs.	491	493	492	493	492	493
Panel B: Restricted sample 2018						
Lottery dummy	0.1418	0.0253	0.1236	0.1745	0.1061	0.063
	(0.098)	(0.110)	(0.088)	(0.101)	(0.100)	(0.106)
	[0.150]	[0.818]	[0.160]	[0.084]	[0.289]	[0.554]
Clustered SE p-value	0.108	0.664	0.367	0.2	0.248	0.451
Fisher p-value	0.161	0.817	0.182	0.0915	0.301	0.546
N. obs.	377	379	378	379	378	379
Panel C: Full sample 2021						
Lottery dummy	0.0151	-0.0073	-0.0383	0.0488	0.0062	-0.003
	(0.100)	(0.102)	(0.097)	(0.095)	(0.100)	(0.097)
	[0.880]	[0.943]	[0.695]	[0.607]	[0.951]	[0.976]
Clustered SE p-value	0.855	0.934	0.637	0.602	0.922	0.974
Fisher p-value	0.8655	0.9415	0.706	0.6075	0.9465	0.9755
N. obs.	465	470	469	467	470	471
Panel D: Restricted sample 2021						
Lottery dummy	0.0255	0.0241	-0.0041	0.046	0.0196	-0.0249
	(0.108)	(0.105)	(0.109)	(0.103)	(0.111)	(0.106)
	[0.814]	[0.819]	[0.970]	[0.656]	[0.860]	[0.814]
Clustered SE p-value	0.775	0.781	0.968	0.677	0.787	0.782
Fisher p-value	0.7965	0.799	0.9735	0.6445	0.8735	0.812
N. obs.	360	365	364	362	365	366

Note – Regressions of child development ASQ-3 scores (reported in units of control-group standard deviations) on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

Table C.2 – ITT estimates on child health, controlling for time

	Child ever hospitalized	Number of times hospitalized	Weight-for-age	Height-for-age
	(1)	(2)	(3)	(4)
Panel A: Full sample 2018				
Lottery dummy	-0.0144	-0.0246	0.0731	-0.8803
	(0.044)	(0.087)	(0.380)	(0.648)
	[0.742]	[0.777]	[0.848]	[0.177]
Control mean	0.3277	0.4966	-0.0304	-0.9767
Clustered SE p-value	0.718	0.794	0.724	0.07
Fisher p-value	0.7055	0.775	0.833	0.178
N. obs.	500	500	225	144
Panel B: Restricted sample 2018				
Lottery dummy	-0.0356	-0.0255	-0.0042	-0.8256
	(0.049)	(0.101)	(0.373)	(0.589)
	[0.467]	[0.801]	[0.991]	[0.164]
Control mean	0.3257	0.4908	0.0624	-0.59
Clustered SE p-value	0.459	0.851	0.989	0.075
Fisher p-value	0.464	0.8105	0.9895	0.2455
N. obs.	383	383	180	117

Note – Regressions of child health variables on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table C.3 – ITT estimates on parenting practices, controlling for time

	Overall parenting score	Parental consistency	Coercive parenting	Positive encouragement	Quality of parent-child relationship
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample 2021					
Lottery dummy	-0.1604	-0.0726	-0.0941	-0.1077	-0.0968
	(0.100)	(0.104)	(0.112)	(0.106)	(0.092)
	[0.111]	[0.486]	[0.402]	[0.310]	[0.292]
Clustered SE p-value	0.099	0.414	0.467	0.153	0.065
Fisher p-value	0.096	0.4655	0.38	0.262	0.297
N. obs.	459	463	467	469	470
Panel B: Restricted sample 2021					
Lottery dummy	-0.1951	-0.1121	-0.1551	-0.0933	-0.0747
	(0.107)	(0.115)	(0.122)	(0.114)	(0.103)
	[0.068]	[0.329]	[0.204]	[0.412]	[0.468]
Clustered SE p-value	0.096	0.33	0.303	0.243	0.079
Fisher p-value	0.08	0.3125	0.1915	0.406	0.48
N. obs.	357	359	362	365	366

Note – Regressions of parenting PAFAS scores (reported in units of control-group standard deviations) on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations.

Table C.4 – ITT estimates on child discipline methods, controlling for time

	Any violent discipline	Psychological aggression	Physical punishment	Severe physical punishment	Attitudes to physical punishment	Any non-violent discipline
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	0.0242	0.0691	-0.0248	0.0146		
	(0.037)	(0.042)	(0.043)	(0.016)		
	[0.517]	[0.097]	[0.565]	[0.369]		
Control mean	0.75	0.6385	0.598	0.0203		
Clustered SE p-value	0.545	0.088	0.361	0.207		
Fisher p-value	0.5225	0.088	0.5755	0.351		
N. obs.	500	500	500	500		
Panel B: Restricted sample 2018						
Lottery dummy	0.0022	0.0494	-0.0744	0.0237		
	(0.045)	(0.049)	(0.049)	(0.020)		
	[0.961]	[0.316]	[0.126]	[0.233]		
Control mean	0.711	0.5963	0.5872	0.0183		
Clustered SE p-value	0.968	0.443	0.187	0.15		
Fisher p-value	0.967	0.312	0.1235	0.1805		
N. obs.	383	383	383	383		
Panel C: Full sample 2021						
Lottery dummy	-0.0489	-0.0107	-0.0869	-0.0166	0.0219	-0.0108
	(0.053)	(0.053)	(0.049)	(0.010)	(0.037)	(0.023)
	[0.359]	[0.840]	[0.078]	[0.097]	[0.552]	[0.637]
Control mean	0.5421	0.4019	0.3832	0.0234	0.1542	0.9626
Clustered SE p-value	0.367	0.836	0.051	0.079	0.464	0.573
Fisher p-value	0.3235	0.8355	0.0645	0.146	0.547	0.598
N. obs.	471	471	471	471	468	471
Panel D: Restricted sample 2021						
Lottery dummy	-0.0521	0.001	-0.0781	-0.0165	0.0294	-0.0008
	(0.058)	(0.057)	(0.052)	(0.012)	(0.039)	(0.020)
	[0.368]	[0.986]	[0.137]	[0.165]	[0.450]	[0.967]
Control mean	0.547	0.3978	0.3702	0.0276	0.1326	0.9669
Clustered SE p-value	0.377	0.987	0.1	0.219	0.308	0.965
Fisher p-value	0.346	0.987	0.14	0.251	0.45	0.9685
N. obs.	366	366	366	366	364	366

Note – Regressions of MICS/UNICEF child discipline indicators on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table C.5 – ITT estimates on family interactions, controlling for time

	Hours caretaker spends with child per week	Has children books	Plays with child	Tells stories to child	Takes child outside	Sings/teaches songs to child
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2018						
Lottery dummy	10.3888	0.0699	-0.0152	0.0024	-0.0413	0.0865
	(3.858)	(0.044)	(0.026)	(0.037)	(0.036)	(0.043)
	[0.007]	[0.112]	[0.563]	[0.950]	[0.250]	[0.046]
Control mean	68.184	0.5236	0.922	0.1952	0.1993	0.6407
Clustered SE p-value	0.034	0.022	0.445	0.922	0.29	0.006
Fisher p-value	0.008	0.0985	0.547	0.9455	0.2425	0.0435
N. obs.	489	500	499	494	500	497
Panel B: Restricted sample 2018						
Lottery dummy	12.1651	0.0772	0.0064	-0.0022	-0.0402	0.0878
	(4.309)	(0.048)	(0.027)	(0.041)	(0.040)	(0.048)
	[0.005]	[0.112]	[0.810]	[0.959]	[0.319]	[0.066]
Control mean	69.0986	0.5275	0.9263	0.1963	0.211	0.6498
Clustered SE p-value	0.019	0.048	0.773	0.941	0.285	0.034
Fisher p-value	0.0045	0.12	0.816	0.959	0.319	0.071
N. obs.	375	383	382	377	383	380

Note – Regressions of family relations indicators on the lottery dummy. Column 2 refers to a variable indicating whether the child has children books available at home. Columns 3-6 refer to a measure of whether the family reports doing the activities everyday with the child. All regressions include a control variable of the time passed between lottery and interview dates. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table C.6 – ITT estimates on caregiver stress, controlling for time

	Stress score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2021						
Lottery dummy	0.0998	-0.0636	0.0166	0.0329	-0.0085	0.0227
	(0.114)	(0.054)	(0.034)	(0.040)	(0.027)	(0.027)
	[0.379]	[0.235]	[0.628]	[0.416]	[0.752]	[0.401]
Control mean		0.6075	0.0935	0.1682	0.0701	0.0607
Clustered SE p-value	0.277	0.14	0.497	0.22	0.659	0.243
Fisher p-value	0.326	0.1975	0.5815	0.3935	0.734	0.374
N. obs.	469	469	469	469	469	469
Panel B: Restricted sample 2021						
Lottery dummy	0.1096	-0.0638	0.0073	0.0445	-0.0159	0.0279
	(0.123)	(0.058)	(0.035)	(0.043)	(0.028)	(0.030)
	[0.374]	[0.268]	[0.833]	[0.306]	[0.574]	[0.350]
Control mean		0.6022	0.0939	0.1657	0.0718	0.0663
Clustered SE p-value	0.275	0.208	0.79	0.181	0.426	0.25
Fisher p-value	0.3515	0.2495	0.8355	0.307	0.586	0.347
N. obs.	364	364	364	364	364	364

Note – Regressions of DASS-21 stress scores (column 1, reported in units of control-group standard deviations) and stress categories (columns 2-6) on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table C.7 – ITT estimates on caregiver depression, controlling for time

	Depression score	Normal	Mild	Moderate	Extremely severe
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample 2021					
Lottery dummy	0.0264	-0.009	0.035	-0.0005	0.0068
	(0.103)	(0.037)	(0.024)	(0.022)	(0.012)
	[0.799]	[0.809]	[0.151]	[0.983]	[0.564]
Control mean		0.8585	0.0377	0.0472	0.0142
Clustered SE p-value	0.701	0.622	0.108	0.976	0.383
Fisher p-value	0.7885	0.8065	0.108	0.9795	0.6085
N. obs.	469	469	469	469	469
Panel B: Restricted sample 2021					
Lottery dummy	0.0411	-0.0068	0.0223	0.0019	0.0027
	(0.113)	(0.040)	(0.026)	(0.021)	(0.013)
	[0.715]	[0.864]	[0.399]	[0.928]	[0.832]
Control mean		0.8611	0.0444	0.0389	0.0167
Clustered SE p-value	0.586	0.702	0.197	0.9	0.676
Fisher p-value	0.7015	0.852	0.396	0.9285	0.816
N. obs.	365	365	365	365	365

Note – Regressions of DASS-21 depression scores (column 1, reported in units of control-group standard deviations) and depression categories (columns 2-6) on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table C.8 – ITT estimates on caregiver anxiety, controlling for time

	Anxiety score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample 2021						
Lottery dummy	0.1176	-0.0912	0.0146	0.0801	0.019	-0.0225
	(0.103)	(0.049)	(0.024)	(0.034)	(0.021)	(0.029)
	[0.255]	[0.064]	[0.541]	[0.020]	[0.375]	[0.441]
Control mean		0.7465	0.0469	0.0751	0.0376	0.0939
Clustered SE p-value	0.09	0.017	0.349	0.003	0.356	0.105
Fisher p-value	0.233	0.045	0.501	0.01	0.3315	0.438
N. obs.	469	469	469	469	469	469
Panel B: Restricted sample 2021						
Lottery dummy	0.1772	-0.1048	-0.0005	0.1088	0.0119	-0.0152
	(0.112)	(0.051)	(0.025)	(0.035)	(0.022)	(0.032)
	[0.114]	[0.041]	[0.983]	[0.002]	[0.583]	[0.632]
Control mean		0.7611	0.0444	0.0667	0.0333	0.0944
Clustered SE p-value	0.068	0.034	0.974	0	0.613	0.402
Fisher p-value	0.106	0.0435	0.9795	0.005	0.57	0.625
N. obs.	364	364	364	364	364	364

Note – Regressions of DASS-21 anxiety scores (column 1, reported in units of control-group standard deviations) and anxiety categories (columns 2-6) on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table C.9 – ITT estimates on access to services, controlling for time

	2018				2021	
	Informed social assistance center	Informed health unit	Child weighted monthly	Child attends childcare	Used a public service (past 4 weeks)	Child attends childcare
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
Lottery dummy	-0.0085	-0.0015	0.0032	-0.0183	0.0621	-0.0568
	(0.037)	(0.017)	(0.045)	(0.036)	(0.050)	(0.046)
	[0.819]	[0.929]	[0.945]	[0.616]	[0.219]	[0.217]
Control mean	0.7568	0.9595	0.4865	0.2399	0.5888	0.4299
Clustered SE p-value	0.619	0.951	0.944	0.703	0.434	0.125
Fisher p-value	0.8225	0.9295	0.9485	0.5995	0.222	0.2065
N. obs.	500	500	500	500	471	471
Panel B: Restricted sample						
Lottery dummy	-0.0217	-0.0103	0.0191	-0.0221	0.0763	-0.0871
	(0.042)	(0.019)	(0.050)	(0.037)	(0.056)	(0.049)
	[0.602]	[0.583]	[0.705]	[0.556]	[0.173]	[0.079]
Control mean	0.7752	0.9725	0.5	0.1881	0.558	0.4586
Clustered SE p-value	0.348	0.651	0.773	0.557	0.302	0.007
Fisher p-value	0.617	0.583	0.7095	0.544	0.1615	0.087
N. obs.	383	383	383	383	366	366

Note – Regressions of indicators related to access to public services on the lottery dummy. All regressions include a control variable of the time passed between lottery and interview dates. Regressions for the 2021 restricted sample include family income dummies. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Clustered SE = standard errors clustered at the municipality level. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

APPENDIX D. Treatment effects bounds estimates on main outcomes for 2021

Table D.1 – Treatment effect bounds estimates on child development

	Overall development score	Communication skills	Gross motor skills	Fine motor skills	Problem solving	Personal and social ability
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Original variable						
Lower bound effect	0.0457	-0.0929	0.0315	0.0349	0.0779	0.0007
	[0.760]	[0.576]	[0.811]	[0.837]	[0.538]	[0.996]
Upper bound effect	0.0767	-0.0271	0.0639	0.0937	0.1156	0.053
	[0.544]	[0.811]	[0.553]	[0.487]	[0.322]	[0.653]
Panel B: Residual after regressing on strata dummies						
Lower bound effect	-0.0044	-0.0337	-0.0186	0.0082	-0.0053	-0.0251
	[0.976]	[0.795]	[0.892]	[0.954]	[0.963]	[0.834]
Upper bound effect	0.025	0.0135	0.0109	0.05	0.019	0.0094
	[0.845]	[0.912]	[0.916]	[0.681]	[0.861]	[0.931]
N. obs. after truncation	411	416	415	413	416	417

Note – Treatment effect bounds estimates on child development ASQ-3 scores (reported in units of control-group standard deviations). Panel B considers as dependent variables the residual after regressing each variable on strata dummies. Bounds are estimated as in Behaghel et al. (2015) using the number of different days in which we attempted interviewing each family to truncate the treatment group. P-values in brackets.

Table D.2 – Treatment effect bounds estimates on parental practices

	Overall parenting score	Parental consistency	Coercive parenting	Positive encouragement	Quality of parent-child relationship
	(1)	(2)	(3)	(4)	(5)
Panel A: Original variable					
Lower bound effect	-0.1388	-0.1174	-0.0364	-0.0933	-0.103
	[0.383]	[0.662]	[0.845]	[0.567]	[0.739]
Upper bound effect	-0.101	-0.0889	-0.0117	-0.0468	-0.0931
	[0.456]	[0.489]	[0.926]	[0.756]	[0.373]
Panel B: Residual after regressing on strata dummies					
Lower bound effect	-0.1237	-0.1061	-0.0711	-0.0701	-0.041
	[0.505]	[0.642]	[0.670]	[0.648]	[0.889]
Upper bound effect	-0.0865	-0.0715	-0.0373	-0.0362	-0.008
	[0.542]	[0.615]	[0.781]	[0.786]	[0.948]
N. obs. after truncation	407	411	413	415	416

Note – Treatment effect bounds estimates on parenting PAFAS scores (reported in units of control-group standard deviations). Panel B considers as dependent variables the residual after regressing each variable on strata dummies. Bounds are estimated as in Behaghel et al. (2015) using the number of different days in which we attempted interviewing each family to truncate the treatment group. P-values in brackets.

Table D.3 – Treatment effect bounds estimates on child discipline methods

	Psychological aggression	Physical punishment	Severe physical punishment	Any violent discipline	Any non-violent discipline	Attitudes to physical punishment
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Original variable						
Lower bound effect	-0.0064	-0.0597	-0.0184	-0.0546	-0.0155	0.0047
	[0.914]	[0.274]	[0.105]	[0.401]	[0.501]	[0.900]
Upper bound effect	0.027	-0.0358	-0.0184	-0.0141	-0.0107	0.0191
	[0.772]	[0.622]	[0.105]	[0.825]	[0.594]	[0.668]
Panel B: Residual after regressing on strata dummies						
Lower bound effect	-0.0036	-0.0697	-0.0141	-0.0467	-0.0166	0.0188
	[0.940]	[0.222]	[0.187]	[0.372]	[0.411]	[0.585]
Upper bound effect	0.0012	-0.0572	-0.0136	-0.0382	-0.014	0.0219
	[0.979]	[0.193]	[0.201]	[0.408]	[0.459]	[0.513]
N. obs. after truncation	417	417	417	417	417	415

Note – Treatment effect bounds estimates on MICS/UNICEF child discipline indicators. Panel B considers as dependent variables the residual after regressing each variable on strata dummies. Bounds are estimated as in Behaghel et al. (2015) using the number of different days in which we attempted interviewing each family to truncate the treatment group. P-values in brackets.

Table D.4 – Treatment effect bounds estimates on caregiver stress

	Stress score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Original variable						
Lower bound effect	0.0831	-0.0886	0.0173	0.0434	-0.0217	0.021
	[0.664]	[0.233]	[0.583]	[0.305]	[0.349]	[0.419]
Upper bound effect	0.1084	-0.06	0.0317	0.0529	-0.0217	0.0257
	[0.342]	[0.337]	[0.473]	[0.210]	[0.349]	[0.335]
Panel B: Residual after regressing on strata dummies						
Lower bound effect	0.0526	-0.0608	0.0264	0.0279	-0.0216	0.0109
	[0.731]	[0.182]	[0.349]	[0.544]	[0.387]	[0.653]
Upper bound effect	0.0746	-0.0505	0.0332	0.0383	-0.0166	0.0128
	[0.478]	[0.355]	[0.317]	[0.283]	[0.444]	[0.591]
N. obs. after truncation	416	416	416	416	416	416

Note – Treatment effect bounds estimates on DASS-21 stress scores (column 1, reported in units of control-group standard deviations) and stress categories (columns 2-6). Panel B considers as dependent variables the residual after regressing each variable on strata dummies. Bounds are estimated as in Behaghel et al. (2015) using the number of different days in which we attempted interviewing each family to truncate the treatment group. P-values in brackets.

Table D.5 – Treatment effect bounds estimates on caregiver depression

	Depression score	Normal	Mild	Moderate	Extremely severe
	(1)	(2)	(3)	(4)	(5)
Panel A: Original variable					
Lower bound effect	0.0049	-0.014	0.0253	0.0063	0.0053
	[0.976]	[0.759]	[0.251]	[0.773]	[0.677]
Upper bound effect	0.0305	-0.0044	0.0349	0.0063	0.0053
	[0.787]	[0.905]	[0.386]	[0.773]	[0.677]
Panel B: Residual after regressing on strata dummies					
Lower bound effect	-0.0059	-0.0065	0.035	-0.0056	0.0027
	[0.970]	[0.842]	[0.076]	[0.847]	[0.848]
Upper bound effect	0.0203	0.0014	0.0372	-0.0002	0.0046
	[0.850]	[0.976]	[0.078]	[0.991]	[0.706]
N. obs. after truncation	415	415	415	415	415

Note – Treatment effect bounds estimates on DASS-21 depression scores (column 1, reported in units of control-group standard deviations) and depression categories (columns 2-5). Panel B considers as dependent variables the residual after regressing each variable on strata dummies. Bounds are estimated as in Behaghel et al. (2015) using the number of different days in which we attempted interviewing each family to truncate the treatment group. P-values in brackets.

Table D.6 – Treatment effect bounds estimates on caregiver anxiety

	Anxiety score	Normal	Mild	Moderate	Severe	Extremely severe
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Original variable						
Lower bound effect	0.1108	-0.0989	0.0014	0.0838	0.0107	-0.0163
	[0.573]	[0.136]	[0.945]	[0.014]	[0.594]	[0.560]
Upper bound effect	0.1337	-0.0797	0.0062	0.0934	0.0107	-0.0115
	[0.224]	[0.132]	[0.802]	[0.019]	[0.594]	[0.701]
Panel B: Residual after regressing on strata dummies						
Lower bound effect	0.0619	-0.0696	0.0068	0.0745	0.0125	-0.0342
	[0.690]	[0.099]	[0.724]	[0.017]	[0.599]	[0.227]
Upper bound effect	0.0873	-0.0614	0.0086	0.0767	0.0165	-0.0304
	[0.428]	[0.235]	[0.671]	[0.010]	[0.385]	[0.239]
N. obs. after truncation	416	416	416	416	416	416

Note – Treatment effect bounds estimates on DASS-21 anxiety scores (column 1, reported in units of control-group standard deviations) and depression categories (columns 2-6). Panel B considers as dependent variables the residual after regressing each variable on strata dummies. Bounds are estimated as in Behaghel et al. (2015) using the number of different days in which we attempted interviewing each family to truncate the treatment group. P-values in brackets.

Table D.7 – Treatment effect bounds estimates on access to services

	Used a public service (past 4 weeks)	Child attends childcare
	(1)	(2)
Panel A: Original variable		
Lower bound effect	0.0135	-0.0535
	[0.850]	[0.357]
Upper bound effect	0.054	-0.0248
	[0.341]	[0.764]
Panel B: Residual after regressing on strata dummies		
Lower bound effect	0.0645	-0.0234
	[0.187]	[0.602]
Upper bound effect	0.0707	-0.0178
	[0.110]	[0.670]
N. obs. after truncation	417	417

Note – Treatment effect bounds estimates on indicators related to access to public services. Panel B considers as dependent variables the residual after regressing each variable on strata dummies. Bounds are estimated as in Behaghel et al. (2015) using the number of different days in which we attempted interviewing each family to truncate the treatment group. P-values in brackets.

APPENDIX E. Balancedness tests for gender heterogeneity subsamples

Table E.1 – Balancedness tests - 2018 female child subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Female caregiver	234	0.94	0.95	0.57	178	0.94	0.98	0.32
Caregiver with incomplete primary education	234	0.23	0.34	0.09	178	0.21	0.37	0.02
Caregiver with complete primary until incomplete secondary education	234	0.39	0.33	0.44	178	0.41	0.30	0.22
Caregiver with at least complete secondary education	234	0.38	0.33	0.35	178	0.38	0.33	0.29
Child age	234	23.23	23.81	0.83	178	21.97	20.74	0.99
Child lives with both parents	234	0.66	0.58	0.57	178	0.63	0.57	0.87
Child lives with mother only	234	0.24	0.33	0.16	178	0.26	0.34	0.30
Pregnancy was planned	232	0.32	0.29	0.59	176	0.31	0.30	0.97
Mother had at most 4 prenatal appointments	234	0.10	0.07	0.36	178	0.12	0.04	0.13
Mother had 5-7 prenatal appointments	234	0.19	0.34	0.01	178	0.19	0.36	0.02
Mother had 8 or more prenatal appointments	234	0.59	0.47	0.14	178	0.58	0.50	0.41
Number of prenatal appointments missing	234	0.11	0.12	0.74	178	0.12	0.10	0.53
Birth weight <2500	234	0.10	0.14	0.23	178	0.08	0.14	0.21
Birth weight 2500-2999	234	0.18	0.22	0.68	178	0.19	0.20	0.89
Birth weight 3000-3999	234	0.62	0.54	0.28	178	0.62	0.57	0.48
Birth weight >4000	234	0.07	0.09	0.63	178	0.07	0.10	0.67
Birth weight missing	234	0.04	0.01	0.28	178	0.05	0.00	0.12
Number of people living in the house	234	4.48	4.33	0.36	178	4.55	4.22	0.14
Number of rooms in the house	233	4.65	4.52	0.99	177	4.66	4.45	0.87
House has electricity	234	0.99	0.99	0.74	178	1.00	0.99	0.36
House has piped water	234	0.98	0.95	0.43	178	0.98	0.97	0.95
House is connected to sewage network	234	0.78	0.81	0.56	178	0.78	0.79	0.84
Family consumes untreated water	234	0.50	0.53	0.50	178	0.50	0.50	0.60
Family owns a computer	234	0.30	0.19	0.12	178	0.31	0.20	0.06
Family owns a tablet	234	0.17	0.13	0.69	178	0.19	0.12	0.51
Family has access to mobile internet	233	0.71	0.73	0.72	177	0.71	0.74	0.83
Family has access to dial-up internet	233	0.12	0.12	0.89	177	0.13	0.10	0.59
Family has access to broadband internet	232	0.50	0.39	0.10	176	0.50	0.38	0.19
Family monthly income until 1 MW	234	0.40	0.46	0.48	178	0.37	0.47	0.17
Family monthly income 1-2 MW	234	0.33	0.25	0.12	178	0.36	0.24	0.03
Family monthly income 2-3 MW	234	0.10	0.09	0.99	178	0.09	0.08	0.74
Family monthly income 3 or more MW	234	0.08	0.09	0.45	178	0.09	0.10	0.95
Family monthly income missing	234	0.09	0.11	0.47	178	0.08	0.12	0.48
Someone in the household benefits from cash transfer programs	232	0.63	0.67	0.54	176	0.61	0.67	0.83
Joint F test				0.37				0.00

Note – Balancedness tests for the female child 2018 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

Table E.2 – Balancedness tests - 2021 female child subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Caregiver lives with the child	229	0.98	0.96	0.12	182	0.99	0.95	0.10
Female caregiver	229	0.99	0.97	0.99	182	0.99	0.96	0.99
Caregiver with incomplete primary education	229	0.19	0.18	0.50	182	0.18	0.19	0.91
Caregiver with complete primary until incomplete secondary education	229	0.27	0.21	0.45	182	0.26	0.22	0.64
Caregiver with at least complete secondary education	229	0.54	0.60	0.22	182	0.57	0.59	0.61
Child age	229	47.82	53.09	0.88	182	51.48	53.29	0.52
Child lives with both parents	229	0.56	0.61	0.15	182	0.53	0.62	0.17
Child lives with mother only	229	0.43	0.37	0.12	182	0.46	0.35	0.08
Number of people living in the house	229	4.25	4.19	0.20	182	4.35	4.15	0.16
Number of rooms in the house	229	5.29	5.17	0.65	182	5.26	5.07	0.36
House has electricity	229	1.00	0.99	0.35	182	1.00	0.99	0.35
House has piped water	229	0.98	0.95	0.25	182	0.98	0.95	0.28
House is connected to sewage network	229	0.65	0.66	0.94	182	0.66	0.67	0.95
Family owns a computer	229	0.34	0.33	0.59	182	0.36	0.31	0.21
Family owns a tablet	229	0.24	0.19	0.22	182	0.26	0.19	0.23
Family has access to mobile internet	229	0.86	0.86	0.96	182	0.88	0.85	0.69
Family has access to dial-up internet	226	0.21	0.24	0.52	180	0.20	0.24	0.53
Family has access to broadband internet	226	0.74	0.74	0.43	180	0.76	0.73	0.19
Someone in the household was infected with covid	228	0.29	0.34	0.01	181	0.24	0.33	0.01
Someone in the house lost their job during the pandemic	229	0.43	0.38	0.35	182	0.44	0.39	0.39
Caregiver had to stop taking care of child because of the pandemic	229	0.07	0.10	0.44	182	0.08	0.12	0.41
Family monthly income until 1 MW	229	0.28	0.42	0.17	182	0.26	0.40	0.09
Family monthly income 1-2 MW	229	0.34	0.36	0.76	182	0.34	0.35	0.65
Family monthly income 2-3 MW	229	0.14	0.11	0.92	182	0.15	0.13	0.92
Family monthly income 3 or more MW	229	0.21	0.09	0.18	182	0.21	0.09	0.12
Family monthly income missing	229	0.03	0.02	0.78	182	0.04	0.02	0.78
Someone in the household benefits from cash transfer programs	229	0.44	0.56	0.61	182	0.43	0.53	0.53
Someone receives Auxílio Emergencial	229	0.60	0.65	0.51	182	0.60	0.62	0.74
Joint F test				0.06				0.00

Note – Balancedness tests for the female child 2021 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

Table E.3 – Balancedness tests - 2018 male child subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Female caregiver	268	0.96	0.94	0.48	207	0.96	0.93	0.21
Caregiver with incomplete primary education	268	0.25	0.36	0.13	207	0.28	0.34	0.49
Caregiver with complete primary until incomplete secondary education	268	0.42	0.37	0.74	207	0.42	0.38	0.73
Caregiver with at least complete secondary education	268	0.32	0.27	0.27	207	0.30	0.28	0.76
Child age	268	21.96	23.09	0.98	207	21.16	20.39	0.43
Child lives with both parents	268	0.62	0.59	0.81	207	0.62	0.61	0.40
Child lives with mother only	268	0.34	0.29	0.18	207	0.34	0.26	0.04
Pregnancy was planned	265	0.28	0.36	0.22	204	0.33	0.33	0.70
Mother had at most 4 prenatal appointments	268	0.07	0.09	0.51	207	0.06	0.09	0.74
Mother had 5-7 prenatal appointments	268	0.20	0.24	0.40	207	0.20	0.23	0.37
Mother had 8 or more prenatal appointments	268	0.64	0.51	0.05	207	0.66	0.52	0.04
Number of prenatal appointments missing	268	0.09	0.15	0.22	207	0.08	0.16	0.11
Birth weight <2500	268	0.04	0.06	0.56	207	0.05	0.08	0.56
Birth weight 2500-2999	268	0.21	0.21	0.71	207	0.24	0.17	0.77
Birth weight 3000-3999	268	0.65	0.61	0.33	207	0.62	0.60	0.38
Birth weight >4000	268	0.08	0.09	0.80	207	0.06	0.11	0.35
Birth weight missing	268	0.02	0.03	0.35	207	0.03	0.04	0.36
Number of people living in the house	268	4.61	4.64	0.94	207	4.52	4.63	0.64
Number of rooms in the house	266	4.86	4.59	0.60	205	4.90	4.65	0.81
House has electricity	268	1.00	0.99	0.36	207	1.00	0.99	0.36
House has piped water	268	0.99	0.98	0.62	207	1.00	0.97	0.06
House is connected to sewage network	268	0.73	0.81	0.40	207	0.73	0.84	0.35
Family consumes untreated water	268	0.51	0.49	0.51	207	0.51	0.47	0.44
Family owns a computer	268	0.25	0.24	0.70	207	0.25	0.25	0.41
Family owns a tablet	268	0.19	0.16	0.53	207	0.19	0.16	0.66
Family has access to mobile internet	266	0.68	0.65	0.56	205	0.67	0.62	0.50
Family has access to dial-up internet	268	0.08	0.08	0.45	207	0.09	0.09	0.31
Family has access to broadband internet	268	0.34	0.44	0.10	207	0.37	0.41	0.37
Family monthly income until 1 MW	268	0.44	0.46	0.90	207	0.43	0.45	0.85
Family monthly income 1-2 MW	268	0.28	0.29	0.61	207	0.27	0.28	0.60
Family monthly income 2-3 MW	268	0.13	0.08	0.13	207	0.14	0.06	0.10
Family monthly income 3 or more MW	268	0.03	0.05	0.42	207	0.04	0.05	0.60
Family monthly income missing	268	0.11	0.12	0.75	207	0.13	0.15	0.57
Someone in the household benefits from cash transfer programs	264	0.57	0.70	0.16	203	0.60	0.67	0.85
Joint F test				0.18				0.04

Note – Balancedness tests for the male child 2018 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

Table E.4 – Balancedness tests - 2021 male child subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Caregiver lives with the child	249	0.96	0.99	0.33	191	0.95	0.99	0.33
Female caregiver	249	0.95	0.98	0.15	191	0.94	0.99	0.08
Caregiver with incomplete primary education	247	0.22	0.24	0.41	189	0.22	0.27	0.24
Caregiver with complete primary until incomplete secondary education	247	0.28	0.36	0.71	189	0.27	0.33	0.73
Caregiver with at least complete secondary education	247	0.50	0.40	0.29	189	0.51	0.40	0.20
Child age	249	50.42	52.32	0.15	191	54.39	53.10	0.16
Child lives with both parents	248	0.69	0.55	0.09	190	0.69	0.58	0.23
Child lives with mother only	248	0.28	0.39	0.35	190	0.29	0.35	0.68
Number of people living in the house	249	4.11	4.25	0.59	191	4.20	4.12	0.78
Number of rooms in the house	249	5.08	5.25	0.14	191	5.24	5.36	0.25
House has electricity	249	1.00	1.00	-	191	1.00	1.00	-
House has piped water	249	0.97	0.94	0.93	191	0.98	0.94	0.35
House is connected to sewage network	249	0.68	0.74	0.40	191	0.67	0.73	0.48
Family owns a computer	249	0.36	0.39	0.61	191	0.41	0.42	0.91
Family owns a tablet	249	0.15	0.14	0.95	191	0.19	0.12	0.51
Family has access to mobile internet	248	0.89	0.87	0.59	190	0.89	0.86	0.44
Family has access to dial-up internet	245	0.26	0.30	0.51	189	0.25	0.30	0.56
Family has access to broadband internet	249	0.74	0.71	0.83	191	0.74	0.71	0.77
Someone in the household was infected with covid	248	0.32	0.26	0.38	191	0.28	0.27	0.70
Someone in the house lost their job during the pandemic	249	0.37	0.49	0.30	191	0.40	0.45	0.51
Caregiver had to stop taking care of child because of the pandemic	248	0.10	0.11	0.54	191	0.12	0.10	0.85
Family monthly income until 1 MW	249	0.37	0.41	0.82	191	0.31	0.39	0.34
Family monthly income 1-2 MW	249	0.36	0.31	0.54	191	0.37	0.32	0.40
Family monthly income 2-3 MW	249	0.12	0.15	0.84	191	0.15	0.15	0.94
Family monthly income 3 or more MW	249	0.12	0.12	0.54	191	0.13	0.13	0.75
Family monthly income missing	249	0.03	0.02	0.66	191	0.04	0.01	0.47
Someone in the household benefits from cash transfer programs	248	0.52	0.52	0.74	190	0.49	0.48	0.93
Someone receives Auxílio Emergencial	249	0.59	0.51	0.69	191	0.55	0.54	0.89
Joint F test				0.25				0.01

Note – Balancedness tests for the male child 2021 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

APPENDIX F. Balancedness tests for caregiver education heterogeneity subsamples

Table F.1 – Balancedness tests - 2018 low education caregiver subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Female caregiver	342	0.95	0.95	0.87	262	0.94	0.95	0.83
Caregiver with incomplete primary education	342	0.37	0.50	0.13	262	0.37	0.50	0.16
Caregiver with complete primary until incomplete secondary education	342	0.63	0.50	0.13	262	0.63	0.50	0.16
Female child	342	0.49	0.41	0.24	262	0.49	0.40	0.34
Child age	342	22.91	23.28	0.79	262	21.60	20.08	0.79
Child lives with both parents	342	0.61	0.57	0.79	262	0.59	0.58	0.69
Child lives with mother only	342	0.33	0.32	0.54	262	0.33	0.31	0.20
Pregnancy was planned	339	0.26	0.29	0.69	259	0.26	0.29	0.53
Mother had at most 4 prenatal appointments	342	0.11	0.09	0.33	262	0.12	0.08	0.12
Mother had 5-7 prenatal appointments	342	0.23	0.33	0.04	262	0.22	0.36	0.02
Mother had 8 or more prenatal appointments	342	0.53	0.45	0.22	262	0.54	0.44	0.17
Number of prenatal appointments missing	342	0.13	0.13	0.86	262	0.12	0.13	0.95
Birth weight <2500	342	0.08	0.08	0.91	262	0.06	0.08	0.67
Birth weight 2500-2999	342	0.20	0.23	0.27	262	0.22	0.19	0.85
Birth weight 3000-3999	342	0.63	0.58	0.20	262	0.62	0.58	0.25
Birth weight >4000	342	0.08	0.10	0.56	262	0.07	0.12	0.17
Birth weight missing	342	0.02	0.01	0.95	262	0.02	0.02	0.95
Number of people living in the house	342	4.78	4.64	0.38	262	4.86	4.58	0.15
Number of rooms in the house	340	4.47	4.39	0.81	260	4.54	4.38	0.53
House has electricity	342	0.99	0.99	0.99	262	1.00	0.99	0.19
House has piped water	342	0.98	0.97	0.54	262	0.99	0.97	0.46
House is connected to sewage network	342	0.77	0.81	0.33	262	0.80	0.82	0.74
Family consumes untreated water	342	0.61	0.58	0.35	262	0.58	0.56	0.63
Family owns a computer	342	0.16	0.15	0.76	262	0.19	0.15	0.77
Family owns a tablet	342	0.14	0.12	0.79	262	0.14	0.12	0.58
Family has access to mobile internet	339	0.68	0.65	0.74	259	0.69	0.62	0.46
Family has access to dial-up internet	341	0.08	0.06	0.92	261	0.10	0.06	0.54
Family has access to broadband internet	340	0.35	0.36	0.65	260	0.35	0.33	0.83
Family monthly income until 1 MW	342	0.52	0.51	0.59	262	0.49	0.53	0.86
Family monthly income 1-2 MW	342	0.28	0.28	0.60	262	0.28	0.27	0.75
Family monthly income 2-3 MW	342	0.06	0.06	0.72	262	0.07	0.03	0.19
Family monthly income 3 or more MW	342	0.05	0.05	0.81	262	0.06	0.05	0.67
Family monthly income missing	342	0.09	0.10	0.86	262	0.10	0.12	0.63
Someone in the household benefits from cash transfer programs	339	0.66	0.79	0.06	259	0.68	0.78	0.49
Joint F test				0.60				0.09

Note – Balancedness tests for the low education caregiver 2018 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

Table F.2 – Balancedness tests - 2021 low education caregiver subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Caregiver lives with the child	234	0.98	0.97	0.92	180	0.98	0.97	0.92
Female caregiver	234	0.98	0.98	0.77	180	0.98	0.99	0.25
Caregiver with incomplete primary education	234	0.43	0.42	0.82	180	0.43	0.45	0.43
Caregiver with complete primary until incomplete secondary education	234	0.57	0.58	0.82	180	0.57	0.55	0.43
Female child	234	0.48	0.36	0.03	180	0.48	0.38	0.14
Child age	234	49.43	51.62	0.43	180	53.32	51.94	0.24
Child lives with both parents	233	0.58	0.52	0.85	179	0.55	0.51	0.84
Child lives with mother only	233	0.39	0.41	0.56	179	0.44	0.40	0.40
Number of people living in the house	234	4.41	4.56	0.74	180	4.62	4.43	0.51
Number of rooms in the house	234	4.83	5.28	0.02	180	4.90	5.37	0.04
House has electricity	234	1.00	1.00	-	180	1.00	1.00	-
House has piped water	234	0.97	0.96	0.79	180	0.98	0.98	0.91
House is connected to sewage network	234	0.66	0.72	0.69	180	0.66	0.74	0.57
Family owns a computer	234	0.20	0.22	0.57	180	0.26	0.22	0.10
Family owns a tablet	234	0.11	0.15	0.42	180	0.11	0.15	0.59
Family has access to mobile internet	234	0.86	0.84	0.43	180	0.87	0.83	0.16
Family has access to dial-up internet	230	0.24	0.29	0.39	178	0.21	0.29	0.49
Family has access to broadband internet	231	0.71	0.71	0.49	178	0.71	0.70	0.12
Someone in the household was infected with covid	233	0.22	0.28	0.15	180	0.16	0.28	0.05
Someone in the house lost their job during the pandemic	234	0.41	0.41	0.90	180	0.47	0.39	0.40
Caregiver had to stop taking care of child because of the pandemic	233	0.08	0.12	0.07	180	0.10	0.13	0.11
Family monthly income until 1 MW	234	0.45	0.48	0.63	180	0.41	0.46	0.25
Family monthly income 1-2 MW	234	0.35	0.33	0.97	180	0.36	0.33	0.88
Family monthly income 2-3 MW	234	0.11	0.10	0.28	180	0.13	0.11	0.17
Family monthly income 3 or more MW	234	0.06	0.06	0.84	180	0.07	0.06	0.79
Family monthly income missing	234	0.02	0.03	0.88	180	0.03	0.03	0.88
Someone in the household benefits from cash transfer programs	234	0.64	0.65	0.80	180	0.59	0.63	0.76
Someone receives Auxílio Emergencial	234	0.61	0.58	0.61	180	0.57	0.58	0.72
Joint F test				0.06				0.00

Note – Balancedness tests for the low education caregiver 2021 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

Table F.3 – Balancedness tests - 2018 high education caregiver subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Female caregiver	160	0.96	0.94	0.90	123	0.96	0.95	0.66
Female child	160	0.56	0.48	0.32	123	0.58	0.45	0.11
Child age	160	22.07	23.70	0.45	123	21.54	21.59	0.43
Child lives with both parents	160	0.69	0.63	0.65	123	0.68	0.62	0.84
Child lives with mother only	160	0.22	0.27	0.60	123	0.23	0.24	0.98
Pregnancy was planned	158	0.38	0.42	0.56	121	0.42	0.39	0.92
Mother had at most 4 prenatal appointments	160	0.04	0.07	0.80	123	0.04	0.05	0.92
Mother had 5-7 prenatal appointments	160	0.14	0.17	0.90	123	0.14	0.12	0.64
Mother had 8 or more prenatal appointments	160	0.76	0.61	0.12	123	0.77	0.68	0.51
Number of prenatal appointments missing	160	0.06	0.15	0.03	123	0.05	0.15	0.12
Birth weight <2500	160	0.06	0.14	0.11	123	0.07	0.15	0.24
Birth weight 2500-2999	160	0.19	0.16	0.75	123	0.19	0.15	0.93
Birth weight 3000-3999	160	0.64	0.58	0.70	123	0.61	0.61	0.95
Birth weight >4000	160	0.06	0.09	0.78	123	0.05	0.06	0.98
Birth weight missing	160	0.06	0.03	0.40	123	0.07	0.03	0.22
Number of people living in the house	160	4.10	4.18	0.95	123	3.91	4.18	0.47
Number of rooms in the house	159	5.27	4.95	0.50	122	5.23	5.02	0.88
House has electricity	160	1.00	1.00	-	123	1.00	1.00	-
House has piped water	160	0.99	0.97	0.30	123	0.98	0.95	0.31
House is connected to sewage network	160	0.72	0.82	0.38	123	0.68	0.82	0.32
Family consumes untreated water	160	0.32	0.33	0.95	123	0.35	0.29	0.54
Family owns a computer	160	0.49	0.38	0.11	123	0.47	0.41	0.34
Family owns a tablet	160	0.25	0.20	0.36	123	0.28	0.20	0.22
Family has access to mobile internet	160	0.72	0.77	0.99	123	0.70	0.77	0.82
Family has access to dial-up internet	160	0.13	0.18	0.61	123	0.12	0.17	0.60
Family has access to broadband internet	160	0.57	0.56	0.70	123	0.60	0.56	0.85
Family monthly income until 1 MW	160	0.24	0.33	0.31	123	0.23	0.30	0.40
Family monthly income 1-2 MW	160	0.36	0.25	0.37	123	0.39	0.24	0.26
Family monthly income 2-3 MW	160	0.22	0.15	0.15	123	0.19	0.15	0.40
Family monthly income 3 or more MW	160	0.07	0.11	0.58	123	0.09	0.14	0.77
Family monthly income missing	160	0.11	0.16	0.18	123	0.11	0.17	0.17
Someone in the household benefits from cash transfer programs	157	0.47	0.44	0.71	120	0.45	0.42	0.70
Joint F test				0.24				0.02

Note – Balancedness tests for the high education caregiver 2018 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

Table F.4 – Balancedness tests - 2021 high education caregiver subsamples

	Full sample				Restricted sample			
Variable	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Caregiver lives with the child	242	0.96	0.98	0.64	191	0.96	0.98	0.87
Female caregiver	242	0.96	0.97	0.54	191	0.95	0.97	0.54
Female child	242	0.52	0.56	0.97	191	0.54	0.56	0.82
Child age	242	48.74	53.70	0.85	191	52.47	54.45	0.48
Child lives with both parents	242	0.67	0.63	0.59	191	0.66	0.69	0.31
Child lives with mother only	242	0.32	0.36	0.63	191	0.32	0.30	0.34
Number of people living in the house	242	3.98	3.87	0.34	191	3.99	3.82	0.24
Number of rooms in the house	242	5.51	5.16	0.07	191	5.56	5.08	0.02
House has electricity	242	1.00	0.99	0.33	191	1.00	0.99	0.33
House has piped water	242	0.98	0.92	0.12	191	0.98	0.91	0.06
House is connected to sewage network	242	0.67	0.68	0.83	191	0.67	0.66	0.77
Family owns a computer	242	0.49	0.50	0.65	191	0.49	0.52	0.70
Family owns a tablet	242	0.27	0.18	0.00	191	0.32	0.15	0.00
Family has access to mobile internet	241	0.90	0.89	0.67	190	0.89	0.88	0.86
Family has access to dial-up internet	239	0.23	0.25	0.69	189	0.24	0.25	0.84
Family has access to broadband internet	242	0.78	0.74	0.59	191	0.79	0.74	0.30
Someone in the household was infected with covid	241	0.39	0.32	0.64	190	0.34	0.32	0.98
Someone in the house lost their job during the pandemic	242	0.39	0.47	0.18	191	0.37	0.46	0.12
Caregiver had to stop taking care of child because of the pandemic	242	0.09	0.09	0.48	191	0.10	0.09	0.54
Family monthly income until 1 MW	242	0.21	0.35	0.08	191	0.17	0.33	0.06
Family monthly income 1-2 MW	242	0.35	0.33	0.49	191	0.35	0.34	0.67
Family monthly income 2-3 MW	242	0.15	0.17	0.48	191	0.18	0.18	0.85
Family monthly income 3 or more MW	242	0.26	0.14	0.15	191	0.25	0.16	0.18
Family monthly income missing	242	0.04	0.01	0.11	191	0.05	0.00	0.09
Someone in the household benefits from cash transfer programs	241	0.32	0.43	0.62	190	0.35	0.38	0.83
Someone receives Auxílio Emergencial	242	0.58	0.58	0.95	191	0.58	0.58	0.76
Joint F test				0.21				0.04

Note – Balancedness tests for the high education caregiver 2021 subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables except the house electricity indicator (excluded for lack of variability).

2 Can Human Capital Investments for At-Risk Youth Affect Welfare Dependency? Experimental Evidence from the Protejo Program

2.1 Introduction

A relatively large body of literature has studied the impacts of training programs on labor market outcomes. Technical and vocational education and training (TVET) has been widespread to serve vulnerable young populations in developing countries (MCKENZIE; ROBALINO, 2010) with the necessary skills to facilitate school-to-work transition, increase employment, especially in the formal market, and improve the quality of work overall.

Examples of such programs in developing countries are the well-known Jóvenes en Acción, in Colombia, and Juventud y Empleo, in the Dominican Republic. Both programs combined vocational and soft skills training with on-the-job experience targeting young people from disadvantaged backgrounds. The evidence, however, suggests largely different impacts between men and women, and there seems to be no pattern about which group benefits the most. For instance, the Colombian program has been shown to have had positive effects formal employment and earnings, especially for women, while also inducing individuals who participated to enroll in tertiary education (ATTANASIO et al., 2011; ATTANASIO et al., 2017; KUGLER et al., 2022). On the other hand, the initiative in the Dominican Republic was found to have no effects on average employment levels, but to increase formality levels among men, and to reduce teenage pregnancy and improve personal skills in the long-term among women (IBARRARAN et al., 2014; ACEVEDO et al., 2017; IBARRARÁN et al., 2019). Alzúa et al. (2016a) also find positive effects of a similar training program in Argentina, the entra21, on formal employment and earnings, which are stronger for men but dissipate after four years.

Tripney et al. (2013) and McKenzie (2017) summarize the evidence from these and other studies of training programs in developing countries. Their general conclusions seem to suggest that, in general, providing training programs to vulnerable youth does little to increase overall employment levels, while it does seem to increase formal employment among beneficiaries. More generally, Card et al. (2018) survey 207 studies on active labor market programs (which include but are not limited to trainings) and find overall effects on employment, especially in medium and long-terms, with larger average effects for females.

While one of the more direct objectives of trainings and, more generally, active labor market programs, is obviously related to increasing employment and otherwise improving employment conditions, it can be argued that an ultimate goal related to these shorter term objectives would be to lift vulnerable populations out of poverty and consequently out of government assistance dependency. This is especially true given that several of the aforementioned initiatives focus on

"at-risk", low-income young populations, who have often not completed their formal education and are at a long span of unemployment, and several studies have found, at best, small effects on employment levels for only a subgroup of beneficiaries.

However, few studies have analyzed whether these initiatives are able to affect the welfare beneficiary status of individuals go undergo trainings (MAWN et al., 2017). Alzúa et al. (2016b) analyze the aforementioned entra21 in Argentina and estimate reductions in receipt of child-related public cash transfers among young women. An older study of the JobStart initiative for school dropouts in the United States also found reductions in welfare receipt among women not living with their children at the beginning of the intervention (CAVE et al., 1993).

We take advantage of randomizations conducted in 2010 to select beneficiaries of a training program in Rio de Janeiro, targeting young people aged 15 to 24 from vulnerable communities, to assess whether this kind of human capital investment is able to affect their dependency of social protection programs. The Protejo program was designed as part of a larger national initiative of the Brazilian federal government concerned with crime prevention and control, and provided 800 hours of training and cultural activities, including both technical and socioemotional components. One key difference between Protejo and other trainings in developing countries is the more focalized approach towards at-risk youth from extremely marginalized backgrounds. Protejo targeted not only low-income and low-education youth, but specifically young populations found to be in contexts exposure to domestic and urban violence. A recent evidence map by Apunyo et al. (2022) concluded that these marginalized populations are understudied among studies of interventions to increase youth employment.

Protejo has been previously studied using this same round of lottery assignments by Lima (2019) and Barros et al. (2019). Their work relied on primary survey data to assess program impacts on a broad range of outcomes related to family formation and social interactions, conflict and victimization, community engagement, political and religious activity, educational perspectives, and noncognitive skills. They also linked the lottery registry to an administrative database containing a complete overview of the Brazilian formal labor market. They find persistent gains in formal employment beginning three years after the end the program for males and four years for females. Another noteworthy result is a statistically significant and positive effect on the probability of having children by the time of the follow-up survey, although they did not investigate differences by subgroups in this case. They find no effects on education or noncognitive skills.

We expand on previous work by Lima (2019) and Barros et al. (2019) in several ways. First, we link the Protejo lottery registry to two novel data sources covering the population of beneficiaries of two nation-wide cash transfer programs, which allows us to identify whether individuals who applied to participate in Protejo benefited from the social protection network up to ten years after the end of the program. We also link Protejo applicants to an administrative

database of deaths in the State of Rio de Janeiro, to observe mortality rates across treatment arms, since one of ultimate goals of the program was to reduce violence and crime. Then, we take advantage of previous linkage to formal employment records to replicate their findings, but we refine the linkage procedure by dropping likely false positives and by including one additional year of employment data. Finally, we use survey data to investigate potential mechanisms affecting individuals' dependency of government welfare benefits.

We find evidence supporting previous results on formal labor market employment, with large and lasting effects for males who were randomly assigned to the Protejo offer-group up to seven years after the end of the intervention, but no significant effects for females. Our estimates indicate that Protejo effectively led men with initial lower educational attainment to be as likely to be employed as untreated males of a higher initial educational level. Concerning welfare programs, we estimate that women have an increased chance of benefiting from two cash transfer programs up to ten years after Protejo, especially among low education and initially childless females. This increase in welfare receipt is consistent with the lack of labor market effects for women and also with our results on family formation, which suggest an increased chance of treated women having children and becoming single mothers. Mediation analyses support the idea that program impacts on welfare receipt for women are partially mediated by increased fertility.

The remainder of this paper is organized as follows. Section 2 presents the main components of the program and details the experimental design on which we rely. In section 3, we describe our main sources of data and test for balancedness in observables across treatment arms. Section 4 presents our estimation strategy. Finally, estimation results are available in section 5 and we conclude in section 6.

2.2 Institutional Background

2.2.1 The 'Protejo' Program

The "Projeto de Proteção de Jovens em Territórios Vulneráveis", or simply Protejo, is a national project implemented by the Brazilian Ministry of Justice starting in 2007, as part of a broader federal program ("Programa Nacional de Segurança Pública com Cidadania", or Pronasci) that aimed at preventing and controlling criminality nationwide. Protejo was designed to help achieve this broader goal by providing training and social integration to vulnerable youth, targeting a population of 15 to 24 year olds under high risk of exposure to domestic and/or urban violence and who generally have low levels of education and income. Between 2008 and 2013, Protejo benefited around 35,000 individuals in 14 Brazilian states, for a total federal investment of over R$ 200 million in that period alone (around US$ 100 million in 2013) (BARROS et al., 2019).

Under its goals of promoting social participation, protecting vulnerable youth and promoting education and work, the 2010 edition of Protejo in Rio de Janeiro (object of our analysis), acted via a combination of technical training and socioemotional development, providing a total of 800 hours of activities to its beneficiaries between November 2010 and July 2011. These hours were divided into 12 modules, of which half were dedicated towards a more technical set of skills, including general domains such as mathematics and Portuguese, information technology, mentoring sessions to build an individual development plan, as well as 160 hours of technical skills training on a specific service occupation (such as handbag manufacturing, hairdressing, sales techniques, among other). The other half of the total load was dedicated mainly towards cultural and soft skills development, including sports and cultural activities, citizenship building and socioemotional development, which encompassed lessons e.g. on self-knowledge, the value of education and work, and family formation.

Selected individuals who actually participated in the program (i.e. attended a minimum of hours defined by the implementation team) were entitled to a monthly stipend of 100 reais, for up to 12 months. We observe whether each individual in the evaluation sample received monthly payments and measure program participation by having received at least one stipend.

2.2.2 Experimental design

Beneficiaries for the 2010 edition of Protejo in Rio de Janeiro were selected via lottery assignment. Potential beneficiaries were identified by trained community leaders and invited to apply to the program. The application form is our main source of baseline data, and randomization blocks were built based on the information gathered at this point. The randomizations took place in October 2010 in 19 communities, within strata defined by gender, socioeconomic vulnerability[1] and educational level[2], for a total of 8 blocks in each community and 152 strata overall.

In 24 of the 152 predefined strata, there was no excess demand, meaning that the number of interested applicants was smaller than the number of vacancies and thus no lottery was conducted. The remaining 128 subgroups had a total of 4,368 applicants, of which 1,941 were randomized into the treatment group. Our complete sample structure is described in Table 1.

In most communities, around 50% of applicants were randomly selected to participate in the program, with few exception being caused by variations in demand, since the number of vacancies was predetermined – such as in Rocinha, with approximately 28% of treated individuals, or Cantagalo and São João do Meriti, with over 67%. The average strata in our evaluation sample

[1] Individuals were classified as "low" or "high" vulnerability based on the following characteristics: i) housing conditions; ii) whether parents were alive; iii) whether the individual had been a victim of violence; iv) whether the individual had been in contact with drugs and v) whether the individual had been indicated to the program by social assistance programs or institutions.

[2] Having low educational level meant having at most completed primary education, while a high educational level was defined as having at least started high school.

Table 1 – Sample structure

Community	Total sample size	N. treated	% treated	Number of strata	Average strata size
Belford Roxo	277	92	33.2	6	46.2
Cantagalo	46	31	67.4	4	11.5
Caxias	144	73	50.7	5	28.8
Complexo do Alemão	165	66	40	4	41.3
Itaboraí	338	150	44.4	8	42.3
Itaguaí	302	150	49.7	8	37.8
Manguinhos	262	124	47.3	6	43.7
Mesquita	325	149	45.8	8	40.6
Nilópolis	285	112	39.3	7	40.7
Niterói	211	100	47.4	8	26.4
Nova Iguaçu	388	150	38.7	8	48.5
Pavão-Pavãozinho	133	75	56.4	8	16.6
Providência	110	59	53.6	5	22
Queimados	276	147	53.3	8	34.5
Rocinha	267	75	28.1	8	33.4
São Gonçalo	414	150	36.2	8	51.8
São João do Meriti	162	110	67.9	7	23.1
Tavares Bastos	75	37	49.3	6	12.5
Vila Kennedy	188	91	48.4	6	31.3
Total	4368	1941	44.4	128	34.1

Note – Each row represents a community where the randomization took place. The total sample size is equal to the total number of individuals in both treatment arms for each lottery draw. N and % treated refer to the absolute and relative number of individuals randomly allocated into the treatment group, respectively. Number of strata is the total number of randomization blocks for each community, and the average strata size is equal to the number of individuals in that row divided by the number of strata.

is composed of roughly 34 individuals.

2.3 Data

2.3.1 Labor market administrative data

We observe whether each individual in our sample is formally employed from 2010 to 2017 using an administrative data set organized annually by the Brazilian Ministry of Labor, the "Relação Anual de Informações Sociais" (RAIS), which gathers the entirety of formal employment records in the country. This same data source spanning 2010 to 2016 has been previously used by Barros et al. (2019), who linked Protejo beneficiaries to RAIS records in two steps: (i) using an unique individual identifier (CPF) whenever available; and (ii) linking records probabilistically based on full name and date of birth. The matching procedure is described in detail in Barros et al. (2019).

We expand on the previous analysis by including data for formal employment in 2017, as well as by dropping dubious matches from the probabilistical linkage. Specifically, we restrict our non-CPF matches to those who fall in the plausible age range of 10 to 30 years of age in the employment records[3]. This results in 122 fewer matches, 120 of which were reported to be older than 30. In total, 3,034 of the 4,368 individuals in our sample are found to be formally employed at some point during the 2010-2017 period.

2.3.2 Records of deaths

We observe the occurrence of deaths among Protejo applicants in our sample by entering individual identification information into Rio de Janeiro's Court of Justice online portal, which allows anyone to consult records of births and deaths in the state of Rio de Janeiro[4]. This consult was conducted at the end of the first semester of 2019, thus we observe deaths that occurred prior to that moment. We were able to certainly identify 22 individuals in our sample who had died up until the end of April 2019, 12 males and 10 females.

Further, we link these records of deaths with administrative data from the Ministry of Health covering all deaths in Brazil (*Sistema de Informação sobre Mortalidade* – SIM/DATASUS). This linkage is completed by comparing dates and municipalities, and we were able to recover the cause of death for all 22 occurrences in our sample, which we use to categorize deaths by violent causes, unspecified causes or disease-related causes, explained in further detail in section 2.5.

2.3.3 Social protection administrative data

The Bolsa Família Program (PBF) was one of the largest conditional cash transfer programs in the world. It was created in 2003 and ended in November 2021, being replaced by a similar initiative titled Auxílio Brasil. Starting in 2013, records of PBF beneficiaries were publicized on a monthly basis by the Brazilian government in its "Portal da Transparência"[5]. These records included the full name of the beneficiary, the municipality of residence, the amount received from the program, and two unique individual identifiers: (i) a partially censored CPF number for 95% of the records, with only 6 out of the 11 available; and (ii) the NIS, used mainly to identify Brazilian citizens in social programs. Further, by entering individuals NIS numbers in the publicly available payment system used to manage PBF transfers (Sistema de Benefícios ao Cidadão, SIBEC), we managed to recover birth dates for approximately 50% of PBF beneficiaries from 2013 to 2019.

We proceeded with a linkage algorithm similar to the one used to link Protejo applicants to formal employment records. First, we matched records with identical names, censored CPFs

[3] The age range of our sample spans from 11 to 25 years of age, with 95.9% of applicants between 15 and 23. We allow for some measurement error by including 10 to 30 year olds.

[4] Available at http://www4.tjrj.jus.br/Portal-Extrajudicial/CNO/http://www4.tjrj.jus.br/Portal-Extrajudicial/CNO/.

[5] Available at: https://www.portaltransparencia.gov.br/.

and birth dates. Next, we matched records with identical names and CPFs, but with no available birth dates, and records with identical names and birth dates but no available CPF. Finally, we performed a probabilistical linkage by measuring the difference between strings with the Levenshtein distance for names and birth dates, but keeping only matches with identical censored CPFs.

In total, 930 of the 4,368 Protejo applicants were found to have received a Bolsa Família transfer at some point from 2013 to 2021. Importantly, though, we only observe PBF "main" beneficiaries, who are mostly women, which explains the 94.2% of females among the 930 matched individuals.

Additionally to Bolsa Família, we also use records for the Auxílio Emergencial Program (PAE), an emergency cash transfer initiative by the Brazilian government designed to address the socioeconomic consequences of the Covid-19 pandemic, starting in April 2020. Once again, we use full names and partially censored CPF numbers to match PAE beneficiaries to Protejo applicants, following the same procedure outlined beforehand, for a total of 1,706 matches, equivalent to 39.1% of the 4,368 individuals in our lottery sample having benefitted from this cash transfer program at some point during 2020.

2.3.4 Primary data

In addition to administrative records related to formal labor market employment and to cash transfer programs, we also use two rounds of primary data collected at the individual level. First, all individuals who applied to participate in the lottery filled an application form in October 2010, which included information on baseline socioeconomic characteristics, such as family income, household size, educational level, relationship status and parenthood. These variables are available for all individuals in the sample, except for a handful of missing responses.

We use baseline information in three ways: (i) to assess balancedness between treatment and control groups; (ii) to conduct heterogeneity analysis by relevant observable characteristics; and (iii) as independent variables, mainly to control for unbalanced characteristics whenever applicable.

Next, approximately two years following the end of the planned 800-hour program, a follow-up survey was conducted with a subsample of lottery participants. This data collection round aimed at measuring much of the same variables observed at baseline, in addition to dimensions such as social interactions, noncognitive skills and community engagement. A total of 1,775 individuals were interviewed, out of an attempted 2,244 subsample of individuals, an 20.9% attrition rate. Barros et al. (2019) tested for differential attrition and showed that, overall, although treated individuals seemed to be more likely to be interviewed, these differences were not statistically significant.

The follow-up survey of 2013 was previously used by Barros et al. (2019) to assess program impacts on the relevant measures of outcomes collected. We make use of the same source of information to expand on their analyses. We do this by considering the possibility of heterogeneity in the results both by educational level and parenthood status[6], and by exploring previously-unused variables related to family formation, both as main outcomes and as potential mediators of effects leading Protejo participants to or from social protection programs.

2.3.5 Balancedness

In all our analyses, we consider female and male groups separately. Balancedness of observable characteristics in each sample can be tested by estimating the following:

$$X_{is} = \alpha + \tau W_{is} + \mu_s + \epsilon_{is} \tag{2.1}$$

where X_{is} is an observed variable of individual i in strata s; W_i is the treatment indicator; μ_s is the fixed-effect of strata s; and ϵ_{is} is a random error term.

Table 2 – Balancedness tests

Variable	Female				Male			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	2783	17.39	17.50	0.44	1568	16.76	16.67	0.22
Household size	2603	4.73	5.66	0.14	1430	5.47	4.75	0.40
Pardo or black	2797	0.74	0.73	0.71	1571	0.72	0.71	0.93
Has a child	2726	0.25	0.19	0.53	1532	0.07	0.05	0.37
Attended school	2775	0.73	0.76	0.79	1560	0.82	0.86	0.67
Satisfied with current educational level	2762	0.55	0.59	0.91	1552	0.61	0.62	0.48
Mother was alive	2789	0.93	0.96	0.47	1562	0.95	0.96	0.54
Mother is literate	2647	0.87	0.89	0.65	1457	0.90	0.92	0.55
Worked in August 2010	2718	0.12	0.11	0.46	1530	0.22	0.18	0.40
Self-esteem indicator	2789	0.58	0.64	0.53	1569	0.61	0.63	0.42
Impulsivity indicator	2790	0.52	0.47	0.28	1567	0.48	0.45	0.38
Victim of violence in the last year	2697	0.17	0.08	0.18	1529	0.14	0.09	0.13
Ever had contact with drugs (including alcohol)	2797	0.39	0.35	0.52	1571	0.42	0.38	0.93
Joint F test				0.67				0.38

Note – Balancedness tests for the female and male samples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

Table 2 shows that the lottery appears to have been successful in creating comparable treatment and control groups, for both the female and male subsamples. P-values for differences in means, controlling for strata fixed effects, show that any discrepancies between treatment and

[6] Because only 96 out of 1,571 males reported having children at the baseline, we explore differences by parenthood status only for females.

control groups are not statistically significant. Our set of observables also seems to be jointly balanced between groups, as per the joint F test.

2.4 Empirical Strategy

We estimate the Intention-to-Treat effect of Protejo on our outcomes of interest with the following preferred specification :

$$y_{is} = \alpha + \tau_{\text{ITT}} W_{is} + \mu_s + \epsilon_{is} \tag{2.2}$$

where y_{is} is some outcome of individual i in strata s; W_i is the treatment indicator, equal to 1 for individuals selected to participate by lottery assignment; μ_s is the fixed-effect of strata s; and ϵ_{is} is a random error term. The parameter of interest in equation 2.2 is denoted by τ_{ITT}, representing the average effect of being offered to participate in Protejo.

Heteroskedasticity-robust standard errors are estimated because randomization took place at the individual level, as preconized by Abadie et al. (2017). Additionally, we also estimate Fisher's p-value via a randomization inference procedure, with the ITT parameter as a test statistic, as described e.g. by Imbens and Rubin (2015), which tests for the sharp null hypothesis of no effect for all observations in the evaluation sample.

2.5 Results

2.5.1 Program participation

We observe whether Protejo applicants participated in the program with administrative records of monthly stipends paid to those who regularly attended classes. We consider that an individual participated if they are recorded to have received at least one stipend. Using this indicator as a dependent variable, we estimate how much lottery assignment was able to influence actual participation, as shown in Table 3. We consider the full sample of randomized individuals in column 1, as well as two sets of subsamples in columns 2-5, one based on educational level at baseline, and one based on parenthood status at baseline for females.

Although compliance was not perfect, we find that randomized assignment was successful in inducing participation among the offer-group. As presented in column 1, contamination in the control group was higher among males, with 34.7% of participation, compared to 27.5% among females. Because participation in the treatment group was similar between both groups, the point estimate for males is slightly smaller at 18.1 percentage points (p.p.). Interestingly, while females of different educational backgrounds and parenthood status show relatively similar participation, variation is much higher comparing males of low or high education, with take-up in the control group jumping from 28.6 to 40.2%, while participation among the treated is still similar in magnitude.

Table 3 – Intention-to-Treat estimates on program participation, by gender and subsample

	Full sample	Educational level		Parenthood status	
		Low	High	Childless	Mother
	(1)	(2)	(3)	(4)	(5)
Panel A: Females					
Lottery dummy	0.2652	0.2442	0.2917	0.2887	0.2108
	(0.021)	(0.029)	(0.032)	(0.025)	(0.045)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Control mean	0.2746	0.2894	0.2647	0.2857	0.2242
Fisher p-value	0.000	0.000	0.000	0.000	0.000
N. obs.	2797	1378	1419	2227	570
Panel B: Males					
Lottery dummy	0.181	0.2245	0.1331		
	(0.027)	(0.036)	(0.041)		
	[0.000]	[0.000]	[0.001]		
Control mean	0.3465	0.2862	0.4019		
Fisher p-value	0.000	0.000	0.002		
N. obs.	1571	887	684		

Note – Regressions of a treatment compliance indicator on the lottery dummy. Column 1 refers to our full registry data. Columns 2 and 3 refer to subsamples of low (no secondary education) and high (at least incomplete secondary education) educational level at baseline. Columns 4 and 5 refer to subsamples based on the parenthood status of females at the baseline. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

2.5.2 Incidence of deaths

Because Protejo's overarching goal was related to crime prevention and protecting at-risk youth, we analyze to which extent the program was able to affect mortality among its participants. In this case, we proceed with a negative binomial regression of an indicator of death on the lottery dummy and strata fixed effects, because of the small number of occurrences in our sample (22 in total among the 4,368 applicants). Incidence rate ratio (IRR) estimates are presented in Table 4.

Surprisingly, we estimate a positive IRR for deaths of any cause (column 1), at a magnitude of 2.72, or a 172% increase in the incidence rate of the treatment group compared to the control. This result is both unexpected and seemingly contrary to the program's objectives. Because part of the workload of Protejo included lessons on subjects such as citizenship building and youth rights, one possible interpretation could be that treated individuals might have started standing up for themselves when faced with injustice, leading to undesirable circumstances within a dangerous environment.

Table 4 – Incidence rate ratio estimates on occurrence of deaths

	Death	Death by causes			
		Violent	Unspecified	Violent or unspecified	Disease-related
	(1)	(2)	(3)	(4)	(5)
Lottery dummy	2.7223	2.3239	0.6948	1.607	4.2827
	(0.527)	(0.844)	(1.295)	(0.74)	(0.689)
	[0.058]	[0.318]	[0.779]	[0.521]	[0.035]
Control mean	0.0025	0.0008	0.0004	0.0012	0.0012
N. obs.	4368	4368	4368	4368	4368

Note – Incidence rate ratio estimates for death (column 1) and death by causes (columns 2-5), estimated with negative binomial regressions controlling for strata fixed effects. Violent deaths in our sample are those whose causes fit one of the following categories: firearm aggression, aggression by cutting/perforating objects, aggression by unspecified means, strangulation/suffocation, drowning or traffic accident. Unspecified deaths are: unspecified fact or events, exposition to unspecified factor leading to trauma, legal intervention, other undefined causes. Disease-related deaths are: pneumonia, epilepsy, heart attack, heart diseases, cancer, HIV-related, sepsis, eclampsia. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

To further understand this phenomenon, we also consider death causes, as recorded in the administrative data. Categorizing causes as "violent", "unspecified" and "disease-related", and also considering a category of "violent or unspecified" to account e.g. for the possibility of misreporting of violent deaths. However, contrary to our previous hypothesis, we find statistically increases only for disease-related deaths, which include e.g. heart conditions, cancer and pneumonia. The estimated coefficient is equivalent to a 328% increase in the incidence rate of disease-related deaths, statistically significant at the 5% level. It could be the case of a spurious result, specially given the small number of death occurrences in our data (11 disease-related deaths). Otherwise, we do not have a plausible explanation for this find.

2.5.3 Formal employment

We now focus on formal employment using administrative information from the RAIS dataset, which comprises yearly records of formal employment in the entire country. We replicate the estimations of Barros et al. (2019) including an additional year of data and with minor adjustments to the matching algorithm that links Protejo applicants to records of formal employment, as described beforehand. Estimates are shown in Table 5.

Overall and as expected, our results are similar to those in Barros et al. (2019). We estimate large gains in formal employment for males, but not for females. The likelihood of ever being employed in the formal market at any point in the post-treatment period increases by 11.2 p.p. for men, or 17.1% compared to the control mean of 65.2% (column 1, Panel B), but the

Table 5 – Intention-to-Treat estimates on formal employment, by gender

	Formally employed	Formally employed by year							
	2011-2017	2010	2011	2012	2013	2014	2015	2016	2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Females									
Lottery dummy	-0.0001	-0.0299	-0.0153	-0.0069	-0.025	-0.0143	0.0315	0.0177	0.0139
	(0.02)	(0.014)	(0.017)	(0.02)	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)
	[0.995]	[0.031]	[0.373]	[0.729]	[0.229]	[0.511]	[0.144]	[0.406]	[0.501]
Control mean	0.6914	0.1466	0.2241	0.3174	0.4185	0.4789	0.4465	0.4152	0.3548
Fisher p-value	1.00	0.047	0.396	0.723	0.237	0.54	0.157	0.409	0.519
N. obs.	2797	2797	2797	2797	2797	2797	2797	2797	2797
Panel B: Males									
Lottery dummy	0.1116	0.0165	0.0176	0.0828	0.0945	0.1068	0.0685	0.0639	0.0955
	(0.026)	(0.018)	(0.022)	(0.026)	(0.027)	(0.028)	(0.028)	(0.028)	(0.027)
	[0.000]	[0.355]	[0.419]	[0.001]	[0.000]	[0.000]	[0.015]	[0.024]	[0.001]
Control mean	0.6518	0.1155	0.198	0.2772	0.3746	0.4422	0.462	0.4175	0.3531
Fisher p-value	0.000	0.356	0.465	0.002	0.001	0.000	0.017	0.022	0.001
N. obs.	1571	1571	1571	1571	1571	1571	1571	1571	1571

Note – Regressions of yearly formal employment indicators on the lottery dummy. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

equivalent estimate for females is practically identical to zero.

Investigating employment records by year, we find that gains for treated males begin in 2012 – the first complete calendar year after the end of the intervention – and are persistent until the end of our available data in 2017, with roughly similar magnitudes throughout. On the other hand, while females are on average unaffected by the treatment, they see a decrease in formal employment in 2010 – the year when the intervention begins – of 3 p.p., or 20.4% compared to the control mean, statistically significant at the 5% level. (column 2, Panel A of Table 5).

These yearly results differ slightly from estimates by Barros et al. (2019), whose estimate for females in 2010 was marginally smaller in magnitude and not significant at the 10% level, and who found positive effects of Protejo on employment of women in 2015 and 2016, while our estimates are 50% smaller in magnitude and indistinguishable from zero. We interpret these differences as consistent with our efforts to improve on the matching algorithm used to link Protejo and RAIS records.

To further investigate how Protejo affects the employability of its beneficiaries, we split our samples based on educational level at baseline, which was one of the characteristics used to stratify lottery draws before the intervention. All observed characteristics are balanced between treatment and control groups for all four samples (available in Tables A.1 and A.2 in Appendix

A). Intention-to-Treat results on formal employment by educational level are presented in Table 6.

Table 6 – Intention-to-Treat estimates on formal employment, by educational level

	Formally employed	Formally employed by year							
	2011-2017	2010	2011	2012	2013	2014	2015	2016	2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A.1: Low education females									
Lottery dummy	-0.0196	-0.0451	-0.0292	-0.0431	-0.0621	-0.0517	-0.0022	-0.0184	0.0045
	(0.029)	(0.016)	(0.021)	(0.025)	(0.027)	(0.029)	(0.028)	(0.028)	(0.027)
	[0.492]	[0.005]	[0.157]	[0.083]	[0.02]	[0.074]	[0.938]	[0.509]	[0.866]
Control mean	0.6392	0.1111	0.1687	0.2538	0.3333	0.4307	0.3964	0.3827	0.3141
Fisher p-value	0.501	0.009	0.206	0.096	0.023	0.078	0.945	0.512	0.901
N. obs.	1378	1378	1378	1378	1378	1378	1378	1378	1378
Panel A.2: High education females									
Lottery dummy	0.0246	-0.0106	0.0023	0.039	0.022	0.033	0.0741	0.0634	0.0258
	(0.029)	(0.024)	(0.029)	(0.032)	(0.033)	(0.033)	(0.033)	(0.033)	(0.032)
	[0.397]	[0.655]	[0.936]	[0.222]	[0.501]	[0.316]	[0.025]	[0.054]	[0.426]
Control mean	0.7262	0.1703	0.261	0.3599	0.4753	0.511	0.4799	0.4368	0.3819
Fisher p-value	0.406	0.74	0.941	0.212	0.539	0.355	0.028	0.065	0.462
N. obs.	1419	1419	1419	1419	1419	1419	1419	1419	1419
Panel B.1: Low education males									
Lottery dummy	0.1291	-0.0214	0.0163	0.074	0.1192	0.1123	0.0626	0.0985	0.1038
	(0.037)	(0.022)	(0.026)	(0.032)	(0.036)	(0.038)	(0.039)	(0.038)	(0.037)
	[0.000]	[0.327]	[0.527]	[0.022]	[0.001]	[0.004]	[0.106]	[0.011]	[0.005]
Control mean	0.5931	0.1034	0.1345	0.1966	0.2828	0.3793	0.4172	0.3655	0.3207
Fisher p-value	0.000	0.398	0.548	0.027	0.002	0.005	0.112	0.014	0.007
N. obs.	887	887	887	887	887	887	887	887	887
Panel B.2: High education males									
Lottery dummy	0.0922	0.0583	0.0191	0.0926	0.0673	0.1008	0.075	0.0257	0.0864
	(0.036)	(0.029)	(0.036)	(0.041)	(0.041)	(0.04)	(0.041)	(0.041)	(0.041)
	[0.01]	[0.042]	[0.596]	[0.023]	[0.101]	[0.013]	[0.068]	[0.536]	[0.034]
Control mean	0.7057	0.1266	0.2563	0.3513	0.4589	0.5	0.5032	0.4652	0.3829
Fisher p-value	0.012	0.047	0.621	0.025	0.117	0.018	0.074	0.576	0.035
N. obs.	684	684	684	684	684	684	684	684	684

Note – Regressions of yearly formal employment indicators on the lottery dummy. Low and high education refer to subsamples with no secondary education and at least incomplete secondary education at baseline, respectively. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Although not statistically significant, point estimates for low and high education females are opposite in direction, negative (-2 p.p.) for women with no secondary education at the baseline and positive (2.5 p.p.) for women with at least incomplete secondary education (column 1). Results by year show the same trend overall, with even statistically significant decreases in formal employment of low education women in some years (2010 and 2012-2014), but increases for

high education women in 2015 and 2016.

For males, while positive effects in formal employment encompass all educational levels with similar trends over the years, the estimated coefficient among low education men is considerably higher, at 12.9 p.p. or 21.8% compared to the control mean, compared to 9.2 p.p. or 13.1% for the high education sample. This difference in magnitudes is more than sufficient for treated low education males to catch up with control (but not with treated) high education males.

Finally, we also investigate the possibility of heterogeneous effects in terms of parenthood status of females at baseline. This subgroup analysis was not accounted for in the study design, but could still provide us with important information to better understand the mechanisms through which Protejo acts on women, especially for outcomes related to family formation and welfare receipt. Treatment and control groups are balanced in observable characteristics (Tables A.3 in Appendix A). Results on formal employment by parenthood status of females are presented in Table 7, but we find no discernible differences other than women who already had children at the baseline not seeing a statistically significant decrease in formal employment during the year when the intervention began (i.e. 2010, column 2), while childless females did (column 2, Panel A).

Table 7 – Intention-to-Treat estimates on formal employment, by parenthood status (females)

	Formally employed	Formally employed by year							
	2011-2017	2010	2011	2012	2013	2014	2015	2016	2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Childless at baseline									
Lottery dummy	0.0132	-0.0302	-0.0272	-0.0024	-0.0189	-0.0118	0.0283	0.0143	0.0254
	(0.023)	(0.015)	(0.019)	(0.022)	(0.024)	(0.025)	(0.025)	(0.025)	(0.024)
	[0.572]	[0.046]	[0.154]	[0.914]	[0.428]	[0.635]	[0.255]	[0.564]	[0.289]
Control mean	0.6935	0.1355	0.2119	0.3119	0.4212	0.4775	0.4547	0.4313	0.3622
Fisher p-value	0.602	0.054	0.167	0.949	0.466	0.637	0.265	0.578	0.31
N. obs.	2227	2227	2227	2227	2227	2227	2227	2227	2227
Panel B: Mother at baseline									
Lottery dummy	-0.0427	-0.0152	0.0188	-0.0165	-0.055	-0.0157	0.0345	0.018	-0.021
	(0.046)	(0.035)	(0.042)	(0.046)	(0.047)	(0.048)	(0.046)	(0.045)	(0.043)
	[0.351]	[0.663]	[0.656]	[0.718]	[0.238]	[0.744]	[0.45]	[0.687]	[0.625]
Control mean	0.6818	0.197	0.2788	0.3424	0.4061	0.4848	0.4091	0.3424	0.3212
Fisher p-value	0.368	0.701	0.63	0.77	0.252	0.752	0.485	0.751	0.67
N. obs.	570	570	570	570	570	570	570	570	570

Note – Regressions of yearly formal employment indicators on the lottery dummy. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

2.5.4 Welfare receipt

Next, we estimate whether Protejo affected access to the Bolsa Família and the Auxílio Emergencial, two nationwide cash transfer programs, the former one of the largest of its kind in the world and the latter a large-scale response to the socioeconomic consequences of the covid-19 pandemic. Results by gender are presented in Table 8.

Table 8 – Intention-to-Treat estimates on welfare receipt, by gender

	Received Auxílio Emergencial	Ever received Bolsa Família	Received Bolsa Família by year								
			2013	2014	2015	2016	2017	2018	2019	2020	2021
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Females											
Lottery dummy	0.0445	0.0479	-0.0265	-0.0265	-0.008	0.0126	0.0191	0.0333	0.0357	0.0309	0.0232
	(0.022)	(0.021)	(0.014)	(0.015)	(0.017)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)
	[0.039]	[0.020]	[0.059]	[0.074]	[0.629]	[0.479]	[0.296]	[0.083]	[0.059]	[0.101]	[0.217]
Control mean	0.3893	0.279	0.1032	0.117	0.1483	0.168	0.1818	0.2054	0.1928	0.1999	0.2037
Fisher p-value	0.041	0.018	0.056	0.068	0.621	0.493	0.307	0.08	0.071	0.105	0.234
N. obs.	2797	2797	2797	2797	2797	2797	2797	2797	2797	2797	2797
Panel B: Males											
Lottery dummy	0.0429	-0.008	-0.0076	-0.0095	-0.0094	-0.0106	-0.0058	-0.0019	0.003	0.0063	0.0069
	(0.027)	(0.011)	(0.006)	(0.006)	(0.007)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
	[0.117]	[0.458]	[0.166]	[0.138]	[0.183]	[0.168]	[0.510]	[0.822]	[0.691]	[0.427]	[0.389]
Control mean	0.3284	0.0314	0.0099	0.0132	0.0149	0.0182	0.0198	0.0198	0.0132	0.0099	0.0099
Fisher p-value	0.126	0.521	0.161	0.153	0.18	0.149	0.523	0.841	0.825	0.492	0.377
N. obs.	1571	1571	1571	1571	1571	1571	1571	1571	1571	1571	1571

Note – Regressions of welfare receipt indicators on the lottery dummy. "Received Auxílio Emergencial" (column 1) is equal to 1 if the individual received the benefit in 2020. "Ever received Bolsa Família" (column 2) is equal to 1 if the individual received the benefit at some point from 2013 until 2021. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Our estimates in columns 1 and 2, Panel A, suggest that females in the offer-group have an increased chance of being beneficiaries of both the Auxílio Emergencial – an estimated increase of 4.4 p.p., or 11.4% compared to the control mean in the 2020 program – and the Bolsa Família programs – at 4.8 p.p., or 17.2% –, statistically significant at the 5% level. Assessing effects on Bolsa Família receipt by year (columns 3-11) shows interesting dynamics, with an initial decrease in 2013 and 2014 (i.e., two years after the intervention), which fades out after 2015 and is reversed into a positive effect as of 2018. This initial decrease in Bolsa Família receipt that is reversed in the medium to long term would be consistent with an initial effect of Protejo on informal employment and/or earnings, which we do not observe with administrative records from the RAIS dataset. However, self-reported labor market outcomes in the 2013 survey, including measures of informality, yield no results to support this hypothesis[7].

[7] ITT estimates on self-reported labor market outcomes were assessed by Barros et al. (2019), who find no statistically significant effects on any of the considered measures. We replicate their findings for all of our subsamples and estimate similarly insignificant results (not shown).

As for males, we do not observe any differences in welfare receipt between treatment and control groups overall (Panel B). Small proportions of Bolsa Família receipt (only 3.14% among males in the control group from 2013-2021) support our previous explanation on its "main" beneficiaries being majoritarily females, and thus our results do not shed light on whether Protejo affects the likelihood of males being part of a family benefited by PBF. However, our estimates are still valid for the likelihood of being a main beneficiary, which does not seem to be affected in this case. Point estimates are negative from 2013 to 2018, consistent with previously-reported gains in formal employment, although they are indistinguishable from zero. Estimates on Auxílio Emergencial receipt are also not statistically significant at the 10% level, although p-values are only slightly bigger than such threshold.

Considering individuals from different educational backgrounds, results suggest that any increases in welfare receipt among women are concentrated among those with lower educational attainment at the baseline, as shown in Panels A.1 and A.2 in Table 9. Bolsa Família receipt among low education females is increased by 5.4 p.p., or 16.8% compared to the control mean, and Auxílio Emergencial receipt is increased by 9.6 p.p. (23.6%), while estimates for high educational females are statistically insignificant. For males, Auxílio Emergencial and overall Bolsa Família (columns 1 and 2) are unaffected, but we do estimate significant increases in Bolsa Família receipt among high education men in 2020 and 2021.

Finally, we reestimate results for females after splitting the sample based on baseline parenthood status, as shown in Table 10. Increases in welfare receipt are observed only for women who did not report having any children before the intervention (Panel A). We estimate increases of 6.01 p.p. (16.4%) in Auxílio Emergencial and of 6.8 p.p. (30.2%) in Bolsa Família receipt among such group, but negative and statistically insignificant coefficients among those who had children at baseline. These results can be partly explained by considerably higher averages in welfare receipt among women who were already mothers initially, meaning that this group can be expected to enter welfare programs regardless of Protejo participation, even if they are not more or less likely to be formally employed (as per our previous results). Still, the fact that we find significant differences between treatment and control groups only among initially-childless females is intriguing and demands further investigation.

2.5.5 Family formation

One mechanism through which Protejo could act on its beneficiaries towards welfare receipt is family formation. Barros et al. (2019) previously found that the program increased by 3.5 p.p. (or 12%) the likelihood of having children or being expecting among the entire follow-up sample (both males and females). They argue that the program could have fostered family relations and family building among treated individuals, leading to emotional stability and consequently increased fertility. It could also be a direct result of the program's 100-hour module

Table 9 – Intention-to-Treat estimates on welfare receipt, by educational level

	Received Auxílio Emergencial	Ever received Bolsa Família	Received Bolsa Família by year								
			2013	2014	2015	2016	2017	2018	2019	2020	2021
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A.1: Low education females											
Lottery dummy	0.0955	0.0536	-0.0388	-0.026	0.0001	0.0156	0.0317	0.0304	0.0317	0.0421	0.0385
	(0.029)	(0.028)	(0.02)	(0.021)	(0.024)	(0.025)	(0.026)	(0.027)	(0.026)	(0.027)	(0.027)
	[0.001]	[0.057]	[0.05]	[0.223]	[0.996]	[0.532]	[0.22]	[0.252]	[0.227]	[0.114]	[0.148]
Control mean	0.4047	0.3196	0.1262	0.144	0.1797	0.2085	0.2209	0.2442	0.2359	0.2428	0.2442
Fisher p-value	0.001	0.052	0.055	0.251	1.000	0.557	0.213	0.272	0.227	0.114	0.163
N. obs.	1378	1378	1378	1378	1378	1378	1378	1378	1378	1378	1378
Panel A.2: High education females											
Lottery dummy	-0.0201	0.0407	-0.0109	-0.0271	-0.0183	0.0088	0.0032	0.0369	0.0408	0.0167	0.0039
	(0.032)	(0.03)	(0.02)	(0.02)	(0.023)	(0.025)	(0.025)	(0.028)	(0.027)	(0.026)	(0.026)
	[0.523]	[0.179]	[0.578]	[0.177]	[0.425]	[0.726]	[0.90]	[0.183]	[0.132]	[0.525]	[0.882]
Control mean	0.3791	0.2518	0.0879	0.0989	0.1273	0.141	0.1557	0.1795	0.1639	0.1712	0.1767
Fisher p-value	0.532	0.175	0.651	0.189	0.451	0.768	0.932	0.174	0.114	0.542	0.935
N. obs.	1419	1419	1419	1419	1419	1419	1419	1419	1419	1419	1419
Panel B.1: Low education males											
Lottery dummy	0.0428	-0.0219	-0.0127	-0.0082	-0.0056	-0.009	-0.0113	-0.0105	-0.0034	-0.0065	-0.0076
	(0.039)	(0.016)	(0.009)	(0.009)	(0.01)	(0.012)	(0.014)	(0.014)	(0.012)	(0.012)	(0.012)
	[0.267]	[0.171]	[0.145]	[0.375]	[0.593]	[0.45]	[0.408]	[0.444]	[0.775]	[0.601]	[0.541]
Control mean	0.3655	0.0448	0.0172	0.0172	0.0172	0.0241	0.031	0.031	0.0207	0.0172	0.0172
Fisher p-value	0.3	0.139	0.123	0.489	0.742	0.583	0.45	0.466	0.784	0.53	0.51
N. obs.	887	887	887	887	887	887	887	887	887	887	887
Panel B.2: High education males											
Lottery dummy	0.0431	0.0073	-0.002	-0.011	-0.0135	-0.0123	0.0002	0.0076	0.0101	0.0204	0.023
	(0.039)	(0.014)	(0.006)	(0.009)	(0.009)	(0.009)	(0.011)	(0.01)	(0.009)	(0.009)	(0.01)
	[0.269]	[0.607]	[0.755]	[0.214]	[0.142]	[0.188]	[0.985]	[0.449]	[0.282]	[0.028]	[0.017]
Control mean	0.2943	0.019	0.0032	0.0095	0.0127	0.0127	0.0095	0.0095	0.0063	0.0032	0.0032
Fisher p-value	0.28	0.611	1.000	0.111	0.124	0.166	1.000	0.522	0.461	0.072	0.034
N. obs.	684	684	684	684	684	684	684	684	684	684	684

Note – Regressions of welfare receipt indicators on the lottery dummy. "Received Auxílio Emergencial" (column 1) is equal to 1 if the individual received the benefit in 2020. "Ever received Bolsa Família" (column 2) is equal to 1 if the individual received the benefit at some point from 2013 until 2021. Low and high education refer to subsamples with no secondary education and at least incomplete secondary education at baseline, respectively. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table 10 – Intention-to-Treat estimates on welfare receipt, by parenthood status (females)

	Received Auxílio Emergencial	Ever received Bolsa Família	Received Bolsa Família by year								
			2013	2014	2015	2016	2017	2018	2019	2020	2021
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Childless at baseline											
Lottery dummy	0.0605	0.0677	-0.0135	-0.0131	0.001	0.026	0.0362	0.0455	0.0563	0.0499	0.0406
	(0.024)	(0.022)	(0.01)	(0.012)	(0.015)	(0.017)	(0.019)	(0.021)	(0.02)	(0.02)	(0.02)
	[0.014]	[0.003]	[0.196]	[0.267]	[0.945]	[0.136]	[0.055]	[0.027]	[0.006]	[0.015]	[0.046]
Control mean	0.3689	0.224	0.0516	0.0657	0.0952	0.1167	0.1328	0.1643	0.1549	0.163	0.1663
Fisher p-value	0.015	0.003	0.26	0.304	1.000	0.132	0.053	0.021	0.005	0.02	0.045
N. obs.	2227	2227	2227	2227	2227	2227	2227	2227	2227	2227	2227
Panel B: Mother at baseline											
Lottery dummy	-0.0041	-0.0323	-0.0692	-0.0714	-0.0403	-0.0253	-0.0359	-0.0091	-0.0334	-0.039	-0.0454
	(0.048)	(0.048)	(0.045)	(0.045)	(0.047)	(0.048)	(0.048)	(0.048)	(0.047)	(0.047)	(0.047)
	[0.932]	[0.498]	[0.128]	[0.114]	[0.395]	[0.599]	[0.456]	[0.849]	[0.479]	[0.409]	[0.337]
Control mean	0.4818	0.5273	0.3364	0.3485	0.3879	0.4	0.403	0.3909	0.3636	0.3667	0.3727
Fisher p-value	1.000	0.491	0.134	0.135	0.451	0.629	0.487	0.843	0.515	0.43	0.388
N. obs.	570	570	570	570	570	570	570	570	570	570	570

Note – Regressions of welfare receipt indicators on the lottery dummy. "Received Auxílio Emergencial" (column 1) is equal to 1 if the individual received the benefit in 2020. "Ever received Bolsa Família" (column 2) is equal to 1 if the individual received the benefit at some point from 2013 until 2021. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

on citizenship building, which included lessons on youth emancipation, family formation and sexuality. To investigate this matter more deeply, we focus on estimating results separately by gender on five previously-unexplored outcomes related to family formation: (i) being expecting (for females) or having a child up to two (incomplete) years of age[8], (ii) getting married after the baseline, (iii) being single at the follow-up (i.e. not married nor living together with a spouse), (iv) getting married after the baseline and having a child up to two years old ("married with children"), and (v) becoming a single parent (i.e. being single and having a 0-2 year old child).

Because data on family formation comes from the 2013 follow-up survey, we need to address concerns with differential attrition and balance of observed characteristics for the interviewed sample. In that occasion, we set out to collect data for a subsample of 2,244 Protejo applicants, of which 1,775 were successfully interviewed. Estimates for differential attrition between treatment and control groups (Table B.1 in Appendix B) show that treated females were less likely to be attriters by 5 p.p. (or 20.7%), especially among the low education group. We also observe slight unbalances in observable characteristics for females overall, low education males and both female subsamples by parenthood status (all available in Appendix C). We proceed by including our full set of variables listed in Table 2 as control variables in all regressions whose results are reported in this subsection[9].

Intention-to-Treat estimates on family formation outcomes by gender are reported in

[8] This threshold on age guarantees that the child was born after the intervention.

[9] Missing values are inputted with the strata average for continues variables and with zero for binary variables, and we create an additional control variable equal to 1 if an individual had any variables inputted.

Table 11. For females, the estimated coefficient for having children after the intervention (column 1) is equivalent to a 6.4 p.p. increase in fertility, or 33.7% compared to the control mean of 18.8%. Interestingly, we do not find any impacts on the likelihood of getting married or being single (columns 2 and 3), but the estimate on becoming a single parent (column 5) is statistically significant, corresponding to a 4.6 p.p. increase, or 62.4%, while the one related to getting married and having children (column 4) is not. For males, estimates are largely imprecise, although the coefficient for being married with children is positive and marginally significant.

Table 11 – Intention-to-Treat estimates on family formation outcomes, by gender

	Had children	Got married	Is single	Married with children	Single parent
	(1)	(2)	(3)	(4)	(5)
Panel A: Females					
Lottery dummy	0.0644	0.0116	-0.0115	0.0109	0.0457
	(0.025)	(0.025)	(0.027)	(0.018)	(0.018)
	[0.011]	[0.647]	[0.673]	[0.556]	[0.01]
Control mean	0.1876	0.2158	0.6792	0.0919	0.0732
Fisher p-value	0.001	0.557	0.646	0.598	0.028
N. obs.	0.021	0.682	0.695	0.608	0.022
	1101	1101	1101	1101	1101
Panel B: Males					
Lottery dummy	0.0155	0.0228	-0.0167	0.0244	-0.0069
	(0.02)	(0.022)	(0.023)	(0.015)	(0.014)
	[0.449]	[0.31]	[0.468]	[0.105]	[0.619]
Control mean	0.0731	0.076	0.9152	0.0263	0.0439
Fisher p-value	0.487	0.322	0.485	0.146	0.647
N. obs.	674	674	674	674	674

Note – Regressions of family formation indicators on the lottery dummy. "Had children" (column 1) is equal to 1 if the individual reports having a child of 0 to 2 years of age at the follow-up. "Got married" (column 2) is equal to 1 if the individual reports having gotten married or moved in with a spouse after the intervention. "Is single" (column 3) is equal to 1 if the individual reports not being married or living with a spouse at the follow-up. "Married with children" (column 4) is equal to 1 if the individual reports having gotten married after the intervention and having a child 0-2 at the follow-up. "Single parent" (column 5) is equal to 1 if the individual reports not being married or living with a spouse and having a child 0-2 at the follow-up. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Allowing for different results by educational level (estimates in Table 12) suggests that women with both low and high educational attainment at baseline have increased fertility at the follow-up, with similar point estimates, although the estimated coefficient for the low education group is more precise and is equivalent to a relatively smaller percentage effect because of a higher control group mean (30.3% versus 44.2%). As for the likelihood of becoming a single

parent, we only observe an increase for low education females, at 6.1 p.p. or an impressive 75.3% compared to the control mean. No statistically significant results are found for males of any educational background.

Finally, comparing ITT effects by parenthood status, we find that women who reported having no children at the baseline are the ones who experience an increased chance both of having children after the intervention and becoming single parents at the time of the follow-up, two years after the end of Protejo implementation. Results are available in Table 13. Estimated coefficients for women who were already mothers at the baseline are all insignificant (Panel B).

Our results on family formation are consistent with previously-shown estimates on welfare receipt. The same subgroups of women who experience increases in the chance of receiving cash transfers from the Bolsa Família and the Auxílio Emergencial programs – i.e., mainly females with no secondary education at the baseline and who reported not having any children at that time – are also the ones who are affected by Protejo towards having children and becoming single parents two years after the end of the intervention.

2.5.6 Mediation analyses

Protejo is a program that encompasses dimensions of technical training, mentoring, vocational education, soft skills and citizenship building, all in one package targeted at youth at-risk from the most vulnerable contexts in disadvantaged communities. The goals of the program were related to increasing social participation, protecting said youth and promoting education and work. As such, the fact that we estimate that the intervention actually increased welfare receipt among beneficiary women raises the question of what are the mechanism through which such effect might have been caused, as opposed to the gains in formal employment that we observe for males.

One potential such mechanism is family formation, as mentioned beforehand. We estimate increases in fertility and in the likelihood of becoming a single parent among the same subgroups of women who experience increased welfare receipt compared to the experimental control group. Having more children might directly lead to welfare programs because it mechanically reduces per capita family income, especially since females do not seem to benefit from the program in terms of labor market outcomes.

To address this argument more directly, we apply the framework proposed by Imai et al. (2010) to estimate the Average Causal Mediation Effect of Protejo on Bolsa Família receipt, as mediated by increased fertility. We do this under their so-called Sequential Ignorability (SI) hypothesis, which demands not only independence of treatment assignment from potential outcomes (satisfied because of the lottery in this study), but also independence between potential outcomes and mediator, conditional on treatment and pre-treatment covariates. In our context,

Table 12 – Intention-to-Treat estimates on family formation outcomes, by educational level

	Had children	Got married	Is single	Married with children	Single parent
	(1)	(2)	(3)	(4)	(5)
Panel A.1: Low education females					
Lottery dummy	0.0662	0.0224	-0.0047	0.0001	0.0612
	(0.034)	(0.034)	(0.036)	(0.026)	(0.024)
	[0.052]	[0.508]	[0.897]	[0.997]	[0.013]
Control mean	0.2188	0.2344	0.6438	0.1156	0.0813
Fisher p-value	0.068	0.548	0.901	0.997	0.022
N. obs.	671	671	671	671	671
Panel A.2: High education females					
Lottery dummy	0.0622	-0.0056	-0.0135	0.0249	0.0254
	(0.038)	(0.037)	(0.042)	(0.025)	(0.026)
	[0.10]	[0.88]	[0.745]	[0.314]	[0.331]
Control mean	0.1408	0.1878	0.7324	0.0563	0.061
Fisher p-value	0.158	0.9	0.766	0.391	0.399
N. obs.	430	430	430	430	430
Panel B.1: Low education males					
Lottery dummy	0.0101	0.0564	-0.0458	0.0232	-0.0131
	(0.031)	(0.035)	(0.036)	(0.024)	(0.022)
	[0.746]	[0.11]	[0.20]	[0.327]	[0.544]
Control mean	0.0944	0.0889	0.9	0.0333	0.0611
Fisher p-value	0.77	0.14	0.239	0.346	0.613
N. obs.	352	352	352	352	352
Panel B.2: High education males					
Lottery dummy	0.0121	-0.0093	0.0126	0.0226	-0.0031
	(0.026)	(0.027)	(0.028)	(0.02)	(0.017)
	[0.642]	[0.734]	[0.654]	[0.271]	[0.855]
Control mean	0.0494	0.0617	0.9321	0.0185	0.0247
Fisher p-value	0.651	0.737	0.662	0.298	0.85
N. obs.	322	322	322	322	322

Note – Regressions of family formation indicators on the lottery dummy. "Had children" (column 1) is equal to 1 if the individual reports having a child of 0 to 2 years of age at the follow-up. "Got married" (column 2) is equal to 1 if the individual reports having gotten married or moved in with a spouse after the intervention. "Is single" (column 3) is equal to 1 if the individual reports not being married or living with a spouse at the follow-up. "Married with children" (column 4) is equal to 1 if the individual reports having gotten married after the intervention and having a child 0-2 at the follow-up. "Single parent" (column 5) is equal to 1 if the individual reports not being married or living with a spouse and having a child 0-2 at the follow-up. Low and high education refer to subsamples with no secondary education and at least incomplete secondary education at baseline, respectively. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

Table 13 – Intention-to-Treat estimates on family formation outcomes, by parenthood status (females)

	Had children	Got married	Is single	Married with children	Single parent
	(1)	(2)	(3)	(4)	(5)
Panel A: Childless at baseline					
Lottery dummy	0.0695	0.0275	-0.0253	0.0167	0.0438
	(0.028)	(0.028)	(0.03)	(0.022)	(0.02)
	[0.014]	[0.334]	[0.397]	[0.44]	[0.027]
Control mean	0.1687	0.1886	0.7618	0.0918	0.067
Fisher p-value	0.025	0.381	0.454	0.481	0.043
N. obs.	834	834	834	834	834
Panel B: Mother at baseline					
Lottery dummy	0.0356	-0.0434	0.0503	-0.0081	0.0169
	(0.063)	(0.065)	(0.067)	(0.042)	(0.042)
	[0.57]	[0.507]	[0.456]	[0.847]	[0.691]
Control mean	0.2462	0.3	0.4231	0.0923	0.0923
Fisher p-value	0.57	0.513	0.48	0.846	0.712
N. obs.	267	267	267	267	267

Note – Regressions of family formation indicators on the lottery dummy. "Had children" (column 1) is equal to 1 if the individual reports having a child of 0 to 2 years of age at the follow-up. "Got married" (column 2) is equal to 1 if the individual reports having gotten married or moved in with a spouse after the intervention. "Is single" (column 3) is equal to 1 if the individual reports not being married or living with a spouse at the follow-up. "Married with children" (column 4) is equal to 1 if the individual reports having gotten married after the intervention and having a child 0-2 at the follow-up. "Single parent" (column 5) is equal to 1 if the individual reports not being married or living with a spouse and having a child 0-2 at the follow-up. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

the second part of the SI hypothesis means that the mediator (having children) must be as good as random for each of the treatment and control groups, i.e. the decision to have children must be exogenous to potential Bolsa Família receipt. The model to be estimated is equivalent to the following two equations:

$$\begin{aligned} M_{is} &= \alpha_1 + \beta_1 W_{is} + \mu_s + \varepsilon_{is} \\ y_{is} &= \alpha_2 + \beta_2 W_{is} + \gamma M_{is} + \mu_s + \epsilon_{is} \end{aligned} \tag{2.3}$$

where M_{is} is the mediator variable for individual i in strata s and everything else is as defined before. The total effect of Protejo on Bolsa Família receipt is then given by $\beta_1\gamma + \beta_2$, i.e., the sum of indirect effect through the mediator $\beta_1\gamma$ and the direct effect β_2, similar to the canonical

approach to mediation by Baron and Kenny (1986). In practice, parameters are estimated using the simulation approach proposed by Imai et al. (2010) to nearly identical results and we also include our full set of controls described in Table 2. Estimates of the mediated (indirect) and total effects, as well as of the proportion of the total effect of Protejo on PBF receipt estimated to be mediated by effects through fertility are presented in Table 14 for the three samples in which we found significant effects in the previous sections, with the accompanying 90% confidence interval[10].

Table 14 – Average Causal Mediation Effect estimates of Protejo on Bolsa Família receipt mediated by having children (females)

	Full sample (1)	Low educational level (2)	Childless at baseline (3)
Average Causal Mediation Effect	0.0096	0.0064	0.0112
	(0.0030, 0.01716)	(0.000062, 0.0152)	(0.0034, 0.0212)
Total effect	0.0544	0.0776	0.0897
	(0.0103, 0.0992)	(0.0183, 0.1405)	(0.0363, 0.1392)
% of total effect mediated	0.1728	0.0815	0.1249
	(0.0885, 0.6046)	(0.0423, 0.2653)	(0.0797, 0.3068)
N. obs.	1101	671	834

Note – Estimates of total and mediated (Average Causal Mediation Effect) ITT effects of Protejo on Bolsa Família receipt, mediated by having a child up to two years of age at the follow-up, estimated based on Imai et al. (2010). Regressions for female subsamples: complete (column 1), low education (column 2) and childless parenthood status (column 3). All regressions include strata fixed effects and our full set of controls described in Table 2. 90% confidence intervals in parentheses.

Results for all of the three subsamples are statistically significant. We estimate that out of a total effect of 5.4 p.p. of Protejo on Bolsa Família receipt for the full sample of females, 0.96 p.p. are mediated by indirect effects on fertility (column 1). However, while there is a statistically significant mediation, the estimated proportion of the total effect that is mediated is only 17.3%, indicating that Protejo is leading treated women to access Bolsa Família partly because of an indirect effect on the probability of having children, but not only through this mechanism. We find evidence of mediation for subsamples of low education females and females with no children at baseline as well, although mediation proportions are smaller.

One other conceivable mechanism for Protejo to lead treated individuals towards welfare programs is by increasing e.g. knowledge of citizens' rights or other skills that could help an individual navigate any government bureaucracy in the way of being a cash transfer beneficiary. However, while compelling, these dimensions are more difficult to measure and we do not have access to data that allows us to explore this argument.

[10] We focus on the confidence interval in this case as it is the preferred metric in Imai et al. (2010).

2.6 Conclusion

In this study, we evaluated the effects of a large-scale human capital intervention aimed at the most vulnerable populations from socioeconomically disadvantaged communities in Rio de Janeiro, Brazil. The Protejo program comprises a series of modules of training and activities encompassing general, technical and soft skills, with a goal of fostering education and work, protecting at-risk youth in low-income and violent settings and promoting social participation. Assignment to receive an offer to participate was randomized among applicants, allowing us to compare treatment and control groups to identify causal effects of the program on relevant outcomes.

Specifically, while several studies have focused on assessing labor market outcomes of training interventions to youth populations, including in developing countries, less is known about the extent to which such initiatives are effective in actually reducing vulnerability, for instance, as measured by benefiting from government programs such as cash transfers to alleviate poverty. Few examples of this type of evidence have found reductions in welfare receipt among specific subgroups of women.

Building on previous evidence of the effects of Protejo program on formal labor market employment, we first focus on the possibility of heterogeneous results both between subgroups of different educational backgrounds and of different parenthood status at the onset of the intervention. Our results are consistent with Protejo only benefiting males in terms of labor market outcomes, as previously shown by Barros et al. (2019). As noted in other works for labor market programs, because we only have access to formal employment records, these results do not necessarily mean an increase in overall employment, but could be caused by a shift from informality to formality.

Next, turning to welfare programs, we estimate that women in the Protejo offer-group are actually more likely to be beneficiaries of two cash transfer programs up to ten years after the end of the intervention. This effect seems particularly relevant for women with initial lower levels of education and who had no children at the beginning of the program. Similarly, these are the same subgroups for which we estimate an increased fertility rate and likelihood of becoming a single mother after two years.

These results seem to contradict the little evidence that exists about the effects of trainings on welfare dependency among women. However, one key difference between Protejo and other initiatives is a large curriculum aimed not only towards technical skills, but also to the promotion of social participation, for example, via lessons on youth rights and emancipation, the delivery of sports and cultural-related activities, and increasing knowledge of sexuality and family formation. Protejo also targeted a quite singular type of audience, even among vulnerable populations, one that specifically was under high risk of exposure to domestic and urban violence.

Overall, following our contributions to an understudied topic within a relatively large body of literature, one major take-away of our findings is related to design aspects of policy making, such as curriculum design and beneficiary selection, especially in programs aimed towards vulnerable populations in developing countries. Individuals react differently to similar interventions based on their characteristics, the experiences they have gone through and the context in which they are inserted, and these factors must be accounted for so as to generate the intended impacts.

APPENDIX A. Balancedness tests for subsamples with full registry data

Table A.1 – Balancedness tests, low educational level subsamples

Variable	Female				Male			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	1370	17.08	17.06	0.27	886	16.34	16.14	0.54
Household size	1272	4.92	6.70	0.18	817	5.19	4.86	0.45
Pardo or black	1378	0.76	0.75	0.42	887	0.72	0.74	0.39
Has a child	1338	0.28	0.26	0.60	864	0.07	0.06	0.51
Attended school	1356	0.73	0.71	0.14	876	0.82	0.87	0.97
Satisfied with current educational level	1357	0.54	0.55	0.37	874	0.59	0.60	0.48
Mother was alive	1371	0.92	0.95	0.27	880	0.95	0.97	0.38
Mother is literate	1289	0.85	0.87	0.29	811	0.89	0.90	0.54
Worked in August 2010	1334	0.13	0.12	0.40	859	0.23	0.22	0.98
Self-esteem indicator	1372	0.55	0.57	0.48	886	0.55	0.58	0.93
Impulsivity indicator	1373	0.53	0.52	0.90	885	0.50	0.51	0.13
Victim of violence in the last year	1324	0.19	0.11	0.37	868	0.16	0.10	0.22
Ever had contact with drugs (including alcohol)	1378	0.39	0.38	0.76	887	0.42	0.37	0.81
Joint F test				0.64				0.29

Note – Balancedness tests for the low education (no secondary education) female and male subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

Table A.2 – Balancedness tests, high educational level subsamples

Variable	Female				Male			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	1413	18.00	17.80	0.93	682	17.44	17.16	0.27
Household size	1331	4.36	4.98	0.17	613	5.93	4.64	0.67
Pardo or black	1419	0.72	0.72	0.73	684	0.71	0.69	0.46
Has a child	1388	0.21	0.14	0.08	668	0.08	0.03	0.05
Attended school	1419	0.72	0.79	0.18	684	0.82	0.86	0.53
Satisfied with current educational level	1405	0.57	0.62	0.24	678	0.64	0.64	0.78
Mother was alive	1418	0.96	0.97	0.72	682	0.95	0.95	0.14
Mother is literate	1358	0.91	0.91	0.47	646	0.93	0.93	0.85
Worked in August 2010	1384	0.12	0.11	0.89	671	0.20	0.15	0.18
Self-esteem indicator	1417	0.64	0.68	0.87	683	0.72	0.69	0.28
Impulsivity indicator	1417	0.50	0.43	0.13	682	0.43	0.39	0.84
Victim of violence in the last year	1373	0.12	0.06	0.30	661	0.12	0.08	0.36
Ever had contact with drugs (including alcohol)	1419	0.40	0.33	0.19	684	0.42	0.40	0.91
Joint F test				0.69				0.27

Note – Balancedness tests for the high education (at least incomplete secondary education) female and male subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

Table A.3 – Balancedness tests, parenthood status subsamples (females)

Variable	Childless at baseline				Mother at baseline			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	2217	16.64	16.95	0.34	566	19.69	19.98	0.76
Household size	2072	4.60	5.50	0.14	531	5.13	6.39	0.44
Pardo or black	2227	0.74	0.73	0.99	570	0.75	0.74	0.67
Has a child	2156	0.00	0.00	-	570	1.00	1.00	-
Attended school	2208	0.84	0.85	1.00	567	0.37	0.38	0.40
Satisfied with current educational level	2201	0.59	0.63	0.29	561	0.43	0.39	0.12
Mother was alive	2221	0.93	0.96	0.40	568	0.93	0.95	0.55
Mother is literate	2108	0.89	0.91	0.46	539	0.82	0.84	0.95
Worked in August 2010	2160	0.11	0.11	0.79	558	0.17	0.14	0.60
Self-esteem indicator	2220	0.60	0.66	0.55	569	0.53	0.56	0.93
Impulsivity indicator	2221	0.52	0.46	0.20	569	0.52	0.50	0.78
Victim of violence in the last year	2147	0.14	0.07	0.45	550	0.24	0.14	0.30
Ever had contact with drugs (including alcohol)	2227	0.35	0.31	0.34	570	0.53	0.53	0.67
Joint F test				0.48				0.68

Note – Balancedness tests for females by parenthood status subsamples. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

APPENDIX B. Differential attrition analyses, follow-up data

Table B.1 – Differential attrition analysis, by gender

	Full sample	Educational level		Parenthood status	
		Low	High	Childless	Mother
	(1)	(2)	(3)	(4)	(5)
Panel A: Females					
Lottery dummy	-0.0499	-0.0719	-0.0148	-0.0392	-0.056
	(0.021)	(0.028)	(0.033)	(0.025)	(0.046)
	[0.019]	[0.009]	[0.657]	[0.111]	[0.224]
Control mean	0.2407	0.2575	0.214	0.2338	0.2586
Fisher p-value	0.034	0.017	0.707	0.156	0.2585
N. obs.	1404	862	542	1054	342
Panel B: Males					
Lottery dummy	0.0203	0.0324	0.0061		
	(0.027)	(0.039)	(0.038)		
	[0.458]	[0.406]	[0.871]		
Control mean	0.1876	0.2035	0.1692		
Fisher p-value	0.4845	0.444	0.8865		
N. obs.	840	451	389		

Note – Regressions of an attrition indicator on the lottery dummy. Column 1 refers to our full registry data. Columns 2 and 3 refer to subsamples of low (no secondary education) and high (at least incomplete secondary education) educational level at baseline. Columns 4 and 5 refer to subsamples based on the parenthood status of females at the baseline. Heteroskedasticity-robust standard errors in parentheses and p-values in brackets. Fisher p-value refers to Fisher's randomization inference test, obtained from 2000 permutations. Control mean is the unconditional mean of the dependent variable among individuals randomized into the control group.

APPENDIX C. Balancedness tests for subsamples with follow-up data

Table C.1 – Balancedness tests, subsample with follow-up data

Variable	Female				Male			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	1096	17.28	17.60	0.04	673	16.69	16.50	0.19
Household size	1027	4.83	7.19	0.21	618	4.49	4.79	0.16
Pardo or black	1101	0.76	0.73	0.17	674	0.73	0.72	0.84
Has a child	1076	0.25	0.25	0.95	658	0.05	0.04	0.54
Attended school	1096	0.74	0.70	0.29	668	0.87	0.89	0.37
Satisfied with current educational level	1088	0.57	0.53	0.19	664	0.68	0.61	0.03
Mother was alive	1098	0.95	0.95	0.78	672	0.96	0.97	0.36
Mother is literate	1041	0.87	0.87	0.66	634	0.91	0.91	0.75
Worked in August 2010	1080	0.12	0.10	0.36	654	0.20	0.16	0.23
Self-esteem indicator	1098	0.58	0.62	0.09	674	0.66	0.64	0.68
Impulsivity indicator	1099	0.51	0.49	0.50	673	0.44	0.47	0.47
Victim of violence in the last year	1068	0.15	0.14	0.44	656	0.07	0.10	0.09
Ever had contact with drugs (including alcohol)	1101	0.40	0.36	0.10	674	0.37	0.38	0.79
Joint F test				0.09				0.16

Note – Balancedness tests for female and male subsamples interviewed at the follow-up. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

Table C.2 – Balancedness tests, low educational level subsamples with follow-up data

Variable	Female				Male			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	667	16.89	17.34	0.02	351	16.13	16.07	0.66
Household size	621	5.07	9.05	0.22	330	4.60	5.06	0.21
Pardo or black	671	0.77	0.72	0.10	352	0.74	0.75	0.80
Has a child	656	0.27	0.31	0.22	342	0.04	0.06	0.45
Attended school	666	0.75	0.67	0.04	346	0.88	0.89	0.79
Satisfied with current educational level	661	0.57	0.52	0.19	346	0.70	0.60	0.04
Mother was alive	669	0.94	0.94	0.86	351	0.95	0.98	0.14
Mother is literate	633	0.85	0.84	0.65	327	0.90	0.88	0.61
Worked in August 2010	660	0.12	0.12	0.79	339	0.24	0.19	0.21
Self-esteem indicator	669	0.56	0.56	0.81	352	0.57	0.59	0.60
Impulsivity indicator	670	0.52	0.54	0.47	352	0.48	0.53	0.26
Victim of violence in the last year	652	0.16	0.17	1.00	345	0.06	0.11	0.11
Ever had contact with drugs (including alcohol)	671	0.40	0.38	0.52	352	0.29	0.37	0.16
Joint F test				0.18				0.02

Note – Balancedness tests for the low education (no secondary education) female and male subsamples interviewed at the follow-up. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

Table C.3 – Balancedness tests, high educational level subsamples with follow-up data

Variable	Female				Male			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	429	17.90	17.99	0.73	322	17.29	16.98	0.17
Household size	406	4.45	4.45	0.90	288	4.37	4.47	0.52
Pardo or black	430	0.74	0.74	0.86	322	0.71	0.69	0.59
Has a child	420	0.22	0.15	0.09	316	0.06	0.03	0.09
Attended school	430	0.71	0.76	0.34	322	0.85	0.89	0.32
Satisfied with current educational level	427	0.56	0.56	0.63	318	0.67	0.61	0.30
Mother was alive	429	0.96	0.97	0.81	321	0.96	0.96	0.87
Mother is literate	408	0.92	0.91	0.91	307	0.93	0.94	0.90
Worked in August 2010	420	0.11	0.08	0.24	315	0.15	0.14	0.72
Self-esteem indicator	429	0.62	0.72	0.02	322	0.75	0.69	0.29
Impulsivity indicator	429	0.49	0.41	0.06	321	0.39	0.40	0.98
Victim of violence in the last year	416	0.13	0.09	0.19	311	0.07	0.10	0.42
Ever had contact with drugs (including alcohol)	430	0.41	0.33	0.07	322	0.45	0.40	0.30
Joint F test				0.22				0.18

Note – Balancedness tests for the high education (at least incomplete secondary education) female and male subsamples interviewed at the follow-up. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

Table C.4 – Balancedness tests, parenthood status subsamples with follow-up data

Variable	Childless at baseline				Mother at baseline			
	N.	Treated	Control	p-value	N.	Treated	Control	p-value
Age	832	16.54	16.83	0.10	264	19.61	20.02	0.24
Household size	778	4.69	8.03	0.18	249	5.28	4.53	0.32
Pardo or black	834	0.76	0.74	0.53	267	0.75	0.68	0.42
Has a child	809	0.00	0.00	-	267	1.00	1.00	-
Attended school	829	0.85	0.82	0.60	267	0.39	0.34	0.42
Satisfied with current educational level	827	0.60	0.59	0.69	261	0.47	0.35	0.06
Mother was alive	831	0.94	0.96	0.34	267	0.96	0.93	0.36
Mother is literate	786	0.89	0.89	0.90	255	0.82	0.79	0.86
Worked in August 2010	816	0.11	0.10	0.62	264	0.16	0.11	0.34
Self-esteem indicator	832	0.59	0.66	0.04	266	0.55	0.51	0.41
Impulsivity indicator	833	0.50	0.48	0.62	266	0.53	0.52	0.63
Victim of violence in the last year	809	0.13	0.12	0.84	259	0.22	0.19	0.60
Ever had contact with drugs (including alcohol)	834	0.36	0.30	0.07	267	0.55	0.53	0.95
Joint F test				0.07				0.08

Note – Balancedness tests for females by parenthood status subsamples interviewed at the follow-up. Values in "Treated" and "Control" columns refer to the unconditional mean of that variable among individuals randomized into treatment and control groups, respectively. Each p-value is obtained from the heteroskedasticity-robust standard errors in a regression of that variable on the lottery dummy. The joint F test is obtained from a regression of the lottery dummy on the full set of variables.

References

ABADIE, A. et al. When should you adjust standard errors for clustering? *NBER Working Paper*, n. w24003, 2017.

ACEVEDO, P. et al. *Living up to expectations: How job training made women better off and men worse off*. [S.l.], 2017.

ALZÚA, M. L.; CRUCES, G.; LOPEZ, C. Long-run effects of youth training programs: Experimental evidence from argentina. *Economic Inquiry*, Wiley Online Library, v. 54, n. 4, p. 1839–1859, 2016.

ALZÚA, M. L.; CRUCES, G.; LÓPEZ, C. Youth training programs beyond employment. 2016. Available from Internet: <http://sedici.unlp.edu.ar/handle/10915/51200>.

ANDREW, A. et al. Effects of a scalable home-visiting intervention on child development in slums of urban india: evidence from a randomised controlled trial. *Journal of child psychology and psychiatry*, Wiley Online Library, v. 61, n. 6, p. 644–652, 2020.

ANDREW, A. et al. Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in colombia. *PLoS medicine*, Public Library of Science San Francisco, CA USA, v. 15, n. 4, p. e1002556, 2018.

APUNYO, R. et al. Interventions to increase youth employment: An evidence and gap map. *Campbell Systematic Reviews*, Wiley Online Library, v. 18, n. 1, p. e1216, 2022.

ARAUJO, M. C. et al. Home visiting at scale and child development. *Journal of Public Economics Plus*, Elsevier, v. 2, p. 100003, 2021.

ATTANASIO, O. et al. Vocational training for disadvantaged youth in colombia: A long-term follow-up. *American Economic Journal: Applied Economics*, v. 9, n. 2, p. 131–43, 2017.

ATTANASIO, O.; KUGLER, A.; MEGHIR, C. Subsidizing vocational training for disadvantaged youth in colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, American Economic Association, v. 3, n. 3, p. 188–220, 2011.

ATTANASIO, O. P. et al. 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*, British Medical Journal Publishing Group, v. 349, 2014.

BARON, R. M.; KENNY, D. A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, American Psychological Association, v. 51, n. 6, p. 1173, 1986.

BARROS, R. P. de et al. Rescuing at-risk youth: Experimental evidence from a human capital investments program in brazil. 2019.

BEHAGHEL, L. et al. Please call again: Correcting nonresponse bias in treatment effect models. *Review of Economics and Statistics*, The MIT Press, v. 97, n. 5, p. 1070–1080, 2015.

BOS, J. et al. Early childhood human capital formation at scale. *Available at SSRN 3906697*, 2021.

BRICKER, D. et al. Ages and stages questionnaire. *Baltimore, MD: Paul H. Brookes*, 1999.

CARD, D.; KLUVE, J.; WEBER, A. What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, Oxford University Press, v. 16, n. 3, p. 894–931, 2018.

CAVE, G. et al. Jobstart: Final report on a program for school dropouts. *MDRC*, 1993. Available from Internet: <https://www.mdrc.org/sites/default/files/full_416.pdf>.

CHAISEMARTIN, C. D.; BEHAGHEL, L. Estimating the effect of treatments allocated by randomized waiting lists. *Econometrica*, Wiley Online Library, v. 88, n. 4, p. 1453–1477, 2020.

CONTI, G.; HECKMAN, J. J.; PINTO, R. The effects of two influential early childhood interventions on health and healthy behaviour. *The Economic Journal*, Oxford University Press, v. 126, n. 596, p. F28–F65, 2016.

CUNHA, F.; HECKMAN, J. J. The technology of skill formation. *American Economic Review*, American Economic Association, v. 97, n. 2, p. 31–47, 2007.

CUNHA, F. et al. Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, Elsevier, v. 1, p. 697–812, 2006.

CUNHA, F.; HECKMAN, J. J.; SCHENNACH, S. M. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, Wiley Online Library, v. 78, n. 3, p. 883–931, 2010.

CURRIE, J.; ALMOND, D. Chapter 15 - human capital development before age five. In: CARD, D.; ASHENFELTER, O. (Ed.). Elsevier, 2011, (Handbook of Labor Economics, v. 4). p. 1315 – 1486. Available from Internet: <http://www.sciencedirect.com/science/article/pii/S0169721811024130>.

CURRIE, J.; THOMAS, D. Does head start make a difference? *American Economic Review*, American Economic Association, v. 85, n. 3, p. 341–64, 1995.

DEEB, A.; CHAISEMARTIN, C. de. *Clustering and External Validity in Randomized Controlled Trials*. arXiv, 2019. Available from Internet: <https://arxiv.org/abs/1912.01052>.

FILGUEIRAS, A. et al. Psychometric properties of the brazilian-adapted version of the ages and stages questionnaire in public child daycare centers. *Early human development*, Elsevier, v. 89, n. 8, p. 561–576, 2013.

GARCES, E.; THOMAS, D.; CURRIE, J. Longer-term effects of head start. *American economic review*, v. 92, n. 4, p. 999–1012, 2002.

GERTLER, P. et al. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, American Association for the Advancement of Science, v. 344, n. 6187, p. 998–1001, 2014.

GERTLER, P. et al. *Effect of the jamaica early childhood stimulation intervention on labor market outcomes at age 31*. [S.l.], 2021.

GRANTHAM-MCGREGOR, S. M. et al. Effects of early childhood supplementation with and without stimulation on later development in stunted jamaican children. *The American journal of clinical nutrition*, Oxford University Press, v. 66, n. 2, p. 247–253, 1997.

HECKMAN, J.; PINTO, R.; SAVELYEV, P. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes †. *American Economic Review*, v. 103, n. 6, p. 2052–2086, 2013. Available from Internet: <http://dx.doi.org/10.1257/aer.103.6.2052>.

HECKMAN, J. J. et al. Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, Wiley Online Library, v. 1, n. 1, p. 1–46, 2010.

IBARRARÁN, P. et al. Experimental evidence on the long-term effects of a youth training program. *ILR Review*, SAGE Publications Sage CA: Los Angeles, CA, v. 72, n. 1, p. 185–222, 2019.

IBARRARAN, P. et al. Life skills, employability and training for disadvantaged youth: Evidence from a randomized evaluation design. *IZA Journal of Labor & Development*, Springer, v. 3, n. 1, p. 10, 2014.

IMAI, K.; KEELE, L.; TINGLEY, D. A general approach to causal mediation analysis. *Psychological methods*, American Psychological Association, v. 15, n. 4, p. 309, 2010.

IMBENS, G. W.; ANGRIST, J. D. Identification and estimation of local average treatment effects. *Econometrica*, [Wiley, Econometric Society], v. 62, n. 2, p. 467–475, 1994. ISSN 00129682, 14680262. Available from Internet: <http://www.jstor.org/stable/2951620>.

IMBENS, G. W.; RUBIN, D. B. *Causal Inference in Statistics, Social, and Biomedical Sciences*. [S.l.]: Cambridge University Press, 2015.

JUNIOR, M. V. W.; RIBEIRO, F. G.; PAESE, L. H. Z. Early childhood home-based programmes and school violence: evidence from brazil. *Development in Practice*, Taylor & Francis, v. 32, n. 2, p. 133–143, 2022.

KUGLER, A. et al. Long-term educational consequences of vocational training in colombia impacts on young trainees and their relatives. *Journal of Human Resources*, University of Wisconsin Press, v. 57, n. 1, p. 178–216, 2022.

LEE, D. S. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, Wiley-Blackwell, v. 76, n. 3, p. 1071–1102, 2009.

LIMA, L. S. *Essays on human capital investments in Brazil*. Tese (Doutorado), 2019.

LOVIBOND, S. H.; LOVIBOND, P. F. *Manual for the depression anxiety stress scales*. [S.l.]: Psychology Foundation of Australia, 1996.

MAWN, L. et al. Are we failing young people not in employment, education or training (neets)? a systematic review and meta-analysis of re-engagement interventions. *Systematic reviews*, Springer, v. 6, n. 1, p. 1–17, 2017.

MCKENZIE, D. How effective are active labor market policies in developing countries? a critical review of recent evidence. *The World Bank Research Observer*, Oxford University Press, v. 32, n. 2, p. 127–154, 2017.

MCKENZIE, D.; ROBALINO, D. Jobs and the crisis. World Bank, Washington, DC, 2010.

PAXSON, C.; SCHADY, N. Cognitive development among young children in ecuador the roles of wealth, health, and parenting. *Journal of Human resources*, University of Wisconsin Press, v. 42, n. 1, p. 49–84, 2007.

RADNER, J. M. et al. Practical considerations for transitioning early childhood interventions to scale: lessons from the saving brains portfolio. *Annals of the New York Academy of Sciences*, Wiley Online Library, v. 1419, n. 1, p. 230–248, 2018.

RIBEIRO, F. G. et al. An empirical assessment of the healthy early childhood program in rio grande do sul state, brazil. *Cadernos de Saúde Pública*, SciELO Brasil, v. 34, 2018.

SANDERS, M. R. et al. Parenting and family adjustment scales (pafas): Validation of a brief parent-report measure for use in assessment of parenting skills and family relationships. *Child Psychiatry & Human Development*, Springer, v. 45, n. 3, p. 255–272, 2014.

SANTANA, L. Adaptação transcultural e validação da parenting and family adjustment scales (pafas)[transcultural adaptation and validation of the parenting and family adjustment scales](unpublished master's thesis). *Universidade Federal da Grande Dourados, Brasil*, 2018.

SHONKOFF, J. P. et al. The developing brain. In: *From neurons to neighborhoods: The science of early childhood development*. [S.l.]: National Academies Press (US), 2000.

SILVA, E. V. da et al. 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*, BMJ Specialist Journals, v. 7, n. 1, p. e007116, 2022.

SYLVIA, S. et al. From quantity to quality: Delivering a home-based parenting intervention through china's family planning cadres. *The Economic Journal*, Oxford University Press, v. 131, n. 635, p. 1365–1400, 2021.

TRIPNEY, J. et al. Technical and vocational education and training (tvet) interventions to improve the employability and employment of young people in low-and middle-income countries: a systematic review. *Campbell systematic reviews*, Wiley Online Library, v. 9, n. 1, p. 1–171, 2013.

VERCH, K. Primeira infância melhor: transformando a atenção aos primeiros anos de vida na américa latina: desafios e conquistas de uma política pública no sul do brasil. *IDB Monograph (Social Sector. Social Protection and Health Division); IDB-MG-548*, Inter-American Development Bank, 2017.

VIGNOLA, R. C. B.; TUCCI, A. M. Adaptation and validation of the depression, anxiety and stress scale (dass) to brazilian portuguese. *Journal of affective disorders*, Elsevier, v. 155, p. 104–109, 2014.

WALKER, S. P. et al. Cognitive, psychosocial, and behaviour gains at age 31 years from the jamaica early childhood stimulation trial. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 63, n. 6, p. 626–635, 2022.

WHO. *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*. [S.l.]: World Health Organization, 2006.