

DISCUSSION PAPER SERIES

IZA DP No. 14575

**The Lasting Effects of Early Childhood
Education on Promoting the Skills and
Social Mobility of Disadvantaged African
Americans**

Jorge Luis García
James J. Heckman
Victor Ronda

JULY 2021

DISCUSSION PAPER SERIES

IZA DP No. 14575

The Lasting Effects of Early Childhood Education on Promoting the Skills and Social Mobility of Disadvantaged African Americans

Jorge Luis García

Clemson University

James J. Heckman

University of Chicago and IZA

Victor Ronda

University of Chicago

JULY 2021

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Lasting Effects of Early Childhood Education on Promoting the Skills and Social Mobility of Disadvantaged African Americans*

This paper demonstrates multiple beneficial impacts of a program promoting intergenerational mobility for disadvantaged African-American children and their children. The program improves outcomes of the first-generation treatment group across the life cycle, which translates into better family environments for the second generation leading to positive intergenerational gains. There are long-lasting beneficial program effects on cognition through age 54, contradicting claims of fadeout that have dominated popular discussions of early childhood programs. Children of the first-generation treatment group have higher levels of education and employment, lower levels of criminal activity, and better health than children of the first-generation control group.

JEL Classification: J13, I28, C93, H43

Keywords: early childhood education, intergenerational mobility, racial inequality, social mobility

Corresponding author:

James J. Heckman
Center for the Economics of Human Development
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
USA
E-mail: jjh@uchicago.edu

* This research is supported by the Buffett Early Childhood Fund, the National Institutes of Health's Eunice Kennedy Shriver National Institute of Child Health and Human Development under award number R37HD065072, and the National Institute of Aging under award numbers R01AG042390 and R01AG053343. The authors thank the researchers of the HighScope Educational Research Foundation's Perry Preschool Project for access to study data and source materials. Years of partnership and collaboration with HighScope have made this work possible. We also thank Meera Mody and Alejandra Campos for research assistance. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health.

1. Introduction

This paper analyzes newly collected data on the original participants of the pioneering Perry Preschool Project (PPP) through age 54 and on their children into their mid-twenties. The program aimed to promote social mobility of disadvantaged African-American children. It has substantial intergenerational multiplier effects and improves social mobility on a number of dimensions. Gains in cognition are sustained through age 54, contradicting claims about cognitive fadeout in the treatment effects of early childhood programs. Enriched early childhood education programs are promising vehicles for promoting social mobility.

Perry is relevant today because it influences the design of current and proposed early childhood education programs. At least 30% of current Head Start programs are based on it (Elango et al., 2016). More than 10% of African-American children born in the 2010s would satisfy the eligibility criteria to participate in PPP (García et al., 2021).

It is well-documented that PPP improved life-cycle outcomes of its original participants through age 40 (e.g., Elango et al., 2016; Heckman et al., 2010b). This paper uses newly collected late midlife measures of skills and life-cycle panel data based on surveys and administrative criminal records to document that the benefits for the treatment-group participants led to better environments for their children. Children of the first generation of treatment participants are more likely than children of the first generation of control participants to grow up in stable two-parent households. Their parents have higher average earnings, lower engagement with the criminal-justice system, and better executive functioning (cognition), socio-emotional skills, and health. Improved home environments are the source of Perry's intergenerational benefits.

PPP did not directly treat the children of the original participants, but nonetheless, it generated positive intergenerational externalities. Children of treated parents are 17 percentage points less likely to have been suspended from school during K-12 education compared to children of control participants. They are also 11 percentage points more likely to be in

good health through young adulthood, 26 percentage points more likely to be employed, and 8 percentage points less likely to be divorced. Children of male treated participants are also 18 percentage points less likely to have been arrested through young adulthood compared to children of male control participants. These estimates are statistically significant and robust across multiple estimators and inferential procedures designed to address methodological challenges inherent to PPP and many other social experiments.

Little is known about the intergenerational impact of early childhood education. Barr and Gibbs (2019) and Rossin-Slater and Wüst (2020) are exceptions. The latter paper exploits differential timing in preschool availability in Denmark during the period 1933-1960 and studies its intergenerational impact on educational attainment at age 25. Barr and Gibbs (2019) study the intergenerational impact of Head Start programs available in the 1970s using a similar design. They study education and a handful of other outcomes, like teenage pregnancy and youth criminality.

Our study is new in its experimental design and detailed knowledge of the life-cycle outcomes of original participants and multiple outcomes of their children. We study intergenerational outcomes across the life cycle from early life (e.g., special education and school suspension) to young adulthood (e.g., employment and marriage). We analyze intergenerational impacts by gender of the original participants and of their children. We use the mediation framework in Heckman et al. (2013) to document that experimentally induced improvements in parenting and reductions in criminal activity of the original participants largely explain the beneficial impact of PPP on their children.¹

¹A companion paper, García et al. (2021), monetizes the treatment effects of PPP through age 54 and finds an annual rate of return of 8.9% after adjusting for the distortion generated by the taxes required to fund the program. García et al. (2021) focus on primarily cost-benefit analysis of the program and do not analyze any of the specific treatment effects reported here. Heckman et al. (2010b) and Heckman and Karapakula (2021) study the impact of PPP on its original participants through age 40. Both studies develop identification, estimation, and inference methods especially suited for tackling the challenges inherent to PPP's design and implementation. Heckman and Karapakula (2021) present additional, relevant material. We use some of the methods explicated in that paper. Heckman et al. (2010a) and Heckman et al. (2013) are related studies. Heckman et al. (2010a) provide estimates of the internal rate of return of PPP using extrapolations informed by original-participant data through age 40. Heckman et al. (2013) provide a mediation framework to document that the short-term impact of PPP on socio-emotional skills largely

This paper proceeds in the following way. Section 2 describes PPP and our data, and summarizes how PPP’s impact on its original participants improved the environments in which their children grew up. Section 3 discusses our methodology and main estimates of PPP’s intergenerational impacts. Section 4 documents and interprets heterogeneity in the intergenerational impacts by gender. Section 5 investigates the mechanisms generating our estimates. Section 6 concludes.

2. The Perry Preschool Project and Its Participants as Parents

2.1 Program Overview

The Perry Preschool Project was a high-quality early childhood education program.² Its curriculum was designed to foster development of cognitive and socio-emotional skills. Children were active learners who planned, executed, and reflected on activities guided by teachers. Children made choices and solved problems. Teachers provided feedback (Schweinhart et al., 1993). Participants lived in the catchment area served by the Perry Elementary School in Ypsilanti, Michigan. In-school surveys, referrals, and canvassing identified an initial pool of participants. Eligibility criteria based on IQ scores and socio-economic status were used to create a pool of 123 disadvantaged African-American children who were randomized into the program (treatment group) or not (control group). Treatment-group children received two years of 2.5-hour preschool sessions during weekdays starting at age three. They also received weekly teacher home visits during the two-year treatment period. Control-group children did not receive any treatment. There were no treatment substitutes available in the area where they lived.³

explains the long-term impacts in age-40 outcomes such as employment and crime. None of these studies use the intergenerational data that we exploit in this paper. Our paper replaces Heckman and Karapakula (2019a), an unpublished manuscript that presents a preliminary analysis of the intergenerational data that we used in this paper. That manuscript remains as a working paper and is not under review.

²The description of PPP in this section takes information from García et al. (2021). We also refer interested readers to Heckman and Karapakula (2019b, 2021) and Heckman et al. (2010b) for extensive details on PPP and its rounds of data collection and to Elango et al. (2016) and Kautz et al. (2014) for a broad discussion of PPP and its relationship with other influential early education and social programs.

³Barnett (1996) reports a total program cost per participant of 21,151 (2017 US dollars) over the two-year life of the program, which ranks PPP in the lower end among programs of its type regarding implementation

Weikart et al. (1978) report that every family that received an offer to participate in PPP accepted it. We thus estimate the average treatment effect for program eligibles. Participants were born in the 1960s. Heckman et al. (2010b) report that 15% of African-American females and 17% of African-American males satisfied its eligibility criteria at the time of its implementation. After participants were randomized, the status of a few participants was swapped. This reassignment potentially compromised the randomization protocol and resulted in an imbalance of baseline characteristics (see Panel a. of Table 1).⁴ We adjust point estimates for this compromise using the methods in Heckman and Karapakula (2019b, 2021).⁵

2.2 Age-54 Follow-up

Panel a. of Table 1 gives the sample size and baseline characteristics of the original PPP study. The first-generation participants were followed in multiple rounds of data collection through age 54. In this paper, we primarily use data from the age-54 follow-up, in which information on their adult children was collected. We supplement these data with baseline information and earlier follow-ups to form panel observations on earnings and crime. Panel b. of Table 1 provides the sample size of original participants in the age-54 follow-up: 83% of the 123 original participants were surveyed; 12% were not surveyed because they were deceased, and 5% were not surveyed for other reasons. Combining survey questions with education and criminal (police and court) administrative records, we observe marriage, earnings, and

cost (Elango et al., 2016).

⁴Intricacies and failures in randomization protocols are not rare in social experiments. However, authors often fail to document and address them methodologically. This failure has sizable empirical consequences (Bruhn and McKenzie, 2009).

⁵The randomization protocol was as follows: 1) Participant status of the younger siblings is the same as that of their older siblings; 2) Those remaining were ranked by their baseline IQ score with odd-ranked and even-ranked subjects assigned to separate groups; 3) Some individuals initially assigned to one group were swapped between groups to balance gender and mean socioeconomic-index scores, with average IQ scores held more or less constant. This generated a minor imbalance in family background variables; 4) A coin toss randomly selected one group as the treatment group and the other as the control group; and 5) Some individuals provisionally assigned to treatment, whose mothers were employed at the time of the assignment, were swapped with control individuals whose mothers were not employed. The reason for this swap was that it was difficult for working mothers to participate in home visits assigned to the treatment group.

criminal histories from enrollment to age 54.⁶

The first part of Panel c. of Table 1 summarizes fertility information of the 102 participants surveyed in the age-54 follow-up. The treatment-control difference in the number of children is small and statistically insignificant. This evidence rules out experimentally induced fertility as an important consideration. The program had minimal impacts on childbearing. Original participants are only asked about their first five children. This does not result in a major loss of information because only a small fraction of first-generation participants report having more than five children. Information losses due to not observing children yet to be born are also a minor issue in the age-54 follow-up, as the vast majority of the original participants are likely to have completed childbearing and adoption.

Appendix A.2 compares the sample of participants with children in the age-54 follow-up to the sample of participants without children. No consistent statistical differences are found between the two samples, although we note that this comparison is not precise because the sample of those without children is very small (9 control-group and 12 treatment-group participants). The mediation analysis reported in Section 5 indicates that fertility variables are minor drivers of the estimated intergenerational impacts.

2.3 Analysis Sample

Our main sample consists of the biological children of original participants. However, a small number of original participants report information on adoptees and stepchildren. They report a total of 10 adopted children (4 in the control group and 6 in the treatment group) and 17 stepchildren (7 in the control group and 10 in the treatment group). We do not include adopted children or stepchildren in the main analysis of this paper because we do not observe important information on their parental origin and age of adoption.

The exclusion of adoptees and stepchildren is a minor issue. The treatment-control

⁶Our empirical strategy accounts for missing data. Appendix Table A.1 shows that missing-data rates due to item non-response are minimal. This paper’s main focus is on intergenerational outcomes. We refer readers interested in the rich panel observation of original-participant outcomes to García et al. (2021).

Table 1. Original Participants (First Generation), Summary Statistics

	<i>Pooled</i>		<i>Male</i>		<i>Female</i>	
	<i>C</i>	<i>(T-C)</i>	<i>C</i>	<i>(T-C)</i>	<i>C</i>	<i>(T-C)</i>
<i>Panel a. Baseline (Age 3)</i>						
IQ	78.54	1.03	77.85	1.37	79.58	0.46
Socioeconomic Index	8.62	0.17	8.65	0.24	8.57	0.09
Mother Works	0.31	-0.22	0.28	-0.22	0.35	-0.23
Mother's Age	28.66	0.92	28.63	0.84	28.71	1.01
Sample Size	65	-7	39	-6	26	-1
<i>Panel b. Age-54 Follow-Up</i>						
Sample Size, Observed	50	2	30	-1	20	3
with Children	41	-1	22	-2	19	1
Sample Size, Not Observed						
Deceased	9	-3	4	0	5	-3
Other Reasons	6	-6	5	-5	1	-1
<i>Panel c. Fertility</i>						
No Children	0.07	0.03	0.10	0.03	0.02	0.02
Children	2.08	0.04	1.87	-0.07	2.40	0.12
>5 Children	0.04	0.04	0.03	0.04	0.05	0.04
Age when Child Born	21.80	1.25	22.82	1.98	20.63	0.67
Age 20 to 35 when Child Born	0.59	0.09	0.68	0.12	0.47	0.08
<i>Panel d. Parenting</i>						
Out of Wedlock when Child Born	0.80	-0.03	0.77	-0.07	0.84	0.01
Cohabitated when Child Grew up	0.61	-0.11	0.59	-0.09	0.63	-0.13
Married through Child's Age 10	0.13	0.19	0.09	0.21	0.18	0.16
Read Daily to Child	0.13	0.13	0.14	0.11	0.11	0.14
<i>Panel e. Education at Age 54</i>						
High School Graduation	0.46	0.31	0.55	0.15	0.37	0.48
College Graduation	0.20	-0.15	0.14	-0.04	0.26	-0.26
<i>Panel f. Average Employment and Income through Child's Age 10</i>						
Fraction of Years Employed	0.44	0.16	0.47	0.16	0.40	0.17
Income (1,000s of 2017 dollars)	18.39	8.58	22.55	11.53	13.57	6.65
<i>Panel g. Crime</i>						
Days in Jail	71.15	-35.52	119.18	-73.83	15.53	10.37
Misdemeanor Arrests	0.90	-0.60	1.45	-0.95	0.26	-0.16
Felony Arrests	0.80	-0.60	1.40	-1.00	0.11	-0.11

Note: Panel a. summarizes basic variables and the sample size at baseline for the original participants. Panel b. summarizes the sample size of the original participants observed and not observed in the age-54 follow-up. The first part of Panel c. summarizes fertility variables for all the original participants observed in the age-54 follow-up. The second part of Panel c. and Panels d. to g. summarize variables at the original participant level for those observed in the age-54 follow-up who report to have children, using information on up to their five eldest children. *C* for sample size rows: number of observations in the control group. *C* for outcome rows: control-group mean for variables at the original-participant level—Panel a., first part of Panel c., and Panels e. and g; and control-group mean in the within original-participant average across up to their five eldest children for variables at the child-of-original-participant level—second part of Panel c. and Panels d. and f. The columns *(T-C)* are constructed analogously to the columns *(C)* for treatment-control differences. We bold *(T-C)* entries for outcome rows when their permutation *p*-values are lower than 0.10. The null hypothesis for each difference is that it is less than or equal to 0. Appendix Table A.1 presents variable definitions and construction details.

difference in the number of adopted children and stepchildren is small and statistically insignificant. The average number of adopted children and stepchildren in the control group is 0.22. The treatment-control average difference in adopted children and stepchildren is 0.09 (p -value > 0.10). Appendix Table A.2 compares the sample of original participants who report having adopted children or stepchildren to those who do not. Appendix Table A.5 compares the outcomes of children analyzed in the main paper with the outcomes of adopted children and stepchildren and finds slight differences. Appendix Table A.8 presents estimates for four samples: a) the sample analyzed in the main text; b) sample of a) plus adopted children; c) sample of a) plus stepchildren; and d) all children. Estimates are barely changed across these samples. Analyzing treatment effects separately for adopted children and stepchildren turns out to be too data demanding because of sample sizes.

Our main analysis sample includes 41 first-generation control and 40 first-generation treatment participants. They have 104 and 110 children, respectively, who constitute the sample of children that we use to assess impacts on the second generation participants. We conduct analyses at the first-generation participant level. Only first-generation participants were randomized. Accordingly, we need to be careful about how we analyze second-generation samples.

We construct intergenerational outcomes as follows. Let \mathcal{I} index first-generation participants and \mathcal{J} index outcomes. Define $Y_{i,j}^{c(i)}$ as the outcome $j \in \mathcal{J}$ of child $c(i)$ of first-generation participant $i \in \mathcal{I}$. The mean outcome j for the children of i is

$$\bar{Y}_{i,j}^c := \frac{1}{\#\mathcal{C}_i} \sum_{c \in \mathcal{C}_i} Y_{i,j}^{c(i)}, \quad (1)$$

where \mathcal{C}_i indexes the children of first-generation participant i .⁷ We define $\bar{Y}_{i,j}^c$ as outcome j for each first-generation participant. $\bar{Y}_{i,j}^c$ is the outcome for the “average child” of i .

⁷ “#” denotes cardinality.

2.4 The First-Generation Participants as Parents

PPP had an impact on the socio-emotional skills of its original treatment participants. Heckman et al. (2013) document that this impact translated into improvements in labor-market, crime, and health outcomes through age 40.⁸ In this subsection, we show that the impact on the skills and health of the original participants persists through their childrearing years up to their late midlife years, and that their improved outcomes translate into better environments for their children to grow up in.

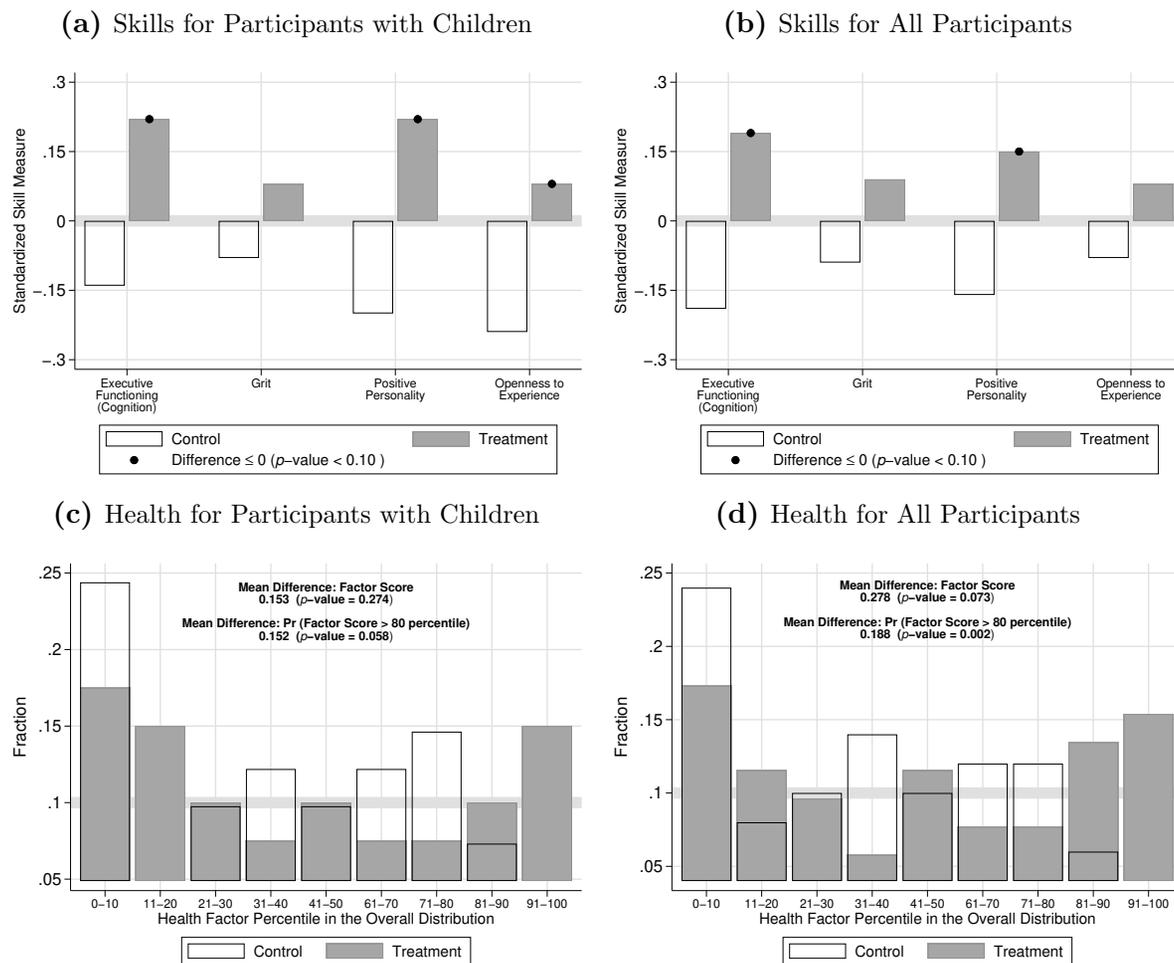
Figures 1a and 1b summarize newly collected data at age 54 on skills of the original participants. They verify that PPP has a long-lasting impact on both cognitive and socio-emotional skills. Ours is the first paper to document the impact of high-quality early education on skills at late midlife. The long-lasting impact on executive functioning challenges the notion of “fadeout” in the treatment effects on skills, specifically on cognition. Previous research claims that the impact of early childhood education on cognitive-test scores disappears (fades out) shortly after the endpoints of interventions (Hojman, 2016; Protzko, 2015). Some authors argue that the fadeout in cognition (and also socio-emotional skills) is real, and not only a measurement artifact (Bailey et al., 2017, 2020). These studies are based on short-run measures. Our results disprove this claim. Our measure of executive functioning uses well-established tests aimed to capture cognition (Raven and Stroop tests). Figure 1 indicates a long-lasting impact on cognition, as well as on socio-emotional skills.

Figures 1c and 1d summarize newly collected data at age 54 on multiple health indicators. We construct a latent variable measuring overall health based on several measures of overall and cardiovascular health.⁹ We analyze how the treatment and control distributions of this health latent fit into the pooled distribution. Treatment shifts rightward the

⁸The impact on adulthood outcomes of the original participants is documented by Conti et al. (2016), Heckman and Karapakula (2019b, 2021), Heckman et al. (2010b), and Heckman et al. (2013).

⁹The measures are listed in Appendix Table A.3. Appendix Table A.3 also summarizes these measures by treatment status for the sample of original participants who report having children in the age-54 follow-up, as well as for all of the original participants.

Figure 1. Age-54 Skills and Health, Original Participants



6

Note: Panel (a) shows the average by treatment status for the measures of the skills in the label for the original participants who reported having children. The measures are latents obtained from items in questionnaires designed to capture each skill. Details are in Appendix A.1. The latents are normalized to have mean 0 and variance 1 among all participants observed in the age-54 follow-up. Examples of items in each skill measure are the following. Executive functioning: Raven-test and Stroop-test items. Grit: self-report of having completed a goal that took years. Positive personality: self reports of reversed measures of disorganized lifestyle and anxiety feelings. Openness to experience: self reports of willingness to take financial risks and measures of openness to new experiences in leisure and other activities. We mark the treatment-group mean when the treatment-control difference has a permutation p -value lower than 0.10. The null hypothesis for the difference is that it is less than or equal to 0. Panel (c) shows how the distribution by treatment status of a health latent, constructed similarly to the skill latents, fits into the overall distribution percentiles. It considers participants who reported having children. Examples of items in the health latent are waist-to-hip ratio, cortisol, cholesterol, chronic pain, and substance-use treatment. Panel (c) also displays the treatment-control mean difference in the factor latent and in the probability of the factor latent being greater than the 80th percentile. We display the permutation p -values for these differences. The null hypothesis for the differences is that they are less than or equal to 0. Panels (b) and (d) are analogous in format to Panels (a) and (c) for all of the original participants.

distribution of the health indicator. Treatment-group members have a higher average health indicator and a higher probability of being healthier than 80% of individuals in the pooled treatment and control sample. The large and long-lasting impact on late midlife health is a new finding. Other studies document impacts on adult health of early childhood education up to age 30 or 40 (Campbell et al., 2014; Conti et al., 2016). Our rich late midlife follow-up allows us to confirm that health impacts are long-lasting. Positive forecasts of the long-run health impact of early childhood education are justified.¹⁰

The second part of Panel c. and Panels d. to g. of Table 1 show that, on average, children of original treatment participants were more likely to grow up with parents who were stably married compared to the average children of original control participants. They were read to more often while growing up.¹¹ Their parents were employed a larger fraction of time, had more education and income, and engaged less in criminal behavior when they were growing up. All of these differences are sizable and statistically significant.¹²

Figure 2 illustrates the evolution of the parental environments in which children of original treatment participants grew up in compared to those of the children of original control-group participants. Panels (a) to (c) show the evolution of marriage, earnings, and arrests over the life-cycle of the original participants who reported having children. Panels (d) to (f) show the same variables throughout the childhood of their “average child.” We plot the mean of the outcome for the “average child” by treatment status and child age.¹³ Children of treatment-group participants are more than 10 percentage points more likely to be born to married parents than children of control-group participants. They are also born to

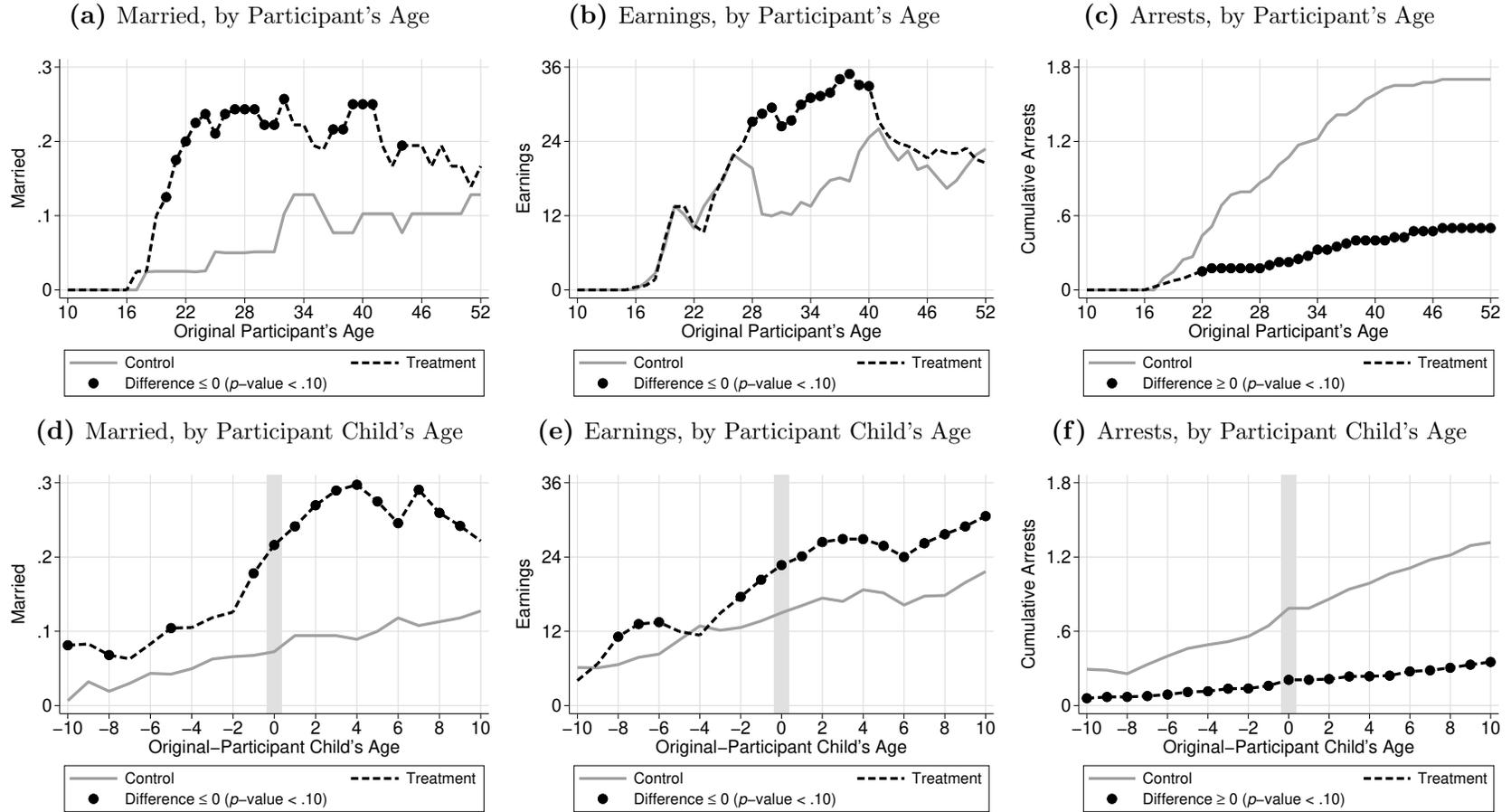
¹⁰Examples of these forecasts are in García and Heckman (2020), García et al. (2020), and García et al. (2021).

¹¹This finding is consistent with Bauer and Schanzenbach (2016), who find that participants of Head Start improve their parenting skills when becoming adults. These authors do not analyze intergenerational outcomes.

¹²The second part of Panel c. and Panels d. to g. only consider first-generation participants who have children. The same is true for Figure 2. We supplement this evidence with longitudinal marriage, earnings, and crime profiles for the full sample in Appendix Figure A.2.

¹³Appendix Figure A.1 is analogous in format to Panels (d) to (f) of Figure 2, but it is based only on the first child of original participants. Panels (d) to (f) of Figure 2 and Appendix Figure A.1 display very similar patterns.

Figure 2. Original-Participant Marriage, Earnings, and Crime by their Age and by their Children’s Age



11

Note: Panel (a) displays the control-group and treatment-group means of a married-status indicator by age of the original participants who reported having children. We mark the treatment-group mean when the treatment-control difference has a permutation p -value lower than 0.10. The null hypothesis for the difference is that it is less than or equal to 0. Panel (b) is analogous in format to Panel (a) for earnings in 1,000s of 2017 USD. Panel (c) is analogous in format to Panel (a) for cumulative violent misdemeanor and felony arrests. For Panel (c) the null hypothesis for the difference is that it is greater than or equal to 0. Panels (d) to (f) are analogous in format to Panels (a) to (c), but they are plotted by age of the children of original participants. For Panels (d) to (f) the outcomes are first averaged within-original-participant across up to five eldest children before constructing control and treatment-control difference means.

parents who, on average, make almost ten thousand dollars *per* year more and have a lower average of cumulative arrests. The advantage of children of treatment-group participants builds up years before they are born. PPP impacts their parents' entire life-cycles, including reducing out-of-wedlock births and criminal activity during youth. This advantage persists throughout their childhoods.

2.5 Intergenerational Outcomes

We analyze the intergenerational outcomes listed in Table 2. We construct these outcomes at the original participant level ($\bar{Y}_{i,j}^c$), including child outcomes ($Y_{i,j}^{c(i)}$). We analyze children of all ages when examining school suspension, special education, crime, and health (Panel b). We only consider children age 18 or older when analyzing high school graduation (Panel c). We impose similar age cutoffs for teenage pregnancy, post-secondary education, college, employment, and marital outcomes. The panels of Table 2 describe the sample size of first-generation and second-generation participants after imposing each age cutoff, and shows that this has minimal consequences on the sample sizes. Second-generation participants are, on average, 28 years old when information on them is reported in the age-54 follow-up (i.e., most of them satisfy all of the age cutoffs imposed). All first-generation participants who report having children have at least one child satisfying all of the age cutoffs except for two.

We adjust $Y_{i,j}^{c(i)}$ to a standard age to account for the variability in child ages. We pool the sample of second-generation treatment and control participants and regress each outcome $Y_{i,j}^{c(i)}$ on age and age squared of the second-generation participant, as well as their gender. We replace $Y_{i,j}^{c(i)}$ with its predicted value from this regression for each individual and outcome. Appendix A.9 shows that our main results remain virtually unchanged when not using age cutoffs or age adjustments.

The treatment-control mean differences for the first generation are the raw source of experimental variation exploited to identify treatment effects. Our estimators adjust these differences to account for compromises in the randomization protocol and for factors pre-

Table 2. Children of the Original Participants (Second Generation), Summary Statistics

	<i>Gender of Parent (Original Participant)</i>					
	<i>Pooled</i>		<i>Male</i>		<i>Female</i>	
	<i>C</i>	<i>(T-C)</i>	<i>C</i>	<i>(T-C)</i>	<i>C</i>	<i>(T-C)</i>
<i>Panel a. Age-54 Follow-Up</i>						
Sample Size, Children of Original Participants Observed	104	6	56	-4	48	10
Age	28.11	0.18	25.75	1.42	30.85	-1.58
<i>Panel b. Children of Any Age</i>						
Never Suspended from School	0.535	0.185	0.614	0.184	0.446	0.195
Never in Special Education	0.832	0.000	0.886	-0.029	0.778	0.030
Never Arrested	0.628	0.097	0.708	0.109	0.536	0.097
Never Addicted	0.926	-0.002	0.924	0.034	0.928	-0.038
In Good Health	0.847	0.087	0.891	0.066	0.796	0.115
Sample Size, Original-Participant Parents	41	-1	22	-2	19	1
Sample Size, Children of Original Participants	104	6	56	-4	48	10
<i>Panel c. Children Age 18 or Older</i>						
High School Graduation Rate	0.756	0.043	0.762	0.128	0.750	-0.042
Sample Size, Original-Participant Parents	40	0	21	-1	19	1
Sample Size, Children of Original Participants	94	10	47	2	47	8
<i>Panel d. Children Age 19 or Older</i>						
Teenage Pregnancy (reversed scale)	0.759	-0.026	0.702	0.043	0.823	-0.101
Sample Size, Original-Participant Parents	40	-1	21	-2	19	1
Sample Size, Children of Original Participants	92	11	45	3	47	8
<i>Panel e. Children Age 21 or Older</i>						
Post-Secondary Education	0.446	0.047	0.425	0.013	0.469	0.075
Sample Size, Original-Participant Parents	40	-3	21	-3	19	0
Sample Size, Children of Original Participants	91	7	44	1	47	6
<i>Panel f. Children Age 23 or Older</i>						
College Graduation	0.184	-0.014	0.208	-0.071	0.158	0.039
Years of Education	12.987	0.150	13.000	-0.044	12.974	0.316
Employed (includes self-employed)	0.424	0.170	0.444	0.199	0.404	0.148
Currently Married	0.136	0.085	0.175	0.027	0.096	0.140
Never Divorced	0.885	0.086	0.904	0.034	0.867	0.133
Sample Size, Original-Participant Parents	39	-4	20	-4	19	0
Sample Size, Children of Original Participants	85	4	40	-1	45	5

Note: Panel a. summarizes the sample generated by information on up to the five eldest children of the original participants and their age. Panel b. summarizes child outcomes for which we do not impose an age cutoff to consider observations. It then summarizes the sample size of the original participants interviewed in the age-54 follow-up who had children, and the size of the sample of children generated. Panels c. to f. are analogous in format to Panel b. after imposing the age cutoffs in the label. *C* for sample size rows: number of observations in the control group. *C* for outcome rows: control-group mean in the within original-participant average across up to their five eldest children. In Panel a. the age variable is summarized at the child level, instead of at the within original-participant average level. The columns (*T-C*) are constructed analogously to the columns (*C*) for treatment-control differences. We bold (*T-C*) entries for outcome rows when their permutation *p*-values are lower than 0.10. The null hypothesis for each difference is that it is less than or equal to 0.

venting us from observing second-generation outcomes. These sources include death and any other reason for not observing first-generation participants in the age-54 follow-up. They also include not observing second-generation outcomes for first-generation participants who do not have children. We treat these three sources equivalently and refer to them jointly as “attrition.”¹⁴

3. Intergenerational Treatment Effects

3.1 Individual-Outcome Treatment Effects

Measurement Framework. We treat average child outcomes as treatment and control outcomes for the original participants, dropping the $c(i)$ and c superscripts for notational simplicity. Let $Y_{i,j}^1$ denote outcome $Y_{i,j}$ when first-generation participant $i \in \mathcal{I}$ is fixed to treatment status ($D_i := 1$) and $Y_{i,j}^0$ when they are fixed to control status ($D_i := 0$). In the framework of Quandt (1958), $Y_{i,j} = Y_{i,j}^1 D_i + Y_{i,j}^0 (1 - D_i)$.

For outcome $j \in \mathcal{J}$, we consider three estimators of the average treatment effect $\mathbb{E}[Y_{i,j}^1 - Y_{i,j}^0]$. The first is the treatment-control *mean difference*. We pool first-generation treatment and control participants and estimate the coefficients in the model

$$Y_{i,j} = \gamma_j + \delta_j D_i + \varepsilon_{i,j}, \tag{2}$$

where $\varepsilon_{i,j}$ is an error term. γ_j is an estimator of the control mean and δ_j is the mean-difference estimator. δ_j identifies the average treatment effect assuming that treatment is randomized without compromises and that attrition is random. Estimates of γ_j and δ_j for the

¹⁴Table 2 reports the number of observations in the sample of children satisfying the age cutoff. When analyzing each outcome, we lose a couple of observations per outcome due to item non-response in the interviews. Item non-response is very minor. For three outcomes, we do not have item-non response cases (never arrested, never addicted, in good health); for four outcomes we have one case (never suspended from school, high school graduation, teenage pregnancy, post-secondary education); for two outcomes we have two cases (college graduation and years of education), for three outcomes we have three cases (never in special education, currently married, never divorced); for one outcome we have four cases (employed). Our estimators account for item non-response as yet another source of attrition. This section provides basic description of the data analyzed. Appendix Tables A.1 and A.4 provide extensive additional details on variable definitions and observations.

observed intergenerational outcomes are in Table 2 (columns \mathbf{C} and $(\mathbf{T}-\mathbf{C})$, respectively). Column (1) of Panel a. in Table 3 reproduces the estimates of δ_j to ease comparison with other estimates.

To address the randomization compromises and attrition patterns described in Section 2 we consider a regression-adjusted mean-difference as a second estimator (*OLS*). We construct the OLS estimator by including the baseline variables in Panel a. of Table 1 in addition to first-generation participant gender as covariates in Equation (2). We do not know the exact form of randomization failure, but we know that baseline variables are only partially balanced across the treatment and control groups. OLS identifies the average treatment effect under the assumption of conditional random assignment to treatment and attrition occurring conditionally at random (i.e., it relaxes the randomization assumption of the mean-difference estimator to be conditional and allows attrition to vary across baseline covariates). Column (2) of Panel a. in Table 3 presents OLS estimates.

Our third estimator is a more general mean-difference adjustment proposed in Heckman and Karapakula (2019b, 2021). It is an augmented inverse-probability weighting (*AIPW*) adapted to the sampling protocol of PPP. This estimator weights Equation (2) by the inverse probability of being treated and having attrited (recall that the reasons for attrition are being dead, not being interviewed in the age-54 follow-up, or not having children). AIPW imputes (missing) counterfactual outcomes for each first-generation participant based on the same baseline variables used as covariates when computing OLS estimates. AIPW is useful for its double-robustness property. It provides a consistent estimator of the average treatment effect if either the weighting scheme or the (imputed) Equation (2) is correctly specified.¹⁵ We use AIPW as our baseline estimator. Panel b. in Table 3 presents AIPW estimates.

We present two supplementary sets of results accounting for compromises in randomization. We first present Lee (2009) bounds for average treatment effects. This method is

¹⁵The identification proofs for the three estimators that we use are standard and we omit them for brevity. Heckman and Karapakula (2019b, 2021) and the Appendix of García et al. (2021) provide detailed proofs.

appropriate for contexts with (conditional) randomized assignment to treatment and sample selection generated by attrition. We refer readers to the source paper for details. Identifying the bounds requires two assumptions: (i) (conditional) randomized assignment to treatment; and (ii) that treatment affects attrition uniformly across the sample (i.e., the probability of being attrited should either increase or decrease as a function of treatment status for all individuals). The first assumption is plausible in our context, as we condition on variables unbalanced due to the compromises in the randomization protocol. Panel b. of Table 1 suggests that the second assumption holds empirically. The first-generation treatment-group participants were more likely to be followed-up at age 54 (either because of death or other reasons). Column (3) of Panel a. in Table 3 presents estimates of the Lee (2009) bounds.

We also present second-generation or child-level OLS estimates of the average treatment effects (i.e., we pool the second-generation participants and regress their outcomes on a variable indicating the treatment status of their parents, as well as on the covariates used in the first-generation OLS model). We refrain from using child-level estimators throughout the paper because only first-generation participants were randomized. We only report child-level OLS estimates for brevity.¹⁶ Column (4) of Panel a. in Table 3 presents child-level OLS estimates.

Inference. The intergenerational outcomes were constructed in a way such that a positive point estimate indicates a beneficial treatment effect.¹⁷ We test if the treatment effect is less than or equal to 0, outcome by outcome. For our baseline AIPW estimator, we present several p -values. Our baseline p -value is permutation-based because it is especially suited for small samples like ours. We also present bootstrap standard errors for all of the estimators considered. In all of our inference, we cluster at the first-generation participant level.¹⁸

¹⁶We use OLS for this exercise because it is a straightforward estimator to communicate, and it addresses our methodological challenges in the most basic manner.

¹⁷This statement is probably not generally true for the outcome “currently married.” We interpret this as a noisy measure of being in a stable relationship. We supplement this with another relationship stability measure, “never divorced,” which is perhaps a more clear-cut beneficial outcome. Empirically, we find no impact on the former variable and a sizable impact on the latter.

¹⁸We follow standard procedures when computing standard errors and p -values. Our standard errors are

Estimates. Panel b. of Table 3 presents our main results. PPP has a beneficial intergenerational impact that is consistent with its impact on the first generation. High-quality early childhood education programs like PPP improve the early-life socio-emotional skills of children. This translates into long-term impacts in labor-market, crime, and health outcomes.¹⁹ School suspension is an indirect measure of early-life socio-emotional skills and PPP has a sizable impact on them for the second generation. The impact on health through young adulthood and longer-term outcomes as employment and marriage stability (never being divorced) are also sizable. For crime, the impact is much stronger for men and we discuss it in Section 4. We reject the null hypothesis of no positive treatment effect on never suspended from school, good health, employment, and no divorce with at least 10% of significance for most p -values considered. The worst-case maximum p -value is a consistent exception. We qualify inference based on it because it is very conservative. Panel a. indicates that the treatment-effect estimates for these four outcomes are remarkably robust when we use other estimators.

We focus only on employment outcomes to interpret our results. We compare the second-generation impacts with the first-generation impacts of PPP and Head Start (a federal early childhood education program targeted toward disadvantaged families like PPP). We estimate that PPP increases the second-generation probability of employment by 25.7 (s.e. 11.5) percentage points. Section 4 documents that the treatment-effect estimate is similar for second-generation male and female participants. Heckman and Karapakula (2021) report

the standard deviation of the empirical bootstrap distribution of each estimator (we cluster the bootstrap drawing at the first-generation participant level). Analytic p -values are asymptotic and robust to heteroskedasticity and arbitrary correlation within first-generation participants (e.g., Liang and Zeger, 1986). They do not account for sampling variation in preliminary estimation stages (e.g., residualization of $Y_{i,j}^{c(i)}$ in Equation (1) or construction of weights in the AIPW estimator). Permutation p -values are calculated as in Lehmann and Romano (2006, Chapter 5). They are especially suited for small-sample-size settings. All bootstrap p -values are calculated as in Hansen (2021, Chapter 10). They account for sampling variation in all estimation stages. We show below that this accounting introduces minor additional variance and use the small-sample-size inference provided by permutation as baseline. Worst-case p -values are as developed in De Haan (1981). We calculate them as in the PPP-specific application of Heckman and Karapakula (2019b, 2021).

¹⁹See Elango et al. (2016) for a survey.

Table 3. Intergenerational Treatment Effects, Main Estimates and Inference

<i>Panel a. Basic Estimates</i>								
	<i>(1)</i> <i>Mean Difference</i>		<i>(2)</i> <i>OLS</i>		<i>(3)</i> <i>Lee Bounds</i>		<i>(4)</i> <i>Child-Level OLS</i>	
	<i>Estimate</i>	<i>S.E.</i>	<i>Estimate</i>	<i>S.E.</i>	<i>Lower</i>	<i>Upper</i>	<i>Estimate</i>	<i>S.E.</i>
Never Suspended from School	0.185	(0.083)	0.140	(0.090)	0.137	0.177	0.134	(0.076)
Never in Special Education	0.000	(0.064)	-0.034	(0.072)	-0.034	-0.002	0.037	(0.073)
Never Arrested	0.097	(0.087)	0.060	(0.087)	0.063	0.094	0.039	(0.079)
Never Addicted	-0.002	(0.042)	0.008	(0.045)	-0.018	0.002	0.003	(0.044)
In Good Health	0.087	(0.054)	0.099	(0.064)	0.060	0.088	0.135	(0.081)
High School Graduation	0.043	(0.076)	0.033	(0.080)	0.036	0.048	0.096	(0.086)
Teenage Pregnancy (reversed scale)	-0.026	(0.080)	-0.089	(0.078)	-0.049	-0.042	-0.120	(0.070)
Post-Secondary Education	0.047	(0.095)	-0.002	(0.102)	0.022	0.080	0.088	(0.099)
College Graduation	-0.014	(0.065)	-0.004	(0.061)	-0.039	0.038	0.025	(0.058)
Years of Education	0.150	(0.327)	0.144	(0.332)	-0.074	0.299	0.418	(0.351)
Employed	0.170	(0.095)	0.225	(0.102)	0.151	0.253	0.207	(0.106)
Currently Married	0.085	(0.073)	0.032	(0.081)	0.058	0.128	-0.005	(0.089)
Never Divorced	0.086	(0.047)	0.047	(0.051)	0.028	0.084	0.053	(0.062)

<i>Panel b. Main Estimates (Estimator: AIPW)</i>								
			<i>p-values</i>					
	<i>Estimate</i>	<i>S.E.</i>			<i>Bootstrap</i>		<i>Worst-Case</i>	
			<i>Analytic</i>	<i>Permutation</i>	<i>Simple</i>	<i>Studentized</i>	<i>De Haan</i>	<i>Maximum</i>
Never Suspended from School	0.174	(0.095)	[0.010]	[0.024]	[0.032]	[0.015]	[0.054]	[0.121]
Never in Special Education	-0.050	(0.082)	[0.176]	[0.738]	[0.277]	[0.151]	[0.292]	[0.520]
Never Arrested	0.093	(0.090)	[0.082]	[0.127]	[0.151]	[0.087]	[0.184]	[0.308]
Never Addicted	0.032	(0.052)	[0.209]	[0.289]	[0.312]	[0.162]	[0.368]	[0.539]
In Good Health	0.112	(0.070)	[0.016]	[0.038]	[0.042]	[0.018]	[0.078]	[0.204]
High School Graduation	0.041	(0.094)	[0.279]	[0.313]	[0.322]	[0.279]	[0.384]	[0.649]
Teenage Pregnancy (reversed scale)	-0.061	(0.085)	[0.181]	[0.761]	[0.270]	[0.134]	[0.363]	[0.559]
Post-Secondary Education	0.008	(0.107)	[0.461]	[0.464]	[0.474]	[0.458]	[0.566]	[0.854]
College Graduation	-0.041	(0.070)	[0.208]	[0.725]	[0.320]	[0.160]	[0.437]	[0.501]
Years of Education	0.079	(0.384)	[0.385]	[0.410]	[0.405]	[0.402]	[0.516]	[0.665]
Employed	0.257	(0.115)	[0.001]	[0.008]	[0.007]	[0.002]	[0.018]	[0.051]
Currently Married	0.022	(0.083)	[0.362]	[0.378]	[0.267]	[0.500]	[0.466]	[0.871]
Never Divorced	0.077	(0.068)	[0.056]	[0.066]	[0.094]	[0.077]	[0.148]	[0.285]

Note: Panel a. presents treatment-effect estimates and standard errors of the average treatment effect for each intergenerational outcome using the mean-difference, OLS, and child-level OLS estimators explained in the text. It also presents the Lee (2009) bounds. For estimation based on the mean-difference and OLS estimators and the Lee (2009) bounds we define the outcomes as summarized in Table 2 (within original-participant averages across up to five eldest children). For the child-level OLS, outcomes are as observed at the child level. Panel b. presents treatment-effect estimates, standard errors, and *p*-values based on our baseline estimator (AIPW). The AIPW estimator and *p*-values are explained in Section 3. The standard errors are bootstrapped and clustered at the first-generation participant level. The null hypothesis for each treatment effect is that it is less than or equal to 0.

an age-40 first-generation impact of PPP on employment of 26.6 percentage points for men (p -value = 0.02) and -1.6 percentage points for women (p -value = 0.50). Our results indicate that not only do the first-generation impacts of PPP have second-generation impacts, but they also widen by gender. The intergenerational impact of PPP on employment is also larger than the first-generation impact of Head Start on the probability of not being idle during young adulthood—7.1 (s.e. 3.8) percentage points (Deming, 2009). PPP has a larger second-generation impact than the first-generation impact of Head Start.²⁰

We detect consistent, significant treatment effects on four out of the thirteen outcomes we study when analyzing the pooled sample of male and female second-generation participants. For two reasons, these results are unlikely to be a consequence of cherry picking. First, we analyze all of the second-generation outcomes observed. If all thirteen treatment effects were 0, we would reject the null hypothesis of no treatment effect for 10% of outcomes by chance when employing a significance level of 10%. We find a positive and statistically significant treatment effect for $4/13 \times 100 \approx 31\%$ of our outcomes. Second, most of our outcomes are interpretable categories of independent interest. Correcting p -values for multiple hypothesis testing by lumping distinct outcomes into common categories, the blocks used to perform step-down procedures as in Romano and Wolf (2005) would contain between one and three outcomes. Any correction would thus be minimal and potentially overly conservative.

3.2 Aggregate Treatment Effects

Table 3 indicates positive treatment effects for the majority of outcomes, which are consistent across the estimators considered (recall that outcomes are constructed in a way such that a positive treatment effect implies a beneficial impact). While not all estimates are statistically significant, the consistent, positive treatment-effect estimates suggest a treatment effect on

²⁰Head Start is of relatively high quality but varies in the effectiveness of the services offered across the United States. Walters (2015) finds that variation in these services or inputs largely explains differences in Head Start’s short-term effects. The inputs include center-based care, home visiting, the HighScope curriculum modeled after PPP, and class size. Walters (2015) documents that Head Start centers which combine center-based care and home visiting, like PPP, are the most effective; he does not investigate long-term effects.

the joint distribution of outcomes. We test this formally.

Tests. Let Δ_j denote the treatment effect on outcome $j \in \mathcal{J}$ (i.e., $\Delta_j := \mathbb{E}[Y_{i,j}^1 - Y_{i,j}^0]$). The fraction of outcomes for which PPP had a beneficial impact is $P := \frac{1}{\#\mathcal{J}} \sum_{j \in \mathcal{J}} \mathbf{1}[\Delta_j > 0]$, where $\mathbf{1}[\cdot] = 1$ if the statement in brackets is true and $\mathbf{1}[\cdot] = 0$ otherwise. Under the null hypothesis of $\Delta_j = 0 \forall j \in \mathcal{J}$ and assuming the validity of asymptotic approximations, the expectation of P is $1/2$. We test the null hypothesis $P = 1/2$ computing the p -value from the empirical bootstrap distribution of P . Similarly, we define the fraction of outcomes for which the treatment effect is positive and statistically significant at the 10% level (P_{10}). Under the same null hypothesis, the expectation of P_{10} is $1/10$ when using a 10% significance level for testing. That is, 10% of outcomes should be “significant” at the 10% level just by chance. We test the null hypothesis $P_{10} = 1/10$. García et al. (2018) develop tests for combining functions P and P_{10} .

We supplement these combining-function tests with a non-parametric test comparing the joint distribution of the outcomes considered. The test is due to Rosenbaum (2005). We provide a basic description of the test and refer interested readers to the source paper for the details. Let $d_{ii'}$ be the Mahalanobis (1936) distance between individuals i and $i' \in \mathcal{I}$ with $i \neq i'$ based on the outcomes observed. There is an optimal non-bipartite pairing of first-generation participants based on $d_{ii'}$ denoted by \mathcal{O} (Derigs, 1988). Under the null hypothesis of no treatment-control difference in the joint distribution of outcomes, treatment-control pairings should be as frequent as treatment-treatment and control-control pairings. The number of treatment-control pairings in \mathcal{O} , denoted by O , is a summary statistic of the non-bipartite optimal pairing. O has an exact distribution that enables us to compute the p -value of interest (small values of O accumulate evidence against the null hypothesis).

Estimates. Column (1) of Panel a. in Table 4 presents tests for the aggregate treatment effect across outcomes in our main sample.²¹ We fail to reject that the number of outcomes

²¹Table 4 is based on the AIPW estimator. Appendix Tables A.6 and A.7 are analogous in format and based on the mean-difference and OLS estimators, respectively.

displaying a positive treatment effect is $1/2$. The non-rejection is however marginal (p -value = 0.128). We reject the null hypothesis that $1/10$ of the treatment effects are positive and significant at the 10% (p -value = 0.000). We thus reject that the four main positive and significant treatment effects are only found by chance. We also reject that the joint distribution of outcomes is the same for the treatment and control groups using the non-parametric exact test of Rosenbaum (2005) (p -value = 0.000). The three tests as a group corroborate that PPP shifts the joint distribution of intergenerational outcomes.

4. Gender Differences in Intergenerational Treatment Effects

The impact of early childhood education is usually found to be greater for boys than for girls in long-term labor market and crime outcomes (Elango et al., 2016). The benefit in terms of education is usually greater for girls.²² The first-generation impact of PPP is consistent with these findings. Panel a. of Table 4 shows a greater intergenerational impact on second-generation male children than on second-generation female children. The three joint-outcome tests indicate this. For instance, we reject the null hypothesis that the treatment effect is less than or equal to 0 using a significance level of 10% for six out of the thirteen outcomes we analyze for second-generation male children. We also reject the null hypothesis that these treatment effects are found by chance. For second-generation female participants, we only reject the single-outcome null for two out of thirteen outcomes. We fail to reject

²²See Elango et al. (2016) for a documentation of gender differences in the impact of several early childhood education programs. Explanations for the gendered impacts include the following. Baker et al. (2008, 2015) establish a harmful impact of lower-quality universal childcare. Kottelenberg and Lehrer (2014) localize this negative impact on boys. Their results indicate that boys are less resilient than girls and putting them in lower-quality environments instead of keeping them at home hurts them; they are consistent with literature supporting greater vulnerability of boys to adverse environments. Golding and Fitzgerald (2017) and Schore (2017) discuss the potential reasons for this greater vulnerability. They are also consistent with literature documenting that boys develop later than girls and thus benefit from an enriched environment (Bertrand and Pan, 2013; Lavigne et al., 1995; Masse and Tremblay, 1997; Nagin and Tremblay, 2001). Autor et al. (2019) show that boys are more affected than girls by household economic shocks. Supplementing boys' environment with high-quality early childhood education is thus more beneficial for them than it is for girls. García et al. (2018, 2019) is an exception in that they find that high-quality early childhood education favors girls more than boys. The authors document that, in their context, there is more scope of improvement in households of girls relative to boys, and thus there is a greater benefit for girls. The greater scope of improvement for girls relative to boys results from fathers being more likely to stay together with mothers and provide for their children when a boy (rather than a girl) is born (e.g., Dahl and Moretti, 2008).

Table 4. Intergenerational Treatment Effects, Estimates by Gender and Aggregate Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>First Generation:</i>	<i>Pooled</i>			<i>Male</i>			<i>Female</i>		
<i>Second Generation:</i>	<i>Pooled</i>	<i>Male</i>	<i>Female</i>	<i>Pooled</i>	<i>Male</i>	<i>Female</i>	<i>Pooled</i>	<i>Male</i>	<i>Female</i>
<i>Panel a. Aggregate Treatment-Effect Tests</i>									
Counts of Treatment Effects									
<i>P</i> : Fraction Positive	10/13	11/13	9/13	7/13	10/13	5/13	8/13	10/13	8/13
	[0.128]	[0.026]	[0.184]	[0.147]	[0.079]	[0.325]	[0.243]	[0.067]	[0.270]
<i>P</i> ₁₀ : Fraction positive and <i>p</i> -value < 0.10	4/13	6/13	2/13	5/13	5/13	2/13	2/13	4/13	1/13
	[0.000]	[0.000]	[0.250]	[0.000]	[0.030]	[0.150]	[0.110]	[0.000]	[0.470]
Rosenbaum (2005) Test <i>p</i> -value	[0.000]	[0.000]	[0.172]	[0.376]	[0.555]	[0.729]	[0.548]	[0.651]	[0.133]
<i>Panel b. Individual-Outcome Treatment Effects</i>									
Never Suspended from School	0.174	0.201	0.123	0.226	0.117	0.164	0.101	0.318	0.066
Never in Special Education	-0.050	0.050	-0.100	-0.066	-0.092	-0.058	-0.028	0.250	-0.160
Never Arrested	0.093	0.088	-0.011	0.178	0.218	-0.005	-0.026	-0.096	-0.019
Never Addicted	0.032	0.064	0.077	0.060	0.022	0.089	-0.006	0.124	0.060
In Good Health	0.112	0.188	0.055	0.108	0.187	-0.006	0.118	0.189	0.142
High School Graduation	0.041	0.011	0.251	0.152	-0.018	0.435	-0.115	0.051	-0.007
Teenage Pregnancy (reversed scale)	-0.061	-0.045	-0.081	-0.030	-0.060	-0.068	-0.105	-0.023	-0.100
Post-Secondary Education	0.008	0.140	0.004	-0.068	0.036	-0.109	0.116	0.287	0.163
College Graduation	-0.041	0.110	-0.109	-0.078	0.082	-0.135	0.011	0.148	-0.073
Years of Education	0.079	0.699	0.131	-0.179	0.310	-0.078	0.444	1.248	0.427
Employed	0.257	0.226	0.214	0.298	0.277	0.361	0.199	0.154	0.007
Currently Married	0.022	-0.054	0.133	-0.003	0.017	0.151	0.058	-0.156	0.107
Never Divorced	0.077	0.089	0.046	0.043	0.075	-0.036	0.125	0.110	0.162

Note: Panel a. presents the fraction of positive treatment-effect estimates and positive treatment-effect estimates that differ statistically from 0 at a 10% level, by gender of the first-generation participant (parent) and second-generation participant (child). We present their bootstrap *p*-values in brackets. The null hypotheses for the fractions are that they are 0.50 and 0.10, respectively. Panel a. also presents the *p*-value for the Rosenbaum (2005) test. The null hypothesis in this test is that the joint distribution of the thirteen outcomes is the same for the treatment and control groups. Details on this test are explained in Section 3. Panel b. presents treatment-effect estimates for our thirteen second-generation outcomes by gender of the first-generation participant (parent) and second-generation participant (child) using our AIPW estimator. When pooling the first and second generations, our baseline estimates are obtained (first column). We bold the treatment-effect estimates when their permutation *p*-values are lower than 0.10. The null hypothesis for each treatment effect is that it is less than or equal to 0.

that these treatment effects are “significant” only by chance. The inference from the exact, non-parametric Rosenbaum (2005) test is consistent with these results.

Our joint tests displayed in Table 4 also show that second-generation male participants benefit similarly from the intergenerational impact of PPP when their parents are first-generation male participants compared to when they are first-generation female participants. Crime is an exception. We find a substantial, positive impact on never being arrested for children of first-generation male participants. The impact on second-generation male children drives this result. PPP reduced criminal activity of first-generation male participants. The intergenerational impact on their sons is consistent with recent studies in economics and sociology finding that parental incarceration (most incarcerated individuals are men) leads to a significant intergenerational increase in behavior issues and teen crime (Dobbie et al., 2018; Haskins, 2014; Murray et al., 2014; Turney and Haskins, 2014). These results are primarily driven by disadvantaged individuals, making this comparison to our study relevant.²³ The second-generation crime impact is also consistent with studies in other fields documenting that early-life environments determine young-adult criminal activity (Henry et al., 1999; Piquero and Moffitt, 2005; Wright et al., 1999). We next investigate how the improvement in early-life environments through impacts on the first-generation drive our intergenerational treatment effects.

5. Intergenerational Treatment-Effect Mechanisms

The treatment effects on first-generation participants translate into better environments for the second-generation participants (see Section 2.4). We use the mediation framework in Heckman et al. (2013) to quantify the role of these better environments in the intergenerational impacts documented in Section 3.

Recall that for outcome $j \in \mathcal{J}$, $Y_{i,j} = Y_{i,j}^1 D_i + Y_{i,j}^0 (1 - D_i)$. We define outcome $j \in \mathcal{J}$

²³Dobbie et al. (2018) is based on Swedish administrative data. A similar study in the Norwegian context is Bhuller et al. (2018). These authors do not find an intergenerational impact of incarceration on child criminal activity. However, they do not provide results for disadvantaged children.

when first-generation participant i is fixed to treatment ($d = 1$) or control ($d = 0$) as follows:

$$Y_{i,j}^d = \alpha_j^d + \boldsymbol{\mu}_j^d \mathbf{M}_i^d + \varepsilon_{i,j}^d, \quad (3)$$

where α_j^d is a constant and $\boldsymbol{\mu}_j^d$ is the coefficient vector associated with the vector of mediators \mathbf{M}_i^d with $\mathbf{M}_i = \mathbf{M}_i^1 D_i + \mathbf{M}_i^0 (1 - D_i)$.²⁴ The error term in Equation (2) is $\varepsilon_i := \varepsilon_{i,j}^1 D_i + \varepsilon_{i,j}^0 (1 - D_i)$. We decompose the average treatment effect into what is explained by mediators and what is explained by any other factors:

$$\begin{aligned} \Delta_j &:= \mathbb{E} [Y_{i,j}^1 - Y_{i,j}^0 | \mathbf{M}_i] \\ &= \underbrace{(\boldsymbol{\mu}_j^1 \mathbf{M}_i^1 - \boldsymbol{\mu}_j^0 \mathbf{M}_i^0)}_{\text{Explained by Mediators}} + \underbrace{(\alpha_j^1 - \alpha_j^0)}_{\text{Explained by Other Factors}}, \end{aligned} \quad (4)$$

allowing both $\boldsymbol{\mu}^d$, the coefficients (technology) associated with the mediators in Equation (3), and \mathbf{M}_i^d , the mediators themselves, to be impacted by treatment. The decomposition in Equation (4) imposes structure on the coefficients γ_j and δ_j in Equation (2): $\gamma_j := \alpha_j^0$ and $\delta_j := (\alpha_j^1 - \alpha_j^0) + (\boldsymbol{\mu}_j^1 \mathbf{M}_i^1 - \boldsymbol{\mu}_j^0 \mathbf{M}_i^0)$.²⁵ Generalizing Equation (4) to a case conditional on baseline characteristics as in the OLS estimator of Section 3 is straightforward. We work with this case to address the methodological concerns discussed in Section 2 (compromises in the randomization protocol and attrition).

We present mediation analyses assuming that \mathbf{M}_i^0 and \mathbf{M}_i^1 have four common elements. These elements are principal components of the variables in the categories described in Table 1. The categories and variables used to construct the principal components as listed in Table 1 (in parentheses) are: fertility (Panel c. except for “No Children”), parenting (Panel d.), education and employment (Panels e. and f.), and crime (Panel g.). We use principal

²⁴The mediators are potentially impacted by treatment.

²⁵Identifying the elements of the decomposition in Equation (4) relies on this structure and on assuming mean independence of the term $\varepsilon_{i,j}^d$ in Equation (3) for $d = 0, 1$ and $j \in \mathcal{J}$. The structure or mean independence assumption are not requirements for identifying the treatment effects discussed in Sections 3 and 4.

components instead of the full set of variables for empirical tractability.²⁶ We standardize the principal components to an in-sample mean of 0 and standard deviation of 1.²⁷

Appendix Table A.14 provides estimates of the elements in Equation (4) for each of the thirteen intergenerational outcomes studied.²⁸ We further decompose the first element (“Explained by Mediators”) into the subcomponents corresponding to each mediator. Figure 3a displays the results for the subset of outcomes with a positive treatment effect. We omit the other outcomes from the visual display because their mediation-component estimates are imprecisely estimated. We omit years of education from the display because the intergenerational impact on education is primarily driven by high school graduation, which is readily displayed. For each outcome, the individual estimated components of Equation (4) are displayed as a fraction of the total sum of the estimated components. For reference, the AIPW estimate in Panel a. of Table 3 is displayed in the figure above each decomposition bar. We also display the permutation p -value corresponding to the level of the estimated treatment effect explained by each mediator—i.e., the p -value for the estimates in $(\boldsymbol{\mu}_j^1 \mathbf{M}_i^1 - \boldsymbol{\mu}_j^0 \mathbf{M}_i^0)$.

Figure 3a summarizes our mediation analysis. There are experimentally induced improvements in mediators summarizing first-generation parent home-environment quality and stability. The figure indicates that this improvement explains almost 50% of the intergenerational impact on the second-generation across the outcomes displayed. Parenting and crime

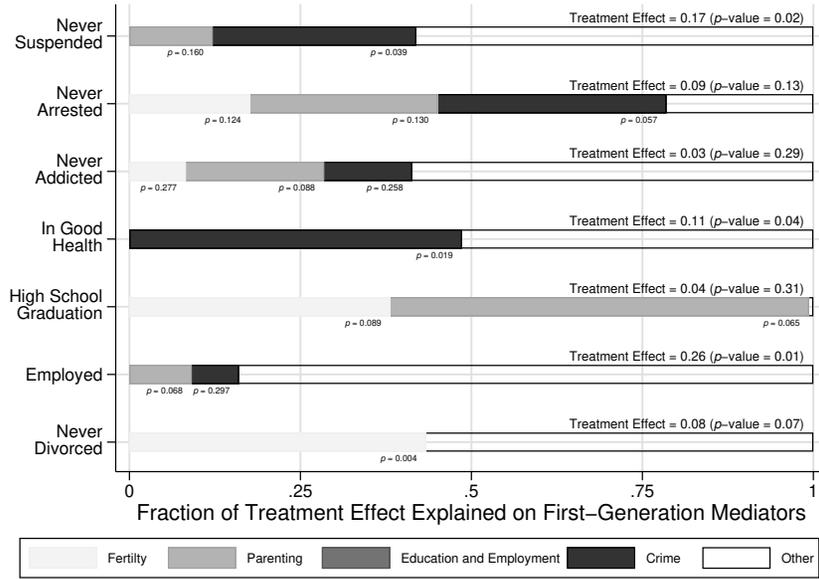
²⁶We use the original-participant outcomes as mediators. The original-participant skills summarized in Section 2.4 are themselves the mediators of these outcomes (Heckman et al., 2013). We do not include original-participant skills explicitly as mediators in this section to avoid overfitting our mediation models.

²⁷The control-group mean and treatment-control mean difference of the principal components are: fertility (control mean = -0.03 , mean difference = 0.06 with p -value = 0.36), parenting (control mean = -0.21 , mean difference = 0.44 with p -value = 0.03), education and employment (control mean = -0.17 , mean difference = 0.36 with p -value = 0.06), and crime (control mean = 0.23 , mean difference = 0.47 with p -value = 0.02). We reverse the scale of the crime principal component for the mediation analysis. Appendix Tables A.10 to A.13 provide details on how we construct each factor.

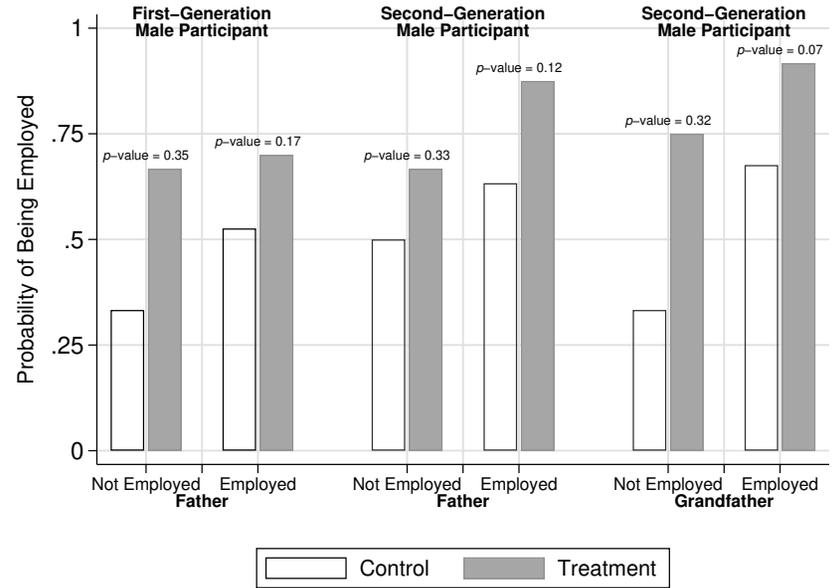
²⁸We obtain these estimates as follows. First, we estimate the principal component for each of the four categories of mediators in the pooled sample of first-generation treatment and control participants. These are the elements in \mathbf{M}_i . Second, we estimate Equation (3) using the control-group sample to obtain estimates of α_j^0 and $\boldsymbol{\mu}_j^0$ and using the treatment-group sample to obtain estimates of α_j^1 and $\boldsymbol{\mu}_j^1$ (one outcome at a time). Third, we construct the empirical counterpart to Equation (4) for each outcome. When displaying results in Figure 3a, we divide each of the individual estimated components of Equation (4) by the sum of all estimated components.

Figure 3. Intergenerational Treatment-Effect Mechanisms

(a) Mediation Analysis



(b) Employment Transitions



Panel (a) decomposes the estimates of the average treatment effect on the second-generation outcome in the label into first-generation mediators impacted by PPP. The component “other” is a residual potentially containing unobserved mediators. Details are in the text of this section. The AIPW estimate and permutation p -value of the treatment effect decomposed are displayed below the bar corresponding to each outcome. We also display the permutation p -value corresponding to the level of the treatment effect explained by each mediator for each outcome. To simplify the display, we omit from the figure mediators explaining a negligible fraction of the treatment effect (we discard negligible mediators outcome by outcome). The null hypothesis for each level explained by mediator and each treatment effect is that they are less than or equal to 0.

Outcomes in Display: For display clarity, we omit in this figure outcomes for which mediation analysis is imprecisely estimated. We include all thirteen intergenerational outcomes analyzed throughout the paper in Appendix Table A.14.

Panel (b) displays the raw probability of being employed for first-generation male participants by their treatment status and by the employment status of their fathers, for second-generation male participants by treatment and employment status of their fathers, and for second-generation male participants by treatment status of their fathers and employment status of their grandfathers. For each employment status, we present the p -value of the treatment-control difference. The null hypothesis for the difference is that it is less than or equal to 0. Employment for fathers of first-generation participants is measured, on average, at age 33. For the first-generation (original) participants, employment is measured, on average, at age 27. For their children, it is measured, on average, at age 28.

are the most salient mediators. Each of them explains a significant fraction of the treatment effects displayed, except for divorce. Education, employment, and fertility play secondary roles. Figure 3a reinforces the suggestion in Section 4 that improvements in the home environment, especially an increase in parental presence due to lower criminal activity in the original-participant treatment group, have long-term intergenerational consequences.²⁹

Figure 3b further explores transmission mechanisms for employment, the only outcome observed across all three generations (parents of the original participants, original participants, and their children). We measure employment around age 30 for all three generations. We only consider males for this exercise to ameliorate well-known labor-force-participation selection issues. We first show that the probability of being employed for first-generation participants increases with treatment, especially for those individuals whose fathers are not employed. We then show an analogous exercise relating first-generation and second-generation participants and find a similar pattern. The relationship also holds between the fathers of the original participants and their grandsons. PPP thus has a sustained intergenerational impact on breaking poverty cycles of non-employment.

6. Summary

The Perry Preschool Project was a pioneering early childhood education program designed to promote the social mobility of disadvantaged African-American children. The principles of the program still guide current practice (Elango et al., 2016). Using newly collected data, we examine its impact on the original participants at age 54 and on their adult children. We find substantial and lasting positive effects for treatment-group members on cognition and beneficial personality traits contradicting claims on fadeout that are based on relatively short-term follow-ups. We also document long-lasting impacts on health using a rich set of measures that include overall health and cardiovascular indicators. The first-generation

²⁹Preliminary analysis indicates that the mediators of treatment effects by gender are quantitatively similar to the mediators in the pooled sample analyzed in this section. We refrain from presenting mediation analysis by gender because it is imprecisely estimated and does not uncover conclusions different than those from this section.

treatment-group members have more stable home lives in terms of marriage and divorce and higher incomes in the child rearing years.

These benefits promote intergenerational mobility of their children. The second generation children of treatment-group members have less special education and fewer school suspensions than the children of control-group members. They are more likely to be employed, married, to graduate high school, and much less likely to engage in crime. There are important differences in impact by gender. The male children of the male treatment group member receive the greatest benefits, consistent with a literature on the adverse effects of disadvantaged environments on boys (Autor et al., 2019). Improvements in parenting and reduced criminal activity are important mediators. As measured by employment, the program attenuates the intergenerational transmission of disadvantage.

References

- Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2019). Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. *American Economic Journal: Applied Economics* 11(3), 338–81.
- Bailey, D., G. J. Duncan, C. L. Odgers, and W. Yu (2017). Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness* 10(1), 7–39.
- Bailey, D. H., G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager (2020). Persistence and Fade-out of Educational-Intervention Effects: Mechanisms and Potential Solutions. *Psychological Science in the Public Interest* 21(2), 55–97.
- Baker, M., J. Gruber, and K. Milligan (2008). Universal Childcare, Maternal Labor Supply, and Family Well-Being. *Journal of Political Economy* 116(4), 709–745.
- Baker, M., J. Gruber, and K. Milligan (2015). Non-Cognitive Deficits and Young Adult Outcomes: The Long-Run Impacts of a Universal Child Care Program. NBER Working Paper w21571, National Bureau of Economic Research.
- Barnett, W. S. (1996). *Lives in the Balance: Age 27 Benefit-Cost Analysis of the High/Scope Perry Preschool Program*. Ypsilanti, MI: High/Scope Press.
- Barr, A. and C. Gibbs (2019). Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood. Ed Working Paper 19-141, Annenberg Institute at Brown University.

- Bauer, L. and D. W. Schanzenbach (2016). The Long-Term Impact of the Head Start Program. Technical report, The Hamilton Project, Brookings.
- Bertrand, M. and J. Pan (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics* 5(1), 32–64.
- Bhuller, M., G. B. Dahl, K. V. Loken, and M. Mogstad (2018). Intergenerational Effects of Incarceration. In *AEA Papers and Proceedings*, Volume 108, pp. 234–40.
- Bruhn, M. and D. McKenzie (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1(4), 200–232.
- Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014). Early Childhood Investments Substantially Boost Adult Health. *Science* 343(6178), 1478–1485.
- Conti, G., J. J. Heckman, and R. Pinto (2016). The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour. *The Economic Journal* 126(596), F28–F65.
- Dahl, G. B. and E. Moretti (2008). The Demand for Sons. *The Review of Economic Studies* 75(4), 1085–1120.
- De Haan, L. (1981). Estimation of the Minimum of a Function Using Order Statistics. *Journal of the American Statistical Association* 76(374), 467–469.
- Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics* 1(3), 111–34.
- Derigs, U. (1988). Solving Non-Bipartite Matching Problems Via Shortest Path Techniques. *Annals of Operations Research* 13(1), 225–61.
- Dobbie, W., H. Grönqvist, S. Niknami, M. Palme, and M. Priks (2018). The Intergenerational Effects of Parental Incarceration. NBER Working Paper w24186, National Bureau of Economic Research.
- Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2016). Early Childhood Education. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 2, Chapter 4, pp. 235–297. Chicago: University of Chicago Press.
- García, J. L., F. Bannhoff, J. J. Heckman, and D. E. Leaf (2021). The Dynastic Benefits of Early Childhood Education. NBER Working Paper w29004, National Bureau of Economic Research.
- García, J. L. and J. J. Heckman (2020). Early Childhood Education and Life-cycle Health. *Health Economics*, 1–23.

- García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). Quantifying the Life-Cycle Benefits of an Influential Early Childhood Program. *Journal of Political Economy* 128(7), 2502–2541.
- García, J. L., J. J. Heckman, and A. L. Ziff (2018). Gender Differences in the Benefits of an Influential Early Childhood Program. *European Economic Review* 109, 9–22.
- García, J. L., J. J. Heckman, and A. L. Ziff (2019). Early Childhood Education and Crime. *Infant Mental Health Journal* 40(1), 141–151.
- Golding, P. and H. E. Fitzgerald (2017). Psychology of Boys At Risk: Indicators From 0-5. *Infant Mental Health Journal* 38(1), 5–14.
- Hansen, B. E. (2021). *Econometrics*. [Online; accessed 31-March-2021].
- Haskins, A. R. (2014). Unintended Consequences: Effects of Paternal Incarceration on Child School Readiness and Later Special Education Placement. *Sociological Science* 1, 141.
- Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review* 103(6), 2052–2086.
- Heckman, J. J. and G. Karapakula (2019a). Intergenerational and Intragenerational Externalities of the Perry Preschool Project. NBER Working Paper w25889, National Bureau of Economic Research.
- Heckman, J. J. and G. Karapakula (2019b). The Perry Preschoolers at Late Midlife: A Study in Design-Specific Inference. NBER Working Paper w25888, National Bureau of Economic Research.
- Heckman, J. J. and G. Karapakula (2021). Using a Satisficing Model of Experimenter Decision-Making to Guide Finite-Sample Inference for Compromised Experiments. *Econometrics Journal* 24(2), C1–C39.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010a). The Rate of Return to the HighScope Perry Preschool Program. *Journal of Public Economics* 94(1-2), 114–128.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b). Analyzing Social Experiments as Implemented: A Reexamination of the Evidence From the HighScope Perry Preschool Program. *Quantitative Economics* 1(1), 1–46.
- Henry, B., A. Caspi, T. E. Moffitt, H. Harrington, and P. A. Silva (1999). Staying in School Protects Boys with Poor Self-Regulation in Childhood from Later Crime: A Longitudinal Study. *International Journal of Behavioral Development* 23(4), 1049–1073.
- Hojman, A. P. C. (2016). *Three Essays on the Economics of Early Childhood Education Programs*. Ph. D. thesis, The University of Chicago.

- Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). Fostering and Measuring Skills: Interventions that Improve Character and Cognition. OECD Education and Social Program Report, OECD.
- Kottelenberg, M. J. and S. F. Lehrer (2014). The Gender Effects of Universal Child Care in Canada: Much Ado About Boys. *Unpublished Manuscript, Department of Economics, Queen's University*.
- Lavigueur, S., R. E. Tremblay, and J.-F. Saucier (1995). Interactional Processes in Families with Disruptive Boys: Patterns of Direct and Indirect Influence. *Journal of Abnormal Child Psychology* 23(3), 359–378.
- Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies* 76(3), 1071–1102.
- Lehmann, E. L. and J. P. Romano (2006). *Testing Statistical Hypotheses*. Springer Science and Business Media.
- Liang, K.-Y. and S. L. Zeger (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* 73(1), 13–22.
- Mahalanobis, P. C. (1936). On the Generalized Distance in Statistics. *Proceedings of the National Institute of Sciences of India* 2(1), 49–55.
- Masse, L. C. and R. E. Tremblay (1997). Behavior of Boys in Kindergarten and the Onset of Substance Use During Adolescence. *Archives of General Psychiatry* 54(1), 62–68.
- Murray, J., C. C. Bijleveld, D. P. Farrington, and R. Loeber (2014). *Effects of Parental Incarceration on Children: Cross-National Comparative Studies*. Washington DC: American Psychological Association.
- Nagin, D. S. and R. E. Tremblay (2001). Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method. *Psychological Methods* 6(1), 18.
- Piquero, A. R. and T. E. Moffitt (2005). Explaining the Facts of Crime: How the Developmental Taxonomy Replies to Farrington's Invitation. *Integrated Developmental and Life-Course Theories of Offending*, 51–72.
- Protzko, J. (2015). The Environment in Raising Early Intelligence: A Meta-Analysis of the Fadeout Effect. *Intelligence* 53, 202–210.
- Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association* 53(284), 873–880.
- Romano, J. P. and M. Wolf (2005). Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. *Journal of the American Statistical Association* 100(469), 94–108.

- Rosenbaum, P. R. (2005). An Exact Distribution-Free Test Comparing Two Multivariate Distributions Based on Adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67(4), 515–530.
- Rossin-Slater, M. and M. Wüst (2020). What is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention. *American Economic Journal: Applied Economics* 12(3), 255–86.
- Schore, A. N. (2017). All Our Sons: The Developmental Neurobiology and Neuroendocrinology of Boys at Risk. *Infant Mental Health Journal* 38(1), 15–52.
- Schweinhart, L. J., H. V. Barnes, and D. P. Weikart (1993). *Significant Benefits: The HighScope Perry Preschool Study Through Age 27*. Ypsilanti, MI: HighScope Press.
- Turney, K. and A. R. Haskins (2014). Falling Behind? Children’s Early Grade Retention After Paternal Incarceration. *Sociology of Education* 87(4), 241–258.
- Walters, C. R. (2015). Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start. *American Economic Journal: Applied Economics* 7(4), 76–102.
- Weikart, D. P., J. T. Bond, and J. T. McNeil (1978). *The Ypsilanti Perry Preschool Project: Preschool Years and Longitudinal Results Through Fourth Grade*. Ypsilanti, MI: HighScope Press.
- Wright, B. R. E., A. Caspi, T. E. Moffitt, and P. A. Silva (1999). Low Self-Control, Social Bonds, and Crime: Social Causation, Social Selection, or Both? *Criminology* 37(3), 479–514.