NBER WORKING PAPER SERIES

THE DYNASTIC BENEFITS OF EARLY CHILDHOOD EDUCATION: PARTICIPANT BENEFITS AND FAMILY SPILLOVERS

Jorge Luis García
Frederik H. Bennhoff
Duncan Ermini Leaf

Working Paper 31555
http://www.nber.org/papers/w31555

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

The research in this paper is supported by the National Institute on Aging of the National Institutes of Health under award number P30AG024968, which funds the Roybal Center for Health Policy Simulation, part of the Leonard D. Schaeffer Center for Health Policy and Economics at the University of Southern California. Part of the collection of data used in this paper was supported by the National Institutes of Health under grant numbers R01HD069609 and R01AG040213, and the National Science Foundation under award numbers SES 1157698 and 1623684. The Health and Retirement Study is sponsored by the National Institute on Aging (grant number NIAU01AG009740) and is conducted by the University of Michigan. 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 or the Leonard D. Schaeffer Center for Health Policy and Economics. We thank the researchers of the HighScope Educational Research Foundation's Perry Preschool Project for access to study data and source materials. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Dynastic Benefits of Early Childhood Education: Participant Benefits and Family Spillovers
Jorge Luis García, Frederik H. Bennhoff, and Duncan Ermini Leaf
NBER Working Paper No. 31555
August 2023
JEL No. C93,H43,I28,J13

ABSTRACT

We demonstrate the social efficiency of investing in high-quality early childhood education using newly collected data from the HighScope Perry Preschool Project. The data analyzed are the longest follow-up of any randomized early childhood education program. Annual observations of participant outcomes up to midlife allow us to provide a cost-benefit analysis without relying on forecasts. Adult outcomes on the participants' children and siblings allow us to quantify spillover benefits. The program generates a benefit-cost ratio of 6.0 (p-value = 0.03). Spillover benefits increase this ratio to 7.5 (p-value = 0.00).

Jorge Luis García
John E. Walker Department of Economics
Clemson University
309-C Wilbur O. and Ann Powers Hall
Clemson, SC 29634
and NBER
jlgarci@clemson.edu

Frederik H. Bennhoff
Department of Economics
University of Zürich
Schönberggasse 1
8001 Zürich
Switzerland
frederik.bennhoff@econ.uzh.ch

Duncan Ermini Leaf
Leonard D. Schaeffer Center for Health Policy and Economics
University of Southern California
635 Downey Way
Los Angeles, CA 90089-3333
USA
dleaf@healthpolicy.usc.edu

A data appendix is available at http://www.nber.org/data-appendix/w31555

1. Introduction

We demonstrate the social efficiency of investing in high-quality early childhood education using newly collected longitudinal data up to midlife (age 54) on the participants of the iconic HighScope Perry Preschool Project (PPP). We also use newly collected data on the adult outcomes of their children and siblings. Our study has three components. First, we describe the participant outcomes at midlife and the adult outcomes of their children and siblings. Second, we provide a life-cycle cost-benefit analysis of the participants' outcomes relying on annual observations up to midlife. Third, we integrate spillover benefits on the children and siblings of the participants into the cost-benefit analysis.

Part of the description of the participant and children data in the first component of our study appears in companion work (García and Heckman, 2023; García et al., 2023). However, the description of the sibling data is new. Consolidating the data description is academically valuable: the newly collected data are the longest follow-up of any randomized early childhood education program. The data are also the first follow-up containing adult outcomes of children and siblings within the literature studying this type of programs. Previous studies provide cost-benefit analyses of PPP (i.e., Belfield et al., 2006; Heckman et al., 2010b; Rolnick and Grunewald, 2003). The second component of this study differs from these previous works in that we rely on annual observations throughout adulthood up to midlife. Previous studies rely on scattered observations (ages 19, 27, and 40) and forecasts. The third component of this study is entirely new to the literature, as no previous work integrates monetized spillover benefits on children and siblings into measures of social efficiency of early childhood education programs.

PPP was conducted in Ypsilanti, Michigan. The program randomized 123 disadvantaged African-American children into treatment and control groups across five cohorts during the early 1960s. 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.

PPP is relevant today. The disadvantaged populations it was designed to serve are still substantial. Heckman et al. (2010a) report that 15% of African-American females and 17% of African-American males born in the US during the 1960s would have been eligible to participate in PPP. The percentage of females and males eligible today is 10%.[1] PPP influences the content of current and proposed early childhood education programs. Its content is the base of 30% of Head Start programs (Elango et al., 2016). Head Start, in turn, is the main federal preschool program in the US and one of the main preschool programs analyzed in the literature.[2] PPP is fundamental in justifying the continuation and expansion of programs like Head Start (e.g., The White House, 2013, 2020, 2021). The academic and policy relevance of understanding the life-cycle benefits of PPP is thus immediate.

We estimate the present value of the program's average life-cycle total benefit at 190 thousand dollars of 2017 by monetizing education, income, crime, and health outcomes. Subtracting the total program cost per participant and the deadweight loss generated by raising taxes to fund the program yields the program's net social benefit: 160 thousand dollars with an associated benefit-cost ratio of 6.0 (p-value = 0.03). These estimates are robust when we use a variety of alternative estimators and inferential procedures, which address methodological challenges inherent to PPP and other social experiments. They explicitly account for PPP's small sample size.

Predicted spillover benefits add an average of 39 thousand dollars (p-value = 0.01) per male child of the participants and 13 thousand dollars (p-value = 0.09) per female child. The average spillover benefit to the male children, who were not directly targeted by the program, almost doubles the total program cost per participant of 21,151 dollars. Spillover benefits to the siblings of the participants are imprecisely estimated, but large in magnitude, especially

[1] This is the percentage of males and females born in households satisfying the eligibility criteria described in Section 2. We use data from US Census Bureau (2010, 2015) for this calculation.

[2] See, for example, Garces et al. (2002), Deming (2009), Ludwig and Miller (2007), Carneiro and Ginja (2014), Kline and Walters (2016), Thompson (2018), De Haan and Leuven (2020), and Bailey et al. (2021).

for male siblings. The overall pattern of the spillovers indicates substantial benefits for the extended dynasty of the program participants. The benefit-cost ratio including participant benefits and spillover benefits to their children and siblings is 7.5 (p-value = 0.00). Benefits are larger for the male participants, male children, and male siblings, which is consistent with studies documenting that boys are more sensitive than girls to adverse environments when growing up. It is also consistent with literature documenting that girls mature faster than boys, and, thus, are more resilient to adverse childhood conditions.[3]

Paper Plan. Section 2 describes PPP, the midlife follow-up and longitudinal data on its participants, and the data on the adult outcomes of their children and siblings. Section 3 discusses the quantification of the life-cycle benefits of the participants and provides a comparison to that obtained in previous studies. Section 4 discusses the quantification of the outcomes of the children and siblings of the participants. Section 5 concludes.

2. The HighScope Perry Preschool Project

PPP was a high-quality early childhood education program. Its curriculum was designed to spur development of cognitive and non-cognitive 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). The participants of the program belonged to 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 socioeconomic status were used to create a pool of 123 disadvantaged children who were randomized into treatment or control.[4]

[3] Studies in economics and other fields document similar gendered patterns (Autor et al., 2019; Bertrand and Pan, 2013; Golding and Fitzgerald, 2017; Masse and Tremblay, 1997; Nagin and Tremblay, 2001; Schore, 2017).

[4] There were three eligibility requirements for participating children (Weikart et al., 1978, pgs.15-18): 1) being African-American; 2) having low IQ at age 3 (70 to 85 points); and 3) belonging to a household scoring less than 11 in a socioeconomic index. For the calculation in the introduction, we use criteria 1) and 3). We do not use criterion 2) because it is not observed in the available data. The socioeconomic index linearly combines three variables: average years of education of the parents present in the household (weight = 0.5); father (if present) or mother employment status (weight = 2); and rooms per person in the household (weight = 2). The employment-status variable had a before-weighting score of 3 if employed in skilled job,

Table 1 summarizes implementation details of the program and describes its programmatic features. For context, it provides a comparison to the Carolina Abecedarian Project (ABC; Ramey and Campbell, 1984), another iconic high-quality early childhood education program.

Table 1. Comprehensive Details of the Perry Preschool and Carolina Abecedarian Projects

	Perry Preschool Project	Carolina Abecedarian Project
Overview		
Years Implemented	1962-1968	1972-1985
Site	Ypsilanti, Michigan	Chapel Hill, North Carolina
Population Targeted	Disadvantaged African Americans	Disadvantaged
Cohorts	5	4
Age at Entry	3	0
Duration	2.5 years	5 years
Sample	58 treatment, 65 control	58 treatment, 56 control (ABC)
Main Treatment Components		
Home Visits	4.5 per month	Not available
Center-based Care	30 weeks per year	50 week per year
	(12.5 hours per week)	(40 hours per week)
Other Treatment Components	None	Basic health check-ups and referrals
		Formula and diapers (also provided to controls)
		Nutrition
Substitutes Attended by Controls	None	Alternative preschools after age 2
		Enrolled at least for one year: 75%
Staff		
Adult-child Ratio	1:5 to 1:6	1:3 (0 to 1), 1:4.5 (1 to 4), 1:5.5 (age 4 to 5)
Teacher Certification	BA	HS graduates mixes with certified staff
Other Specialists	No	Physician, nurse, social worker
Curriculum Targets		
Cognitive Development	Yes	Yes
High-risk Behavior	No	Yes
Language Development	Yes	Yes
Motor Development	No	Yes
Non-cognitive Development	Yes	Yes
School Readiness	Yes	Yes
Task Orientation	No	Yes
Total Cost per Participant		
	21,151 (2017 USD)	95,071 (2017 USD)

Note: Details of the Perry Preschool and Carolina Abecedarian Projects (PPP and ABC). **Source:**: Barnett (1996), Schweinhart et al. (1993), and Weikart et al. (1978) for PPP and Appendix A of García et al. (2020) for ABC.

Weikart et al. (1978) report that every family that received an offer to participate in the program accepted it. Thus, we estimate the average treatment effect for program eligibles. After randomization, a few treatment and control participants were reassigned, and this generated an imbalance in baseline characteristics (see Table 2).[5] We discuss potential biases

2 if employed in semi-skilled job, and 1 if employed in unskilled job or unemployed.

[5]The 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 and even-ranked subjects assigned to separate groups; 3) Some individuals initially assigned to one group were swapped between groups to balance sex and mean socioeconomic-index scores, with average Stanford–Binet IQ scores held "more or less" constant. This generated a minor imbalance in family background variables;

and correct point estimates accordingly. Failures in implementing randomization protocols are not rare in social experiments. Authors often fail to document and address them, with substantial consequences for inference (Bruhn and McKenzie, 2009).

2.1 Midlife Follow-Up and Annual Longitudinal Data on the Participants

Table 2 provides a general description of the raw data on the participants that we use throughout the paper. Estimating the social efficiency of investing in PPP requires monetizing the outcomes observed in these data. We discuss such monetization in Section 3. We provide detailed variable definitions in Appendix Table A.1.

Panel *a* of Table 2 briefly describes the participants at baseline, based on the initial sample of 123 participants. It shows the imbalance in maternal employment that we address below. Treatment and control participants were followed up during multiple rounds of data collection through midlife. In the midlife follow-up, participants were, on average, 54 years old; 102 or 83% of the 123 initial participants were surveyed; 12% of the initial participants were not surveyed because they were deceased. Accounting for death, an outcome of its own, our results below monetize the life-cycle outcomes of 95% of the initial participants.

Panels *b* and *c* of Table 2 show that, at midlife, treatments are, on average, doing better than controls in terms of their skills and education. For example, the non-cognitive skills of the treatments are, on average, 0.5 standard deviations larger than those of the controls. Their high school graduation rate is 17 percentage points higher. Panel *c* shows that mental health impacts are also substantial. For physical health, we construct an aggregate that generally summarizes the conditions we use to monetize the life-cycle benefits in Section 3. These conditions include measures of prevalence of diabetes, being bedridden during the last year, requiring home assistance, or having been treated for substance abuse. We reverse the

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 provided to the treatment group. See Heckman et al. (2010a) and Heckman and Karapakula (2021) for more details on the randomization issues and methods to address them, which we use in this paper.

scale of the aggregate. For female participants, PPP increases this aggregate by 0.4 standard deviations (p-value = 0.01). For male participants, the increase is half as much: 0.2 (p-value = 0.23). Campbell et al. (2014) report a similar gendered gradient in the impact on health in their analysis of ABC. We elaborate on gender differences in the benefits of PPP below.

We do not directly monetize cognitive or non-cognitive skills. While increased skills may have a social value *per se*, we interpret skills as inputs of the monetized outcomes (i.e., education, income, health, and crime). Therefore, annual skill observations are not necessary for our computations. For education, we assume the levels reported at midlife are final. We further assume that the individuals obtained these levels at the standard ages at which individuals obtain schooling grades in the US and reconstruct their longitudinal trajectories. For income and crime, we construct annual longitudinal profiles up to age 54.

At ages 19, 27, 40, and 54, participants report their income in the current and preceding years. We supplement this information with rich retrospective information on employment histories. We also use administrative criminal records to set the labor income of incarcerated individuals to 0. This construction allows us to observe annual labor income up to age 54 for 86% of the treatment group and 81% of the control group.[6] Our aggregate measures of PPP's social efficiency consider the total gain in income, which breaks into transfers that the government would have provided to individuals if they had lower labor income (transfers),[7] higher state and federal taxes paid by individuals, and after-tax labor income.[8]

[6]Our baseline estimates below do not interpolate to fill in missing data; they rely solely on the available observations. Appendix Table A.14 shows that our estimates are robust to using interpolation and applying the interpolation and extrapolation method in García et al. (2020). Their method imputes missing observations by non-parametrically matching participants to individuals in the National Longitudinal Survey of the Young 1979 (Bureau of Labor Statistics, 2015) for interpolating and extrapolating (forecasting) labor income between ages 20 and 60 (the assumed age of retirement).

[7]In principle we should add the benefit of reduced cost of taxation to fund these transfers. Preliminary calculations indicate that doing so barely budges our estimates.

[8]Income increases for treatment-group members. The increase generates deadweight from adjustments to avoid taxation. The income benefits that we monetize are net of this avoidance behavior. An alternative calculation would include the income increase in the absence of taxation. This alternative would provide the "true" income benefit due to gains in skills and other forms of human capital generated by PPP. It is also important to note that if the marginal rate of substitution between leisure and consumption equals the marginal wage rate, labor income should not be counted into the program benefits. Search frictions and other frictions usually break this equality and justify quantifying income benefits.

Details on the calculation of the corresponding components are in Appendix A5.2. These calculations account for the greater marriage stability generated by treatment. Panel *d* of Table 2 describes this greater stability. It also displays the average treatment-control difference in life-cycle labor income focusing on the period where most gains appear, ages 20 to 40. In this period, the average labor income of treatment-group participants is more than 50% larger than the average of their control-group counterparts.

For crime, we collect information up to midlife using several administrative sources.[9] Any reduction in crime lowers the cost to the criminal justice system (i.e., police, court, and correctional costs). It also reduces the costs to crime victims. Below, we quantify the first component based on the (observed) number of arrests, not on the (unobserved) number of crimes committed. To illustrate the impact on crime based on our data construction, Panel *d* of Table 2 displays the average treatment-control difference in the accumulated days in jail or prison up to midlife. Though the estimate is imprecise, it indicates that, at midlife, treatment reduces an average of almost two years in prison or jail.

2.2 Data on the Children and Siblings of the Participants

The program participants provided information on their children during the midlife follow-up. Most of these children were adults by then. Their average age was 28.2 years old. Questions allowing us to construct the outcomes in Panel *a* of Table 3 were asked to them. We do not have any other information source for the outcomes of the children of the participants.

We may selectively observe the outcomes of the participants' children. First, as mentioned before, part of the sample is not observed during the midlife follow-up. Our estimators below address this issue. Second, treatment may impact the number of children participants

[9] Participants followed up at age 54 provided their consent for us to search their criminal records. The records were collected by searching the electronic systems of the Michigan State Police LEIN, county district and circuit records, Detroit Recorder's Court, and the Federal Court in Detroit. Records of a few additional criminal incidents were obtained by searching the county social services records. Records for the subjects living out of state were requested from state criminal information offices. Information on juvenile offenses was obtained from the County's juvenile court records. Prison records were collected through the Michigan Department of Corrections.

Table 2. PPP Participants at Baseline and at the Midlife Follow-up

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Female Participants			**Male Participants**			**All Participants**		
	Control Mean	Mean Difference	*p*-value	Control Mean	Mean Difference	*p*-value	Control Mean	Mean Difference	*p*-value
Panel a. Baseline									
IQ	79.577	0.463	0.818	77.846	1.366	0.431	78.538	1.031	0.374
Socioeconomics Index	8.569	0.087	0.862	8.648	0.239	0.500	8.617	0.171	0.533
Mother Does not Work	0.654	0.226	0.057	0.718	0.221	0.000	0.692	0.221	0.000
Mother's Birth Year	1960	-0.232	0.651	1960	0.210	0.598	1960	0.031	1.000
Panel b. Midlife Skills									
Cognitive	0.000	0.571	0.000	0.000	0.144	0.273	0.000	0.333	0.013
Non-Cognitive	0.000	0.601	0.000	0.000	0.416	0.036	0.000	0.499	0.000
Panel c. Midlife Education and Health									
High-School Graduate	0.400	0.470	0.000	0.667	-0.046	0.771	0.560	0.171	0.003
College Graduate	0.100	0.030	0.593	0.033	0.001	0.757	0.060	0.017	0.572
Physical Health	0.000	0.366	0.009	0.000	0.155	0.234	0.000	0.252	0.011
Mental Health	0.000	0.078	0.348	0.000	0.501	0.000	0.000	0.314	0.017
Panel d. Longitudinal Outcomes									
Married	0.281	0.077	0.242	0.298	0.046	0.266	0.291	0.059	0.153
Labor Income (2017 USD)	15,403.069	8,052.526	0.010	17,390.682	6,205.737	0.043	16,595.637	6,938.495	0.000
Household Labor Income (2017 USD)	28,158.885	15,772.945	0.000	24,173.385	8,105.323	0.050	25,767.584	11,665.389	0.000
Total Days in Jail or Prison	203.100	-180.578	0.002	2,417.167	-576.994	0.272	1,531.540	-495.328	0.210
Never Arrested	0.650	0.220	0.052	0.333	0.080	0.357	0.460	0.155	0.029
Panel e. Midlife Fertility									
Any Children	0.950	-0.080	0.606	0.733	-0.044	0.800	0.820	-0.051	0.631
Number of Children	2.950	-0.124	0.944	2.233	-0.268	0.660	2.520	-0.174	0.718
> 5 Children	0.100	-0.057	0.577	0.067	-0.032	1.000	0.080	-0.042	0.407
Age at Onset	20.895	0.505	0.734	24.136	1.464	0.446	22.634	0.866	0.479
Panel f. Midlife Siblings									
Any Siblings	1.000	0.000	1.000	0.967	-0.036	0.632	0.980	-0.018	1.000
Number of Siblings	4.650	0.350	0.601	5.467	0.464	0.709	5.140	0.379	0.626
Up to 5 Eldest	3.611	0.345	0.393	3.682	0.118	0.794	3.650	0.225	0.384
Fraction Older	0.667	0.246	0.003	0.773	-0.013	1.000	0.725	0.108	0.191
Fraction Older; < 5 Years Apart	0.667	0.116	0.378	0.682	-0.042	0.894	0.675	0.033	0.809

Note: Each panel displays the control-group mean and treatment-control average difference (mean difference) of the variable or outcome in the label. It also displays the permutation p-value associated with the null hypothesis that the mean difference is equal to 0 (Panels a, e, and f) or less than 0 (Panels b, c, and d). Panel a is based on the baseline sample of 123 participants. Panels b to f are based on the participants observed in the midlife follow-up, which has a base sample of 102 participants. Cognitive skill, non-cognitive skill, physical health, and mental health measures are principal component factors of several measures. We standardize the principal component factors using their corresponding in-sample control-group means and standard deviations. Married is the fraction of years married between ages 20 and 40, labor income is the average earnings from labor income between ages 20 and 40 (2017 USD), household labor income is the previous variable in addition to average spouse's labor income between ages 20 and 40 (in 2017 USD) if married, total days in jail or prison and never arrested are the outcomes as observed in the midlife follow-up. Fertility variables are based on questions regarding the total number of children during the midlife follow-up. The first and second sibling variables are based on questions regarding the total number of siblings during the midlife follow-up. The subsequent variables are delimited to the five eldest siblings.

have, as well as birth-timing decisions. Third, participants were only asked about their five eldest children. Panel e of Table 2 describes the fertility outcomes reported at midlife. There are no treatment-control differences in the probability of having any children, the number of children, or the probability of having more than five children. Among those who report having children, there are no sizable or statistically significant differences in the age at onset of fertility. Therefore, experimentally induced fertility or information losses due to only observing up to the five eldest children are a minor concern.

Of the 102 participants observed at the midlife follow-up, 81 report having children. These 81 participants report information on 214 children (104 children of controls, 110 children of treatments; 108 female children, 106 male children). When monetizing child outcomes below, we use the outcomes of each child as observed. For describing the outcomes in Table 3, we construct within-participant mean intergenerational outcomes. Let $\mathcal{I}$ index participants and $\mathcal{J}$ index outcomes. Define $Y_{i,j}^{c(i)}$ as the outcome $j \in \mathcal{J}$ of child $c(i)$ of participant $i \in \mathcal{I}$. The mean outcome j for the children of i is

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

where $\mathcal{C}_i$ indexes the children of participant i and "#" denotes cardinality. $\bar{Y}_{i,j}^{c}$ is the outcome for the "average child" of i. When constructing each outcome j, we consider children $c(i)$ who satisfy the age thresholds indicated in Panel a of Table 3. Imposing these thresholds improves the interpretability of the outcomes.

Panel a of Table 3 summarizes the intergenerational outcomes. Detailed variable definitions are in Appendix Table A.2. The panel indicates substantial intergenerational spillovers. García et al. (2023) and García and Heckman (2023) discuss these intergenerational outcomes. They do not monetize them. García and Heckman (2023) document that the beneficial intergenerational impact of PPP is such that the intergenerational correlation of education and crime, which is substantial for the controls, disappears for the treatments.

Table 3. Spillover Outcomes: Children and Siblings of the Participants of PPP

	(1)	(2)	(3)	(4)	(5)	(6)
	Female Children/Siblings			**Male Children/Siblings**		
	Control Mean	Mean Difference	p-value	Control Mean	Mean Difference	p-value
Panel a. Children						
Age	29.105	-1.238	0.426	27.962	0.707	0.734
High-School Graduate (Age $\geq$ 18)	0.742	0.133	0.026	0.672	-0.015	0.582
College Graduate (Age $\geq$ 23)	0.311	-0.087	0.846	0.036	0.078	0.063
Employed (Age $\geq$ 23)	0.411	0.089	0.218	0.478	0.189	0.040
Never Arrested (Age $\geq$ 18)	0.776	0.060	0.210	0.369	0.137	0.089
In Good Health (Age $\geq$ 18)	0.849	0.101	0.030	0.817	0.122	0.006
Not a Parent (14 $\leq$ Age $\leq$ 22)	0.833	0.121	0.234	1.000	0.000	1.000
Never Divorced (Age $\geq$ 23)	0.856	0.106	0.016	0.926	0.074	0.028
Panel b. Siblings						
Age	55.420	-0.990	0.273	56.903	-1.745	0.043
High-School Graduate	0.853	0.004	0.494	0.867	0.068	0.173
College Graduate	0.441	-0.132	0.833	0.267	0.125	0.157
Employed	0.531	-0.248	0.998	0.347	0.189	0.074
Employed or Retired	0.625	-0.067	0.721	0.514	0.261	0.000
Never Arrested	0.882	-0.144	0.927	0.393	0.201	0.043
In Good Health	0.647	0.035	0.384	0.544	0.166	0.071

Note: Panel *a* displays the control-group mean and treatment-control mean difference (MD) for the intergenerational outcome in the label. Intergenerational outcomes are for the average child. We construct them by averaging within program participants across up to their five eldest children. For each mean difference, we present the permutation p-value associated with the null hypothesis that the mean difference is equal to 0 (age) or less than or equal to 0 (all other outcomes). Panel *b* is analogous in format to Panel *a* for the intragenerational outcomes. We construct them by averaging within program participants across up to five eldest siblings who are both older than the participant and at most five years apart in age.

Participants also answered questions about their siblings.[10] These questions had the same format as the questions about their children. We only analyze the outcomes of the older "pre-treatment" siblings who were born before PPP began. Focusing on the older siblings avoids having to deal with impacts of the program on fertility and subsequent family planning.[11] Out of these pre-treatment siblings, we do not consider siblings who are more than five years older than the participants.[12] Information reported on the siblings is essentially complete (education and crime careers are typically completed by midlife), although not longitudinal. Using a construction analogous to that based on Equation (1), we construct within-participant mean intragenerational outcomes and describe outcomes for the "average sibling" in Panel *b* of Table 3. This paper is the first to present the spillover impacts on the siblings of the participants of PPP.

In practice, PPP did not impact the number of "post-treatment" siblings the participants had (see Table 2). We still leave post-treatment siblings out of the analysis out of extreme caution. PPP did have an impact on other parental-decision margins. García and Heckman (2023) document that the treatment-group children received more parental investment than their control-group counterparts, as measured by the Parental Attitude Research Instrument (PARI, Loewenstein, 1973). More generally, there is evidence indicating that, for the parents of most children, the childcare services provided by early childhood education programs are a complement of their parental investment (e.g., Dougan et al., 2022; García et al., 2018; Gelber and Isen, 2013; Heckman and Mosso, 2014).[13]

[10]The same questions were asked about children and siblings. However, item non-response does not allow us to construct fertility and marriage variables for the siblings.

[11]Some participants belonging to the first couple of cohorts have younger siblings who are original participants themselves in the later cohorts. In Section 4, we only count the benefits on "pre-treatment" siblings once per household, independently of the number of participants in that household.

[12]We do not consider these older siblings because, in our monetization of sibling spillovers, we aim to avoid forecasting outcomes out of sample for individuals over 60 years old. Using this criterion, we drop nineteen treatment-group siblings and nine control-group siblings. We monetize the outcomes of 58 male siblings of the participants and 50 female siblings.

[13]These results on the joint determination of fertility and child services such as parental investment speak to a rich literature on the joint determination of "quality" and "quantity" of children (e.g., Becker and Lewis, 1973) and the simultaneity of fertility and parental investments concerning child education, health, and other outcomes (e.g., Ehrlich and Yin, 2013).

Panel b of Table 3 shows that, for male siblings, the impacts are substantial. There are spillover impacts on education, employment, never being arrested, and being in good health. We follow García et al. (2023) and focus on examining "employment" to further interpret our spillover estimates. We first compare the impact on the average child of the participants to the impact on the participants themselves and to the impact of Head Start. Heckman and Karapakula (2021) report that PPP increases the probability of being employed at age 40 for the male participants by 26.6 percentage points (p-value = 0.02). We find that PPP increases the probability of being employed for the average male child by 18.9 percentage points (p-value = 0.040) and for the average male sibling by the same amount (p-value = 0.040). The spillover impact on the probability of participants' male children and siblings not being idle during young adulthood. This spillover impact is larger than the corresponding impact of Head Start on its participants of 10.0 (s.e. 4.9) percentage points (Deming, 2009).

For the average female child of the participants, there is also a positive impact on the probability of being employed of 8.9 percentage points. However, it is not significant (p-value = 0.22). The impact on the employment probability of the female participants reported in Heckman and Karapakula (2021) is close to 0: -1.6 percentage points (p-value = 0.50), and the impact on the female siblings is negative and sizable: -25 percentage points (p-value = 0.99). The employment variable is 1 when individuals are employed and 0 when they are retired, unemployed, or out of the labor force. Because siblings are older, retirement may be a positive outcome. It may mean that they can retire early due to the availability of savings. When considering being employed or retired as an outcome, the impact on female siblings becomes less negative: -6.7 percentage points (p-value = 0.721).

Women are less attached to the labor force than men (Goldin et al., 2022). Therefore, for female children, who were around their 30s when their parents were interviewed at midlife, it is expected that impacts on employment would be smaller than those on their male counterparts. The non-positive impacts for participants and their siblings could seem at odds with the program benefiting female participants and their siblings. However, it is essential to note that several outcomes are positive for female participants (see Table 2). More generally, selection on unobservables has evolved since the 1970s in a way such that women of the highest (unobserved) ability have transitioned from being more likely to select out of the labor force (as well as to select out of obtaining higher levels of education) to select in (Mulligan and Rubinstein, 2008). This change has been gradual. Older cohorts (e.g., siblings of female participants) tended to select out more than subsequently younger cohorts (female participants and their children). Thus, the negative impact on education and employment outcomes is not as surprising for the relatively old siblings of the participants.[14]

3. Quantifying the Life-Cycle Benefits and Costs of the Participants

As before, let $\mathcal{J}$ index outcomes (in this case, the set of monetized outcomes for the participants: education, income, crime, and health). The discounted (to the date of initial program participation) average life-cycle benefit or present value of treatment from outcome $j \in \mathcal{J}$ is

$$\Pi_j := \sum_{a \in \mathcal{A}} \beta^{a-3} \cdot \mathbb{E}\left[Y^1_{j,a} - Y^0_{j,a}\right], \tag{2}$$

where $\mathcal{A}$ indexes age and β is the discount factor associated with a discount rate ρ. $Y^d_{j,a}$ is the monetary value of outcome $j \in \mathcal{J}$ at age $a \in \mathcal{A}$ when treatment status is fixed to $d = 0$ (control) or $d = 1$ (treatment). $Y^d_{j,a}$ totals the monetary value of the outcome across subcategories (e.g., across different types of crime). $\mathbb{E}\left[Y^1_{j,a} - Y^0_{j,a}\right]$ is the average treatment effect for outcome $j \in \mathcal{J}$ at age $a \in \mathcal{A}$. Income is reported in dollars. We monetize other outcomes using procedures described below. Equation (2) is the life-cycle present value that outcome $j \in \mathcal{J}$ generates. The present value of the program's life-cycle total benefit, net

[14] Our monetization accounts for these negative impacts as given. Mulligan and Rubinstein (2008) explain that the relatively older cohorts of high-ability women who tended to select out of the labor force likely specialized in household production. Limitations in the data on the siblings of the participants do not allow us to account for potential gains through the marriage market, which could offset the negative impact on education and employment. Of course, the negative impact on the probability of never being arrested is not necessarily explained by this selection phenomenon. Crimes committed by women are generally minor so the monetized value of criminal activity by female siblings of participants ends up being negligible.

social benefit (NSB), and benefit-cost ratio (BCR) are

$$\Pi := \sum_{j \in \mathcal{J}} \Pi_j \tag{3}$$

$$\text{NSB} := \Pi - C \tag{4}$$

$$\text{BCR} := \frac{\Pi}{C}, \tag{5}$$

where C is the total program cost per participant. Appendix A4 discusses other measures of social efficiency (e.g., the marginal value of public funds Hendren and Sprung-Keyser, 2020).

We estimate $\mathbb{E}\left[Y^1_{j,a} - Y^0_{j,a}\right]$ using age-wise regression-adjusted treatment-control mean differences (OLS), based on the baseline variables in Table 2.[15] Appendix A2 provides details on our estimator, as well as its alternatives. We test whether the net social benefit is positive (i.e., if the average life-cycle total benefit of the program less its total cost per participant is greater than zero), which is equivalent to testing if the benefit-cost ratio is greater than or equal to one. Our main inferences are based on bias-corrected accelerated (BCA) bootstrap p-values (Efron, 1987), corresponding to our one-sided hypotheses. Appendix A3 provides additional details on our inference procedures.

3.1 Major Sources of Benefits Across the Life Cycle

For education, we monetize essentially full records, including K-12 outcomes. PPP reduced participation in special education and K-12 grade retention and increased high school graduation rates, especially for female participants. The benefit due to the decrease in treatment-group special education and K-12 grade retention is positive. It is essentially offset by the

[15] This adjustment accounts for the randomization compromises discussed in Section 2. The randomization compromises would attenuate our estimates towards 0 if working mothers who induced the treatment-status swaps and their children had relatively high unobserved ability. This bias would be a standard case of ability bias (e.g., Blackburn and Neumark, 1993). The attenuation would be reinforced if attrition and missing data were more prevalent in the treatment group due to unexplained factors. Indeed, the treatment-group participants are more likely to be observed in the midlife follow-up than their control-group counterparts. García et al. (2023) use the bounding method in Lee (2009) to detect whether attenuation due to greater re-contact with the treatment group during the midlife follow-up biases OLS estimates as the ones we use in this paper. They do not find evidence of such biases.

cost due to the increase in treatment-group formal schooling.

Figure 1 summarizes the cumulative present value of the other major sources of the program's benefits.[16] For income and crime, our monetization starts at age 16. For health, it starts at age 18. We do not display education costs or health benefits before age 21 graphically because their contribution is negligible. We display the cumulative benefit by age segments, together with the accumulated benefit at each segment's endpoint. We provide extensive details on the construction of each component in Appendices A5.1 (education), A5.2 (income), A5.3 (crime),[17] and A5.4 (health).[18]

Figure 1a shows that the cumulative present value due to labor income for men is negligible before age 26. After that, there is an increase up to age 40 that reaches 57.6 thousand dollars. Education and criminal activity are two potential explanations. The first explana-

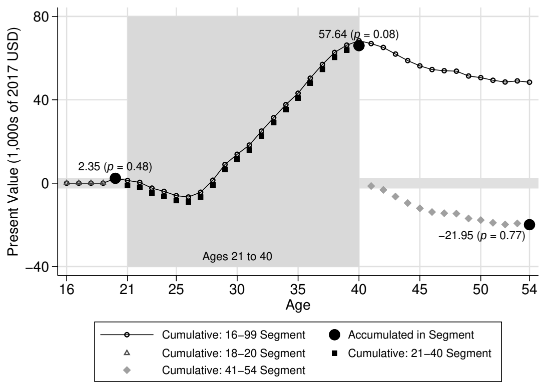
[16]We set benefits to 0 after participants die (which we observe up to the midlife follow up), or, in the case of the health forecasts, after we predict that they die or turn 99 years old.

[17]Crime reduction generates a substantial benefit. It lowers the cost to the criminal justice system. It also reduces the costs to crime victims. We quantify the first component based on the (observed) number of arrests, not on the (unobserved) number of crimes committed. This biases downward the estimated benefit from crime reduction because the treatment-induced reduction in crimes is likely larger than the treatment-induced reduction in arrests made for those crimes. We quantify the second component by inflating the number of arrests to address the disparity between the number of arrests and the number of crimes committed, which is standard practice in the criminology literature. We then quantify the material losses to the crime victims (property and medical bills). We include costs of the criminal justice system and the monetary costs to crime victims, which we document thoroughly in Appendix A5.3, in the main paper. We exclude the quality-adjusted life year (QALY) cost to victims in the main paper. We provide estimates that include this subjective cost in Appendix Table A.13.

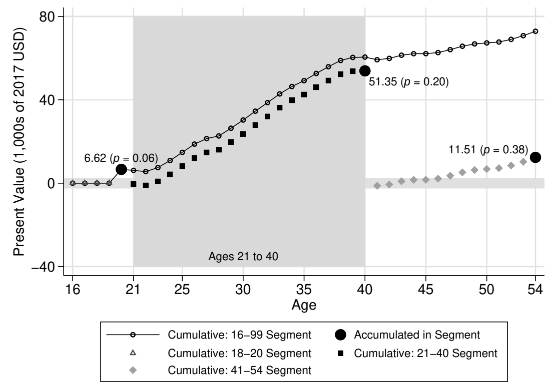
[18]Treatment effects on health are expected to substantially materialize after age 54 due to lasting improvements in treatment-group health behavior. Health conditions are unlike labor income or crime, for which substantial treatment effects appear before age 50 (see Figures 1e and 1f). We monetize health outcomes from age 18 to (realized or forecasted) death. Health is the only outcome we model and forecast. In sum, our monetization adapts well-established health forecasting models as follows. First, we estimate health-transition probability models in nationally representative datasets. Second, based on these estimates, we predict probabilities for the PPP participants based on their observed health conditions during the age 40 and 54 follow-ups. Those conditions are summarized in Table 2. Third, we map the predicted transition probabilities into quality-adjusted life years. In a given year, the average treatment-control difference in a quality-adjusted life year is the benefit from improved health. Fourth, we net out from this benefit medical costs, which we predict annually using health-focused nationally representative datasets. Additional details are in Appendix A5.4. García and Heckman (2020) study in detail models for forecasting health benefits of early childhood education programs. Our calculations of the health benefits do not account for the potential impact of health on the marginal utility of consumption. Introducing consumption dynamics into the health models is complex and would introduce additional uncertainty, mainly because we do not observe consumption for the participants of PPP. That said, our health forecasts allow the health-state transition probabilities, which we then monetize, to be a function of income and employment status.

Figure 1. Cumulative Present Value of Benefits Accrued by the Participants of PPP

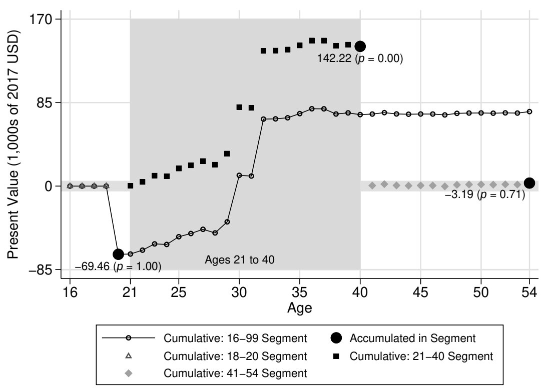
(a) Labor Income, Male Participants



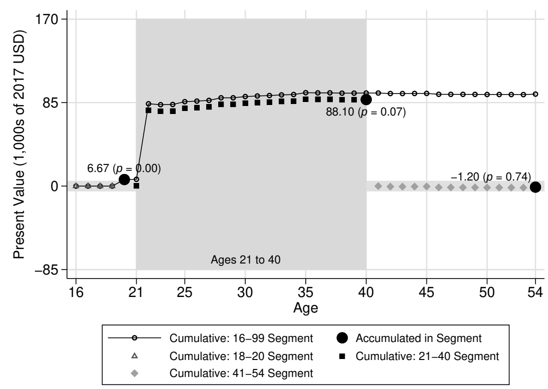
(b) Labor Income, Female Participants



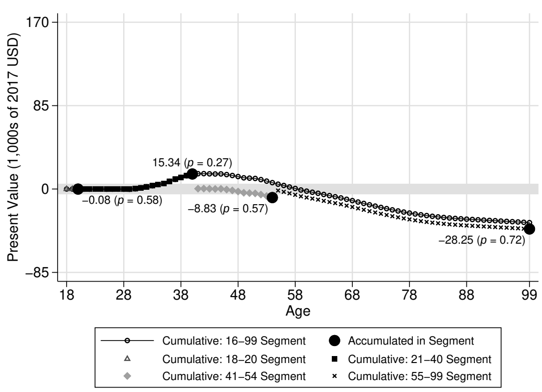
(c) Crime, Male Participants



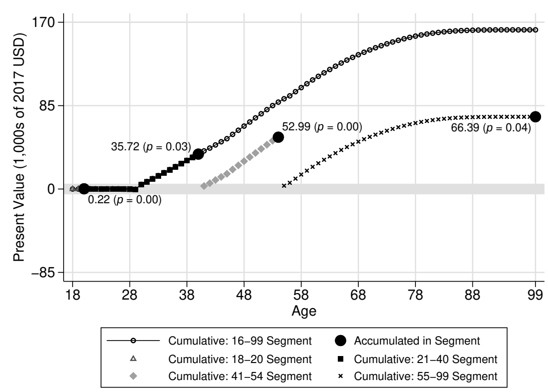
(d) Crime, Female Participants



(e) Health, Male Participants



(f) Health, Female Participants



Note: Panel (a) shows the average life-cycle treatment-control difference or present value of the Perry Preschool Project due to labor income for the male participants. The present value is shown as cumulative for the labeled age segments. We also present the accumulated present value by segment together with its p-value for segments after age 21. The null hypothesis for the accumulated present value is that it is less than or equal to 0. We use the OLS estimator discussed in Section 3 and discount to the year in which the program started using a 3% rate. We display bias-corrected accelerated bootstrap p-values. Panel (b) to (f) are analogous in format to Panel (a) for the the outcomes and participants in the label. We do not monetize outcomes after age 54 for labor income and crime. For health, we monetize outcomes until (realized or forecasted) death or age 99.

tion lacks empirical support: Table 2 indicates a relatively small and statistically insignificant impact on education for males. The second explanation is plausible: treatment has a sizable negative impact on criminal activity at midlife. Education and criminal careers end around age 40, making treatment-control differences measured at midlife valid explanations of the life-cycle pattern in Figure 1a. Criminal careers are in their prime during the late 20s and early 30s (Farrington, 1986). This coincides with the age range when the labor-income gap opens between treatment and control males, suggesting that the second explanation is the primary driver of the pattern in the figure.

Figure 1b shows that the cumulative present value due to labor income for women differs statistically from 0, beginning at age 20. It continues to increase up to age 40. Between ages 21 and 40, the cumulative present value amounts to 51 thousand dollars. Education is the primary plausible explanation for women. The treatment-group high-school graduation rate is 47 percentage points higher than the control-group rate of 40%. Additionally, the pattern of the labor-income treatment effect for women starts increasing soon after the education treatment effect realizes.

Figure 1c validates our interpretation of how the life-cycle pattern of the labor-income treatment effect emerges. The present value due to crime for men starts increasing right around the same age as the present value due to income. For crime, the pattern is spiked because costly crimes occur in scattered years. These costly crimes involve long sentences, which incapacitate control-group participants from earning labor income. The present value due to crime is negative between ages 16 and 21. This is mainly due to a couple of crimes that are very costly in our quantification (violent rapes). The increase in the present value from ages 21 to 40 is due to a variety of crimes committed by several individuals in the control group—in the age-54 follow-up, 67% of the control-group males had been arrested for at least one time. The treatment-group rate is 8 percentage points lower. Figure 1d indicates that the positive present value due to female crime is driven by an isolated increase at age 22, which is due to a recorded arrest for violent murder by a control-group individual.

The present value due to female crime is otherwise minor.

Figures 1a to 1d indicate that the life-cycle benefits from income and crime are minor after age 40. We thus limit our life-cycle analysis to ages 16 to 40 for these two outcomes in Table 4, as well as in the remainder of the paper (note that these are the crucial years for childrearing). The empirical bootstrap distributions of our estimates indicate that adding data from the segment after age 40 increases the variance of our estimates, while barely changing the point estimates. Least absolute-value shrinkage and selection operator (LASSO) generalizations of Equation (A.1) further justify limiting the life cycle to ages 16 to 40 for income and crime (see Appendix A2). The segment between ages 16 and 40 includes a substantial negative component due to male crime before age 20, generated by a couple of isolated crimes. It also accounts for the increase in the female crime present value due to an isolated violent murder. These two components, due to isolated events, essentially offset each other in the computation of our preferred estimates in Table 4.

Figures 1e and 1f indicate that women drive the present value of health. To forecast health, we estimate health transitions in nationally representative datasets, which we apply to predict the transitions in the PPP sample after several harmonization steps across several data sources. Initialization of the predicted transitions is based on the health conditions generally summarized by the health aggregates in Table 2. The health forecasts include benefits (quality-adjusted life year gains given annual better overall health) net of medical costs—consistent with the literature, our forecasts predict that higher-income individuals (those in the treatment groups) incur higher governmental and private medical expenditures (Pashchenko and Porapakkarm, 2016).

Table 4 presents our baseline OLS estimates for the pooled sample of male and female participants. It aggregates the life-cycle trajectories in Figure 1 into present values. Results are reported in 2017 US dollars and are discounted to the year in which the program started using a 3% rate. For the net social benefit and benefit-cost ratio, we present two estimates.

The first estimate sets the cost to the total program cost per participant of 21,151. The second estimate multiplies this cost by 1.5, thereby subtracting from the net social benefit the welfare loss that would be generated by collecting the taxes required to fund the program. The second estimate assumes a deadweight loss of 50 cents per dollar, which is on the upper end of the range reported in the literature (Feldstein, 1999).

PPP generates a present value of the program's life-cycle total benefit of 189.9 thousand dollars, which differs statistically from 0 at a 5% level. This present value is generated by an increase of 18% in the average present value for the treatment group, relative to its control-group counterpart. Subtracting the cost from the total benefit yields the net social benefit. In the baseline cost scenario, we estimate a net social benefit of 168.7 thousand dollars. Accounting for the deadweight loss, we estimate a net social benefit of 158.2. Both estimates differ from 0 statistically at a 5% level. Their corresponding benefit-cost ratios, which are also statistically significant, indicate that the program generates between 6.0 and 9.0 dollars of benefit per dollar of cost. The former estimate of 6.0 is our preferred baseline estimate.[19]

Elango et al. (2016) document gender differences in the impact of early childhood education programs. There are various explanations for the gendered impacts. They all point towards boys benefiting more than girls from a supplemented environment while growing up. Baker et al. (2008, 2015) find harmful impacts of a low-quality universal childcare policy. Kottelenberg and Lehrer (2014) show that this negative impact is driven by boys, who appear less resilient than girls. Putting boys in low-quality environments instead of keeping them at home hurts them. This explanation is consistent with literature in other fields, which supports a greater vulnerability of boys to adverse environments (Golding and Fitzgerald, 2017; Schore, 2017). Related studies document that boys develop later than girls and thus benefit from an enriched environment (Bertrand and Pan, 2013; Lavigueur et al.,

[19]Appendix Table A.15 shows that our baseline estimates in Table 4 are robust to the use of alternative estimators and inference procedures. Appendix Table A.16 further analyzes the sensitivity of our estimates to variation in externally supplied parameters that we employ to estimate benefits (e.g., school-year or incarceration costs, value of a statistical life, discount factor).

Table 4. Life-Cycle Present Value, Net Social Benefit, and Benefit-Cost Ratio for the Participants of PPP

Present Values in 1,000s of 2017 USD	**Estimate**	(%Δ)
Education Present Value		
Total	0.27	(-.29%)
Income Present Value		
Transfers†	5.90	
Federal Taxes	10.96	
State Taxes	2.74	
After-Tax Labor Income	43.43	
Total	**61.58**	(50%)
Crime Present Value		
Criminal Justice System Cost	19.21	
Monetary Cost to Victims	60.13	
Total	***79.34***	(-47%)
Health Present Value		
Government Expenditure	-2.00	
Private Expenditure	-9.00	
Quality-Adjusted Life Years	59.66	
Total	48.69	(4.1%)
Total Present Value (Π)	***189.88***	(18%)
	[43.40 , 409.48]	
Net Social Benefit $(\boldsymbol{\Pi - C})$		
Baseline Program Cost	***168.73***	
	[22.25 , 388.33]	
Subtract Deadweight Loss (50%)	***158.15***	
	[11.68 , 377.75]	
Benefit-Cost Ratio $\left(\frac{\boldsymbol{\Pi}}{\boldsymbol{C}}\right)$		
Baseline Program Cost	***8.98***	
	[2.05 , 19.36]	
Subtract Deadweight Loss (50%)	***5.98***	
	[1.37 , 12.91]	

Note: This table displays the net present value due to the benefits of the four components that we quantify. It then displays our preferred estimate of the present value of the program's life-cycle total benefits—Π in Equation (3)—and the corresponding estimates of the net social benefit and the benefit-cost ratio—as defined in Equations (4) and (5). We display two-sided bias-corrected accelerated bootstrap 90% confidence intervals in brackets, clustered at the household level. We bold (italicize) estimates when they are significant at the 10% (5%) based on bias-corrected accelerated bootstrap p-values. The null hypothesis for the total present value is that it is less than or equal to 0. The null hypothesis for the net social benefit is the same. The null hypothesis for the benefit-cost ratio is that it is less than or equal to 1. The estimates rely on the OLS estimator explained in Section 3, which adjusts for compromises in the randomization protocol, attrition, and item non-response. The estimates are discounted to the year in which the program started using a rate of 3%. We show the net social benefit and the benefit-cost ratio using the baseline program cost (21,151 of 2017 US dollars), as well as using the baseline program cost multiplied by 1.5 to account for the deadweight that would be generated by collecting the taxes required to fund the program.
%Δ: For the total present value, we show in parentheses the percentage change in the average present value for the treatment group relative to the average present value for the control group.
†Transfers that the government would have provided to individuals had they not increased their labor income due to treatment. This component is decomposed from the observed before-tax labor income, not counted as an additional gain.

1995; Masse and Tremblay, 1997; Nagin and Tremblay, 2001). Autor et al. (2019) show that household economic shocks affect boys more strongly than girls.

3.2 Comparison to Previous Studies

Table 5 compares the present value of the major sources of benefits we obtain for the pooled sample of male and female participants to counterparts obtained in the literature. It also compares the net social benefit and the benefit-cost ratio. There are two aspects to consider when comparing the results from the different studies: observation span and frequency. Rolnick and Grunewald (2003) report benefits relying on observation for ages 3 to 27 and on forecasts between ages 28 and 65. Belfield et al. (2006) and Heckman et al. (2010b) rely on observation for ages 3 to 40 and on forecast for ages 40 to 65. Differently from us, the three studies rely on scattered rather than annual observations.

For illustration, we compare the benefits due to income we obtain to those reported by in Heckman et al. (2010b). Heckman et al. (2010b) rely on scattered observations at ages 19, 27, and 40. To interpolate (i.e., obtain annual observations for the years between 19 and 40 for which there are no annual observations), they estimate income dynamics models in observational nationally representative datasets. They apply these models to predict the income of the treatment and control participants and obtain the corresponding treatment-control differences. To extrapolate, they proceed similarly. They overestimate the benefits from income by 36 thousand dollars. This overestimation is not trivial. It amounts to 1.7 times the total cost of the program per participant.

In their study of forecast methods to quantify the benefits of social programs when annual observations are lacking, García et al. (2020) show that methods like those used by Heckman et al. (2010b) often miss that the inflection point in income profiles is around age 40 for relatively disadvantaged individuals. They show that this inflection point is relatively early in life, even for treatment groups (who remain relatively disadvantaged relative to the overall population, even after treatment). They also show that profiles for relatively disadvantaged individuals are often flatter than those of the average individual in nationally representative datasets. Heckman et al. (2010b) do not account for the earlier inflection point and flatter profile, which are observed when using annual data (see Figure 1). In the case of PPP, the main reason for the earlier inflection point and flatter profile is that a large fraction of the benefits due to income come from reductions in criminal activity (and, therefore, availability for being employed), especially for men. Criminal careers usually end by age 30, so these benefits gradually taper off (some control-group individuals remain incarcerated for several years without income during their 20s and 30s). Indeed, after age 40, the gain for male participants of PPP becomes negative, though statistically insignificant.

While the aggregate benefit measures in Rolnick and Grunewald (2003) and Heckman et al. (2010b) (i.e., their estimates of the net social benefit and benefit-cost ratio) closely align with ours, the present values by major source of benefit underscore that such alignment is coincidental. These studies either overestimate crime or income benefits.[20] Our aggregate estimates align with theirs after we account for health benefits, which, in practice, they set to 0. Of course, once we add the child and sibling spillovers discussed below, our benefit-cost ratios increase in magnitude. They also increase in statistical precision, providing more certainty regarding the degree of social efficiency of investing in high-quality early childhood education programs.

4. Quantifying Intergenerational and Intragenerational Spillover Benefits

The results so far indicate that the treated individuals can provide better environments as parents. The literature suggests that their male children receive the greatest benefits, as they are most sensitive to adverse environments during childhood (e.g., Autor et al., 2019). We monetize the intergenerational program impacts next. We focus on predicting education, income, and crime present values. Limited information makes it impossible to

[20] The estimate of crime benefits in Heckman et al. (2010b) is almost identical to ours. However, they include QALY costs to victims, and we do not. When including QALY costs, our baseline benefit-cost ratio of 5.98 reported in Table 4 increases to 11.48 (see Appendix Table A.13). Again, the alignment between our study and theirs is coincidental.

Table 5. Survey of Estimates of Social-Efficiency Criteria for the Participants PPP

Present Value in 1,000s of 2017 USD	(1)	(2)	(3)	(4)
	Rolnick and Grunewald (2003)	Belfield et al. (2006)	Heckman et al. (2010b)	This Paper
Benefits				
Education	10.8	10.0	5.2	0.3
Income (from labor and transfers)	52.4	91.0	97.3	61.6
Crime	120.4	239.1	79.5	79.3
Health				48.7
Total	183.6	340.1	182.0	189.9
Total Program Cost	21.2	21.1	21.2	21.2
Net Social Benefit	162.4	319.0	160.8	168.7
Benefit-Cost Ratio				
Baseline	8.7	16.1	9.2	9.0
			[3.2 , 14.9]	[2.1 , 19.4]
with Deadweight Loss (DL)			6.6	6.0
			[2.1 , 11.1]	[1.4 , 12.9]
with DL and Child and Sibling Spillovers				7.5
				[4.7 , 11.5]

Note: Column (1) displays the average present-value benefits, total program cost, net social benefit, and benefit-cost ratio as reported in the source labeled. Estimates are present values in 1,000s of 2017 US dollars, discounting to the year in which the program started with a discount rate of 3%. When a benefit is left blank, it is not quantified by the authors and thus assumed to be 0 in the corresponding calculation. When a benefit-cost ratio or [90% confidence interval] is left blank, it is not reported by the authors. Columns (2) to (4) are analogous in format to Column (1) for the sources labeled. In Columns (3) and (4), costs incurred by the government are multiplied by 1.5 in the calculation with deadweight loss, thus assuming a deadweight loss of 50 cents per dollar of governmental expenditure. Details on the estimates in Column (4) are discussed throughout this paper. Benefit-cost ratios accounting for spillovers include average present-value benefits on up to the five eldest children of the program participants, as well as up to five eldest siblings both older than the program participants and at most five years apart in age.

quantify intergenerational health benefits. We also monetize the benefits to the siblings of the participants, focusing on education, income, and crime for the same reasons.

We extend the frameworks of Heckman et al. (2013) and García et al. (2020) to use mediation analysis as a prediction tool. Prediction of child and sibling life-cycle total spillover benefits is an exploratory exercise supplementing our main analysis. For two reasons, it requires several strong assumptions not previously invoked. First, we have only a single observation of child and sibling outcomes (obtained from the age-54 interview with the participants). Data cannot be validated as carefully as it is possible to do with the panel observations on the participants. Second, children and siblings of participants were born in different years. We thus require age adjustments. Predictions are prone to bias due to cohort effects.

4.1 Intergenerational Spillover Benefits

Let $\boldsymbol{\mathcal{Y}}$ denote an individual's vector of present values due to the monetized outcomes (i.e., education, income, crime).[21] This individual could either be in the treatment or control group. Our goal is to predict the average treatment-control difference in $\boldsymbol{\mathcal{Y}}$, referred to as present value throughout the paper, for the children of the participants. Let $\boldsymbol{Y}$ denote a vector of predictors of $\boldsymbol{\mathcal{Y}}$ measured at early adulthood, which are potentially affected by treatment. We use the superindices h, p, and c for variables observed for households, participants, and children of participants, respectively. We use the same letters for subindexing the variance clustering of error terms. We postulate the relationships

$$\boldsymbol{\mathcal{Y}}^{\mathrm{p}} = \boldsymbol{\chi}\boldsymbol{Y}^{\mathrm{p}} + \boldsymbol{\varepsilon}_{\mathrm{ph}} \quad (6)$$

$$\boldsymbol{\mathcal{Y}}^{\mathrm{c}} = \boldsymbol{\chi}\boldsymbol{Y}^{\mathrm{c}} + \boldsymbol{\varepsilon}_{\mathrm{cph}}, \quad (7)$$

[21]When monetizing the present value due to education of the children of the participants, we update the education costs (see Appendix A5.1.3 for details). When monetizing the present value due to income, we adjust for real wage growth using a 0.98% per annum rate calculated from US Census Bureau (2020). We do not perform adjustments for crime because the information is limited to what we use for the participants. We do not perform any adjustments when monetizing sibling outcomes in Section 4.2.

where $\boldsymbol{\varepsilon}_{\text{ph}}$ and $\boldsymbol{\varepsilon}_{\text{cph}}$ are mean-independent error terms by assumption. $\boldsymbol{\chi}$ is a matrix of coefficients indicating the relevant elements in $\boldsymbol{Y}$ for predicting $\boldsymbol{\mathcal{Y}}$ (predictors may differ by outcome). These relationships assume structural invariance of $\boldsymbol{\chi}$ across generations and experimental groups. We observe $\boldsymbol{\mathcal{Y}}^{\text{p}}$, $\boldsymbol{Y}^{\text{p}}$, and $\boldsymbol{Y}^{\text{c}}$. We do not observe $\boldsymbol{\mathcal{Y}}^{\text{c}}$.

We assume that a vector $\boldsymbol{\theta}^{\text{p}}$ mediates the predictors of the intergenerational present values:

$$\boldsymbol{Y}^{\text{c}} = \boldsymbol{\Gamma}\boldsymbol{\theta}^{\text{p}} + \boldsymbol{f}(a) + \boldsymbol{\eta}_{\text{cph}}, \tag{8}$$

where $\boldsymbol{\Gamma}$ is a matrix of coefficients and $\boldsymbol{\eta}_{\text{cph}}$ is an error-term vector, which is independent of treatment status D by assumption (i.e., mediators $\boldsymbol{\theta}^{\text{p}}$ fully explain the treatment effect on predictors $\boldsymbol{Y}^{\text{c}}$). The term $\boldsymbol{f}(a)$ is a matrix of second-degree age polynomials that age-adjust $\boldsymbol{Y}^{\text{c}}$ so that it is measured at the same age as $\boldsymbol{Y}^{\text{p}}$ (participants have children of different ages). The system of mediators is

$$\boldsymbol{\theta}^{\text{p}} = \boldsymbol{\Delta}D + \boldsymbol{Q}\boldsymbol{W}^{\text{p}} + \boldsymbol{\upsilon}_{\text{h}}, \tag{9}$$

where $\boldsymbol{\Delta}$ is a vector of treatment-effect coefficients on the mediators. $\boldsymbol{Q}$ is a matrix of coefficients, $\boldsymbol{W}^{\text{p}}$ is the vector of the baseline variables used in our estimations of Section 3, and $\boldsymbol{\upsilon}_{\text{h}}$ is an error-term vector, which is conditionally independent of D by assumption (i.e., $\boldsymbol{\upsilon}_{\text{h}} \perp\!\!\!\perp D|\boldsymbol{W}^{\text{ph}}$).

We combine Equations (6) to (9) using standard properties of conditional expectations to obtain the reduced form of the prediction model:

$$\hat{\boldsymbol{\mathcal{Y}}}^{\text{c}} = \underbrace{(\boldsymbol{\chi}\boldsymbol{\Gamma}\boldsymbol{\Delta})}_{=:\boldsymbol{\pi}} D + \boldsymbol{\chi}\boldsymbol{\Gamma}\left(\boldsymbol{\theta}^{\text{p}} - \hat{\boldsymbol{\Delta}}D\right) + \boldsymbol{\chi}\boldsymbol{f}(a) + \boldsymbol{\xi}_{\text{cph}}, \tag{10}$$

where $\hat{\boldsymbol{\mathcal{Y}}}^{\text{c}}$ is the prediction of $\boldsymbol{\mathcal{Y}}^{\text{c}}$ based on the estimate of $\boldsymbol{\chi}$ from Equation (6) and the prediction of $\boldsymbol{Y}^{\mathrm{c}}$ from Equation (8), $\left(\boldsymbol{\theta}^{\mathrm{p}} - \hat{\boldsymbol{\Delta}} D\right)$ is the vector of mediators residualized from treatment, and $\boldsymbol{\xi}_{\mathrm{cph}}$ is an error term that converges in probability to 0. $\boldsymbol{\pi}$ is the object of interest (i.e., the predicted treatment-control difference in $\boldsymbol{\mathcal{Y}}^{\mathrm{c}}$ or predicted present value due to the relevant intergenerational outcome). In this framework, $\boldsymbol{\pi}$ summarizes the intergenerational transmission of the program into education, income, and crime.

We estimate Equation (6) using the outcome present values from Section 3 as the dependent variable $\boldsymbol{\mathcal{Y}}^{\mathrm{p}}$. We use an indicator of ever arrested at age 22 as the only predictor in $\boldsymbol{Y}^{\mathrm{p}}$ for crime. For income, we use indicators of high-school completion, some college, college completion, ever arrested at age 22, employment at age 25, and the interaction of employment at age 25 and high-school graduation. For education, we use indicators of ever in special education, high-school completion, some college, and college completion.

We form $\boldsymbol{Y}^{\mathrm{c}}$ using the same variables contained in $\boldsymbol{Y}^{\mathrm{p}}$ to estimate Equation (8).[22] We estimate Equations (8) and (9) using the following mediators $\boldsymbol{\theta}^{\mathrm{p}}$: average income from ages 20 to 40, an index of variables measuring neighborhood quality observed in the age-40 and age-54 follow-ups, cumulative days in jail up to age 54, marriage stability up to child's age 21, an index of variables measuring parenting observed in the age-40 and age-54 follow-ups, and an index of cognitive and non-cognitive skills observed in the age-54 follow-up.[23] Table 2 and García et al. (2023) report substantial treatment effects on these mediators. Estimation of Equations (6), (8), and (9) provides the elements for estimating $\boldsymbol{\pi}$ from Equation (10).

The prediction model characterizes the spillover benefits to the children. Modeling the

[22]Our choice of predictors is based on two criteria: 1) A small amount of missing values; and 2) Predictive power as suggested by economic theory. For crime, predictors other than ever arrested barely budge our estimates. Additionally, we verify that, when regressing each of the predictors on the relevant mediators and a treatment indicator, the null hypothesis that the coefficient associated with a treatment indicator is 0 holds empirically. García et al. (2020) suggest verifying this null hypothesis as an empirical test of the assumption that mediators fully explain the treatment effect on the predictors. We verify that null hypothesis for each predictor and thus exclude D from Equation (8).

[23]We use the same mediators for predicting all of the intergenerational present values. When participants have more than one child and the mediators are child specific (e.g., marriage stability up to child's age 21), we use the average across their children as the mediator. The analysis of the children of the participants assumes that missing information occurs randomly conditional on the baseline variables $\boldsymbol{W}^{\mathrm{p}}$. Missing information could occur either because of attrition, item non-response, or because some participants do not have children.

intergenerational transmission of the program through its impact on the mediators $\boldsymbol{\theta}^{\mathrm{p}}$ and their subsequent impact on the child predictors $\boldsymbol{Y}^{\mathrm{c}}$ is more efficient than directly estimating the reduced form in Equation (10). The efficiency gain results from stating the variance clustering of the error term in each equation of the prediction model.[24]

4.2 Intragenerational Spillover Benefits

We use the same method and predictors used to monetize child impacts when monetizing sibling impacts. We use as mediators the income, jail time, and skill mediators considered when monetizing child impacts. These are the mediators that have the most potential to impact the siblings' lives. For example, by having larger incomes, treatment participants can provide social insurance to their siblings. By having lower incarceration rates, they require less help from their siblings and extended families for surviving life in prison—incarceration is an enormous burden to the immediate and extended families of incarcerated individuals (Grinstead et al., 2001; Roberts, 2004). We do not age-adjust the sibling predictors (older siblings are observed at midlife, and their main outcomes are unlikely to change thereafter).[25]

[24]We could make separate predictions for each outcome. For efficiency, we estimate Equation (6) as a seemingly unrelated regression (SUR) system of the three outcomes (education, income, and crime). Similarly, we estimate Equation (8) as a SUR system of predictors, Equation (9) as a SUR system of mediators, and Equation (10) as a SUR system of (predicted) present values. We bootstrap the estimation of these equations at the level indicated by their error terms. In principle, we do not require to account for sampling variation in the estimation of Equation (10) because $\boldsymbol{\xi}_{\mathrm{cph}}$ converges in probability to 0. However, for some children of the participants, we do not need to age-adjust the ever arrested indicator at age 22 (we observe the variable at that age 22 in some cases). In those cases, the additional term $\boldsymbol{\chi}\boldsymbol{\eta}_{\mathrm{cph}}$ appears in Equation (10). We thus bootstrap the estimation of this equation as well.

[25]We observe a small number of mediators when predicting sibling impacts. Unobserved mediators are likely in this case. Let s index siblings and expand equation Equation (9) into two subsystems, one of observed mediators and one of unobserved mediators. Denote by $*$ the elements of the former subsystem and by $\sim$ the elements of the latter. We use the same notation for coefficients as in Section 4.1 for brevity. The reduced form of the prediction model for the siblings is

$$\hat{\boldsymbol{\mathcal{Y}}}^{\mathrm{s}} = \underbrace{\left(\boldsymbol{\chi}\boldsymbol{\Gamma}^{*}\boldsymbol{\Delta}^{*} + \boldsymbol{\chi}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Delta}}\right)}_{=:\boldsymbol{\pi}^{\mathrm{siblings}}} D + \boldsymbol{\chi}\boldsymbol{\Gamma}^{*}\left(\boldsymbol{\theta}^{\mathrm{c}} - \hat{\boldsymbol{\Delta}}D\right) + \boldsymbol{\chi}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{Q}}\boldsymbol{W}^{\mathrm{h}} + \boldsymbol{\chi}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\nu}}_{\mathrm{h}} + \boldsymbol{\chi}\boldsymbol{\eta}_{\mathrm{sph}}. \tag{11}$$

Some participants share siblings. In that case, we average their baseline variables and mediators and denote the vectors of averages by $\boldsymbol{W}^{\mathrm{h}}$ and $\boldsymbol{\theta}^{\mathrm{h}}$. Estimation proceeds in the same way as in Section 4.1. The error term of Equation (11) does not converge in probability to 0 as in Equation (10). We resample our bootstrap as in Section 4.1. In addition, we account for sampling variation in the estimation of Equation (11) by bootstrapping its estimation clustering at the household level due to the presence of $\boldsymbol{\chi}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\nu}}_{\mathrm{h}} + \boldsymbol{\chi}\boldsymbol{\eta}_{\mathrm{sph}}$.

4.3 Estimates

Table 6 presents estimates of the predicted present value for the children of the participants for education, income, and crime—i.e., the elements of $\boldsymbol{\pi}$ in Equation (10). It also presents the corresponding total. We display estimates by sex of the participants and their children.

The present value due to education is generally negative. Education of the children of the treatment-group participants is greater than that of controls, raising direct costs. However, in our prediction model, education mediates income. Thus, the program generates a sizable intergenerational present value due to income for both male and female children. For male children, there is also a major present value due to crime of 17 thousand dollars (p-value = .04). The total present value across outcomes is 39 thousand dollars (p-value = 0.01) for male children and 13 thousand (p-value = 0.09) for female children.

The gender differences in the program's intergenerational benefits are similar in magnitude and outcomes to the gender differences in the benefits for its participants. This reinforces our interpretation of the greater vulnerability of boys to disadvantaged environments and thus greater benefits from supplementing their families while they grow up. Of special relevance is the large present value due to crime for male children of male participants, which amounts to 24 thousand dollars (p-value = .01). The program decreases crime for the male participants. This decrease transmits across generations, which is consistent with studies in economics and sociology on the intergenerational transmission of incarceration of disadvantaged men (e.g., Dobbie et al., 2018; Haskins, 2014; Murray et al., 2014; Turney and Haskins, 2014).

The benefits to the siblings of the participants are imprecisely estimated. However, they are large in magnitude and consistent with the gendered pattern of the intergenerational benefits. Together, the results in Table 6 indicate substantial benefits for the extended dynasty of children and siblings of the participants. We recalculate the benefit-cost ratio of 6.0 reported in Table 4, using the child and sibling spillovers reported in Table 6 and the

Table 6. Predicted Life-Cycle Present Value for the Children and Siblings of the Participants of PPP

Present Values in 1,000s of 2017 USD

	All Participants	
	Male Children	Female Children
Education	-1.45	-0.88
Income	23.54	12.86
Crime	16.74	0.67
Total	38.83	12.65
(p-value)	(0.006)	(0.092)
	Male Participants	
	Male Children	Female Children
Education	-0.76	1.03
Income	3.49	-0.99
Crime	24.08	0.39
Total	26.81	0.43
(p-value)	(0.082)	(0.480)
	Female Participants	
	Male Children	Female Children
Education	-2.31	-2.16
Income	35.28	22.87
Crime	6.43	1.62
Total	39.40	22.32
(p-value)	(0.204)	(0.153)
	All Participants	
	Male Siblings	Female Siblings
Education	-5.16	0.80
Income	57.05	9.98
Crime	9.21	0.69
Total	61.09	11.46
(p-value)	(0.180)	(0.359)

Note: This table summarizes our estimates of the program's predicted life-cycle present value for the children of the participants for education, income, and crime—i.e., the elements of $\boldsymbol{\pi}$ in Equation (10). It also presents the total of the three present values. We present results by gender of the participants and their children. We also present the estimates for the siblings of the participants. For the total of the outcome present values, we display bias-corrected accelerated bootstrap p-values. The null hypothesis for the total present value is that it is less than or equal to 0. We use the estimator and inference described in Sections 4.1 and 4.2. The estimates are discounted to the year in which the program started using a rate of 3%.

children and siblings of participants. The estimate increases to 7.5 (p-value $= 0.00$).

5. Summary

This paper quantifies the life-cycle benefits of the HighScope Perry Preschool Project for the participants and the spillover benefits accruing to their children and siblings. Using a discount rate of 3%, the average present value of the monetized program outcomes amounts to 190 thousand dollars of 2017. Subtracting the total program cost per participant from this total present value yields the program's net social benefit of 170 thousand dollars. Subtracting the deadweight generated by collecting the taxes required to fund the program, the net social benefit decreases to 160 thousand. The benefit-cost ratio corresponding to this net social benefit is 6.0 (p-value $= 0.03$). Our estimates survive extensive sensitivity analyses. Undiscounted, as recently suggested in Congressional Budget Office (2018), our benefit-cost ratio estimate increases to 17.2 (p-value $= 0.047$). The program provided inputs additional to parental investment that developed children. The simultaneous provision of such inputs together with parental investment had long-term impacts, supporting theoretical work on the long-term implications of simultaneous early-life investments on long-term human capital formation and accumulation (Ehrlich and Yin, 2013; Heckman and Mosso, 2014).

For the first time in an evaluation of early childhood education programs, our analysis accounts for spillover benefits on participants' (untreated) children and (untreated) siblings. Spillover estimates show substantial benefits for the extended dynasty. For example, the program more than pays for itself through the average benefit of 39 thousand dollars (p-value $= 0.01$) that it generates for the male children of its participants. Our study suggests that spillover benefits to children and siblings of participants, for which research is incipient despite some exceptions (e.g., Barr and Gibbs, 2022; García et al., 2023; Rossin-Slater and Wüst, 2020), are relevant in quantifying the benefits of early childhood education. Human-capital spillovers are an essential byproduct of these programs, as they are engines of economic development (Ehrlich and Pei, 2020).

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, M. J., S. Sun, and B. Timpe (2021). Prep School for Poor Kids: The Long-run Impacts of Head Start on Human Capital and Economic Self-sufficiency. *American Economic Review 111*(12), 3963–4001.

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 (2022). Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood. *Journal of Political Economy 130*(12), 3253–85.

Becker, G. S. and H. G. Lewis (1973). On the Interaction between the Quantity and Quality of Children. *Journal of Political Economy 81*(2, Part 2), S279–S288.

Belfield, C. R., M. Nores, S. Barnett, and L. Schweinhart (2006). The High/Scope Perry Preschool Program Cost–Benefit Analysis Using Data from the Age-40 Followup. *Journal of Human Resources 41*(1), 162–190.

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.

Blackburn, M. L. and D. Neumark (1993). Omitted-ability Bias and the Increase in the Return to Schooling. *Journal of Labor Economics 11*(3), 521–544.

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.

Bureau of Labor Statistics (2015). National Longitudinal Surveys: The NLSY79. Website, `http://www.bls.gov/nls/nlsy79.htm`.

Campbell, F. A., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. P. Pungello, and Y. Pan (2014). Early Childhood Investments Substantially Boost Adult Health. *Science 343*(6178), 1478–1485.

Carneiro, P. and R. Ginja (2014). Long-term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy 6*(4), 135–73.

Congressional Budget Office (2018). How CBO Produces Fair-Value Estimates of the Cost of Federal Credit Programs: A Primer.

De Haan, M. and E. Leuven (2020). Head Start and the Distribution of Long-term Education and Labor Market Outcomes. *Journal of Labor Economics 38*(3), 727–765.

Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics 1*(3), 111–134.

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.

Dougan, W., J. L. García, and I. Polovnikov (2022). Childcare and Parenting as Skill Determinants: Evidence from a Large-scale Trial of an Early-Childhood Education Program. SSRN Working Paper 4234499, Social Science Research Network.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association 82*(397), 171–185.

Ehrlich, I. and Y. Pei (2020). Human Capital as Engine of Growth: The Role of Knowledge Transfers in Promoting Balanced Growth within and across Countries. *Asian Development Review 37*(2), 225–263.

Ehrlich, I. and Y. Yin (2013). Equilibrium Health Spending and Population Aging in a Model of Endogenous Growth: Will the GDP Share of Health Spending Keep Rising? *Journal of Human Capital 7*(4), 411–447.

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.

Farrington, D. P. (1986). Age and Crime. *Crime and justice 7*, 189–250.

Feldstein, M. (1999). Tax Avoidance and the Deadweight Loss of the Income Tax. *Review of Economics and Statistics 81*(4), 674–680.

Garces, E., D. Thomas, and J. Currie (2002). Longer-term Effects of Head Start. *American Economic review 92*(4), 999–1012.

García, J. L., F. Bennhoff, 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 3*(S1), 1–23.

García, J. L. and J. J. Heckman (2023). Parenting Promotes Social Mobility Within and Across Generations. *Annual Review of Economics 15*(1), 349–388.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). Quantifying the Life-Cycle Benefits of a Prototypical Early Childhood Program. *Journal of Political Economy 128*(7), 2502–2541.

García, J. L., J. J. Heckman, and V. Ronda (2023). The Lasting Effects of Early Childhood Education on Promoting the Skills and Social Mobility of Disadvantaged African Americans. *Journal of Political Economy 6*(131), 1477–1506.

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.

Gelber, A. and A. Isen (2013). Children's Schooling and Parents' Behavior: Evidence from the Head Start Impact Study. *Journal of Public Economics 101*, 25–38.

Goldin, C., S. P. Kerr, and C. Olivetti (2022). When the Kids Grow Up: Women's Employment and Earnings across the Family Cycle. NBER Working Paper w30323, National Bureau of Economic Research.

Golding, P. and H. E. Fitzgerald (2017). Psychology of Boys At Risk: Indicators From 0-5. *Infant Mental Health Journal 38*(1), 5–14.

Grinstead, O., B. Faigeles, C. Bancroft, and B. Zack (2001). The Financial Cost of Maintaining Relationships with Incarcerated African American Men: A Survey of Women Prison Visitors. *Journal of African American Men 6*, 59–69.

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. 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. Q. Yavitz (2010a). Analyzing Social Experiments as Implemented: A Reexamination of the Evidence From the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b). The Rate of Return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1–2), 114–128.

Heckman, J. J. and S. Mosso (2014). The Economics of Human Development and Social Mobility. *Annual Reviews of Economics 6*(1), 689–733.

Heckman, J. 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.

Hendren, N. and B. Sprung-Keyser (2020). A Unified Welfare Analysis of Government Policies. *Quarterly Journal of Economics 135*(3), 1209–1318.

Kline, P. and C. R. Walters (2016). Evaluating Public Programs with Close Substitutes: The Case of HeadStart. *Quarterly Journal of Economics 131*(4), 1795–1848.

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.

Loewenstein, S. (1973). Conclusion of a Study of PARI as a Research Tool. *Child Welfare 52*(6), 400–402.

Ludwig, J. and D. L. Miller (2007). Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *Quarterly Journal of Economics 122*(1), 159–208.

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.

Mulligan, C. B. and Y. Rubinstein (2008). Selection, Investment, and Women's Relative Wages over Time. *Quarterly Journal of Economics 123*(3), 1061–1110.

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.

Pashchenko, S. and P. Porapakkarm (2016). Medical Spending in the US: Facts from the Medical Expenditure Panel Survey Data Set: Medical Spending in the US. *Fiscal Studies 37*(3-4), 689–716.

Ramey, C. T. and F. A. Campbell (1984). Preventive Education for High-risk Children: Cognitive Consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency 85*(5).

Roberts, D. E. (2004). The Social and Moral Cost of Mass Incarceration in African American Communities. *Stanford Law Review 56*(5), 1271–1305.

Rolnick, A. and R. Grunewald (2003). Early Childhood Development: Economic Development with a High Public Return. *The Region 17*(4), 6–12.

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. et al. (1993). *Significant Benefits: The High/Scope Perry Preschool Study through Age 27. Monographs of the High/Scope Educational Research Foundation, No. 10.* ERIC.

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.

The White House (2013). Fact Sheet: President Obama's Plan for Early Education for all Americans. Website, `https://obamawhitehouse.archives.gov/the-press-office/2013/02/13/fact-sheet-president-obama-s-plan-early-education-all-americans`.

The White House (2020). The Build Back Better Framework. Website, `https://www.whitehouse.gov/build-back-better/`.

The White House (2021). The Biden-Harris Administration Immediate Priorities. Website, `https://www.whitehouse.gov/priorities/`.

Thompson, O. (2018). Head Start's Long-run Impact Evidence from the Program's Introduction. *Journal of Human Resources 53*(4), 1100–1139.

Turney, K. and A. R. Haskins (2014). Falling Behind? Children's Early Grade Retention After Paternal Incarceration. *Sociology of Education 87*(4), 241–258.

US Census Bureau (2010). 2010 American Community Survey.

US Census Bureau (2015). 2015 American Community Survey.

US Census Bureau (2020). Mean Income Received by Each Fifth and Top 5 Percent of All Families: 1966 to 2019. Current Population Survey, Annual Social and Economic Supplements (CPS ASEC).

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.