The Dynastic Benefits of Early Childhood Education

Jorge Luis García
Frederik H. Bennhoff
Duncan Ermini Leaf
James J. Heckman

DISCUSSION PAPER SERIES

The Dynastic Benefits of Early Childhood Education

Jorge Luis García
Clemson University

Frederik H. Bennhoff
University of Chicago

Duncan Ermini Leaf
University of Southern California

James J. Heckman
University of Chicago and IZA

JUNE 2021

ABSTRACT

The Dynastic Benefits of Early Childhood Education*

This paper monetizes the life-cycle intragenerational and intergenerational benefits of the Perry Preschool Project, a pioneering high-quality early childhood education program implemented before Head Start that targeted disadvantaged African-Americans and was evaluated by a randomized trial. It has the longest follow-up of any experimentally evaluated early childhood education program. We follow participants into late midlife as well as their children into adulthood. Impacts on the original participants and their children generate substantial benefits. Access to life-cycle data enables us to evaluate the accuracy of widely used schemes to forecast life-cycle benefits from early-life test scores, which we find wanting.

Corresponding author:
James J. Heckman
Center for the Economics of Human Development
Department of Economics
The University of Chicago
1126 East 59th Street
Chicago, IL 60637
USA

E-mail: jjh@uchicago.edu

1. Introduction

This paper uses newly collected data to study the life-cycle benefits through late midlife (age 54) of the Perry Preschool Project (PPP) participants, and the benefits on their adult children and on their siblings. PPP is an influential early childhood education program targeted to promote the social mobility of disadvantaged African-American children. PPP is relevant today because it influences the design of current and proposed early childhood education programs. At least 30% of Head Start programs are based on it (Elango et al., 2016). PPP was evaluated by a randomized trial. Data are collected on the original participants, their siblings, and their children.

By late midlife, program treatment effects on the original participants have largely materialized. Our paper is the first to assess the long-run benefits of investing in high-quality early childhood education using actual data rather than imputed future outcomes as is common in the literature.[1] We thus avoid the bias arising from cohort effects which potentially plague synthetic-cohort approaches using imputed data.

PPP increases the education and earnings of participants, reduces participation in crime, and improves health and healthy behaviors. We monetize the life-cycle treatment effects of the program and calculate benefit-cost ratios and internal rates of return to it. We also monetize the impacts on the participants' siblings and adult children. We assess the accuracy of widely used imputation and forecasting approaches and find them wanting. Our analysis goes beyond documenting short-term gains in cognitive-test scores, school readiness, or early-life social behavior.[2] We do not focus solely on gains in specific adult outcomes or at specific

---

[1]Belfield et al. (2006) and Heckman et al. (2010b) present cost-benefit analyses of PPP. Their studies are limited in terms of outcomes analyzed and in that they rely on forecast or observation that does not go beyond early adulthood. García et al. (2020) forecast the life-cycle benefits of two other influential early childhood education programs (the Abecedarian Project and the Carolina Approach to Responsive Education, Campbell et al., 2008). Their paper develops a principled method for forecasting life-cycle benefits when follow-ups through adulthood are not possible. No current study provides intergenerational or intragenerational cost-benefit analysis.

[2]See Almond and Currie (2011), Cascio (2021), Cunha et al. (2006), and Elango et al. (2016) for literature surveys.

ages as in Campbell et al. (2014), Gray-Lobe et al. (2021), Havnes and Mogstad (2011), and Heckman et al. (2010a). We analyze a broad array of life-cycle trajectories for a variety of dimensions of achievement.

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 and there were no treatment substitutes available in the area where they lived. The participants of PPP were representative of the broader disadvantaged African-American population in the US. 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. We calculate that the analogous percentages for females and males eligible today are 10%.[3] Treatment and control participants were followed up in multiple rounds of data collection through age 54. Attrition remains low in both groups. In the age-54 follow up, when data on the siblings and adult children of the original participants were collected, 83% of the 123 participants were surveyed; 12% of the participants were not surveyed because they were deceased. Accounting for death, we monetize the life-cycle outcomes of 95% of the original participants.

Table 1 summarizes our estimates of benefit-cost ratios. It presents them for the original participants, and for the extended dynasty of original participants, their siblings, and their children.[4] Using one-sided tests, benefit-cost ratios exceed one statistically at a 5% significance level.[5] PPP generates nine dollars of benefits per dollar invested in it. This falls

---

[3]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.

[4]We have limited information on the siblings and adult children of the original participants so estimates of benefit-cost ratios for them are imprecise. We add their benefits to those of the original participants.

[5]We presume that benefits are greater than costs and thus use one-sided tests. This presumption follows a large body of literature (see Elango et al., 2016).

Table 1. Benefit-Cost Ratio of the Perry Preschool Project, Main Estimates

	[1] 1st Generation	[2] Dynasty	[3] Extended Dynasty
	(Original Participants)	([1] + Child Benefits)	([2] + Sibling Benefits)
Baseline Program Cost	*9.10*	*10.65*	*13.52*
[s.e.]	[5.04]	[5.35]	[6.30]
(*p*-value)	(0.032)	(0.025)	(0.019)
Add Deadweight Cost	*6.07*	*7.10*	*9.01*
[s.e.]	[3.36]	[3.57]	[4.20]
(*p*-value)	(0.037)	(0.030)	(0.022)

Note: This table summarizes our main estimates of the benefit-cost ratio of PPP. We provide estimates for two program-cost scenarios: 1) the baseline program cost (21,151 of 2017 US dollars); and 2) the baseline program cost multiplied by 1.5 to account for the deadweight loss that would be generated by collecting taxes to fund the program. The standard errors in brackets are bootstrapped and clustered at the household level. The *p*-values in parentheses are based on the bias-corrected accelerated bootstrap. The null hypothesis for each benefit-cost ratio is that it is less than or equal to 1.

to six dollars per dollar invested adjusting for the deadweight generated by collecting taxes to fund the program. Adding the extended-dynasty benefits, each dollar invested in PPP generates fourteen dollars not counting the deadweight costs and nine dollars accounting for deadweight costs.

The paper proceeds in the following way. Section 2 describes PPP and the data collected on its participants, their siblings, and their children. Section 3 discusses our strategy for monetizing and aggregating the life-cycle benefits of the original participants. Section 4 presents our estimates and multiple sensitivity checks. It reports the internal rate of return to PPP. Section 5 demonstrates that the *ad hoc* forecasts of life-cycle benefits based on short-term cognitive-test gains used in recent studies substantially underestimate benefits. In contrast, recently proposed principled forecasts based on economic models are very accurate. Section 6 monetizes the outcomes of the siblings and children of the original participants. Section 7 concludes.

2. The 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 the program (treatment group) or not (control group).[6] Barnett (1996) reports a total program cost per participant of 21,151 (2017 US dollars). We use this as the baseline total program cost, which includes operating (teacher salaries and administration) and capital (classroom and facilities) costs.

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 the program eligibles. After randomization, a few treatment and control participants were reassigned, and this generated an imbalance in baseline characteristics (see Table 2).[7] We correct point estimates accordingly. Failures in randomization protocols are not rare in social experiments. Authors often fail to document and address them methodologically, and this has sizable empirical consequences (Bruhn and McKenzie, 2009).

---

[6]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, 2 if employed in semi-skilled job, and 1 if employed in unskilled job or unemployed.

[7]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; 4) A coin toss randomly selected one group as the treatment group and the other as the control group; and 5) Some individuals provisionally assigned to treatment, whose mothers were employed at the time of the assignment, were swapped with control individuals whose mothers were not employed. The reason for this swap was that it was difficult for working mothers to participate in home visits assigned to the treatment group. See Heckman et al. (2010a) and Heckman and Karapakula (2019, 2021) for more details on the problems and methods to address them, which we use in this paper.

Table 2. Perry Preschool Project, Summary Statistics

Variable	Men and Women C	(T − C)	Men C	(T − C)	Women C	(T − C)
Sample and Baseline Variables						
Sample Size, Age 3 (baseline)	65	−7	39	−6	26	−1
Sample Size, Age 54	50	2	30	−1	20	3
IQ, Age 3 (baseline)	78.54	1.03	77.85	1.37	79.58	0.46
Socioeconomic Index, Age 3 (baseline)	8.62	0.17	8.65	0.24	8.57	0.09
Mother Works, Age 3 (baseline)	0.31	**−0.22**	0.28	**−0.22**	0.35	**−0.23**
Mother Age, Age 3 (baseline)	28.66	0.92	28.63	0.84	28.71	1.01
Education						
Years of Retention	0.48	−0.17	0.46	−0.07	0.50	−0.30
Years in Special Education	0.40	−0.04	0.28	0.14	0.58	−0.30
High School Graduation	0.78	0.08	0.85	−0.06	0.69	**0.27**
College Graduation	0.14	−0.04	0.10	−0.01	0.19	−0.07
Income (1,000s of 2017 USD)						
Age 20 to 27 Average	15.16	2.58	17.39	1.19	11.67	4.96
Age 28 to 40 Average	15.60	**10.16**	13.78	**10.90**	18.85	8.36
Age 41 to 55 Average	20.61	−1.53	21.60	−2.34	19.08	−0.23
Crime						
Total Arrests	10.26	−3.35	14.13	−3.79	4.46	−2.06
Felony Arrests	3.28	−0.91	5.10	−1.07	0.54	−0.38
Misdemeanor Arrests	6.98	**−2.43**	9.03	−2.72	3.92	−1.68
Total Violent Crime Arrests	2.08	**−1.30**	3.08	**−1.77**	0.58	**−0.50**
Health						
Number of Deceased	9	−3	4	0	5	−3
Age at Death	37.44	3.72	35.00	6.00	39.40	2.10
Good Health	0.48	0.11	0.49	0.15	0.46	0.06
Works Out	0.34	**0.16**	0.38	0.10	0.27	**0.25**
Siblings						
Number of Living Siblings per Participant	2.35	0.04	2.30	0.06	2.43	0.02
Age	56.84	−1.12	57.14	**−2.01**	56.35	−0.07
Male Fraction	0.48	0.10	0.55	−0.03	0.40	**0.26**
Ever in Special Education	0.06	**0.09**	0.07	0.10	0.04	0.07
High School Graduation	0.81	0.06	0.85	0.05	0.76	0.08
College Graduation	0.17	−0.05	0.21	−0.11	0.12	0.04
Employed or Retired	0.19	0.06	0.18	0.04	0.20	0.09
Ever Arrested	0.33	0.01	0.38	−0.14	0.28	0.17
Children						
Number of Living Children per Participant	2.37	−0.01	2.21	−0.21	2.62	0.22
Age	28.12	0.16	27.28	−0.11	29.08	0.21
Male Fraction	0.52	−0.05	0.55	−0.07	0.48	−0.01
Child in Good Health (at Age 23)	0.82	0.09	0.87	0.05	0.77	0.13
Ever in Special Education (Age 18)	0.23	−0.01	0.06	0.05	0.57	−0.16
High School Graduation (Age 20)	0.47	0.12	0.55	0.15	0.43	0.13
College Graduation (Age 25)	0.11	0.02	0.12	−0.03	0.08	0.07
Employed (Age 25)	0.37	**0.14**	0.34	**0.19**	0.39	0.11
Ever Arrested (Age 22)	0.24	−0.07	0.19	−0.08	0.25	−0.06

Note: This table summarizes the data in this paper. Column C displays the average in the control group unless sample size or another count is specified. $T−C$ displays the corresponding treatment-control difference. For the participants, the education variables combine administrative records and self reports, the income and health variables are self-reported, and the crime variables are based on administrative records. All of the program-participant variables cover the age range 0–54 unless specified otherwise. The sibling and child variables are based on questions answered by the participants in the age-54 follow-up. For the siblings, information on up to five living siblings is available. We only consider older siblings who were born before PPP started and are at most five years older than the participants. All sibling variables are reported as observed. For the children of participants, information on up to the five eldest living children is available. The number, age, and sex of the children are reported as observed. The remaining child variables are age-adjusted as explained in Appendix A3.7. We bold the treatment-control differences that statistically differ from zero at least with 10% significance. We use analytic standard errors clustered at the household level in this basic data description for computational and replication simplicity. The inference on the main results in the paper is based on bootstrap methods as explained below.

Table 2 summarizes the data available for analysis. Follow-up surveys with high response rates were conducted for both the treatment and control participants when they were, on average, 15, 19, 27, 40, and 54 years old. School, police, and court records allow us to observe full education and criminal histories through age 54. By this age, participants have completed their education and criminal careers. García et al. (2020) document that this is an accurate assumption for the US. We observe current and retrospective earnings at the ages sampled. We also observe data on health behaviors across the life cycle.[8]

Table 2 also shows data on the children of the participants. This information is based on questions asked during the age-54 follow-up. Interviews asked information about up to five eldest living children; 21% of the participants report having no children, and 8% report having more than five children.[9] Table 2 reports the number and sex of the children observed. We age-adjust child outcomes because participants have children of different ages (see Appendix A3.7 for details). We use age-adjusted variables when monetizing the outcomes of the children of the original participants in Section 6.

Original participants were also asked for information about up to five living siblings. We only consider older "pre-treatment" siblings who were born before PPP was administered. Doing so prevents contaminating the empirical analysis by affecting the fertility choices of the parents of the original participants.[10] Additionally, we drop siblings who were more than five years older than the original participants.[11] Information reported on the siblings is essentially complete (education and crime careers are completed by midlife).

---

[8]Table 2 shows that, in most variables, the average treatment-control differences have the "expected sign," although not all of them are statistically significant. Table 2 is an inspection of the raw data. Our focus is on combining the overall effect across outcomes into aggregate statistics.

[9]García et al. (2021) document that there are no economically or statistically significant differences in the level and timing of fertility between the original treatment and control participants. They provide several arguments indicating that selective fertility has minor impact on PPP's intergenerational treatment effects.

[10]Some original participants belonging to the first couple of cohorts have younger siblings who are original participants themselves in the later cohorts. In Section 6, we only count the benefits on "pre-treatment" siblings once per household, independently of the number of original participants that they have.

[11]This aims to increase the plausibility of the monetization of sibling impacts in Section 6. We drop nineteen treatment-group siblings and nine control-group siblings.

3. Monetizing and Aggregating the Life-Cycle Benefits and Costs of PPP

3.1 Measurement Framework

Let $\mathcal{J}$ index the set of monetized outcomes for the original participants (education, income, crime, health). The program's discounted average life-cycle benefit or present value of treatment from outcome $j \in \mathcal{J}$ is

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

where $\mathcal{A}$ indexes age and β is the discount factor associated with a discount rate ρ. $Y_{j,a}^d$ 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). $Y_{j,a}^d$ totals the monetary value of the outcome across subcategories (e.g., across different types of crime). $\mathbb{E}\left[Y_{j,a}^1 - Y_{j,a}^0\right]$ 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 (1) is the average benefit that outcome $j \in \mathcal{J}$ generates. The present value of the program's life-cycle total benefits is

$$\Pi := \sum_{j \in \mathcal{J}} \Pi_j. \tag{2}$$

The benefit-cost ratio is

$$\text{Benefit-Cost Ratio} := \frac{\Pi}{\text{Total Program Cost}}. \tag{3}$$

We use three estimators of the average treatment effect $\mathbb{E}\left[Y_{j,a}^1 - Y_{j,a}^0\right]$ in Equation (1). The first is the average treatment-control difference (mean difference). For each outcome $j \in \mathcal{J}$, we pool the treatment and control groups at all observed ages to estimate

$$Y_{j,a} = \delta_{j,a}^0 + \delta_{j,a}^1 D + \varepsilon_{j,a} \quad \text{for } a \in \mathcal{A}, \tag{4}$$

7

where $Y_{j,a}$ is the value of the observed outcome. In the framework of Quandt (1958), $Y_{j,a} :=$ $D \cdot Y_{j,a}^1 + (1 - D) \cdot Y_{j,a}^0$ where D indicates treatment status. $\delta_{j,a}^0$ and $\delta_{j,a}^1$ are age-specific coefficients and $\varepsilon_{j,a}$ is an error term. The coefficient $\delta_{j,a}^1$ is the mean difference for outcome $j \in \mathcal{J}$ at age $a \in \mathcal{A}$. Our first estimator of treatment effects uses sample counterparts of $\delta_{j,a}^1$. Mean-difference estimators identify the average treatment effect if treatment is randomized.

To address the randomization compromises described in Section 2 as well as the problem of missing data, we use regression-adjusted mean differences as a second estimator (OLS). We do not know the exact form of randomization failure. However, we know that baseline variables are only partially balanced across the treatment and control groups (see Section 2). We adjust the average treatment-control difference by including the baseline variables in Table 2 as regressors in Equation (4). Doing so allows missing-data patterns to vary across covariates. We use this estimator for our baseline estimates because it appropriately accounts for the methodological challenges in hand, and it is straightforward to replicate.

Our third estimator follows the robust conservative methodology of Heckman and Karapakula (2019, 2021) applied to PPP data. It uses the augmented inverse-probability estimator (AIPW), a more general mean-difference adjustment than OLS. AIPW weights Equation (4) using the inverse probability of being treated and having non-missing data. It imputes (missing) counterfactual outcomes for each participant. Weighting and imputation are based on the same baseline variables used for OLS. AIPW estimators have a double-robustness property. For the estimator to be consistent, it suffices that either the weighting scheme or the imputation equation is correctly specified. Thus, if the linear correction of OLS is inadequate but the weighting scheme is correct, we would detect this from comparing AIPW to OLS estimates. Appendix A3.1 provides further details on these estimators.

3.2 Inference

We test if the average life-cycle total benefit of the program is less than or equal to its total cost (i.e., if the benefit-cost ratio is less than or equal to one). We also test if the average

8

life-cycle benefit by outcome or overall total is less than or equal to zero.

Our main inferences are based on bias-corrected accelerated bootstrap confidence intervals (BCAs, Efron, 1987). This procedure provides confidence intervals that converge more quickly when compared to standard inference methods (i.e., it provides an asymptotic refinement). It also corrects for outcome skewness. The asymptotic refinement is a relevant addition to our inference strategy because our sample is small. The skewness correction is important because our outcome distributions are asymmetrical. Labor-income and crime outcomes are naturally bounded below by zero and some exhibit extreme values. This translates into non-normality and left-skewness in the distribution of our estimators, as indicated by their empirical bootstrap distributions.[12] Our results throughout the paper show that the correction offered by the BCAs bites. Rule-of-thumb approaches that divide point estimates by standard errors often understate the statistical significance of our estimates.

We display bootstrap standard errors clustered at the household level, defining households as individual-sibling clusters. In all of our inference procedures we stratify to keep the size of the treatment and control groups constant. We account for sampling variation in all estimation stages required by our estimators (i.e., we form the empirical bootstrap distributions of our estimators by computing all estimation stages in each bootstrap draw). Section 4.2 compares our main inference to inference based on standard asymptotic methods and to inference based on alternative bootstrap procedures. Appendix A3.2 provides step-by-step explanations of all of the inference procedures that we present throughout the paper.

3.3 Formal Education

PPP reduced participation in special education and K-12 grade retention and increased high school graduation rates, especially for female participants (see Table 2). The benefit due to

---

[12]See Cameron et al. (2008) for a comparison of several, related inference procedures and an empirical demonstration of the advantage of asymptotic refinements as the one in our BCAs. The advantage of the inference provided by the BCAs with respect to other methods like bias-corrected percentile-t (studentized) bootstrap confidence intervals is that it explicitly accounts for skewness as a small-sample feature.

the decrease in treatment-group special education and K-12 grade retention is positive. The benefit due to the increase in treatment-group formal schooling is negative.

We observe records for K-12 and special education for 118 of the 123 participants, and directly quantify the benefit generated by the program. We assign an annual cost of 8,665 (2017 US dollars) per year of K-12 education. We obtain this figure from the 1975-1976 school-year cost in Grant and Lind (1978). For the same period, Kakalik et al. (1981) report that a year of special education was 2.17 times more costly than a regular year of K-12. We apply this multiplier and assign an annual cost of 18,803 (2017 US dollars) per year of K-12 special education. We proceed in a similar fashion for college education. We combine enrollment and matriculation records (see Appendix A3.3). We assign each college year its annual cost in 1982-83 (13,768 of 2017 US dollars) using information from Grant and Snyder (1986). We choose 1982-1983 because most individuals in the sample who graduated from high school do so in that year.

3.4 Income

At ages 19, 27, 40, and 54, participants report their labor income in the current and preceding years. We supplement this information by setting to zero the labor income of incarcerated participants. With this supplementation, we observe the labor income of 86% of the treatment-group participants and 81% of the control-group participants across the available ages. Our baseline estimates do not interpolate to fill in missing data; they rely solely on the available observations. In Section 4, we document that our estimates are robust to using interpolation and to applying the method in García et al. (2020). The latter method imputes missing observations by non-parametrically matching the PPP 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).

We break down the average treatment-control difference into transfers that the govern-

ment would have provided to individuals if they had lower labor income (transfers), state and federal taxes, and after-tax labor income.[13] We calculate the federal and state taxes for each individual using the historical tax rates in US Census Bureau (2019)—the calculation considers their observed marital status. We then estimate the average increase in state and federal taxes paid due to the increase in the labor income of the treatment participants.

We quantify the reduction in transfers from multiple means-tested programs. This reduction is a consequence of higher labor income. We report the benefit from transfers using sample observations. In the interviews at ages 27, 40, and 54, individuals report their 15-year histories as beneficiaries of social programs (e.g., temporary assistance to needy families, food stamps, child care subsidies, supplemental security income, unemployment insurance, general welfare assistance, disability payments, aid to families with dependent children, any other).[14] Figures 1a and 1b display the cumulative average benefit from after-tax labor income and from labor-income taxes and transfers over the life cycle. Some benefits accrue earlier (e.g., crime) than other benefits (e.g., health). Appendix A3.4 provides details on the quantification of the outcomes in this subsection.

3.5 Crime

PPP reduces crime. This reduction generates a substantial benefit. It lowers the cost to the criminal justice system (i.e., police, court, and correctional costs). 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

---

[13]Treatment-participant income increases. The increase generates deadweight from adjustments to avoid taxation. The income benefits that we monetize are net of this 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.

[14]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.

Figure 1a. Cumulative Present Value of the Perry Preschool Project Over the Life Cycle, Men
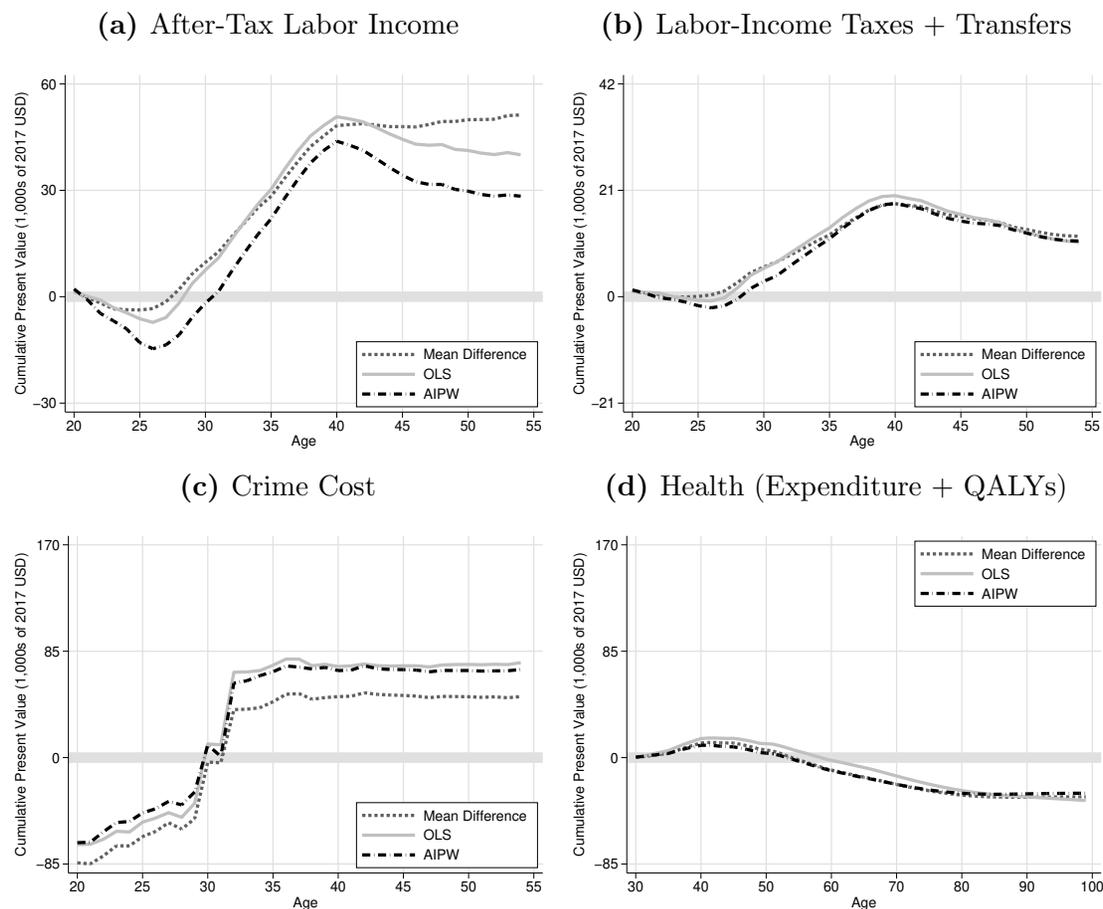
(a) After-Tax Labor Income

(b) Labor-Income Taxes + Transfers

(c) Crime Cost

(d) Health (Expenditure + QALYs)

Note: Panel (a) A point in each of the lines represents the cumulative present value from after-tax labor income (endpoint included) for male participants in 1,000s of 2017 US dollars. Results using the three estimators discussed in Section 3.1 are displayed. All values are discounted to the year in which the program started using a 3% rate. Panel (b) Analogous in format to Panel (a) for the sum of federal and state labor-income taxes and social-program transfers from the government to individuals. Panel (c) Analogous in format to Panel (a) for crime costs related to the criminal justice system and monetary costs to victims (without quality-adjusted life years for the crime victims). Panel (d) Analogous in format to Panel (a) for the value of the governmental and private medical expenditures plus quality health-adjusted life years, assigning a value of statistical life of 100,000 (2017 US dollars).

Figure 1b. Cumulative Present Value of the Perry Preschool Project Over the Life Cycle, Women
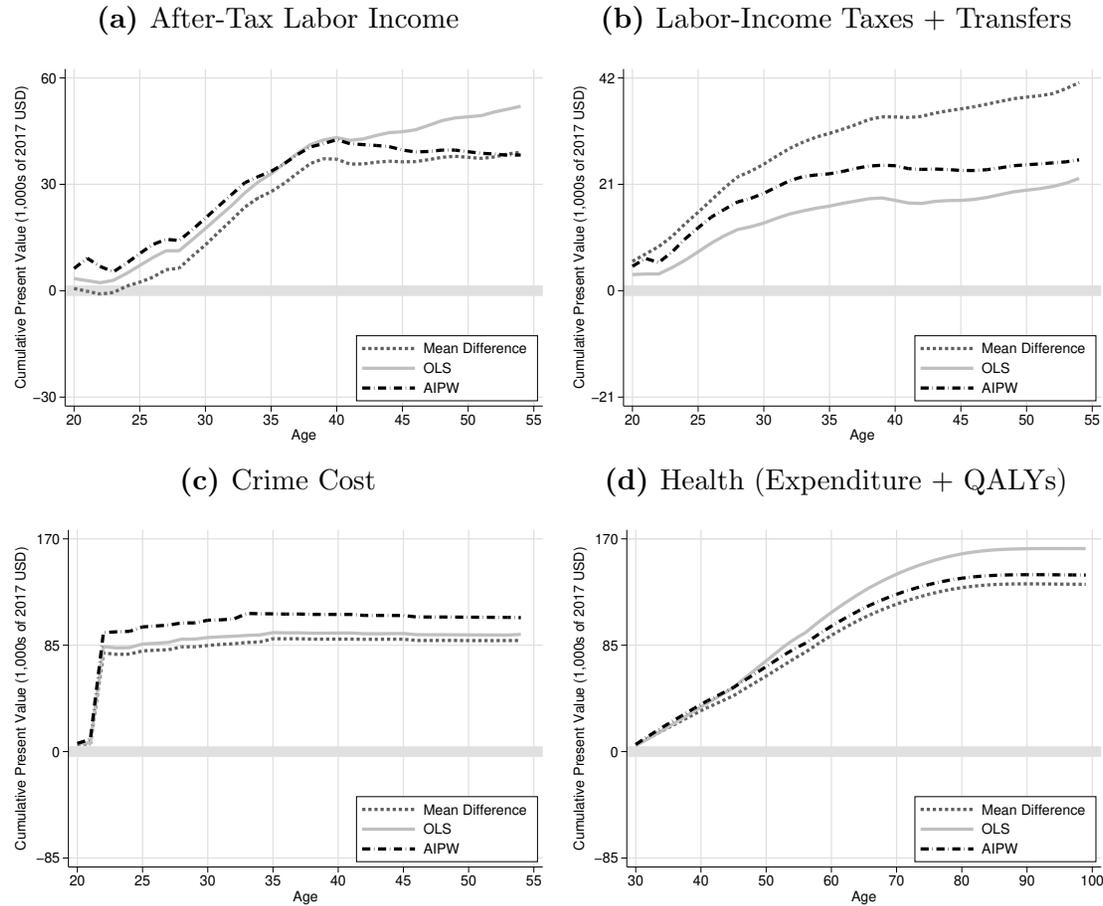
(a) After-Tax Labor Income

(b) Labor-Income Taxes + Transfers

(c) Crime Cost

(d) Health (Expenditure + QALYs)

Note: Panel (a) A point in each of the lines represents the cumulative present value from after-tax labor income (endpoint included) for female participants in 1,000s of 2017 US dollars. Results using the three estimators discussed in Section 3.1 are displayed. All values are discounted to the year in which the program started using a 3% rate. Panel (b) Analogous in format to Panel (a) for the sum of federal and state labor-income taxes and social-program transfers from the government to individuals. Panel (c) Analogous in format to Panel (a) for crime costs related to the criminal justice system and monetary costs to victims (without quality-adjusted life years for the crime victims). Panel (d) Analogous in format to Panel (a) for the value of governmental and private medical expenditures plus quality health-adjusted life years, assigning a value of statistical life of 100,000 (2017 US dollars).

is likely larger than the treatment-induced reduction in arrests made for those crimes.

We follow Hunt et al. (2017) and Hunt et al. (2019) and monetize the police and court costs per arrest by type of crime using estimates specific to Michigan. The police and court cost per arrest range from 2,371 (2017 US dollars) for drug-related crimes, driving offenses, vandalism, and similar crimes to 367,107 (2017 US dollars) for murder. We take correctional costs from United States Department of Justice (2010): 26,323 (2017 US dollars) per year spent in prison or jail, and 1,330 (2017 US dollars) per year of probation (due to monitoring). The prison cost is specific to Michigan and the probation cost is national.

Victims of crime often sustain mental and physical healthcare costs, their property is damaged or stolen, become unable to work, or lose quality of life. To calculate the cost to the victims of crime, we proceed as follows. First, we follow standard practice in the criminology literature and inflate the number of arrests to address the disparity between the number of arrests and the number of crimes committed. Second, we monetize the material and quality-of-life losses to the crime victims. For the first step, we use several nationally representative datasets to estimate victimization-arrest inflation ratios. For the second step, we use the unit crime costs in Miller et al. (2020). Our estimates are conservative: under weak assumptions, possible measurement errors of victim costs are classical.[15] Section 4.3 analyzes the sensitivity of our estimates to the several choices of monetization methodology in this section. We consider a scenario where the cost to crime victims is lower because stolen goods could be considered a transfer (without destruction of goods) between criminals and the victims.[16] Figures 1a and 1b display the cumulative average benefit from crime reduction over the life cycle. In these figures, we include the costs of the criminal justice system and the monetary costs to crime victims. We exclude the quality-adjusted life year (QALY)

---

[15]The strongest assumption is that, for a finely chosen crime typology and by treatment status, mean clearance rates are constant in criminal propensity and victim costs. We believe this assumption is justified since variation of underlying factors of clearance rates (e.g., police effort) is well captured by our detailed typology which differentiates by seriousness and nature of crime.

[16]The monetary cost to the victims includes medical and mental-health costs, loss of work, and lost and damaged property. These costs vary by crime type. Lost and damaged property is only a fraction of the total. We cannot separate "lost" from "damaged" in the sources that we use for monetizing crime.

cost to victims in the main paper. We provide estimates including this cost in Appendix Table A4.1. Appendix A3.5 provides details on how we monetize crime outcomes.

3.6 Health

Treatment effects on health are expected to appear after age 54 due to lasting improvements in treatment-group health behavior. Health conditions are unlike labor income or crime, for which treatment effects largely materialize before age 50 (see Figures 1a and 1b). We monetize health outcomes from age 18 to (realized or forecasted) death. Health is the only outcome for which we rely on modeling and forecasts rather than direct observation. To forecast future health outcomes and predict their consequences (including costs and benefits) we model health behaviors and biomarkers as predictors of subsequent health conditions.

We simulate the health outcomes of participants in the age range between the age-27 and age-54 interviews using the Future Adult Model (FAM, Goldman, Ermini Leaf, and Tysinger, 2018). FAM simulates health-state transitions in the Panel Study of Income Dynamics (PSID, Institute for Social Research at the University of Michigan, 2019). We harmonize the PPP interviews at ages 27, 40, and 54 with the PSID. We use the age-27 interview as the initial period when starting the simulation. We use importance-weighted sampling to generate a distribution of outcomes simulated by FAM that is conditional on the outcomes observed in the PPP interviews at ages 40 and 54 (Ermini Leaf, 2017). We simulate the health outcomes after age 54 using the Future Elderly Model (FEM, Ermini Leaf et al., 2020; Goldman et al., 2018). FEM is analogous to FAM but is simulated in the Health and Retirement Study RAND HRS Version P public-use dataset and 2006-2014 biomarker sensitive datasets.[17] We harmonize the datasets and simulate the health outcomes as we do with FAM. FAM does not simulate death because we observe this outcome in the age-54 interview. FEM simulates death as an outcome. If an individual dies, we assign all of their outcomes a monetary value of zero. We simulate the life-cycle trajectories of the PPP

---

[17]RAND HRS Version P (2016); University of Michigan and the National Institute on Aging (2013, 2015a,b, 2016, 2017).

participants 1,000 times. We use the average across simulations as point estimates of our predicted health outcomes.

We compute governmental medical expenditures through programs like Medicaid and Medicare and private medical expenditures using the Medical Expenditure Panel Survey (MEPS) and the Medicare Current Beneficiary Survey (MCBS). We harmonize the PPP data up to the age-27 interview with MEPS data and predict costs using regression models estimated in the 2000-2010 MEPS waves.[18] We compute costs between age 27 and Medicare entry by applying MEPS-based models to the FAM and FEM predicted outcomes for the PPP participants. We assume that the participants enter Medicare at age 65 or after claiming disability for two consecutive years. After they enter Medicare, we follow an analogous procedure to compute their costs replacing MEPS with the 2007-2010 waves of MCBS.[19] We adjust the predicted values for average annual real medical cost growth using Congressional Budget Office (2007).

We do not estimate health QALYs before age 27 because the data required are not available in the PPP surveys. QALY gains generated by PPP are expected to be small when the participants are young. We estimate QALYs after age 27 using the model in Goldman et al. (2018) adapted to the PSID for the outcomes of the FAM simulation. We proceed as in Goldman et al. (2018) for the outcomes of the FEM simulation. We use a yearly value of statistical life of 100,000 (2017 US dollars). QALYs adjust this yearly value to account for the burden of disease. Figures 1a and 1b display the cumulative average benefit from health over the life cycle, adding up governmental and private medical expenditures and QALYs. In addition to accounting for sampling variation as with all other monetized outcomes, our standard errors account for simulation error in the health predicted outcomes (i.e., forecasting error in the simulations of the life-cycle health trajectories). We provide details on this inference procedure in Appendix A3.2.

---

[18]Agency for Healthcare Research and Quality (2003, 2004a,b, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014).

[19]Centers for Medicare and Medicaid Services, U.S. Department of Health (2011a,b, 2012, 2013).

4. Estimates

Table 3 summarizes our results by outcome category and overall. We display the average life-cycle difference between the treatment and control groups in the present value of the monetized outcomes. Estimates are based on the OLS estimator (see Section 3.1). Values are in 2017 US dollars and are discounted to the year in which the program started (age 3) using a 3% rate. We also present the percentage change in the present value of the treatment group relative to the control group by outcome and two estimates of the benefit-cost ratio. The percentage change quantifies each outcome's contribution to the overall total present value. The first estimate of the benefit-cost ratio is the ratio of the overall total benefit over the total program cost, which is 21,151 (2017 US dollars). The second estimate employs the same overall total benefit but multiplies the total program cost by 1.5 to account for the welfare loss that would be generated by collecting the taxes required to fund the program. We assume a deadweight loss of 50 cents per dollar, which is in the upper end of the range reported in the literature (Feldstein, 1999).

Treated individuals spent fewer years in special education and K-12 grade retention and increased their years in formal education. The benefits from these education categories essentially offset each other. The higher earnings of the treatment group generate an increase in the state and federal earnings tax base. These higher earnings save the government 5.9 thousand of 2017 USD in transfers through means-tested safety-net programs. We estimate an average benefit from after-tax labor income of 43.4 thousand. Better criminal behavior in the treatment group decreases criminal justice costs and monetary costs to victims by 79.3 thousand, which statistically exceeds zero with 5% significance. Health expenditure increases because individuals in the treatment group have higher income. Higher-income individuals incur higher governmental and private medical expenditures, consistent with the literature (Pashchenko and Porapakkarm, 2016). The better health conditions far outweigh health expenditures when quantified using QALYs. The gain in the health QALYs is due to

17

Table 3. Life-Cycle Present Value and Benefit-Cost Ratio for the Original Participants of the Perry Preschool Project

Present Values in 1,000s of 2017 USD	Estimate	(%Δ)
Education Present Value		
Total	0.27	(-.29%)
[s.e.]	[2.60]	
Income Present Value		
Transfers[†]	5.90	
Federal Taxes	10.96	
State Taxes	2.74	
After-Tax Labor Income	43.43	
Total	**61.58**	(50%)
[s.e.]	[32.27]	
Crime Present Value		
Criminal Justice System Cost	19.21	
Monetary Cost to Victims	60.13	
Total	***79.34***	(-47%)
[s.e.]	[65.10]	
Health Present Value		
Government Expenditure	-2.00	
Private Expenditure	-9.00	
QALY	59.66	
Total	48.69	(4.1%)
[s.e.]	[78.50]	
Overall Total Present Value	***189.88***	(18%)
[s.e.]	[108.51]	
Benefit-Cost Ratio		
Baseline Program Cost	***8.98***	
[s.e.]	[5.13]	
Add Deadweight Loss (50%)	***5.98***	
[s.e.]	[3.42]	

Note: This table summarizes our preferred estimates of the monetized average life-cycle treatment-control difference or present value of the Perry Preschool Project for the original participants, as well as the corresponding benefit-cost ratios. We bold (italicize) the present values per outcome total and overall total and benefit-cost ratios when they are significant at the 10% (5%) level based on their bias-corrected accelerated bootstrap confidence intervals. The null hypothesis for the present value per outcome is that it is less than or equal to 0. The null hypothesis for the overall total is the same. The null hypothesis for the benefit-cost ratio is that it is less than or equal to 1. The present values are adjusted for compromises in the randomization protocol, attrition, and item non-response. They are in 2017 US dollars and discounted to the year in which the program started using a rate of 3%. We show the present value per outcome, the overall total (addition of outcome totals), the benefit-cost ratio using the baseline program cost (21,151 of 2017 US dollars), and the benefit-cost ratio multiplying by 1.5 the baseline program cost to account for the deadweight that would be generated by collecting the taxes required to fund the program. The standard errors in brackets are bootstrapped and clustered at the household level.

%Δ: For the outcome total and overall total present values, 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.

better health behavior, especially for women (see Table 2). Table 2 shows that this better health behavior leads to a lower mortality rate at age 54, which persists throughout the health simulation and partly explains the difference in the health gains between men and women (see Figures 1a and 1b).

Table 3 includes the benefits for which the monetization does not introduce major uncertainty (education, labor income, criminal justice and health-system state costs, and crime victimization costs). The monetization of these outcomes is based on data and well-documented costs. This specification produces a benefit-cost ratio of 9.0 (s.e. 5.1) when only considering the total program cost and of 6.0 (s.e. 3.4) when considering the deadweight loss that would be generated by collecting the taxes required to fund the program. Both estimates exceed one statistically at a level of 5% significance. Table 3 does not include benefits due to reductions in the QALY cost to crime victims. We report estimates including these QALY costs as supplementary material in Appendix Table A4.1 because they are based on conjectures in the literature (e.g., Miller et al., 2020) and not on a model. The supplementary specification, including this crime QALY costs, produces a benefit-cost ratio of 17.2 (s.e. 10.8, p-value $< .05$) when only accounting total program cost and of 11.5 (s.e. 7.2, p-value $< .05$) when factoring in the deadweight loss.

4.1 The Age Profile of the Life-Cycle Benefits

Our life-cycle monetization of benefits starts at age 16 for income and crime. For education, we monetize essentially full records, including K-12 outcomes. For health, our monetization starts at age 18 as explained in Section 3.6. Figure 2 displays the age profile of the accrued benefits by outcome. We omit education costs and health benefits before age 21 because their contribution turns out to be negligible. We display the cumulative benefits by age segments, together with the accumulated benefit at each segment's endpoint. For example, the 21 to 40 segment starts with a negative age-21 benefit because some treatment participants are still full-time students. Cumulative benefits become positive at age 26. They reach a present

value of 56 thousand 2017 USD at age 40.

Figure 2 shows that most income and crime benefits accumulate between ages 21 and 40. PPP participants belong to a disadvantaged, credit-constrained population. Treatment boosts human capital and increases availability to work due to not being imprisoned. It increases income and relaxes credit constraints. Hai and Heckman (2017) show that individuals are the most credit constrained during their 20s and 30s. That is the age range when PPP's benefits materialize. These benefits also impact children of treated participants because they increase parental resources at childrearing ages. García et al. (2021) show that children of treated participants do better than children of control participants in part because they grow up in better environments, which mediate the intergenerational impacts monetized in Section 6.

Figure 2 also indicates that the life-cycle benefits from income and crime are virtually zero after age 40. We thus limit our life-cycle analysis to ages 16 to 40 for these two outcomes in Table 3, as well as in the remainder of the paper. The empirical bootstrap distributions of our estimates indicate that adding data from the segment after age 40 greatly increases the variance of our estimates, while barely changing the point estimates. Least absolute-value shrinkage and selection operator (LASSO) generalizations of Equation (4) further justify limiting the life cycle to ages 16 to 40 for income and crime (see Appendix A3.1). For education, we use the full records. For health, we consider the full age range until realized or forecasted death. Health outcomes throughout the life cycle are of interest and differ from zero after age 40, especially for women (see Figure 1b).

The core benefits of the program accumulate between ages 21 and 40. In this period, PPP generates 9.5 dollars not counting the deadweight costs (s.e. 3.1) and 6.3 dollars when counting the deadweight costs (s.e. 2.1). In the other segments, benefits are negligible. The benefit-cost ratios counting deadweight costs are the following: 16 to 20 segment, -1.1 (s.e. 0.9); 41 to 54 segment, 0.3 (s.e. 1.2); 55-99 (or death) segment, 0.4 (s.e. 0.8) (see Appendix

Figure 2. Life-Cycle Present Value by Outcome and Age Segments

(a) Income
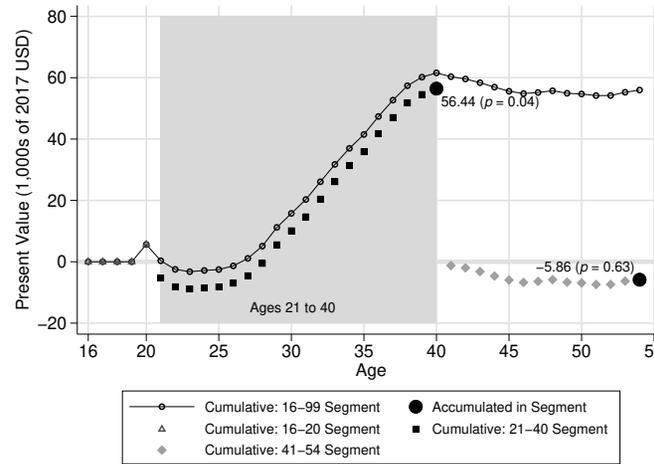
(b) Crime
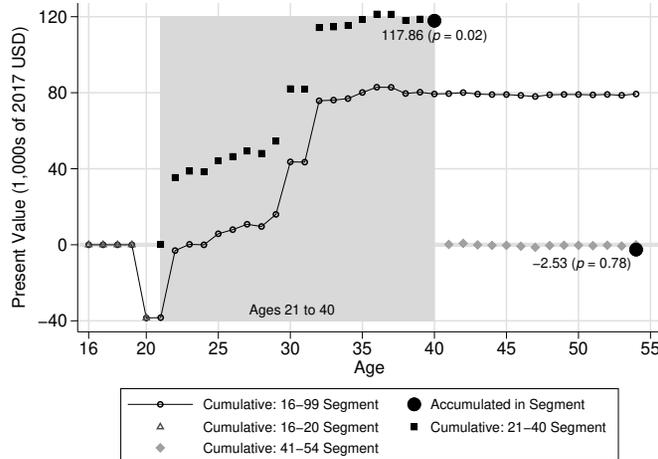
(c) Health
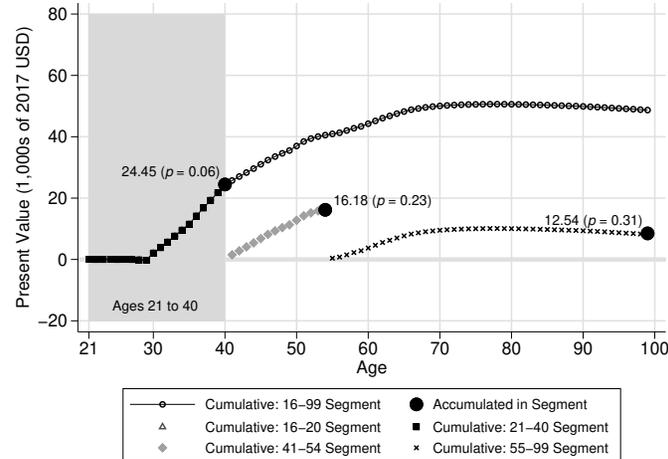
Note: Panel (a) shows the average life-cycle treatment-control difference or present value of the Perry Preschool Project due to income. The present value is shown as cumulative for the labeled age segments. We also present the accumulated present value by segment together with its p-value for segments after age 21. The null hypothesis for the accumulated present value is that it is less than or equal to 0. Present values are based on the OLS estimator discussed in Section 3.1 and are discounted to the year in which the program started using a 3% rate. Panels (b) and (c) are analogous in format to Panel (a) for crime (without quality-adjusted life years for the crime victims) and health outcomes (governmental and private medical expenditures plus quality health-adjusted life years). For income and crime, we do not monetize outcomes after age 54.

Table A4.2).

4.2 Sensitivity Analysis of Estimators and Inference Methods

Table 4 analyzes whether our main results in Table 3 are robust to the use of alternative estimators and inference procedures. First, we show that the average life-cycle present value by outcome is very similar across the three estimators in Section 3.1. There are no qualitative changes across estimates obtained using the different estimators. The magnitudes are essentially aligned. The same is true for the overall total and the two benefit-cost ratio estimates we present for each estimator. We also show that our inference regarding the overall social efficiency of PPP would have been the same had we used alternative procedures. These alternatives include standard analytic asymptotic p-values and various bootstrap procedures. The bootstrap procedures are based on regular "simple" bootstrap, trimmed bootstrap in which we trim the bottom and top 1% of the empirical bootstrap distributions before computing inference to assess the support issues when using AIPW discussed in Appendix A3.2, and studentized bootstrap p-values suggested in Heckman and Karapakula (2019, 2021). Appendix A3.2 provides details on all our inference procedures, although we rely in the text on standard inferential procedures (e.g., Hansen, 2021).

4.3 Sensitivity Analysis of Externally Supplied Parameters

Table 5 analyzes the sensitivity of our main results to externally supplied parameters that we employ to estimate benefits. We take these parameters as given from different sources and cannot account for their uncertainty in our inference. We thus explore how our preferred estimate of the benefit-cost ratio in Table 3 changes when perturbing them. Column [1] shows the value of the parameter used when obtaining our preferred estimate in Table 3. The perturbations consider pessimistic and optimistic scenarios. In the pessimistic (optimistic) scenarios, we multiply the parameter by 1.5 (0.5) if it represents a cost and by 0.5 (1.5) if it represents a benefit. We perturb one parameter at a time. We reestimate the benefit-cost ratio for each individual perturbation. Under the several perturbations, the reestimated

Table 4. Life-Cycle Present Value and Benefit-Cost Ratio, Robustness Checks for Estimators and Inference

	Main Specification		Alternative *p*-values				Maximum *p*-value
	Estimate	*p*-value	Analytic	Bootstrap			
Present Values in 1,000s of 2017 USD				Simple	Trimmed	Studentized	
Panel (a): MD							
Education Present Value	1.58	.280	.270	.275	.263	.276	.280
Income Present Value	60.83	.042	.025	.026	.018	.037	.042
Crime Present Value	70.24	.001	.138	.140	.124	.144	.144
Health Present Value	37.15	.260	.283	.286	.271	.285	.286
Overall Total Present Value	169.80	.027	.043	.052	.039	.060	.060
Benefit-Cost Ratios							
Baseline Program Cost	8.03	.043	.066	.077	.062	.089	.089
Add Deadweight Loss (50%)	5.35	.057	.081	.093	.077	.100	.100
Panel (b): OLS (Table 3)							
Education Present Value	0.27	.441	.456	.458	.455	.439	.458
Income Present Value	61.58	.058	.017	.028	.020	.035	.058
Crime Present Value	79.34	.000	.105	.111	.094	.120	.120
Health Present Value	48.69	.197	.225	.268	.251	.256	.268
Overall Total Present Value	189.88	.016	.024	.040	.029	.048	.048
Benefit-Cost Ratios							
Baseline Program Cost	8.98	.024	.039	.060	.046	.066	.066
Add Deadweight Loss (50%)	5.98	.034	.049	.072	.058	.075	.075
Panel (c): AIPW							
Education Present Value	0.88	.369	.352	.366	.356	.354	.369
Income Present Value	67.93	.033	.079	.019	.013	.030	.079
Crime Present Value	74.62	.096	.107	.288	.165	.141	.288
Health Present Value	41.02	.240	.285	.287	.272	.261	.287
Overall Total Present Value	184.45	.049	.040	.122	.053	.059	.122
Benefit-Cost Ratios							
Baseline Program Cost	8.72	.063	.061	.151	.076	.085	.151
Add Deadweight Loss (50%)	5.81	.081	.074	.167	.090	.093	.167

Note: This table presents the baseline present values by outcome, overall total, benefit-cost ratios, and *p*-values corresponding to our main specification (Table 3). We provide estimates for the three estimators in Section 3.1. The baseline *p*-value corresponds to the bootstrap bias-corrected accelerated confidence intervals that we present throughout the paper. The alternative *p*-values are analytic asymptotic or based on the different bootstrap procedures indicated in the label. The column labeled maximum *p*-value displays the maximum *p*-value among all of the *p*-values in each row. Details on all of the inference procedures are in Appendix A3.2. The null hypothesis for the present value per outcome is that it is less than or equal to 0. The null hypothesis for the overall total benefit is the same. The null hypothesis for the benefit-cost ratio is that it is less than or equal to 1.

Table 5. Benefit-Cost Ratio, Robustness Checks for Externally Supplied Parameters

Baseline Benefit-Cost (BC) Ratio:		Pessimistic Scenarios		Optimistic Scenarios	
5.98 [3.42]	Baseline Parameter [1]	Parameter Multiplier [2]	Reestimated BC-Ratio [3]	Parameter Multiplier [4]	Reestimated BC-Ratio [5]
Education Costs					
K-12	$8,665	×1.5	*5.90* [3.41]	×0.5	*6.07* [3.43]
Special Education	$18,803	×0.5	*5.93* [3.41]	×1.5	*6.04* [3.43]
College	$13,768	×1.5	*6.02* [3.41]	×0.5	*5.95* [3.43]
Crime Costs					
Incarceration	Listed in Appendix A3.5	×0.5	*5.87* [3.34]	×1.5	*6.10* [3.51]
Cost to Victim		×0.5	*5.04* [3.06]	×1.5	*6.93* [3.92]
Criminal Justice		×0.5	*5.80* [3.38]	×1.5	*6.17* [3.46]
Victimization Inflation		×0.5	*5.04* [3.06]	×1.5	*6.93* [3.92]
Health Benefits					
Value of Statistical Life	$100,000	×0.5	*5.04* [2.68]	×1.5	*6.92* [4.44]
Cash-Flow Discount Rate	3%	×$\frac{5}{3}$	*3.18* [1.86]	×0	*17.15* [11.63]

Note: Column [1] presents the externally supplied parameters that we use for monetizing treatment effects in Table 3. These parameters are *per annum* costs or benefits. We then present pessimistic (optimistic) scenarios where the parameters are multiplied by 1.5 (0.5) if they represent a cost and multiplied by 0.5 (1.5) if they represent a benefit. We reestimate our preferred estimate (baseline) of the benefit-cost ratio when perturbing individually each parameter (one at a time). The standard errors in brackets are bootstrapped and clustered at the household level. We bold (italicize) the benefit-cost ratios when they are significant at the 10% (5%) level based on their bias-corrected accelerated bootstrap confidence intervals. The null hypothesis for the benefit-cost ratio is that it is less than or equal to 1.

benefit-cost ratios remain statistically significant and similar in magnitude in most cases. Even rather substantial perturbations do not generate a drastic deviation from our preferred benefit-cost ratio estimate.[20]

4.4 Internal Rates of Return

An alternative way to measure the benefits of PPP uses internal rates of return. We do not use them as main parameters because their existence and uniqueness depend on a single-crossing property that is standard for obtaining the roots of a polynomial (the benefit can only go from negative to positive a single time across the life cycle; see Hirshleifer, 1958). Because our sensitivity analyses are extensive, there are several cases where this property fails. Additionally, computing bootstrap inference would require that the single-crossing property holds in each bootstrap draw. The condition could fail to hold due to life-cycle segments with sizable, negative treatment effects.

For our main results in Table 3, the single-crossing property holds for each outcome we analyze.[21] We provide a visual representation in Figure 3. For each outcome, we display the average life-cycle benefit using the three estimators discussed in Section 3.1. We mark the intersection with the total program cost (21,151 of 2017 US dollars) and with the total program cost multiplied by 1.5 to account for the deadweight that would be generated by collecting taxes required to fund the program. The intersections are the corresponding internal rates of return. The internal rate of return is at least 4.0% when considering each outcome separately. When considering all of the benefits that we monetize, the internal rate of return more than doubles to at least 8.9%. As a benchmark, we note that the average annual return to ten-year stock-market investments in the US during the last 140 years was

---

[20]Appendix Table A4.2 also analyzes the sensitivity of the results in Table 3 to choices other than those regarding the estimator, inference, or externally supplied parameters. We impute missing labor income by linearly interpolating or by using the method in García et al. (2020) (see also Section 3.4). We also set the benefits from each outcome (one at a time) to zero. Our main inferences are robust to these perturbations, and the benefit-cost ratio estimates remain well above one in magnitude and statistical significance.

[21]Education is an exception. Recall that its contribution turns out to be negligible. Education benefits are offset by its costs.

Figure 3. Life-Cycle Present Value and Internal Rates of Return of the Perry Preschool Project

(a) Before-Tax Labor Income

(b) Crime

(c) Health (Expenditure + QALYs)

(d) All Outcomes
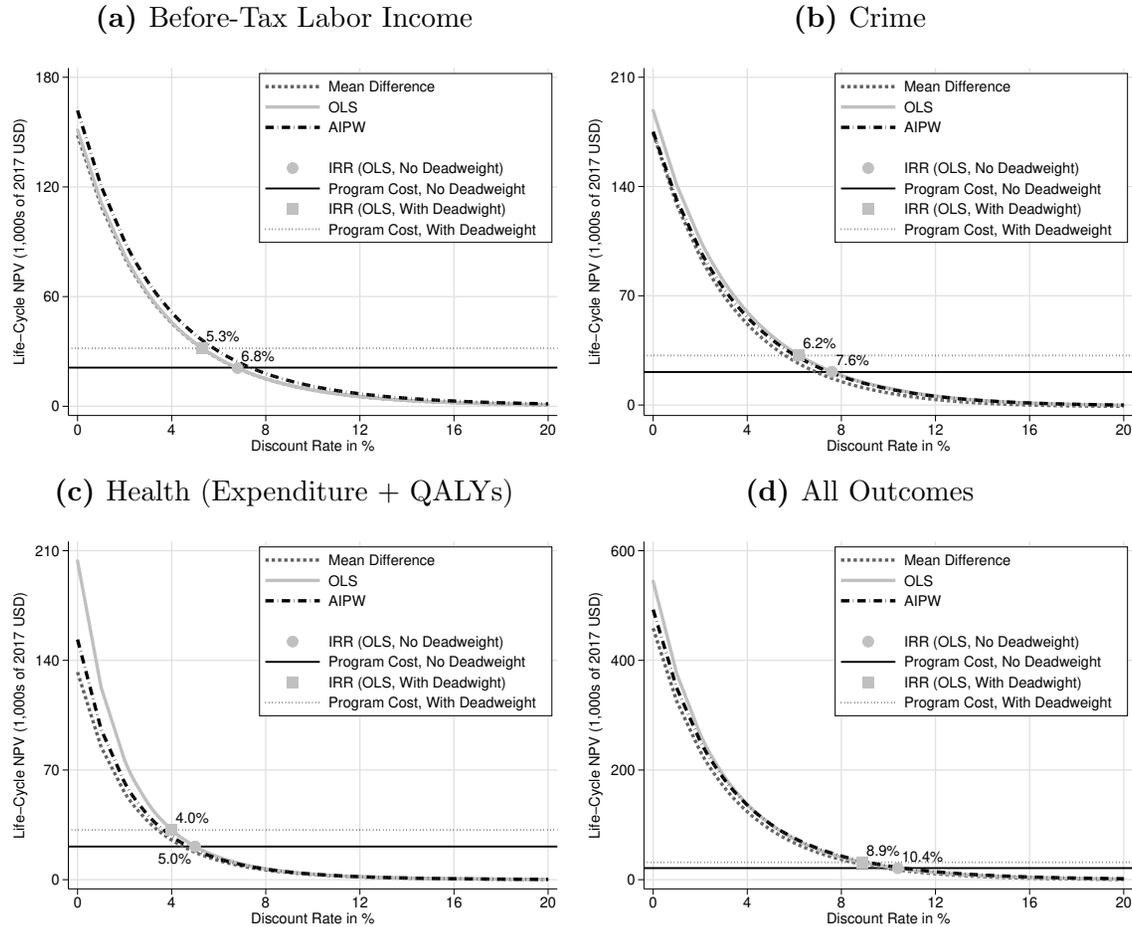
Note: Panel (a) A point in each of the lines represents the average life-cycle treatment-control difference or present value from before-tax labor income in 1,000s of 2017 US dollars, using the discount rate indicated in the horizontal axis. Results using the three estimators discussed in Section 3.1 are displayed. We mark the intersection with the total program cost (21,151 of 2017 US dollars), and with the total program cost multiplied by 1.5 to account for the deadweight that would be generated by collecting the taxes required to fund the program. The intersections are the corresponding internal rates of return. Panel (b) Analogous in format to Panel (a) for crime costs related to the criminal justice system and monetary cost to victims (without quality-adjusted life years for the crime victims). Panel (c) Analogous in format to Panel (a) for the value of governmental and private medical expenditures plus quality health-adjusted life years, assigning a value of statistical life of 100,000 (2017 US dollars). Panel (d) Analogous in format to Panel (a) for all of the benefits that we monetize (education, income, crime, and health).

very similar (9.2%, S&P Global Market Intelligence, 2020). Another relevant comparison is the 9.4% average annual stock-market return from 1960 (around when PPP started) to 2020 (French, 2021). The evidence on internal rates of return supplements the benefit-cost ratio estimates, and indicates that investing in PPP is socially efficient.

5. Comparing Forecast Methods with Actual Data

The recent literature in economics evaluating the impacts of early childhood education forecasts before-tax life-cycle *earnings* from short-run measures of test scores, joined with non-experimental datasets relating later-life earnings to test scores (e.g., Chetty et al., 2011; Deming, 2009; Duncan and Sojourner, 2013; Ganimian et al., 2021; Kline and Walters, 2016). These authors do not forecast full benefits of programs. They also do not eliminate any spurious bias between the non-experimental measures of test scores and earnings that is removed by conducting experiments. Table 6 compares our estimates of the life-cycle benefit based on (observed) before-tax labor income to forecasted earnings.

Kline and Walters (2016) estimate the treatment effect of Head Start on the average of two post-treatment cognitive-test scores at age 4.5.[22] They report an increase of 0.247 of a standard deviation. They follow Chetty et al. (2011) and assume that a one standard-deviation increase in a cognitive-test score produces an increase of 13.1% in annual earnings at age 27. They assume that the age-27 earnings gain recurs every year until retirement. They discount the life-cycle projection of earnings to age 3 (when Head Start begins), which is the same age to which we discount. They estimate a value of 438,000 (2014 USD), which they adjust to 343,392 (2014 USD) to account for the fact that Head Start beneficiaries are less skilled and have flatter earnings profiles. We convert their figure to 2017 dollars. Kline and Walters (2016) calculate the life-cycle benefit from labor income of Head Start as

---

[22]They use data from children in the Head Start Impact Study (HSIS), a one-year randomized trial of the Head Start Program.

$$\underbrace{355,578}_{\text{projection}} \times \underbrace{0.131}_{\text{``return to IQ''}} \times \underbrace{0.247}_{\text{IQ gain}} \approx 11,505 \text{ (2017 USD)}.[23]$$

To apply the method in Kline and Walters (2016) to our data, we estimate the treatment effect in the average of two post-treatment cognitive-test scores (ages 5 and 6) and find an increase of 0.46 of a standard deviation. We calculate the corresponding gain: $355,578 \times 0.131 \times 0.46 \approx 21,527$ (2017 USD). We take the life-cycle projection of earnings and "return to IQ" as parameters and calculate standard errors based on the standard error of the treatment effect on the average of the cognitive-test scores. The Kline and Walters (2016) method generates a life-cycle present value of roughly 1/3 of the estimate from our empirically based procedure. Such underestimation is expected because the main mediators of the treatment effects on adult outcomes as labor income are non-cognitive skills, not cognitive skills (Heckman et al., 2013). A prediction based solely on cognitive-test scores neglects important determinants of earnings.

Table 6 also compares actual data to forecasts from the method of García et al. (2020). This is a principled approach using economic models that incorporate exogenous boosts in cognition and non-cognitive skills developed to forecast the life-cycle benefits of an entirely different early childhood education program. It uses the causal effect of psychological measurements on earnings. It performs remarkably well when adapted to the PPP sample. One reason for its good fit is that it utilizes well-established dynamic models of earnings and also uses the treatment effects of both cognitive and non-cognitive skills to forecast equations. Adding program benefits beyond earnings obviously leads to larger estimates than are obtained focusing only on earnings.

---

[23]The appendix of Kline and Walters (2016) provides more details. The authors interpret their calculation as a marginal life-cycle benefit from enrolling an additional student to Head Start, although it is really an average benefit. Our estimates throughout the paper represent the average benefit. In disadvantaged, homogeneous samples like PPP and HSIS, the average benefit is likely a good approximation of the marginal benefit. Kline and Walters (2016) motivate their calculation with the 13.1% "return to IQ" from Chetty et al. (2011). They round this down to 10% in their main results, intending to be conservative. The results using 10% and 13.1% are similar. Both underestimate program benefits.

Table 6. Comparing Methods for Estimating the Life-Cycle Present Value from Labor Income

Method	Description of Future-Earnings Estimation	Age Range	Present Value
Kline and Walters (2016)	Forecast using "return to IQ"	Life-cycle Projection	21.53 [5.11]
García et al. (2020)	Forecast using economic models	16-60	62.97 [40.80]
This Paper	Observation	16-40	61.58 [32.27]

Note: This table compares methods for obtaining the average life-cycle treatment-control difference or present value from labor income in 1,000s of 2017 USD. The method in Kline and Walters (2016) is explained in Section 5. The method in García et al. (2020) and the method in this paper are explained in Section 3.4. For both, we display OLS estimates taken from Appendix Table A4.2 and Table 3, respectively. The standard errors in brackets are bootstrapped and clustered at the household level.

6. Adding Intergenerational and Intragenerational Impacts

It is of interest to examine if there are dynastic benefits to PPP. To this end, we first monetize the benefits to the children of the original participants focusing on their education, income, and criminal activity. The limited life-cycle information on them makes it difficult to credibly value their health benefits.[24] We also monetize the benefits to the siblings of the original participants. For much of the same reasons as for children, we focus on their education, income, and criminal activity. Monetization of child and siblings outcomes is an exploratory exercise that supplements our main analysis. For two reasons this requires strong assumptions. First, we have only one observation of child and sibling outcomes (obtained from the age-54 interview with the original participants), although we have retrospective data at that interview. Data cannot be validated with the same consistency as possible with the panel observations on the original participants. Second, children and siblings of original

---

[24]For education, the PPP participants report valid predictors for their children's life-cycle behavior, such as "ever in special education" or "high school graduation." For income, these variables are good predictors. For crime, "ever arrested in youth" is also a good predictor of adult criminal activity (Henry et al., 1999; Piquero and Moffitt, 2005; Wright et al., 1999). PPP participants report whether their children are "in good health." This variable is limited in predictive power, thus compromising estimates of child health benefits. Nonetheless Table 2 suggests a positive intergenerational impact on health. Excluding health benefits attenuates our estimates of the intergenerational impact of PPP.

participants were born in different years and this requires adjustment. We observe their outcomes at different ages so assessments of life-cycle consequences are less precise.

6.1 Monetizing Child Impacts

We use the mediation framework in Heckman et al. (2013) to predict missing outcomes. In doing so, we proceed as follows. We use the sample of original participants to forecast the life-cycle present value of outcomes using early-adulthood indicators of behavior. We apply these estimated relationships to the child sample. This approach ignores possible cohort effects and the fact that some data are missing for the children. It follows a well-established synthetic cohort methodology (Heckman et al., 2006).

For example, let $\mathcal{Y}_{\mathrm{crime}}$ denote the life-cycle present value of a crime outcome (e.g., criminal justice system cost) and let y_{crime} denote an early-adult crime behavior (e.g., "ever arrested at age 22") that we measure for most of the children. We estimate the relationship

$$\mathcal{Y}_{\mathrm{crime}} = \chi^0 + \chi^1 \, y_{\mathrm{crime}} + \upsilon_{\mathrm{crime}}, \tag{5}$$

for the original participants where χ^0 and χ^1 are parameters and $\upsilon_{\mathrm{crime}}$ is a mean-zero forecast error. The predicted life-cycle present value for the child is $\widehat{\mathcal{Y}}_{\mathrm{crime}}^{\mathrm{child}} = \widehat{\chi}^0 + \widehat{\chi}^1 \, y_{\mathrm{crime}}^{\mathrm{child}}$, where $y_{\mathrm{crime}}^{\mathrm{child}}$ is the crime behavior of the child reported by the parent in the age-54 interview. To standardize age differences in the observation of $y_{\mathrm{crime}}^{\mathrm{child}}$ (children of the participants were born in different years), we replace $y_{\mathrm{crime}}^{\mathrm{child}}$ with its predicted value at a common age for all of the children of the participants.[25] We proceed in an analogous way in monetizing education and income.[26] For education, the early-adulthood behaviors that we use as predictors are

---

[25]We use the Probit predictions summarized in Table 2. We show that the results in this section (Table 7) are very similar when replacing the Probit predictions with either linear-probability or Logit predictions in Appendix Tables A4.3 and A4.4.

[26]We update the special education, K-12, and college costs to figures corresponding to twenty years after the figures reported for the participants in Section 3.3 before estimating Equation (5). This update accounts for growth in education costs. We adjust income for real growth using the rates in US Census Bureau (2020). More details on this adjustment are in Appendix A3.7. We do not perform adjustments for crime because the information is limited to what we use for the participants in Section 3.5.

"ever in special education" for predicting the benefits from time spent in special education, and "high school graduation," "some college," and "college graduation" for predicting the benefits from time spent in formal education. For crime, we use "ever arrested in youth." For income, we use the same forecast variables used to forecast educational outcomes in addition to an indicator of employment status. Table 2 summarizes the early-adulthood predictor variables used.

Appendix A3.7 gives a detailed description of our procedure. Identification relies on four key assumptions: 1) Missing values occur randomly, conditional on baseline variables (missing values occur either because of item non-response or because some participants do not have children); 2) Equation (5) is valid across the two generations (i.e., no cohort effects); 3) Behaviors used for predicting the present value in Equation (5) mediate the treatment effect (and are measured at the appropriate age). For example, in the case of crime, these conditions imply that $\mathcal{Y}_{\text{crime}}$ is mean-independent of treatment status once y_{crime} is controlled for. We use an OLS mean-difference adjustment as in Section 3 to compute treatment effects on the predicted present values.[27]

6.2 Monetizing Sibling Impacts

We use essentially the same method employed to monetize child impacts to monetize sibling benefits. Recall from Section 2 that 1) we only consider older siblings born before PPP started who are at most five years older than the original participants;[28] and 2) we do not require age adjustments of the behavior measures (older siblings are observed at midlife, and their main outcomes are unlikely to change thereafter).[29] We use slightly different predictors

---

[27]The present values for the first generation in Table 7 are identical to the present values using the OLS estimator reported in Table 3, except for income. We use the method for monetizing income in García et al. (2020) for the first generation in this section. This method makes the computation of the bootstrapped standard errors easier because its interpolation and extrapolation allow for wider support. Appendix Table A4.2 shows that either method for monetizing income yields similar point estimates. The change in the method for monetizing income implies a very small difference between Table 3 and Table 7 (from which we take the estimates for Table 1) in the estimates of overall total and benefit-cost ratios.

[28]Misspecification due to cohort effects should be negligible in this case.

[29]The only exception is the employed or retired indicator. We age-adjust this variable using OLS. We pool the sample of original participants and their siblings and regress an employed or retired indicator on

for the siblings to find better fit in the relationships estimated in the original-participant sample. For education, we use the interaction of education categories (no high school, high school, some college, college) with an indicator of "ever being suspended from school." For income and crime, we use an employment or retirement indicator, education categories, and an interaction term for "ever being suspended from school" with "ever being arrested." Table 2 summarizes the sibling behaviors used in the predictions.

6.3 Estimates

Table 7 presents our estimates of dynasty and extended-dynasty benefits.[30] It expands the estimates initially summarized in Table 1. The benefit from education is negative for the second generation (children). This is because the education of the children of the treatment-group participants increases, raising direct costs. The same is true for the siblings. However, in our forecast models, education predicts or mediates income. Thus, the program's sizable intergenerational and intragenerational impact on education increases the average life-cycle income of the children by 27 thousand (2017 USD) and of the average life-cycle income of the siblings by 57 thousand (a large fraction of the big difference between the intergenerational and intragenerational impacts is explained by discounting). Similarly, the lower arrest rates of the children and siblings of treated PPP participants generate a predicted positive life-cycle benefit from better crime behavior of 6.7 and 5.2 thousand, respectively.

The benefits for the children add 32.7 thousand (2017 USD) to the benefit of the program, while the benefits for the siblings add 60.6 thousand. The full-dynasty life-cycle overall present value of the program increases to 286 thousand, which is statistically significant at the 5% level. Combining the evidence for the first and second generations with that of the siblings, the extended-dynasty benefit-cost ratio is a statistically significant 13.5 not counting deadweight costs and 9.0 counting deadweight costs.

---

a second-degree age polynomial. We use the age-60 prediction as the relevant behavior. We report the raw variable in Table 2 because this age adjustment is minor.

[30]Our bootstrap inference draws households of participants together with all their children and siblings.

Table 7. Dynastic Present Value and Benefit-Cost Ratio of the Perry Preschool Project

Present Value in 1,000s of 2017 USD	[1] 1st Generation (Original Participants)	[2] Siblings (Intragenerational)	[3] Children (Intergenerational)	[4] Dynasty ([1]+[3])	[5] Extended Dynasty ([1]+[2]+[3])
Education Present Value	0.27	-1.42	-0.98	-0.71	-2.13
[s.e.]	[2.57]	[2.98]	[0.79]	[2.78]	[4.26]
Income Present Value	*64.25*	*56.79*	26.98	*91.23*	*148.03*
[s.e.]	[29.79]	[33.22]	[27.57]	[43.69]	[58.47]
Crime Present Value	*79.34*	5.23	**6.74**	*86.08*	*91.31*
[s.e.]	[67.86]	[41.81]	[5.98]	[70.15]	[85.24]
Health Present Value	48.69	N/A*	N/A*	48.69	48.69
[s.e.]	[75.23]			[75.23]	[75.23]
Overall Total Present Value	*192.55*	60.61	32.74	*225.30*	*285.90*
[s.e.]	[106.58]	[55.59]	[29.71]	[113.12]	[133.15]
Benefit-Cost Ratio					
Baseline Program Cost	*9.10*	2.87	1.55	*10.65*	*13.52*
[s.e.]	[5.04]	[2.63]	[1.40]	[5.35]	[6.30]
Add Deadweight Loss (50%)	*6.07*	1.91	1.03	*7.10*	*9.01*
[s.e.]	[3.36]	[1.75]	[0.94]	[3.57]	[4.20]

Note: This table summarizes the monetized average life-cycle treatment-control difference of the Perry Preschool Project for the original participants (first generation), their siblings, their children (second generation), the dynasty (addition of the first and second generations), and the extended dynasty (addition of the first generation, their siblings, and the second generation). All figures are discounted to the year in which the program started using a rate of 3%. For the participants, we consider our main estimates using the OLS estimator. For the siblings and the children of the participants, we use the method explained in Section 6. We display the present value per outcome, the overall total (addition of outcome totals), the benefit-cost ratio using the baseline program cost (21,151 of 2017 US dollars), and the benefit-cost ratio multiplying by 1.5 the baseline program cost to account for the deadweight that would be generated by collecting taxes required to fund the program. The standard errors in brackets are bootstrapped and clustered at the original-participant household level. We bold (italicize) the present values per outcome and overall totals and benefit-cost ratios when they are significant at the 10% (5%) level based on their bias-corrected accelerated bootstrap confidence intervals. The null hypothesis for the present value per outcome is that it is less than or equal to 0. The null hypothesis for the overall total is the same. The null hypothesis for the benefit-cost ratio is that it is less than or equal to 1.

*N/A means not available (health is not monetized for the siblings and the children).

7. Summary

This paper monetizes and aggregates the life-cycle benefits of the Perry Preschool Project for the original participants and the additional benefits accruing to their siblings and children. The program increases labor income and reduces crime and the cost of the criminal justice system. It also improves health and healthy behaviors. Treatment-group members have higher medical expenditure because they have higher income.[31] The ratio of program benefits to the total program cost is 9.1 (s.e. 5.1). Accounting for the deadweight generated by collecting taxes required to fund the program, the estimate decreases to 6.1 (s.e. 3.4)—our preferred estimate. These estimates discount the benefits of the program to the year in which the program started using a rate of 3%. Recent arguments indicate that benefits of social programs should be discounted at a rate less than 3%.[32] Undiscounted, our preferred estimate of the benefit-cost ratio is 17.2 (s.e. 11.6). We show that these estimates are robust to extensive sensitivity analyses of the different assumptions used to produce results.

We use our long-run measures of intervention outcomes to assess the validity of current, widely used procedures of forecasting life-cycle earnings using cognitive-test scores measured at early ages. We find that this approach is flawed for two reasons: 1) Cognitive-test scores are a weak reed to lean on. Many other strongly predictive factors, primarily social and emotional skills, are not used in this approach; and 2) It ignores a myriad of other benefits of early intervention besides earnings. In our data, this approach substantially understates long-term program benefits. We show that the methodology of García et al. (2020) developed and applied to analyze a different early childhood education program works well in application to the Perry Preschool Project.

For the first time, we account for the impacts of an early childhood education program on the siblings and the children of the original participants. While our intragenerational and

---

[31]This may be a cost but also a benefit because of higher-quality medical treatments. We are conservative in treating it as a cost. However, we measure the enhanced benefits of such medical treatments using QALYs.

[32]Congressional Budget Office (2018).

intergenerational estimates are crude, the extended-dynasty cost-benefit analysis boosts our preferred estimate of the benefit-cost ratio of PPP from 6.1 (s.e. 3.4) to 9.0 (s.e. 4.2).

References

Agency for Healthcare Research and Quality (2003). MEPS HC-050: 2000 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2004a). MEPS HC-060: 2001 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2004b). MEPS HC-070: 2002 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2005). MEPS HC-079: 2003 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2006). MEPS HC-089: 2004 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2007). MEPS HC-097: 2005 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2008). MEPS HC-105: 2006 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2009). MEPS HC-113: 2007 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2010). MEPS HC-121: 2008 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2011). MEPS HC-129: 2009 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2012). MEPS HC-138: 2010 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2013). MEPS HC-147: 2011 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality (2014). MEPS HC-155: 2012 Full Year Consolidated Data File. Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality.

Almond, D. and J. Currie (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives 25*(3), 153–172.

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.

Belfield, C. R., M. Nores, W. S. Barnett, and L. J. 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.

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`.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics 90*(3), 414–427.

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.

Campbell, F. A., B. H. Wasik, E. Pungello, M. Burchinal, O. Barbarin, K. Kainz, J. J. Sparling, and C. T. Ramey (2008). Young Adult Outcomes of the Abecedarian and CARE Early Childhood Educational Interventions. *Early Childhood Research Quarterly 23*(4), 452–466.

Cascio, E. (2021). Early Childhood Education in the United States: What, When, Where, Who, How, and Why. NBER Working Paper w28722, National Bureau of Economic Research.

Centers for Medicare and Medicaid Services, U.S. Department of Health (2011a). Medicare Current Beneficiary Survey, 2007 Cost and Use Data. U.S. Department of Health and Human Services.

Centers for Medicare and Medicaid Services, U.S. Department of Health (2011b). Medicare Current Beneficiary Survey, 2008 Cost and Use Data. U.S. Department of Health and Human Services.

Centers for Medicare and Medicaid Services, U.S. Department of Health (2012). Medicare Current Beneficiary Survey, 2009 Cost and Use Data. U.S. Department of Health and Human Services.

Centers for Medicare and Medicaid Services, U.S. Department of Health (2013). Medicare Current Beneficiary Survey, 2010 Cost and Use Data. U.S. Department of Health and Human Services.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *Quarterly Journal of Economics 126*(4), 1593–1660.

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

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam.

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

Duncan, G. J. and A. J. Sojourner (2013). Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps? *Journal of Human Resources 48*(4), 945–968.

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

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.

Ermini Leaf, D. (2017). Unified Method for Markov Chain Transition Model Estimation Using Incomplete Survey Data.

Ermini Leaf, D., B. Tysinger, D. P. Goldman, and D. N. Lakdawalla (2020). Predicting Quantity and Quality of Life with the Future Elderly Model. *Health Economics*.

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

French, K. R. (2021). French-Fama Research Data Factors. Website, https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Ganimian, A. J., K. Muralidharan, and C. R. Walters (2021). Augmenting State Capacity for Child Development in India. NBER Working Paper w28780, National Bureau of Economic Research.

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 (2021). Early Childhood Education Promotes the Intergenerational Mobility of Disadvantaged African Americans. Technical report, The University of Chicago *Mimeo*.

Goldman, D. P., D. Ermini Leaf, J. Sullivan, and B. Tysinger (2018). The Future Elderly Model: Technical Documentation. Technical report.

Goldman, D. P., D. Ermini Leaf, and B. Tysinger (2018). The Future Adult Model: Technical Documentation. Technical report.

Grant, W. V. and C. G. Lind (1978). Digest of Education Statistics 1977-78.

Grant, W. V. and T. D. Snyder (1986). Digest of Education Statistics, 1985-86.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2021). The Long-Term Effects of Universal Preschool in Boston. NBER Working Paper w28756, National Bureau of Economic Research.

Hai, R. and J. J. Heckman (2017). Inequality in Human Capital and Endogenous Credit Constraints. *Review of Economic Dynamics 25*, 4–36.

Hansen, B. E. (2021). *Econometrics*. [Online; accessed 31-March-2021].

Havnes, T. and M. Mogstad (2011). No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes. *American Economic Journal: Economic Policy 3*(2), 97–129.

Heckman, J. J. and G. Karapakula (2019). The Perry Preschoolers at Late Midlife: A Study in Design-Specific Inference. NBER Working Paper w25888, National Bureau of Economic Research.

Heckman, J. J. and G. Karapakula (2021). Using a Satisficing Model of Experimenter Decision-Making to Guide Finite-Sample Inference for Compromised Experiments. *Econometrics Journal*.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. *Handbook of the Economics of Education 1*, 307–458.

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., 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.

Henry, B., A. Caspi, T. E. Moffitt, H. Harrington, and P. A. Silva (1999). Staying in School Protects Boys with Poor Self-Regulation in Childhood from Later Crime: A Longitudinal Study. *International Journal of Behavioral Development 23*(4), 1049–1073.

Hirshleifer, J. (1958). On the Theory of Optimal Investment Decision. *Journal of Political Economy 66*(4), 329–352.

Hunt, P., J. Anderson, and J. Saunders (2017). The Price of Justice: New National and State-Level Estimates of the Judicial and Legal Costs of Crime to Taxpayers. *American Journal of Criminal Justice 42*(2), 231–254.

Hunt, P. E., J. Saunders, and B. Kilmer (2019). Estimates of Law Enforcement Costs by Crime Type for Benefit-Cost Analyses. *Journal of Benefit-Cost Analysis 10*(1), 95–123.

Institute for Social Research at the University of Michigan (2019). Panel Study of Income Dynamics, Public Use Dataset. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan.

Kakalik, J. S., W. S. Furry, M. A. Thomas, and M. F. Carney (1981). *The Cost of Special Education: Summary of Study Findings*. Rand Corporation.

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.

Miller, T. R., M. A. Cohen, D. I. Swedler, B. Ali, and D. V. Hendrie (2020). Incidence and Costs of Personal and Property Crimes in the United States, 2017. *Available at SSRN 3514296*.

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.

Piquero, A. R. and T. E. Moffitt (2005). Explaining the Facts of Crime: How the Developmental Taxonomy Replies to Farrington's Invitation. *Integrated Developmental and Life-Course Theories of Offending*, 51–72.

Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association 53*(284), 873–880.

RAND HRS Version P (2016). Produced by the RAND Center for the Study of Aging, With Funding From the National Institute on Aging and the Social Security Administration.

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.

S&P Global Market Intelligence (2020). S&P 500 Returns to Halve in the Coming Decade. Website, `https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/s-p-500-returns-to-halve-in-coming-decade-8211-goldman-sachs-59439981`.

United States Department of Justice (2010). National Judicial Reporting Program, 2006.

University of Michigan and the National Institute on Aging (2013). Health and Retirement Study, (2006 Biomarker Data) Sensitive Dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).

University of Michigan and the National Institute on Aging (2015a). Health and Retirement Study, (2010 Biomarker Data) Sensitive Dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).

University of Michigan and the National Institute on Aging (2015b). Health and Retirement Study, (2012 Biomarker Data) Sensitive Dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).

University of Michigan and the National Institute on Aging (2016). Health and Retirement Study, (RAND HRS Version P) Public Use Dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).

University of Michigan and the National Institute on Aging (2017). Health and Retirement Study, (2014 Biomarker Data) Sensitive Dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).

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

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

US Census Bureau (2019). Historical Income Tables H-1 and H-3.

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.

Wright, B. R. E., A. Caspi, T. E. Moffitt, and P. A. Silva (1999). Low Self-Control, Social Bonds, and Crime: Social Causation, Social Selection, or Both? *Criminology 37*(3), 479–514.