Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program

Jorge Luis García

Clemson University

James J. Heckman

American Bar Foundation and University of Chicago

Duncan Ermini Leaf

University of Southern California

María José Prados

University of Southern California

This paper quantifies and aggregates the multiple lifetime benefits of an influential high-quality early-childhood program with outcomes measured through midlife. Guided by economic theory, we supplement experimental data with nonexperimental data to forecast the life-cycle benefits and costs of the program. Our point estimate of the internal rate of return is 13.7%, with an associated benefit/cost ratio of 7.3. We account for model estimation and forecasting error and present estimates from extensive sensitivity analyses. This paper is a template for synthesizing experimental and nonexperimental data using economic theory to estimate the long-run life-cycle benefits of social programs.

This research was supported in part by grants from the Robert Wood Johnson Foundation's Policies for Action program, the American Bar Foundation, the Buffett Early Childhood Fund, and the Pritzker Children's Initiative; by the National Institutes of Health from the Eunice Kennedy Shriver National Institute of Child Health and Human Development under

Electronically published May 21, 2020
[*Journal of Political Economy*, 2020, vol. 128, no. 7]

I. Introduction

A large body of evidence documents that high-quality early-childhood programs boost the skills of disadvantaged children.[1] Much of this research reports treatment effects of programs on cognitive test scores, school readiness, and measures of early-life social behavior. A few studies analyze longer-term benefits in terms of completed education, adult health, crime, and labor income.[2] Rigorous evidence on their long-term social efficiency is scarce.[3]

This paper investigates the social benefits and costs of an influential pair of closely related early-childhood programs conducted in North Carolina that targeted disadvantaged children. The Carolina Abecedarian Project (ABC) and the Carolina Approach to Responsive Education (CARE)—henceforth ABC/CARE—were evaluated by randomized control trials. Both programs were launched in the 1970s. Participants were followed through their mid-30s. The programs started early in life (at

awards R37HD065072 and R01HD054702; and by the National Institute on Aging under awards R01AG042390 and P30AF024968. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. The authors wish to thank Frances Campbell, Craig and Sharon Ramey, Margaret Burchinal, Carrie Bynum, and the staff of the Frank Porter Graham Child Development Institute at the University of North Carolina at Chapel Hill for the use of data and source materials from ABC and CARE. Years of partnership and collaboration have made this work possible. We thank Bryan Tysinger of the Leonard D. Schaeffer Center for Health Policy and Economics at the University of Southern California for help adapting the FAM to make the health projections used in this paper. We also thank Andrés Hojman, Yu Kyung Koh, Sylvi Kuperman, Stefano Mosso, Rodrigo Pinto, Joshua Shea, Jake Torcasso, and Anna Ziff for help on work related to this paper. For very helpful comments on various versions of the paper, we thank the editor, Harald Uhlig, and four anonymous referees, Stéphane Bonhomme, Neil Cholli, Flávio Cunha, Steven Durlauf, David Figlio, Dana Goldman, Ganesh Karapakula, Magne Mogstad, Sidharth Moktan, Tanya Rajan, Azeem Shaikh, Jeffrey Smith, Chris Taber, Matthew Tauzer, Evan Taylor, Ed Vytlacil, Jim Walker, Chris Walters, and Matt Wiswall. We benefited from helpful comments received at the Leonard D. Schaeffer Center for Health Policy and Economics in December 2016, and at the University of Wisconsin, February 2017. We thank Peg Burchinal, Carrie Bynum, Frances Campbell, and Elizabeth Gunn for information on the implementation of the ABC and CARE and assistance in data acquisition. For information on childcare in North Carolina, we thank Richard Clifford and Sue Russell. The set of codes to replicate the computations in this paper are posted in a repository. Interested parties can request to download all the files. The address of the repository is https://github.com/jorgelgarcia/abccare-cba. To replicate the results in this paper, contact the first author, who will put you in contact with the appropriate individuals to obtain access to restricted data. The appendix for this paper is posted on http://cehd.uchicago.edu/ABC_CARE. Data are provided as supplementary material online.

[1] See Cunha et al. (2006), Almond and Currie (2011), and Elango et al. (2016) for surveys.

[2] Examples include Heckman et al. (2010a), Campbell et al. (2014), and Havnes and Mogstad (2011).

[3] Belfield et al. (2006) and Heckman et al. (2010b) present a life-cycle cost-benefit analysis of the Perry Preschool Program. Our approach is more comprehensive in terms of the outcomes analyzed, in terms of providing a general methodology that can be replicated to assess the social efficiency of other programs, and in the procedure used to obtain standard errors of estimates of the parameters summarizing social efficiency.

8 weeks of life) and engaged participants until age 5. They generated numerous positive treatment effects.[4] Parents of participants (primarily mothers) received free childcare that facilitated parental employment and adult education. We find that the program has a 13.7% (SE [standard error]: 3%) per annum tax-adjusted internal rate of return and a 7.3 (SE : 1.84) tax-adjusted benefit/cost ratio.

The program is a prototype for many programs planned or in place today.[5] About 19% of all African American children would be eligible for ABC/CARE today.[6] Implementation of the ABC/CARE program in disadvantaged populations would be an effective, socially efficient policy for promoting social mobility.[7]

This paper addresses a fundamental problem that arises in evaluating social programs. Few program evaluations have complete life-cycle histories of participants. In our data, the oldest experimental subject is in his/her mid-30s. At issue is determining the life-cycle impact of the program. We forecast the full life-cycle benefits and costs of the program, using nonexperimental data guided by economic theory. Our approach is guided by the economics of the problem, in stark contrast to mechanical forecasting approaches devoid of economics (e.g., Abadie. Diamond, and Hainmueller 2010 and Chetty et al. 2011).[8]

[4] A companion paper by García, Heckman, and Ziff (2018) reports these treatment effects. Participants in ABC/CARE benefit in terms of cognitive and socioemotional skills, education, employment and labor income, and risky behavior and health. The parents of participants benefit in terms of labor income and education.

[5] Programs inspired by ABC/CARE have been (and are currently being) launched around the world. Sparling (2010) and Ramey, Sparling, and Ramey (2014) list numerous programs based on the ABC/CARE approach. The programs are the Infant Health and Development Program in eight different cities in the United States (Gross, Spiker, and Haynes 1997); Early Head Start and Head Start (Schneider and McDonald 2007); the Johns Hopkins Cerebral Palsy Study in the United States (Sparling 2010); the Classroom Literacy Interventions and Outcomes study (Sparling 2010); the Massachusetts Family Child Care Study (Collins et al. 2010); the Healthy Child Manitoba Evaluation (Healthy Child Manitoba 2015); the Abecedarian Approach within an Innovative Implementation Framework (Jensen and Nielsen 2016); and Building a Bridge into Preschool in Remote Northern Territory Communities in Australia (Scull et al. 2015). Current Educare programs in the United States are also based on ABC/CARE (Yazejian and Bryant 2012; Educare 2014). Appendix A.8 (apps. A–H are available online) lists these Educare programs, all of which implement curricula based on ABC/CARE.

[6] Forty-three percent of African American children were eligible at its inception.

[7] García and Heckman (2016) estimate that if ABC/CARE were implemented on the current stock of eligible children, the intrablack gap (black disadvantaged relative to black advantaged) in high school graduation, years of education, employment, and labor income at age 30 for females would be reduced by 110%, 76%, 22%, and 30%, respectively. It would eradicate the intrablack high school graduation gap, reduce the years-of-education gap to 0.12 years, reduce the employment gap to 14 percentage points, and reduce the labor income gap to $4,075 (2014 USD). For males, the program would eradicate the intrablack high school graduation gap, reduce the years-of-education gap to 0.18 years, and reduce the employment gap to 9 percentage points.

[8] Ridder and Moffitt (2007) discuss data combination methods. These methods are related to the older "surrogate marker" literature in biostatistics (see, e.g., Prentice

As a by-product, we also address the problem of aggregating evidence from the multiplicity of treatment effects found in ABC/CARE. We estimate economically interpretable aggregates: internal rates of return and benefit/cost ratios that monetize the large array of benefits and costs generated. In constructing these aggregates, we account for model estimation and forecasting error and the welfare cost of taxation to fund programs. Our estimates survive extensive sensitivity analyses.

We construct synthetic cohorts from nonexperimental samples. The cohorts are chosen to approximate the life cycles of experimentals in their postexperimental years. To make these approximations, we formulate and estimate production functions on nonexperimental samples that predict program treatment effects and assess their within-sample forecast accuracy in experimental samples. Some of the inputs of the estimated production functions are changed by treatment and are measured in both experimental and nonexperimental samples. If the production functions mapping inputs to outputs across cohorts are unaffected by treatment (i.e., are "treatment invariant"), we can safely use them to forecast treatment effects at older ages, provided that we accurately forecast the path of future inputs. We test and do not reject treatment invariance comparing outcomes in experimental samples with those forecasted in nonexperimental samples that overlap in age.[9] We forecast experimental treatment effects, using our estimated production functions applied to nonexperimental data with inputs and outputs. We conduct extensive sensitivity analyses for our baseline forecasting models, examining, for example, alternative assumptions about cohort effects that might characterize nonexperimental samples. We compare the outcomes of our approach with a cruder matching procedure. Estimates from it replicate those from our more sophisticated procedure.

Our analysis is a template for estimating the life-cycle gains of social experiments for which there is less than full lifetime follow-up. Supplementing experimental data with nonexperimental data enhances the information available from social experiments. Using economic theory and econometric methods to generate empirically concordant forecasts enhances the credibility of the procedure.

The quest for long-run estimates from experiments with short-term follow-up has recently led to application of informal procedures for estimating long-term benefits using short-term measures of childhood test

1989). However, as noted below, accounting for endogeneity is an integral part of the models we estimate, although it is not considered in the statistics literature. That literature does not provide testable predictions for validation of its forecasts, as we do.

[9] See Hurwicz (1962) for the definition of treatment (policy) invariance (his definition of "structural" relationships implies policy invariance). We build on the methodology of Heckman, Pinto, and Savelyev (2013), who relate intermediate and long-term outcomes in a mediation analysis. However, they do not construct out-of-sample forecasts, as we do in this paper.

scores (e.g., Chetty et al. 2011; Kline and Walters 2016). We show by example that these procedures can give very misleading estimates of true life-cycle program benefits by focusing on earnings, not counting the full array of benefits generated, and relying solely on test scores to predict future earnings.

This paper proceeds in the following way. Section II describes the ABC/CARE program. Section III discusses our methodology for forecasting life-cycle outcomes and the evidence supporting our assumptions. We examine in detail how we forecast life-cycle labor income. Section IV discusses how we forecast other life-cycle outcomes. Section V reports baseline estimates of internal rates of return and benefit/cost ratios and reports an array of sensitivity analyses. Section VI uses our estimates to examine the predictive validity of widely used informal forecasting methods and to examine the reliability of forecasts based only on measures available early on in the experiment. Section VII summarizes our findings.

II. ABC/CARE: Background

ABC and CARE were enriched childcare programs that targeted the early years of disadvantaged, predominately African American children in the area of Chapel Hill, North Carolina.[10] Appendix A describes these programs in detail. We summarize their main features here.

These early-childhood programs went well beyond providing regular care. They were high-quality, educationally focused childcare centers. Their goal was to enhance the life skills of disadvantaged children. They supported language, motor, and cognitive development as well as socioemotional competencies considered crucial for school success, including task orientation, the ability to communicate, independence, and prosocial behavior (Sparling 1974; Ramey et al. 1976, 1985; Wasik et al. 1990; Ramey, Sparling, and Ramey 2012). All treatment children received medical checkups that conformed with the American Academy of Pediatrics. Parents were notified if children had medical issues. The first cohort of the ABC control group received similar services, but not the later cohorts (Henderson et al. 1982; Campbell et al. 2014).

The design and implementation of ABC and CARE were very similar. Both had two phases. The first and main phase lasted from birth until age 5. In this phase, children were randomly assigned to treatment. The second phase of the study took place in the first three years of public schooling and supported children's academic development. It attempted to enhance parental involvement in the education of the children. A home

[10] Both ABC and CARE were designed and implemented by researchers at the Frank Porter Graham Center of the University of North Carolina in Chapel Hill.

visit took place every two weeks and provided parents home activities to complement the skills taught at school. The visitor facilitated communication between the teachers and the parents (Campbell and Ramey 1995). Children were assigned to this treatment through a second-stage randomization. The first phase of CARE, from birth until age 5, had an additional treatment arm of home visits designed to improve home environments (Wasik et al. 1990).

ABC recruited four cohorts of children born between 1972 and 1976. CARE recruited two cohorts of children born between 1978 and 1980. For both programs, families of potential participants were referred to researchers by local social service agencies and hospitals at the beginning of the mother's last trimester of pregnancy. Eligibility was determined by a score on a childhood "risk index" of disadvantage.[11]

Our analysis uses data from the first phase and pools the ABC treatment group with the CARE treatment group that received center-based childcare. We do not use the data from the CARE group that received home visits only in the early years. Campbell et al. (2014) test and do not reject the hypothesis that the CARE data through age 5 (without home visits) and the ABC data through age 5 come from the same distribution.

The initial ABC sample consisted of 120 families. Because of attrition and nonresponse, the study sample was reduced to 114 subjects: 58 in the treatment group and 56 in the control group. In CARE, the initial sample had 65 families: 23 were randomized to a control group, 25 to a family education treatment group, and 17 to a center-based childcare treatment group that followed ABC protocols.[12] We use standard weighting methodologies to account for attrition, nonresponse, and missing data (see app. C.2).

For both programs, data were frequently collected on cognitive and socioemotional skills, home environments, family structure, and family economic characteristics from birth until age 8. Further follow-ups were collected at ages 12, 15, 21, and 30. In addition, there is information from administrative criminal records and from a full in-person medical assessment that included a survey and collection of biospecimens and measurements when the subjects were in their mid-30s. Appendix A.7 provides exact details on the timing of the different data collections.[13] Many

[11] See app. A.2 for details on the construction of the "risk index" used to determine eligibility.

[12] There were no randomization compromises in CARE. During preschool, five subjects attrited (three in the treatment group, one in the family education group, and one in the control group). Details on attrition and nonresponse are presented in app. A.3.

[13] We also document the balance in observed baseline characteristics across the treatment and control groups, after dropping the individuals for whom we have no crime or health information. There are substantial missing data for these outcomes, which we address using the methodology exposited in app. C.

control-group children in both ABC and CARE attended alternative formal childcare arrangements (75% and 74%, respectively).[14]

III. Forecasting Life-Cycle Costs and Benefits: Methodology

Our forecasting procedure builds on the methodology of Heckman, Pinto, and Savelyev (2013). Experimental outcomes are modeled as the outputs of treatment-invariant technologies. Treatment affects inputs only and not the relationship between inputs and outputs. We measure how treatment affects inputs within sample and forecast future input paths resulting from experimental manipulations. We use nonexperimental data to estimate the technologies associated with treatment and use the experimental data to examine the hypothesis of treatment invariance of the technology and the ability of the estimated technology to reproduce experimental results using the input changes induced by the experiment. Given policy invariance, these input changes define the intervention and link it to changes in child environments that occur outside the experiment.

To formalize our procedure, the following notation is useful. We use $W = 1$ to indicate that parents referred to the program participate in the randomization protocol; $W = 0$ indicates otherwise. The term R indicates randomization into treatment ($R = 1$) or control ($R = 0$), and D indicates whether or not a family attends the program; $D = R$ implies compliance with the initial randomization protocol. Lowercase variables denote realizations of random variables. We suppress individual subscripts to avoid notational clutter.

Individuals are eligible to participate in the program if their baseline background variables $\boldsymbol{B} \in \mathcal{B}_0$, where $\mathcal{B}_0$ is the set of scores on the childhood risk index that determines program eligibility. Because all of the eligible individuals with the option to participate chose to do so ($W = 1$, and $D = R$), we can safely interpret the treatment effects generated by the experiment as average treatment effects for the eligible population and not just average treatment effects for the treated.[15]

[14] The alternative arrangements were generally of lower quality than ABC/CARE (see app. A.6.1 for details). In our main analysis, we compare treatment- and control-group children, irrespective of take-up of alternatives. In app. F, we address the problem of substitution bias (Heckman 1992; Heckman et al. 2000; Kline and Walters 2016). We disaggregate our analysis to distinguish treatment effects by type of alternative selected by the control group.

[15] All providers of health care and social services (referral agencies) in the area of the ABC/CARE study were informed of the programs. They referred mothers whom they considered disadvantaged. Eligibility was corroborated before randomization. Conversations with the program staff indicate that all but one of the referred mothers attended and agreed to participate in the initial randomization (Ramey, Sparling, and Ramey 2012).

Define $\boldsymbol{Y}_a^1$ as the outcome vector at age a for the treated. Then $\boldsymbol{Y}_a^0$ is the age-a outcome vector for the controls. At age a, the vector of average treatment effects for the population for which $\boldsymbol{B} \in \mathcal{B}_0$ is

$$\boldsymbol{\Delta}_a \coloneqq \mathbb{E}\left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_a^0 | W = 1\right] = \mathbb{E}\left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_a^0 | \boldsymbol{B} \in \mathcal{B}_0\right]. \tag{1}$$

Randomization identifies this parameter in the experimental sample.

A. *Using Economic Models to Make Forecasts*

This paper uses economic models to generate unbiased, out-of-sample forecasts of $\boldsymbol{\Delta}_a$. We use a structural production function (mediation) model for treatment ($D = 1$) and control ($D = 0$) outcomes at age a in sample $k \in \{\text{e}, \text{n}\}$, where "e" denotes membership in the experimental sample and "n" denotes membership in a nonexperimental (auxiliary) sample. The vector of production functions for output is

$$\boldsymbol{Y}_{k,a}^d = \boldsymbol{\phi}_{k,a}^d(\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k) + \varepsilon_{k,a}^d, \tag{2}$$

$d \in \{0, 1\}$, $k \in \{\text{e}, \text{n}\}$, and $a \in \{1, ..., \bar{A}\}$, where $\boldsymbol{\phi}_{k,a}^d(\cdot, \cdot)$ is a vector of structural production relationships mapping inputs $\boldsymbol{X}_{k,a}^d$, $\boldsymbol{B}_k$ into outputs, holding the error term $\varepsilon_{k,a}^d$ fixed.[16] The vector $\boldsymbol{B}_k$ consists of baseline variables not affected by treatment, and $\boldsymbol{X}_{k,a}^d$ are variables potentially affected by treatment.[17] We denote as $\bar{A}$ the oldest age through which benefits are projected. In the experiment we analyze, participants are observed through age $a^* < \bar{A}$.

The relationship between the inputs $\boldsymbol{X}_{k,a}^d$, $\boldsymbol{B}_k$ and outputs $\boldsymbol{Y}_{k,a}^d$ can, in principle, differ between experimental and nonexperimental samples, although in our data this is not the case. Equation (2) characterizes the outcomes of the two treatment regimes in any sample, including a nonexperimental sample with no direct empirical counterpart for the case $d = 1$. We present conditions for identifying and estimating $\boldsymbol{\phi}_{k,a}^d(\cdot, \cdot)$ in nonexperimental samples. A crucial condition is that for fixed values of inputs $\boldsymbol{X}_{k,a}^d = \boldsymbol{x}$, $\boldsymbol{B}_k = \boldsymbol{b}$, there are no differences in the technologies and in the distributions of $\varepsilon_{k,a}^d$ across treatment regimes and samples. We first formalize this assumption and then remark on its content.

ASSUMPTION A-1 (Structural invariance). For all $\boldsymbol{x}, \boldsymbol{b} \in \text{supp}(\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k)$, $k \in \{\text{e}, \text{n}\}$, and $a \in \{1, ..., \bar{A}\}$,

$$\begin{aligned} \boldsymbol{\phi}_{k,a}^0(\boldsymbol{x}, \boldsymbol{b}) &= \boldsymbol{\phi}_{k,a}^1(\boldsymbol{x}, \boldsymbol{b}) \\ &= \boldsymbol{\phi}_a(\boldsymbol{x}, \boldsymbol{b}), \end{aligned} \tag{3a}$$

[16] Fixing and conditioning are fundamentally different concepts. See Haavelmo (1943) and Heckman and Pinto (2015) for discussions. The "do" operator in Pearl (2009) is an example of fixing.

[17] Using the Quandt (1972) switching regression notation, we can write outputs and inputs generated by treatment as $\boldsymbol{Y}_{k,a} = (1 - D)\boldsymbol{Y}_{k,a}^0 + (D)\boldsymbol{Y}_{k,a}^1$ and $\boldsymbol{X}_{k,a} = (1 - D)\boldsymbol{X}_{k,a}^0 + (D)\boldsymbol{X}_{k,a}^1$.

where $\phi_a(\boldsymbol{x}, \boldsymbol{b})$ is the common structural function (across d and k) generating the deterministic portion of the effect of $\boldsymbol{B}_k = \boldsymbol{b}$, $\boldsymbol{X}^d_{k,a} = \boldsymbol{x}$ on outcomes and

$$\begin{aligned} F^0_{k,a}(\cdot|\mathbf{Fix}\,\boldsymbol{X}^d_{k,a} = \boldsymbol{x}, \boldsymbol{B}_k = \boldsymbol{b}) &= F^1_{k,a}(\cdot|\mathbf{Fix}\,\boldsymbol{X}^d_{k,a} = \boldsymbol{x}, \boldsymbol{B}_k = \boldsymbol{b}) \\ &= F_a(\cdot|\mathbf{Fix}\,\boldsymbol{X}^d_{k,a} = \boldsymbol{x}, \boldsymbol{B}_k = \boldsymbol{b}), \end{aligned} \tag{3b}$$

where $F^j_{k,a}(\boldsymbol{z}|\mathbf{Fix}\,\boldsymbol{\Omega} = \boldsymbol{\omega})$ is the distribution of $\boldsymbol{Z}$ for $\boldsymbol{\Omega}$ fixed at $\boldsymbol{\omega}$ and $F_a(\boldsymbol{z}|\mathbf{Fix}\,\boldsymbol{\Omega} = \boldsymbol{\omega})$ is the age-a distribution of the errors associated with the production functions, assumed to be common across treatment regimes and samples, given $\boldsymbol{X}^d_{k,a} = \boldsymbol{x}$ and $\boldsymbol{B}_k = \boldsymbol{b}$.

We clarify assumption A-1 with two remarks.

Remark R-1 (Two distinct aspects of structural invariance). Assumption A-1 has two distinct aspects that can be resolved further into two separate assumptions: (i) structural relationships and distributions evaluated at the same arguments have identical values for treatment and control groups in the experimental sample, and (ii) analogous conditions hold across the experimental and nonexperimental samples. Condition ii enables analysts to simulate treatment and control outcomes in nonexperimental samples.

Remark R-2 (Accounting for cohort effects). A second aspect of assumption A-1, that the structural relationships and distributions are identical in the experimental and nonexperimental samples for $a \geq a^*$, embeds an implicit assumption about the absence of cohort effects in the technology and distribution of errors in the postsample period for the experimental sample. In particular, a structural function $\phi^d_{\mathrm{n},a}(\boldsymbol{x}, \boldsymbol{b})$ or distribution $F^d_{\mathrm{n},a}(\boldsymbol{z}|\mathbf{Fix}\,\boldsymbol{\Omega} = \boldsymbol{\omega})$ applied out of sample to subjects who are older than those in the experimental sample is a valid tool for forecasting the outcomes in the experimental sample at ages currently out of the age range of the experiment. Note that this does not mean that there are no cohort effects in the outcome of interest, $\boldsymbol{Y}^d_{k,a}$. Instead, it means that there are no cohort effects in the mapping between $\boldsymbol{X}_{k,a}$, $\boldsymbol{B}_k$ and $\boldsymbol{Y}^d_{k,a}$, $k \in \{\mathrm{e}, \mathrm{n}\}$.[18]

Testing assumption A-1 using the experimental and nonexperimental samples without imposing parametric assumptions requires common support conditions over the age range $a \leq a^*$. This requirement is captured by assumption A-2.

Assumption A-2 (In-sample support conditions). For $a \leq a^*$ and $d \in \{0, 1\}$,

$$\mathrm{supp}(\boldsymbol{Y}^d_{\mathrm{e},a}, \boldsymbol{X}^d_{\mathrm{e},a}, \boldsymbol{B}_\mathrm{e}, \varepsilon_{\mathrm{e},a}) \subseteq \mathrm{supp}(\boldsymbol{Y}_{\mathrm{n},a}, \boldsymbol{X}_{\mathrm{n},a}, \boldsymbol{B}_\mathrm{n}, \varepsilon_{\mathrm{n},a}). \tag{4}$$

To forecast out-of-sample input paths in order to forecast out-of-sample treatment effects, we require assumption A-3.

[18] Note that the function $\phi^d_{k,a}$ could include polynomial age trends as arguments in order to capture, e.g., work experience (age − schooling − 6).

Assumption A-3 (Out-of-sample forecast support). (i) The support of the nonexperimental data contains the support of the future values that experimentals could experience. (ii) We can accurately forecast the distributions of future values of inputs and errors. Condition i is a version of assumption A-2 for $a > a^*$.

Remark R-3 (Requirements for accurate forecasts). This assumes that the analyst can make accurate adjustments for cohort effects. We require only distributions of forecast variables to make accurate forecasts.

To be specific, we now consider how one can use this framework to forecast life-cycle labor income.

B. Forecasting Labor Income

1. Step 1: Constructing a Synthetic Cohort

We use the Children of the National Longitudinal Survey of Youth (CNLSY) to construct a synthetic cohort from ages 21 to 29, using similarity with the baseline variables in the experimental samples ($\boldsymbol{B}$). We use both the National Longitudinal Survey of Youth 1979 (NLSY79) and the Panel Study of Income Dynamics (PSID) to construct a synthetic cohort from ages 29 to 67. Whenever we use the NLSY79 and the PSID together, we combine samples. Thus, we use three nonexperimental data sets to obtain information across the life cycle. We satisfy support conditions (eq. [4]).

Because we do not observe each element of the eligibility index discussed in section II, we approximate $\boldsymbol{B}_n \in \mathcal{B}_0$. We delimit the sample to include observations satisfying the following criteria: (i) for NLSY79: black, labor income less than \$300,000 (2014 USD) in any given year, birth year between 1957 and 1965; (ii) for PSID: black, labor income less than \$300,000 (2014 USD), birth year between 1945 and 1981; and (iii) for CNLSY: black, labor income less than \$300,000 (2014 USD) in any given year, birth year between 1978 and 1983.

We weight individuals in the nonexperimental samples according to their resemblance to individuals in the experimental sample.[19] We match on baseline pretreatment match variables: year of birth, gender, and number of siblings at baseline. All are available in the nonexperimental data sets. This procedure generates a synthetic cohort in the nonexperimental sample for subsequent analysis in our structural forecasting procedure.

By design, there is no treatment effect in the nonexperimental sample. Matching to the experimental sample is executed using baseline variables not affected by treatment in the experimental sample. Figure 1 demonstrates that the synthetic cohort is comparable to the control group of

[19] We use Mahalanobis's matching algorithm 1 and weights derived from the Mahalanobis distances that downweight dissimilar observations. See app. C for details.

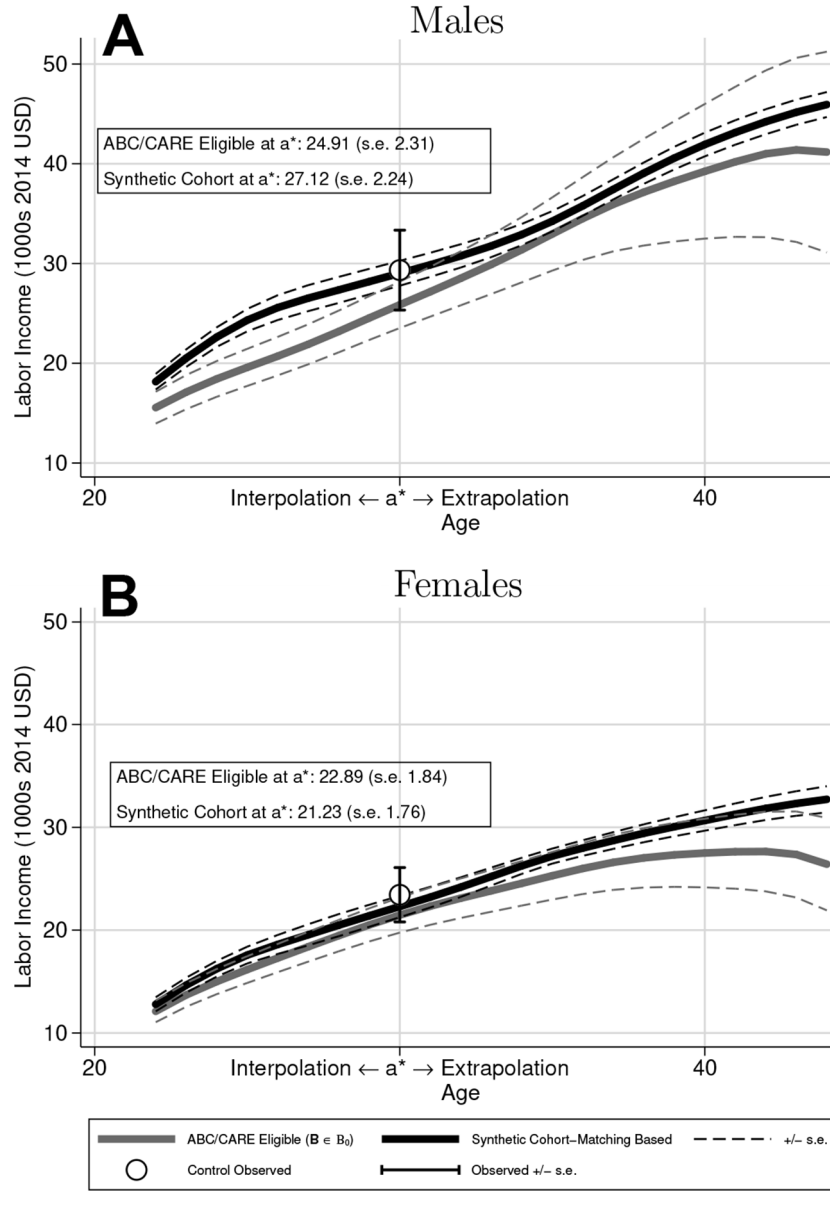

FIG. 1.—Labor income profile, disadvantaged individuals and synthetic cohort constructed by matching in the auxiliary samples. Forecast labor income for males (*A*) and females (*B*) in the auxiliary samples for whom $\boldsymbol{B} \in \mathcal{B}_0$, i.e., ABC/CARE eligible; for the synthetic cohort we construct on the basis of the method proposed in section III. We combine data from the Panel Study of Income Dynamics (PSID), the National Longitudinal Survey of Youth 1979 (NLSY79), and the Children of the National Longitudinal Survey of Youth 1979 (CNLSY79). We highlight the observed labor income at oldest a^* (age 30) for the ABC/CARE control-group participants. We stop at age 45 for want of data to compute the childhood risk index defining $\boldsymbol{B} \in \mathcal{B}_0$ in the auxiliary samples. Standard errors (s.e.) are based on the empirical bootstrap distribution. a^* is the oldest age in the experimental sample.

the experiment. At age 30, average observed labor income for the control group in the experimental sample coincides with average labor income for the synthetic cohort.[20]

2. Step 2: Establishing Exogeneity of Inputs

Forecasting does not require that we take a position on the exogeneity of $\boldsymbol{X}_{k,a}^{d}$ for $k \in \{\mathrm{e}, \mathrm{n}\}$ and $a \in \{1, \ldots \bar{A}\}$ with respect to labor income. Estimated structural models with biased parameters can still give reliable forecasts if relationships between observed and unobserved variables are the same within sample and in the forecast sample.[21] However, exogeneity facilitates the use of economic theory to interpret treatment effects, to forecast outcomes in samples where inputs are manipulated differently than in the experimental sample, and to test the validity of the construction of our synthetic cohort. Exogeneity also makes identification of $\boldsymbol{\phi}_{k,a}^{d}(\cdot, \cdot)$ in the nonexperimental sample straightforward. Assumption A-4 formalizes a strong form of the exogeneity condition.

Assumption A-4 (Exogeneity). Let $\{1, \ldots, \bar{A}\}$ index the periods of a life cycle. For all a, $a' \in \{1, \ldots, \bar{A}\}$ and for d, $d' \in \{0, 1\}$,

$$\varepsilon_{k,a}^{d} \perp\!\!\!\perp \boldsymbol{X}_{k,a'}^{d'} | \boldsymbol{B}_{k} = \boldsymbol{b} \tag{5}$$

for all $\boldsymbol{b}$ in the support of $\boldsymbol{B}_{k}$, $k \in \{\mathrm{e}, \mathrm{n}\}$, where $\boldsymbol{M} \perp\!\!\!\perp \boldsymbol{N} | \boldsymbol{Q}$ denotes independence of $\boldsymbol{M}$ and $\boldsymbol{N}$, given $\boldsymbol{Q}$. Below we discuss how this condition can be weakened and unbiased forecasts can still be obtained.

To appreciate the benefit of assumption A-4, consider the following example. Suppose that years of education is a component of $\boldsymbol{X}_{k,a'}^{d'}$. The joint distribution of $\varepsilon_{k,a}^{d}$ and $\boldsymbol{X}_{k,a'}^{d'}$ can differ substantially across experimental and nonexperimental samples. In the experimental sample, years of education is boosted by treatment, which is randomly assigned. In the nonexperimental samples, however, there is no experimental variation, and exogeneity does not hold. Individuals with high observed levels of education could have high values of $\varepsilon_{k,a}^{d}$ due to omitted ability. This creates a fundamentally different dependence between $\varepsilon_{k,a}^{d}$ and $\boldsymbol{X}_{k,a'}^{d'}$ in the nonexperimental sample. Assumption A-4 avoids this problem in making forecasts. Below, we establish that our forecasts based on this assumption are concordant with forecasts from approaches that do not require assumption A-4.

[20] We observe labor income in the experimental samples at ages 21 and 30. We use the data at age 21 to initialize our forecasting model and hence cannot use it for testing our forecast.

[21] See Liu, Moon, and Schorfheide (2016). Conditions C-1 to C-3 in app. C.3 spell out the requirements.

3. Step 3: Determining Inputs

Analysis of the nonexperimental data reveals that the inputs determining labor income under assumption A-4 are average PIAT (Peabody Individual Achievement Test) achievement score at ages 5–7, completed education, labor income at age 21, and lagged labor income. Table C.3 (tables A.1–A.33, B.1, C.1–C.13, D.1, E.1–E.8, G.1–G.8, and H.1 are available online) shows that these variables predict labor income. Appendix C.3.6 shows that there is common support across data sets. We report tests for endogeneity of these variables in the experimental and auxiliary samples used in this paper in appendix C.3.7.[22] After conditioning on $\boldsymbol{X}^d_{k,a}$ and $\boldsymbol{B}_k$, we do not reject the null hypothesis of exogeneity. Accordingly, we use ordinary least squares in making our baseline estimates.

Table 1 displays the treatment effects of the program for these inputs.[23] Assignment to treatment has statistically and economically significant causal effects on the inputs generating final outcome treatment effects. Our forecasted treatment effects are based on these program-induced changes in inputs. For females, treatment increases the average PIAT score by almost one-third of a standard deviation.[24] For males, the effect is almost half of a standard deviation. The program substantially boosts high school graduation for females and college graduation for both males and females. We use years of education attained to summarize both effects in a measure that is comparable across genders. Labor income at age 21 for girls is not greatly boosted by the program. This arises in part because treated girls are more likely to be enrolled in college at age 21 and thus do not work at that age. The program boosts annual labor income, especially for males, for whom the average treatment effect at age 30 is almost $20,000 (2014 USD).[25]

[22] These tests are based on the assumption that $\varepsilon^d_{k,a}$ for $k \in \{\mathrm{e}, \mathrm{n}\}$ is characterized by a factor structure. The factors are predicted by measurements of cognitive and noncognitive skills. We use estimated factors as control functions. We do not reject the null of exogeneity. See app. C.3.7. Factor structure models are widely used in structural estimation of production functions of skills during early childhood. See, e.g., Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010).

[23] We first present raw treatment-control mean differences. As we report in table 1, the treatment effects are substantial across multiple outcomes. In some cases, this finding is at odds with what other studies report (Ramey et al. 1985; Clarke and Campbell 1998; Campbell et al. 2001, 2002, 2008, 2014). The difference is explained, mainly, by the fact that we consider effects by gender. Only Campbell et al. (2014) consider treatment effects by gender. They focus on health effects and find that men have many more positive effects, especially in cardiovascular and metabolic conditions, compared to women. This is consistent with the results we report below.

[24] The test is standardized to an in-sample standard deviation of 15 units.

[25] Table 1 displays age-21 and age-30 labor income because labor income is observed at ages 21 and 30 in the experimental sample. Labor income at age 30 is an input in our methodology only after age 30.

TABLE 1
SUMMARY OF TREATMENT EFFECTS FOR INPUTS GENERATING LABOR INCOME ($X^d_{k,a}$)

	FEMALES		MALES	
INPUTS	Control Mean	Average Treatment Effect	Control Mean	Average Treatment Effect
PIAT scores	95.63	**4.92**	93.46	**7.70**
High school graduation	.51	**.25**	.61	.07
College graduation	.08	.13	.12	**.17**
Years of education	11.76	**2.14**	12.90	**.66**
Labor income at 30	23,443.42	2,547.50	29,340.31	**19,809.74**

NOTE.—This table shows the control-group level and the raw mean difference between treatment and control (average treatment effects), by gender. PIAT scores have a sample mean of 100 and a standard deviation of 15. High school and college graduation are expressed in rates. Labor income is in 2014 USD. Average treatment effects in boldface are statistically significant at the 10% level.

4. Step 4: Testing the Empirical Implications of Assumption A-1

Under assumption A-4, we build on Heckman, Pinto, and Savelyev (2013) to test for invariance. Condition (3a) of assumption A-1, combined with assumptions A-2 and A-4 and the normalization $\mathbb{E}(\varepsilon^d_{k,a}) = 0$ for all $a \in \{1, ..., \bar{A}\}$, generates the following testable implications:

$$\mathbb{E}\left[\boldsymbol{Y}^1_{\mathrm{e},a} | \boldsymbol{X}^1_{\mathrm{e},a} = \boldsymbol{x}, \boldsymbol{B}_{\mathrm{e}} = \boldsymbol{b}, D = 1\right] = \mathbb{E}[\boldsymbol{Y}^0_{\mathrm{e},a} | \boldsymbol{X}^0_{\mathrm{e},a} = \boldsymbol{x}, \boldsymbol{B}_{\mathrm{e}} = \boldsymbol{b}, D = 0] \quad (6a)$$

and

$$\mathbb{E}[\boldsymbol{Y}^d_{\mathrm{e},a} | \boldsymbol{X}^d_{\mathrm{e},a} = \boldsymbol{x}, \boldsymbol{B}_{\mathrm{e}} = \boldsymbol{b}, D = d] = \mathbb{E}[\boldsymbol{Y}_{\mathrm{n},a} | \boldsymbol{X}_{\mathrm{n},a} = \boldsymbol{x}, \boldsymbol{B}_{\mathrm{n}} = \boldsymbol{b}], \quad (6b)$$

for $d \in \{0, 1\}$ where $\boldsymbol{Y}_{\mathrm{n},a}$ is the counterpart of $\boldsymbol{Y}_{\mathrm{e},a}$ in the nonexperimental sample.

Note, however, that if the only goal is to construct unbiased forecasts of mean treatment effects, the minimal requirement is that experimental treatment effects should equal differences in the conditional means of forecasts formed on the nonexperimental samples evaluated at $\boldsymbol{X}_{\mathrm{n},a} = \boldsymbol{x}^1$ and $\boldsymbol{X}_{\mathrm{n},a} = \boldsymbol{x}^0$ respectively:

$$\begin{aligned} &\mathbb{E}\left[\boldsymbol{Y}^1_{\mathrm{e},a} | \boldsymbol{X}^1_{\mathrm{e},a} = \boldsymbol{x}^1, \boldsymbol{B}_{\mathrm{e}} = \boldsymbol{b}, D = 1\right] - \mathbb{E}[\boldsymbol{Y}^0_{\mathrm{e},a} | \boldsymbol{X}^0_{\mathrm{e},a} = \boldsymbol{x}^0, \boldsymbol{B}_{\mathrm{e}} = \boldsymbol{b}, D = 0] \\ &\quad = \mathbb{E}\left[\boldsymbol{Y}_{\mathrm{n},a} | \boldsymbol{X}_{\mathrm{n},a} = \boldsymbol{x}^1, \boldsymbol{B}_{\mathrm{n}} = \boldsymbol{b}\right] - \mathbb{E}[\boldsymbol{Y}_{\mathrm{n},a} | \boldsymbol{X}_{\mathrm{n},a} = \boldsymbol{x}^0, \boldsymbol{B}_{\mathrm{n}} = \boldsymbol{b}]. \end{aligned} \quad (6c)$$

Rather than imposing condition (6c), we test sufficient conditions for it to hold: that is, we test conditions (6a) and (6b). We test condition (6a) across treatment regimes and condition (6b) for $d = 0$ at age 30, where we observe labor income in the experimental sample for both

the treatment and the control groups. Assuming linearity, if condition (6a) holds, then the coefficient associated with D, denoted by τ, should be zero in

$$\boldsymbol{Y}^{d}_{\mathrm{e},30} = \tau \cdot D + \boldsymbol{B}_{\mathrm{e}} \cdot \gamma^{d}_{\mathrm{e},30} + \boldsymbol{X}^{d}_{\mathrm{e},30} \cdot \beta^{d}_{\mathrm{e},30} + \varepsilon^{d}_{\mathrm{e},a}. \tag{7a}$$

Failing to reject the null hypothesis H_0: $\tau = 0$ is equivalent to failing to reject invariance across treatment regimes.

Panel A of table 2 displays estimates of the coefficients of equation (7a) for labor income at age 30 by gender. We do not reject the null hypothesis that the technology is invariant across treatment regimes for either gender.[26] Panel B of table 2 reports estimates for the remaining coefficients in equation (7a). Years of education is strongly boosted by ABC/CARE (see table 1).

Define $K = \boldsymbol{1}(k = \mathrm{e})$ as an indicator of whether an observation comes from the experimental sample. The coefficient on K, denoted by π, should be zero in the following linear technology (i.e., H_0: $\pi = 0$ if condition [6b] is true):

$$\boldsymbol{Y}_{k,30} = \pi \cdot K + \boldsymbol{B}_{k} \cdot \gamma_{k,30} + \boldsymbol{X}_{k,30} \cdot \beta_{k,30} + \varepsilon_{k,a}. \tag{7b}$$

Panel C of table 2 displays estimates of the parameters of equation (7b) for labor income at age 30 for males and females. Estimates of π are small and not statistically significant.[27] We do not reject the null hypothesis that the technologies are invariant across samples, so the data are consistent with invariance.

We use analogous procedures to test condition (3b) of assumption A-1. First, we test invariance in the distributions of $\varepsilon^{d}_{k,a}$ across treatment regimes within the experimental sample for labor income at age 30. Then, invoking invariance across treatment regimes, we test invariance across the experimental and nonexperimental residual distributions. Residuals are generated from the estimated forecasting model in equation (2), assuming a linear technology. We adjust the residuals for model estimation error as explained in step 8 of appendix C.7.1.

With the empirical counterparts of $\varepsilon^{d}_{k,a}$ in hand, we implement two tests to compare the distributions in table 3: t-tests of mean comparisons

[26] Note that after background variables and the intermediate inputs are accounted for, average labor income is \$2,213 (2014 USD) higher in the control group for females. This value is relatively small in the context of annual labor income at age 30 and given that the average in the control group is \$23,443 (2014 USD). The same holds for the males, where the treatment-control difference is \$232 net of inputs and the average for the control group is \$29,340 (2014 USD).

[27] The averages of labor income in the experimental and nonexperimental samples are, respectively, \$24,584 and \$40,007 for females and \$24,098 and \$32,717 (2014 USD) for males.

TABLE 2
TESTING INVARIANCE IN TECHNOLOGIES ($\phi^d_{k,a}(\boldsymbol{x}, \boldsymbol{b})$) OF LABOR INCOME AT AGE 30

	FEMALES		MALES	
	Coefficient	*p*-Value	Coefficient	*p*-Value
	A. Invariance across Treatment Regimes			
D	−2,212.806	.586	231.606	.969
	B. Precision of Estimated Coefficients of Eq. (7a)			
$\boldsymbol{B}_c$:				
Mother's education (at birth)	−957.0972	.387	1,850.201	.358
$\boldsymbol{X}^d_{c,30}$:				
PIAT score (5–7)	5.726	.975	327.186	.338
Years of education (30)	2,356.143	.006	4,474.721	.018
Labor income (21)	.218	.320	.322	.175
R^2	.281		.207	
Observations	52		50	
	C. Invariance across Experimental and Nonexperimental Samples			
K	−142.631	.965	1,887.575	.654
	D. Precision of Estimated Coefficients of Counterpart to Eq. (7b) in the Nonexperimental Sample			
$\boldsymbol{B}_k$:				
Mother's education (at birth)	−229.481	.631	427.224	.459
$\boldsymbol{X}^d_{k,30}$:				
PIAT score (5–7)	266.1971	.002	219.220	.044
Years of education (30)	4,263.156	.000	4,434.173	.000
Labor income (21)	.355	.000	.685	.000
R^2	.221		.182	
Observations	829		746	

NOTE.—Panels A and B show estimates of the coefficients in eq. (7a) for labor income at age 30 by gender within the experimental sample. D denotes the treatment indicator ($D = 0$ for control-group participants and $D = 1$ for treatment-group participants). $\boldsymbol{B}_c$ consists of baseline variables not affected by treatment (mother's education at birth), and $\boldsymbol{X}^d_{k,30}$ is age-30 intermediate inputs. We drop labor income observations above the 95th percentile to avoid precision issues. Panels C and D show estimates of the coefficients in eq. (7b) for labor income at age 30 by gender, pooling the experimental treatment and control groups and the synthetic cohort at age 30. K denotes membership to the experimental or nonexperimental sample ($K = 0$ indicates the synthetic cohort in the nonexperimental sample and $K = 1$ the experimental sample). $\boldsymbol{B}_k$ consists of baseline variables not affected by treatment (mother's education at birth), and $\boldsymbol{X}^d_{k,30}$ is age-30 intermediate inputs.

and Kolmogorov-Smirnov tests of equality of distributions. We do not reject equality of treatment and control distributions of $\varepsilon^d_{k,a}$ for both females and males (panel A). As before, we pool the experimental treatment and control groups and the synthetic cohort to test invariance across samples by gender. Except for one hypothesis test of condition (6b) for males, we do not reject the null hypothesis of invariance across samples; see panel B.

TABLE 3
TESTING INVARIANCE IN DISTRIBUTIONS OF THE ERROR TERMS ($F^d_{k,a}$) OF LABOR INCOME AT AGE 30

	FEMALES			MALES		
	t-Statistic	*p*-Value	K-S *p*-Value	*t*-Statistic	*p*-Value	K-S *p*-Value
	A. Invariance across Treatment Regimes					
Equality in means	1.075	.287		−.0390	.969	
Equality in distributions			.272			.632
	B. Invariance across Experimental and Nonexperimental Samples					
Equality in means	.054	.957		−.226	.822	
Equality in distributions			.481			.046

NOTE.—Panel A: tests for equality in distributions of residuals within the experimental sample across treatment regimes at age 30 in labor income by gender. Panel B: tests for equality in distributions of residuals across the experimental and nonexperimental samples pooling the experimental treatment and control groups and the synthetic cohort at age 30 for labor income by gender. Residuals are the relevant outcome net of mother's education at birth, average PIAT score at ages 5–7, years of education at age 30, and labor income at age 21. Residuals are adjusted for estimation error as explained in step 6 of app. C.7.1. Tests are a *t*-test of equality in means and the Kolmogorov-Smirnov (K-S) test.

5. Step 5: Accounting for Estimation Error, Forecast Error, and Plausible Ranges of Externally Supplied Parameters

We obtain standard errors from the empirical bootstrap distribution. Our inference accounts for each step of our estimation procedure, as well as forecast error. We conduct sensitivity analyses for externally supplied parameters. A step-by-step recipe for accounting for parameter uncertainty is presented in appendix C.7. The forecasted present value of the gain induced by treatment using the estimates displayed in figure 2 is $133,032 (SE: $76,634; 2014 USD). We explore the estimates from alternative forecasting models in appendix C.6.[28] When males and females are pooled and when the samples are separated by gender, the present value gains remain within a range that does not change our inference that the program had substantial lifetime benefits.

6. Step 6: Validating Forecasts

Invariance across treatment regimes and samples is the essential ingredient for constructing valid forecasts. Figure 2 displays our forecasted labor

[28] As both a referee and various discussants of our paper have pointed out, our identification and estimation strategies do not impose cross-equation restrictions. We use different data sets to identify and estimate the $\phi_a(\boldsymbol{x}, \boldsymbol{b})$ for each outcome (e.g., labor income, health, crime). Thus, the predictor variables that we are able to use differ across outcomes, and we cannot conduct joint estimations.

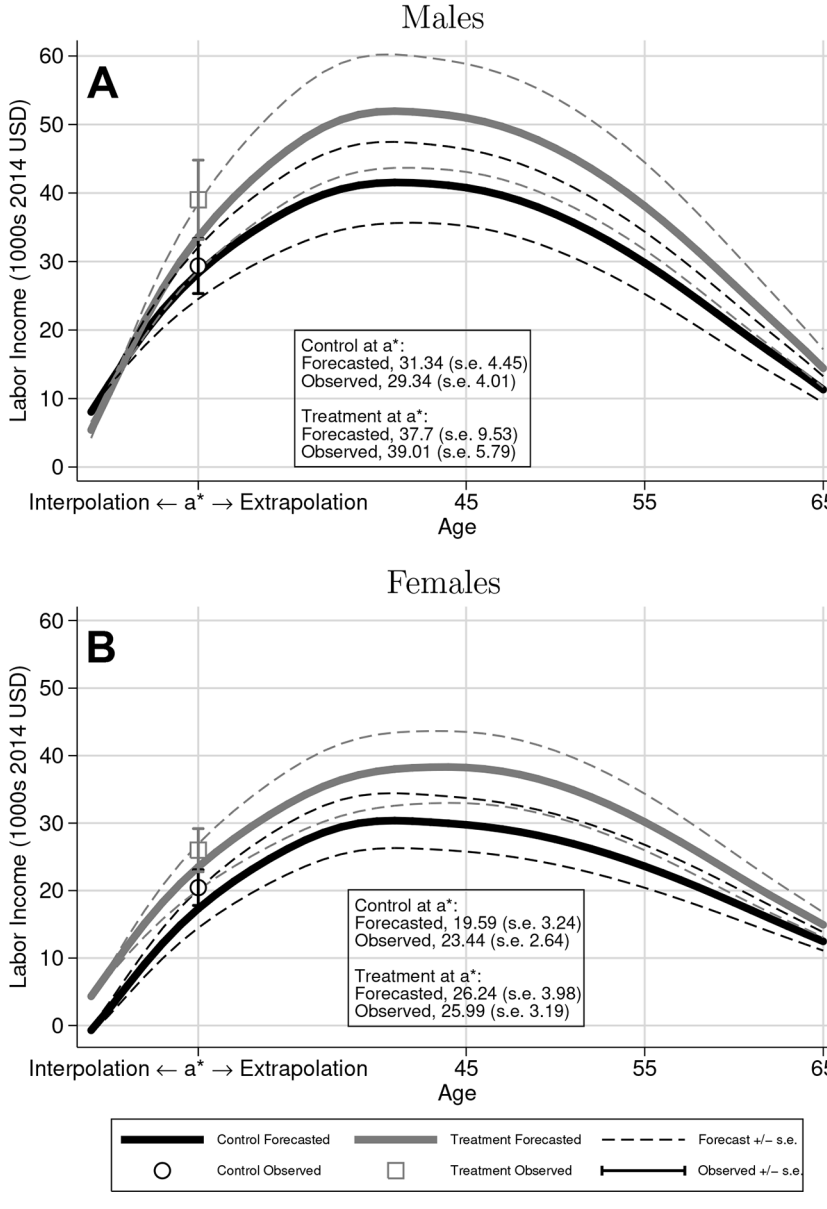

FIG. 2.—Forecast life-cycle labor income profiles for ABC/CARE males (*A*) and females (*B*) by treatment status, based on the method proposed in sec. III. We combine data from the Panel Study of Income Dynamics (PSID), the National Longitudinal Survey of Youth 1979 (NLSY79), and the Children of the National Longitudinal Survey of Youth 1979 (CNLSY79). We highlight the observed labor income at a^* (age 30) for the ABC/CARE control- and treatment-group participants. Our forecasts go up to age 67, the age of assumed retirement. Standard errors (s.e.) are the standard deviations of the empirical bootstrap distribution. See app. C for a discussion of our choice of predictors and a sensitivity analysis on those predictors. We underpredict labor income for both males and females. These differences, however, are not statistically significant (and labor income is a relatively minor component of the overall analysis for females).

income profiles. Forecasted and actual labor income are closely aligned in both the treatment and the control regimes.[29] Computing the net present value, the internal rate of return, and benefit/cost ratios is straightforward once age-by-age forecasts are available.[30]

7. Alternative Forecasting Models for Labor Income

An alternative nonparametric forecasting method compresses the whole forecasting procedure. Under assumptions A-1 and A-4, we can use matching on baseline variables and variables affected by treatment to construct counterparts to the experimental treatment and control groups in the nonexperimental sample.[31] Matching is a nonparametric estimation procedure for conditional mean functions. Matching creates direct counterparts in the auxiliary sample for each member of the experimental sample. Instead of estimating a model for the life-cycle profile of labor income and forecasting from it, we directly use the counterpart matched profiles.[32] This is an intuitively appealing nonparametric estimator of life-cycle program treatment effects that is valid under exogeneity (Heckman and Navarro 2004).

This matching procedure is fundamentally different from what is used to construct the synthetic cohort in step 1. In the main analysis of this paper, we match on baseline variables not affected by treatment to construct a synthetic cohort with $\boldsymbol{B} \in \mathcal{B}_0$. Using these samples, we estimate production functions on this cohort to forecast out-of-sample treatment effects. In contrast, in the analysis of this subsection, we match both on baseline variables ($\boldsymbol{B}$) and on variables affected by treatment ($\boldsymbol{X}^d_{k,a}$), compressing the construction of the synthetic cohort and the estimation of the production functions for out-of-sample forecasts to a single, nonparametric procedure.

Table C.12 shows that there is close agreement between nonparametric estimates based on matching and the more parametric model-based approach previously presented. This reassuring concordance is consistent with exogeneity of inputs and structural invariance.[33]

[29] The content in fig. 2 is sufficient but not necessary to calculate the gain of the program due to labor income. It would be sufficient to forecast the difference between the treatment and control groups. Forecasting the levels, however, provides us with additional testable implications. It also allows us to easily account for forecasting error and to verify that the life-cycle profiles that we estimate are comparable to observed profiles for similar socioeconomic groups. The pattern of life-cycle labor income we generate is typical for that of low-skilled workers (Gladden and Taber 2000; Sanders and Taber 2012; Blundell, Graber, and Mogstad 2015; Lagakos et al. 2018).

[30] Some practical details involved in doing this are in apps. C.4 and C.5.

[31] Heckman et al. (1998) discuss this procedure.

[32] See app. C.3.5 for details.

[33] Note that the nonparametric estimates are more tightly estimated because there are fewer steps in estimation. We are conservative in using the less precisely estimated forecasts in our main analysis.

IV. Forecasting Other Life-Cycle Benefits

In this section, we adapt the methodology described in section III to forecast the net benefits of the program arising from enhanced parental income, health, and reduced crime. In the text, we focus on forecasting health benefits and briefly discuss forecasts of crime and parental labor income.[34] Procedures for forecasting the benefits and costs of education are reported in appendix D.

A. Health

One contribution of this paper is forecasting and monetizing the life-cycle benefits of the enhanced health and reduced health costs of participants, using a version of equation (2). The model recognizes that (i) health outcomes such as diabetes, heart disease, and death are absorbing states and (ii) there is no obvious terminal time period for benefits and costs except death, which we forecast.

We adapt the Future Adult Model (FAM), a forecasting model for health conditions and costs developed by Goldman et al. (2015).[35] We forecast health outcomes of program participants from their mid-30s up to their projected age of death.[36] Our version of FAM passes a variety of specification tests and accurately forecasts health outcomes and health behaviors.[37]

Our methodology has four steps; extensive details are provided in appendix G: (i) follow an adapted version of the steps in section III to predict the health state occupancy probabilities for the ABC/CARE subjects; (ii) estimate quality-adjusted-life-year (QALY) models using the Medical Expenditure Panel Survey (MEPS) and the PSID;[38] (iii) estimate medical cost models using MEPS and the Medicare Current Beneficiary Survey (MCBS), allowing estimates to differ by health state and observed characteristics; and (iv) forecast the medical expenditures and QALYs that correspond to the simulated individual health trajectories.[39]

Our application of FAM uses the information on age-30 observed characteristics and a mid-30s health survey, allowing us to account for components that are important for forecasting health outcomes. The models

[34] Apps. E and C.3.8 provide further documentation.

[35] App. G discusses the FAM methodology in detail. It is not a competing-risks model but forecasts vectors of incidence and costs of disease one category at a time, using univariate models.

[36] The simulation starts at the age at which we observe the subjects' age-30 follow-up.

[37] Goldman et al. (2015) present tests of the model assumptions and predictive performance for population aggregate health and health behavior outcomes.

[38] QALY is a measure that reweighs a year of life according to its quality, given the burden of disease (Dolan 1997; Shaw, Johnson, and Coons 2005).

[39] As part of step 1, we impute some of the variables used to initialize the FAM models (see app. G.1.6.1).

forecast the probability of having any of the major disease categories and health states at age $a + 1$ based on the state of health as summarized by major disease categories at age a.[40]

Using the occupancy probabilities for each health outcome at each age, we take a Monte Carlo draw for each subject. Each simulation depends on each individual's health history and characteristics. For every simulated trajectory of health outcomes, we forecast the life-cycle medical expenditure, using the models estimated from MEPS and MCBS. We estimate the expected life-cycle medical expenditure by taking the mean of each individual's simulated life-cycle medical expenditure.

The models estimated using MCBS represent medical costs in the years 2007–10. The MEPS estimation captures costs during 2008–10. To account for real medical cost growth after 2010, we adjust each model's forecast, using the method described in appendix G.2.3. The same procedure is applied to calculate QALYs. We compute QALYs on the basis of a widely used health-related quality-of-life measure (EQ-5D) available in MEPS. We then apply this model to the PSID data. QALYs monetize the health of an individual at each age. Although there is not a clear age-by-age treatment effect on QALYs, there is a statistically and substantively significant difference in the accumulated present value of the QALYs between the treatment and control groups.[41]

We estimate three models of medical spending: (i) Medicare spending (annual medical spending paid by parts A, B, and D of Medicare), (ii) private spending (medical spending paid by a private insurer or paid out of pocket by the individual), and (iii) all public spending other than Medicare. Each medical spending model includes the variables we use to forecast labor and transfer income, together with current health, risk factors, and functional status as explanatory variables.

We also calculate medical expenditures before age 30 (see app. G.2.4). The ABC/CARE interviews at ages 12, 15, 21, and 30 have information related to hospitalizations at different ages and number of births before age 30. We combine this information, along with individual and family demographic variables, to use MEPS to forecast medical spending for each age.

[40] See tables G.1–G.3 for a summary. Our forecasts are based on 2-year lags, because of data limitations in the auxiliary sources we use to simulate the FAM. For example, if the individual is 30 (31) years old in the age-30 interview, we simulate the trajectory of her health status at ages 30 (31), 32 (33), 34 (35), and so on until her projected death. Absorbing states are an exception. For example, heart disease at age a does not enter into the estimation for heart disease at age $a + 1$, because it is an absorbing state: once a person has heart disease, she carries it through the rest of her life.

[41] Our baseline estimation assumes that each year of life is worth $150,000 (2014 USD). Our estimates are robust to substantial variation in this assumption, as we show in app. H.

B. *Crime*

To estimate the life-cycle benefits and costs of ABC/CARE on crime, we use rich data obtained from public records. Two previous studies consider the impacts of ABC on crime: Clarke and Campbell (1998) use administrative crime records up to age 21 and find no statistically significant treatment effects; Barnett and Masse (2002, 2007) analyze self-reported crime at age 21. They lacked access to the longer-term, administrative data that we use and report weak treatment effects on crime. Our study improves on this research in two ways: (i) we use administrative data on the accumulated number of crimes that participants commit through their mid-30s; and (ii) we use microdata specific to the states in which participants grew up, as well as other national data sets, to forecast criminal activity from the mid-30s to 50. We forecast using methods standard in the criminology literature (Cohen and Bowles 2010; McCollister, French, and Fang 2010). See appendix E for a complete discussion of our crime forecasts.

C. *Parental Labor Income*

ABC/CARE offers childcare to the parents of treated children for more than 9 hours a day for 5 years, 50 weeks a year. Only 27% of participant mothers of children reported living with a partner at baseline. This barely changed during the course of the experiment (see app. A). The childcare component generates substantial treatment effects on maternal labor force participation and parental labor income.[42] In addition, subsidized childcare induced wage growth due to enhanced parental educational attainment and through wage growth due to work experience.

We observe parental labor income at eight different ages for the participants through age 21.[43] To estimate the profile of parental earnings over the entire life cycle, we use two different approaches in appendix C.3.8: (i) an approach based on projections using standard Mincer equations and (ii) an approach based on the analysis of section III.

Any childcare inducements of the program likely benefit parents who, at baseline, did not have any other children who were not eligible for program participation. Additional childcare responsibilities would weaken the childcare effects of ABC/CARE, especially if younger siblings are

[42] There is also an effect on maternal school enrollment. Some of the mothers of participants decided to enroll in school and further their education. This could be one of the reasons why they make more money afterward. We quantify the social cost of additional education in app. D.

[43] The ages at which parental labor income is observed are 0, 1.5, 3.5, 4.5, 8, 12, 15, and 21. When the ABC/CARE subjects were age 21, their mothers were, on average, 41 years old.

We linearly interpolate parental labor income for ages for which we do not have observations between 0 and 21.

present. In appendix C.3.8, we show that the treatment effect for discounted parental labor income is much larger when participant children have no siblings at baseline. Treatment effects weaken when children who have siblings younger than 5 years old are compared to children who have siblings age 5 years or older.[44]

D. Program Costs

The yearly cost of the program was $18,514 per participant, in 2014 USD. We improve on previous cost estimates by using primary-source documents.[45] Appendix B discusses program costs in detail.

V. Estimates and Sensitivity Analysis

Figure 3 summarizes our findings. It displays the discounted (using a 3% discount rate) life-cycle benefits and costs of the program (2014 USD) pooled across genders, over all outcome categories, and for separate categories as well.[46] These benefits are the inputs of our baseline estimates for the annual internal rates of return and benefit/cost ratios. We conduct extensive sensitivity and robustness analyses to produce ranges of plausible values for the estimates of the internal rate of return (8.0%–18.3%) and benefit/cost ratio (1.52–17.40). We document that no single component of benefits drives our estimates.

The costs of the program are substantial, as frequently been noted by critics.[47] But so are the benefits, which far outweigh the costs. The individual gains in labor income, parental labor income, crime, and health are at least as large in magnitude as the costs. As a consequence, our measures of social efficiency remain statistically and economically significant even after we eliminate the benefits from any one of the four main components that we monetize.

[44] These patterns persist when the ABC/CARE sample is split by gender, but the estimates are not precise because the samples become too small. See app. C.3.8.

[45] Our calculations are based on progress reports written by the principal investigators and related documentation recovered in the archives of the research center where the program was implemented. We display these sources in app. B. The main component is staff costs. Other costs arise from nutrition and services that the subjects receive when they were sick, diapers during the first 15 months of their lives, and transportation to the center. The control-group children also receive diapers, during approximately 15 months, and iron-fortified formula. The costs are based on sources describing ABC treatment for 52 children. We use the same cost estimates for CARE, for which there is less information available. The costs exclude any expenses related to research or policy analysis. A separate calculation by the implementers of the program indicates an almost identical amount (see app. B).

[46] At discount rates of 0%, 3%, and 7%, the estimates for the benefit/cost ratios are 17.40 (SE:5.90), 7.33 (SE:1.84), and 2.91 (SE:0.59), respectively. We report estimates for discount rates between 0% and 15% in app. H.1.

[47] See, e.g., Fox Business News (2014) and Whitehurst (2014).

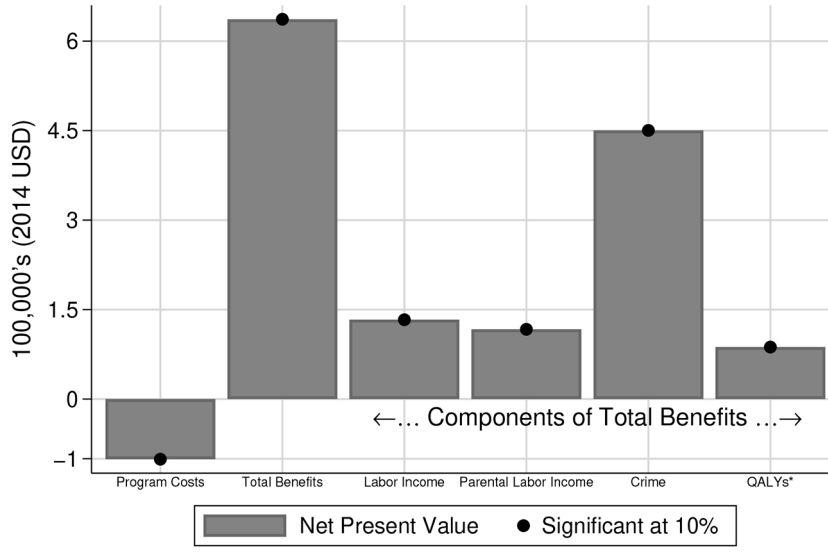

FIG. 3.—Net present value of main components of the pooled (over gender) cost/benefit analysis over the life cycle per program participant, treatment vs. control. This figure displays the life-cycle net present values per program participant of the main components of the cost/benefit analysis of ABC/CARE from birth to forecasted death, discounted to birth at a rate of 3%. By "net" we mean that each component represents the total value for the treatment group minus the total value for the control group. Program costs: the total cost of ABC/CARE, including the welfare cost of taxes to finance it. Total benefits: the benefits for all of the components we consider. Labor income: total individual labor income of program participants from age 21 to retirement (assumed to be at age 67). Parental labor income: total parental labor income of the parents of the participants when the participants were aged 1.5–21. Crime: the total cost of crime (judicial and victimization costs). To simplify the display, the following components are not shown in the figure: (1) cost of alternative preschool paid by the control-group children's parents; (2) social welfare costs of transfer income from the government; (3) disability benefits and social security claims; (4) costs of increased individual and maternal education (including special education and grade retention); (5) total medical public and private costs. Inference is based on nonparametric, one-sided *p*-values from the empirical bootstrap distribution. Dots indicate point estimates significant at the 10% level. Any gain corresponds to better health conditions until forecasted death, with $150,000 (2014 USD) as the base value for a year of life.

When males and females are pooled, the program is socially efficient: the internal rate of return and the benefit/cost ratio are 13.7% and 7.3, respectively. These estimates are statistically significant, even after accounting for sampling variation and forecast and estimation error in the experimental and auxiliary samples and the tax costs of financing the program.[48]

[48] We obtain the reported standard errors by bootstrapping all steps of our empirical procedure, including variable selection, imputation, model selection steps, and forecast error (see app. C.7).

TABLE 4
SENSITIVITY ANALYSIS FOR BENEFIT/COST RATIOS

	Pooled		Males		Females	
Baseline[a]	**7.33** (1.84)		**10.19** (2.93)		**2.61** (.73)	
	No IPW	No IPW, No Controls	No IPW	No IPW, No Controls	No IPW	No IPW, No Controls
Specification	**7.31** (1.81)	**7.99** (2.18)	**9.80** (2.69)	**8.83** (2.72)	**2.57** (.72)	**2.82** (.68)
	To Age 21	To Age 30	To Age 21	To Age 30	To Age 21	To Age 30
Forecast span	**1.52** (.36)	**3.19** (1.04)	**2.23** (.61)	**3.84** (1.60)	1.46 (.36)	**1.81** (.50)
	Vs. Stay at Home	Vs. Alt. Presch.	Vs. Stay at Home	Vs. Alt. Presch.	Vs. Stay at Home	Vs. Alt. Presch.
Counterfactuals	**5.44** (1.86)	**9.63** (3.10)	3.30 (2.95)	**11.46** (3.16)	**5.79** (1.37)	**2.28** (.76)
	0%	100%	0%	100%	0%	100%
Deadweight loss	**11.01** (2.79)	**5.50** (1.37)	**15.38** (4.35)	**7.59** (2.23)	**3.83** (1.04)	**2.01** (.59)
	0%	7%	0%	7%	0%	7%
Discount rate	**17.40** (5.90)	**2.91** (.59)	**25.45** (10.42)	**3.78** (.79)	**5.06** (2.82)	**1.49** (.32)

	Mincer Life Cycle	Life-Cycle Prediction	Mincer Life Cycle	Life-Cycle Prediction	Mincer Life Cycle	Life-Cycle Prediction
Parental income	**7.63**	**7.73**	**10.46**	**10.63**	**2.98**	**3.12**
	(1.84)	(1.92)	(2.94)	(2.95)	(.76)	(.85)
	.5% Annual Decay	.5% Annual Growth	.5% Annual Decay	.5% Annual Growth	.5% Annual Decay	.5% Annual Growth
Labor income	**7.01**	**7.66**	**9.58**	**10.79**	**2.51**	**2.71**
	(1.80)	(1.90)	(2.66)	(3.24)	(.70)	(.75)
	Drop Major Crimes	Halve Costs	Drop Major Crimes	Halve Costs	Drop Major Crimes	Halve Costs
Crime	**4.24**	**5.18**	**7.41**	**7.12**	**2.61**	**2.47**
	(1.10)	(1.22)	(3.43)	(2.41)	(.67)	(.66)
	Drop All	Double Value of Life	Drop All	Double Value of Life	Drop All	Double Value of Life
Health (QALYs)	**6.48**	**8.19**	**9.14**	**11.23**	**2.20**	**3.03**
	(1.79)	(2.13)	(2.73)	(3.40)	(.69)	(1.04)

NOTE.—This table displays results of sensitivity analyses of our baseline benefit/cost ratio calculation to the perturbations indexed in the rows. IPW (inverse probability weighting) adjusts for attrition and item nonresponse (see app. C.2 for details). Control variables are Apgar scores at 1 and 5 minutes and high-risk index (see app. C.8 for details on how we choose these controls). When forecasting up to ages 21 and 30, we consider all benefits and costs up to those ages. As counterfactuals we consider treatment vs. next best (baseline), treatment vs. stay at home, and treatment vs. alternative preschools (alt. presch.; see app. F for a discussion). Deadweight loss is the loss implied by any public expenditure (0% is no loss, and 100% is a one-dollar loss per dollar spent). Discount rate is the rate to discount benefits to child's age 0 (in all calculations). For parental labor income, see app. C.3.8 for details on the two alternative forecasts (Mincer and life cycle). For labor income, 0.5 annual growth (decay) is an annual wage growth (decay) due to cohort effects. Major crimes are rape and murder; half costs takes half of victimization and judiciary costs. For health (QALYs), "drop all" sets the value of life equal to zero. Standard errors obtained from the empirical bootstrap distribution are in parentheses. Values in boldface are significant at 10%, using one-sided tests. For details on the null hypothesis, see table H.1.

[a] Characteristics of the baseline calculation: IPW and controls, life span up to predicted death, treatment vs. next best, 50% marginal tax (50% deadweight loss), discount rate of 3%, parental income 0–21 (child's age), labor income predicted from age 21 to 65, all crimes (full costs), value of life at $150,000 (2014 USD).

We conduct an extensive set of sensitivity analysis. Table 4 displays the results of a sensitivity analysis of the estimates of the benefit/cost ratio to alternative plausible assumptions. Table 5 presents the corresponding internal rates of return. Our estimates are not driven by our methods for accounting for attrition and item nonresponse, by the conditioning variables, by the functional forms of projection equations used when computing the net present values, or by values of externally set parameters, such as the value of life introduced in our predictions of crime and health costs.[49] Although the internal rate of return remains relatively high for participant outcome measures only up to ages 21 or 30, the benefit/cost ratios indicate that accounting for benefits that go beyond age 30 is important. The return to each dollar is at most 3/1 when considering only benefits up to age 30 (see the columns in the forecast span rows).

Accounting for the treatment substitutes available to controls also matters. Males benefit the most from ABC/CARE relative to attending alternative formal childcare, while females benefit the most from ABC/CARE relative to staying at home. We explore these differences further in appendix F.

Our baseline estimates account for the deadweight loss caused by distortionary taxes collected to fund programs, plus the direct costs associated with collecting taxes.[50] We assume a marginal tax rate of 50%.[51] Our estimates are robust to dropping it to 0% or doubling it to 100% (deadweight loss columns). Our baseline estimate of benefit/cost ratios is based on a discount rate of 3%. Not discounting roughly doubles our benefit/cost ratios, while they remain statistically significant using a higher discount rate of 7% (discount rate columns).

Parental labor income effects induced by the childcare subsidy are an important component of the benefit/cost ratio.[52] We take a conservative approach in our baseline estimates and do not account for potential shifts in parental labor income profiles due to education and work experience subsidized by childcare (see the discussion in sec. IV.C). Our baseline estimates rely solely on parental labor income when participant children are aged 0–21. Alternative approaches considering the gain for the parents

[49] See app. C for a detailed discussion.

[50] When the transaction between the government and an individual is a direct transfer, we consider 0.5 dollars as the cost per each transacted dollar. We do not weight the final recipient of the transaction (e.g., transfer income). When the transaction is indirect, we classify it as government spending as a whole and consider its cost as 1.5 per dollar spent (e.g., public education).

[51] Feldstein (1999) estimates that the deadweight loss caused by increasing existing tax rates (marginal deadweight loss) may exceed 2 dollars per dollar of revenue generated. We use a more conservative value (0.5 dollars per dollar of revenue generated). In tables 4 and 5 and app. H.2, we explore the robustness of this choice of the welfare cost and find little sensitivity.

[52] There is no inconsistency between the weak female treatment effects on wages at age 30 and the high lifetime net-present-value treatment effect for earnings, given life-cycle wage growth attributable to enhanced inputs (education, PIAT scores, etc.).

through age 67 generate an additional increase in the gain due to parental labor income (see parental labor income columns).[53]

Individuals in ABC/CARE could experience positive cohort effects that might (i) make them more productive and therefore experience wage growth or (ii) experience a negative shock such as an economic crisis and therefore experience a wage decline. Our estimates are robust when we vary annual growth or decay rates in labor income between −0.5% and 0.5%. This is consistent with the range of values in Lagakos et al. (2018).

We also examine the sensitivity of our estimates to (i) dropping the most costly crimes, such as murder and rape,[54] and (ii) halving the costs of victimization and judiciary costs related to crime. The first sensitivity check is important because we do not want our estimates to be based on a few exceptional crimes. The second is important because estimates of victimization costs are controversial because they are subjective (see app. E.3). Our benefit/cost estimates are robust to these adjustments, even though crime is a major component of them. We also examine the sensitivity with respect to our main health component, QALYs. This is an important component because healthier individuals survive longer, and treatment improves health conditions. Since this component is largely realized later in life and thus is heavily discounted, improvements in future medical care have a negligible effect on the estimated life-cycle benefits. Dropping this component or doubling the value of life does not have a major impact on our calculations.

Figure 4 summarizes the results from our extensive sensitivity analyses reported in table H.1, including the case where only one of the many streams we consider is the source of the benefit. We calculate the estimates with all possible combinations of the main benefit and cost streams. Our measures of economic efficiency remain statistically and economically significant even after we eliminate the benefits from any one of the four main components that we monetize. Overall, our sensitivity analyses indicate that no single category of outcomes drives the social efficiency of the program. Rather, it is the life-cycle benefits across multiple dimensions of human development.

VI. Assessing Recent Benefit-Cost Analyses

We use our analysis to examine the empirical foundations of the approach to benefit/cost analysis taken in a prototypical study of Kline

[53] If labor markets operate without frictions and the marginal rate of substitution between leisure and consumption equals the marginal wage rate, parental labor income should not be valued at the margin. The bottom box in fig. 4 shows that the benefit/cost ratio and the internal rate of return remain sizable in magnitude and statistically significant if we omit parental income from the benefits attributed to the program.

[54] Two individuals in the treatment group were convicted of rape, and one individual in the control group was convicted of murder.

TABLE 5
SENSITIVITY ANALYSIS FOR INTERNAL RATE OF RETURN (%)

	Pooled		Males		Females	
Baseline[a]	**13.7** (3.3)		**14.7** (4.2)		**10.1** (6.0)	
	No IPW	No IPW, No Controls	No IPW	No IPW, No Controls	No IPW	No IPW, No Controls
Specification	**13.2** (2.9)	**14.0** (3.1)	**13.9** (3.7)	**13.0** (4.3)	9.6 (6.0)	**10.0** (4.9)
	To Age 21	To Age 30	To Age 21	To Age 30	To Age 21	To Age 30
Forecast span	**8.8** (4.5)	**12.0** (3.4)	**11.8** (4.8)	12.8 (4.7)	**10.7** (5.8)	**11.7** (5.2)
	Vs. Stay at Home	Vs. Alt. Presch.	Vs. Stay at Home	Vs. Alt. Presch.	Vs. Stay at Home	Vs. Alt. Presch
Counterfactuals	**9.4** (4.2)	**15.6** (4.3)	6.0 (3.6)	**15.8** (5.0)	**13.4** (5.7)	8.8 (7.0)
	0%	100%	0%	100%	0%	100%
Deadweight loss	**18.3** (4.7)	**11.2** (3.1)	**19.4** (6.2)	**12.1** (3.9)	**17.7** (12.4)	7.1 (4.2)
	Mincer Life Cycle	Life-Cycle Prediction	Mincer Life Cycle	Life-Cycle Prediction	Mincer Life Cycle	Life-Cycle Prediction
Parental income	**15.2** (4.0)	**14.5** (6.4)	**16.0** (5.1)	**14.5** (6.4)	**13.3** (8.2)	**12.3** (9.9)

	.5% Annual Decay	.5% Annual Growth	.5% Annual Decay	.5% Annual Growth	.5% Annual Decay	.5% Annual Growth
Labor income	**13.5**	**13.8**	**14.5**	**14.8**	**9.9**	**10.3**
	(3.4)	(3.2)	(4.3)	(4.1)	(6.0)	(6.0)
	Drop Major Crimes	Halve Costs	Drop Major Crimes	Halve Costs	Drop Major Crimes	Halve Costs
Crime	**10.7**	**11.6**	**12.0**	**11.9**	**10.1**	**9.9**
	(4.4)	(3.8)	(5.3)	(4.9)	(6.0)	(6.0)
	Drop All	Double Value of Life	Drop All	Double Value of Life	Drop All	Double Value of Life
Health (QALYs)	**12.8**	**13.5**	**13.5**	**14.4**	8.8	9.3
	(4.6)	(3.6)	(5.6)	(4.6)	(6.4)	(6.1)

NOTE.—This table displays results of sensitivity analyses of our baseline internal rate of return calculation to the perturbations indexed in the different rows. IPW (inverse probability weighting) adjusts for attrition and item nonresponse (see app. C.2 for details). Control variables are Apgar scores at 1 and 5 minutes and high-risk index (see app. C.8 for details on how we choose these controls). When forecasting up to ages 21 and 30, we consider all benefits and costs up to those ages. As counterfactuals we consider treatment vs. next best (baseline), treatment vs. stay at home, and treatment vs. alternative preschools (see app. F for a discussion). Deadweight loss is the loss implied by any public expenditure (0% is no loss, and 100% is a one-dollar loss per dollar spent). Discount rate is the rate to discount benefits to child's age 0 (in all calculations). For parental labor income, see app. C.3.8 for details on the two alternative forecasts (Mincer and life cycle). For labor income, 0.5 annual growth (decay) is an annual wage growth (decay) due to cohort effects. Major crimes are rape and murder; half costs takes half of victimization and judiciary costs. For health (QALYs), "drop all" sets the value of life equal to zero. Standard errors obtained from the empirical bootstrap distribution are in parentheses. Values in boldface are significant at 10%, using one-sided tests. For details on the null hypothesis, see table H.1.

[a] Characteristics of the baseline calculation: IPW and controls, life span up to predicted death, treatment vs. next best, 50% marginal tax (50% deadweight loss), discount rate of 3%, parental income 0–21 (child's age), labor income predicted from age 21 to 65, all crimes (full costs), value of life at $150,000 (2014 USD).

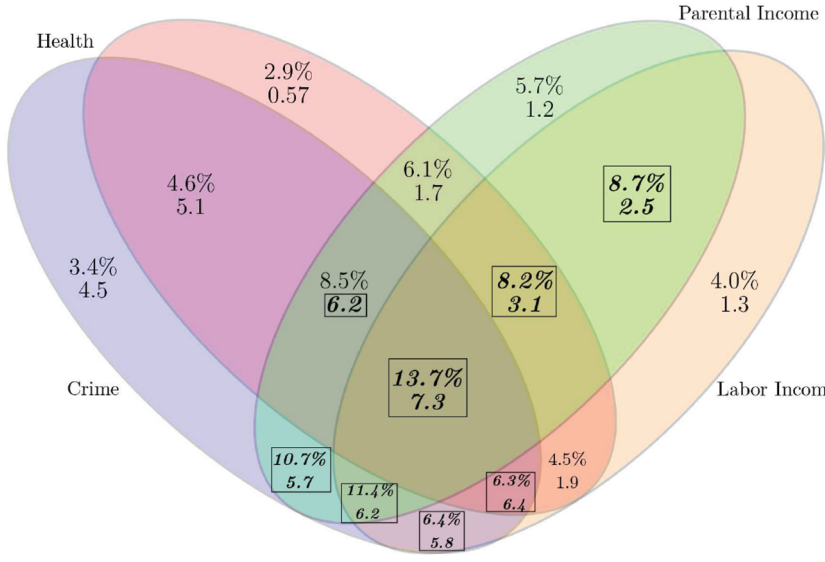

FIG. 4.—Benefit/cost ratio and internal rate of return when accounting for different combinations of the main benefits. This figure presents all possible combinations of accounting for the benefits from the four major categories in our analysis. The nonoverlapping areas present estimates arising from a single category as the benefit. Where multiple categories overlap, we account for benefits from each of the overlapping categories. The other components remain constant across all calculations and are the same as in fig. 3. Health combines QALYs and health expenditure. Inference is based on nonparametric, one-sided *p*-values from the empirical bootstrap distribution. We put boxes around point estimates that are statistically significant at the 10% level.

and Walters (2016), which in turn is based on estimates used in Chetty et al. (2011).[55] Although this approach is widely emulated, we show that it offers imprecise approximations of benefit/cost ratios with questionable validity.

Kline and Walters (2016) use data from the Head Start Impact Study (HSIS) and report a benefit/cost ratio between 1.50 and 1.84.[56] Their analysis proceeds in three steps: (i) calculate program treatment effects on cognitive test scores measured around age 5;[57] (ii) monetize this gain, using the return to cognitive test scores measured between ages 5 and 7, in terms of net present value of labor income at age 27, using the estimates of

[55] Examples of application of this approach include Attanasio, Kugler, and Meghir (2011), Behrman, Parker, and Todd (2011), and Lafortune, Rothstein, and Schanzenbach (2018).

[56] HSIS is a one-year-long randomized evaluation of Head Start (Puma et al. 2010).

[57] They use an index based on the Peabody Picture Vocabulary and Woodcock Johnson III Tests.

Chetty et al. (2011);[58] and (iii) calculate the benefit/cost ratio based on this gain and their own calculations of the program's cost.[59]

To analyze how our estimates compare to those based on this method, we present a series of estimates in column 2 of table 6. For purposes of comparison, column 3 of table 6 shows the analogous estimates based on our samples and forecasts.

For the first estimate, we calculate the benefit/cost ratio, using both the "return to IQ" and the net present value of labor income at age 27 reported in Chetty et al. (2011). This calculation is the same type of calculation as that used in Kline and Walters (2016). In the second exercise, we perform a similar exercise but use our own estimate of the net present value of labor income at age 27.[60] In this exercise, the standard errors account for variation in the return because we calculate the return in every bootstrapped resample. In that sense, our approach more accurately accounts for the underlying uncertainties when compared to the approach of Kline and Walters (2016), who do not account for estimation error in reporting standard errors. The estimated return is smaller because our sample is much more disadvantaged than that used by Chetty et al. (2011).

The remaining exercises in table 6 increase the age range over which we calculate the net present value of labor income or consider the value of all of the components we analyze throughout the paper, in addition to labor income. The more inclusive the benefits measured and the longer the horizon over which they are measured, the greater the benefit/cost ratio. The final reported estimate, 7.3, is our baseline estimate that incorporates all of the components across the life cycles of the subjects.

Our methodology provides a more accurate estimate of the benefits and costs of the ABC/CARE program. We better quantify the effects of the experiment by considering the full array of benefits over the whole life cycle. We also better approximate the uncertainty of our estimates by considering the sampling error in both the experimental and auxiliary samples, the forecast error due to interpolation and extrapolation, and the sensitivity of the estimates to externally specified parameters.

[58] The Chetty et al. (2011) return is based on Stanford Achievement Tests.

For this comparison exercise, we interpret the earnings estimated in Chetty et al. (2011) to be equivalent to labor income. Calculations from Chetty et al. (2011) indicate that a 1 standard deviation gain in achievement scores at age 5 implies a 13.1% increase in the net present value of labor income through age 27. This is based on combining information from Project Star and administrative data at age 27.

[59] Their calculation assigns the net present value of labor income through age 27 of $385,907.17 to the control-group participants, as estimated by Chetty et al. (2011).

All monetary values that we provide in this section are in 2014 USD. We discount the value provided by Chetty et al. (2011) to the age of birth of the children in our sample (first cohort).

[60] This allows us to compute our own "return to IQ" and impute it to the treatment-group individuals.

TABLE 6
EXAMINING THE VALIDITY OF RECENT AD HOC METHODS FOR FORECASTING IN LIGHT OF THE ANALYSIS OF THIS PAPER

Age, Source	Component (1)	Kline and Walters (2016) Method (2)	Authors' Method (3)
27:			
Chetty et al. (2011)	Labor income	.58 (.28)	
ABC/CARE calculated	Labor income	.09 (.04)	1.09 (.04)
34:			
ABC/CARE calculated	Labor income	.37 (.04)	.15 (.05)
ABC/CARE calculated	All	1.21 (.05)	3.20 (1.04)
Life cycle:			
ABC/CARE calculated	Labor income	1.56 (.08)	1.55 (.76)
ABC/CARE calculated	All	3.80 (.29)	7.33 (1.84)

NOTE.—This table displays benefit/cost ratios based on the methodology in Kline and Walters (2016) and on our own methodology. Age is the age at which we stop calculating the net present value. Component refers to the item used to compute net present value ("all" refers to the net present value of all the components). Standard errors (in parentheses) are based on the empirical bootstrap distribution.

In the concluding portion of this section, we consider how well researchers would do if they aim to forecast life-cycle benefits applying our methodology but lacking access to data through the mid-30s. For example, suppose that a researcher had access to inputs of the production function that are experimentally shifted by the treatment, but only up to age 21. Studying the performance of our forecasting procedure in this case is an additional examination of robustness.

Researchers seeking to implement a formal forecasting methodology may face severe data limitations. Often, data are lacking beyond early ages (e.g., age 5). Our analysis can inform researchers on how precise such forecasts might be if they are based only on data from earlier segments of the life cycle. The first row of table 6 gives information on how well test scores at ages 5–7 predict the benefit-cost ratio of the program based only on labor income through age 27. For reasons discussed above, this produces a poor approximation.

We present two additional analyses. First, we analyze the case where we limit $\boldsymbol{X}_{k,a}^{d}$ to PIAT scores and years of education but use only data up to age 21 (the previous wave of the survey). As before, we use the average PIAT achievement score at ages 5–7. An analysis based on these two measures requires only data through the age at which education is completed—generally long before age 30. Second, we conduct an analysis where we limit $\boldsymbol{X}_{k,a}^{d}$ to contain only lagged labor income at age 21. Researchers often have access to a test score as well as a measure of schooling or may have only a measure of labor income early in life with which they could initialize a forecast of autoregressive labor income. For both of these cases, we produce forecasts for the entire life cycle of labor income.

In the first analysis, we use PIAT scores and years of education to predict income from age 21 to age 65 (assumed age of retirement). In the second analysis, we use an autoregressive model of earnings—see, for example, Meghir and Pistaferri (2011)—to predict labor income from age 21 to age 65. As in our baseline forecasting procedure, we are able to initialize this forecast in the experimental sample because we observe labor income at age 21.[61] We otherwise follow the steps in section III.B and reestimate the forecasting functions in the nonexperimental sample, as researchers would do if they were to forecast with these subsets of predictor variables. We then use the estimated functions to forecast in the experimental sample.

Figure 5 shows the results from analyses by displaying labor income forecasts analogous to figure 2 for the two sets of predictor variables. We do not vary the construction of the nonexperimental samples used for forecasting or the background variables that we include. The forecasts show that PIAT scores and years of education used together provide a forecast that is imprecise but in the ballpark of the baseline forecast in figure 2. Lagged income alone does not suffice to accurately forecast treatment-control life-cycle differences. This makes sense because we initialize the forecast at age 21, where the treatment-control difference in labor income is not informative because many of the subjects—especially those in treatment—are still in school. Thus, while a forecast based on short-term test scores does not produce an accurate forecast, a short-term test score together with education is close to being on the mark.

VII. Summary

This paper presents a template for constructing economically interpretable summaries of the multiple treatment effects generated from a randomized evaluation of a high-quality, widely emulated early-childhood program with follow-up through the mid-30s. We go beyond the usual practice of reporting batteries of treatment effects. We report the costs and monetize the treatments across numerous domains. We estimate the tax-adjusted internal rate of return and the benefit/cost ratio of the program to assess the social efficiency of the program.

We use auxiliary information and structural economic models to guide monetization of treatment effects and to extrapolate the measured benefits and costs to the full life cycles of participants. We account for model

[61] We do not present results based solely on a short-term test score such as the PIAT because the results are extremely imprecise, which is consistent with the results in table 4 and warns practitioners against forecasting based on a short-term test score, although this is commonly done (e.g., Kline and Walters 2016). Our test scores are measured later than theirs (at ages 3–4) and are likely more precise.

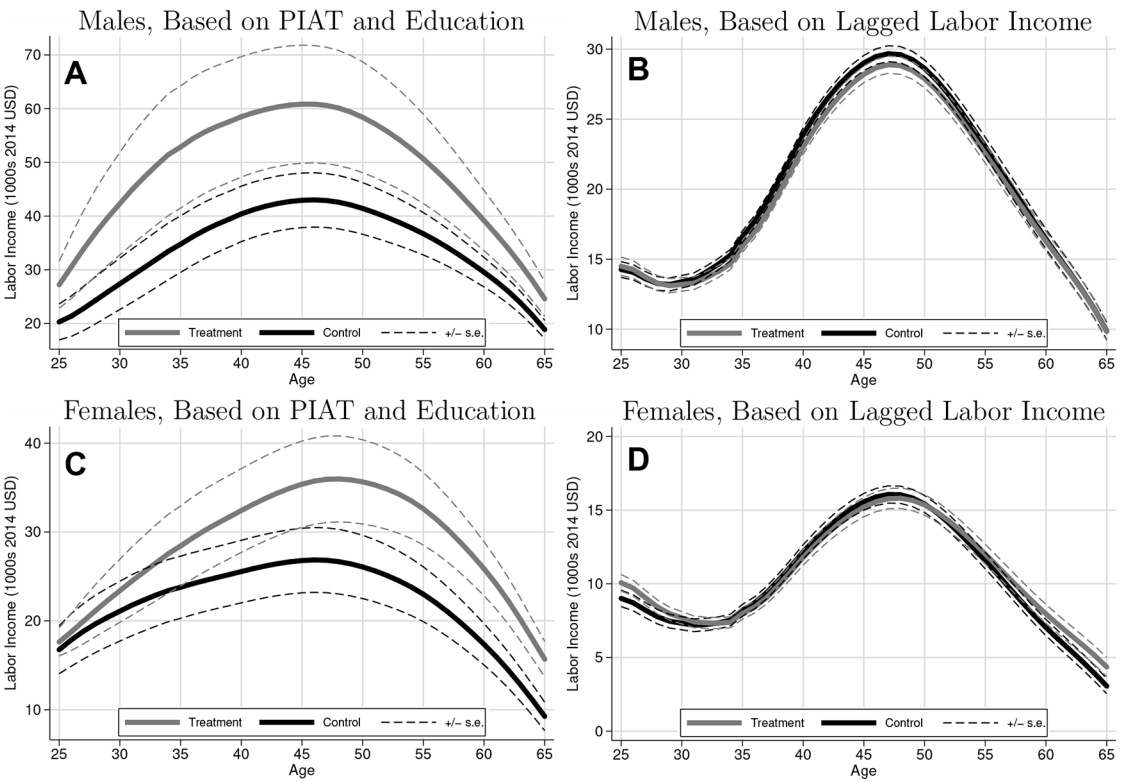

FIG. 5.—Forecast life-cycle labor income profiles for ABC/CARE males (*A*, *B*) and females (*C*, *D*) by treatment status, analogous to the forecasts in fig. 2 but using either PIAT scores (at age 5–7) and years of education (*A*, *C*) or lagged labor income (*B*, *D*) as predictors. Figure 2 uses PIAT scores (5–7), years of education as predictors, and lagged labor income.

estimation and forecast error and conduct extensive sensitivity analyses of our estimates to alternative assumptions and methodologies. Under a variety of plausible assumptions, we estimate that the tax-adjusted internal rate of return of the program ranges from 8% to 18.3%. These estimates demonstrate the social profitability of ABC/CARE. We show that forecasts from a robust nonparametric matching strategy are close to those from our structural approach.

We conclude with a cautionary note. The program we study was targeted to a relatively homogenous, disadvantaged, and predominately African American population in a university town in North Carolina. Generalization of our findings to other populations should proceed with caution.[62] In particular, there is no basis for using this study to argue for universal application of ABC/CARE across all socioeconomic groups. However, the essential features of the ABC/CARE approach are currently in wide use in a variety of early-childhood intervention programs that target disadvantaged children. In this sense, our analysis has lessons of general interest for disadvantaged populations. Our study indicates what is possible and that the possibilities are substantial.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *J. American Statis. Assoc.* 105 (490): 493–505.

Almond, Douglas, and Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *J. Econ. Perspectives* 25 (3): 153–72.

Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial." *American Econ. J.: Appl. Econ.* 3 (3): 188–220.

Barnett, W. Steven, and Leonard N. Masse. 2002. "A Benefit-Cost Analysis of the Abecedarian Early Childhood Intervention." Tech. report, Nat. Inst. Early Educ. Res., New Brunswick, NJ.

———. 2007. "Comparative Benefit-Cost Analysis of the Abecedarian Program and Its Policy Implications." *Econ. Educ. Rev.* 26 (1): 113–25.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2011. "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of PROGRESA/Oportunidades." *J. Human Resources* 46 (1): 93–122.

Belfield, Clive R., Milagros Nores, W. Steven Barnett, and Lawrence J. Schweinhart. 2006. "The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup." *J. Human Resources* 41 (1): 162–90.

Blundell, Richard, Michael Graber, and Magne Mogstad. 2015. "Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family." *J. Public Econ.* 127:58–73.

Campbell, Frances A., Gabriella Conti, James J. Heckman, et al. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science* 343 (6178): 1478–85.

[62] Especially problematic are forecasts over the supports of our samples.

Campbell, Frances A., Elizabeth P. Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey. 2001. "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment." *Developmental Psychology* 37 (2): 231–42.

Campbell, Frances A., and Craig T. Ramey. 1995. "Cognitive and School Outcomes for High-Risk African-American Students at Middle Adolescence: Positive Effects of Early Intervention." *American Educ. Res. J.* 32 (4): 743–72.

Campbell, Frances A., Craig T. Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson. 2002. "Early Childhood Education: Young Adult Outcomes from the Abecedarian Project." *Appl. Developmental Sci.* 6 (1): 42–57.

Campbell, Frances A., Barbara Wasik, Elizabeth Pungello, et al. 2008. "Young Adult Outcomes of the Abecedarian and CARE Early Childhood Educational Interventions." *Early Childhood Res. Q.* 23 (4): 452–66.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Q.J.E.* 126 (4): 1593–660.

Clarke, Stevens H., and Frances A. Campbell. 1998. "Can Intervention Early Prevent Crime Later? The Abecedarian Project Compared with Other Programs." *Early Childhood Res. Q.* 13 (2): 319–43.

Cohen, Mark A., and Roger Bowles. 2010. "Estimating Costs of Crime." In *Handbook of Quantitative Criminology*, edited by Alex R. Piquero and David Weisburd, 143–61. New York: Springer.

Collins, Ann, Barbara D. Goodson, Jeremy Luallen, Alyssa Rulf Fountain, Amy Checkoway, and Abt Associates. 2010. "Evaluation of Child Care Subsidy Strategies: Massachusetts Family Child Care Study." Tech. Report OPRE 2011-1, Office Planning, Res. and Evaluation, Admin. Children and Families, US Dept. Health and Human Services, Washington, DC.

Cunha, Flávio, and James J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *J. Human Resources* 43 (4): 738–82.

Cunha, Flávio, James J. Heckman, Lance J. Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, vol. 1, edited by Eric A. Hanushek and Finis Welch, 697–812. Amsterdam: North-Holland.

Cunha, Flávio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

Dolan, Paul. 1997. "Modeling Valuations for EuroQol Health States." *Medical Care* 35 (11): 1095–108.

Educare. 2014. "A National Research Agenda for Early Education." Tech. report, Educare Learning Network Res. and Evaluation Committee, Chicago.

Elango, Sneha, Jorge Luis García, James J. Heckman, and Andrés Hojman. 2016. "Early Childhood Education." In *Economics of Means-Tested Transfer Programs in the United States*, vol. 2, edited by Robert A. Moffitt, 235–97. Chicago: Univ. Chicago Press (for NBER).

Feldstein, Martin. 1999. "Tax Avoidance and the Deadweight Loss of the Income Tax." *Rev. Econ. and Statis.* 81 (4): 674–80.

Fox Business News. 2014. "Head Start Has Little Effect by Grade School?" Video. http://video.foxbusiness.com/v/3306571481001/head-start-has-little-effect-by-gradeschool/?#sp=show-clips.

García, Jorge Luis, and James J. Heckman. 2016. "How Would a National Implementation of Early Childhood Interventions Narrow the Intra-black and Black-White Outcome Gaps?" Manuscript, Dept. Econ., Univ. Chicago.

García, Jorge Luis, James J. Heckman, and Anna L. Ziff. 2018. "Gender Differences in the Benefits of an Influential Early Childhood Program." *European Econ. Rev.* 109:9–22.

Gladden, Tricia, and Christopher Taber. 2000. "Wage Progression among Less Skilled Workers." In *Finding Jobs: Work and Welfare Reform*, edited by David E. Card and Rebecca M. Blank, 160–92. New York: Russell Sage Found.

Goldman, Dana P., Darius Lakdawalla, Pierre-Carl Michaud, et al. 2015. "The Future Elderly Model: Technical Documentation." Tech. report, Univ. Southern California.

Gross, Ruth T., Donna Spiker, and Christine W. Haynes, eds. 1997. *Helping Low Birth Weight, Premature Babies: The Infant Health and Development Program*. Redwood City, CA: Stanford Univ. Press.

Haavelmo, Trygve. 1943. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica* 11 (1): 1–12.

Havnes, Tarjei, and Magne Mogstad. 2011. "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes." *American Econ. J.: Econ. Policy* 3 (2): 97–129.

Healthy Child Manitoba. 2015. "Starting Early, Starting Strong: A Guide for Play-Based Early Learning in Manitoba: Birth to Six." Tech. report, Healthy Child Manitoba, Winnipeg.

Heckman, James J. 1992. "Randomization and Social Policy Evaluation." In *Evaluating Welfare and Training Programs*, edited by Charles F. Manski and Irwin Garfinkel, 201–30. Cambridge, MA: Harvard Univ. Press.

Heckman, James J., Neil Hohmann, Jeffrey Smith, and Michael Khoo. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *Q.J.E.* 115 (2): 651–94.

Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra E. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–98.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Q. Yavitz. 2010a. "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program." *Quantitative Econ.* 1 (1): 1–46.

———. 2010b. "The Rate of Return to the HighScope Perry Preschool Program." *J. Public Econ.* 94 (1–2): 114–28.

Heckman, James J., and Salvador Navarro. 2004. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." *Rev. Econ. and Statis.* 86 (1): 30–57.

Heckman, James J., and Rodrigo Pinto. 2015. "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs." *Econometric Rev.* 34 (1–2): 6–31.

Heckman, James J., Rodrigo Pinto, and Peter A. Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *A.E.R.* 103 (6): 2052–86.

Henderson, Frederick W., Albert M. Collier, Margaret A. Sanyal, et al. 1982. "A Longitudinal Study of Respiratory Viruses and Bacteria in the Etiology of Acute Otitis Media with Effusion." *New England J. Medicine* 306 (23): 1377–83.

Hurwicz, Leonid. 1962. "On the Structural Form of Interdependent Systems." In *Logic, Methodology and Philosophy of Science*, edited by Ernest Nagel, Patrick Suppes, and Alfred Tarski, 232-39. Stanford, CA: Stanford Univ. Press.

Jensen, Bente, and Marlene Nielsen. 2016. "Abecedarian Programme, within an Innovative Implementation Framework (APIIF). A Pilot Study." http://pure.au.dk/portal/da/projects/abecedarian-programme-within-an-innovative-implementation-framework-apiif-a-pilot-study-en-dansk-version-af-abecedarian–en-interaktionistisk-tilgang-til-laeringsrelationer-mellem-professionelle-og-03-aarige-boern-i-dagtilbud(5ade06a4-56d0-49f8-8ed1-1177deca7537).html.

Kline, Patrick, and Christopher Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Q.J.E.* 131 (4): 1795–848.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. 2018. "School Finance Reform and the Distribution of Student Achievement." *American Econ. J.: Appl. Econ.* 10 (2): 1–26.

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. 2018. "Life Cycle Wage Growth across Countries." *J.P.E.* 126 (2): 797–849.

Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide. 2016. "Forecasting with Dynamic Panel Data Models." Res. Paper No. 17-02, Inst. New Econ. Thinking, Univ. Southern California.

McCollister, Kathryn E., Michael T. French, and Hai Fang. 2010. "The Cost of Crime to Society: New Crime-Specific Estimates for Policy and Program Evaluation." *Drug and Alcohol Dependence* 108 (1–2): 98–109.

Meghir, Costas, and Luigi Pistaferri. 2011. "Earnings, Consumption and Life Cycle Choices." In *Handbook of Labor Economics*, vol. 4, edited by Orley C. Ashenfelter and David Card, 773–854. Amsterdam: North-Holland.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*, 2nd ed. New York: Cambridge Univ. Press.

Prentice, Ross L. 1989. "Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria." *Statis. Medicine* 8 (4): 431–40.

Puma, Michael, Stephen Bell, Ronna Cook, and Camilla Heid. 2010. "Head Start Impact Study: Final Report." Tech. report, Office Planning, Res. and Evaluation, Admin. Children and Families, US Dept. Health and Human Services, Washington, DC.

Quandt, Richard E. 1972. "A New Approach to Estimating Switching Regressions." *J. American Statis. Assoc.* 67 (338): 306–10.

Ramey, Craig T., Donna M. Bryant, Joseph J. Sparling, and Barbara H. Wasik. 1985. "Project CARE: A Comparison of Two Early Intervention Strategies to Prevent Retarded Development." *Topics Early Childhood Special Educ.* 5 (2): 12–25.

Ramey, Craig T., Albert M. Collier, Joseph J. Sparling, et al. 1976. "The Carolina Abecedarian Project: A Longitudinal and Multidisciplinary Approach to the Prevention of Developmental Retardation." In *Intervention Strategies for High-Risk Infants and Young Children*, edited by Theodore Tjossem, 629–55. Baltimore, MD: Univ. Park.

Ramey, Craig T., Joseph J. Sparling, and Sharon L. Ramey. 2012. *Abecedarian: The Ideas, the Approach, and the Findings*. Los Altos, CA: Sociometrics.

———. 2014. "Interventions for Students from Low Resource Environments: The Abecedarian Approach." In *Essentials of Planning, Selecting and Tailoring Interventions for Unique Learners*, edited by Jennifer T. Mascolo, Vincent C. Alfonso, and Dawn P. Flanagan, 415–48. Hoboken, NJ: Wiley.

Ridder, Geert, and Robert Moffitt. 2007. "The Econometrics of Data Combination." In *Handbook of Econometrics*, vol. 6B, edited by James J. Heckman and Edward E. Leamer, 5469–547. Amsterdam: Elsevier.

Sanders, Carl, and Christoper Taber. 2012. "Life-Cycle Wage Growth and Heterogeneous Human Capital." *Ann. Rev. Econ.* 4:399–425.

Schneider, Barbara, and Sarah-Kathryn McDonald, eds. 2007. *Scale-Up in Education*, vol. 1. *Ideas in Principle*. Lanham, MD: Rowman & Littlefield.

Scull, Janet, John Hattie, Jane Page, et al. 2015. "Building a Bridge into Preschool in Remote Northern Territory Communities." https://education.unimelb.edu.au/research/projects/building-a-bridge-into-preschool-in-remote-northern-territory-communities.

Shaw, James W., Jeffrey A. Johnson, and Stephen Joel Coons. 2005. "US Valuation of the EQ-5D Health States: Development and Testing of the D1 Valuation Model." *Medical Care* 43 (3): 203–20.

Sparling, Joseph J. 1974. "Synthesizing Educational Objectives for Infant Curricula." Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, April.

———. 2010. "Highlights of Research Findings from the Abecedarian Studies." Tech. report, Teaching Strategies, Bethesda, MD. https://www2.teachingstrategies.com/content/pageDocs/Abecedarian-Research-Findings-Highlights.pdf.

Wasik, Barbara H., Craig Ramey, Donna M. Bryant, and Joseph J. Sparling. 1990. "A Longitudinal Study of Two Early Intervention Strategies: Project CARE." *Child Development* 61 (6): 1682–96.

Whitehurst, Grover J. 2014. Testimony given to the Health, Education, Labor, and Pensions Committee of the United States Senate, April 10. https://www.help.senate.gov/imo/media/doc/Whitehurst.pdf.

Yazejian, Noreen, and Donna M. Bryant. 2012. "Educare Implementation Study Findings." Tech. report, Frank Porter Graham Child Development Inst., Univ. North Carolina, Chapel Hill.