

Quantifying the Life-cycle Benefits of an Influential Early Childhood Program*

Jorge Luis García
John E. Walker Department of Economics
Clemson University
Social Science Research Institute
Duke University

Duncan Ermini Leaf
Leonard D. Schaeffer Center
for Health Policy and Economics
University of Southern California

James J. Heckman
American Bar Foundation
Center for the Economics
of Human Development
The University of Chicago

María José Prados
Dornsife Center for
Economic and Social Research
University of Southern California

First Draft: January 5, 2016
This Draft: August 2, 2018

*This research was supported in part by grants from the Robert Wood Johnson Foundation's Policies for Action program, NICHD R37HD065072, the American Bar Foundation, the Buffett Early Childhood Fund, the Pritzker Children's Initiative, NICHD R01HD054702, NIA R01AG042390, and by the National Institute On Aging of the National Institutes of Health under Award Number P30AG024968. 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 Chapel Hill for the use of data and source materials from the Carolina Abecedarian Project and the Carolina Approach to Responsive Education. Years of partnership and collaboration have made this work possible. Andrés Hojman, Yu Kyung Koh, Sylvi Kuperman, Stefano Mosso, Rodrigo Pinto, Joshua Shea, Jake Torcasso, and Anna Ziff contributed to the analysis in this paper. 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 Future Adult Model to make the health projections used in this paper. For helpful comments on various versions of the paper, we thank the editor, Harald Uhlig, two anonymous referees, Stéphane Bonhomme, Neal 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 Vytlačil, 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 Carolina Abecedarian Project and the Carolina Approach to Responsive Education 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.

Abstract

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 non-experimental 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 non-experimental data using economic theory to estimate the long-run life-cycle benefits of social programs.

Keywords: Early childhood education, life-cycle benefits, long-term forecasts, program evaluation, rates of return, cost-benefit analysis

JEL codes: J13, I28, C93

Jorge Luis García
John E. Walker
Department of Economics
Clemson University
228 Surrine Hall
Clemson, SC 29630
Phone: 773-449-0744
Email: jorge.ggmenendez@gmail.com

James J. Heckman
Center for the Economics
of Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Phone: 773-702-0634
Email: jjh@uchicago.edu

Duncan Ermini Leaf
Leonard D. Schaeffer Center for
Health Policy & Economics
University of Southern California
635 Downey Way
Los Angeles, CA 90089
Phone: 213-821-6474
Email: dleaf@healthpolicy.usc.edu

María José Prados
Dornsife Center for
Economic and Social Research
University of Southern California
635 Downey Way
Los Angeles, CA 90089
Phone: 213-821-7969
Email: prados@usc.edu

1 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 essentially identical 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 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. The program has a 13.7% (s.e. 3%) per-annum tax-adjusted internal rate of return and a 7.3 (s.e. 1.84) tax-adjusted benefit/cost ratio.

The program is a prototype for many programs planned or in place today.⁵ About 19%

¹See [Cunha et al. \(2006\)](#), [Almond and Currie \(2011\)](#), [Duncan and Magnuson \(2013\)](#), and [Elango et al. \(2016\)](#) for surveys.

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

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 and in providing a general methodology that can be replicated to assess the social efficiency of other programs.

⁴A companion paper, [García et al. \(2018\)](#), reports these treatment effects. Participants in ABC/CARE benefit in terms of both cognitive and socio-emotional skills, education, employment and labor income, and risky behavior and health. The parents of participants benefit in terms of labor income and education.

⁵Programs inspired by ABC/CARE have been (and are currently being) launched around the world. [Sparling \(2010\)](#) and [Ramey et al. \(2014\)](#) list numerous programs based on the ABC/CARE approach. The programs are: Infant Health and Development Program (IHDP) in eight different cities in the U.S. ([Spiker et al., 1997](#)); Early Head Start and Head Start. ([Schneider and McDonald, 2007](#)); John’s Hopkins Cerebral Palsy Study in the U.S. ([Sparling, 2010](#)); Classroom Literacy Interventions and Outcomes (CLIO) study. ([Sparling, 2010](#)); Massachusetts Family Child Care Study ([Collins et al., 2010](#)); Healthy Child Manitoba Evaluation ([Healthy Child Manitoba, 2015](#)); 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 U.S. are also based on

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 non-experimental data guided by economic theory.⁸

As a byproduct, 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 non-experimental samples. The cohorts are chosen to approximate the life cycles of experimentals in their post-experimental years. We formulate and estimate production functions that predict program treatment effects and assess their within-sample forecast accuracy. Some of the inputs of the estimated production functions are changed by treatment and are measured in both experimental and non-experimental

ABC/CARE (Educare, 2014; Yajejian and Bryant, 2012). Appendix A.8 lists these Educare programs, all of which implement curricula based on ABC/CARE.

⁶43% of African-American children were eligible at its inception.

⁷García and Heckman (2016) estimate that if ABC/CARE were implemented on the current stock of eligible children, the intra-black 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 intra-black 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 USD (2014). For males, the program would eradicate the intra-black high-school graduation gap, reduce the years of education gap to 0.18 years, and reduce the employment gap to 9 percentage points.

⁸Ridder and Moffitt (2007) discuss data combination methods. These methods are related to the older “surrogate marker” literature in biostatistics (see e.g., Prentice, 1989). However, as noted below, exogeneity 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.

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 in experimental and non-experimental samples that overlap in age.⁹ We forecast experimental treatment effects using our estimated production functions applied to non-experimental data with inputs and outputs suitably adjusted for cohort effects.

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 non-experimental 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 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 2](#) describes the ABC/CARE program. [Section 3](#) discusses our methodology for forecasting life-cycle outcomes and the evidence supporting our assumptions. To make matters concrete, we consider in detail how we forecast life cycle labor income. [Section 4](#) discusses how we forecast other life-cycle outcomes. [Section 5](#) reports baseline estimates of internal rates of return and benefit/cost ratios and reports an array of sensitivity analyses. [Section 6](#) uses our estimates to examine the predictive validity of widely used informal forecasting methods. [Section 7](#) summarizes our findings.

⁹See [Hurwicz \(1962\)](#) for the definition treatment (policy) invariance. We build on the methodology of [Heckman et al. \(2013\)](#), who relate intermediate and long-term outcomes in a mediation analysis. However, they do not construct out-of-sample forecasts.

2 ABC/CARE: Background

The Carolina Abecedarian Project and the Carolina Approach to Responsive Education (ABC/CARE) were enriched childcare programs that targeted the early years of disadvantaged, predominately African-American children in Chapel Hill, Durham, and Raleigh in North Carolina.¹⁰ Appendix A describes these programs in detail. We summarize their main features here.

The goal of these programs was to enhance the life skills of disadvantaged children. They supported language, motor, and cognitive development as well as socio-emotional competencies considered crucial for school success including task orientation, the ability to communicate, independence, and pro-social behavior.¹¹ They also provided free health screenings to treatment-group members with costs of medications borne by the parents.

The design and implementation of ABC and CARE were very similar. Both had two phases, the first of which lasted from birth until age 5. In this phase, children were randomly assigned to treatment. The second phase of the study consisted of home visits from ages 5 through 8 that were assigned by 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.¹²

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.¹³

Our analysis uses data from the first phase and pools the ABC treatment group with that of the CARE treatment groups. We do not use the data from the CARE group that

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

¹¹Sparling (1974); Ramey et al. (1976, 1985); Wasik et al. (1990); Ramey et al. (2012).

¹²Wasik et al. (1990).

¹³See Appendix A.2 for details on the construction of the “risk index” used to determine eligibility.

only received home visits 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. Due to attrition and non-response, 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.¹⁴ We use standard weighting methodologies to account for attrition, non-response, and missing data.¹⁵

For both programs, data were collected annually on cognitive and socio-emotional skills, home environments, family structure, and family economic characteristics from birth until age 8. After age 8, data on cognitive and socio-emotional skills, education, and family economic characteristics were collected at ages 12, 15, 21, and 30.¹⁶ In addition, there is information from administrative criminal records, and from a physician-administered medical survey, when the subjects were in their mid 30s.¹⁷ Many control-group children in both ABC and CARE attended alternative formal childcare arrangements (75% and 74% respectively).¹⁸

3 Forecasting Life-cycle Costs and Benefits: Method

To explain our forecasting procedure, the following notation is useful. $W = 1$ indicates that parents referred to the program participate in the randomization protocol. $W = 0$

¹⁴There were no randomization compromises in CARE. During preschool, 5 subjects attrited (3 in the treatment group, 1 in the family education group, and 1 in the control group). Details on attrition and non-response are presented in [Appendix A.3](#).

¹⁵See [Appendix C.2](#).

¹⁶At age 30, measures of cognitive skills are not available for both ABC and CARE.

¹⁷See [Appendix A.7](#) for a more comprehensive description of the data. There, we 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 is substantial missing data for these outcomes, which we address using the methodology exposted in [Appendix C](#).

¹⁸The alternative arrangements were generally lower quality than ABC/CARE (see [Appendix A.6.1](#) for details). In our main analysis, we compare treatment- and control-group children, irrespective of take-up of alternatives. In [Appendix 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.

indicates otherwise. R indicates randomization into treatment ($R = 1$) or control ($R = 0$). D indicates whether or not a family attends the program. $D = R$ implies compliance with the initial randomization protocol. Lowercase variables d and r denote realizations of D and R , respectively. We suppress individual subscripts to avoid notational clutter.

Individuals are eligible to participate in the program if their baseline background variables $\mathbf{B} \in \mathcal{B}_0$, 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.¹⁹

Define \mathbf{Y}_a^1 as the outcome vector at age a for the treated. \mathbf{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 $\mathbf{B} \in \mathcal{B}_0$ is:

$$\Delta_a := \mathbb{E} [\mathbf{Y}_a^1 - \mathbf{Y}_a^0 | W = 1] = \mathbb{E} [\mathbf{Y}_a^1 - \mathbf{Y}_a^0 | \mathbf{B} \in \mathcal{B}_0]. \quad (1)$$

Randomization identifies this parameter in the experimental sample.

3.1 Using Economic Models to Make Forecasts

This paper uses economic models to generate unbiased, out-of-sample forecasts of Δ_a . We use a structural production function (mediation) model for treatment ($D = 1$) and control ($D = 0$) outcomes at age a in sample $k \in \{e, n\}$, where e denotes membership in the experimental sample and n denotes membership in a non-experimental (auxiliary) sample. The vector of production functions for output $\mathbf{Y}_{k,a}^d$ is:

$$\mathbf{Y}_{k,a}^d = \phi_{k,a}^d(\mathbf{X}_{k,a}^d, \mathbf{B}_k) + \boldsymbol{\varepsilon}_{k,a}^d, \quad (2)$$

¹⁹All 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 et al., 2012).

$d \in \{0, 1\}, k \in \{e, n\}, a \in \{1, \dots, \bar{A}\}$ where $\phi_{k,a}^d(\cdot, \cdot)$ is a vector of structural production relationships mapping inputs $\mathbf{X}_{k,a}^d, \mathbf{B}_k$ into outputs holding the error term $\varepsilon_{k,a}^d$ fixed.²⁰ \mathbf{B}_k are baseline variables not affected by treatment. $\mathbf{X}_{k,a}^d$ are variables potentially affected by treatment. \bar{A} is 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 $\mathbf{X}_{k,a}^d, \mathbf{B}_k$ and outputs $\mathbf{Y}_{k,a}^d$ can, in principle, differ between experimental and non-experimental 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 non-experimental sample with no direct empirical counterpart for the case $d = 1$. We present conditions for identifying and estimating $\phi_{k,a}^d(\cdot, \cdot)$ in non-experimental samples. A crucial condition is that for fixed values of inputs $\mathbf{X}_{k,a}^d = \mathbf{x}, \mathbf{B}_k = \mathbf{b}$ 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 $\mathbf{x}, \mathbf{b} \in \text{supp}(\mathbf{X}_{k,a}^d, \mathbf{B}_k), k \in \{e, n\}$, and $a \in \{1, \dots, \bar{A}\}$

$$\begin{aligned} \phi_{k,a}^0(\mathbf{x}, \mathbf{b}) &= \phi_{k,a}^1(\mathbf{x}, \mathbf{b}) \\ &= \phi_a(\mathbf{x}, \mathbf{b}), \end{aligned} \tag{3a}$$

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

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

where $F_{k,a}^j(\mathbf{z} | \mathbf{Fix} \Omega = \omega)$ is the distribution of \mathbf{Z} for Ω fixed at ω and $F_a(\mathbf{z} | \mathbf{Fix} \Omega = \omega)$ is the age- a distribution of the errors of the production functions, assumed to be common across treatment regimes and samples given $\mathbf{X}_{k,a}^d = \mathbf{x}$ and $\mathbf{B}_k = \mathbf{b}$. \square

We clarify Assumption A–1 with two remarks.

²⁰Fixing 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.

Remark R–1 *There are 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 non-experimental samples. Condition (ii) enables analysts to simulate treatment and control outcomes in non-experimental 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 non-experimental samples, embeds an implicit assumption about the absence of cohort effects in the post-sample period for the experimental sample. In particular, a structural function $\phi_{n,a}^d(\mathbf{x}, \mathbf{b})$ or distribution $F_{n,a}^d(\mathbf{z} | \mathbf{Fix} \Omega = \omega)$, which we identify and estimate in the non-experimental sample with subjects who are older than those in the experimental sample, is a valid tool for forecasting the experimental sample at older 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, $\mathbf{Y}_{k,a}^d$. Instead, it means that there are no cohort effects in the **mapping** between $\mathbf{X}_{k,a}$, \mathbf{B}_k and $\mathbf{Y}_{k,a}^d$, $k \in \{e, n\}$.

To test Assumption A–1 in the experimental and non-experimental samples and to make accurate forecasts at age $a > a^*$, we require the common support conditions that we now formalize. They are tested in Appendix C.3.6.

Assumption A–2 *Support and Input Forecast Conditions* For $a \in \{1, \dots, \bar{A}\}$ and $d \in \{0, 1\}$

$$\text{supp}(\mathbf{Y}_{e,a}^d, \mathbf{X}_{e,a}^d, \mathbf{B}_e, \boldsymbol{\varepsilon}_{e,a}) \subseteq \text{supp}(\mathbf{Y}_{n,a}^d, \mathbf{X}_{n,a}^d, \mathbf{B}_n, \boldsymbol{\varepsilon}_{n,a}). \quad (4)$$

We also require that the path of inputs in the post-treatment lives of the experimental treatments and controls have counterparts in the non-experimental sample. This includes accounting for cohort effects.

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

3.2 Forecasting Labor Income

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 (\mathbf{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 PSID together, we combine samples. Thus we use three non-experimental datasets to obtain information across the life cycle. We satisfy support conditions.²¹

Because we do not observe each element of the eligibility index discussed in Section 2, we approximate $\mathbf{B}_n \in \mathcal{B}_0$. We delimit the sample to include observations satisfying the following criteria: (i) **NLSY79**: Black, labor income less than \$300,000 (2014 USD) in any given year, birth year between 1957 and 1965; (ii) **PSID**: Black, labor income less than \$300,000 (2014 USD), birth year between 1945 and 1981; and (iii) **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 non-experimental samples according to their resemblance to individuals in the experimental sample.²² We match on baseline pre-treatment match variables: year of birth, gender, and number of siblings at baseline. All are available in the non-experimental datasets. This procedure generates a synthetic cohort in the non-experimental sample for subsequent analysis in our structural forecasting procedure.

By design, there is no treatment effect in the non-experimental 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 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.²³

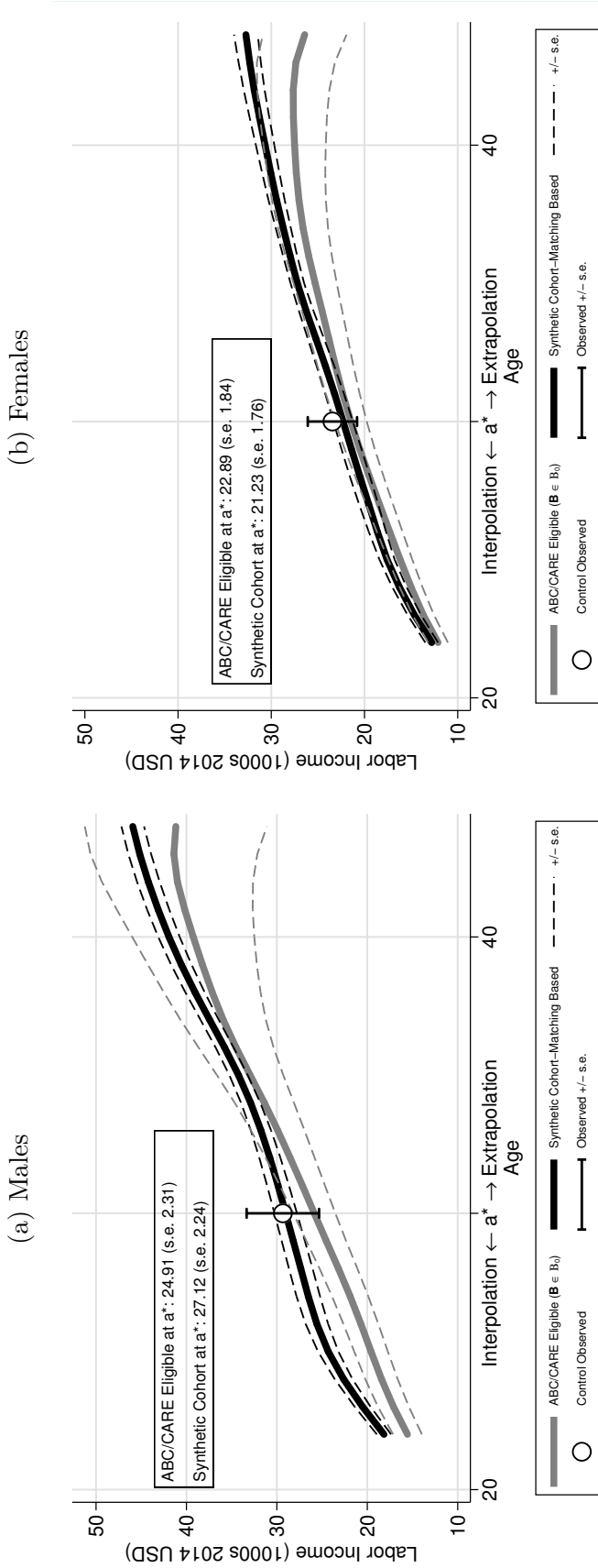
Step 2. Establishing Exogeneity of Inputs. Forecasting does not require that we take a position on the exogeneity of $\mathbf{X}_{k,a}^d$ for $k \in \{e, n\}$ and $a \in \{1, \dots, \bar{A}\}$ with respect

²¹See Appendix C.3.6.

²²We use Mahalanobis' matching Algorithm 1 and weights derived from the Mahalanobis distances that downweight dissimilar observations. See Appendix C for details.

²³We 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.

Figure 1: Labor Income Profile, Disadvantaged Individuals and Synthetic Cohort Constructed by Matching in the Auxiliary Samples



Note: Panel (a) displays the forecast labor income for males in the auxiliary samples for whom $B \in \mathcal{B}_0$, i.e., ABC/CARE eligible, and for the synthetic cohort we construct based on the method proposed in this section. 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 $B \in \mathcal{B}_0$ in the auxiliary samples. Panel (b) displays the analogous figure for females. Standard errors are based on the empirical bootstrap distribution. a^* is the oldest age in the experimental sample.

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.²⁴ 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 $\phi_{k,a}^d(\cdot, \cdot)$ in the non-experimental sample straightforward. Assumption A-3 formalizes a strong form of the exogeneity condition.

Assumption A-3 Exogeneity Let $\{1, \dots, \bar{A}\}$ index the periods of a life cycle. For all $a, a' \in \{1, \dots, \bar{A}\}$ and for $d, d' \in \{0, 1\}$,

$$\varepsilon_{k,a}^d \perp\!\!\!\perp \mathbf{X}_{k,a'}^{d'} | \mathbf{B}_k = \mathbf{b} \quad (5)$$

for all \mathbf{b} in the support of \mathbf{B}_k , $k \in \{e, n\}$, where “ $\mathbf{M} \perp\!\!\!\perp \mathbf{N} | \mathbf{Q}$ ” denotes independence of \mathbf{M} and \mathbf{N} given \mathbf{Q} . \square

Below we discuss how this condition can be weakened and unbiased forecasts can still be obtained.

To appreciate the benefit of Assumption A-3, consider the following example. Suppose that years of education is a component of $\mathbf{X}_{k,a'}^{d'}$. The joint distribution of $\varepsilon_{k,a}^d$ and $\mathbf{X}_{k,a'}^{d'}$ can differ substantially across experimental and non-experimental samples. In the experimental sample, years of education are boosted by treatment, which is randomly assigned. In the non-experimental samples, however, there is no experimental variation. 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 $\mathbf{X}_{k,a'}^{d'}$ in the non-experimental sample. Assumption A-3 avoids this problem when making forecasts. Below, we establish that our forecasts based on this assumption are concordant with forecasts from approaches that do not necessarily require Assumption A-3.

²⁴See Liu et al. (2016). Conditions C-1 to C-3 in Appendix C.3 spell out the requirements.

Step 3. Determining Inputs. Analysis of the non-experimental data reveals that the inputs determining labor income under Assumption A-3 are: average PIAT achievement score from ages 5 to 7, completed education, labor income at age 21, and lagged labor income. Appendix Table C.3 shows that these variables predict labor income. Appendix C.3.6 shows that there is common support across datasets. We report tests for endogeneity of these variables in the experimental and auxiliary samples used in this paper in Appendix C.3.7.²⁵ After conditioning on $\mathbf{X}_{k,a}^d$ and \mathbf{B}_k , we do not reject the null hypothesis of exogeneity. Accordingly, we use OLS for making our baseline estimates.

Table 1 displays the treatment effects of the program for these inputs.²⁶ 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, it increases the average PIAT score by almost one third of a standard deviation.²⁷ 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).²⁸

²⁵These tests are based on the assumption that $\varepsilon_{k,a}^d$ for $k \in \{e, n\}$ is characterized by a factor structure. The factors are predicted by measurements of cognitive and non-cognitive skills. We use estimated factors as control functions. We do not reject the null of exogeneity. See Appendix 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 et al. (2010).

²⁶We 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 if compared to women. This is consistent with the results we report below.

²⁷The test is standardized to an in-sample standard deviation of 15 units.

²⁸Table 1 displays age-21 and age-30 labor income because labor income is observed at ages 21 and 30 in

Table 1: Summary of Treatment Effects for Inputs Generating Labor Income ($\mathbf{X}_{k,a}^d$)

Inputs	Females		Males	
	Control Mean	Average Treatment Effect	Control Mean	Average Treatment Effect
PIAT Scores	95.63	4.92	93.46	7.70
High School Graduation	0.51	0.25	0.61	0.07
College Graduation	0.08	0.13	0.12	0.17
Years of Education	11.76	2.14	12.90	0.66
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 are bolded when statistically significant at the 10% level.

Step 4. Testing the Empirical Implications of Assumption A-1. Under Assumption A-3, we build on Heckman et al. (2013) to test for invariance. Condition (3a) of Assumption A-1 combined with Assumptions A-2 and A-3 and the normalization $\mathbb{E}(\varepsilon_{k,a}^d) = 0$ for all $a \in \{1, \dots, \bar{A}\}$, generates the following testable implications:

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

and

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

for $d \in \{0, 1\}$ where $\mathbf{Y}_{n,a}$ is the counterpart of $\mathbf{Y}_{e,a}$ in the non-experimental sample.

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 the non-experimental samples evaluated at $\mathbf{X}_{n,a} = \mathbf{x}^1$ and $\mathbf{X}_{n,a} = \mathbf{x}^0$ respectively:

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

Equations (6a) and (6b) are sufficient conditions for Equation (6c) to hold. We test

the experimental sample. Labor income at age 30 is an input in our methodology only after age 30.

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, the coefficient associated with D , denoted by τ , should be zero in

$$Y_{e,30}^d = \tau \cdot D + \mathbf{B}_e \cdot \boldsymbol{\gamma}_{e,30}^d + \mathbf{X}_{e,30}^d \cdot \boldsymbol{\beta}_{e,30}^d + \varepsilon_{e,a}^d. \quad (6a)$$

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 (6a) 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.²⁹ Panel (b) of Table 2 reports estimates for the remaining coefficients in Equation (6a). Years of education is strongly boosted by ABC/CARE (see Table 1).

Define $K = \mathbf{1}$ ($k = e$) 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)

$$Y_{k,30} = \pi \cdot K + \mathbf{B}_k \cdot \boldsymbol{\gamma}_{k,30} + \mathbf{X}_{k,30} \cdot \boldsymbol{\beta}_{k,30} + \varepsilon_{k,a}. \quad (6b)$$

Panel (c) in Table 2 displays estimates of the parameters of Equation (6b) for labor income at age 30 for males and females. Estimates of π are small and not statistically significant.³⁰ We do not reject the null hypothesis that the technologies are invariant across samples so the data are consistent with invariance.

²⁹Note that after accounting for background variables and the intermediate inputs, 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).

³⁰The averages of labor income in the experimental and non-experimental sample for females and males are \$24,584 and \$40,007, and \$24,098 and \$32,717 (2014 USD), respectively.

Table 2: Testing Invariance in Technologies ($\phi_{k,a}^d(\mathbf{x}, \mathbf{b})$) of Labor Income at Age 30

	Females		Males	
	coefficient	<i>p</i> -value	coefficient	<i>p</i> -value
Panel (a). Invariance Across Treatment Regimes				
<i>D</i> (treatment indicator)	-2,212.806	0.586	231.606	0.969
Panel (b). Precision of Estimated Coefficients of (6a)				
<i>B_e</i> (baseline variables)				
Mother's Education (at birth)	-957.0972	0.387	1,850.201	0.358
<i>X_{e,30}^d</i> (age-30 inputs)				
PIAT (5-7)	5.726	0.975	327.186	0.338
Years of Education (30)	2,356.143	0.006	4,474.721	0.018
Labor Income (21)	0.218	0.320	0.322	0.175
<i>R</i> ²	0.281		0.207	
Observations	52		50	
Sample: Experimental Treatment and Control Groups at Age 30				
Panel (c). Invariance Across Experimental and Non-Experimental Samples				
<i>K</i> (treatment indicator)	-142.631	0.965	1,887.575	0.654
Panel (d). Precision of Estimated Coefficients of Counterpart to (6b) in the Non-experimental Sample				
<i>B_k</i> (baseline variables)				
Mother's Education (at birth)	-229.481	0.631	427.224	0.459
<i>X_{k,30}^d</i> (age-30 inputs)				
PIAT (5-7)	266.1971	0.002	219.220	0.044
Years of Education (30)	4,263.156	0.000	4,434.173	0.000
Labor Income (21)	0.355	0.000	0.685	0.000
<i>R</i> ²	0.221		0.182	
Observations	829		746	
Sample: Experimental Treatment and Control Groups and Non-Experimental Synthetic Cohort at Age 30				

Note for Panel (a) and (b): Estimates of the coefficients in Equation (6a) 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). *B_e* is comprised of baseline variables not affected by treatment (mother's education at birth) and *X_{e,30}^d* is age-30 intermediate inputs. We drop labor income observations above the 95th percentile to avoid precision issues.

Note for Panel (c) and (d): Estimates of the coefficients in Equation (6b) 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 non-experimental sample (*K* = 0 synthetic cohort in the non-experimental sample and *K* = 1 experimental sample). *B_k* is comprised of baseline variables not affected by treatment (mother's education at birth) and *X_{k,30}^d* is age-30 intermediate inputs.

We use analogous procedures to test Condition (3b) of Assumption A-1. First, we test invariance in the distributions of $\varepsilon_{k,a}^d$ 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 non-experimental 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.

Table 3: Testing Invariance in Distributions of the Error Terms ($F_{k,a}^d$) of Labor Income at Age 30

	Females		Males	
Panel (a). Invariance Across Treatment Regimes				
Equality in means	<i>t</i> -stat	<i>p</i> -value	<i>t</i> -stat	<i>p</i> -value
	1.075	0.287	-0.0390	0.969
Equality in distributions	K-S <i>p</i> -value		K-S <i>p</i> -value	
	0.272		0.632	
Sample: Experimental Treatment and Control Groups at Age 30				
Panel (b). Invariance Across Experimental and Non-Experimental Samples				
Equality in means	<i>t</i> -stat	<i>p</i> -value	<i>t</i> -stat	<i>p</i> -value
	0.054	0.957	-0.226	0.822
Equality in distributions	K-S <i>p</i> -value		K-S <i>p</i> -value	
	0.481		0.046	
Sample: Experimental Treatment and Control Groups and Non-Experimental Synthetic Cohort at Age 30				

Note for Panel (a): Tests for equality in distributions of residuals within the experimental sample across treatment regimes at age 30 in labor income by gender. Residuals are the relevant outcome net of mother’s education at birth, average PIAT test from ages 5 to 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 Appendix C.7.1. Tests are a *t*-test of equality in means and the Kolgomorov-Smirnov test.

Note for Panel (b): Tests for equality in distributions of residuals across the experimental and non-experimental 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 test from ages 5 to 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 Appendix C.7.1. Tests are a *t*-test of equality in means and the Kolgomorov-Smirnov test.

With the empirical counterparts of $\varepsilon_{k,a}^d$ in hand, we implement two tests to compare the distributions in Table 3: *t*-tests of mean comparisons and Kolgomorov-Smirnov tests of

equality of distributions. We do not reject equality of treatment and control distributions of $\varepsilon_{k,a}^d$ 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).

Steps 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 (s.e. \$76,634) in 2014 USD. We explore the estimates from alternative forecasting models in Appendix C.6.³¹ When pooling males and females and when separating the samples by gender, the present value gains remain within a range that does not change our inference that the program had substantial lifetime benefits.

Step 6. Validating Forecasts. Invariance across treatment regimes and samples is the essential ingredient for constructing valid forecasts. Figure 2 displays our forecasted labor income profiles. Forecasted and actual labor income are closely aligned in both the treatment and the control regimes.³² Computing the net present value, the internal rate of return and benefit/cost ratios is straightforward once age-by-age forecasts are available.³³

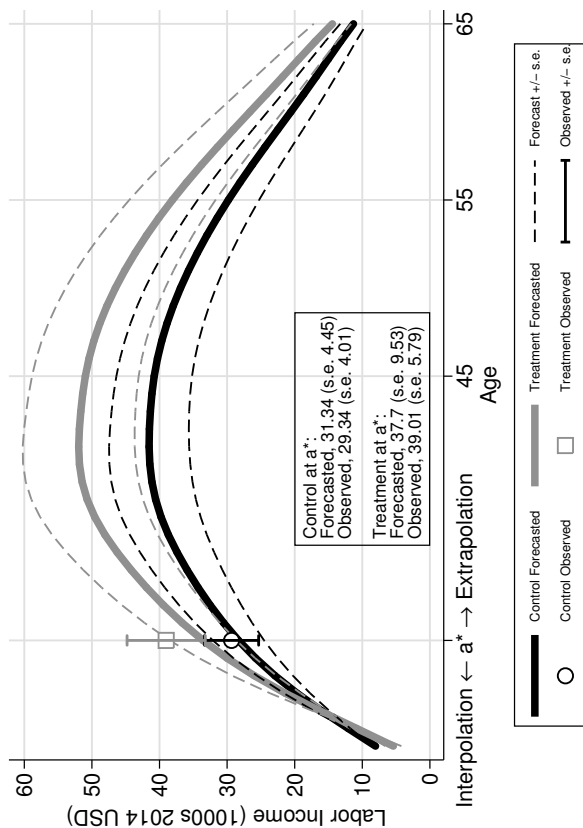
³¹As both a referee and various discussants of our paper have pointed out, our identification and estimation strategies do not impose cross-sectional restrictions. We use different datasets to identify and estimate the $\phi_a(\mathbf{x}, \mathbf{b})$ for each outcome (e.g., labor income, health, crime). Thus, the predictor variables that we are able to use differ across outcomes and we cannot conduct joint estimations.

³²The content in Figure 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 socio-economic groups. The pattern of life-cycle labor income we generate is typical for that of low-skilled workers (Blundell et al., 2015; Gladden and Taber, 2000; Sanders and Taber, 2012; Lagakos et al., 2016).

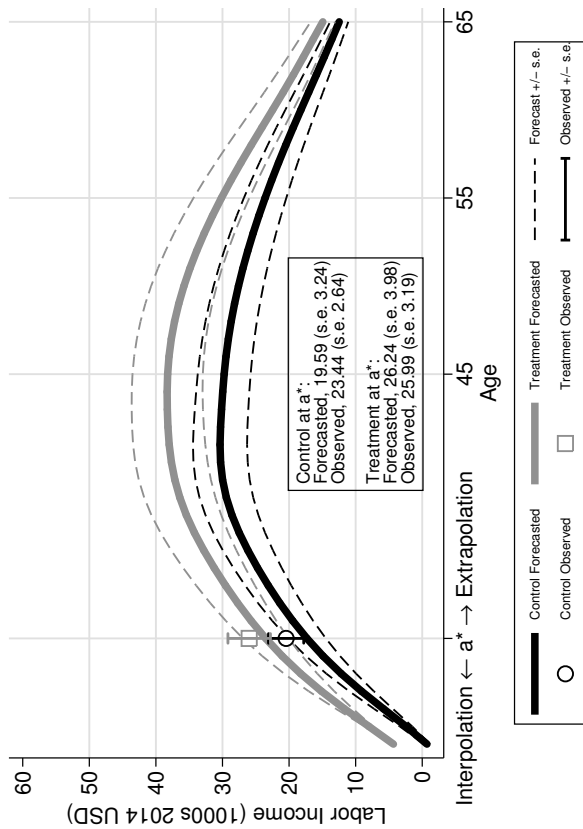
³³Some practical details involved in doing this are in Appendices C.4 and C.5.

Figure 2: Forecasted Labor Income Profiles for ABC/CARE Participants

(a) Males



(b) Females



Note: Panel (a) displays the forecast life-cycle labor income profiles for ABC/CARE males by treatment status, based on the method proposed in Section 3. 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. Panel (b) displays the analogous figure for females. Our forecasts go up to age 67, the age of assumed retirement. Standard errors are the standard deviations of the empirical bootstrap distribution. See Appendix C for a discussion of our choice of predictors and a sensitivity analysis on those predictors. We under-predict 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).

3.2.1 Alternative Forecasting Models for Labor Income

An alternative non-parametric forecasting method compresses the whole forecasting procedure. Under Assumptions A-1 and A-3, we can use matching on baseline variables and *variables affected by treatment* to construct counterparts to the experimental treatment and control groups in the non-experimental sample.³⁴ Matching is a non-parametric 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.³⁵ This is an intuitively appealing non-parametric estimator of life-cycle program treatment effects that is valid under exogeneity (Heckman and Navarro, 2004).

This is a fundamentally different matching procedure than 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 $\mathbf{B} \in \mathcal{B}_0$. 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 (\mathbf{B}) and on variables affected by treatment ($\mathbf{X}_{k,a}^d$) compressing the construction of the synthetic cohort and the estimation of the production functions for out-of-sample forecasts to a single, non-parametric procedure.

Table C.12 shows that there is close agreement between non-parametric 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.³⁶

³⁴Heckman et al. (1998) discuss this procedure.

³⁵See Appendix C.3.5 for details.

³⁶Note that the non-parametric 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.

4 Forecasting Other Life-cycle Benefits

In this section, we adapt the methodology described in Section 3 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.³⁷ Procedures for forecasting the benefits and costs of education are reported in Appendix D.

4.1 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, or 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 Dana Goldman and coauthors (Goldman et al., 2015).³⁸ We forecast health outcomes of program participants from their mid 30s up to their projected age of death.³⁹ Our version of FAM passes a variety of specification tests and accurately forecasts health outcomes and health behaviors.⁴⁰

Our methodology has four steps—extensive details are provided in Appendix G: (i) follow an adapted version of the steps in Section 3 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;⁴¹ (iii) estimate medical

³⁷Appendices E and C.3.8 provide further documentation.

³⁸Appendix 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.

³⁹The simulation starts at the age in which we observe the subjects' age-30 follow-up.

⁴⁰Goldman et al. (2015) present tests of the model assumptions and predictive performance for population aggregate health and health behavior outcomes.

⁴¹QALY is a measure which reweighs a year of life according to its quality given the burden of disease (Dolan, 1997; Shaw et al., 2005).

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.⁴²

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 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 .⁴³

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 the MEPS and the 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 to 2010. The MEPS estimation captures costs during 2008 to 2010. 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 QALY based on a widely-used health-related Quality-of-Life measure available in MEPS.⁴⁴ 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

⁴²As part of step (i), we impute some of the variables used to initialize the FAM models (see Appendix G.1.6.1).

⁴³See Tables G.1 to G.3 for a summary. Our forecasts are based on two-year lags, due to 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 in 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.

⁴⁴HRQoL measure EQ-5D.

and the control groups.⁴⁵

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 30.⁴⁶ 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.

4.2 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; (ii) we use micro-data specific to the states in which participants grew up, as well as other national datasets, to forecast criminal activity from the mid 30s to 50. We forecast using methods standard in the criminology literature.⁴⁷ See [Appendix E](#) for a complete discussion of our crime forecasts.

⁴⁵Our 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 [Appendix H](#).

⁴⁶See [Appendix G.2.4](#).

⁴⁷[Cohen and Bowles \(2010\)](#) and [McCollister et al. \(2010\)](#).

4.3 Parental Labor Income

ABC/CARE offers childcare to the parents of treated children for more than nine hours a day for five 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 Appendix A). The childcare component generates substantial treatment effects on maternal labor force participation and parental labor income.⁴⁸ 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.^{49,50} 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 3.

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 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 comparing children who have siblings younger than 5 years old to children who have siblings age 5 years or older.⁵¹

⁴⁸There 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 Appendix D.

⁴⁹The ages at which parental labor income is observed are 0, 1.5, 3.5, 4.5, 8, 12, 15, and 21. At age 21 the mothers of the ABC/CARE subjects were, on average, 41 years old.

⁵⁰We linearly interpolate parental labor income for ages for which we do not have observations between 0 and 21.

⁵¹These patterns persist when splitting the ABC/CARE sample by gender, but the estimates are not precise because the samples become too small. See Appendix C.3.8.

4.4 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.⁵² Appendix B discusses program costs in detail.

5 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.⁵³ 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.⁵⁴ 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 eliminating the benefits from any one of the four main components that we monetize.

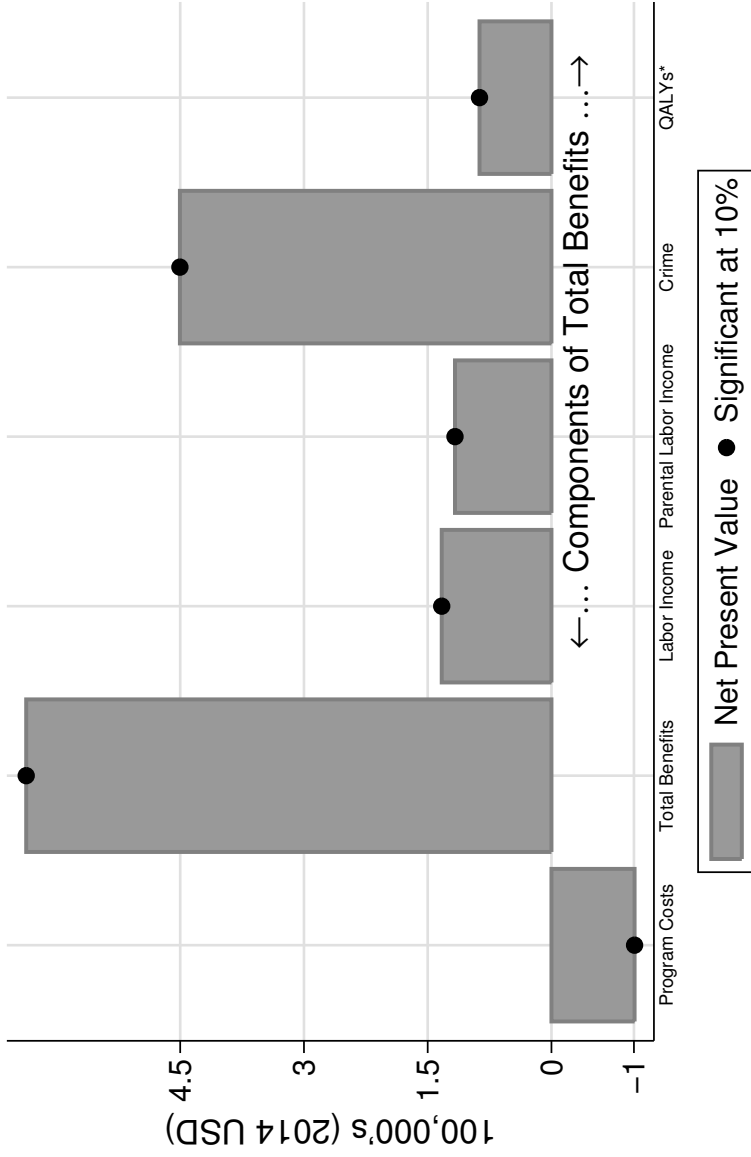
Pooling males and females, the program is socially efficient: the internal rate of return

⁵²Our 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 Appendix 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 costs 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 almost an identical amount (see Appendix B).

⁵³Using discount rates of 0%, 3%, and 7%, the estimates for the benefit/cost ratios are 17.40 (s.e. 5.90), 7.33 (s.e. 1.84), and 2.91 (s.e. 0.59), respectively. We report estimates for discount rates between 0% and 15% in Appendix H.1.

⁵⁴See, e.g., Fox Business News (2014) and Whitehurst (2014).

Figure 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



Per-annum Rate of Return: Males and Females 13.7% (s.e. 3%).
 Benefit-cost Ratio: Males and Females 7.3 (s.e. 1.8).

Note: 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 from age 21 to the retirement of program participants (assumed to be at age 67). Parental labor income: total parental labor income of the parents of the participants from when the participants were ages 1.5 to 21. Crime: the total cost of crime (judicial and victimization costs). To simplify the display, the following components are not shown in the figure: (i) cost of alternative preschool paid by the control-group children’s parents; (ii) the social welfare costs of transfer income from the government; (iii) disability benefits and social security claims; (iv) costs of increased individual and maternal education (including special education and grade retention); (v) total medical public and private costs. Inference is based on non-parametric, one-sided *p*-values from the empirical bootstrap distribution. Dots (●) indicate point estimates significant at the 10% level. *QALYs refers to the quality-adjusted life years. Any gain corresponds to better health conditions until forecasted death, with \$150,000 (2014 USD) as the base value for a year of life.

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.⁵⁵

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 non-response, 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.⁵⁶ Although the internal rate of return remains relatively high when using 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 only considering 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.⁵⁷ We assume a marginal tax rate of 50%.⁵⁸ Our estimates are robust to dropping it to 0% or

⁵⁵We obtain the reported standard errors by bootstrapping *all* steps of our empirical procedure, including variable selection, imputation, model selection steps, and forecast error (see Appendix C.7).

⁵⁶See Appendix C for a detailed discussion.

⁵⁷When the transaction between the government and an individual is a direct transfer, we consider 0.5 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).

⁵⁸Feldstein (1999) estimates that the deadweight loss caused by increasing existing tax rates (marginal deadweight loss) may exceed two dollars per dollar of revenue generated. We use a more conservative value (0.5 dollars per dollar of revenue generated). In Tables 4, 5, and in Appendix H.2, we explore the robustness of this choice of the welfare cost and find little sensitivity.

Table 4: Sensitivity Analysis for Benefit/Cost Ratios

	Pooled		Males		Females	
	No IPW	and No Controls	No IPW	and No Controls	No IPW	and No Controls
Baseline	7.33 (s.e. 1.84)		10.19 (s.e. 2.93)		2.61 (s.e. 0.73)	
<i>Baseline: IPW and Controls, Life-span up to predicted death, Treatment vs. Next Best, 50% Marginal tax 50% (deadweight loss), Discount rate 3%, Parental income 0 to 21 (child's age), Labor Income predicted from 21 to 65, All crimes (full costs), Value of life 150,000.</i>						
Specification	7.31 (1.81)	7.99 (2.18)	9.80 (2.69)	8.83 (2.72)	2.57 (0.72)	2.82 (0.68)
Forecast Span	to Age 21 1.52 (0.36)	to Age 30 3.19 (1.04)	to Age 21 2.23 (0.61)	to Age 30 3.84 (1.60)	to Age 21 1.46 (0.36)	to Age 30 1.81 (0.50)
Counter-factuals	vs. Stay at Home 5.44 (1.86)	vs. Alt. Presch. 9.63 (3.10)	vs. Stay at Home 3.30 (2.95)	vs. Alt. Presch. 11.46 (3.16)	vs. Stay at Home 5.79 (1.37)	vs. Alt. Presch. 2.28 (0.76)
Deadweight-loss	0% 11.01 (2.79)	100% 5.50 (1.37)	0% 15.38 (4.35)	100% 7.59 (2.23)	0% 3.83 (1.04)	100% 2.01 (0.59)
Discount Rate	0% 17.40 (5.90)	7% 2.91 (0.59)	0% 25.45 (10.42)	7% 3.78 (0.79)	0% 5.06 (2.82)	7% 1.49 (0.32)
Parental Income	Mincer Life-cycle 7.63 (1.84)	Life-cycle Prediction 7.73 (1.92)	Mincer Life-cycle 10.46 (2.94)	Life-cycle Prediction 10.63 (2.95)	Mincer Life-cycle 2.98 (0.76)	Life-cycle Prediction 3.12 (0.85)
Labor Income	.5% Annual Decay 7.01 (1.80)	.5% Annual Growth 7.66 (1.90)	.5% Annual Decay 9.58 (2.66)	.5% Annual Growth 10.79 (3.24)	.5% Annual Decay 2.51 (0.70)	.5% Annual Growth 2.71 (0.75)
Crime	Drop Major Crimes 4.24 (1.10)	Halve Costs 5.18 (1.22)	Drop Major Crimes 7.41 (3.43)	Halve Costs 7.12 (2.41)	Drop Major Crimes 2.61 (0.67)	Halve Costs 2.47 (0.66)
Health (QALYs)	Drop All 6.48 (1.79)	Double Value of Life 8.19 (2.13)	Drop All 9.14 (2.73)	Double Value of Life 11.23 (3.40)	Drop All 2.20 (0.69)	Double Value of Life 3.03 (1.04)

Note: This table displays sensitivity analyses of our baseline benefit/cost ratio calculation to the perturbations indexed in the different rows. The characteristics of the *baseline* calculation are in the table header. IPW: adjusts for attrition and item non-response (see Appendix C.2 for details). Control variables: Apgar scores at ages 1 and 5 and a high-risk index (see Appendix 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 these ages, respectively. Counterfactuals: we consider treatment vs. next best (baseline), treatment vs. stay at home, and treatment vs. alternative preschools (see Appendix F for a discussion). Deadweight loss is the loss implied by any public expenditure (0% is no loss and 100% is one dollar loss per dollar spent). Discount rate: rate to discount benefits to child's age 0 (in all calculations). Parental labor income: see Appendix C.3.8 for details on the two alternative forecasts (Mincer and Life-cycle). Labor Income: 0.5 annual growth (decay) is an annual wage growth (decay) due to cohort effects. Crime: major crimes are rape and murder; half costs takes half of victimization and judiciary costs. Health (QALYs): "drop all" sets the value of life equal to zero. Standard errors obtained from the empirical bootstrap distribution are in parentheses. Bolded *p*-values are significant at 10% using one-sided tests. For details on the null hypothesis see Table H.1.

Table 5: Sensitivity Analysis for Internal Rate of Return

	Pooled			Males			Females		
Baseline	13.7% (s.e. 3.3%)	14.7% (s.e. 4.2%)	10.1% (s.e. 6.0%)						
<i>Baseline: IPW and Controls, Life-span up to predicted death, Treatment vs. Next Best, 50% Marginal tax 50% (deadweight loss), Discount rate 3%, Parental income 0 to 21 (child's age), Labor Income predicted from 21 to 65, All crimes (full costs), Value of life 150,000.</i>									
Specification	No IPW 13.2% (2.9%)	and No Controls 14.0% (3.1%)	No IPW 13.9% (3.7%)	and No Controls 13.0% (4.3%)	No IPW 9.6% (6.0%)	and No Controls 10.0% (4.9%)			
Forecast Span	to Age 21 8.8% (4.5%)	to Age 30 12.0% (3.4%)	to Age 21 11.8% (4.8%)	to Age 30 12.8% (4.7%)	to Age 21 10.7% (5.8%)	to Age 30 11.7% (5.2%)			
Counterfactuals	vs. Stay at Home 9.4% (4.2%)	vs. Alt. Presch. 15.6% (4.3%)	vs. Stay at Home 6.0% (3.6%)	vs. Alt. Presch. 15.8% (5.0%)	vs. Stay at Home 13.4% (5.7%)	vs. Alt. Presch. 8.8% (7.0%)			
Deadweight-loss	0% 18.3% (4.7%)	100% 11.2% (3.1%)	0% 19.4% (6.2%)	100% 12.1% (3.9%)	0% 17.7% (12.4%)	100% 7.1% (4.2%)			
Parental Income	Mincer Life-cycle 15.2% (4.0%)	Life-cycle Prediction 14.5% (6.4%)	Mincer Life-cycle 16.0% (5.1%)	Life-cycle Prediction 14.5% (6.4%)	Mincer Life-cycle 13.3% (8.2%)	Life-cycle Prediction 12.3% (9.9%)			
Labor Income	.5% Annual Decay 13.5% (3.4%)	.5% Annual Growth 13.8% (3.2%)	.5% Annual Decay 14.5% (4.3%)	.5% Annual Growth 14.8% (4.1%)	.5% Annual Decay 9.9% (6.0%)	.5% Annual Growth 10.3% (6.0%)			
Crime	Drop Major Crimes 10.7% (4.4%)	Halve Costs 11.6% (3.8%)	Drop Major Crimes 12.0% (5.3%)	Halve Costs 11.9% (4.9%)	Drop Major Crimes 10.1% (6.0%)	Halve Costs 9.9% (6.0%)			
Health (QALYs)	Drop All 12.8% (4.6%)	Double Value of Life 13.5% (3.6%)	Drop All 13.5% (5.6%)	Double Value of Life 14.4% (4.6%)	Drop All 8.8% (6.4%)	Double Value of Life 9.3% (6.1%)			

Note: This table displays sensitivity analyses of our baseline internal rate of return calculation to the perturbations indexed in the different rows. The characteristics of the *baseline* calculation are in the table header. IPW: adjusts for attrition and item non-response (see Appendix C.2 for details). Control variables: Apgar scores at ages 1 and 5 and a high-risk index (see Appendix 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 these ages, respectively. Counterfactuals: we consider treatment vs. next best (baseline), treatment vs. stay at home, and treatment vs. alternative preschools (see Appendix F for a discussion). Deadweight loss is the loss implied by any public expenditure (0% is no loss and 100% is one dollar loss per dollar spent). Parental labor income: see Appendix C.3.8 for details on the two alternative forecasts (Mincer and Life-cycle). Labor Income: 0.5 annual growth is an annual wage growth due to cohort effects; only benefit assumes labor income is the only benefit of the program. Crime: major crimes are rape and murder; half costs takes half of victimization and judiciary costs. Health (QALYs): “drop all” sets the value of life equal to zero. Bolded *p*-values are significant at 10% using one-sided tests. For details on the null hypothesis see Table H.1.

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.⁵⁹ 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 Section 4.3). Our baseline estimates rely solely on parental labor income when participant children are ages 0 to 21. Alternative approaches considering the gain for the parents through age 67 generate an additional increase in the gain due to parental labor income (see parental labor income columns).⁶⁰

Individuals in ABC/CARE could experience positive cohort effects that might (i) make them more productive and therefore experience wage growth; (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 and decay rates in labor income between -0.5% and 0.5% . This is consistent with the range of values in [Lagakos et al. \(2016\)](#).

We also examine the sensitivity of our estimates to (i) dropping the most costly crimes such as murders and rapes,⁶¹ 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 Appendix E.3). Our

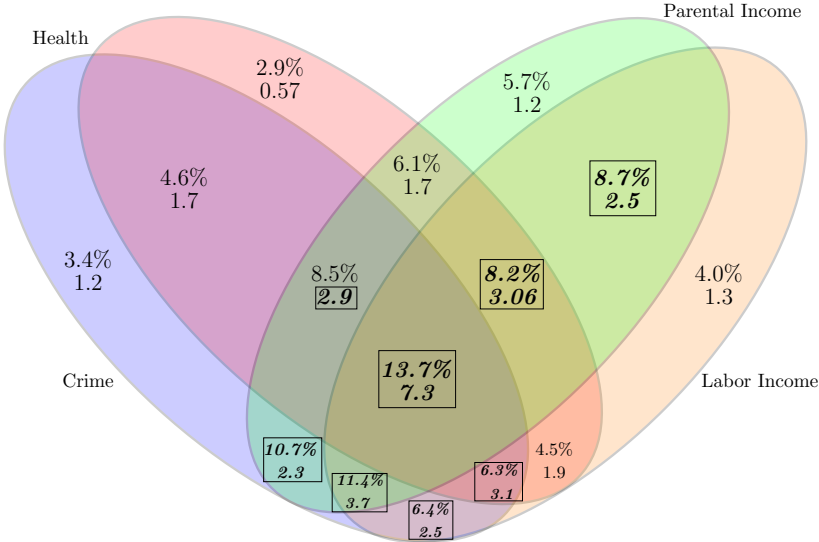
⁵⁹There is no inconsistency between the weak female treatment effects on wages at age 30 and the high lifetime net present value treatment effect for earnings given life-cycle wage growth attributable to enhanced inputs (education, PIAT scores, etc.).

⁶⁰If 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 of Figure 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.

⁶¹Two individuals in the treatment group were convicted of rape and one individual in the control group was convicted of murder.

benefit/cost estimates are robust to these adjustments, even though crime is a major component of it. We also examine the sensitivity with respect to our main health component: quality-adjusted life years. 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: Benefit/Cost Ratio and Internal Rate of Return when Accounting for Different Combinations of the Main Benefits



Note: This figure presents all possible combinations of accounting for the benefits from the four major categories in our analysis. The non-overlapping 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 Figure 3. Health combines QALYs (quality-adjusted life years) and health expenditure. Inference is based on non-parametric, one-sided p -values from the empirical bootstrap distribution. We put boxes around point estimates that are statistically significant at the 10% level.

Figure 4 summarizes the results from our extensive sensitivity analyses reported in Table H.1 of Appendix H, 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 eliminating 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.

6 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 and Walters \(2016\)](#), which in turn is based on estimates taken from [Chetty et al. \(2011\)](#).⁶² Although widely emulated, this approach offers an imprecise approximation 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 1.84.⁶³ Their analysis proceeds in three steps: (i) calculate program treatment effects on cognitive test scores measured around age 5;⁶⁴ (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 [Chetty et al. \(2011\)](#);^{65,66,67} and (iii) calculate the benefit/cost ratio based on this gain and their own calculations of the program's cost.^{68,69}

To analyze how our estimates compare to those based on this method, we present a series of estimates in the fourth column of [Table 6](#). For purposes of comparison, the fifth column of [Table 6](#) shows the analogous estimates based on our samples and forecasts.

⁶²Examples of application of this approach include [Attanasio et al. \(2011\)](#), [Behrman et al. \(2011\)](#), and [Lafortune et al. \(2018\)](#).

⁶³HSIS is a one-year-long randomized evaluation of Head Start ([Puma et al., 2010](#)).

⁶⁴They use an index based on the Peabody Picture Vocabulary and Woodcock Johnson III Tests.

⁶⁵The [Chetty et al. \(2011\)](#) return is based on Stanford Achievement Tests.

⁶⁶For this comparison exercise, we interpret the earnings estimated in [Chetty et al. \(2011\)](#) to be equivalent to labor income.

⁶⁷Calculations 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.

⁶⁸Their 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. \(2011\)](#).

⁶⁹All 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).

Table 6: Examining the Validity of Recent *Ad Hoc* Methods for Forecasting in Light of the Analysis of This Paper

(1)	(2)	(3)	(4)	(5)
Age	NPV Source	Component	Kline and Walters (2016) Method	Authors' Method
27	Chetty et al. (2011)	Labor income	0.58 (s.e. 0.28)	1.09 (s.e. 0.04)
	ABC/CARE-calculated	Labor income	0.09 (s.e. 0.04)	
34	ABC/CARE-calculated	Labor income	0.37 (s.e. 0.04)	0.15 (s.e. 0.05)
	ABC/CARE-calculated	All	1.21 (s.e. 0.05)	3.20 (s.e. 1.04)
Life-cycle	ABC/CARE-calculated	Labor income	1.56 (s.e. 0.08)	1.55 (s.e. 0.76)
	ABC/CARE-calculated	All	3.80 (s.e. 0.29)	7.33 (s.e. 1.84)

Note: This table displays benefit/cost ratios based on the methodology in Kline and Walters (2016) and based on our own methodology. Age: age at which we stop calculating the net present value. NPV Source: source where we obtain the net present value. Component: item used to compute net present value (all refers to the net present value of all the components). Kline and Walters (2016) Method: estimate based on these authors' methodology. Authors' Method: estimates based on our methodology. Standard errors are based on the empirical bootstrap distribution.

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 27.⁷⁰ In this exercise, the standard errors account for variation in the return because we calculate the return in every bootstrapped re-sample. 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 are: (i) increase the age range over which we calculate the net-present value of labor income; or (ii) consider the value of all 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 bene-

⁷⁰This allows us to compute our own “return to IQ” and impute it to the treatment-group individuals.

fit/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 both the sampling error in the experimental and auxiliary samples, the forecast error due to the interpolation and extrapolation, and the sensitivity of the estimates to externally specified parameters.

7 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 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 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 conclude with a cautionary note. The program we study was targeted to a disadvantaged, predominately African-American population in a university town in North Carolina. Blind generalization of our findings to other populations is clearly not warranted. In particular, there is no support in this study for universal application of ABC/CARE across all socio-economic 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 applicability to disadvantaged populations. Our study indicates what is possible and that the possibilities are substantial.

References

- Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives* 25(3), 153–172.
- Attanasio, O., A. Kugler, and C. Meghir (2011). Subsidizing vocational training for disadvantaged youth in Colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics* 3(3), 188–220.
- Barnett, W. S. and L. N. Masse (2002). A benefit-cost analysis of the Abecedarian Early Childhood Intervention. Technical report, National Institute for Early Education Research, New Brunswick, NJ.
- Barnett, W. S. and L. N. Masse (2007, February). Comparative benefit-cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review* 26(1), 113–125.
- Behrman, J. R., S. W. Parker, and P. E. Todd (2011). Do conditional cash transfers for schooling generate lasting benefits? a five-year followup of PROGRESA/Oportunidades. *Journal of Human Resources* 46(1), 93–122.
- Belfield, C. R., M. Nores, W. S. Barnett, and L. J. Schweinhart (2006). The High/Scope Perry Preschool Program: Cost-benefit analysis using data from the age-40 followup. *Journal of Human Resources* 41(1), 162–190.
- Blundell, R., M. Graber, and M. Mogstad (2015, July). Labor income dynamics and the insurance from taxes, transfers, and the family. *Journal of Public Economics* 127, 58–73.
- Campbell, F. A., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. P. Pungello, and Y. Pan (2014). Early childhood investments substantially boost adult health. *Science* 343(6178), 1478–1485.
- Campbell, F. A., E. P. Pungello, S. Miller-Johnson, M. Burchinal, and C. T. Ramey (2001, March). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology* 37(2), 231–242.
- Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science* 6(1), 42–57.

- Campbell, F. A., B. Wasik, E. Pungello, M. Burchinal, O. Barbarin, K. Kainz, J. Sparling, and C. Ramey (2008). Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly* 23(4), 452–466.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011, November). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics* 126(4), 1593–1660.
- Clarke, S. H. and F. A. Campbell (1998). Can intervention early prevent crime later? The Abecedarian Project compared with other programs. *Early Childhood Research Quarterly* 13(2), 319–343.
- Cohen, M. A. and R. Bowles (2010). Estimating costs of crime. In A. R. Piquero and D. Weisburd (Eds.), *Handbook of Quantitative Criminology*, Chapter 8, pp. 143–161. New York: Springer.
- Collins, A., B. D. Goodson, J. Lualaba, A. R. Fountain, A. Checkoway, and Abt Associates Inc. (2010, June). Evaluation of child care subsidy strategies: Massachusetts Family Child Care study. Technical Report OPRE 2011-1, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.
- Cunha, F. and J. J. Heckman (2008, Fall). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43(4), 738–782.
- Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.
- Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78(3), 883–931.
- Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care* 35(11), 1095–1108.
- Duncan, G. J. and K. Magnuson (2013). Investing in preschool programs. *Journal of Economic Perspectives* 27(2), 109–132.
- Educare (2014). A national research agenda for early education. Technical report, Educare Learning Network Research & Evaluation Committee, Chicago, IL.
- Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2016). Early childhood education. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 2, Chapter 4, pp. 235–297. Chicago: University of Chicago Press.
- Feldstein, M. (1999, November). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics* 81(4), 674–680.

- 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-grade-school/?#sp=show-clips>.
- García, J. L. and J. J. Heckman (2016). How would a national implementation of early childhood interventions narrow the intra-black and black-white outcome gaps? University of Chicago, Department of Economics.
- García, J. L., J. J. Heckman, and A. L. Ziff (2018). Gender differences in the effects of early childhood education. *European Economics Review*, *Forthcoming*. Unpublished.
- Gladden, T. and C. Taber (2000). Wage progression among less skilled workers. In D. E. Card and R. M. Blank (Eds.), *Finding Jobs: Work and Welfare Reform*, Chapter 4, pp. 160–192. New York: Russell Sage Foundation.
- Goldman, D. P., D. Lakdawalla, P.-C. Michaud, C. Eibner, Y. Zheng, A. Gailey, I. Vaynman, J. Sullivan, B. Tysinger, and D. Ermini Leaf (2015). The Future Elderly Model: Technical documentation. Technical report, University of Southern California.
- Haavelmo, T. (1943, January). The statistical implications of a system of simultaneous equations. *Econometrica* 11(1), 1–12.
- Havnes, T. and M. Mogstad (2011, May). No child left behind: Subsidized child care and children’s long-run outcomes. *American Economic Journal: Economic Policy* 3(2), 97–129.
- Healthy Child Manitoba (2015, April). Starting early, starting strong: A guide for play-based early learning in Manitoba: Birth to six. Technical report, Healthy Child Manitoba, Winnipeg, Manitoba.
- Heckman, J. J. (1992). Randomization and social policy evaluation. In C. F. Manski and I. Garfinkel (Eds.), *Evaluating Welfare and Training Programs*, Chapter 5, pp. 201–230. Cambridge, MA: Harvard University Press.
- Heckman, J. J., N. Hohmann, J. Smith, and M. Khoo (2000, May). Substitution and dropout bias in social experiments: A study of an influential social experiment. *Quarterly Journal of Economics* 115(2), 651–694.
- Heckman, J. J., H. Ichimura, J. Smith, and P. E. Todd (1998, September). Characterizing selection bias using experimental data. *Econometrica* 66(5), 1017–1098.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010a, July). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics* 1(1), 1–46.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b, February). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics* 94(1–2), 114–128.

- Heckman, J. J. and S. Navarro (2004, February). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics* 86(1), 30–57.
- Heckman, J. J. and R. Pinto (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews* 34(1–2), 6–31.
- Heckman, J. J., R. Pinto, and P. A. Savelyev (2013, October). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 2052–2086.
- Hurwicz, L. (1962). On the structural form of interdependent systems. In E. Nagel, P. Suppes, and A. Tarski (Eds.), *Logic, Methodology and Philosophy of Science*, pp. 232–239. Stanford University Press.
- Jensen, B. and M. Nielsen (2016). Abecedarian programme, within an innovative implementation framework (APIIF). A pilot study. Website, [http://pure.au.dk/portal/en/projects/abecedarian-programme-within-an-innovative-implementation-framework-apiif-a-pilot-study\(8d4da9a9-f2ff-44db-87cc-1278d1e7006c\).html](http://pure.au.dk/portal/en/projects/abecedarian-programme-within-an-innovative-implementation-framework-apiif-a-pilot-study(8d4da9a9-f2ff-44db-87cc-1278d1e7006c).html) (Accessed 8/1/2016).
- Kline, P. and C. Walters (2016). Evaluating public programs with close substitutes: The case of Head Start. *Quarterly Journal of Economics* 131(4), 1795–1848.
- Lafortune, J., J. Rothstein, and D. W. Schanzenbach (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics* 10(2), 1–26.
- Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2016). Life-cycle wage growth across countries. Forthcoming, *Journal of Political Economy*.
- Liu, L., H. R. Moon, and F. Schorfheide (2016, December). Forecasting with dynamic panel data models. Technical report, USC-INET Research Paper No. 17-02.
- McCollister, K. E., M. T. French, and H. 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.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). New York: Cambridge University Press.
- Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* 8(4), 431–440.
- Puma, M., S. Bell, R. Cook, and C. Heid (2010). Head Start Impact Study: Final report. Technical report, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.

- Ramey, C. T., D. M. Bryant, J. J. Sparling, and B. H. Wasik (1985). Project CARE: A comparison of two early intervention strategies to prevent retarded development. *Topics in Early Childhood Special Education* 5(2), 12–25.
- Ramey, C. T., A. M. Collier, J. J. Sparling, F. A. Loda, F. A. Campbell, D. A. Ingram, and N. W. Finkelstein (1976). The Carolina Abecedarian Project: A longitudinal and multidisciplinary approach to the prevention of developmental retardation. In T. Tjossem (Ed.), *Intervention Strategies for High-Risk Infants and Young Children*, pp. 629–655. Baltimore, MD: University Park Press.
- Ramey, C. T., J. J. Sparling, and S. L. Ramey (2012). *Abecedarian: The Ideas, the Approach, and the Findings* (1 ed.). Los Altos, CA: Sociometrics Corporation.
- Ramey, S. L., C. T. Ramey, and R. G. Lanzi (2014). Interventions for students from impoverished environments. In J. T. Mascolo, V. C. Alfonso, and D. P. Flanagan (Eds.), *Essentials of Planning, Selecting and Tailoring Interventions for Unique Learners*, pp. 415–48. Hoboken, NJ: John Wiley & Sons, Inc.
- Ridder, G. and R. Moffitt (2007). The econometrics of data combination. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 5469–5547. Amsterdam: Elsevier.
- Sanders, C. and C. Taber (2012, September). Life-cycle wage growth and heterogeneous human capital. *Annual Review of Economics* 4, 399–425.
- Schneider, B. and S.-K. McDonald (Eds.) (2007). *Scale-Up in Education*, Volume 1: Ideas in Principle. Lanham, Maryland: Rowman & Littlefield Publishers.
- Scull, J., J. Hattie, J. Page, J. Sparling, C. Tayler, A. King, V. Nossar, A. Piers-Blundell, and C. Ramey (2015). Building a bridge into preschool in remote Northern Territory communities.
- Shaw, J. W., J. A. Johnson, and S. J. Coons (2005). Us valuation of the EQ-5D health states: Development and testing of the D1 valuation model. *Medical Care* 43(3), 203–220.
- Sparling, J. (2010). Highlights of research findings from the Abecedarian studies. Technical report, Teaching Strategies, Inc., Center on Health and Education, Georgetown University, and FPG Child Development Institute, University of North Carolina at Chapel Hill, Bethesda, MD.
- Sparling, J. J. (1974). Synthesizing educational objectives for infant curricula. Annual Meeting of the American Educational Research Association.
- Spiker, D., C. W. Haynes, and R. T. Gross (Eds.) (1997). *Helping Low Birth Weight, Premature Babies: The Infant Health and Development Program*. Redwood City, CA: Stanford University Press.

Wasik, B. H., C. Ramey, D. M. Bryant, and J. J. Sparling (1990, December). A longitudinal study of two early intervention strategies: Project CARE. *Child Development* 61(6), 1682–1696.

Whitehurst, G. J. (2014, April). Testimony given to the Health, Education, Labor, and Pensions Committee of the U.S. Senate.

Yazejian, N. and D. M. Bryant (2012). Educare implementation study findings. Technical report, Frank Porter Graham Child Development Institute, Chapel Hill, NC.