

Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program*

Jorge Luis García

Department of Economics
The University of Chicago

James J. Heckman

American Bar Foundation
Department of Economics
The University of Chicago

Duncan Ermini Leaf

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

María José Prados

Dornsife Center for
Economic and Social Research
University of Southern California

First Draft: January 5, 2016

This Draft: May 26, 2017

*This research was supported in part by current grants from the Robert Wood Johnson Foundation's Policies for Action program, NICHD R37HD065072 and the American Bar Foundation and by previous grants from the Buffett Early Childhood Fund, the Pritzker Children's Initiative, NICHD R01HD054702 and NIA R01AG042390. 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. Collaboration with Andrés Hojman, Yu Kyung Koh, Sylvi Kuperman, Stefano Mosso, Rodrigo Pinto, Joshua Shea, Jake Torcasso, and Anna Ziff on related work has strengthened the analysis in this paper. Collaboration with Bryan Tysinger of the Leonard D. Schaeffer Center for Health Policy and Economics at the University of Southern California on adapting the Future America Model is gratefully acknowledged. For helpful comments on various versions of the paper, we thank Stéphane Bonhomme, Flávio Cunha, Steven Durlauf, David Figlio, Dana Goldman, Ganesh Karapakula, Magne Mogstad, Sidharth Moktan, Tanya Rajan, Azeem Shaikh, Jeffrey Smith, Chris Taber, Matthew Tauzer, 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. For information on the implementation of the Carolina Abecedarian Project and the Carolina Approach to Responsive Education and assistance in data acquisition, we thank Peg Burchinal, Carrie Bynum, Frances Campbell, and Elizabeth Gunn. 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 any of the authors, 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 the experimentally evaluated life-cycle benefits of a widely implemented early childhood program targeting disadvantaged families. We join experimental data with non-experimental data using economic models to forecast its life-cycle benefits. Our baseline estimate of the internal rate of return (benefit/cost ratio) is 13.7% (7.3). We conduct extensive sensitivity analyses to account for model estimation error, forecasting error, and judgments made about the empirical magnitudes of non-market benefits. We examine the performance of widely used, *ad hoc* estimates of long-term benefit/cost ratios based on short-term measures of childhood test scores and find them wanting.

Keywords: Childcare, early childhood education, life-cycle benefits, long-term forecasts, rates of return.

JEL codes: J13, I28, C93

Jorge Luis García
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Phone: 773-449-0744
Email: jorgelgarcia@uchicago.edu

James J. Heckman
Department of Economics
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

This paper estimates the life-cycle benefits of an influential early childhood program targeted to disadvantaged children that is currently being replicated around the world.¹ It has substantial impacts on the lives of participant children and their mothers. Monetizing benefits and costs across multiple domains, we estimate a baseline internal rate of return of 13.7% per annum and a baseline benefit/cost ratio of 7.3. 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 the benefit/cost ratio (1.52, 17.40). Consistent with the previous literature, there are substantial differences in economic returns that favor males.²

Our analysis contributes to a growing literature on the value of early-life programs for disadvantaged children.³ The full array of life cycle benefits and not just short-term batteries of treatment effects are relevant for life cycle policy analysis. However, long-term evidence on the effectiveness of these programs is limited given the short-term follow-up for most studies.⁴ For want of follow-up data, many studies of early childhood programs report treatment effects for a few outcomes collected at early ages after program completion, e.g.

¹Programs 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: IHDP in eight different cities around the U.S. (Spiker et al., 1997); Early Head Start and Head Start in the U.S. (Schneider and McDonald, 2007); John’s Hopkins Cerebral Palsy Study in the U.S. (Sparling, 2010); Classroom Literacy Interventions and Outcomes (CLIO) study in the U.S. (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 ABC/CARE (Educare, 2014; Yajejian and Bryant, 2012). Appendix A.7 lists around 25 Educare programs, all of which implement curricula based on ABC/CARE. In this appendix, we also list the precise similarities across these programs. Our estimates are likely lower bounds for the returns of Educare. Evidence from ABC/CARE is highly relevant for contemporary policy discussions because their main components are present in a variety of current interventions. About 19% of all African-American children would be eligible for ABC/CARE today and 43% of African-American children were eligible at its inception.

²García et al. (2017).

³See, e.g., Currie (2011) and Elango et al. (2016).

⁴The major source of evidence is from the Perry Preschool Program (see Schweinhart et al., 2005 and Heckman et al., 2010a,b), the Carolina Abecedarian Project (ABC) and the Carolina Approach to Responsive Education (CARE) (Ramey et al., 2000, 2012), and the Infant Health and Development Program (IHDP) (Gross et al., 1997; Duncan and Sojourner, 2013). IHDP was inspired by ABC/CARE (Gross et al., 1997).

IQ scores or school readiness measures.⁵ Other studies, such as [Kline and Walters \(2016\)](#), use early measures fit on auxiliary data to project life-cycle estimates of labor income. We show how misleading this practice can be. We analyze labor income as well as a number of non-market benefits. Because of the plausible ranges of values for the monetary value of non-market benefits found in the literature, we conduct an extensive sensitivity analysis for this component of our estimates. We offer our analysis as a template for conducting analyses of the long-term benefits of social programs evaluated by random assignment but with less than full lifetime follow-up.

We analyze the costs and benefits from two virtually identical early childhood programs evaluated by randomized trials conducted in North Carolina: the Carolina Abecedarian Project (ABC) and the Carolina Approach to Responsive Education (CARE)—henceforth ABC/CARE. Both programs were launched in the 1970s and have follow-ups through the mid 30s. The programs start early (at 8 weeks of life) and engaged participants until age 5. [García et al. \(2017\)](#) analyze the treatment effects of the program on a variety of outcomes up to the mid-30's (e.g. participation in crime, labor income, IQ, schooling, and increased maternal education and labor income arising from the implied childcare subsidy). This paper forecasts these benefits over the life cycle to estimate rates of return and benefit/cost ratios.⁶

Analyzing the life-cycle benefits of programs with a diverse array of outcomes across multiple domains and periods of life is both challenging and rewarding. Doing so highlights the numerous ways through which early childhood programs enhance the lives of the mothers and the adult capabilities of the subjects. We use a variety of measures to characterize program benefits. Instead of reporting individual treatment effects or categories of treatment effects, our benefit/cost analyses account for all measured aspects of these programs, including the welfare costs of taxes used to publicly finance them. We report the sensitivity of our

⁵See, e.g., [Weiland and Yoshikawa \(2013\)](#).

⁶The parental labor income we observe is aggregated across the parents. Only 27% of the mothers lived with a partner at baseline, so we refer to the gain in parental labor income as a gain in mother's labor income.

estimates to the inclusion and exclusion of the various components of costs and benefits.⁷

A fundamental problem in evaluating the life-cycle benefits of any intervention with limited follow-up is assessing out-of-sample future costs and benefits. All solutions to this problem are based on versions of a synthetic cohort approach using the outcomes of older cohorts who did not have access to the program, but are otherwise comparable to the treated and control subjects, to proxy treatment effects of individuals when they are older.⁸

Two issues arise in using synthetic cohorts: (i) cohort effects; and (ii) the measured outcomes (e.g. test scores, education) used to forecast future outcomes might bear a different relationship in experimental samples than in non-experimental samples. The variation generated by random assignment is generally not the same as the variation generated in non-experimental data. We address issue (ii) by using policy-invariant structural models to combine experimental data through the mid 30's with information from multiple auxiliary panel data sources to forecast life-cycle benefits and costs. We test and do not reject the hypothesis of structural invariance.⁹

Our analysis is simplified by the fact that all of the families offered participation in the program took the offer. We create a synthetic treatment group by applying and extending the methodology in Heckman et al. (2013). Program treatment effects are produced through experimentally induced changes in intermediate inputs of a stable (across treatment regimes) production function for outcomes. We use production functions for adult outcomes to make out-of-sample forecasts based on inputs influenced by treatment, and test for the presence of misalignment in the error structures across experimental and non-experimental groups.

⁷Barnett and Masse (2002, 2007) present a cost/benefit analysis of ABC through age 21, before many benefits are realized. They report a benefit/cost ratio of 2.5, but give no standard errors or sensitivity analyses for their estimate. They do not disaggregate by gender. For want of the data collected on health at the mid 30s, they do not account for health benefits. They use self-reported crime data (unlike the administrative crime data later collected that we analyze) and ignore the welfare costs of financing the program. We use cost data from primary sources that were not available to them.

⁸Mincer (1974) addresses this problem using a synthetic cohort approach and provides evidence on its validity. See also the discussion of the synthetic cohort approach in Heckman et al. (2006).

⁹Ridder and Moffitt (2007) provide a valuable discussion of data combination methods. These methods are related to the older “surrogate marker” literature in biostatistics (see e.g., Prentice, 1989). However, that literature ignores the issue of misalignment of experimental and non-experimental predictors and outcomes and the use of structural models to correct it, which we address in this paper.

We account for sampling uncertainty arising from combining data, estimating parameters of behavioral equations, and simulation error. We conduct sensitivity analyses on outcomes for which sampling uncertainty is not readily quantified. Our approach to combining multiple data sets is of interest in its own right as a guide for forecasting the life-cycle benefits of social programs.

We utilize and build on the analysis of program treatment effects of ABC/CARE in [García et al. \(2017\)](#). They note that control-group substitution is a central feature of the experiment.¹⁰ Roughly 75% of the control-group children in ABC/CARE enroll in some form of lower-quality alternative childcare outside of the home.¹¹ [García et al. \(2017\)](#) define and estimate parameters accounting for the choices taken by control-group families. They find pronounced gender differences in treatment effects comparing high-quality treatment with lower-quality alternatives. Males benefit much less from lower-quality alternative childcare arrangements compared to females, a result consistent with the literature on the greater vulnerability of male children when removed from their mothers, even for short periods.¹² Estimated life-cycle benefits depend on alternatives to treatment.

We contribute to the literature on the effectiveness of early childhood programs by considering their long-term benefits on health. We estimate the savings from life-cycle medical costs and from improvements in the quality of life.¹³ There are benefits for participants in terms of reduced participation in crime, increased life-cycle labor income, reduced special education costs and enhanced educational attainment. The program subsidizes maternal childcare, thereby facilitating maternal employment, labor income, and educational attainment.

Figure 1 previews 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

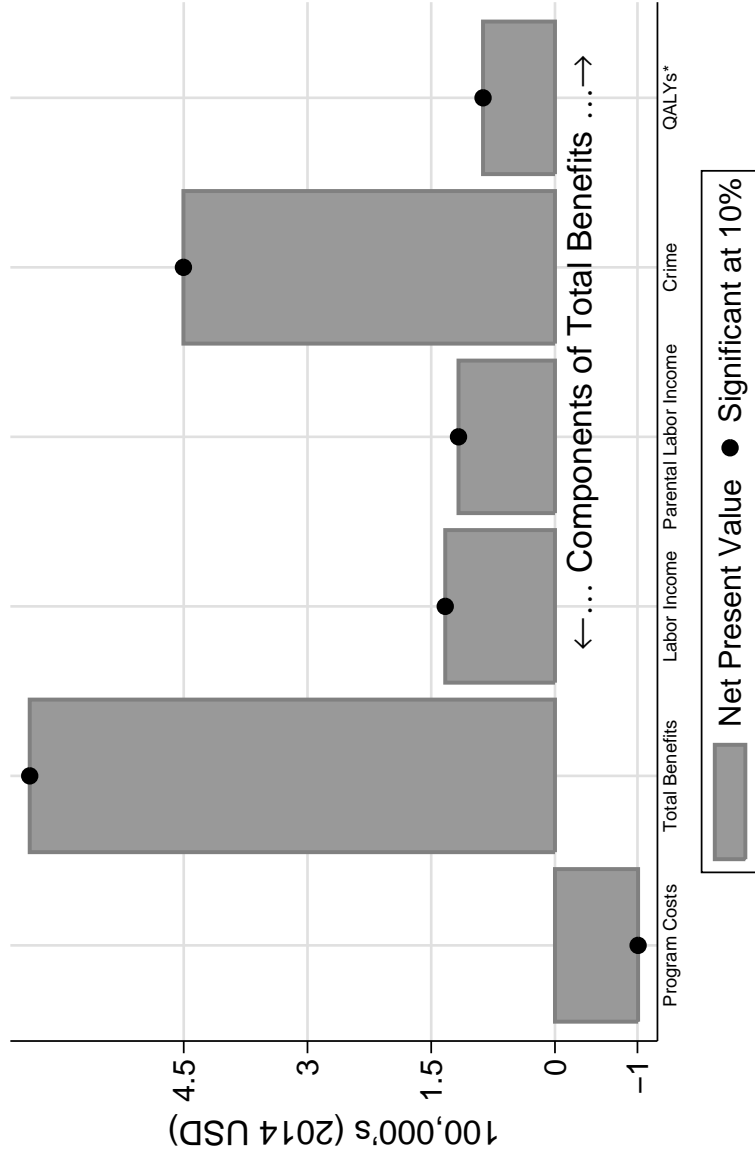
¹⁰See [Heckman \(1992\)](#), [Heckman et al. \(2000\)](#), and [Kline and Walters \(2016\)](#).

¹¹We refer to alternatives as alternative childcare or alternative preschool centers. See Appendix A.5 for a precise description of these alternatives.

¹²See [Kottelenberg and Lehrer \(2014\)](#), [Baker et al. \(2015\)](#), [Schore \(2017\)](#), and [García et al. \(2017\)](#).

¹³[Campbell et al. \(2014\)](#) show the substantial adult (mid 30s) health benefits of ABC but do not present a cost/benefit analysis of their results.

Figure 1: Net Present Value of Main Components of the Pooled (Over Gender) Cost/Benefit Analysis Over the Life Cycle per Program Participant, Treatment vs. Next Best



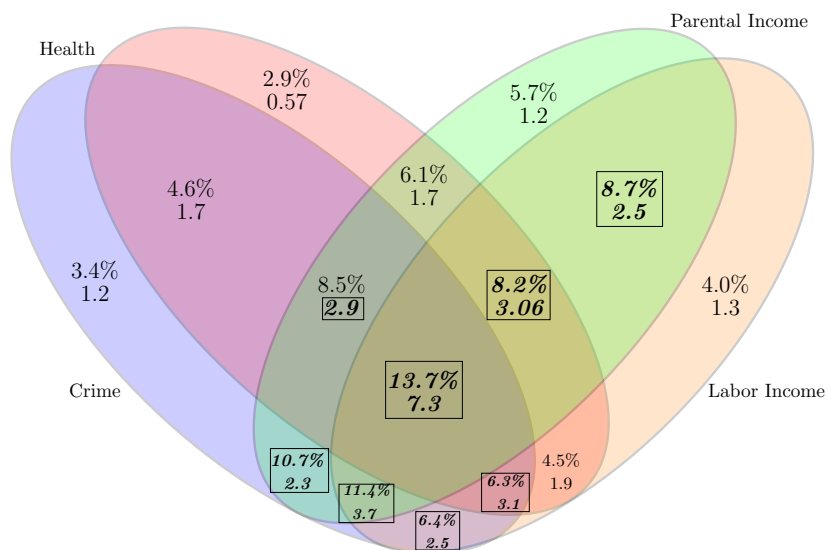
Per-annum Rate of Return: Males and Females 13.7% (s.e. 3%).
 Benefit-cost Ratio: Males and Females 7.3 (s.e. 1.8).

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. We 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.

categories, and for separate categories as well.¹⁴ We report separate estimates by gender later in this paper and find substantial differences in gender effects. The costs of the program are substantial, as has frequently been noted by critics.¹⁵ But so are the benefits, which far outweigh the costs.

Figure 2: 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 accounting for 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 1. 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 2 summarizes the results from our extensive sensitivity analyses. It shows the estimated rates of return and benefit/cost ratios when we calculate our estimates under the assumption that 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.

¹⁴The baseline discount rate of 3% is an arbitrary decision. In Table 3 and Table 5, we report benefit/cost ratios using other discount rates. Using 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 G.1.

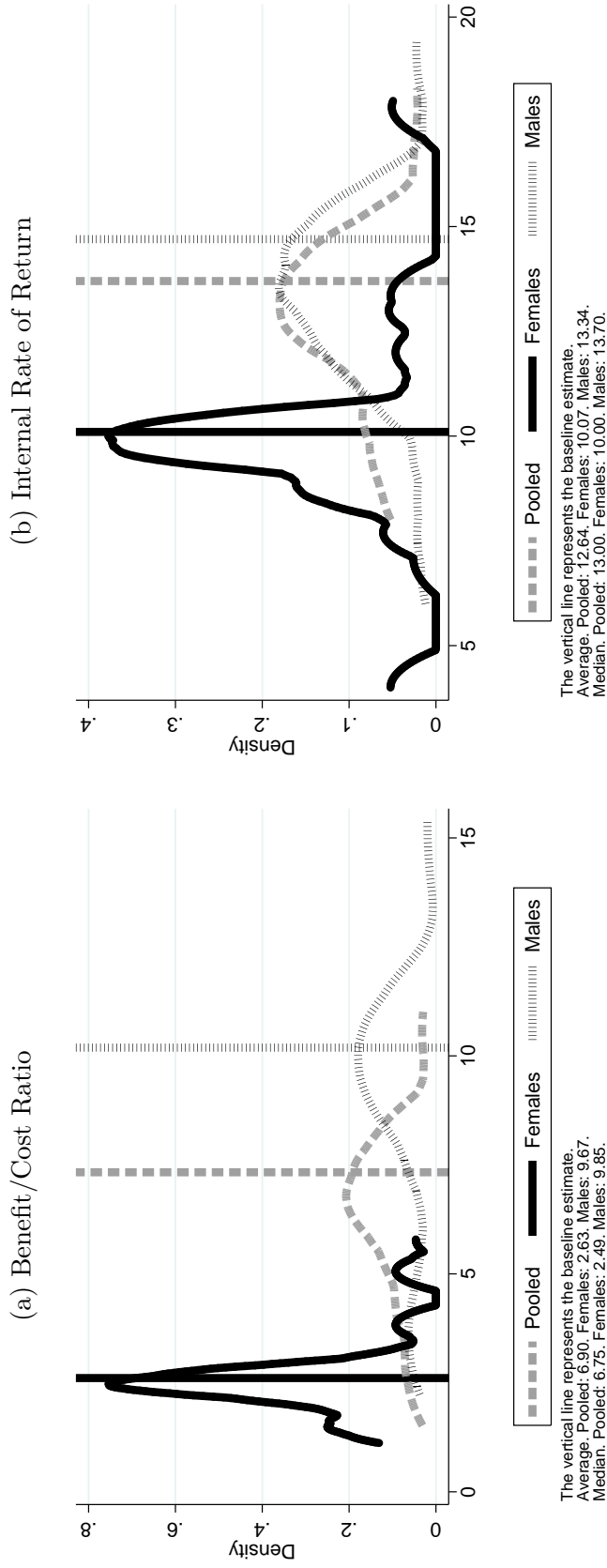
¹⁵See, e.g., Fox Business News (2014) and Whitehurst (2014).

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. No single component drives our results. We report estimates from a variety of specifications of regressors and functional forms. Our estimates are robust to different plausible assumptions about the values of non-market outcomes, for which conventional estimates of standard errors are not available.

Figure 3 summarizes the range of all of the estimates that we generate in this paper. The overall benefit/cost ratio (internal rate of return) ranges from 1.52 to 17.40 (8 to 18.3) for the full sample. They range from 2.23 to 25.45 (6 to 19.4) for the males, and from 1.12 to 5.79 (4 to 18) for the females. Benefits from reductions in criminality and increased labor income are pronounced for males, contributing to their larger estimates relative to the females estimates. However, when we omit crime from our analysis, we still estimate substantial returns for males. The returns of the program are higher for males relative to attending a low-quality preschool; the returns are higher for females relative to those who stay at home.

This paper justifies and interprets these estimates. The paper unfolds in the following way. Section 2 discusses the ABC/CARE program. Section 3 presents our notation and the definitions of the treatment effects reported in García et al. (2017) that are the basis for the estimates reported in this paper. They show that the program had substantial impacts on multiple domains. This paper reports treatment effects in economically meaningful metrics. Section 3 presents our methods for forecasting life-cycle outcomes and the evidence supporting the assumptions that justify them. Section 4 discusses how we monetize life-cycle outcomes. Section 5 reports our estimates of benefit/cost ratios and rates of return and the outcomes from a variety of robustness checks. Section 6 examines the predictive validity of *ad hoc* forecast methods currently used in the literature to estimate the long-run benefits of programs with short-term follow up. Section 7 summarizes the paper.

Figure 3: Distribution of Benefit/Cost and Internal Rate of Return Estimates



Note: Panel (a) displays the distribution of estimates of the benefit/cost ratio that we estimate throughout the paper. Vertical lines indicate the baseline estimates, presented in Figure 1. Panel (b) presents the analogous figure for the internal rate of return. See Figure G.1 in Appendix G for the analogous plots of net present values.

2 Background and Data Sources

ABC/CARE targeted disadvantaged, predominately African-American children in Chapel Hill/Durham, North Carolina.¹⁶ [García et al. \(2017\)](#) and [Appendix A](#) describe these programs in detail. Here, we summarize their main features.

The goal of these programs is to enhance the life skills of disadvantaged children. Both programs support 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.¹⁷ The programs provide health screenings to treatment group members, but health care costs are paid by parents.

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.¹⁸

The design and implementation of ABC and CARE are very similar. Both have two phases, the first of which lasts from birth until age 5. In this phase, children are randomly assigned to treatment. The second phase of the study consists of child academic support through home visits from ages 5 through 8. The first phase of CARE, from birth until age 5, has an additional treatment arm of home visits designed to improve home environments.¹⁹ Our analysis uses the first phase and pools the CARE treatment group with the ABC treatment group. We do not use data from the CARE group that only receive home visits in

¹⁶Both ABC and CARE were designed and implemented by researchers at the Frank Porter Graham Center of the University of North Carolina in Chapel Hill.

¹⁷[Sparling \(1974\)](#); [Ramey et al. \(1976, 1985\)](#); [Wasik et al. \(1990\)](#); [Ramey et al. \(2012\)](#).

¹⁸See [Appendix A.2](#) for details on the construction of the index. It weights the following variables (listed from the most to the least important according to the index): maternal and paternal education, family income, father’s presence at home, lack of maternal relatives in the area, siblings behind appropriate grade in school, family on welfare, father in unstable job, low maternal IQ, low siblings’ IQ, social agency indicates that the family is disadvantaged, one or more family members has sought a form of professional help in the last three years, and any other special circumstance detected by program’s staff.

¹⁹[Wasik et al. \(1990\)](#).

the early years. [Campbell et al. \(2014\)](#) test and do not reject the hypothesis that the data sets used in this paper have a common structure.

For both programs, from birth until age 8, data were collected annually on cognitive and socio-emotional skills, home environments, family structure, and family economic characteristics. After age 8, data on cognitive and socio-emotional skills, education, and family economic characteristics were collected at ages 12, 15, 21, and 30.²⁰ In addition, we have access to administrative criminal records and a physician-administered medical survey when the subjects were in their mid 30s.²¹

Randomization for ABC/CARE was conducted on child pairs matched on family characteristics. Siblings and twins were jointly randomized into either treatment or control groups.²² Randomization pairing was based on the childhood risk index, maternal education, maternal age, and gender of the subject.²³ Dropouts are evenly balanced and are predominantly related to the health of the child and mobility of families rather than dissatisfaction with the program.²⁴

Seventy-five percent of the ABC control group and 74% of the CARE control (but no children from the families offered treatment) attended alternative (to home) childcare.²⁵ Those who enrolled generally stayed enrolled. As control children age, they are more likely to enter alternative childcare (see [Appendix A.5](#)). Children in the control group who are enrolled in alternative early childcare programs are less economically disadvantaged at baseline compared to children who stay at home. On average, they are children with mothers

²⁰At age 30, measures of cognitive skills are unavailable for both ABC and CARE.

²¹See [Appendix A.6](#) 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 attrition for these data collections. Further, the methodology we propose addresses missing data in either of these two outcome categories.

²²For siblings, this occurred when two siblings were close enough in age such that both of them were eligible for the program.

²³We do not know the original pairs.

²⁴The 22 dropouts in ABC include four children who died, four children who left the study because their parents moved, and two children who were diagnosed as developmentally delayed. See [Table A.3](#) for details. All eligible families agreed to participate. Dropping out occurs *after* randomization.

²⁵See [Heckman et al. \(2000\)](#) on the issue of substitution bias in social experiments.

who were more likely to be working at baseline.²⁶ Parents of control-group girls were much more likely than parents of control-group boys to use alternative childcare if assigned to the control group.²⁷

3 Forecasting Life-cycle Costs and Benefits

This section discusses the model and data sources used to forecast life cycle profiles.

3.1 Parameters Underlying Our Analysis

We begin by summarizing the treatment parameters underlying our analysis.²⁸ Toward this end, we define three indicator variables: $W = 1$ indicates that the parents referred to the program participate in the randomization protocol, $W = 0$ indicates otherwise. R indicates randomization into the treatment group ($R = 1$) or to the control group ($R = 0$). D indicates attending the program, i.e., $D = R$ implies compliance with the initial randomization protocol.

Individuals are eligible to participate in the program if baseline background variables $\mathbf{B} \in \mathcal{B}_0$. \mathcal{B}_0 is the set of scores on the risk index that determines program eligibility. Because all of the eligible persons given the option to participate choose 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 population for which $\mathbf{B} \in \mathcal{B}_0$ and not just treatment effects for the treated (**TOT**).²⁹

²⁶The difference is statistically significant at 10%.

²⁷Most of the alternative childcare centers received federal subsidies and were subject to the federal regulations of the era (Appendix A.5.1). See Department of Health, Education, and Welfare (1968); North Carolina General Assembly (1971); Ramey et al. (1977); Ramey and Campbell (1979); Ramey et al. (1982); Burchinal et al. (1997). They had relatively low quality compared to ABC/CARE.

²⁸See García et al. (2017) for a full discussion of these parameters and the presentation of their estimates.

²⁹All 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. Our conversations with the program staff indicate that the encouragement from the referral agencies was such that most referred mothers attended and agreed to participate in the initial randomization (Ramey et al., 2012).

Denote potential outcome j , $j \in \mathcal{J}_a$ at age $a \in [1, \dots, A]$ under treatment status $d \in \{0, 1\}$ (treatment or control) in the sample $k \in \{e, n\}$, by $Y_{k,j,a}^d$. The set \mathcal{J}_a indexes the outcomes of interest measured at age $a \in [1, \dots, A]$.

All treatment group children have the same exposure. Although it would be ideal to analyze control children by the length of their exposure to alternative environments, data limitations lead us to simplify the analysis of the control substitution by creating two categories. “ H ” indicates that the control child is in home care throughout the entire length of the program. “ C ” indicates that the control child is in alternative childcare for any amount of time.³⁰ This separation of the counterfactual setting allows us to calculate the returns of ABC/CARE in comparison to (i) staying at home (H), (ii) alternative childcare (C), or (iii) the “next best” option, which pools across the full control group.

3.2 Using Auxiliary Data Sources to Forecast Out-of-Sample Outcomes

The goal of this paper is to quantify the multiple benefits of ABC/CARE in terms of benefit/cost ratios and rates of return. We rely on auxiliary data to forecast the costs and benefits of the program over the life cycle after the measurement phase of the study ends.³¹

This section explains our strategy for constructing out-of-sample treatment effects.³² Our approach builds on the analysis of Heckman et al. (2013), who show that, for the Perry Preschool program, the effect of treatment on outcomes operates through its effects on measured inputs in a stable production function rather than through shifts in the production function. If this is also true for ABC/CARE, this feature greatly facilitates our projection analysis. We test and do not reject the hypothesis that treatment works by shifting inputs

³⁰This assumption is consistent with the finding in García et al. (2017) that once parents decide to enroll their children in alternative childcare arrangements, the children stay enrolled up to age 5. They also find little sensitivity of the estimates of treatment effects to the choice of different, related categorizations.

³¹See Appendix C.3.2 for an overview of the auxiliary datasets that we use.

³²Appendix C.7 gives details of our step-by-step procedure and states its identification and estimation strategy in the Generalized Method of Moments framework.

Table 1: Summary of Forecast Methodology to Construct Life-cycle Costs and Benefits

Component	Subject's Age at Prediction	Baseline Prediction Method	Variables Used to Construct Synthetic Experimental Groups	Variables Used to Predict	Auxiliary Samples Used
Program Costs	0 to 5	Observed (source documents)	N/A	N/A	N/A
Costs of Alternative Preschools	0 to 5	Estimated from Location & Time Relevant Documents	N/A	N/A	N/A
Education Costs (includes special education and grade retention)	up to 30	Level is Observed (Per Level Cost taken from NCES)	N/A	N/A	N/A
Labor Income or Transfer Income	21 to 30	Based on Prediction Model in the Auxiliary Sample	Birth-year; Gender; Siblings at Birth	Gender; Mother's Education; at Birth; PIAT Math (5 to 7); Education (30) Labor Income (21) Lagged Outcome	CNLSY
Labor Income or Transfer Income	30 to 67	Based on Prediction Model in the Auxiliary Sample	Birth-year; Gender; Siblings at Birth	Gender; Education (30); Labor Income (30); Lagged Outcome	Pooled NLSY79 and PSID
Parental Labor Income	0 to 21	Linear Interpolation (Observed Values at Ages 1.5, 2.5, 3.5, 8, 12, 15, 21)	N/A	N/A	N/A
Crime (Arrests and Sentences)	up to Mid 30s	Observed* (Combines Administrative and Self-reported Data)	N/A	N/A	N/A
Crime (Arrests and Sentences)	Mid 30s to 50	Based on Prediction Model in the Auxiliary Sample (One Prediction per Arrest or Sentence)	Use Full Auxiliary Sample to Predict Control and Treatment Outcomes	Lagged Crime Outcomes (all outcomes listed in Table E.1)	NCJDPS
Victimization Inflation	up to Age 50	Impute national victims-arrests ratio	Use Full Auxiliary Samples to Impute	N/A	NCVS; NJRP; UCERS (vary by crime)
Health Costs	before Age 30	Based on Prediction Model in the Auxiliary Sample	Use Full Auxiliary Sample to Predict	Age-specific (four follow-ups available) detailed in Table F.6	MEPS
Health Transitions (includes disability claims)	30 to Death	Based on Prediction Model in the Auxiliary Sample	Use Full Auxiliary Samples to Predict	Gender; Education (30); Lagged Health Outcomes as indicated in Table F.1	PSID and HRS (only for mortality)
Health Costs	30 to Death	Based on Prediction Model in the Auxiliary Sample	Use Full Auxiliary Samples to Predict	Age; Gender; Race; Education (30); Marital Status (30); Disease Conditions; Labor Income (30)	MEPS MCBS (if Medicaid eligible)
QALYs	30 to Death	Based on Prediction Model in the Auxiliary Sample	Use Full Auxiliary Samples to Predict	ADL and IADL counts; Disease Conditions	PSID and MEPS
Deadweight-loss	0 to Death	.50 cents per each government-spent dollar	N/A	N/A	N/A

Note: This table summarizes our methodology for forecasting the costs and benefits of each component that we consider. Abbreviations: ADL: Activities for Daily Living; IADL: Instrumental Activities for Daily Living; CNLSY: Children of the National Longitudinal Survey of the Youth 1979; HRS: Health and Retirement Study; NCES: National Center of Education Statistics; NCDPS: North Carolina Department of Public Safety Data; NLSY79: National Longitudinal Survey of the Youth 1979; MEPS: Medical Expenditure Panel Survey; MCBS: Medicare Current Beneficiary Survey; NJRP: National Judicial Reporting Program; NVS: National Crime Victimization Survey; PSID: Panel Study of Income Dynamics; UCERS: Uniform Crime Reporting Statistics. N/A: not applicable. A quality-adjusted life year (QALY) weighs a year of life according to its quality given the burden of disease. Suppose we assign a value of \$150,000 (2014 USD) to each year of life. A QALY of \$150,000 denotes a year of life in the absence of disease (perfect health). The value of QALY for an individual in a given year is smaller than \$150,000 when there is positive burden of disease, as worse health conditions imply a lower quality of life. When an individual dies, her QALY equals zero. There are extreme combinations of disease and disability that may generate negative QALYs, although this is unusual. Because we quantify labor income in addition to other components, this value corresponds solely to monetizing the value of life net of what individuals produce in terms of economic output. The benefit/cost ratio and internal rate of return remain significant after removing this component entirely (see Table 5).

*When not observed, we impute based on the national arrest-sentence ratio from NJRP and UCERS. We assume that a lack of criminal records before the mid 30s implies no crime after the mid 30s.

and not the production function using the ABC/CARE data.

Table 1 presents the outcomes for which we conduct these analyses. It also summarizes the methodology and auxiliary samples used to make our predictions. We focus on labor income to illustrate our approach, but a similar methodology is used to forecast other outcomes.^{33,34} We first present an intuitive summary of our approach. We formalize these intuitions in the next section, and in the Appendix. The remaining sections apply the methodology to other outcomes besides labor income.

We have data on control- and treatment-group members through age a^* . We can identify treatment effects within the experimental sample at these ages. We cannot identify these treatment effects at ages for which we lack information on participant outcomes. Instead, we forecast post- a^* treatment effects, which are required to construct counterfactual life-cycle profiles.

Making valid forecasts of out-of-sample treatment effects does not require making valid forecasts of separate out-of-sample treatment and control profiles. Only valid forecasts of their difference are required. We compare the predictive power of constructed treatment and control groups through age a^* . We also analyze the performance of forecasts of separate treatment and control profiles after age a^* . Doing so allows us to test the validity of our methodology by comparing (within the support of the experimental sample) outcomes by treatment status for the experimental control and treatment groups with those from the synthetic control and treatment groups we generate. Comparisons between the experimental

³³We do not monetize the loss of leisure and household production that individuals experience from working (this applies both to the individuals in the program and to their parents). The reasons for this are two-fold: (i) we lack information on intensive-margin labor supply; and (ii) different labor supply models and market structures have different implications with respect to the value of non-market time. In addition, our data are not well-suited for estimating a structural model of labor supply. We note, however, that the benefit/cost ratio and internal rate of return are both statistically significant and substantial after removing any benefits from labor income entirely (see Table 5). This exercise corresponds to a one-to-one loss of leisure given the gain in labor income, i.e. for each additional dollar an individual makes, she loses the same dollar of (monetized) leisure and household production.

³⁴Our calculations are based on labor income gross of tax, because we want to quantify the effects of the program on the gross output that an individual is able to produce. A rise in gross labor income increases the taxable base and has an implied increase in deadweight loss. We do not quantify that deadweight loss because: (i) we do not have enough information to make full use of standard tax simulators; and (ii) we are not able to vary the standard tax simulators to assess estimation uncertainty.

control group and the synthetic control group are particularly compelling because neither group receives treatment.³⁵ Because all persons in the experimental sample who are offered treatment accept it, it is straightforward to construct synthetic control groups in auxiliary samples using only eligibility criteria.

There are two distinct stages in our analysis. In Stage I, we construct samples of individuals in the auxiliary samples with characteristics similar to those of the individuals in the experimental sample. The minimal set of characteristics includes the background variables $\mathbf{B} \in \mathcal{B}_0$. We use a coarse form of matching based on Algorithm 1 in Appendix C.3.3. In Stage II, we build models using these samples to forecast out-of-sample outcomes separately for the treated and the controls.

Stage II is implemented by a three-step procedure. In Step 1, we use the experimental sample to conduct mediation analyses relating the vector of outcomes at age a for person i ($\mathbf{Y}_{i,a}^d$) for $a \leq a^*$ to predictor variables (and interactions) that are affected by treatment ($\mathbf{X}_{i,a}^d$), as well as background variables (\mathbf{B}_i). It turns out that we accurately predict within-sample treatment effects as well as the levels of treatment- and control-group profiles using this approach. In Step 2, we construct counterpart forecasts of treatment and control outcomes using the auxiliary samples. We compare the constructed counterparts to the actual samples for ages $a \leq a^*$. In Step 3, we use the estimated dynamic relationships fit on the constructed samples to forecast the post- a^* outcomes.

Under exogeneity of the predictor variables and structural invariance, the two stages can be compressed into a single, one-stage, non-parametric matching procedure.³⁶ In Appendix C.3.4.1 we compare the estimates from matching with those from our main approach and find close agreement between the two approaches and for different assumptions about the serial correlation processes of the outcome equations.

³⁵Although some might be attending other centers. We control for participation in Head Start in our auxiliary samples. Doing so does not substantially alter our estimates. The raw difference in the net present values between treatment and control labor income is 81,230 (123,210) for females (males). This is compared to 55,720 (153,140) for the baseline estimates. We underestimate the treatment effect by not conditioning on Head Start.

³⁶See Heckman et al. (1998) for an example.

Figure 4 previews the results of applying the two-stage approach, displaying the life-cycle labor income profiles for the treatment and control groups. It also compares the realized labor income with the model-predicted labor income at a^* . There is close agreement of the constructed profiles within the age group of the experimental sample. 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).³⁷

We conduct a further check on the validity of our procedure. In the experimental sample all of the parents of children with characteristics $\mathbf{B} \in \mathcal{B}_0$ agree to participate in the program. Because the auxiliary samples have no treatment group members, we can evaluate our procedure by comparing the labor incomes of individuals in the auxiliary samples for whom $\mathbf{B} \in \mathcal{B}_0$ to the labor incomes of individuals in our constructed synthetic control group. Figure 5 makes this comparison. It plots the average labor incomes of individuals in our auxiliary sample for whom $\mathbf{B} \in \mathcal{B}_0$ alongside those of the constructed synthetic control group from ages 20 to 45. It also displays the labor income of the experimental control group at a^* (age 30).³⁸ The agreement is reassuringly close.

3.3 Constructing Out-of-Sample Counterfactuals

We now formalize our analytical framework and its underlying assumptions. To avoid notational clutter, we henceforth suppress individual i subscripts. Our analysis is based on a causal (structural) model for treatment ($d = 1$) and control ($d = 0$) counterfactual outcomes for outcome j measured at age a in sample $k \in \{e, n\}$, where e denotes membership in the experimental sample and n denotes membership in the auxiliary sample:

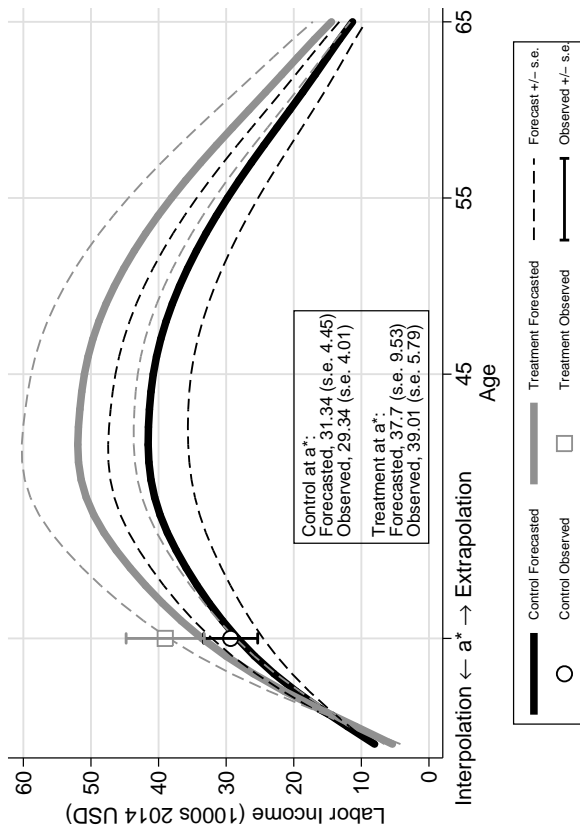
$$Y_{k,j,a}^d = \phi_{k,j,a}^d(\mathbf{X}_{k,a}^d, \mathbf{B}_k) + \varepsilon_{k,j,a}^d, \quad j \in \mathcal{J}_a, \quad (1)$$

³⁷For details on the variables used to construct the forecasts, see Appendix C.

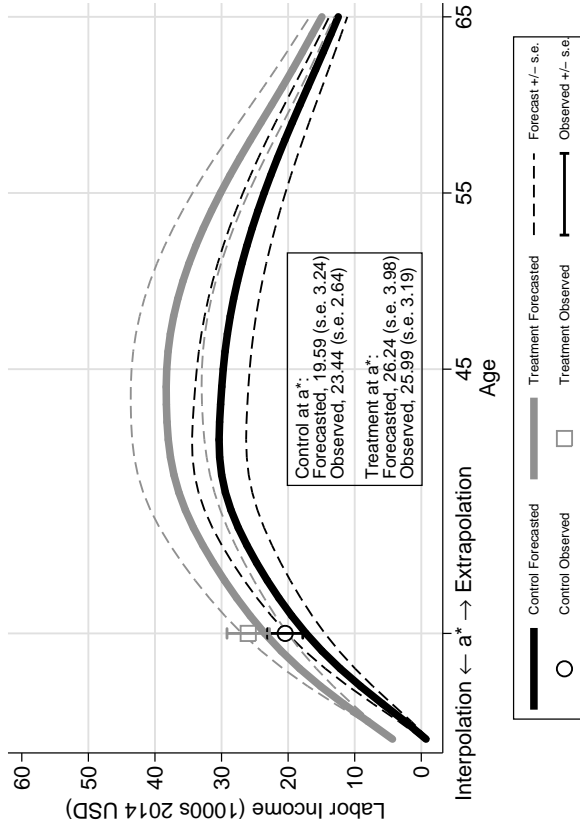
³⁸The graphs stop at age 45 because we do not observe all of the components of the risk index determinants of eligibility after age 45 in the auxiliary samples. We use only a subset of this index to make life-cycle projections. These variables are effective predictors over the age range for which the full set of \mathbf{B} is available.

Figure 4: Forecasted Labor Income Profiles for ABC/CARE Participants

(a) Males



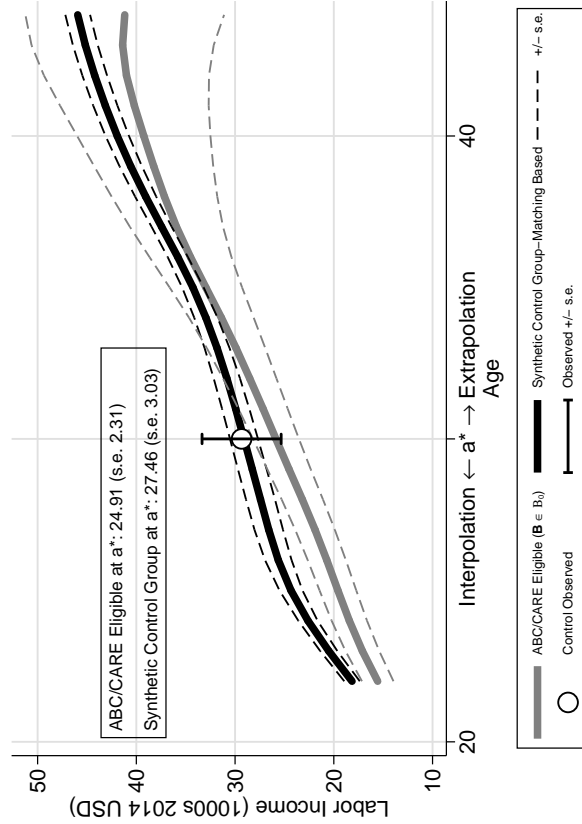
(b) Females



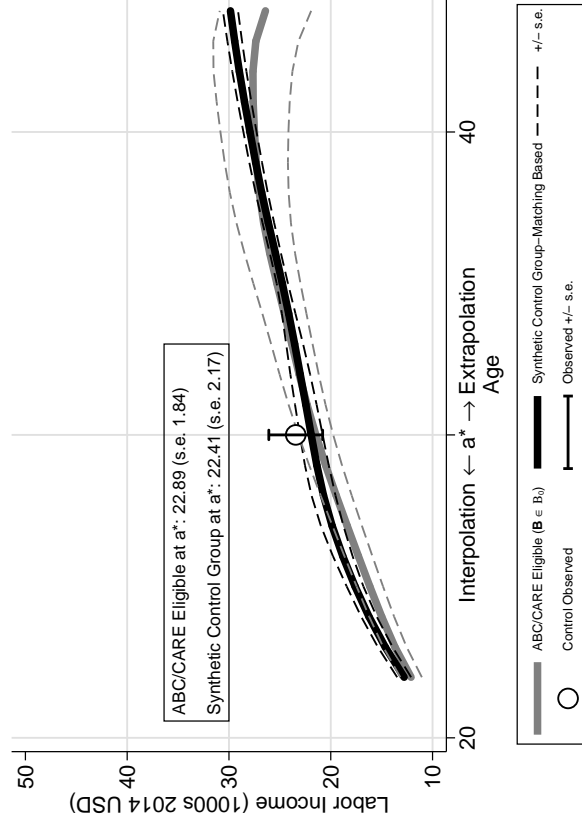
Note: Panel (a) displays the forecasted life-cycle labor income profiles for ABC/CARE males by treatment status, based on the method proposed in Section 3.2. We combine data from the Panel Study of Income Dynamics (PSID), the National Longitudinal Survey of Youth 1979 (NLSY79), and the Children of the 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 based on the empirical bootstrap distribution. See Condition C-3 for a discussion of testing the difference in the observed and forecasted treatment effects (as opposed to the difference in the observed and forecasted levels). 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).

Figure 5: Labor Income Profile, Disadvantaged Individuals Synthetic Control Group in the Auxiliary Samples

(a) Males



(b) Females



Note: Panel (a) displays the forecasted labor income for males in the auxiliary samples for whom $\mathbf{B} \in \mathcal{B}_0$, i.e., ABC/CARE eligible, and for the synthetic control group 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 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 $\mathbf{B} \in \mathcal{B}_0$ in the auxiliary samples. Panel (b) displays the analogous figure for females. Standard errors are based on the empirical bootstrap distribution.

where $\phi_{k,j,a}^d(\cdot, \cdot)$ is an invariant structural production relationship mapping inputs $\mathbf{X}_{k,a}^d, \mathbf{B}_k$ into output $Y_{k,j,a}^d$ holding error term $\varepsilon_{k,j,a}^d$ fixed.³⁹ We normalize $\varepsilon_{k,j,a}^d$ to have mean zero. Among the $\mathbf{X}_{k,a}^d$ are variables caused by treatment, including lagged dependent variables. In this general framework, the relationships between the dependent and right-hand side variables in Equation (1) do not necessarily coincide across the samples, $k \in \{e, n\}$.

Let \mathbf{Y}_k^d denote the vector of all outcomes at all ages for $k \in \{e, n\}$, when treatment status is fixed to d . Similarly, \mathbf{X}_k^d is the vector of all causal predictors of \mathbf{Y}_k^d at all ages. Both \mathbf{Y}_k^d and \mathbf{X}_k^d include the full set of possible outcomes over the life cycle, even though they are not observed after age a^* . The background variables may have different distributions in the two samples. We denote the joint distribution of these vectors conditional on $\mathbf{B}_k = \mathbf{b}$ by $F_{\mathbf{Y}_k^d, \mathbf{X}_k^d | \mathbf{B}_k = \mathbf{b}}(\cdot, \cdot)$.

In the experimental sample, parents of eligible children ($\mathbf{B}_e \in \mathcal{B}_0$), always agree to participate in the program ($W_e = 1$) and accept treatment ($R_e = D_e$). We assume that this condition holds in the auxiliary sample. Given this condition, we can use D_e and R_e interchangeably and apply a standard [Quandt \(1972\)](#) switching regression model to write the outputs and inputs generated by treatment as

$$\begin{aligned} Y_{k,j,a} &= (1 - D_k) Y_{k,j,a}^0 + (D_k) Y_{k,j,a}^1, & (2) \\ \mathbf{X}_{k,a} &= (1 - D_k) \mathbf{X}_{k,a}^0 + (D_k) \mathbf{X}_{k,a}^1. &^{40} \end{aligned}$$

$j \in \mathcal{J}_a, a \in \{1, \dots, A\}, \quad k \in \{e, n\}$

The fact that $D_e = R_e$ allows us to use experimental data (for $a \in \{1, \dots, a^*\}$) to identify the distribution of $Y_{e,j,a}^d$ (i.e., $Y_{e,j,a}^d$ when fixing treatment status (d)).

3.3.1 Accounting for Age, Period, and Cohort Effects

The auxiliary data (n) come from older cohorts not exposed to the program, for whom we observe more complete segments of their life cycles. We do not observe what treatment

³⁹Fixing and conditioning are fundamentally different concepts. See [Haavelmo \(1943\)](#) and [Heckman and Pinto \(2015\)](#) for discussions. Our analysis applies the methodology of these papers.

⁴⁰We keep the conditioning on $\mathbf{B} \in \mathcal{B}_0$ implicit.

status d would have been in the auxiliary data. Even if we did, we do not know if cohort (c) or time (t) effects make the experiences of the individuals in the auxiliary sample different from the experiences of the individuals in the experimental sample.

To formalize this problem, and our solution to it, let $Y_{j,k,a,c,t}^d$ be outcome j for sample k at age a for birth cohort c at time t when treatment is fixed to d . We make the following assumption. It allows us to circumvent the problem by assuming cohort and time effects operate identically across the experimental (e) and non-experimental (n) samples in the following sense:

Assumption A–1 *Alignment of Cohort and Time Effects*

For experimental sample cohort c_e and auxiliary sample cohort c_n :

$$Y_{e,a,c_e,t_e}^d = Y_{n,a,c_n,t_n}^d \tag{3}$$

for $d \in \{0, 1\}$, $a \geq a^$, where t_e, t_n are the years for which cohorts c_e, c_n are observed, where $t_e = t_n + c_e - c_n$, and t_n is the year that the age a outcome is observed for cohort n ($t_n = a + c_n$). \square*

Notice that Y_{n,a,c_n,t_n}^d is the outcome for treatment status d in the auxiliary sample. Assumption A–1 does not rule out cohort or period effects. However, it rules out any *differences* in cohort and time effects of the auxiliary sample and the experimental sample when they reach the age of those in the auxiliary sample.

We henceforth drop the “ c ” and “ t ” sub-indices. The out-of-sample year effect for the experimental sample is assumed to be the same as for the auxiliary sample measured at year t_n . We can weaken Assumption A–1 if there is prior knowledge about year and/or cohort effects or if we can parameterize estimable functions of c and t .⁴¹ In the sensitivity analyses reported below, we examine plausible alternative assumptions about cohort and time effects drawing on results in the empirical literature. Examples include a rate of decay in labor income due to a time effect as a crisis or a cohort effect due to skill depreciation. We also

⁴¹See Heckman and Robb (1985). For health, cohort effects could be very substantial (e.g., the growth of medical costs) and we account for this as explained in Section 4.1 and Appendix F.

consider wage increases due to alternative scenarios of gains in productivity, or time effect due to plausible changes in the structure of medical costs.

3.3.2 Support Conditions

We require that the support of the auxiliary sample contains the support of the experimental sample. This assumption allows us to find counterpart values of $\mathbf{X}_{k,a}^d$, \mathbf{B} , and $\mathbf{Y}_{k,a}$ in the control and experimental samples.

Assumption A–2 *Support Conditions*

For $a \in \{1, \dots, A\}$, the support of $(\mathbf{Y}_{e,a}^d, \mathbf{X}_{e,a}^d, \mathbf{B}_e)$ in the experimental sample is contained in the support of $(\mathbf{Y}_{n,a}^d, \mathbf{X}_{n,a}^d, \mathbf{B}_n)$ in the auxiliary sample:

$$\text{supp}(\mathbf{Y}_{e,a}^d, \mathbf{X}_{e,a}^d, \mathbf{B}_e) \subseteq \text{supp}(\mathbf{Y}_{n,a}^d, \mathbf{X}_{n,a}^d, \mathbf{B}_n), \quad d \in \{0, 1\}. \quad \square \quad (4)$$

This assumption is straightforward to test for ages $a \leq a^*$. It is satisfied in our samples, as shown in Appendix C.3.5.

3.3.3 Conditions for Valid Out-of-Sample Forecasts

A strong sufficient condition for identifying the distribution of life-cycle profiles of individuals in the experimental sample using individuals in the auxiliary samples is Condition C–1:

Condition C–1 *Equality of Distributions Across the Experimental and Auxiliary Samples*

$$F_{\mathbf{Y}_e^d, \mathbf{X}_e^d | \mathbf{B}_e = \mathbf{b}}(\cdot, \cdot) = F_{\mathbf{Y}_n^d, \mathbf{X}_n^d | \mathbf{B}_n = \mathbf{b}}(\cdot, \cdot), \quad d \in \{0, 1\} \quad (5)$$

for $\mathbf{Y}_e^d, \mathbf{X}_e^d | \mathbf{B}_e = \mathbf{b}$ and $\mathbf{Y}_n^d, \mathbf{X}_n^d | \mathbf{B}_n = \mathbf{b}$ contained in the support of the experimental sample $\text{supp}(\mathbf{Y}_e^d, \mathbf{X}_e^d, \mathbf{B}_e)$.

Since we are only interested in means for benefit/cost analysis, we can get by with a weaker requirement for conditional means, which has testable implications, as we show below:

Condition C–2 *Equality in Conditional Expectations Across the Experimental and Auxiliary Samples*

$$\mathbb{E} [Y_e^d | X_e^d = \mathbf{x}, B_e = \mathbf{b}] = \mathbb{E} [Y_n^d | X_n^d = \mathbf{x}, B_n = \mathbf{b}], \quad d \in \{0, 1\} \quad (6)$$

for $d \in \{0, 1\}$ over $\text{supp} (Y_{e,a}^d, X_{e,a}^d, B_e)$.

Since we are primarily interested in treatment effects, we can get by with an even weaker condition:

Condition C–3 *Equality in Mean Treatment Effects Across the Experimental and Auxiliary Samples*

$$\mathbb{E} [Y_e^1 - Y_e^0 | B_e = \mathbf{b}] = \mathbb{E} [Y_n^1 - Y_n^0 | B_n = \mathbf{b}] \quad (7)$$

over $\text{supp} (Y_{e,a}^d, B_e)$.⁴²

We could simply invoke Condition C–2 or C–3 and be done. Our approach is to examine and test (when possible) assumptions that justify the treatment effects, and Condition C–2 is useful for doing so.

3.3.4 Exogeneity

Conditions C–1 to C–3 do not require that we take a position on the exogeneity of X_k^d , $k \in \{e, n\}$. However, exogeneity facilitates the use of economic theory to generate and interpret treatment effects, to test the validity of our synthetic control groups, and to find auxiliary sample counterparts to treatments and controls. It also facilitates matching, one of the methods used in this paper to construct synthetic treatment and control groups.⁴³ For these purposes, we assume:

⁴²We test the difference between the observed and forecasted treatment effect, because as we describe above, an equal difference (as opposed to equal levels) is sufficient. Even though the point estimate of the difference is 2,037.08 (–2,256.09) for males (females, we find that this difference is not statistically different than 0 for both genders).

⁴³See Heckman and Navarro (2004).

Assumption A–3 Exogeneity

For all $a, a'' \in \{1, \dots, A\}$ and for $d, d' \in \{0, 1\}$,

$$\varepsilon_{k,j,a}^d \perp\!\!\!\perp \mathbf{X}_{k,a''}^{d'} \mid \mathbf{B}_k = \mathbf{b} \quad (8)$$

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

Assumption A–3 is much stronger than needed. It justifies Conditions C–1 and C–2, but these conditions only require that endogeneity processes are governed by the same relationships in the experimental and auxiliary samples. However, when this assumption and our other assumptions hold, we generate testable implications of our forecasting model.

To appreciate the benefit of Assumption A–3, consider the following example. Say we want to predict future labor income and we use years of education as a main component of $\mathbf{X}_{k,a''}^{d'}$. The joint distribution of $\varepsilon_{k,j,a}^d$ and $\mathbf{X}_{k,a''}^{d'}$ could differ substantially across experimental and non-experimental samples. In the experimental sample, years of education are increased by treatment, which is randomly assigned. In the non-experimental samples, however, there is no treatment. Individuals with high observed levels of education could very well have a high value of $\varepsilon_{k,j,a}^d$ (e.g., ability bias). Assumption A–3 avoids this problem when making forecasts. However, under further assumptions, it is testable and fixable.

In Appendix C.3.6, we test Assumption A–3 for a variety of outcomes and fail to reject the null of exogeneity. We discuss these tests in Section 3.3.6. In Appendix C.6, we analyze standard panel data models for the outcome equations as well as instrumental variable approaches to account for lagged dependent variables and serial correlation. Our estimates are robust even when we allow for different failures of Assumption A–3. We also present non-parametric matching estimates. Their near-coincidence with the matching estimates with those from the structural model is further support of the validity of Assumption A–3.

3.3.5 Structural Invariance

We assume that the variables $\mathbf{X}_{k,a}^d$ fully summarize treatment in the sense that any effect that treatment has on outcomes operates through the inputs, $\mathbf{X}_{k,a}^d$, and not through shifts in the production function relating inputs to outputs (see Heckman et al., 2013). Assumption A–4 formalizes this condition.

Assumption A–4 *Structural Invariance*

For all $\mathbf{x}, \mathbf{b} \in \text{supp}(\mathbf{X}_{e,a}^d, \mathbf{B}_e)$, $k \in \{e, n\}$

$$\begin{aligned} \phi_{k,j,a}^0(\mathbf{x}, \mathbf{b}) &= \phi_{k,j,a}^1(\mathbf{x}, \mathbf{b}) \\ &=: \phi_{j,a}(\mathbf{x}, \mathbf{b}), \end{aligned} \tag{9}$$

$\phi_{k,j,a}^d(\mathbf{x})$ is the function generating the causal effect of setting $\mathbf{X}_{k,a}^d = \mathbf{x}$ holding $\varepsilon_{k,j,a}^d$ fixed for $a \in \{1, \dots, A\}$ for any outcome $j \in \mathcal{J}_a$. \square

This assumption has two distinct aspects which could be broken down into two separate assumptions: (i) the structural functions evaluated with the same arguments have identical values for treatment and control groups in the experimental sample, and (ii) the structural relationships are identical in the experimental and auxiliary samples. As previously noted, exogeneity is not needed to justify Conditions C–1 through C–3. But in the absence of exogeneity, the relationship between the inputs, $\mathbf{X}_{k,a}^d$, and the errors, $\varepsilon_{k,a}^d$, likely differs across experimental (e) and non-experimental (n) samples because randomization imparts a source of exogenous variation to $\mathbf{X}_{e,a}^d$ that is absent in non-experimental samples.

3.3.6 Testing Exogeneity

In Appendix C.3.6, we report tests for endogeneity in the experimental and auxiliary samples used in this paper. We assume that $\varepsilon_{k,j,a}^d, k \in \{e, n\}$ follows a factor structure.⁴⁴ We provide evidence supporting exogeneity in both samples for the predictor variables used

⁴⁴Factor structure models are widely used in structural estimation of production functions of skills during early childhood. See, e.g., Cunha and Heckman (2008); Cunha et al. (2010) and Agostinelli and Wiswall (2016).

in our empirical analyses. Once we condition on $\mathbf{X}_{k,a}^d$ and \mathbf{B}_k , we do not reject the null hypothesis of exogeneity.

3.3.7 Testable Implications

Assumption A-4 combined with Assumption A-3, Equation (2), and the assumption $\mathbb{E}(\varepsilon_{k,j,a}^d) = 0$ for all $a \in \{1, \dots, A\}$ generate testable restrictions for our forecasting models. Exogeneity and invariance enable us to jointly test the two aspects of structural invariance for $a \leq a^*$, when $Y_{k,j,a}$ is observed in both the experimental and auxiliary samples:

$$\mathbb{E}[Y_{e,j,a}^1 | \mathbf{X}_{e,a}^1 = \mathbf{x}, \mathbf{B}_e = \mathbf{b}, D = 1] = \mathbb{E}[Y_{e,j,a}^0 | \mathbf{X}_{e,a}^0 = \mathbf{x}, \mathbf{B}_e = \mathbf{b}, D = 0] \quad (10)$$

and

$$\mathbb{E}[Y_{e,j,a}^d | \mathbf{X}_{e,a}^d = \mathbf{x}, \mathbf{B}_e = \mathbf{b}, D = d] = \mathbb{E}[Y_{n,j,a} | \mathbf{X}_{n,a} = \mathbf{x}, \mathbf{B}_n = \mathbf{b}] \text{ for } d \in \{0, 1\}. \quad (11)$$

Under our assumptions, 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 $\mathbf{X}_{n,a} = \mathbf{x}^0$:

$$\begin{aligned} & \mathbb{E}[Y_{e,j,a}^1 | \mathbf{X}_{e,a}^1 = \mathbf{x}^1, \mathbf{B}_e = \mathbf{b}, D = 1] - \mathbb{E}[Y_{e,j,a}^0 | \mathbf{X}_{e,a}^0 = \mathbf{x}^0, \mathbf{B}_e = \mathbf{b}, D = 0] = \\ & \mathbb{E}[Y_{n,j,a} | \mathbf{X}_{n,a} = \mathbf{x}^1, \mathbf{B}_n = \mathbf{b}] - \mathbb{E}[Y_{n,j,a} | \mathbf{X}_{n,a} = \mathbf{x}^0, \mathbf{B}_n = \mathbf{b}]. \end{aligned} \quad (12)$$

In Appendix C.3.7, we test and do not reject all three hypotheses, singly and jointly for $a \leq a^*$.⁴⁵

3.3.8 Summarizing the Implications of Exogeneity and Structural Invariance

Collecting results, we obtain the following theorem:

⁴⁵This holds when pooling males and females and when testing separately by gender (see Appendix C.3.7).

Theorem 1 *Valid Out-of-Sample Forecasts*

Under Assumptions A-1-A-4, Conditions C-2 and C-3 hold for any value of $(\mathbf{X}_{k,a}^d, \mathbf{B}_k)$.

This is an immediate consequence of the cited assumptions. \square

3.3.9 Using Matching to Construct Virtual Treatment and Comparison Groups

Under exogeneity assumption A-3 and invariance condition A-4 we can use matching to construct counterparts to the experimental treatment and control groups in the auxiliary sample.⁴⁶ Doing so compresses the two stages of constructing a comparison group and creating forecasts into one stage. Matching in this fashion creates direct counterparts in the auxiliary samples for each member of the experimental samples. It is an intuitively appealing non-parametric estimator that is valid under exogeneity (Heckman and Navarro, 2004).

We discuss this approach in Appendix C.3.3. Matching is a non-parametric estimation procedure for conditional mean functions. There is close agreement between non-parametric estimates based on matching and more parametric model-based approaches used in most of this paper (see Appendix C.3.4.1).

⁴⁶Heckman et al. (1998) use this procedure.

Table 2: Net Present Value of Labor Income and Cost/Benefit Analysis Under Different Specifications for Labor Income Process

	Specification 1: ("Baseline") $\lambda_1 \neq 0$ $\rho \neq 0$ $f = 0$			Specification 2: $\lambda_1 \neq 0$ $\rho \neq 0$ $f = 0$			Specification 3: $\lambda_1 \neq 0$ $\rho \neq 0$ $f = 0$			Specification 4: $\lambda_1 \neq 0$ $\rho \neq 0$ $f \neq 0$			Specification 5: Non-parametric matching		
	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C
Pooled	636,674 (183,224)	0.14 (0.03)	7.33 (1.84)	154,547 (187,036)	0.15 (0.12)	7.31 (5.15)	268,179 (211,089)	0.26 (0.14)	12.68 (5.81)	46,953 (25,323)	0.05 (0.02)	2.22 (0.7)	132,924 (11,253)	0.13 (0.01)	6.28 (0.31)
Males	919,049 (287,442)	0.15 (0.04)	10.19 (2.93)	200,509 (160,988)	0.11 (0.05)	9.35 (5.51)	456,078 (358,534)	0.25 (0.12)	21.26 (12.28)	74,775 (54,752)	0.04 (0.02)	3.49 (1.88)	196,530 (20,210)	0.11 (0.01)	9.16 (0.69)
Females	161,759 (72,355)	0.10 (0.06)	2.61 (0.73)	79,441 (99,416)	0.19 (0.28)	4.64 (3.19)	31,303 (168,160)	0.07 (0.48)	1.83 (5.4)	19,959 (34,142)	0.05 (0.1)	1.17 (1.1)	69,317 (4,350)	0.17 (0.0 1)	4.05 (0.14)

Note: This table displays the net present value of labor income in 2014 USD (treatment - control) using the five different specifications for forecasts that are explained below. Specification 1 is our baseline estimate. It also presents the calculation of the internal rate of return and the benefit/cost ratio of the program using these different net present values. Specification 1: forecast based on lagged outcome; no serial autocorrelation; and no fixed effect. Specification 2: forecast based on lagged outcome; arbitrary serial autocorrelation; and no fixed effect. Specification 3: forecast based on lagged outcome; first-order serial autocorrelation; and no fixed effects. Specification 4: forecast based on lagged outcome; no serial autocorrelation; and fixed effect. Specification 5: Forecast based on non-parametric matching estimates described in Section 3.3.9.

Model: With the exception of Specification 5, the different specifications are particular cases of the following model:

$$\begin{aligned}
 Y_{k,j,a} &= \lambda_0 + \lambda_1 Y_{k,j,a-1} + \lambda_2 \mathbf{X}_{k,a} + \varepsilon_{k,j,a} \\
 \varepsilon_{k,j,a} &= \underbrace{f}_{\text{Person-Specific Effect}} + \underbrace{\omega_{k,j,a}}_{\text{Possibly Serially Correlated Component}} \\
 \omega_{k,j,a} &= \rho \omega_{k,j,a-1} + \underbrace{U_{k,j,a}}_{\text{Independent Innovation}}
 \end{aligned}
 \tag{13}$$

3.3.10 Exploring the Impact of Other Forecast Models

In Appendix C.6, we analyze the consequences of using more general forecast models that allow us to relax exogeneity Assumption A-3. We estimate standard panel data specifications. We summarize the results from these exercises for the case of forecasting labor income in Table 2 with a comprehensive note at the base of the table. Forecasts of present values exhibit little sensitivity across these different methods, and do not differ substantially from the prediction baseline model.

4 Monetizing Specific Components of the Benefits and Costs

This section discusses how we calculate the benefits and costs of health, parental income, crime, and the costs of the program. We discuss our measures of educational attainment and costs in Appendix D.

4.1 Health

A major contribution of this paper is forecasting and monetizing the life-cycle benefits of enhanced participant health using a version of Equation (1) including a full vector of lagged dependent variables for different indicators of health status. This requires adapting the models of Section 3.3, as well as estimating a dynamic model of competing health risks (Kalbfleisch and Prentice, 1980). Three additional issues arise: (i) health outcomes such as diabetes or heart disease are absorbing states; (ii) health outcomes are highly interdependent within and across time periods; and (iii) there is no obvious terminal time period for benefits and costs except death, which is endogenous.⁴⁷

Our auxiliary model for health is an adaptation of the Future America Model (FAM).

⁴⁷For example, we extrapolate labor income until the retirement age of 67. However, for health, we need to forecast an age of death for each individual before extrapolating.

This model forecasts health outcomes from the subjects’ mid 30s up to their projected age of death (Goldman et al., 2015).⁴⁸ Appendix F discusses the FAM methodology in detail.

FAM passes a variety of specification tests and accurately forecasts health outcomes and healthy behaviors.⁴⁹ We initialize the health forecast model using the same variables that we use to forecast labor and transfer income, along with the initial health conditions as listed in Table F.1.

Our methodology has five steps: (i) estimate age-by-age health state transition probabilities using the Panel Study of Income Dynamics (PSID); (ii) match these transition probabilities to the ABC/CARE subjects based on observed characteristics; (iii) estimate quality-adjusted life year (QALY) models using the Medical Expenditure Panel Survey (MEPS) and the PSID; (iv) estimate medical cost models using the MEPS and the Medicare Current Beneficiary Survey (MCBS), allowing estimates to differ by health state and observed characteristics; and (v) forecast the medical expenditures and QALYs that correspond to the simulated individual health trajectories.⁵⁰

Our microsimulation model starts with information on observed characteristics at age 30, along with the information on observed characteristics available at this age. Restricting it to the individuals for whom we have information from the mid 30s health survey allows us to account for components that are important for forecasting health outcomes. The models forecast the probability of being in any of the states in the horizontal axis of Table F.1 at age $a + 1$ based on the state at age a , which is described by the vertical axis of the table.⁵¹ Absorbing states are an exception. For example, heart disease at age a does not enter in the estimation of transitions for heart disease at age $a + 1$ because it is an absorbing state: once

⁴⁸The simulation starts at the age in which we observe the subjects’ age-30 follow-up.

⁴⁹In Appendix F.4, we present tests of the model’s assumptions and predictive performance for population aggregate health and healthy behavior outcomes.

⁵⁰As an intermediate step between (i) and (ii), we impute some of the variables used to initialize the FAM models (see Appendix F.3.1).

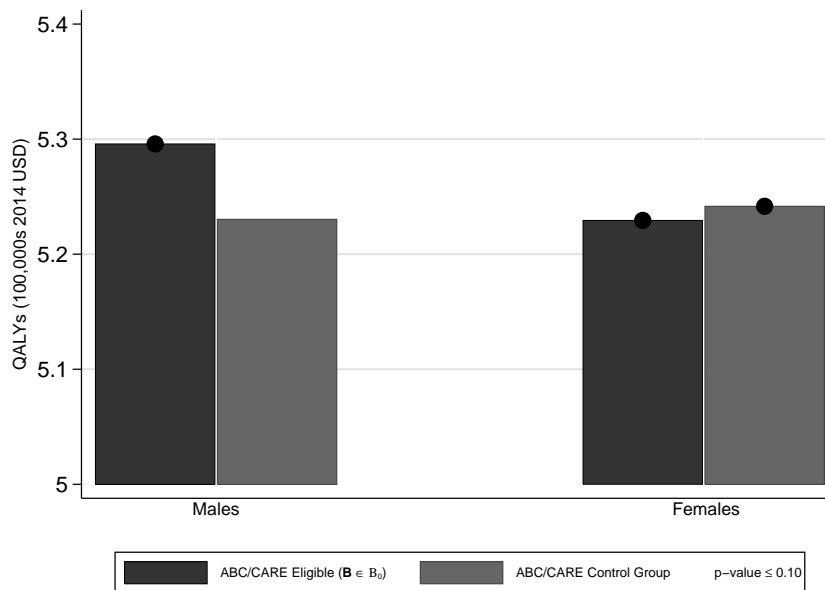
⁵¹In practice, the forecasts are based on two-year lags, due to data limitations in the auxiliary sources we use to simulate the FAM. For example, if the individual is 30 (31) years old in the age-30 interview, we simulate the trajectory of her health status at ages 30 (31), 32 (33), 34 (35), and so on until her projected death.

a person has heart disease, she carries it through the rest of her life.

At each age, once we obtain the transition probability for each health outcome, we make a Monte-Carlo draw for each subject. Each simulation depends on each individual’s health history and on her particular 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 then obtain an estimate of 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 2007–2010. The MEPS estimation captures costs during 2008–2010. To account for real medical cost growth after 2010, we adjust each model’s forecast using the method described in Appendix F.3.4.

Figure 6: Mean Quality-Adjusted Life Years: Forecasts and Comparison to PSID



Note: This figure displays the life-cycle net present value of forecasted quality-adjusted life years (QALYs) for ABC/CARE males and females in the control group. The forecasts are based on combining data from the Panel Study of Income Dynamics (PSID), the Health Retirement Study (HRS), and the Medical Expenditure Panel Survey (MEPS). For each gender, we display a comparison to disadvantaged males and females in the PSID, where disadvantaged is defined as being African American and having at most 12 years of education. QALYs are the quality-adjusted life years accounting for the burden of disease. The black dots indicate that the bars are significantly greater than 0 at the 10% level. This is calculated using the empirical bootstrap distribution.

The same procedure is applied to calculate quality-adjusted life years (QALYs). We

compute a QALY model based on a widely-used health-related Quality-of-Life (HRQoL) measure (EQ-5D), available in MEPS.⁵² We then estimate this model using the PSID.

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 expenditure before age 30 (see Appendix F.3.5). 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.

QALYs are crucial for our cost/benefit analysis because they monetize the health of an individual at each age. Figure 6 plots our estimates of mean QALYs together with a PSID comparison for the control sample in an exercise analogous to that used to produce Figure 4.⁵³ Although there is not a clear age-by-age treatment effect on QALYs, there is a statistically and substantively significant difference in the accumulated present value of the QALYs between the treatment and the control groups. The QALYs for female individuals in the control group match the QALYs of disadvantaged individuals in the PSID. For males, use of the PSID auxiliary sample to construct controls understates the net benefits of ABC/CARE.⁵⁴

⁵²For a definition and explanation of this instrument, see Dolan (1997); Shaw et al. (2005).

⁵³In our baseline estimation, we assume 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 G.

⁵⁴In Appendix F we further discuss and justify the parameterizations required to obtain estimates of QALYs. We only consider QALYs starting at age 30. Tysinger et al. (2015) examine the sensitivity to these parameterizations and discuss alternative micro-simulations monetizing health condition.

4.2 Forecasting 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 mothers of children reported living with a partner at baseline and this status 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 as reported in [García et al. \(2017\)](#). This arises from wage growth due to parental educational attainment and more work experience.⁵⁵

We observe parental labor income at eight different ages for the experimental subjects up through age 21.^{56,57} An ideal approach would be to estimate the profile over the full life-cycle of mothers. We propose two different approaches for doing this in Appendix C.3.8: (i) an approach based on parameterizing parental labor income using standard Mincer equations; and (ii) an approach based on the analysis of Section 3.3. In Section 5, we present estimates using the labor income through age 21 and using these two alternatives for projecting future labor income after age 21. The benefits of the program increase when considering the full life-cycles of mothers using either approach.

Any childcare inducements of the program likely benefit parents who, at baseline, did not have any other children. If they did, then they might have had to take care of other children anyway, weakening the childcare-driven effect, especially if there are younger siblings present. In Appendix C.3.8, we show that the treatment effect for discounted parental labor income is much higher when there are no siblings of the participant children at baseline. The effect also weakens when comparing children who have siblings younger than 5 years old to children who have siblings 5 years old or older.⁵⁸

⁵⁵There is also an effect on maternal school enrollment. Some of the mothers decided to further enroll in school, obtained a higher degree, and that could be one of the reasons why they make more money afterward. We report these treatment effects in Appendix A. We quantify the cost of this additional education using the strategy in Appendix D.

⁵⁶The 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.

⁵⁷We linearly interpolate parental labor income for ages for which we do not have observations between 0 and 21.

⁵⁸These patterns persist when splitting the ABC/CARE sample by gender, but the estimates are not

4.3 Crime

To estimate the life-cycle benefits and costs of ABC/CARE related to criminal activity, we use rich data on crime outcomes obtained from public records.⁵⁹ See Appendix E for a more complete discussion. We consider the following types of crime: arson, assault, burglary, fraud, larceny, miscellaneous (which includes traffic and non-violent drug crimes which can lead to incarceration), murder, vehicle theft, rape, robbery, and vandalism. We use administrative data that document: (i) youth arrests, gathered at the age-21 follow-up; (ii) adult arrests, gathered at the mid 30s follow-up; and (iii) sentences, gathered at the mid 30s follow-up. We also use self-reported data on adult crimes, gathered in the age-21 and age-30 subject interviews. Because none of these sources capture all criminal activity, it is necessary to combine them to more completely approximate the crimes the subjects committed. We also use several auxiliary datasets to complete the life-cycle profile of criminal activity and compute the costs of the committed crimes (see Appendix E.1).

We follow four steps to estimate the costs of crime.

1. *Count arrests and sentences.* We start by counting the total number of sentences for each individual and type of crime (arson, assault, etc.) up to the mid 30s, matching crimes across data sources, to construct the total number of arrests for each individual and type of crime up to the mid 30s.⁶⁰ For individuals missing arrest data,⁶¹ we impute

precise because the samples become too small. See Appendix C.3.8.

⁵⁹Two 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 differences between the treatment and the control groups. [Barnett and Masse \(2002, 2007\)](#) account for self-reported crime at age 21. They find weak effects based on self-reports, but they lack access to longer term, administrative data. The novelty of our study with respect to crime does not only consist of using administrative data allowing us to know the accumulated number of crimes that the children commit once they were in their mid 30s. It is also novel because we use micro-data specific to the state in which these individuals grew up, as well as other national datasets, to forecast criminal activity from the mid 30s to 50.

⁶⁰In practice, we count all offenses (an arrest might include multiple offenses). This gives the correct number of victims for our estimations. The youth data have coarser categories than the rest of the data: violent, property, drug, and other. To match these data with the adult data, we assume that all property crimes were larcenies and that all violent crimes are assaults. In the ABC/CARE sample, assault is the most common type of violent crime, and larceny/theft is the most common property crime.

⁶¹About 10% of the ABC/CARE sample has missing arrest data. We fail to reject the null hypothesis of

the number of arrests by multiplying the number of sentences for each type of crime by a national arrest-sentence ratio for the respective crime.⁶²

2. *Construct forecasts.* Based on the sentences observed before the mid 30s, we forecast the sentences that the ABC/CARE subjects will have after their mid 30s. Data from the North Carolina Department of Public Safety (NCDPS), which provide life-cycle sentences of individuals in North Carolina, are used to estimate sentences incurred after the mid 30s from sentences incurred before then. Applying these models to the ABC/CARE data, we forecast the number of future sentences for each subject up to age 50.⁶³ We then add these estimates to the original number of sentences, getting an estimate of the life-cycle sentences. Adding these estimates increases the total count of crimes by 30%-50%.
3. *Estimate number of victims from the crimes.* We only observe crimes that resulted in consequences in the justice system: crimes that resulted in arrests and/or sentences. To include unobserved crimes, we use victimization inflation (VI).⁶⁴ We start by constructing a VI ratio, which is the national ratio of victims to arrests for each type of crime.⁶⁵ Then, we estimate the number of victims from the crimes committed by ABC/CARE subjects as their total arrests multiplied by the VI ratio.⁶⁶
4. *Find total costs of crimes.* We use the estimates of the cost of crimes for victims from [McCollister et al. \(2010\)](#) to impute the total victimization costs. For crimes resulting

no differences in observed characteristics between the treatment- and control-group participants for whom we observe arrests data (see Appendix A.6).

⁶²This arrest-sentence ratio is constructed using the National Crime Victimization Survey (NJR) and the Uniform Crime Reporting Statistics (UCRS).

⁶³We assume that individuals with no criminal records before their mid 30s commit no crimes thereafter.

⁶⁴Previous papers using this method include [Belfield et al. \(2006\)](#) and [Heckman et al. \(2010b\)](#).

⁶⁵We assume that each crime with victims is counted separately in the national reports on arrests, even for arrests that might have been motivated by more than one crime. This victim-arrest ratio is constructed using the NJRP and the National Crime Victimization Survey (NCVS).

⁶⁶Additionally, we can calculate an analogous estimate of the number of crime victims using sentences, based on the VI ratio and the national arrest-sentence ratio. These estimates are very similar, as shown in Appendix E.3. To improve precision, the estimates in the rest of our paper are based on the average of the two calculations.

in arrests and/or sentences, we consider criminal justice system costs as well, such as police costs.⁶⁷ Finally, we construct the total costs of incarceration for each subject using the total prison time and the cost of a day in prison.⁶⁸

4.4 Program Costs

The yearly cost of the program was \$18,514 per participant in 2014 USD. We improve on previous cost estimates using primary-source documents.⁶⁹ Appendix B discusses the program costs in detail. This section reports benefit/cost and rate of return analyses underlying Figure 1. Appendix G displays an extensive sensitivity analysis of each of the components we consider. It includes scenarios in which all of our assumptions hold and scenarios in which they are violated, providing bounds for our estimates.

5 Estimating the Benefit/Cost Ratio and the Internal Rate of Return

We first present our main estimates and then conduct sensitivity analyses.⁷⁰ Table 3 gives our baseline estimates of benefit/cost ratios and the sensitivity of the estimates to alternative

⁶⁷To be able to assign costs to each type of crime, we assume that the cost of the justice system depends on the number of offenses of each type, rather than on the number of arrests. While this could very slightly overestimate justice system costs, these costs only represent about 5% of the total crime costs.

⁶⁸Appendix G examines the sensitivity of our crime costs quantification to different assumptions. Section 5 and Appendix G examine the sensitivity of our overall assessment of ABC/CARE results to the quantification of crime that we explain in this section.

⁶⁹Our 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).

⁷⁰See Appendices C.4 and C.5 for more details on these estimations. We note that one limitation of our analysis is that data limitations prevent us from jointly estimating all of the behavioral relationships considered in this paper. Different components of the net benefits are estimated on different data sets.

assumptions. Table 4 presents the corresponding internal rates of return. Pooling males and females, the results indicate that the program is socially efficient: the internal rate of return and the benefit/cost ratio are 13.7% and 7.3. *Our baseline estimates indicate that the program generates a benefit of 7.3 dollars for every dollar spent on it.* These estimates are statistically significant, even after accounting for sampling variation, serial correlation, and forecast error in the experimental and auxiliary samples and the tax costs of financing the program.⁷¹ These benefits arise despite the fact that ABC/CARE was much more expensive than other early childhood education programs because the program used more services over a longer time period.

We accompany these estimates with an extensive set of sensitivity checks of statistical and economic interest. Our estimates are not driven by our methods for accounting for attrition and item non-response, by the conditioning variables, or the functional forms of projection equations used when computing the net-present values.⁷² 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 only (forecast span columns). Accounting for the treatment substitutes available to controls also matters. Males benefit the most from ABC/CARE relative to attending alternative childcare, while females benefit the most from ABC/CARE relative to staying at home. We explore this difference below.

Our baseline estimates account for the deadweight loss caused by distortionary taxes to fund programs, plus the direct costs associated with collecting taxes.⁷³ We assume a marginal tax rate of 50%.⁷⁴ Our estimates are robust to dropping it to 0% or doubling it

⁷¹We 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.8).

⁷²See Appendix C for a detailed discussion.

⁷³When 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 each dollar spent (e.g., public education).

⁷⁴Feldstein (1999) reports that the deadweight loss caused by increasing existing tax rates (marginal

Table 3: Sensitivity Analysis for Benefit/Cost Ratios

	<i>Pooled</i>			<i>Males</i>			<i>Females</i>		
Baseline	7.33 (s.e. 1.84)	7.99 (2.18)	9.80 (2.69)	10.19 (s.e. 2.93)	8.83 (2.72)	2.57 (0.72)	2.61 (s.e. 0.73)	2.57 (0.72)	2.82 (0.68)
<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	<i>No IPW</i>	<i>and No Controls</i>	<i>No IPW</i>	<i>and No Controls</i>	<i>No IPW</i>	<i>and No Controls</i>	<i>No IPW</i>	<i>and No Controls</i>	<i>and No Controls</i>
Prediction Span	<i>to Age 21</i>	<i>to Age 30</i>	<i>to Age 21</i>	<i>to Age 30</i>	<i>to Age 21</i>	<i>to Age 30</i>	<i>to Age 21</i>	<i>to Age 30</i>	<i>to Age 30</i>
Counterfactuals	<i>vs. Stay at Home</i>	<i>vs. Alt. Presch.</i>	<i>vs. Stay at Home</i>	<i>vs. Alt. Presch.</i>	<i>vs. Stay at Home</i>	<i>vs. Alt. Presch.</i>	<i>vs. Stay at Home</i>	<i>vs. Alt. Presch.</i>	<i>vs. Alt. Presch.</i>
Deadweight-loss	0%	100%	0%	100%	0%	100%	0%	100%	100%
Discount Rate	17.40 (5.90)	2.91 (0.59)	15.38 (4.35)	5.50 (1.37)	3.78 (0.79)	3.83 (1.04)	5.06 (2.82)	3.83 (1.04)	2.01 (0.59)
Parental Income	7.63 (1.84)	7.73 (1.92)	10.46 (2.94)	7.73 (1.92)	10.63 (2.95)	2.98 (0.76)	2.98 (0.76)	2.98 (0.76)	3.12 (0.85)
Labor Income	7.01 (1.80)	7.66 (1.90)	9.58 (2.66)	7.66 (1.90)	10.79 (3.24)	2.51 (0.70)	2.51 (0.70)	2.51 (0.70)	2.71 (0.75)
Crime	4.24 (1.10)	5.18 (1.22)	7.41 (3.43)	5.18 (1.22)	7.12 (2.41)	2.61 (0.67)	2.61 (0.67)	2.61 (0.67)	2.47 (0.66)
Health (QALYs)	6.48 (1.79)	8.19 (2.13)	9.14 (2.73)	8.19 (2.13)	11.23 (3.40)	2.20 (0.69)	2.20 (0.69)	2.20 (0.69)	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.9 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 Section 3.1 for a discussion). Deadweight loss is the loss implied by any public expenditure (0% is no loss and 100% is one dollar less per each 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 5.

Table 4: Sensitivity Analysis for Internal Rate of Return, ABC/CARE

	<i>Pooled</i>		<i>Males</i>		<i>Females</i>	
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%)
Prediction Span	<i>to Age 21</i> 8.8% (4.5%)	<i>to Age 30</i> 12.0% (3.4%)	<i>to Age 21</i> 11.8% (4.8%)	<i>to Age 30</i> 12.8% (4.7%)	<i>to Age 21</i> 10.7% (5.8%)	<i>to Age 30</i> 11.7% (5.2%)
Counter-factuals	<i>vs. Stay at Home</i> 9.4% (4.2%)	<i>vs. Alt. Presch.</i> 15.6% (4.3%)	<i>vs. Stay at Home</i> 6.0% (3.6%)	<i>vs. Alt. Presch.</i> 15.8% (5.0%)	<i>vs. Stay at Home</i> 13.4% (5.7%)	<i>vs. Alt. Presch.</i> 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	<i>Mincer Life-cycle</i> 15.2% (4.0%)	<i>Life-cycle Prediction</i> 14.5% (6.4%)	<i>Mincer Life-cycle</i> 16.0% (5.1%)	<i>Life-cycle Prediction</i> 14.5% (6.4%)	<i>Mincer Life-cycle</i> 13.3% (8.2%)	<i>Life-cycle Prediction</i> 12.3% (9.9%)
Labor Income	<i>.5% Annual Decay</i> 13.5% (3.4%)	<i>.5% Annual Growth</i> 13.8% (3.2%)	<i>.5% Annual Decay</i> 14.5% (4.3%)	<i>.5% Annual Growth</i> 14.8% (4.1%)	<i>.5% Annual Decay</i> 9.9% (6.0%)	<i>.5% Annual Growth</i> 10.3% (6.0%)
Crime	<i>Drop Major Crimes</i> 10.7% (4.4%)	<i>Halve Costs</i> 11.6% (3.8%)	<i>Drop Major Crimes</i> 12.0% (5.3%)	<i>Halve Costs</i> 11.9% (4.9%)	<i>Drop Major Crimes</i> 10.1% (6.0%)	<i>Halve Costs</i> 9.9% (6.0%)
Health (QALYs)	<i>Drop All</i> 12.8% (4.6%)	<i>Double Value of Life</i> 13.5% (3.6%)	<i>Drop All</i> 13.5% (5.6%)	<i>Double Value of Life</i> 14.4% (4.6%)	<i>Drop All</i> 8.8% (6.4%)	<i>Double Value of Life</i> 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.9 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 Section 3.1 for a discussion). Deadweight loss is the loss implied by any public expenditure (0% is no loss and 100% is one dollar loss per each 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 5.

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 is an important component of the benefit/cost ratio. We take a conservative approach in our baseline estimates and do not account for potential shifts in profiles in parental labor income due to education and work experience subsidized by childcare (see the discussion in Section 4.2). 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 increase in the gain due to parental labor income (see parental labor income columns).

As noted in Section 3, the baseline estimates ignore cohort effects. Individuals in ABC/CARE could experience positive cohort effects that might (i) make them more productive and therefore experience wage growth (Lagakos et al., 2016); (ii) experience a negative shock such as an economic crisis and therefore experience a wage decline (Jarosch, 2016). Our estimates are robust when we vary annual growth and decay rates between -0.5% and 0.5% .⁷⁵

deadweight loss) may exceed two dollars per each dollar of revenue generated. We use a more conservative value (0.5 dollars per each dollar of revenue generated). In Tables 3, 4, and 5 and in Appendix G.2, we explore the robustness of this decision and find little sensitivity.

⁷⁵We account for cohort effects in health as explained in Section 4.1.

Table 5: Cost/Benefit Analysis of ABC/CARE, Summary

Removed Component	Females			Males			Pooled		
	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C
None	161,759	10.1% (6%)	2.61 (0.73)	919,049	14.7% (4%)	10.19 (2.93)	636,674	13.7% (3%)	7.33 (1.84)
Parental Income	148,854	4% (2%)	1.12 (0.65)	107,907	11% (3%)	9.10 (2.92)	116,953	9% (3%)	6.17 (1.87)
Subject Labor Income	41,908	9% (6%)	2.21 (0.66)	238,105	13% (5%)	7.75 (2.23)	133,032	13% (4%)	6.03 (1.77)
Subject Transfer Income	419	10% (6%)	2.61 (0.73)	-7,265	15% (4%)	10.26 (2.93)	-4,372	14% (3%)	7.38 (1.84)
Subject QALY	42,102	9% (6%)	2.20 (0.69)	106,218	14% (6%)	9.14 (2.73)	87,181	13% (5%)	6.48 (1.79)
Medical Expenditures	-16,037	9% (6%)	2.77 (0.76)	-42,038	15% (3%)	10.61 (2.89)	-31,221	14% (3%)	7.65 (1.85)
Alternative Preschools	16,691	8% (5%)	2.45 (0.73)	13,434	14% (4%)	10.05 (2.92)	14,659	12% (3%)	7.19 (1.84)
Education Costs	1,457	10% (6%)	2.59 (0.72)	-7,852	15% (4%)	10.26 (2.93)	-4,518	14% (3%)	7.37 (1.86)
Crime Costs	31,668	10% (6%)	2.34 (0.62)	638,923	9% (5%)	4.08 (2.18)	450,368	8% (4%)	3.06 (1.01)
Deadweight Loss		18% (12%)	3.83 (1.04)		19% (6%)	15.38 (4.35)		18% (5%)	11.01 (2.79)
0% Discount Rate			5.06 (2.82)			25.45 (10.42)			17.40 (5.90)
7% Discount Rate			1.49 (0.32)			3.78 (0.79)			2.91 (0.59)

Note: This table presents the estimates of the net present value (NPV) for each component, and the internal rate of return (IRR) and the benefit/cost ratio (B/C) of ABC/CARE for different scenarios based on comparing the groups randomly assigned to receive center-based childcare and the groups randomly assigned as control in ABC/CARE. The first row represents the baseline estimates. The other rows present estimates for scenarios in which we remove the NPV estimates of the component listed in the first column. The category “Alternative Preschools” refers to the money spent in alternatives to treatment from the control-group children parents. QALYs refers to the quality-adjusted life years. Any gain corresponds to better health conditions through the age of death. The quantity listed in the NPV columns is the component we remove from NPV when computing the calculation in each row. All the money figures are in 2014 USD and are discounted to each child’s birth, unless otherwise specified. For the B/C ratio we use a discount rate of 3%, unless otherwise specified. We test the null hypotheses $IRR = 3\%$ and $B/C = 1$ —we select 3% as the benchmark null because that is the discount rate we use. Inference is based on non-parametric, one-sided p -values from the empirical bootstrap distribution. We highlight point estimates significant at the 10% level. Total cost of the program per child is \$92,570 (2014).

We also examine the sensitivity of our estimates to (i) dropping the most costly crimes such as murders and rapes,⁷⁶ 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 victimization costs are somewhat subjective (see Appendix E.3). Our cost/benefit 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

⁷⁶Two individuals in the treatment group were convicted of rape and one individual in the control group was convicted of murder.

important component because healthier individuals survive longer, and treatment improves health conditions. It is important to note that this component largely accumulates later in life and therefore is heavily discounted. Dropping the component or doubling the value of life does not have a major impact on our calculations.

The estimates are robust when we conduct a rather drastic sensitivity analysis by removing components of the cost/benefit analysis entirely (see Table 5 and Figure 2).⁷⁷ Even when completely removing the gain associated with crime for males, the program is socially efficient—both the internal rate of return and the benefit/cost ratio are substantial. Parental labor income and crime are the components for which the internal rate of return and the benefit/cost ratio are the most sensitive. The reason for this sensitivity to parental labor income is that the amount is substantial and it is not heavily discounted because it accumulates during the first 21 years of the subjects’ lives. Crime occurs later in life and its benefits are discounted accordingly. The amount due to savings in crime is large, so removing it diminishes both the internal rate of return and the benefit/cost ratio (but they remain statistically significant).

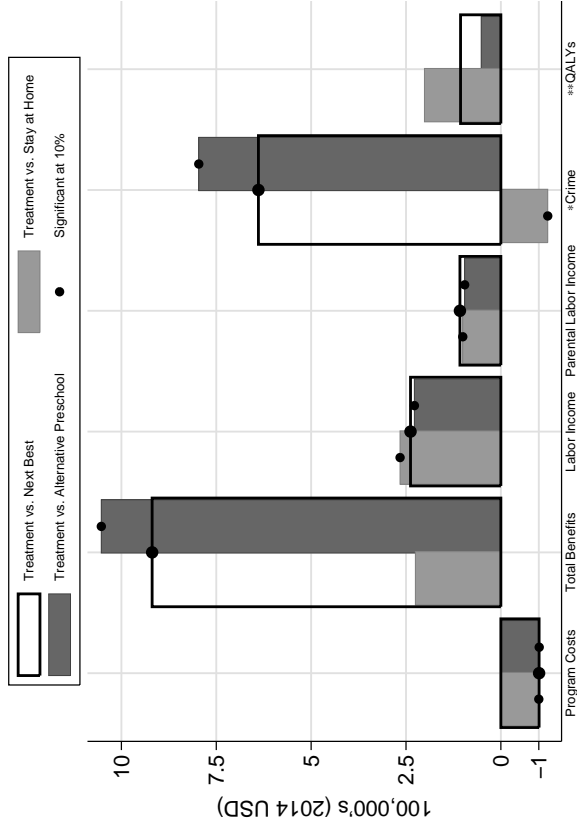
In Appendix C.6, we investigate how sensitive our forecast model is to a variety of perturbations: different autocorrelation processes in the forecast errors, functional forms of the prediction equations, forecasts without lagged variables, etc. Our estimates are robust to using different forecast models.

Overall, our sensitivity analyses indicate that no single category of outcomes drives the social efficiency of the program (see Figures 2 and 3). Rather, it is the life-cycle benefits across multiple dimensions of human development.

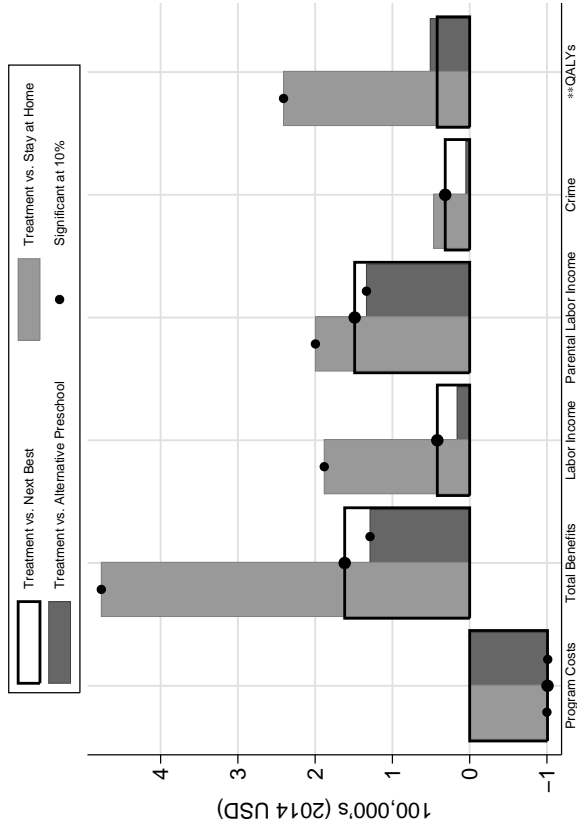
⁷⁷In Appendix G.3, we present exercises that are not as drastic as removing the whole component, but instead remove fractions of it.

Figure 7: Life-cycle Net Present Value of Main Components of the CBA

(a) Males



(b) Females



Note: This figure displays the life-cycle net present values 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%. "Treatment vs. Control": compares the treatment to the control group. "Treatment vs. Stay at Home": compares the treatment group to those subjects who stayed at home. "Treatment vs. Alternative Preschool": compares the treatment group to those subjects who attended alternative preschools. The latter two are based on matching estimators that account for selection on observable variables. 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 net benefits: are for *all* of the components we consider. Labor income: total individual labor income from ages 20 to the retirement of program participants (assumed to be 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 parents of control group children; (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. We indicate point estimates significant at the 10% level.

*The treatment vs. stay at home net present value is sizable and negative (-\$123,498,.2); its standard error is \$62,745.72.

**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.

6 Using Our Estimates to Understand 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. \(2011\)](#). Although widely emulated, this approach offers an imprecise approximation of benefit/cost ratios with questionable validity. Examples of application of this approach include [Attanasio et al. \(2011\)](#), [Behrman et al. \(2011\)](#), and [Deshpande and Yue \(2017\)](#).

[Kline and Walters \(2016\)](#) use data from the Head Start Impact Study (HSIS) and report a benefit/cost ratio between 1.50 and 1.84.⁷⁸ Their analysis proceeds in three steps: (i) calculate program treatment effects on IQ measured around age 5⁷⁹; (ii) monetize this gain using the return to the IQ 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. \(2011\)](#).^{80,81,82}; and (iii) calculate the benefit/cost ratio based on this gain and their own calculations of the program’s cost.^{83,84}

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

In the first exercise, 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

⁷⁸HSIS is a one-year-long randomized evaluation of Head Start.

⁷⁹They use an index based on the Peabody Picture Vocabulary and Woodcock Johnson III Tests.

⁸⁰The [Chetty et al. \(2011\)](#) return is based on Stanford Achievement Tests.

⁸¹For this comparison exercise, we interpret the earnings estimated in [Chetty et al. \(2011\)](#) to be equivalent to labor income.

⁸²Calculations from [Chetty et al. \(2011\)](#) indicate that a 1 standard deviation gain in IQ 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.

⁸³Their 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. \(2011\)](#).

⁸⁴All 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: Alternative Cost/Benefit Analyses Calculations

Age	NPV Source	Component	Kline and Walters (2016) Method	Authors' Method
27	Chetty et al. (2011)	Labor income	0.58 (s.e. 0.28)	
	ABC/CARE-calculated	Labor income	0.09 (s.e. 0.04)	1.09 (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.

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 27.⁸⁵ 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 is a valid account of underlying uncertainties when compared to Kline and Walters (2016), who do not account for estimation error in reporting standard errors. The return is smaller because our sample is much more disadvantaged than that of Chetty et al. (2011).

The remaining exercises are similar, but we (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 benefit/cost ratio. The final reported estimate, 7.33, is our baseline estimate that incorporates all of the components across the life cycle of the subjects.

Our methodology provides a more accurate estimate of the net present value (and the return to IQ) of the program. We better quantify the effects of the experiment by considering

⁸⁵This allows us to compute our own "return to IQ" and impute it to the treatment-group individuals.

benefits over the whole life cycle. We also better approximate the statistical uncertainty of our estimates by considering both the sampling error in the experimental and auxiliary samples and the forecast error due to the interpolation and extrapolation. Proceeding in this fashion enables us to examine the sensitivity of each of the components in our study that we monetize.

7 Summary

This paper goes beyond analyzing batteries of short-term of treatment effects on test scores to evaluate early education programs—a standard practice in the current literature. Based on the large array of treatment effects from ages 0 to mid 30s reported in [García et al. \(2017\)](#), we combine experimental and non-experimental datasets to forecast long-term outcomes. We use these forecasts to compute statistics summarizing the social efficiency of the influential and widely implemented program that we evaluate.

Our analysis accounts for multiple sources of statistical and modeling uncertainty. We go beyond current analyses that only quantify a single, short-term treatment effect to forecast labor income without providing a sense of statistical and modeling uncertainty. We quantify long-term costs and benefits and provide a life-cycle analysis. We quantify and monetize health outcomes, a novel outcome in the evaluation of early childhood programs.

We demonstrate the benefits of economic and econometric theory, and auxiliary data sets, for evaluating the long-term benefits of a social experiment. Our forecasts rely on economic models and on testable assumptions about functional forms and endogeneity of inputs that we test. We show the robustness of our estimates through extensive empirical sensitivity analyses. We produce baseline estimates and then provide a wide array of estimates, for readers to analyze what the estimates would be under assumptions that they found most plausible. Our estimates of the internal rate of return (benefit/cost ratio) range from 8.0% to 18.3% (1.52 to 17.40). Investing in this program is highly socially profitable.

References

- Agostinelli, F. and M. Wiswall (2016, July). Estimating the technology of children’s skill formation. Working Paper, W. P. Carey School of Business, University of Wisconsin-Madison, Arizona State University.
- 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.
- Baker, M., J. Gruber, and K. Milligan (2015, September). Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program. Working Paper 21571, National Bureau of Economic Research.
- 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.
- Burchinal, M. R., F. A. Campbell, D. M. Bryant, B. H. Wasik, and C. T. Ramey (1997, October). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development* 68(5), 935–954.
- 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.
- 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.

- Collins, A., B. D. Goodson, J. Luallen, 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, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78(3), 883–931.
- Currie, J. (2011). Inequality at birth: Some causes and consequences. *American Economic Review* 101(3), 1–22.
- Department of Health, Education, and Welfare (1968). Federal interagency day care requirements, pursuant to Sec. 522 (D) of the Economic Opportunity Act. Technical report, U.S. Department of Labor, Washington, DC.
- Deshpande, M. and L. Yue (2017). Who is screened out? application costs and the targeting of disability program. Unpublished.
- Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care* 35(11), 1095–1108.
- Duncan, G. J. and A. J. Sojourner (2013). Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *Journal of Human Resources* 48(4), 945–968.
- 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., J. J. Heckman, and A. L. Ziff (2017). Gender differences in the effects of early childhood education. 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.
- Gross, R. T., D. Spiker, and C. W. Haynes (1997). *Helping Low Birth Weight, Premature Babies: The Infant Health and Development Program*. Stanford, CA: Stanford University Press.
- Haavelmo, T. (1943, January). The statistical implications of a system of simultaneous equations. *Econometrica* 11(1), 1–12.
- 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., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, Chapter 7, pp. 307–458. Amsterdam: Elsevier.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010a, August). 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). Causal analysis after Haavelmo. *Econometric Theory* 31(1), 115–151.
- 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.

- Heckman, J. J. and R. Robb (1985, October-November). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30(1–2), 239–267.
- Jarosch, G. (2016). Searching for job security and the consequences of job loss. Forthcoming, *Econometrica*.
- 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).
- Kalbfleisch, J. D. and R. L. Prentice (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- 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.
- Kottelenberg, M. J. and S. F. Lehrer (2014, August). The gender effects of universal child care in Canada: Much ado about boys? Unpublished manuscript, Department of Economics, Queen’s University.
- Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2016). Life-cycle wage growth across countries. Forthcoming, *Journal of Political Economy*.
- 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.
- Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press for National Bureau of Economic Research.
- North Carolina General Assembly (1971). Chapter 803, House bill 100. In *North Carolina General Assembly 1971 Session*, North Carolina. An Act to protect children through licensing of day-care facilities and other limited regulation.
- Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* 8(4), 431–440.
- Quandt, R. E. (1972, June). A new approach to estimating switching regressions. *Journal of the American Statistical Association* 67(338), 306–310.
- 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. and F. A. Campbell (1979, February). Compensatory education for disadvantaged children. *The School Review* 87(2), 171–189.

- Ramey, C. T., F. A. Campbell, M. Burchinal, M. L. Skinner, D. M. Gardner, and S. L. Ramey (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science* 4(1), 2–14.
- 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., M. C. Holmberg, J. H. Sparling, and A. M. Collier (1977). An introduction to the Carolina Abecedarian Project. In B. M. Caldwell and D. J. Stedman (Eds.), *Infant Education: A Guide for Helping Handicapped Children in the First Three Years*, Chapter 7, pp. 101–121. New York: Walker and Company.
- Ramey, C. T., G. D. McGinness, L. Cross, A. M. Collier, and S. Barrie-Blackley (1982). The Abecedarian approach to social competence: Cognitive and linguistic intervention for disadvantaged preschoolers. In K. M. Borman (Ed.), *The Social Life of Children in a Changing Society*, Chapter 7, pp. 145–174. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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.
- Schore, A. N. (2017, January/February). All our sons: The developmental neurobiology and neuroendocrinology of boys at risk. *Infant Mental Health Journal* 38(1), 15–52.
- Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, and M. Nores (2005). *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*, Volume 14 of *Monographs of the HighScope Educational Research Foundation*. Ypsilanti, MI: High/Scope Press.
- 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.
- Tysinger, B., D. Ermini Leaf, and D. P. Goldman (2015). The Future America Model: Technical Documentation. Technical report, Working Document.
- 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.
- Weiland, C. and H. Yoshikawa (2013). Impacts of a prekindergarten program on children’s mathematics, language, literacy, executive function, and emotional skills. *Child Development* 84(6), 2112–2130.
- 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.