NBER WORKING PAPER SERIES

# THE LIFE-CYCLE BENEFITS OF AN INFLUENTIAL EARLY CHILDHOOD PROGRAM

Jorge Luis García
James J. Heckman
Duncan Ermini Leaf
María José Prados

Working Paper 22993
http://www.nber.org/papers/w22993

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2016

This research was supported in part by: the Buffett Early Childhood Fund; the Pritzker Children‘s Initiative; the Robert Wood Johnson Foundation‘s Policies for Action program; the Leonard D. Schaeffer Center for Health Policy and Economics; NIH grants NICHD R37HD065072, NICHD R01HD054702, NIA R24AG048081, and NIA P30AG024968; Successful Pathways from School to Work, an initiative of the University of Chicago‘s Committee on Education and funded by the Hymen Milgrom Supporting Organization; the Human Capital and Economic Opportunity Global Working Group, an initiative of the Center for the Economics of Human Development and funded by the Institute for New Economic Thinking; and the American Bar Foundation. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders, the official views of the National Institutes of Health, nor the views of the National Bureau of Economic Research. 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, Sidharth Moktan, Tanya Rajan, Azeem Shaikh, Jeffrey Smith, Matthew Tauzer, and Ed Vytlacil. We also benefited from helpful comments received at seminars at the Russell Sage Foundation in September, 2016 and at the Leonard D. Schaeffer Center for Health Policy and Economics in December, 2016. 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 Web Appendix for this paper is posted on http://cehd.uchicago.edu/ABC_CARE.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

The Life-cycle Benefits of an Influential Early Childhood Program
Jorge Luis García, James J. Heckman, Duncan Ermini Leaf, and María José Prados
NBER Working Paper No. 22993
December 2016
JEL No. C93,I28,J13


## ABSTRACT

This paper estimates the long-term benefits from an influential early childhood program targeting disadvantaged families. The program was evaluated by random assignment and followed participants through their mid-30s. It has substantial beneficial impacts on health, children's future labor incomes, crime, education, and mothers' labor incomes, with greater monetized benefits for males. Lifetime returns are estimated by pooling multiple data sets using testable economic models. The overall rate of return is 13.7% per annum, and the benefit/cost ratio is 7.3. These estimates are robust to numerous sensitivity analyses.



Jorge Luis García
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago IL 60637
jorgelgarcia@uchicago.edu

James J. Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and IZA
and also NBER
jjh@uchicago.edu

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



A data appendix is available at http://www.nber.org/data-appendix/w22993

This paper estimates the large array of life-cycle benefits of an influential early childhood program targeted to disadvantaged children. The program has substantial impacts on the lives of its participants. Monetizing benefits and costs across multiple domains, we estimate a rate of return of 13.7% per annum and a benefit/cost ratio of 7.3. There are substantial differences across genders favoring males.

Our analysis contributes to a growing literature on the value of early-life programs for disadvantaged children.[1] Long-term evidence on their effectiveness is surprisingly limited.[2] For want of follow-up data, many studies of early childhood programs report few outcomes for early ages after program completion, e.g. IQ scores, school readiness measures.[3] Yet it is the long-term returns that are relevant for policy analysis.

We analyze the costs and benefits from two virtually identical early childhood programs evaluated by randomized trials conducted in North Carolina. The programs are the Carolina Abecedarian Project (ABC) and the Carolina Approach to Responsive Education (CARE)—henceforth ABC/CARE. Both were launched in the 1970s and have long-term follow-ups through the mid 30's. The programs started early (at 8 weeks of life) and engaged participants to age 5. We analyze their impacts on a variety of life outcomes such as health, the quality of life,[4] participation in crime, labor income, IQ, schooling, and

[1] See, e.g., Currie (2011) and Elango et al. (2016).

[2] The 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).

[3] See, e.g., Kline and Walters (2016) and Weiland and Yoshikawa (2013).

[4] Throughout this paper, *we refer to health-related quality of life as quality of life.* It is

increased parental labor income arising from subsidized childcare.[5]

Evidence from these programs is relevant for contemporary policy discussions because their main components are present in a variety of current interventions.[6] About 19% of all African-American children are eligible for these programs today.[7]

Analyzing the 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 adult capabilities. We use a variety of measures to characterize program benefits. Instead of reporting only 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 to publicly finance them. We display the sensitivity of our estimates excluding various components of costs and benefits.[8]

---

the weight attached to each year of life as a function of disease burden, as we discuss further below.

[5]The 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.

[6]Programs 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—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). Educare programs are also based on ABC/CARE (Educare, 2014; Yazejian and Bryant, 2012).

[7]43% of African-American children were eligible in 1972. (Author's calculation using the Panel Study of Income Dynamics (PSID).)

[8]Barnett and Masse (2002, 2007) present a cost/benefit analysis for ABC through age 21, before many benefits are realized. They report a benefit/cost ratio of 2.5, but give

A fundamental problem in evaluating any intervention is assessing out-of-sample future costs and benefits. 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 and are otherwise comparable to treated and control persons to proxy future treatment effects.[9] Using this approach, we combine experimental data through the mid 30's with information from multiple auxiliary panel data sources to predict benefits and costs over the lifetimes of participants.[10]

Our analysis is simplified by the fact that all eligible families offered participation in the program took the offer. This motivates a revealed preference approach to constructing synthetic control groups. We develop a synthetic treatment group drawing on and extending the analysis of Heckman et al. (2013). They show that program treatment effects are produced through changes in inputs in a stable (across treatment regimes) production function for outcomes. We use the estimated production function to make out-of-sample predictions based on inputs caused by treatment.

We account for sampling uncertainty arising from combining data, estimating parameters of behavioral equations, and simulating models. We con-

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 30's, 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 not available to them.

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

[10] Ridder 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).

duct sensitivity analyses for outcomes for which sampling uncertainty is not readily quantified. Our approach to combining multiple data sets and analyzing blocks of outcomes is of interest in its own right as a template for evaluating other programs with numerous long-run outcomes using intermediate outcome measures.

Our analysis accounts for control group substitution.[11] Roughly 75% of the control-group children in ABC/CARE enroll in some form of lower quality alternative childcare outside of the home.[12] We define and estimate parameters accounting for the choices taken by the control groups in our study.

We find pronounced gender differences in treatment effects comparing high quality treatment with lower quality alternatives. Males benefit much less from alternative market childcare arrangements compared to females, a result consistent with the literature on the vulnerability of male children when removed from their mothers, even for short periods.[13]

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.[14] There are benefits for participants in terms of reduced crime, gains in life-cycle labor income, reduced special education costs and enhanced educational attainment. The program subsidizes the childcare of the mothers of participants and facilitates their employment and earnings.

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

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

[13] See Kottelenberg and Lehrer (2014) and Baker et al. (2015).

[14] 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 summarizes the main findings of this paper. It displays the discounted (using a 3% discount rate) life-cycle benefits of the program and costs (2014 USD), overall and disaggregated by category.[15] We report separate estimates by gender, and for the pooled sample of males and females. Costs are substantial, as has frequently been noted by critics.[16] But so are the benefits, which far outweigh the costs.

Table 1 summarizes results from numerous sensitivity analyses that we conduct throughout the paper. We set to zero the net present value of each of the four main components of our analysis and recalculate our cost-benefit analysis. Our estimates are statistically and economically significant even after eliminating the benefits from anyone of the four main components that we monetize. No single component drives our results.

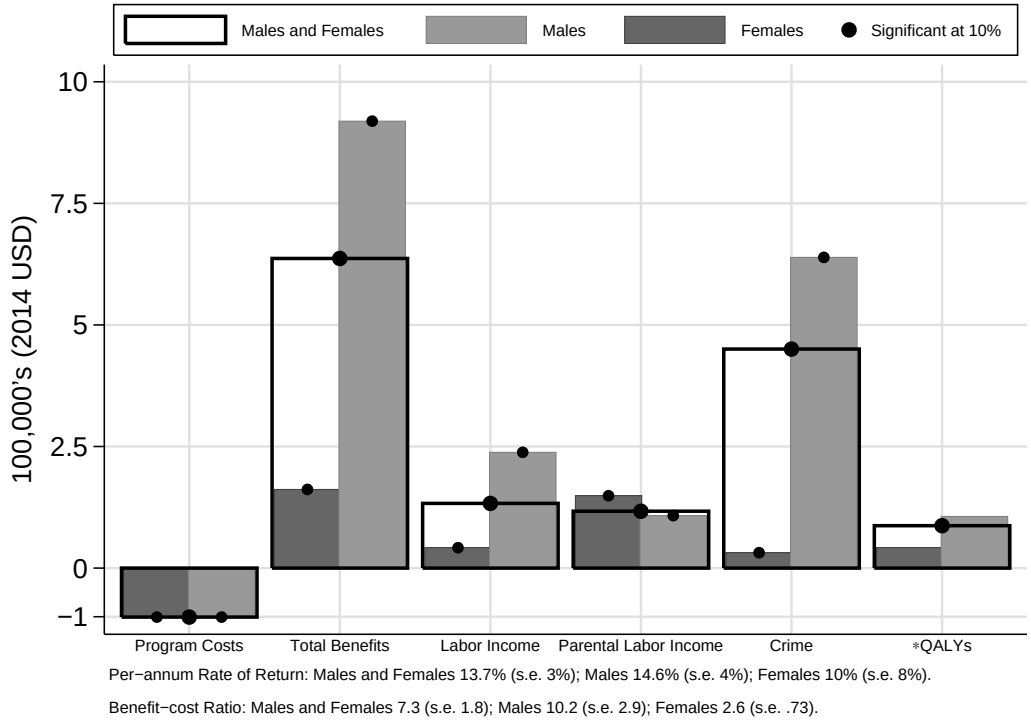
The rest of the paper justifies and interprets these estimates. We proceed in the following way. Section 1 discusses the ABC/CARE intervention. Section 2 presents our notation and the definitions of the treatment effects estimated in this paper. Section 3 discusses our approaches to inference for vectors of treatment effects. We use combining functions that summarize the *number* of beneficial outcomes, as well as the number of statistically significant beneficial outcomes. Section 4 reports estimated treatment effects.

Our analysis of treatment effects establishes that the program had substantial impacts on multiple domains. It motivates our benefit-cost ratio and

[15]The baseline discount rate of 3% is an arbitrary decision. In Table 7 and Table 9, 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 I.

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

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




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 predicted 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 net benefits: 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 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 predicted death, with $\$150,000$ (2014 USD) as base value for a year of life.

Table 1: Summary of Sensitivity Analyses for Cost/Benefit Analysis of the Program, Treatment vs. Next Best

| Component Set to Zero: | None | | | Labor Income | | | Parental Labor Income | | | Crime | | | *QALYs | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample: | | | | | | | | | | | | | | | |
| Pooled | ✓ | | | ✓ | | | ✓ | | | ✓ | | | ✓ | | |
| Male | | ✓ | | | ✓ | | | ✓ | | | ✓ | | | ✓ | |
| Female | | | ✓ | | | ✓ | | | ✓ | | | ✓ | | | ✓ |
| IRR | **0.13** | **0.13** | 0.10 | **0.11** | **0.12** | 0.10 | **0.09** | **0.11** | 0.04 | **0.09** | 0.08 | 0.09 | **0.12** | **0.12** | 0.10 |
| | (0.05) | (0.06) | (0.08) | (0.06) | (0.06) | (0.08) | (0.03) | (0.05) | (0.02) | (0.05) | (0.04) | (0.08) | (0.06) | (0.07) | (0.08) |
| B/C Ratio | **6.29** | **11.10** | **2.45** | **4.86** | 8.22 | **2.16** | **5.36** | 10.36 | 1.34 | **3.02** | 4.24 | 1.74 | **5.38** | 9.90 | **2.32** |
| | (2.11) | (6.35) | (0.79) | (2.18) | (5.35) | (0.70) | (2.11) | (6.36) | (0.69) | (1.14) | (2.72) | (0.72) | (2.04) | (6.13) | (0.76) |

Note: This table presents estimates of the internal rate of return (IRR) and the benefit-cost ratio (B/C Ratio) of ABC/CARE in scenarios where we set the net-present value of the estimated gain generated by the program of different components. "None" refers to the baseline estimation, where we do not set any of the components to zero. By "net" we mean that each component represents the total value for the treatment group minus the total value for the control group. For details on the construction of each component, see Figure 1. Inference is based on non-parametric, one-sided $p$-values from the empirical bootstrap distribution. For the B/C ratio we use a discount rate of 3%. We test the null hypotheses IRR = 3% and B/C = 1—we select 3% as the benchmark null for the IRR because that is the baseline discount rate that we use in this paper. 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 predicted death, with $\$150,000$ (2014 USD) as base value for a year of life.

internal rate of return analysis to summarize these effects using economically meaningful metrics. Section 5 presents our approaches for predicting life-cycle outcomes and the evidence supporting the assumptions that justify these approaches. Section 6 reports our estimates of benefit/cost ratios and rates of return. It reports outcomes from a variety of robustness checks. Section 7 summarizes the paper.

# 1 Background and Data Sources

## 1.1 Overview

ABC/CARE targeted disadvantaged, predominately African-American children in Chapel Hill/Durham, North Carolina.[17] Table 2 compares the two virtually identical programs. Appendix A describes these programs in detail. Here, we summarize their main features.

The goal of these programs was to enhance the early-life skills of disadvantaged children. Both programs supported language, motor, and cognitive development as well as socio-emotional competencies considered crucial for school success including task-orientation, ability to communicate, independence, and pro-social behavior.[18]

The programs individualized treatment. Each child's progress was recorded and learning activities were appropriately adjusted every 2 to 3 weeks. Environments were organized to promote pre-literacy and provide access to a

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

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

rich set of learning tools.[19] The curriculum emphasized active learning experiences, dramatic play, and basic concepts of order and category ("pre-academic skills"), as well as discipline and the ability to interact with and respect others. At later ages (3 through 5), the program focused on the development of "socio-linguistic and communicative competence."[20]

ABC recruited four cohorts of children born between 1972 and 1976. CARE recruited two cohorts of children, born between 1978 and 1980. The recruitment processes for each study were identical. Potential participant families 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.[21]

As shown in Table 2, the design and implementation of ABC and CARE were very similar. ABC had two phases, the first of which lasted from birth until age 5. In this phase, children were randomly assigned to treatment. The second phase of the study consisted of child academic support through home visits from ages 5 through 8. CARE consisted of two treatment phases as well that were very similar to ABC. The first phase of CARE from birth until age

[19]The "LearningGames" approach was implemented by infant and toddler caregivers in 1:1 child-adult interactions. Each "LearningGames" activity states a developmentally-appropriate objective, the necessary materials, directions for teacher behavior, and expected child outcome.

[20]Ramey et al. (1977); Haskins (1985); Ramey and Haskins (1981); Ramey and Campbell (1979); Ramey and Smith (1977); Ramey et al. (1982); Sparling and Lewis (1979, 1984).

[21]See Appendix A for details on the construction for the index used. The index weighs 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 in welfare, father in unstable job, maternal IQ, 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.

5, had an additional treatment arm of home visits designed to improve home environments.[22] Participation in the second phase was randomized in ABC, but not in CARE.

Our analysis is based on the first phase and pools the CARE treatment group with the ABC treatment group. The second-phase treatment of ABC/CARE had little impact on participants (for evidence, see Campbell et al., 2014 and García et al., 2016). Campbell et al. (2014) establish the validity of pooling the data on second phase treatments and controls with the first phase controls in ABC.

We do not use the data on the CARE group that only received home visits in the early years. Campbell et al. (2014) and García et al. (2016) show that there is no statistically significant effect of this component.

For the treatment phase that we analyze, the center received the treated children from 7:45 a.m. to 5:30 p.m, five days a week and fifty weeks a year. As we argue below, in practice the center had a very relevant childcare component that caused gains in parental labor income.

For both programs, from birth until the age of 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.[23] In addition, we have access to administrative criminal records and a physician-administered medical survey at the mid 30's. This allows us to study the long-term effects of the programs

[22] Wasik et al. (1990).

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

Table 2: ABC and CARE, Program Comparison

| | ABC | CARE | ABC = CARE ? |
|---|---|---|---|
| **Program Overview** | | | |
| Years Implemented | 1972–1982 | 1978–1985 | |
| First-phase Treatment | Birth to 5 years old | Birth to 5 years old | ✓ |
| Second-phase Treatment | 5 to 8 years old | 5 to 8 years old | ✓ |
| Initially Recruited Sample | 121* | 67 | |
| # of Cohorts | 4 | 2 | |
| **Eligibility** | Socio-economic disadvantage according to a multi-factor index (see Appendix A) | Socio-economic disadvantage according to a multi-factor index (see Appendix A) | ✓ |
| **Control** | | | |
| N | 54 | 23 | |
| Treatment Given | Diapers from birth to age 3, unlimited formula from birth to 15 months | Diapers from birth to age 3, unlimited formula from birth to 15 months | ✓ |
| Control Substitution | 75% | 74% | |
| **Treatment** | Center-based childcare | Center-based childcare and family education | |
| **Center-based Childcare** | | | |
| N | 53 (participated) | 17 | |
| Intensity | 6.5–9.75 hours a day for 50 weeks per year | 6.5–9.75 hours a day for 50 weeks per year | ✓ |
| Components | Stimulation, medical care, nutrition, social services | Stimulation, medical care, nutrition, social services | ✓ |
| Staff-to-child Ratio | 1:3 during ages 0–1 | 1:3 during ages 0–1 | ✓ |
| | 1:4–5 during age 1–4 | 1:4–5 during age 1–4 | ✓ |
| | 1:5–6 during ages 4–5 | 1:5–6 during ages 4–5 | ✓ |
| Staff Qualifications | Range of degrees beyond high school; experience in early childcare | Range of degrees beyond high school; experience in early childcare | ✓ |
| **Home Visitation** | | | |
| N | (not part of the program) | 27 | |
| Intensity | | Home visits lasting 1 hour. 2–3 per month during ages 0–3. 1–2 per month during ages 4–5 | |
| Curriculum | | Social and mental stimulation; parent-child interaction | |
| Staff-to-child Ratio | | 1:1 | |
| Staff Qualifications | | Home visitor training | |
| **School-age Treatment** | | | |
| N | 46 | 39 | |
| Intensity | Every other week | Every other week | ✓ |
| Components | Parent-teacher meetings | Parent-teacher meetings | ✓ |
| Curriculum | Reading and math | Reading and math | ✓ |
| Staff Qualifications | Range of degrees beyond high school; experience in early childcare | Range of degrees beyond high school; experience in early childcare | ✓ |

Note: This table compares the main elements of ABC and CARE, summarized in this section. A ✓ indicates that ABC and CARE had the same feature. A blank space indicates that the indicated component was not part of the program.
* As documented in Appendix A.2, there were losses in the initial samples due to death, parental moving, and diagnoses of mental pathologies for the children.

along multiple dimensions of human development.[24]

## 1.2 Randomization Protocol and Compromises

Randomization for ABC/CARE was conducted on child pairs matched on family background. Siblings and twins were jointly randomized into either treatment or control groups.[25] Randomization pairing was based on a risk index, maternal education, maternal age, and gender of the subject.[26] ABC collected an initial sample of 121 subjects. We characterize each missing observation in Appendix A. In Appendix G.3, we document that our estimates are robust when we adjust for missing data using standard methods, described in Appendix C.2. We conduct the same analysis for the CARE sample. 22 subjects in ABC did not stay in the program through age 5. Dropouts are evenly balanced and are primarily related to the health of the child and mobility of families and not to dissatisfaction with the program.[27]

[24] See 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, once we drop the individuals for whom we have no crime or health information, for which there is substantial attrition. Further, the methodology we propose addresses missing data in either of these two outcome categories.

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

[26] We do not know the original pairs.

[27] The 22 dropouts include four children who died, four children who left the study because their parents moved, and two children who were diagnosed as developmentally delayed. Details are in Table A.2. Everyone offered the program was randomized to either treatment or control. All eligible families agreed to participate. Dropping out occurs *after* randomization.

## 1.3 Control Group Substitution

In ABC/CARE, many control group members (but no children from families offered treatment) attended alternative (to home) childcare or preschool centers.[28] The figure is 75% for ABC and 74% for CARE.

Figure 2a shows the cumulative distribution of the proportion of time in the first five years that control subjects were enrolled in alternatives. Figure 2b shows the dynamics of enrollment. Those who enroll generally stay enrolled. As control children age, they are more likely to enter 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. Disadvantage is measured by maternal education, maternal IQ, Apgar scores, and the high-risk index defining ABC/CARE eligibility. Children who attend alternatives have fewer siblings. On average, they are children of mothers who are more likely to be working at baseline.[29] Parents of girls are much more likely to use alternative childcare if assigned to the control group.[30]

Most of the alternative childcare centers received federal subsidies and were subject to the federal regulations of the era.[31] They had relatively low

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

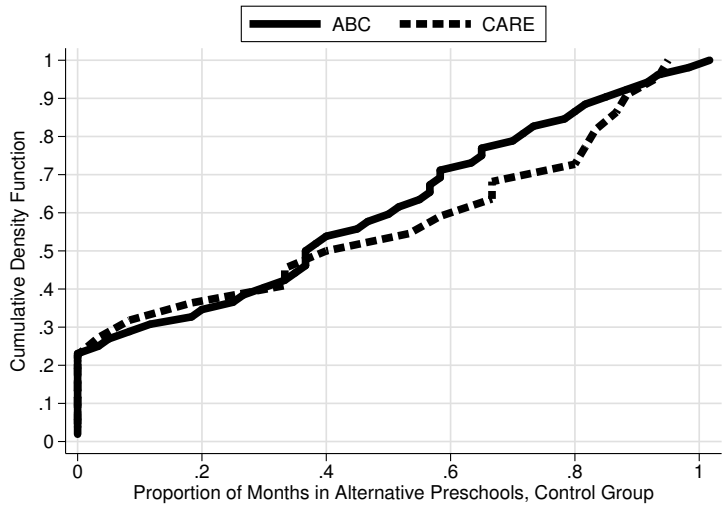
[29] Statistically significant at 10%.

[30] See Table A.4 in Appendix A for tests of differences across these variables between children in the control group who attended and who did not attend alternative preschools.

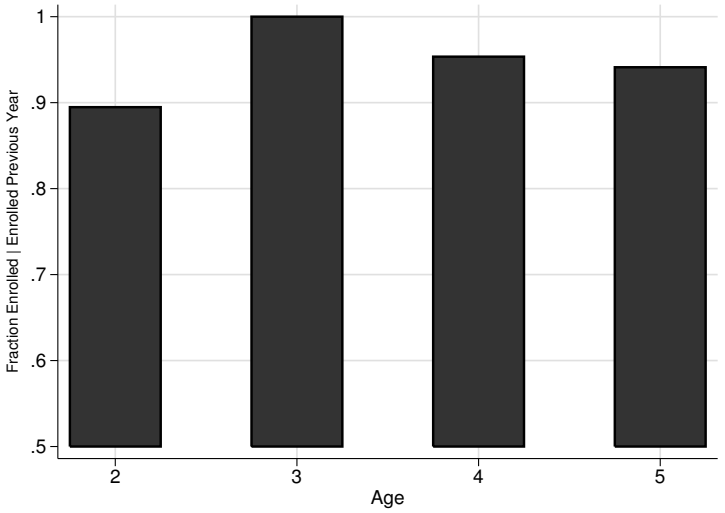
[31] Appendix A.5.1 discusses the federal standards of that day. See Department of Health, Education, and Welfare (1968); North Carolina General Assembly (1971); Ramey et al. (1977); Ramey and Campbell (1979); Ramey et al. (1982); Burchinal et al. (1997).

Figure 2: Control Substitution Characteristics, ABC/CARE Control Group

(a) Cumulative Enrollment




(b) Enrollment Dynamics




Note: Panel (a) displays the cumulative distribution function of enrollment in alternatives. Panel (b) displays the fraction of ABC/CARE control-group children enrolled in alternatives, conditional on being enrolled in the previous age (at least one month).

quality compared to ABC/CARE.[32] The access of control-group children to alternative programs affects the interpretation of estimated treatment effects, as we discuss next.

## 2 Parameters Estimated in This Paper

Random assignment to treatment does not guarantee that conventional treatment effects answer policy-relevant questions. In this paper, we define and estimate three parameters that address different policy questions.

Life cycles consist of $A$ discrete periods. Treatment occurs in the first $\bar{a}$ periods of life $[1, \ldots, \bar{a}]$. We have data through age $a^* > \bar{a}$. We lack follow-up data on the remainder of life $(a^*, \ldots, A]$. 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 compliance in the initial randomization protocol, i.e., $D = R$ implies compliance into the initial randomization protocol.

Individuals are eligible to participate in the program if baseline background variables $\boldsymbol{B} \in \mathcal{B}_0$. $\mathcal{B}_0$ is the set of scores on the risk index that determines program eligibility. As it turns out, in the ABC/CARE study, all of the eligible persons given the option to participate choose to do so ($W = 1$, and $D = R$). There are very few dropouts. *Ex ante*, parents perceived that ABC/CARE was superior to other childcare alternatives. Thus,

[32] When we compare ABC/CARE treatment to these alternatives, ABC/CARE has substantial treatment effects. Further, as we argue below, parents perceived that ABC/CARE was superior to the alternatives.

we can safely interpret the treatment effects generated by the experiment as average treatment effects for the population for which $\boldsymbol{B} \in \mathcal{B}_0$ and not just treatment effects for the treated (**TOT**).[33]

Let $\boldsymbol{Y}_a^1$ be the outcome vector at age $a$ for the treated. $\boldsymbol{Y}_a^0$ is the age-$a$ outcome vector for the controls. In principle, life-cycle outcomes for the treatments and controls can depend on the exposures to various alternatives at each age. It would be desirable to estimate treatment effects for each possible exposure but our samples are too small to make credible estimates for very detailed exposures.

All treatment group children have the same exposure. We simplify the analysis of the controls 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.[34] We test the sensitivity of our estimates to the choice of different categorizations in our empirical analysis in Appendix H.

We thus compress a complex reality into two counterfactual outcome states

[33] All providers of health care and social services (referral agencies) in the area of the ABC/CARE study were informed of the programs. They referred mothers whom they considered disadvantaged. Eligibility was corroborated before randomization. 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).

[34] This assumption is consistent with Figure 2b. Once parents decide to enroll their children in alternative childcare arrangements, the children stay enrolled up to age 5.

at age $a$ for control group members:

$$\begin{aligned} \boldsymbol{Y}^0_{a,H} \quad &: \quad \textbf{Subject received home care exclusively} \\ \boldsymbol{Y}^0_{a,C} \quad &: \quad \textbf{Subject received some alternative childcare}. \end{aligned}$$

We define $V$ as a dummy variable indicating participation by control-group children in an alternative preschool. $V = 0$ denotes staying at home. The outcome when a child is in control status is

$$\boldsymbol{Y}^0_a := (1 - V)\, \boldsymbol{Y}^0_{a,H} + (V)\, \boldsymbol{Y}^0_{a,C}. \tag{1}$$

One parameter of interest addresses the question: what is the effect of the program as implemented? This is the effect of the program compared to the next best alternative as perceived by the parents (or the relevant decision maker) and is defined by

$$\boldsymbol{\Delta}_a := \mathbb{E}\left[\boldsymbol{Y}^1_a - \boldsymbol{Y}^0_a | W = 1\right] = \mathbb{E}\left[\boldsymbol{Y}^1_a - \boldsymbol{Y}^0_a | \boldsymbol{B} \in \mathcal{B}_0\right], \tag{2}$$

where the second equality follows because everyone eligible wants to participate in the program. For the sample of eligible persons, this parameter addresses the effectiveness of the program relative to the quality of all alternatives available when the program was implemented, including staying at home.

It is fruitful to ask: what is the effectiveness of the program with respect to a counterfactual world in which the child stays at home full time? The associated causal parameter for those who would choose to keep the child at

home is:

$$\boldsymbol{\Delta}_a\left(V=0\right) := \mathbb{E}\left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_a^0 | V=0, W=1\right] := \mathbb{E}\left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_{a,H}^0 | V=0, \boldsymbol{B} \in \mathcal{B}_0\right]. \tag{3}$$

It is also useful to assess the average effectiveness of a program relative to attendance in an alternative preschool for those who would choose an alternative:

$$\boldsymbol{\Delta}_a\left(V=1\right) := \mathbb{E}\left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_a^0 | V=1, W=1\right] := \mathbb{E}\left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_{a,C}^0 | V=1, \boldsymbol{B} \in \mathcal{B}_0\right]. \tag{4}$$

Random assignment to treatment does not directly identify (3) or (4). Econometric methods are required to identify these parameters. We characterize the determinants of choices and our strategy for controlling for selection into "$H$" and "$C$" below.[35]

# 3 Summarizing Multiple Treatment Effects

ABC/CARE has rich longitudinal data on multiple outcomes over multiple periods of the life cycle. Summarizing these effects in an interpretable way is challenging.[36] Simpler, more digestible summary measures are useful for understanding our main findings. This section discusses our approach to sum-

[35] Appendix H displays results with alternative definitions of $V$ (i.e., different thresholds define if a child attended alternative preschool). The results are robust to the various definitions. What matters is whether any out-of-home child care is being used ($V > 0$), and not the specific value of $V$.

[36] Appendix G presents step-down $p$-values for the blocks of outcomes that are used in our benefit/cost analysis which we summarize in this section (Lehmann and Romano, 2005 and Romano and Shaikh, 2006). We follow the algorithm in Romano and Wolf (2016).

marizing vectors of treatment effects using combining functions that count the proportion of treatment effects by different categories of outcomes.

Consider a block of $N_l$ outcomes indexed by set $Q_l = \{1, \ldots, N_l\}$. Let $j \in Q_l$ be a particular outcome within block $l$. Associated with it is a mean treatment effect

$$\Delta_{j,a} := \mathbb{E}\left[Y^1_{j,a} - Y^0_{j,a} | \boldsymbol{B} \in \mathcal{B}_0\right], j \in Q_l. \tag{5}$$

We assume that outcomes can be ordered so that $\Delta_{j,t} > 0$ is beneficial.[37] We summarize the estimated effects of the program on outcomes within the block by the count of positive impacts within block $l$:

$$C_l = \sum_{j=1}^{N_l} 1(\hat{\Delta}_{j,a} > 0). \tag{6}$$

The proportion of beneficial outcomes in block $l$ is $C_l/N_l$.[38]

Let $\mathcal{L}$ be the set of blocks. Under the null hypothesis of no treatment effects for all $j \in Q_l, l \in \mathcal{L}$, and assuming the validity of asymptotic approximations, $C_l/N_l$ should be centered around 1/2. We bootstrap to obtain $p$-values for the null for each block and over all blocks.[39] We also count the beneficial treatment effects that are statistically significant in the sets of outcomes across each of the groups indexed by the set $Q_l$. Using a 10% significance level, on average 10% of all outcomes should be "significant" at the 10% level even if

[37] All but 5% of the outcomes we study can be ranked in this fashion. See Appendix G for a discussion.

[38] In our empirical application we consider all the outcomes as a block, and then different blocks grouped according to common categories—e.g., skills, health, crime.

[39] Bootstrapping allows us to account for dependence across outcomes in a general way.

there is no treatment effect of the program. We provide evidence against both null hypotheses.[40] Combining counts across all blocks enables us to avoid (i) arbitrarily picking outcomes that have statistically significant effects—"cherry picking"; or (ii) arbitrarily selecting blocks of outcomes to correct the $p$-values when accounting for multiple hypothesis testing.[41,42]

# 4 Estimated Treatment Effects and Combining Functions

ABC/CARE has a multiplicity of treatment effects corresponding to all of the measures collected in the multiple waves of the longitudinal surveys. Reporting these treatment effects in the text would overwhelm the reader. Here we report estimates of the main treatment effects that underlie our benefit/cost and rate of return analyses.[43] These treatment effects are monetized in Section 5 to present an economically justified aggregate measure.

Evidence from ABC/CARE and many other early childhood programs is often criticized because of their small sample sizes.[44] An extensive analysis reported in Campbell et al. (2014) shows that asymptotic inference and small

[40] In this case, we perform a "double bootstrap" procedure to first determine significant treatment effects at 10% level and then calculate the standard error of the count.

[41] We present $p$-values for these hypotheses and a number of combining functions by outcome categories in Appendix G.

[42] In Appendix G we present yet another alternative. We calculate a "latent" outcome out of the set of outcomes within a block and perform inference on this latent. The results point to beneficial effects of the program in this case as well.

[43] Appendix G reports treatment effects and step-down $p$-values for all the outcomes analyzed. These account for multiple hypothesis testing as in Lehmann and Romano (2005) and Romano and Shaikh (2006).

[44] See, e.g., Murray (2013).

sample permutation-based inference closely agree when applied to ABC/CARE data. For this reason, we use large sample inference throughout this paper.[45]

## 4.1 Estimated Treatment Effects

Tables 3 and 4 present the following estimates, for males and females respectively. Column (1) gives sample mean differences in outcomes between treatment and control groups. Column (2) adjusts the differences for attrition and controls for background variables. Both are estimates of the parameter defined in equation (2). Column (3) shows the mean difference between the full treatment-group and the control-group children who did not attend alternatives. Column (4) gives standard matching estimates for the parameter defined in equation (3).[46] Column (5) gives mean differences between the full treatment-group and control-group children who attended alternatives. Column (6) gives matching estimates for the parameter of equation (4).

The results for females show that ABC/CARE has substantial effects on education when comparing treatment outcomes to those from the next best alternative. High school graduation increases between 13 and 25 percentage points, depending on the estimate that we consider; college graduation increases 13 percentage points; and the average years of schooling increase between 2.1 and 1.8 years. Employment at age 30 increases between 13 and 8 percentage points. ABC/CARE has substantial impacts on human capi-

[45]For precise details on the construction of the inference procedures used throughout the paper, see Appendix C.8.

[46]In Appendix G.1.1, we provide details on: (i) the kernel matching estimator that we use; (ii) the matching variables that we use; and (iii) a sensitivity analysis to these matching variables.

tal accumulation and employment. The results strengthen when we compare treatment with the alternative of staying at home.

The results for males are somewhat different from those for females. Treatment has substantial effects when compared to next best alternative. The effects are positive for a variety of health indicators, including drug use and hypertension. The effects on employment and labor income are also substantial. The increase in employment at age 30 ranges from 11 to 19 percentage points. Labor income at age 30 increases between 19 and 24 thousand of 2014 USD after treatment. The effects strengthen when comparing treatment to alternative preschool. Separation from the mother and being placed in relatively low quality childcare centers have more deleterious consequences for males than for females.[47]

The results hold using alternative definitions of control substitution (see Appendix H). They remain statistically significant or are borderline statistically insignificant when computing two-tailed $p$-values (see Appendix H).

The estimates contrasting the effects for females and males in (3) and (5) are not based on matching; the estimates in (4) and (6) are. For the matching estimates, we rely on observed, baseline characteristics. In Appendix G.1, we explain our choice of these variables and we make a thorough analysis to conclude that there is little sensitivity to the choice of these variables.[48]

[47] This is consistent with the evidence in Baker et al. (2015) and Kottelenberg and Lehrer (2014).

[48] We also present this sensitivity analysis changing the variables used to condition while estimating treatment effects and changing the variables used to construct the weights to account for attrition.

Table 3: Treatment Effects on Selected Outcomes, Males

| Category | Variable | Age | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|---|---|
| Parental Income | Parental Labor Income | 3.5 | 1,036 | 494 | 73.862 | 1,462 | 123 | 690 |
| | | | (0.374) | (0.411) | (0.474) | (0.390) | (0.479) | (0.417) |
| | | 12 | 7,085 | 9,625 | 18,050 | 12,639 | 6,620 | 5,383 |
| | | | **(0.092)** | **(0.020)** | **(0.038)** | **(0.074)** | **(0.098)** | (0.139) |
| | | 15 | 8,488 | 4,495 | 5,540 | 4,805 | 2,885 | 4,345 |
| | | | **(0.071)** | (0.221) | (0.243) | (0.264) | (0.354) | (0.296) |
| | | 21 | 12,732 | 8,809 | 122 | -933 | 10,784 | 10,283 |
| | | | **(0.005)** | **(0.098)** | (0.448) | (0.456) | **(0.056)** | **(0.041)** |
| Education | Graduated High School | 30 | 0.073 | 0.044 | 0.116 | 0.083 | 0.040 | 0.063 |
| | | | (0.262) | (0.375) | **(0.001)** | (0.346) | (0.407) | (0.317) |
| | Graduated 4-year College | 30 | 0.170 | 0.138 | 0.149 | 0.099 | 0.135 | 0.143 |
| | | | **(0.055)** | (0.128) | (0.216) | (0.338) | (0.154) | (0.130) |
| | Years of Education | 30 | 0.525 | 0.541 | 1.010 | 0.777 | 0.351 | 0.344 |
| | | | (0.151) | (0.163) | (0.998) | (0.136) | (0.280) | (0.256) |
| Labor Income | Employed | 30 | 0.119 | 0.196 | 0.108 | 0.040 | 0.237 | 0.261 |
| | | | (0.128) | **(0.025)** | **(0.001)** | (0.383) | **(0.025)** | **(0.013)** |
| | Labor Income | 30 | 19,810 | 24,365 | 25,220 | 20,611 | 23,072 | 21,836 |
| | | | **(0.091)** | **(0.092)** | (0.998) | (0.122) | (0.107) | **(0.094)** |
| Crime | Total Felony Arrests | Mid-30s | 0.196 | 0.685 | 1.523 | 1.340 | 0.481 | 0.188 |
| | | | (0.368) | (0.183) | **(0.064)** | **(0.026)** | (0.284) | (0.410) |
| | Total Misdemeanor Arrests | Mid-30s | -0.501 | -0.244 | -0.298 | -0.034 | -0.246 | -0.507 |
| | | | (0.171) | (0.289) | (0.314) | (0.422) | (0.329) | (0.168) |
| Health | Self-reported drug user | Mid-30s | -0.333 | -0.438 | -0.673 | -0.557 | -0.326 | -0.330 |
| | | | **(0.019)** | **(0.002)** | **(0.000)** | **(0.000)** | **(0.039)** | **(0.023)** |
| | Systolic Blood Pressure (mm Hg) | Mid-30s | -9.791 | -13.275 | 14.196 | 14.976 | -24.166 | -18.559 |
| | | | (0.113) | **(0.049)** | **(0.013)** | **(0.000)** | **(0.000)** | **(0.011)** |
| | Diastolic Blood Pressure (mm Hg) | Mid-30s | -10.854 | -14.134 | -9.709 | -8.741 | -18.387 | -13.987 |
| | | | **(0.032)** | **(0.004)** | **(0.049)** | **(0.032)** | **(0.000)** | **(0.007)** |
| | Hypertension | Mid-30s | -0.291 | -0.377 | -0.120 | -0.074 | -0.492 | -0.434 |
| | | | **(0.042)** | **(0.009)** | (0.302) | (0.353) | **(0.006)** | **(0.006)** |

Note: This table shows the treatment effects for categories outcomes that are important for our benefit/cost analysis. Systolic and diastolic blood pressure are measured in terms of mm Hg. Each column present estimates for the following parameters: (1) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}^0 | \boldsymbol{B} \in \mathcal{B}_0\right]$ (no controls); (2) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}^0 | \boldsymbol{B} \in \mathcal{B}_0\right]$ (controls); (3) $\mathbb{E}\left[\boldsymbol{Y}^1 | R = 1\right] - \mathbb{E}\left[\boldsymbol{Y}^0 | R = 0, V = 0\right]$ (no controls); (4) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}_H^0 | \boldsymbol{B} \in \mathcal{B}_0\right]$ (controls); (5) $\mathbb{E}\left[\boldsymbol{Y}^1 | R = 1\right] - \mathbb{E}\left[\boldsymbol{Y}^0 | R = 0, V = 1\right]$ (no controls); (6) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}_C^0 | \boldsymbol{B} \in \mathcal{B}_0\right]$ (controls). We account for the following background variables ($\boldsymbol{B}$): ABC/CARE indicator; Apgar scores at minutes 1 and 5, and the high-risk index. We define the high-risk index in Appendix A and explain how we choose the control variables in Appendix G.1. Columns (2), (4), and (6) correct for item non-response and attrition using inverse probability weighting as we explain in Appendix C.2. Inference is based on non-parametric, one-sided $p$-values from the empirical bootstrap distribution. We highlight point estimates significant at the 10% level. See Appendix H for two-sided $p$-values.

Table 4: Treatment Effects on Selected Outcomes, Females*

| Category | Variable | Age | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|---|---|
| Parental Income | Parental Labor Income | 3.5 | 2,756 | 2,986 | 6,864 | 8,584 | 1,521 | 3,773 |
| | | | (0.189) | (0.213) | (0.122) | **(0.045)** | (0.332) | (0.154) |
| | | 12 | 13,633 | 19,592 | 28,328 | 26,489 | 15,343 | 18,678 |
| | | | **(0.054)** | **(0.027)** | **(0.027)** | **(0.009)** | **(0.064)** | **(0.019)** |
| | | 15 | 8,565 | 7,159 | 2,713 | 8,441 | 7,465 | 10,487 |
| | | | **(0.060)** | (0.137) | (0.480) | (0.345) | (0.134) | **(0.064)** |
| | | 21 | 5,708 | 8,670 | 45,697 | 25,142 | 6,251 | 3,943 |
| | | | (0.136) | (0.140) | **(0.000)** | **(0.000)** | (0.224) | (0.261) |
| Education | Graduated High School | 30 | 0.253 | 0.131 | 0.553 | 0.595 | -0.026 | 0.066 |
| | | | **(0.009)** | (0.152) | **(0.003)** | **(0.000)** | (0.413) | (0.320) |
| | Years of Education | 30 | 2.143 | 1.843 | 3.861 | 3.923 | 1.163 | 1.409 |
| | | | **(0.001)** | **(0.002)** | **(0.000)** | **(0.000)** | **(0.054)** | **(0.017)** |
| Labor Income | Employed | 30 | 0.131 | 0.081 | 0.381 | 0.340 | -0.010 | 0.070 |
| | | | **(0.096)** | (0.206) | **(0.039)** | **(0.057)** | (0.465) | (0.264) |
| | Labor Income | 30 | 2,548 | 1,884 | 15,094 | 13,096 | -2,677 | -2,122 |
| | | | (0.335) | (0.382) | **(0.056)** | **(0.022)** | (0.330) | (0.363) |
| Crime | Total Felony Arrests | Mid-30s | -0.328 | -0.351 | -0.944 | -0.965 | -0.059 | 0.004 |
| | | | **(0.077)** | **(0.087)** | **(0.095)** | **(0.095)** | (0.287) | (0.500) |
| | Total Misdemeanor Arrests | Mid-30s | -0.973 | -0.737 | -2.010 | -2.451 | -0.269 | -0.201 |
| | | | **(0.057)** | (0.134) | (0.134) | (0.120) | (0.273) | (0.289) |
| Health | Self-reported drug user | Mid-30s | -0.033 | 0.004 | -0.114 | -0.101 | 0.020 | 0.033 |
| | | | (0.381) | (0.478) | (0.273) | (0.323) | (0.443) | (0.406) |
| | Systolic Blood Pressure (mm Hg) | Mid-30s | -2.899 | -5.407 | -0.488 | -0.822 | -6.239 | -6.784 |
| | | | (0.307) | (0.241) | (0.488) | (0.457) | (0.249) | (0.170) |
| | Diastolic Blood Pressure (mm Hg) | Mid-30s | -0.002 | -0.179 | 4.091 | 4.122 | -1.347 | -2.160 |
| | | | (0.483) | (0.438) | (0.245) | (0.222) | (0.392) | (0.339) |
| | Hypertension | Mid-30s | 0.172 | 0.085 | 0.077 | 0.162 | 0.102 | 0.107 |
| | | | (0.111) | (0.293) | (0.331) | (0.245) | (0.299) | (0.255) |

Note: This table shows the treatment effects for categories outcomes that are important for our benefit/cost analysis. Systolic and diastolic blood pressure are measured in terms of mm Hg. Each column present estimates for the following parameters: (1) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}^0 | \boldsymbol{B} \in \mathcal{B}_0\right]$ (no controls); (2) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}^0 | \boldsymbol{B} \in \mathcal{B}_0\right]$ (controls); (3) $\mathbb{E}\left[\boldsymbol{Y}^1 | R = 1\right] - \mathbb{E}\left[\boldsymbol{Y}^0 | R = 0, V = 0\right]$ (no controls); (4) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}^0_H | \boldsymbol{B} \in \mathcal{B}_0\right]$ (controls); (5) $\mathbb{E}\left[\boldsymbol{Y}^1 | R = 1\right] - \mathbb{E}\left[\boldsymbol{Y}^0 | R = 0, V = 1\right]$ (no controls); (6) $\mathbb{E}\left[\boldsymbol{Y}^1 - \boldsymbol{Y}^0_C | \boldsymbol{B} \in \mathcal{B}_0\right]$ (controls). We account for the following background variables ($\boldsymbol{B}$): ABC/CARE indicator; Apgar scores at minutes 1 and 5, and the high-risk index. We define the high-risk index in Appendix A and explain how we choose the control variables in Appendix G.1. Columns (2), (4), and (6) correct for item non-response and attrition using inverse probability weighting as we explain in Appendix C.2. Inference is based on non-parametric, one-sided $p$-values from the empirical bootstrap distribution. We highlight point estimates significant at the 10% level. See Appendix H for two-sided $p$-values.
*For females, we do not report graduation from a four-years college because we lack of common support to compute estimates for some parameters.

## 4.2 Estimated Combining Functions

We next report estimates of the proportion of beneficial effects by block and overall.[49] The analysis is based on treatment effect (2). Figure 3 displays the results from this analysis: ABC/CARE positively impacted a large percentage of the outcomes. We show the counts for treatment compared to the next best alternative chosen by parents in Figure 3a. Proportionately more outcomes are beneficial for females, but the proportions are high for both groups and well above the benchmark of 1/2. In Tables G.4 to G.12 of Appendix G, we document a large and precisely determined fraction of beneficial treatment effects well above one half for both genders for categories of outcomes spanning the life cycle through the mid 30's.

Using an $\alpha$-level of significance, one would expect to find that $\alpha$% of the treatment effects are "statistically significant," even if the null hypothesis of no effect of the program is true simply by chance. At a 10% level of significance, 46% are statistically significant for females and 28% for males (see Figure 3b).

Figures 3c and Figure 3d adjust the count in Figure 3a to analyze more clearly defined counterfactuals: treatment compared to staying at home and treatment compared to alternative preschool. These comparisons indicate that girls and boys benefit differently from alternatives to high quality treatment. Compared across all categories, girls benefit more from treatment when compared to staying at home (as opposed to attending alternative childcares), while males benefit more from treatment when compared to attending an al-

[49] We consider a total of 95 outcomes that we classify in Appendix G. These are the outcomes that most clearly relate to the treatment offered by the program.

ternative childcare arrangement (as opposed to staying at home).

# 5 Predicting and Monetizing Life-cycle Costs and Benefits

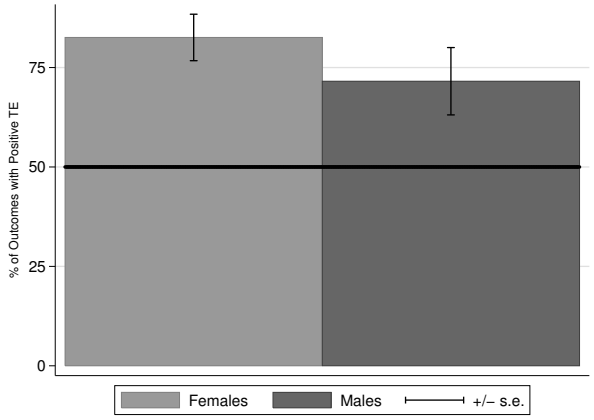
The major goal of this paper is to summarize the multiple benefits of ABC/CARE using benefit/cost and rate of return analyses. We rely on auxiliary data to predict the costs and benefits of the program over the life cycle after the measurement phase of the study ends.

This section explains our strategy for constructing out-of-sample treatment effects.[50] Our approach starts from and extends the analysis of Heckman et al. (2013), who show, in a setting similar to ours, that the effect of treatment on outcomes operates through its effects on inputs in a stable production function rather than through shifts in the production function. Table 5 presents the outcomes for which we conduct these analyses, and summarizes the methodology and auxiliary samples used. We initially focus on labor income to illustrate our approach, but a similar methodology is used to predict

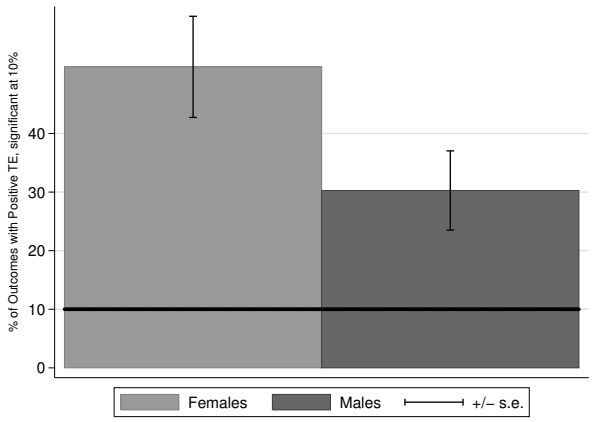
[50] Appendix C.7 gives details of our step-by-step procedure and state its identification and estimation strategy in the Generalized Method of Moments framework.

Figure 3: Positively Impacted Outcomes, ABC/CARE Males and Females

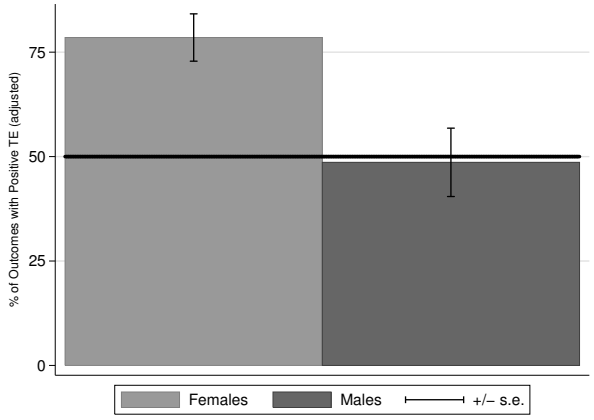
(a) Treatment vs. Next Best



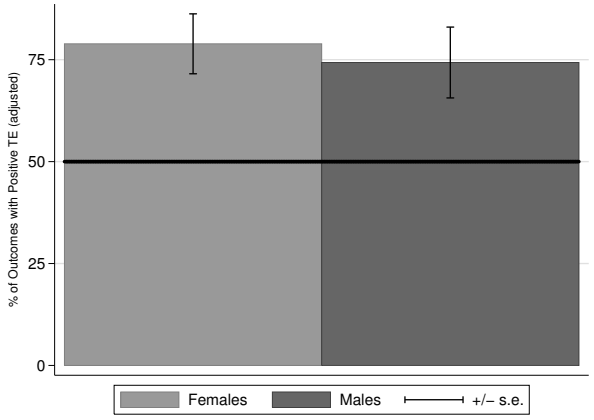

(b) Treatment vs. Next Best, Significant at 10% Level




(c) Treatment vs. Stay at Home




(d) Treatment vs. Alternative Preschool




Note: Panel (a) percentage of outcomes displaying a positive treatment effect, comparing treatment to next best. Panel (b) percentage of outcomes displaying a positive and statistically significant treatment effect (10% significance level). Panel (c) displays the percentage of outcomes with a positive treatment effect, comparing treatment to staying at home. Panel (d) displays the percentage of outcomes with a positive treatment effect, comparing treatment to alternative childcare arrangements. Standard errors are based on the empirical bootstrap distribution. For Panel (b) we perform a "double bootstrap" procedure to first determine significant treatment effects at 10% level and then calculate the standard error of the count.

Table 5: Summary of Prediction 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 |
| Alternative Preschools Costs | 0 to 5 | Imputed 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-30's | Observed* (Combines Administrative and Self-reported Data) | N/A | N/A | N/A |
| Crime (Arrests and Sentences) | Mid-30's 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) | NCDPS |
| Victimization Inflation | up to Age 50 | Impute national victims-arrests ratio | Use Full Auxiliary Samples to Impute | N/A | NCVS; NJRP; UCRS (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 6 | 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 predicting 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; QALY's: Quality-adjusted Life Years;UCRS: Uniform Crime Reporting Statistics. *When not observed, we impute based on the national arrest-sentence ratio from NJRP and UCRS. We assume that criminal records before the mid-30's implies no crime after the mid 30's. N/A not applicable.

the other outcomes.[51,52]

## 5.1 Using Auxiliary Data Sources to Predict Out-of-Sample Outcomes

We first present an informal summary of our approach. The next section gives a formal justification and reports tests of key assumptions. The remaining sections give applications 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. We lack information on participant outcomes afterward. Post-$a^*$ treatment effects are required to construct counterfactual life-cycle profiles.

Making valid predictions of out-of-sample treatment effects does not require making valid predictions of separate out-of-sample treatment and control profiles. Only valid predictions of their difference is required.

Nonetheless, in this paper we focus on making valid predictions of sep-

[51] We do not monetize the loss of leisure and household production that individuals suffer from working more (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 have different implications with respect to leisure-time allocations. In addition, our data are not well-suited for estimating a structural model. We note, however, that the benefit-cost ratio and internal rate of return are both statistically and substantially significant after removing labor income entirely (see Table 9). 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.

[52] Our 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 manipulate the standard tax simulators to assess estimation uncertainty.

arate treatment and control post-$a^*$ profiles. 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 control group and the synthetic control group are particularly compelling. By design, neither group receives treatment. In our data, all persons offered treatment accept it, so 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 comparison group members in the auxiliary samples with the same or similar characteristics as the experimental group members. The minimal set of characteristics includes the background variables $\boldsymbol{B} \in \mathcal{B}_0$. We use a coarse form of matching based on Algorithm 1 in Appendix C.3.3. In Stage II, we build models in these samples to predict out-of-sample outcomes separately for the outcomes of the treated and the controls.

Specifically, we adopt 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 $i$ ($\boldsymbol{Y}^d_{i,a}$) for $a \leq a^*$ to predictor variables (and interactions) that are affected by treatment ($\boldsymbol{X}^d_{i,a}$), as well as background variables ($\boldsymbol{B}_i$).[53] It turns out that we accurately predict within-sample treatment effects as well as levels of treatment and control profiles using this approach. In Step 2, we construct counterpart predictions of treatment and control outcomes using the

[53]To avoid notational clutter, we henceforth suppress individual $i$ subscripts.

auxiliary samples. We compare these 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 predict the post-$a^*$ outcomes.

Under exogeneity of the predictor variables and structural invariance (defined below), the two stages can be compressed into a single, one-stage, nonparametric matching procedure.[54] In Appendix C.3.4.1 we compare the estimates from matching with those from our main approach. We find close agreement between the two approaches (see Appendix C.3.4.1) and for different assumptions about the serial correlation processes of the outcome equations.

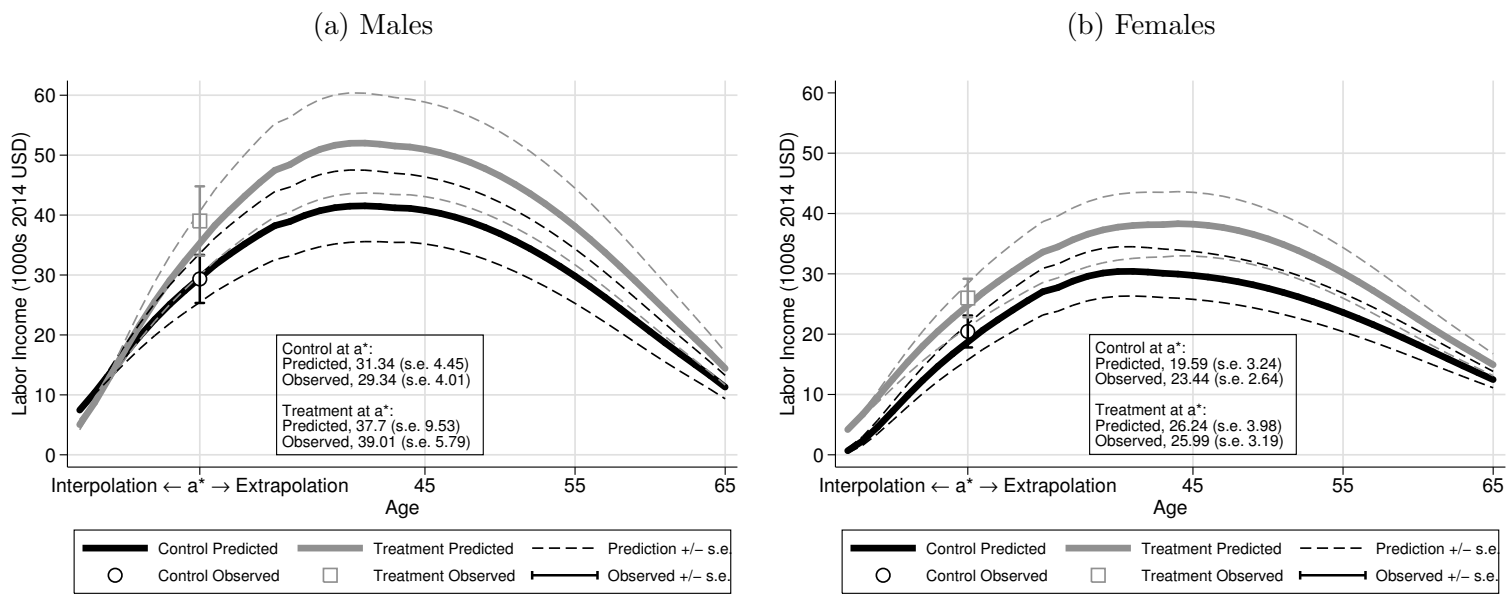
Figure 4 previews the outputs from our approach, displaying the life-cycle labor income profiles for the treatment and control groups. It also compares the realized labor income to the model-predicted labor income at $a^*$. There is close agreement of the constructed profiles within sample. The pattern of life-cycle labor income we generate is typical for low-skilled workers (Blundell et al., 2015; Gladden and Taber, 2000; Sanders and Taber, 2012; Lagakos et al., 2016).[55]

We conduct a further check on the validity of our procedure. In the experimental sample all of the parents of children with characteristics $\boldsymbol{B} \in \mathcal{B}_0$ 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 $\boldsymbol{B} \in \mathcal{B}_0$ 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

[54] See Heckman et al. (1998) for an example.

[55] For details on the variables used to construct the predictions, see Appendix C.

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




Note: Panel (a) displays the predicted life-cycle labor income profiles for ABC/CARE males by treatment status, 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- and treatment-group participants. Panel (b) displays the analogous figure for females. Our predictions go up to age 67, age of assumed retirement. Standard errors are based on the empirical bootstrap distribution. See Appendix C for a discussion of our choice of predictors and a sensitivity analysis on those predictors. For the control-group males, we over predict labor income; for the treatment-group males we under predict labor income. The over and under predictions are not statistically significant (and they underestimate the predicted treatment effect on labor income at age 30). For the control-group females, we under predict labor income; for treatment-group females, we over predict labor income. This overestimates the predicted treatment effect on labor income at age 30. This overestimation, however, is not statistically significant (and labor income is a relatively minor component of the overall analysis for females).

in our auxiliary sample for whom $\boldsymbol{B} \in \mathcal{B}_0$ 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).[56] The agreement is reassuringly close. We now formalize our approach.

## 5.2 Constructing Out-of-Sample Counterfactuals

We now present our analytical framework and its underlying assumptions. Our analysis is based on a causal model for treatment ($d = 1$) and control ($d = 0$) outcomes for measure $j$ 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:

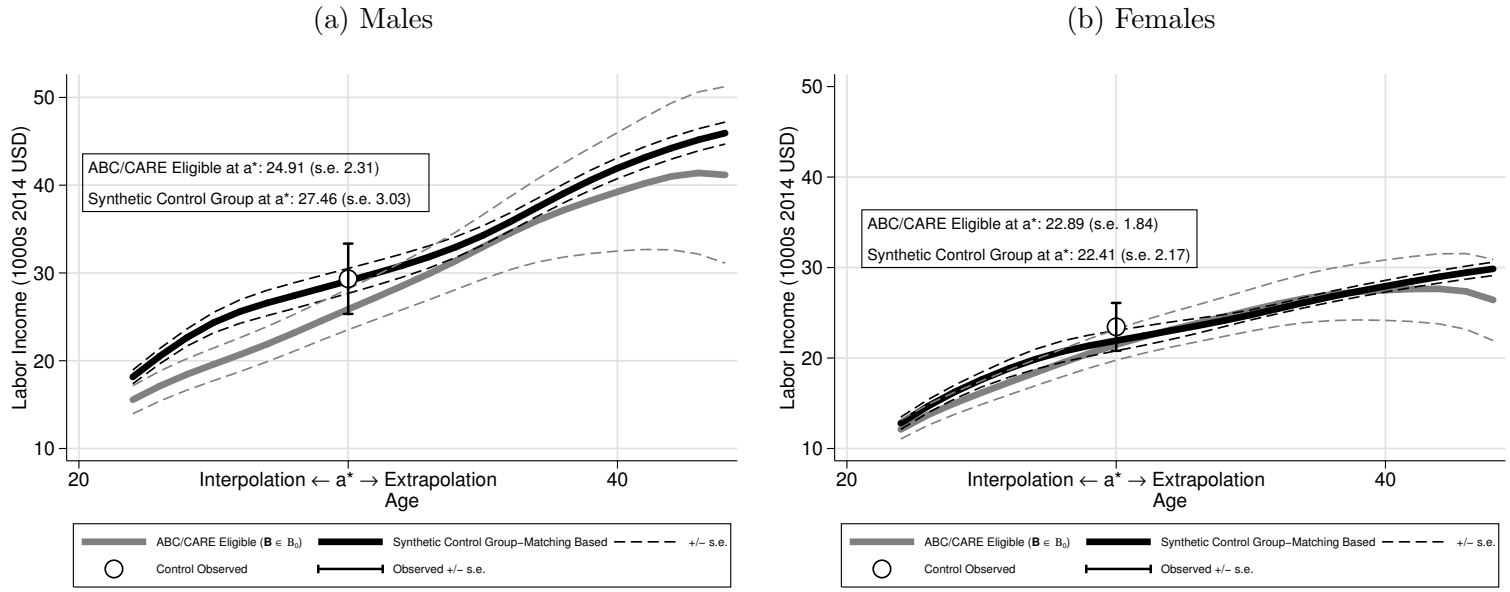
$$Y^d_{k,j,a} = \phi^d_{k,j,a}(\boldsymbol{X}^d_{k,a}, \boldsymbol{B}_k) + \varepsilon^d_{k,j,a}, \quad j \in \mathcal{J}_a. \tag{7}$$

$\phi^d_{k,j,a}(\cdot,\cdot)$ is an invariant structural production relationship mapping inputs $\boldsymbol{X}^d_{k,a}, \boldsymbol{B}_k$ into output $Y^d_{k,j,a}$ holding error term $\varepsilon^d_{k,j,a}$ fixed.[57] We normalize $\varepsilon^d_{k,j,a}$ to have mean zero. Among the $\boldsymbol{X}^d_{k,a}$ are variables caused by treatment, including lagged dependent variables. In this general framework, the relationships between the dependent and right-hand side variables in (7) do not necessarily coincide across the samples, $k \in \{e, n\}$.

[56] The 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 $\boldsymbol{B}$ is available.

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

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




Note: Panel (a) displays the predicted labor income for males in the auxiliary samples for whom $\boldsymbol{B} \in \mathcal{B}_0$, 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 High-Risk Index defining $\boldsymbol{B} \in \mathcal{B}_0$ in the auxiliary samples. Panel (b) displays the analogous figure for females. Standard errors are based on the empirical bootstrap distribution.

Let $\boldsymbol{Y}_k^d$ denote the vector of all outcomes at all ages for $k \in \{e, n\}$, when treatment status is fixed to $d$. Similarly, $\boldsymbol{X}_k^d$ is the vector of all causal predictors of $\boldsymbol{Y}_k^d$ at all ages. Both $\boldsymbol{Y}_k^d$ and $\boldsymbol{X}_k^d$ include the full set of possible outcomes over the life cycle, even though they are not sampled (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 $\boldsymbol{B}_k = \boldsymbol{b}$ by $F_{\boldsymbol{Y}_k^d, \boldsymbol{X}_k^d | \boldsymbol{B}_k = \boldsymbol{b}}(\cdot, \cdot)$.

In the experimental sample, parents of eligible children ($\boldsymbol{B}_e \in \mathcal{B}_0$), 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

$$
\begin{aligned}
Y_{k,j,a} &= (1 - D_k)\, Y_{k,j,a}^0 + (D_k)\, Y_{k,j,a}^1, \\
&\qquad\qquad j \in \mathcal{J}_a, a \in \{1, \ldots, A\}, \quad k \in \{e, n\} \\
\boldsymbol{X}_{k,a} &= (1 - D_k)\, \boldsymbol{X}_{k,a}^0 + (D_k)\, \boldsymbol{X}_{k,a}^1.
\end{aligned} \tag{8}
$$

(We keep the conditioning on $\boldsymbol{B} \in \mathcal{B}_0$ implicit.)

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

### 5.2.1 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 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 auxiliary-sample individuals different from the experiences of the individuals in the experimental sample.

To formalize this problem, let $Y^d_{j,k,a,c,t}$ 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 amounts to avoiding the problem by saying cohort and time effects operate identically across the $e$ and $n$ samples in the following sense:

**Assumption A–1** ***Alignment of Cohort and Time Effects***
*For experimental sample cohort $c_e$ and auxiliary sample cohort $c_n$:*

$$Y^d_{e,a,c_e,t_e} = Y^d_{n,a,c_n,t_n} \tag{9}$$

*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 $n$ ($t_n = a + c_n$).* □

Notice that $Y^d_{n,a,c_n,t_n}$ is the synthetic outcome for treatment status $d$ in the auxiliary sample. This assumption does not rule out cohort or period effects. However, it rules out any *differences* in cohort and time effects for the auxiliary group counterparts and the experimental groups when they reach the age of the auxiliary group.

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 counterpart 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 param-

eterize estimable functions of $c$ and $t$.[58] In the sensitivity analyses reported below, we examine plausible alternative assumptions about cohort and time effects.

### 5.2.2 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 $\boldsymbol{X}_{k,a}^d$, $\boldsymbol{B}$, and $\boldsymbol{Y}_{k,a}$ in the control and experimental samples.

**Assumption A–2** ***Support Conditions***
*For $a \in \{1, \ldots, A\}$, the support of $\left(\boldsymbol{Y}_{e,a}^d, \boldsymbol{X}_{e,a}^d, \boldsymbol{B}_e\right)$ in the experimental sample is contained in the support of $\left(\boldsymbol{Y}_{n,a}^d, \boldsymbol{X}_{n,a}^d, \boldsymbol{B}_n\right)$ in the auxiliary sample:*

$$supp(\boldsymbol{Y}_{e,a}, \boldsymbol{X}_{e,a}^d, \boldsymbol{B}_e) \subseteq supp(\boldsymbol{Y}_{n,a}, \boldsymbol{X}_{n,a}^d, \boldsymbol{B}_n), \quad d \in \{0,1\}. \quad \square \tag{10}$$

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

### 5.2.3 Conditions for Valid Out-of-Sample Predictions

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***

$$F_{\boldsymbol{Y}_e^d, \boldsymbol{X}_e^d | \boldsymbol{B}_e = \boldsymbol{b}}(\cdot, \cdot) = F_{\boldsymbol{Y}_n^d, \boldsymbol{X}_n^d | \boldsymbol{B}_n = \boldsymbol{b}}(\cdot, \cdot), \quad d \in \{0,1\} \tag{11}$$

[58] See Heckman and Robb (1985). For health, cohort effects could be very substantial (e.g. medical costs growth) and we account for this as explained in Section 5.4 and Appendix F.

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

Since we are only interested in means for cost-benefit analysis, we can get by with the following 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***

$$\mathbb{E}\left[\boldsymbol{Y}_e^d | \boldsymbol{X}_e^d = \boldsymbol{x}, \boldsymbol{B}_e = \boldsymbol{b}\right] = \mathbb{E}\left[\boldsymbol{Y}_n^d | \boldsymbol{X}_n^d = \boldsymbol{x}, \boldsymbol{B}_n = \boldsymbol{b}\right], \quad d \in \{0, 1\} \tag{12}$$

*for* $d \in \{0, 1\}$ *over supp* $\left(\boldsymbol{Y}_{e,a}^d, \boldsymbol{X}_{e,a}^d, \boldsymbol{B}_e\right)$.

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***

$$\mathbb{E}\left[\boldsymbol{Y}_e^1 - \boldsymbol{Y}_e^0 | \boldsymbol{B}_e = \boldsymbol{b}\right] = \mathbb{E}\left[\boldsymbol{Y}_n^1 - \boldsymbol{Y}_n^0 | \boldsymbol{B}_n = \boldsymbol{b}\right] \tag{13}$$

*over supp* $\left(\boldsymbol{Y}_{e,a}^d, \boldsymbol{B}_e\right)$.

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 them, and Condition C–2 is useful for doing so.

### 5.2.4 Exogeneity

Conditions C–1 to C–3 do not require that we take a position on the exogeneity of $\boldsymbol{X}_k^d$, $k \in \{e, n\}$. 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.[59] For these purposes, we assume:

**Assumption A–3** ***Exogeneity***
*For all* $a, a'' \in \{1, \ldots, A\}$ *and for* $d, d' \in \{0, 1\}$,

$$\varepsilon^{d}_{k,j,a} \perp\!\!\!\perp \boldsymbol{X}^{d'}_{k,a''} | \boldsymbol{B}_k = \boldsymbol{b} \tag{14}$$

*for all* $\boldsymbol{b}$ *in the support of* $\boldsymbol{B}_k$, $k \in \{e, n\}$, *for all outcomes* $j \in \mathcal{J}_a$, *where "*$\boldsymbol{M} \perp\!\!\!\perp \boldsymbol{N} | \boldsymbol{Q}$*" denotes independence of* $\boldsymbol{M}$ *and* $\boldsymbol{N}$ *given* $\boldsymbol{Q}$. □

We test and do not reject this Assumption for a variety of outcomes in Appendix C.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, but our estimates are robust even when we allow for different failures of Assumption A–3.

### 5.2.5 Structural Invariance

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

**Assumption A–4** ***Structural Invariance***

[59] It also facilitates matching, one of the methods used in this paper. See Heckman and Navarro (2004).

*For all* $\boldsymbol{x}, \boldsymbol{b} \in supp(\boldsymbol{X}^d_{e,a}, \boldsymbol{B}_e), k \in \{e, n\}$

$$\begin{aligned} \phi^0_{k,j,a}(\boldsymbol{x}, \boldsymbol{b}) &= \phi^1_{k,j,a}(\boldsymbol{x}, \boldsymbol{b}) \\ &=: \phi_{j,a}(\boldsymbol{x}, \boldsymbol{b}), \end{aligned} \tag{15}$$

$\phi^d_{k,j,a}(\boldsymbol{x})$ *is the function generating the causal effect of setting* $\boldsymbol{X}^d_{k,a} = \boldsymbol{x}$ *holding* $\varepsilon^d_{k,j,a}$ *fixed for* $a \in \{1, \ldots, A\}$ *for any outcome* $j \in \mathcal{J}_a$. □

This assumption has two distinct messages: (i) the structural functions evaluated at the same arguments have identical values for treatment and control groups in the experimental sample. It also says (ii) that the structural relationships are identical in the experimental and auxiliary samples. As previously noted, exogeneity is not needed to justify any of the Conditions C–1 through C–3. But in the absence of exogeneity, the relationship between the $\boldsymbol{X}^d_{k,a}$ and the errors $\boldsymbol{\varepsilon}^d_{k,a}$ likely differs across experimental ($e$) and non-experimental ($n$) samples because randomization imparts a source of exogenous variation to the $\boldsymbol{X}^d_{e,a}$ not present in non-experimental samples. Assumption A–4 combined with Assumption A–3, Equation (8), and the assumption of a zero mean for the errors ($\mathbb{E}(\varepsilon^d_{k,j,a}) = 0$) for all $a \in \{1, \ldots, A\}, d \in \{0, 1\}$ and $k \in \{e, n\}$ enable us to write:

$$\mathbb{E}\left[Y^d_{k,j,a} | \boldsymbol{X}^d_{k,a} = \boldsymbol{x}, \boldsymbol{B}_k = \boldsymbol{b}, D = d\right] = \mathbb{E}\left[Y_{k,j,a} | \boldsymbol{X}^d_{k,a} = \boldsymbol{x}, \boldsymbol{B}_k = \boldsymbol{b}\right], \tag{16}$$

for $a \in \{1, \ldots, A\}$, $k \in \{e, n\}$, and $d \in \{0, 1\}$.

### 5.2.6 Testable Implications

Equation (16) relates outcomes for $Y^d_{k,j,a}$ to treatment effects for $\boldsymbol{X}^d_{k,a}$, together with the background variables $\boldsymbol{B}_k$. It is possible to test A–4 within the experimental sample ($a \leq a^*$). The test consists of asking if $\boldsymbol{X}^d_{e,a}, \boldsymbol{B}$ predict the within-experimental sample treatment effects for $Y_{e,j,a}$. Under the null hypothesis that A–4 is correct, a separate indicator variable for treatment status ($d$) is irrelevant when computing $\mathbb{E}\left[Y_{e,j,a}|\boldsymbol{X}^d_{k,a} = \boldsymbol{x}, \boldsymbol{B}_k = \boldsymbol{b}, D = d\right]$. In Appendix C.3.7, we test and do not reject the null hypotheses.[60]

Exogeneity and invariance enable us to test additional assumptions. By Equation (16), we can write:

$$\mathbb{E}\left[Y_{e,j,a}|\boldsymbol{X}^d_{e,a} = \boldsymbol{x}, \boldsymbol{B}_e = \boldsymbol{b}\right] = \mathbb{E}\left[Y_{n,j,a}|\boldsymbol{X}^d_{n,a} = \boldsymbol{x}, \boldsymbol{B}_n = \boldsymbol{b}\right] \tag{17}$$

for $d \in \{0, 1\} \quad a \in \{1, \ldots, A\}$.

Relationship (17) is testable for $a \leq a^*$, when $Y_{k,j,a}$ is observed in both the experimental and auxiliary samples. We do not reject the null hypotheses of no differences in the conditional mean functions in the experimental and auxiliary samples conditioning on $\boldsymbol{X}^d_{k,a}$ and $\boldsymbol{B}_k$, $k \in \{e, n\}$.[61]

### 5.2.7 Summarizing the Implications of Exogeneity and Structural Invariance

Collecting results, we obtain the following theorem:

[60] This holds both when pooling males and females and when testing separately by gender (see Appendix C.3.7).

[61] This holds when pooling males and females and when testing by gender (see Appendix C.3.7).

**Theorem 1** ***Valid Out-of-Sample Predictions***
*Under Assumptions A–1-A–4, Conditions C–2 through C–3 hold for any value of* $\left(\boldsymbol{X}_{k,a}^{d}, \boldsymbol{B}_{k}\right)$.
This is an immediate consequence of the cited assumptions. □

### 5.2.8 Testing for Endogeneity

In Appendix C.3.6, we report tests for endogeneity in the experimental and auxiliary samples used in this paper. We follow Heckman et al. (2013) and assume that the $\varepsilon_{k,j,a}^{d}$ obey a factor structure, $k \in \{e, n\}$. We develop that framework and provide evidence supporting exogeneity in both samples for the predictor variables used in our empirical analyses.

### 5.2.9 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 treatment and control groups in the auxiliary sample.[62] Doing so compresses the two stages of constructing a comparison group and creating predictions into one stage. Matching in this fashion creates direct auxiliary counterparts for each member of the experimental samples. It is an intuitively appealing estimator.

We discuss this approach in Appendix C.3.3. Matching is a non-parametric estimator of the conditional mean functions. There is close agreement between non-parametric estimates based on matching and more parametric model-based approaches like the one we use throughout the main text (see Ap-

[62] Heckman et al. (1998) use this procedure.

pendix C.3.4.1).

### 5.2.10 Summarizing Our Approach and What We Do in This Paper

Using the data sources listed in Table 5, we execute our analysis in two stages. In Stage I, we construct comparison samples using background variables ($\boldsymbol{B}_n$). In Stage II, we use the samples so constructed to create post-$a^*$ prediction models based on $\boldsymbol{B}$ and $\boldsymbol{X}_e^d$. We repeat that the structural invariance and exogeneity assumptions discussed in this section are *sufficient* conditions for justifying Conditions C–2 and C–3. We can clearly weaken these assumptions. For example, creating mean difference counterfactuals only requires exogeneity and structural invariance in mean differences, and not in individual level equations. However, the conditions stated for level equations by treatment status have the advantage of being testable on certain subsamples of the data. They also justify matching to make valid predictions. Appendix C.6 reports estimates from a variety of alternative approaches accounting for serial correlation in the errors and from non-parametric matching.

## 5.3 Predicting 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 the treatment effects in maternal labor force participation and parental labor income reported in Tables 3 and 4 and Appendix G.

We observe parental labor income at eight different times for the experimental subjects up through age 21.[63,64] As Tables 3 and 4 show, treatment effects on parental labor income are sustained through the child's age 21. This presumably arises from wage growth due to parental attainment of further education and/or more work experience. 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 5.2. In Section 6, 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 younger.[65]

[63]The 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 in ABC/CARE were, on average, 41 years old.

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

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

## 5.4 Health

We predict and monetize health outcomes based on a version of Equation (7) including a full vector of lagged dependent variables for different indicators of health status. This requires adapting the models of Section 5.2. 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.[66]

Our auxiliary model for health is an adaptation of the Future America Model (FAM). This model predicts health outcomes from the subjects' mid-30s up to their projected age of death (Goldman et al., 2015).[67] Appendix F discusses the FAM methodology in detail.

In Appendix F, we present tests of its assumptions and its predictive performance for population aggregate health and healthy behavior outcomes. FAM passes a variety of specification tests and accurately predicts health outcomes and healthy behaviors. We initialize the health prediction model using the same variables that we use to predict labor and transfer income, along with the initial health conditions as listed in Table 6.

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)

[66]For example, for income we extrapolate up to the retirement age of 67. However, for health, we need to predict an age of death for each individual.

[67]The simulation starts at the age in which we observe the subject's age-30 follow-up.

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) predict the medical expenditures and QALYs that correspond to the simulated individual health trajectories.[68]

Our microsimulation model starts with health predictions for 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 predicting health outcomes. The models predict the probability of being in any of the states in the horizontal axis of Table 6 at age $a+1$ based on the state at age $a$, which is described by the vertical axis of the table.[69] 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 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. Thus, each simulation depends on each individual's health history and on their particular characteristics. For every simulated trajectory of health outcomes, we predict the lifetime medical expenditure using the models estimated from the MEPS and the MCBS. We then obtain an estimate of the expected lifetime medical ex-

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

[69] In practice, the predictions 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.

Table 6: Health State Transitions, Age $a$ as Predictor of Age $a+1$

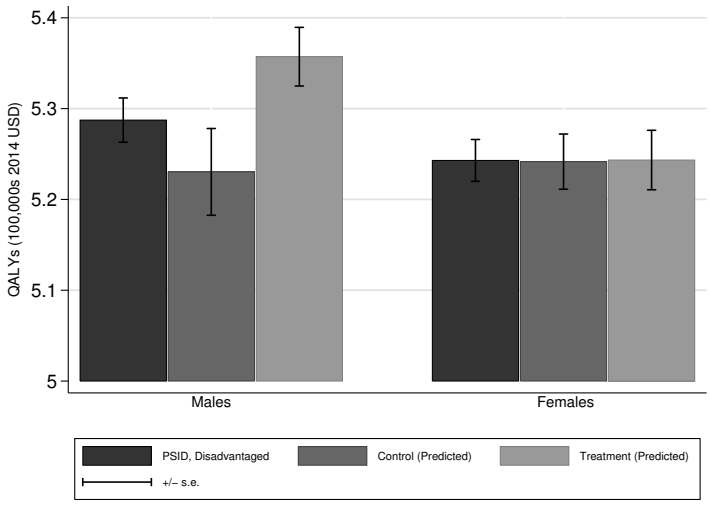
| Age $a$ | Age $a+1$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Heart Disease | Hyper-tension | Stroke | Lung Disease | Diabetes | Cancer | Disability | Mortality | Smoking | Obesity | Health Insurance | DI Claim | SS Claim | SSI Claim |
| Heart Disease | | | × | | | | × | × | × | | × | × | × | × |
| Hypertension | × | | × | | | | × | × | | | × | × | × | × |
| Stroke | | | | | | | × | × | | | × | × | × | × |
| Lung Disease | | | | | | | × | × | × | | × | × | × | × |
| Diabetes | × | × | × | | | | × | × | × | | × | × | × | × |
| Cancer | | | × | | | | × | × | | | × | × | × | × |
| Disability | | | | | | | × | × | | | × | × | × | × |
| Smoking | × | × | × | × | × | × | × | × | × | | × | × | × | × |
| BMI | × | × | × | × | × | × | × | | × | × | | | | |
| Physical Activ. | × | × | × | × | × | × | × | | × | | | | | |
| Binge Drinking | | | | | | | | × | × | | | | | |
| DI Claim | | | | | | | | | | | × | × | × | × |
| SS Claim | | | | | | | | | | | × | | | × |
| SSI Claim | | | | | | | | | | | | | | × |

Note: This table illustrates how health outcomes at age a predict health outcomes at age $a+1$. The crosses indicate if we use the age a outcome to predict the age $a+1$ outcome. DI Claim: payroll-tax funded, federal insurance program for individuals who work long enough paying Social Security taxes; SS Claim: old-age survivors and disability insurance program collected through payroll taxes by the Internal Revenue Service; SSI Claim: claims of stipends for low-income people who are older than 65 years old, blind, or disabled.

penditure by taking the mean of each individual's simulated lifetime 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 prediction using the method described in Appendix F.

Figure 6: Quality Adjusted Life Years: Predictions and Comparison to PSID




Note: This figure displays the life-cycle net-present value of predicted quality-adjusted life years (QALYs) for ABC/CARE males and females, respectively, by treatment status. The predictions are based on combining data from the Panel Study of Income Dynamics (PSID), the Health Retirement Study, and the Medical Expenditure Panel Survey (MEPS). For each gender, we display a comparison to disadvantaged males and females in the Panel Study of Income Dynamics (PSID), where disadvantaged is defined as being Black and having 12 years of education or less. QALYs are the quality-adjusted life years gain due to better health conditions. Standard errors are based on the empirical bootstrap distribution.

The same procedure is applied to calculate quality-adjusted life years (QALYs).[70] We compute a QALY model based on a widely-used health-related

[70] A quality-adjusted life year (QALY) reweights 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

Quality-of-Life (HRQoL) measure (EQ-5D), available in MEPS.[71] 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) out-of-pocket spending (medical spending paid directly by the individual); and (iii) all public spending other than Medicare. Each medical spending model includes the variables we use to predict labor and transfer income, together with current health, risk factors, and functional status as explanatory variables.

We also calculate medical expenditure before age 30. 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 predict medical spending for each age.

QALYs are crucial for our benefit-cost analysis because they monetize the health of an individual at each age. Figure 6 shows our estimation of QALYs together with a PSID comparison, in an exercise analogous to that used to produce Figure 4.[72] Although there is not a clear age-by-age treatment effect

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 lower QALYs. 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 9).

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

[72] In 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 I.

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 individuals in the control group match the QALYs of disadvantaged individuals in the PSID.[73]

## 5.5 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 police records.[74] 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-30's follow-up; and (iii) sentences, gathered at the mid-30's 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.

[73] In Appendix F we further discuss and justify the parameterizations required to obtain estimates of QALYs. Tysinger et al. (2015) examine the sensitivity to these parameterizations and discuss alternative micro-simulations monetizing health condition.

[74] Two previous studies consider the impacts of ABC on crime: Clarke and Campbell (1998) use administrative crime records up to age 21, and find no statistically significant 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, but they lack access to longer term, administrative data.

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 30's, matching crimes across data sources, to construct the total number of arrests for each individual and type of crime up to the mid 30's.[75] For individuals missing arrest data,[76] we impute 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.[77]

2. *Construct predictions.* Based on the sentences observed before age mid 30's, we predict the sentences that the ABC/CARE subjects will have after age mid 30's. Data from the North Carolina Department of Public Safety (NCDPS), which provide lifetime sentences of individuals in North Carolina, are used to estimate sentences incurred after age mid 30's from sentences incurred before age mid 30's. Applying these models to the ABC/CARE data, we predict the number of future sentences for each subject up to age 50.[78] We then add these estimates to the original number of sentences, getting an estimate of the lifetime sentences.

[75] In 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.

[76] About 10% of the ABC/CARE sample has missing arrest data. We fail to reject the null hypothesis of no differences in observed characteristics between the treatment- and control-group participants for whom we observe arrests data (see Appendix A.6).

[77] This arrest-sentence ratio is constructed using the National Crime Victimization Survey (NJRP) and the Uniform Crime Reporting Statistics (UCRS).

[78] We assume that individuals with no criminal records before age mid 30's commit no crimes after age mid 30's.

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).[79] We start by constructing a VI ratio, which is the national ratio of victims and arrests for each type of crime.[80] Then, we estimate the number of victims from the crimes committed by ABC/CARE subjects as their total arrests multiplied by the VI ratio.[81]

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 in arrests and/or sentences, we consider justice system costs as well, such as police costs.[82] Finally, we construct the total costs of incarceration for each subject using the total prison time and the cost of a day in prison.[83]

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

[80] We 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).

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

[82] To 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, the costs only represent about 5% of the total crime costs.

[83] Appendix I examines the sensitivity of our crime costs quantification to different assumptions. Section 6 and Appendix I we examine the sensitivity of our overall assessment of ABC/CARE results to the quantification of crime that we explain in this section.

### 5.6 Education

Follow-up data on educational attainment were collected through age 30. In Appendix D, we show that using auxiliary data sources, education up to this age is an accurate predictor of lifetime educational attainment. Therefore, we do not predict educational attainment beyond age 30. To monetize the costs of education, we consider the public costs of K-12 education and the public and private costs of post-secondary education, including vocational programs and community college. Other costs of education include grade retention and special education. Previous analyses of ABC pay special attention to special education, arguing that savings due to a reduction in this category are substantial (Barnett and Masse, 2002, 2007).[84] This category is much less important in our calculation.[85]

## 6 Benefit/Cost Analysis

This section reports benefit/cost and rate of return analyses underlying Figure 1. Appendix I 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.

[84] Their analyses do not include CARE.

[85] Pooling males and females the net gain due to a reduction in special education is $\$9,724.4$ (2014 USD) (s.e. $\$8,608.1$). For males the gain is $\$14,694.9$ (2014 USD) (s.e. $\$11,065.4$) and for females it is $\$4,077.5$ (2014 USD) (s.e. $\$14,892.0$). This quantity is discounted to the child's birth.

## 6.1 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.[86]

## 6.2 Benefit/Cost Estimates

Table 7 presents our baseline estimates of benefit/cost ratios, Table 8 presents the analogous 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. *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 prediction error in the experimental and auxiliary samples and the tax costs of financing the program.[87] These benefits arise despite the fact that ABC/CARE was much more expensive than other early childhood education programs—the treatment involved more services over a longer time period (Elango et al., 2016).

[86] Our calculations are based on progress reports written by the principal investigators and related documentation recovered in the archives of the research center where the program was implemented. We display these sources in 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).

[87] We obtain the reported standard errors by bootstrapping *all* steps of our empirical procedure, including variable selection, imputation, model selection steps, and prediction error (see Appendix C.8).

We accompany these estimates with a 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 or by the conditioning variables we use when computing the net-present values. Although the internal rate of return remains relatively high when using participant outcome measures up to ages 21 or 30, the benefit/cost ratios indicate that accounting for benefits that go beyond age 30 is important. The return to each dollar is at most 3/1 when considering benefits up to age 30 only (prediction span columns). Accounting for the treatment substitutes available to controls also matters. Males benefit the most from ABC/CARE relative to attending alternative childcare centers, 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 the government using distortionary taxes to fund programs, plus the direct costs associated with collecting taxes.[88] Our baseline estimate assumes that the marginal tax rate is 50%.[89] Our estimates are robust to dropping it to 0% or doubling it to 100% (deadweight loss columns). Our baseline estimate of benefit/cost ratios is based on a discount rate of 3%. Not discounting roughly doubles our benefit/cost ratios, while they remain statistically significant using

[88] When the transaction between the government and an individual is a direct transfer, we consider 0.5 as a cost per each transacted dollar as 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).

[89] Feldstein (1999) reports that the deadweight loss caused by increasing existing tax rates (marginal 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 7, 8, and 9 and in Appendix I, we explore the robustness of this decision and find little sensitivity.

Table 7: Sensitivity Analysis for Benefit/Cost Ratios

| | *Pooled* | | *Males* | | *Females* | |
|---|---|---|---|---|---|---|
| **Baseline** | **7.33** (s.e. 1.84) | | **10.19** (s.e. 2.93) | | **2.61** s.e. 0.73) | |
| *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.* | | | | | | |
| **Specification** | *No IPW* | *and No Controls* | *No IPW* | *and No Controls* | *No IPW* | *and No Controls* |
| | **7.31** | **7.99** | **9.80** | **8.83** | **2.57** | **2.82** |
| | (1.81) | (2.18) | (2.69) | (2.72) | (0.72) | (0.68) |
| **Prediction Span** | *to Age 21* | *to Age 30* | *to Age 21* | *to Age 30* | *to Age 21* | *to Age 30* |
| | **1.52** | **3.19** | **2.23** | **3.84** | 1.46 | **1.81** |
| | (0.36) | (1.04) | (0.61) | (1.60) | (0.36) | (0.50) |
| **Counter-factuals** | *vs. Stay at Home* | *vs. Alt. Presch.* | *vs. Stay at Home* | *vs. Alt. Presch.* | *vs. Stay at Home* | *vs. Alt. Presch.* |
| | **5.44** | **9.63** | **3.30** | **11.46** | **5.79** | **2.28** |
| | (1.86) | (3.10) | (2.95) | (3.16) | (1.37) | (0.76) |
| **Deadweight-loss** | *0%* | *100%* | *0%* | *100%* | *0%* | *100%* |
| | **11.01** | **5.50** | **15.38** | **7.59** | **3.83** | **2.01** |
| | (2.79) | (1.37) | (4.35) | (2.23) | (1.04) | (0.59) |
| **Discount Rate** | *0%* | *7%* | *0%* | *7%* | *0%* | *7%* |
| | **17.40** | **2.91** | **25.45** | **3.78** | **5.06** | **1.49** |
| | (5.90) | (0.59) | (10.42) | (0.79) | (2.82) | (0.32) |
| **Parental Income** | *Mincer Life-cycle* | *Life-cycle Prediction* | *Mincer Life-cycle* | *Life-cycle Prediction* | *Mincer Life-cycle* | *Life-cycle Prediction* |
| | **7.63** | **7.73** | **10.46** | **10.63** | **2.98** | **3.12** |
| | (1.84) | (1.92) | (2.94) | (2.95) | (0.76) | (0.85) |
| **Labor Income** | *.5% Annual Decay* | *.5% Annual Growth* | *.5% Annual Decay* | *.5% Annual Growth* | *.5% Annual Decay* | *.5% Annual Growth* |
| | **7.01** | **7.66** | **9.58** | **10.79** | **2.51** | **2.71** |
| | (1.80) | (1.90) | (2.66) | (3.24) | (0.70) | (0.75) |
| **Crime** | *Drop Major Crimes* | *Halve Costs* | *Drop Major Crimes* | *Halve Costs* | *Drop Major Crimes* | *Halve Costs* |
| | **4.24** | **5.18** | **7.41** | **7.12** | **2.61** | **2.47** |
| | (1.10) | (1.22) | (3.43) | (2.41) | (0.67) | (0.66) |
| **Health (QALYs)** | *Drop All* | *Double Value of Life* | *Drop All* | *Double Value of Life* | *Drop All* | *Double Value of Life* |
| | **6.48** | **8.19** | **9.14** | **11.23** | **2.20** | **3.03** |
| | (1.79) | (2.13) | (2.73) | (3.40) | (0.69) | (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 G for details on how we choose these controls). When predicting 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 2 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). 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 predictions (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 9.

Table 8: Sensitivity Analysis for Internal Rate of Return, ABC/CARE

| | ***Pooled*** | | ***Males*** | | ***Females*** | |
|---|---|---|---|---|---|---|
| **Baseline** | **13.7%** (s.e. 3.3%) | | **14.7%** (s.e. 4.2%) | | **10.1%** (s.e. 6.0%) | |
| *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.* | | | | | | |
| **Specification** | *No IPW* | *and No Controls* | *No IPW* | *and No Controls* | *No IPW* | *and No Controls* |
| | **13.2%** | **14.0%** | **13.9%** | **13.0%** | 9.6% | **10.0%** |
| | (2.9%) | (3.1%) | (3.7%) | (4.3%) | (6.0%) | (4.9%) |
| **Prediction Span** | *to Age 21* | *to Age 30* | *to Age 21* | *to Age 30* | *to Age 21* | *to Age 30* |
| | **8.8%** | **12.0%** | **11.8%** | (12.8%) | **10.7%** | **11.7%** |
| | (4.5%) | (3.4%) | (4.8%) | 4.7% | (5.8%) | (5.2%) |
| **Counter-factuals** | *vs. Stay at Home* | *vs. Alt. Presch.* | *vs. Stay at Home* | *vs. Alt. Presch.* | *vs. Stay at Home* | *vs. Alt. Presch.* |
| | **9.4%** | **15.6%** | 6.0% | **15.8%** | **13.4%** | 8.8% |
| | (4.2%) | (4.3%) | (3.6%) | (5.0%) | (5.7%) | (7.0%) |
| **Deadweight-loss** | *0%* | *100%* | *0%* | *100%* | *0%* | *100%* |
| | **18.3%** | **11.2%** | **19.4%** | **12.1%** | **17.7%** | 7.1% |
| | (4.7%) | (3.1%) | (6.2%) | (3.9%) | (12.4%) | (4.2%) |
| **Parental Income** | *Mincer Life-cycle* | *Life-cycle Prediction* | *Mincer Life-cycle* | *Life-cycle Prediction* | *Mincer Life-cycle* | *Life-cycle Prediction* |
| | **15.2%** | **14.5%** | **16.0%** | **14.5%** | **13.3%** | **12.3%** |
| | (4.0%) | (6.4%) | (5.1%) | (6.4%) | (8.2%) | (9.9%) |
| **Labor Income** | *.5% Annual Decay* | *.5% Annual Growth* | *.5% Annual Decay* | *.5% Annual Growth* | *.5% Annual Decay* | *.5% Annual Growth* |
| | **13.5%** | **13.8%** | **14.5%** | **14.8%** | **9.9%** | **10.3%** |
| | (3.4%) | (3.2%) | (4.3%) | (4.1%) | (6.0%) | (6.0%) |
| **Crime** | *Drop Major Crimes* | *Halve Costs* | *Drop Major Crimes* | *Halve Costs* | *Drop Major Crimes* | *Halve Costs* |
| | **10.7%** | **11.6%** | **12.0%** | **11.9%** | **10.1%** | **9.9%** |
| | (4.4%) | (3.8%) | (5.3%) | (4.9%) | (6.0%) | (6.0%) |
| **Health (QALYs)** | *Drop All* | *Double Value of Life* | *Drop All* | *Double Value of Life* | *Drop All* | *Double Value of Life* |
| | **12.8%** | **13.5%** | **13.5%** | **14.4%** | 8.8% | 9.3% |
| | (4.6%) | (3.6%) | (5.6%) | (4.6%) | (6.4%) | (6.1%) |

Note: This table displays 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 G for details on how we choose these controls). When predicting 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 2 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 predictions (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. N/A: 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 9.

a higher discount rate of 7% (discount rate columns).

Parental labor income 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 5.3). Our baseline estimates rely solely on observed parental labor income when participant children were 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 (parental labor income columns). As noted in Section 5, our estimates ignore any 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 $-0.5\%$ to $0.5\%$.[90]

[90] We account for cohort effects in health as explained in Section 5.4.

Table 9: Cost/benefit Analysis of ABC/CARE, Summary

| | Females | | | Males | | | Pooled | | |
|---|---|---|---|---|---|---|---|---|---|
| Removed Component | NPV | IRR | B/C | NPV | IRR | B/C | NPV | IRR | B/C |
| None | 161,759 | **10.1%** | **2.61** | 919,049 | **14.7%** | **10.19** | 636,674 | **13.7%** | **7.33** |
| | | (6%) | (0.73) | | (4%) | (2.93) | | (3%) | (1.84) |
| Parental Income | 148,854 | 4% | 1.12 | 107,907 | **11%** | **9.10** | 116,953 | **9%** | **6.17** |
| | | (2%) | (0.65) | | (3%) | (2.92) | | (3%) | (1.87) |
| Subject Labor Income | 41,908 | 9% | **2.21** | 238,105 | **13%** | **7.75** | 133,032 | **13%** | **6.03** |
| | | (6%) | (0.66) | | (5%) | (2.23) | | (4%) | (1.77) |
| Subject Transfer Income | 419 | **10%** | **2.61** | -7,265 | **15%** | **10.26** | -4,372 | **14%** | **7.38** |
| | | (6%) | (0.73) | | (4%) | (2.93) | | (3%) | (1.84) |
| Subject QALY | 42,102 | 9% | **2.20** | 106,218 | **14%** | **9.14** | 87,181 | **13%** | **6.48** |
| | | (6%) | (0.69) | | (6%) | (2.73) | | (5%) | (1.79) |
| Medical Expenditures | -16,037 | 9% | **2.77** | -42,038 | **15%** | **10.61** | -31,221 | **14%** | **7.65** |
| | | (6%) | (0.76) | | (3%) | (2.89) | | (3%) | (1.85) |
| Alternative Preschools | 16,691 | 8% | **2.45** | 13,434 | **14%** | **10.05** | 14,659 | **12%** | **7.19** |
| | | (5%) | (0.73) | | (4%) | (2.92) | | (3%) | (1.84) |
| Education Costs | 1,457 | **10%** | **2.59** | -7,852 | **15%** | **10.26** | -4,518 | **14%** | **7.37** |
| | | (6%) | (0.72) | | (4%) | (2.93) | | (3%) | (1.86) |
| Crime Costs | 31,668 | 10% | **2.34** | 638,923 | **9%** | 4.08 | 450,368 | **8%** | **3.06** |
| | | (6%) | (0.62) | | (5%) | (2.18) | | (4%) | (1.01) |
| Deadweight Loss | | **18%** | **3.83** | | **19%** | **15.38** | | **18%** | **11.01** |
| | | (12%) | (1.04) | | (6%) | (4.35) | | (5%) | (2.79) |
| 0% Discount Rate | | | **5.06** | | | **25.45** | | | **17.40** |
| | | | (2.82) | | | (10.42) | | | (5.90) |
| 7% Discount Rate | | | **1.49** | | | **3.78** | | | **2.91** |
| | | | (0.32) | | | (0.79) | | | (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 rest of the 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 B/C 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 $92,570$.

We also examine the sensitivity of our estimates to (i) dropping the most costly crimes such as murders and rapes;[91] and (ii) halving the costs of victimization and judiciary costs related to crime. The first sensitivity check is

[91] Two individuals in the treatment group committed a rape and one individual in the control group committed a murder.

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). Our benefit/cost analysis is robust to these adjustments, even when crime is a major component. Lastly, we examine the sensitivity with respect to our main health component: quality-adjusted life years. This is an important component because relatively healthier individuals survive longer and healthier, and treatment improves health conditions. It is important to note that this component largely accumulates later in life and therefore it 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 drastic sensitivity analysis by removing components of the benefit/cost analysis entirely (Table 9).[92] 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 the 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 a child's life. 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 also investigate how sensitive our prediction model

[92] In Appendix I, we present exercises that are not as drastic as removing the whole component, but instead remove fractions of it.

is to a variety of perturbations: different autocorrelation processes in the prediction errors, predictions without lagged variables, etc. Our estimates are robust to using different prediction models.

*Overall, our sensitivity analyses indicate that no single category of outcomes drives the social efficiency of the program. Rather, it is the life-cycle benefits across multiple dimensions of human development.*

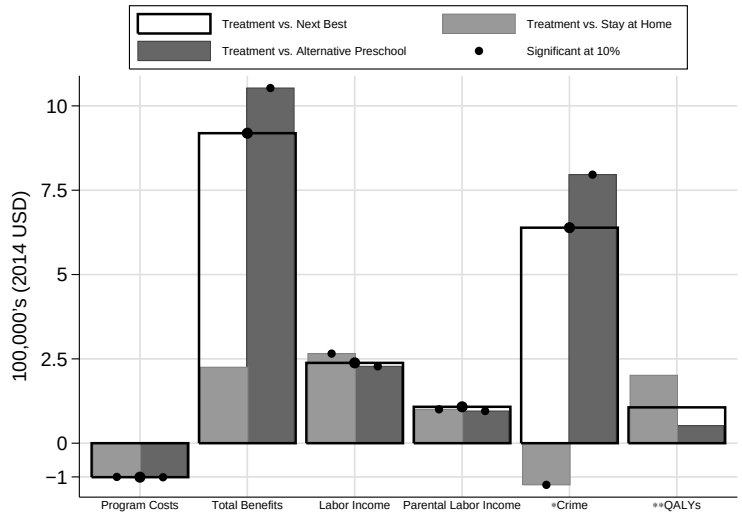
In Appendix J we use our analysis to examine the empirical foundations of recent cost-benefit studies of early childhood programs that use short term estimates of experimental test score gains coupled with auxiliary estimates of the impact of test scores on earnings (see, e.g., work by Kline and Walters, 2016). We show that this approach applied to the ABC data greatly understates the true benefit-cost ratio because (a) earnings are only predicted through a young age (27) and (b) benefits extend beyond earnings. The difference is sizable. Applying the Kline and Walters method to the ABC/CARE experiment, we would find a cost-benefit ratio of 1.4 compared to our estimate of 7.3.

## 6.3 Possible Explanations for Gender Differences

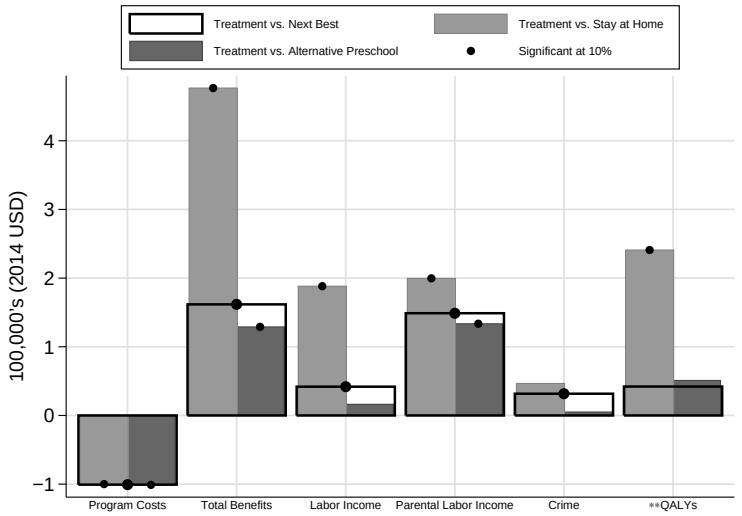
The benefit/cost ratio and internal rate of return calculations both indicate that males and females benefit *differently* from the program compared to the alternatives "$H$" and "$C$". There are two complementary stories that help to explain this difference. First, gender differences could exist as a consequence of the outcomes monetized, and not because of the particular counterfactuals that we estimate. Males have relatively high benefits from the outcomes that we are able to monetize. Labor income and crime are prime examples of this.

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 predicted 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 predicted death, with $\$150,000$ (2014 USD) as base value for a year of life.

Females are less likely to work than males. While all males supply labor in our sample at age 30, not all females do. We are not able to quantify household production benefits for either males or females. This is an important omission for females who decide to stay at home instead of work. For example, we lack data on their children's outcomes.

ABC/CARE has treatment effects on crime for females for a number of categories (see Appendix G). However, males are much more likely to commit crimes that are more costly to the victims, to the criminal justice system, and to society (Cohen and Bowles, 2010; Barak et al., 2015). ABC/CARE also has treatment effects on crime for females for a number of categories (see Appendix G). However, males commit crimes that are much more expensive to society. These two categories are examples of why the magnitudes of the gains are much higher for males than they are for females.

For health, there are also substantial gender differences. Both males and females have substantial gains: males benefit on more standard measures of physical health, and females benefit on a set of mental health measures (see Appendix G). We quantify both components (see Section 5.4 and Appendix F).

There is a second factor at work. There are substantial differences between males and females in one counterfactual: treatment vs. alternative preschools. The estimated treatment effects are very similar across genders for treatment compared to those staying at home full time. Males benefit much more from treatment relative to alternative preschools compared to their benefits from treatment relative to staying at home. This result is consistent with findings noted elsewhere: (i) stark gender differences resulting from attending low

quality childcare (Kottelenberg and Lehrer, 2014; Baker et al., 2015); and (ii) females are less sensitive to uncertain environments (see, e.g., Autor et al., 2015).

Our evidence does not indicate that the program has no benefits for females. When compared to staying at home, there is a gain of 4.93 dollars per each dollar invested. When we decompose the net-present value for each of the components that we monetize, we find substantial benefits for females across a variety of categories, including health and crime. For males, the magnitudes are noticeably increased when comparing outcomes from treatment to outcomes from attending alternative preschools (see Figure 7).

# 7 Summary

This paper studies two influential early childhood programs evaluated by the method of randomized control trials with long term follow-up through the mid 30's. These programs are emulated in a variety of active early childhood programs around the world. We document outcomes across multiple life domains. We estimate a statistically significant aggregate benefit/cost ratio of 7.3 and a rate of return of 13.7% per annum, even after adjusting for the welfare costs of financing the program through taxation.

To reach these conclusions, we address a number of empirical challenges: (i) control group substitution; (ii) extrapolating lifetime benefits beyond the experimental period; and (iii) the multiplicity of hypotheses tested. Our approach serves as a template for evaluations of programs with partial follow up

over the life cycle.

Benefits differ substantially by gender. Females have more beneficial treatment effects than males, but the monetized value of the male treatment effects is greater. There are substantial effects on health and (health-related) quality of life as well as crime for males. For females, the benefits are concentrated in education, employment, and minor crimes. The effects for females are stronger compared to the alternative of staying at home. The effects for males are stronger compared to the alternative of participation in alternative childcare arrangements. The program subsidizes maternal employment and has a strong causal effect on maternal labor income. We demonstrate the long-term multiple benefits of these programs.

# References


Autor, D. H., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2015, October). Family disadvantage and the gender gap in behavioral and educational outcomes. Working Paper WP-15-16, Institute for Policy Research, Northwestern University, Evanston, IL.

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.

Barak, G., P. Leighton, and A. Cotton (2015). *Class, Race, Gender, and Crime: The Social Realities of Justice in America*. Rowman and Littlefield.

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.

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.

Clarke, S. H. and F. A. Campbell (1998). Can intervention early prevent crime later? The Abecedarian Project compared with other programs. *Early Childhood Research Quarterly 13*(2), 319–343.

Cohen, M. A. and R. Bowles (2010). Estimating costs of crime. In A. R. Piquero and D. Weisburd (Eds.), *Handbook of Quantitative Criminology*, Chapter 8, pp. 143–161. New York: Springer.

Collins, A., B. D. Goodson, J. 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.

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.

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, A. Hojman, Y. K. Koh, J. Shea, and A. Ziff (2016). Documentation of full ABC/CARE treatment effects. Unpublished Manuscript, University of Chicago.

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.

Haskins, R. (1985, June). Public school aggression among children with varying day-care experience. *Child Development 56*(3), 689–703.

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 (Accessed 8/1/2016).

Kline, P. and C. Walters (2016). Evaluating public programs with close substitutes: The case of Head Start. Forthcoming, *Quarterly Journal of Economics*.

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.

Lehmann, E. and J. P. Romano (2005, June). Generalizations of the familywise error rate. *Annals of Statistics 33*(3), 1138–1154.

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.

Murray, C. (2013). *Giving Kids a Fair Chance*, pp. 63–67. MIT Press.

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. and R. Haskins (1981). The modification of intelligence through early experience. *Intelligence 5*(1), 5–19.

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. and B. J. Smith (1977, January). Assessing the intellectual consequences of early intervention with high-risk infants. *American Journal of Mental Deficiency 81*(4), 318–324.

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.

Romano, J. P. and A. M. Shaikh (2006, August). Stepup procedures for control of generalizations of the familywise error rate. *Annals of Statistics 34*(4), 1850–1873.

Romano, J. P. and M. Wolf (2016). Efficient computation of adjusted $p$-values for resampling-based stepdown multiple testing. *Statistics and Probability Letters 113*, 38–40.

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.

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. and I. Lewis (1979). *LearningGames for the First Three Years: A Guide to Parent/Child Play*. New York: Walker and Company.

Sparling, J. and I. Lewis (1984). *LearningGames for Threes and Fours*. New York: Walker and Company.

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. E. 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.