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

Jorge Luis García
Department of Economics
The University of Chicago

James J. Heckman
American Bar Foundation
Department of Economics
The University of Chicago

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

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

Contents

List of Figures

List of Tables

A   Background[1]

A.1   Overview

The Carolina Abecedarian Project (ABC) and the Carolina Approach to Responsive Education (CARE) were high-quality early childhood education programs. Each had two phases of randomized controlled design. They were both implemented at the Frank Porter Graham Center (FPGC) of the University of North Carolina in Chapel Hill. ABC served four cohorts of children born between 1972 and 1977, and CARE served two cohorts of children born between 1977 and 1980. In this section of the appendix, we expand on important details of the eligibility requirements, the randomization protocol, and the programmatic contents of both programs. Table A.1. offers a visual summary and comparison of the two programs.

A.2   Eligibility Criteria and Populations Served

The mothers of the ABC and CARE subjects were generally recruited during the last trimester of pregnancy. Potential families were referred by local social service agencies and hospitals. Eligibility was determined by a score of 11 or more on a weighted 13-factor High-risk Index (HRI). Table A.2 details the items of the HRI for ABC.

The HRI for CARE was similar to that of ABC—it also contained 13 weighted variables and a score of 11 or above was required to be considered eligible. The items for maternal and paternal education levels have the same categories and weights as the ABC HRI. The other identical items are having an absent father, school-age siblings performing lower than the norm based on grade-level or achievement tests, a record of father's unstable job history or unskilled labor, social agencies indicating a high level of need, and other circumstances

---

[1]Sylvi Kuperman greatly assisted us in preparing this section of the appendix.

Table A.1: 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	56	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	58	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**			
Treatment N	49	39	
Control N	47	22	
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.

** We do not analyze this sample in this paper and mention it only for completeness.

Table A.2: High-risk Index for ABC

	Item	Response	Weight
1	Maternal education (years of education)	6	8
		7	7
		8	6
		9	3
		10	2
		11	1
		12	0
2	Paternal education (years of education)	same as maternal education	
3	Year family income (2014 USD)	$5,663.54 or less	8
		$5,663.54-$11,327.08	7
		$11,327.08-$16,990.62	6
		$16,990.62-$22,654.16	5
		$22,654.16-$28,317.70	4
		$28,317.70-$33,981.24	0
4	Father's absence from the household for reason other than health or death	Yes	3
5	Lack of maternal relatives in the area	Yes	3
6	Siblings in school age one or more grades behind age-appropriate level or low scores on school-administered achievement tests	Yes	3
7	Received payments from welfare agencies within the past 3 years	Yes	3
8	Father's work unstable or unskilled and semi-skilled labor	Yes	3
9	Maternal or paternal IQ 90 or below	Yes	3
10	Sibling with an IQ score 90 or below	Yes	3
11	Relevant social agencies indicate that family is in need of assistance	Yes	3
12	One or more family members has sought professional help in the past 3 years	Yes	1
13	Special circumstances not included in any of the above that are likely contributors to cultural or social disadvantage	Yes	1

Note: This table shows the High-risk Index (HRI) for ABC. A score of 11 or more determined eligibility (Ramey and Smith, 1977; Ramey and Campbell, 1984, 1991; Ramey et al., 2000). The weighting scale aimed to establish the relative importance of each item in the index (Ramey and Smith, 1977). Race was not considered for eligibility; however, 98% of the families who agreed to participate were African-American(Ramey and Smith, 1977; Ramey and Campbell, 1979).

related to cultural or social disadvantage. The specification of the following items were changed between the ABC and CARE HRI. The weight associated with household income depended on the number of individuals in the family for CARE and the income categories range from less than \$11,327.08 to \$76,457.80 (2014 USD) or more. In the CARE HRI, it is determined if payments were received from welfare agencies in the past 5 years instead of the past 3 years. Similarly, it determines if any family member has sought counseling in the past 5 years instead of the past 3 years. The threshold for maternal or paternal IQ is 85 in the CARE HRI instead of 90 as in the ABC HRI. It does not have an item related to the absence of maternal relatives in the area, but replaces that item with asking if any member of the mother or father's immediate family has received services for the mentally disabled (the weight for this item is 3).[2]

All subjects were substantially disadvantaged (see Figure A.1 and Figure A.2). Maternal age when the subject was born was, on average, 19.9 years in ABC and 21.1 years in CARE. Approximately half of the mothers of both treatment-group and control-group subjects in ABC were 19 years or younger and one third were 17 years or younger. In CARE, approximately half of the mothers of both treatment-group and control-group subjects were 20 years or younger and one third were 17.2 years or younger. Mean maternal IQ score in ABC was approximately 85, one standard deviation below the national mean. In CARE, the mean maternal IQ score was approximately 87. Only 25% of the ABC subjects lived with both biological parents, and more than 50% lived with extended families in multi-generational households (61% of treatment-group subjects and 56% of control-group subjects).[3] About 79% of subjects did not have a father in the home in both ABC and CARE.

---

[2]Ramey et al. (1985).
[3]Ramey and Campbell (1991); Campbell and Ramey (1994).

Figure A.1: High-risk Index Distribution, ABC
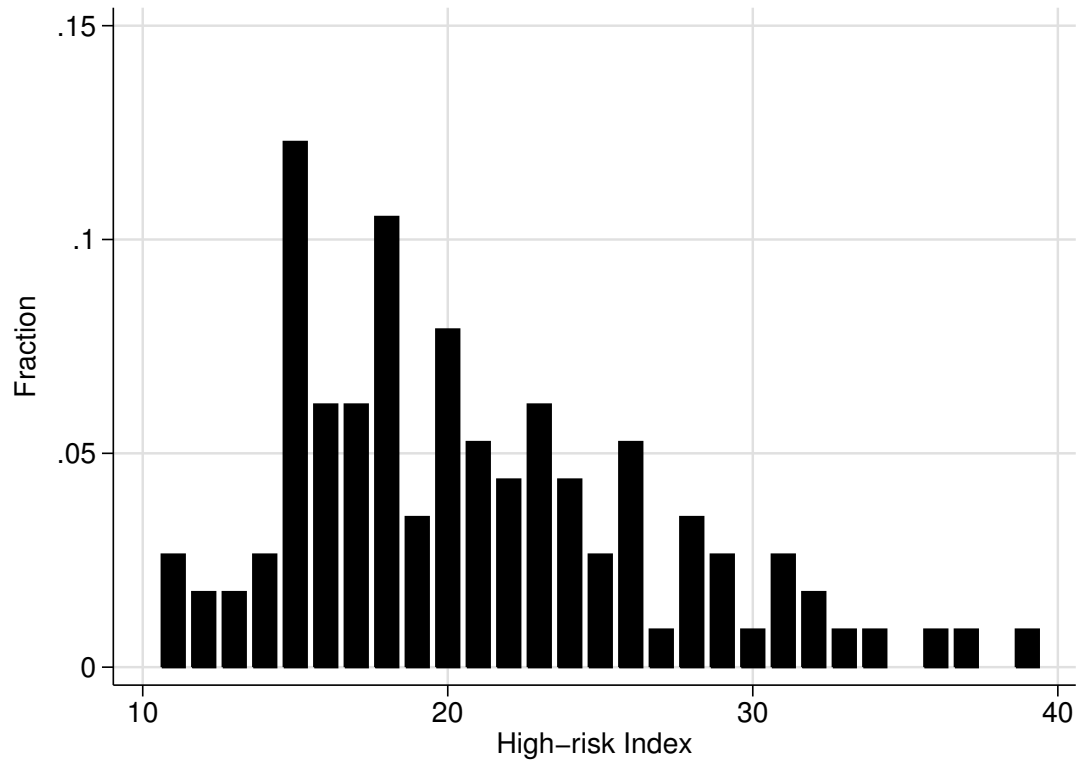
Note: This plot shows the distribution of the High-risk Index (HRI) for ABC, which determined eligibility. Subjects were eligible if they had a score of 11 or more.

Figure A.2: High-risk Index Distribution, CARE
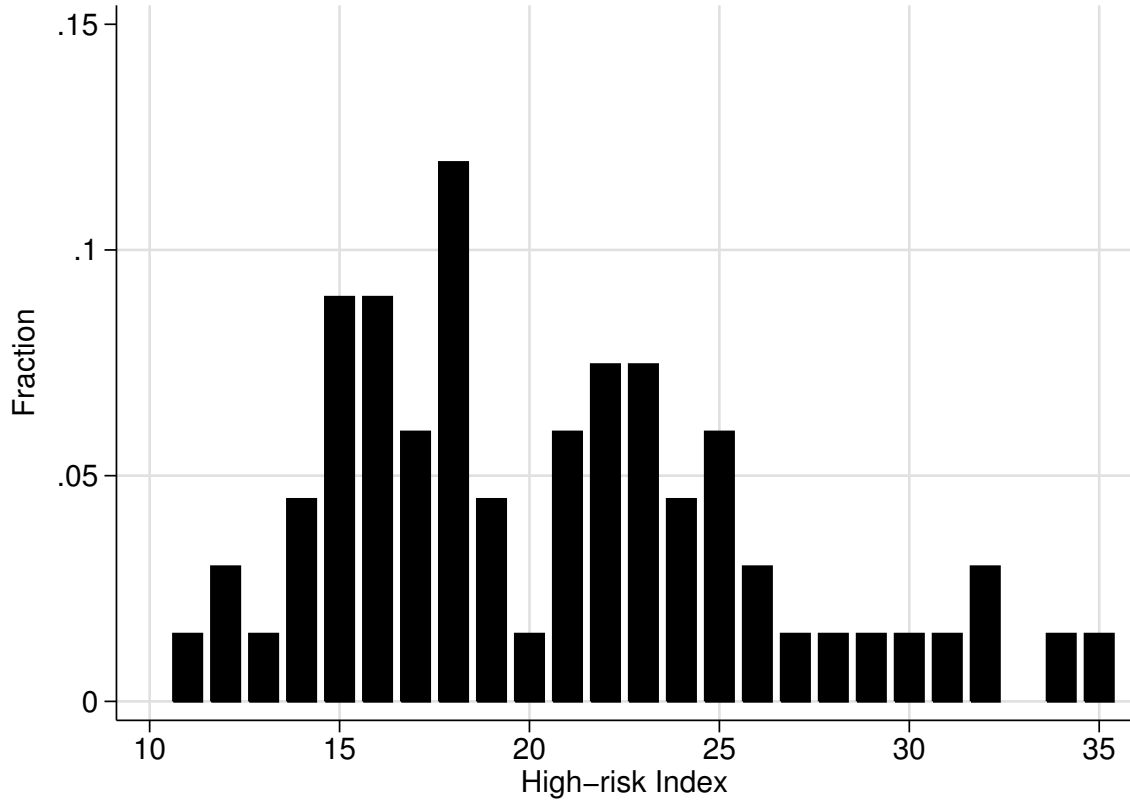
Note: This plot shows the distribution of the High-risk Index (HRI) for CARE, which determined eligibility. Subjects were eligible if they had a score of 11 or more.

A.3 Randomization Protocol and Compromises

Randomization compromises throughout ABC's and CARE's implementations pose a challenge for evaluating the programs' effects. We discuss each case of compromise in detail. Figure A.3 and Figure A.4 are flow charts that depict the sample from the first-phase randomization through the last data follow-up accounting for all cases of attrition and non-compliance.

Although most randomization compromises occurred at early stages, our methodology also accounts for the fact that a few subjects were not in the sample either for the second-phase randomization or for the adult follow-ups. In Appendix A.7, we describe the sample reductions that attrition at different stages of the study generates and test potential differences between the subjects who completed data follow-ups and the subjects who did not.

Figure A.3: Randomization Protocol and Treatment Compliance, ABC

112 families were paired-matched; 114 children

56 pairs
+ 1 twin
+ 1 sibling

Each pair was randomly assigned
but siblings stayed together

Twins
Siblings

Preschool
Randomization [R]

Preschool Treatment Group [R=1]
N=58 children

[R=1]   [R=0]

Preschool Control Group [R=0]
N=56 children

Abandoned study=4
Died before age 5=2
Developmentally=2
delayed

Died before age 5=2
Withdrew before age 5=1

[R=0]
Replacement=1

[R=1]
Replacement=7

Crossover=1
Non-randomly reassigned=2

[R=0;D=1]=3   [R=0;D=0]=51

[R=1;D=0]=7

Non-compliance=3

[R=1;D=1]=50

Participated in ABC Program [D=1]
N=56 children

Not participated in ABC Program [D=0]
N=54 children

Not=3
Followed

School-Age
Randomization [SR]

Not=3
Followed

[R=1;D=1;SR=No]=1   [R=1;D=1;SR=1]=25   [R=0;D=0;SR=1]=24   [R=1;D=1;SR=0]=24   [R=0;D=0;SR=0]=23   [R=0;D=0;SR=No]=4

Temporary Attrited=5

Temporary Attrited=4

School-Age Treatment Group [SR=1]
N=49 children

School-Age Control Group [SR=0]
N=47 children

No Show=3

Participated
N=46 children

Not Participated
N=47 children

Attrited=1

Adult Follow-Up (Age 21)

[R=1; SR=1]=25   [R=0; SR=1]=24   [R=1; SR=0]=19   [R=0; SR=0]=27   [R=1; SR=No]=4

Details on Figure A.3: Sources: Ramey et al. (1976); Ramey and Smith (1977); Ramey and Campbell (1979, 1984), internal documentation of the program, and own calculations. Note: The variable R represents randomization into treatment, $[R = 1]$, or control, $[R = 0]$, groups. After the original randomization, some subjects died or withdrew from the program early in life and were replaced. R also includes those replacements. Arrows pointing outside of the diagram indicate subjects who left the study permanently. The variable D represents participation in the preschool-age program. The variable SR represents randomization into the *school-age* program, $[SR = 1]$, or out of it, $[SR = 0]$. Some subjects were not randomized at school age, $[SR = No]$. We use the term "temporarily attrited" for subjects who did not participate in the study at school age, but were later interviewed in the age-21 followup.

Figure A.4: Randomization Protocol and Treatment Compliance, CARE

Preschool Randomization [R]
N=67 children; 65 families

[R=2]

[R=1]

[R=0]

Center-based & F. Visits Group [R=2]
N=17 children

Family Visits Group [R=1]
N=27 children; 25 families

Control Group [R=0]
N=23 children

Abandoned study=1
Moved=2

[R=2;P=2]=14

Died=1

[R=1;D=1]=25

Temporary=1
Attrited

Moved=1

[R=0;D=0]=22

Participated [D=2]
N=14 children

Participated [D=1]
N=25 children

Not Participated [D=0]
N=22 children

[R=2;D=2]=14

[R=1;D=1]=25

[R=0;D=0]=22

School-Age Treatment Group
N=39 children

School-Age Control Group
N=22 children

Participated
N=39 children

Not Participated
N=22 children

Attrited=1

Attrited=3

Adult Follow-Up (Age 21)

[R=2]=13

[R=1]=26

[R=0]=19

Details on Figure A.4: Sources: Wasik et al. (1990), internal documentation of the program, and own calculations. Note: The variable R represents randomization into center-based childcare and family education, $[R = 2]$, family education, $[R = 1]$, or control, $[R = 0]$. Arrows pointing outside of the diagram indicate subjects who left the study permanently. The variable D represents participation in the corresponding group of the preschool-age program. The variable SR represents those who participated in the *school-age* program, $[SR = 1]$, or did not, $[SR = 0]$. Unlike in ABC, there was no second-phase randomization in CARE. Rather, those in the center-based childcare and family education group and those in the family education group were automatically assigned to receive the school-age treatment. We use the term "temporarily attrited" for subjects who did not participate in the study at school age, but were later interviewed in the age-21 followup.

A.3.1 ABC

Both the first and second phases of randomization were conducted at the family level, so pairs of siblings and twins were jointly randomized into either treatment or control groups.[4] Although we know that pairing was based on HRI, maternal IQ, maternal education, maternal age, and gender of the subject, we do not know the original pairs. The study collected an initial sample of 120 families, with a total of 123 potentially eligible children. After removing 2 subjects who were mentally disabled, 4 subjects who left the study for unspecified reasons, and 3 subjects who died before 6 months of age, there are 114 subjects from 112 families who are part of the initial sample. Including these 9 subjects, twenty two subjects did not complete the first-phase of treatment as initially assigned by the randomization. We classify these subjects into eight categories in Table A.3 and describe them next.[5]

First, there were four subjects assigned to treatment who left the study before any data on them was collected. In our main methodology, we assume that they are missing at random.

Second, four subjects died before age 5—two of them initially assigned to treatment and two of them initially assigned to control. For all of them, we observe baseline characteristics and any other data collected before their death. For methodological purposes, they represent cases of program attrition when we do not observe their outcomes.

Third, three subjects in the treatment group did not comply to treatment status. They are different from the four subjects who left the study before any data collection because we observe data collected for them from birth to age 8. Afterward, the program staff chose not to follow them anymore.[6] Therefore, these subjects remain in treatment sample until age 8 or before. After, they represent cases of program attrition, given that we do not observe

---

[4]Sibling pairs occurred when the two siblings were close enough in age such that both of them were eligible for the program.

[5]In Appendix C, we compare the observed baseline characteristics of the subjects in Table A.3 to the observed baseline characteristics of the subjects who complied to the initial treatment assignment. We find little evidence of differences.

[6]Informal conversations with the program's staff do not indicate a clear reason for this.

them anymore.

Table A.3: Randomization Compromises, ABC

Case	Initial Assignment	Compromise Description	Age at Departure	Data Availability	Methodology Assumption
1	Treatment	Left the study	0	None	Missing at random
2	Treatment	Left the study	0	None	Missing at random
3	Treatment	Left the study	0	None	Missing at random
4	Treatment	Left the study	0	None	Missing at random
5	Control	Death (age 0), heart disease	0	Baseline; before dead	Attrition after death
6	Control	Death (age 0), heart disease	0	Baseline; before dead	Attrition after death
7	Treatment	Death (age 0), SIDS	0	Baseline; before dead	Attrition after death
8	Treatment	Death (age 4), pedestrian accident	4	Baseline; before dead	Attrition after death
9	Treatment	Non-compliance	Do not depart	Baseline; before age 8	Attrition after age 8
10	Treatment	Non-compliance	Do not depart	Baseline; before age 8	Attrition after age 8
11	Treatment	Non-compliance	Do not depart	Baseline; before age 8	Attrition after age 8
12	Control	Crossover from control to treatment	Do not depart	Baseline; before age 8	Attrition after age 8
13	Treatment	3 months of treatment	3 months	Baseline; after age 2	Same as treatment group
14	Treatment	10 months of treatment	10 months	Baseline; after age 2	Same as treatment group
15	Treatment	6 months of treatment	6 months	Baseline; after age 2	Same as treatment group
16	Treatment	9 months of treatment	9 months	Baseline; after age 2	Same as treatment group
17	Control	Left study at 54 months	54 months	Baseline; before 54 months	Attrition after 54 months
18	Treatment	Developmentally delayed at 6 months	6 months	No data after diagnosis	Dropped (non-eligible)
19	Treatment	Developmentally delayed at 3 years	3	No data after diagnosis	Dropped (non-eligible)
20	Control	Crossover from control to treatment	Do not depart	Baseline, before age 8	Attrition after age 8
21	Control	Crossover from control to treatment	Do not depart	Baseline, before age 8	Attrition after age 8

Note: This table describes the various randomization compromises in ABC. For each subject, we display: the nature of the compromise, the available data, and the methodological assumption when accounting for non-compliance and program attrition. The case numbers do not have anything to do with individual identifiers of program participants.

Fourth, one subject initially assigned to control was enrolled into treatment. The mother wanted to work and the program staff decided to admit her child into center-based care.[7] Both in terms of data collection and in terms of methodological purposes, this subject is analogous to the subjects in the third case.[8]

Fifth, four subjects in the treatment group did not complete treatment in its entirety. They were treated for at most 10 months. Except for follow-ups during childhood, which our main results do not use, we observe most of the data for these subjects. We avoid taking a stand on how beneficial the program was at each age, because we do not have a way to document this. Therefore, we assume that they were treated as other subjects in the treatment group.[9]

Sixth, the family of one subject in the control group moved at age 54 months. We observe data before the family moved, so we consider the subject as part of the control group in any estimation before this event. Afterwards, we do not observe any data on the subject, so we consider her a case of program attrition.

Seventh, two subjects initially assigned to treatment status were diagnosed as developmentally delayed after 6 and 36 months of treatment. No data for them are available after the diagnosis. We drop them from the sample because they were not eligible to be part of the program.

Eighth, two subjects initially assigned to the control group were admitted into treatment. Local authorities requested this because the children were considered highly at risk. Data on them are available from birth to age 8. Although they crossed over from the control group to the treatment group, we consider them to be members of the control group who attrited after age 8.

Analysis of each of these cases leads to the following conclusions. For four subjects, we

---

[7]Correspondence with the program officers stating this permission is available under request from the authors.

[8]The sensitivity analysis finding little evidence when adjusting for non-compliance includes this case.

[9]If anything, this downward biases the effects of the program we estimate.

do not have data to assess them as cases of program attrition, though sensitivity analyses suggest that the treatment effects of the program persist after assigning them the same outcome as the subjects who did the worst in the treatment group. For the subjects who did not comply with treatment, adjusting our estimates for non-compliance when data are available makes little difference. The remaining 14 subjects who did not complete treatment as initially assigned represent various cases of program attrition, for which we propose a correction methodology in Appendix C.2.

To increase the number of subjects in the sample, the program officers recruited additional subjects who were added to the program before the subjects were 6 months old. Our calculations indicate that there were eight initial replacements. We cannot distinguish in the data the subjects who were initially randomized from the replacement children and there is no documentation on how these subjects were recruited.[10] After the various compromises, the sample consisted of 114 subjects: 58 in the treatment group and 56 in the control group. The observed characteristics for each subjects indicate that they were eligible for the program; all subjects in the sample have an HRI of 11 or above.

Prior to the second phase of randomization, 3 subjects in the first-phase control group and 3 subjects in the first-phase treatment group could not be located for follow-up. One subject in the control group and eight subjects in the treatment group of the first phase did not participate in the second phase but later agreed to participate in the data collections during adulthood. This yields a sample of 96 subjects in the second phase: 49 in treatment and 47 in control. After the second-phase randomization, three subjects in the treatment group chose not to participate in the program, while all subjects in the control group adhered to their randomization status.

---

[10]Three replacements are reported in Ramey and Campbell (1979). Three are documented in correspondence with the program officers, which is available from the authors upon request. The other two replacements are implied by the number of subjects who participated in the randomization protocol in each cohort.

A.3.2 CARE

The randomization protocol in CARE had no major compromises.[11] Of the 65 initial families, 23 were randomized to a control group, 25 to the family education treatment group (we do not consider this group in our combined ABC/CARE sample), and 17 to the family education and center-based childcare treatment group. Two families in the family education treatment group had twins who were jointly randomized, as in ABC. We document four cases of program attrition (see Table A.4).[12] For methodological purposes, we consider these subjects analogous to their corresponding cases in ABC. We do not present exercises to evaluate the sensitivity to non-compliance because there was none in CARE. Figure A.4 illustrates CARE's randomization protocol and the presence of subjects throughout the data follow-ups.

---

[11]Wasik et al. (1990); Burchinal et al. (1997).

[12]In Appendix C, we compare the observed baseline characteristics of the subjects in Table A.4 to the observed baseline characteristics of the subjects who complied to the initial treatment assignment. We find little evidence of differences.

Table A.4: Randomization Compromises, CARE

Case	Initial Assignment	Compromise Description	Age at Departure	Data Availability	Methodology Assumption
1	Family education	Death (age 0), unknown causes	0	Baseline	Attrition after dead
2	Center-based Childcare	Left study at age 5	5	Baseline; before age 5	Attrition after age 5
3	Control	Move at 11 months old	11 months	Baseline; before 11 months	Attrition after 11 months
4	Center-based Childcare	Move at 5 months	5 months	Baseline; before 5 months	Attrition after 5 months
5	Center-based Childcare	Move at age 5	5	Baseline; before age 5	Attrition after 5

Note: This table describes the various randomization compromises in CARE. For each subjects, we display: the nature of the compromise, the available data, and the methodological assumption when accounting for non-compliance and program attrition. The case numbers do not have anything to do with individual identifiers of program participants.

19

A.4 Chronology of Attrition

Over time, various subjects in both ABC and CARE left the studies for different reasons. We describe the chronology of this attrition. In Appendix C, we describe how we adjust for this attrition in the methodology.

A.4.1 ABC

Before randomization, there was an initial pool of 123 subjects from 120 families. Four subjects left the study even before baseline data could be collected. In addition, 3 subjects died before 6 months of age and 2 subjects were diagnosed as mentally disabled making them ineligible for the study. The sample at the beginning of the study is thus 114 subjects (56 in the control group, 58 in the treatment group).

Before age 5, an additional subject died. Public officials requested that 2 control-group subjects be moved to the treatment group due to extremely impoverished home conditions. Three treatment-group subjects and 1 control-group subject did not comply with the randomization. The researchers at FPG continued collecting data on these 3 cases of non-compliance until age 8 when the researchers decided to halt data collection on these non-compliant cases. These cases were also described in Appendix A.3 due to the fact that they are randomization compromises.

After age 5, 1 subject moved away from the study area and 7 subjects attrited from the study. At age 21, 99 subjects participated in the data collection (51 in the control group, 48 in the treatment group).

Between age 21 and 30, 2 subjects died and 4 of the 7 attrited subjects were recovered leading to an age-30 sample of 101 subjects (52 in the control group, 49 in the treatment group).

Many subjects were not included in the age-34 data collection because they had to travel to

20

go to a clinic rather than having at-home visitors, which is how the previous data collections were done. Because of this inconvenience to the subjects, data are only available on 72 subjects (34 in the control group, 38 in the treatment group).

A.4.2 CARE

We describe the chronology of attrition for the two experimental groups that we use in the data (the control group and the group that received center-based childcare and family education). Before randomization, there were 40 subjects. Four moved away from the study area. Between the start of the program and age 5, there was no attrition. These cases are also described in Appendix A.3 due to the fact that they are randomization compromises.

Four subjects attrited after age 5 and were not recovered at later data collections. In the age-34 medical sweep, data were collected on 22 subjects (11 in the control group, 11 in the treatment group).

A.5 Program Description and Content

A.5.1 Goals

The original goals of treatment were to prevent mental retardation by enhancing overall development from birth, in turn fostering school-readiness for an at-risk population.[13] Additional curriculum goals were to (i) support language, motor, and cognitive development; (ii) minimize high-risk behaviors; and (iii) develop socio-emotional competencies considered crucial for school success including task-orientation, communicative competence, independence, and prosocial behavior.[14] Implementation of ABC's and CARE's educational treatments

---

[13]Note that the clinical understanding of mental retardation was once associated with disadvantages that hindered early-life development (Noll and Trent, 2004).

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

evolved each successive year as program staff evaluated ongoing outcome data.[15]

A.5.2 Daily Schedule

For both ABC and CARE, FPGC was open to families from 7:45 a.m. to 5:30 p.m., 5 days per week and 50 weeks per year.[16] Subjects were offered free transportation to and from the center. A driver and second adult staffed each vehicle (one van and two station wagons) equipped with child safety seats.[17] Approximately 65% of treated ABC families utilized the free transportation.[18] Vehicles typically arrived by 9:00 a.m. to the center and departed around 3:45 p.m.[19] At FPGC, ABC and CARE treatment-group subjects received breakfast, lunch, and a snack planned by a nutritionist.[20] Meals were catered by off-site kitchens. Infants received iron-fortified formula until doctors advised adding solid food. The control-group subjects also received an unlimited amount of iron-fortified formula until approximately 15 months of age.[21]

A.5.3 Program Staff and Physical Space

To promote trust in FPGC within the subjects' families, staff were recruited from the local community.[22] Infant and toddler caregivers and preschool teachers demonstrated varied educational backgrounds ranging from high school graduation to master's degrees. Their average professional working experience with young children was 7 years.[23] All classroom staff participated in extensive training and were closely observed by FPGC's academic staff, as part of a broad variety of ongoing clinical and social research related to early childhood

---

[15]Ramey et al. (1975); Finkelstein (1982); McGinness (1982); Haskins (1985).

[16]Ramey et al. (1976, 1985).

[17]Ramey and Campbell (1979); Kuperman (2015).

[18]Barnett and Masse (2002).

[19]Ramey et al. (1977).

[20]Haskins (1985); Bryant et al. (1987); Ramey et al. (1977).

[21]Campbell et al. (2014); Kuperman (2015).

[22]Ramey et al. (1977); Bryant et al. (1987); Feagans (1996); Kuperman (2015).

[23]Ramey et al. (1982, 1985); Wasik et al. (1990).

education, psychology, and health. In ABC, child-caregiver ratios varied by age: 3:1 for infants up to 13 to 15 months of age; 4:1 for toddlers up to 36 months; and 5:1 or 6:1 for children aged 3 to 5 years, depending on cohort size.[24] Child-caregiver ratios were similar in CARE.[25]

The ABC and CARE staff included a program director, a secretary, 12 to 14 teachers and assistant teachers, 3 administrative staff members, and a transportation supervisor.[26] Lead caregivers and teachers had bachelor's or master's degrees. Teacher aides, recruited from the local community, held high school diplomas (at minimum) and were comparatively well-compensated in the childcare field. They remained a stable treatment component throughout the study. After 1980, following revisions to FIDCR regarding minimum requirements for early childhood education staff, several teacher aides pursued and received undergraduate degrees and became lead teachers. All classroom staff were supervised daily, received weekly mentoring, and professional development from outside consultants..[27]

Infant nurseries, toddler rooms, and preschool classrooms were housed on different floors of FPGC. Early reports indicate that FPGC allocated two floors to ABC, but later reports indicate the use of three floors.[28] Two infant nurseries were staffed by five adults in a suite of four adjoining rooms: two sleeping rooms contained seven cribs each, while the other two rooms were designated for activities.[29] The four rooms opened into a large, shared space with feeding tables, an area for food preparation, and a couch.[30] Offices for the medical staff, along with two examining rooms and facilities for laboratory tests were located around the corner from the infant nurseries.[31] Two multi-age toddler rooms were located one floor

---

[24]Ramey et al. (1977); Ramey and Campbell (1979); Ramey et al. (1982).
[25]Burchinal et al. (1997); Ramey et al. (1985).
[26]Ramey et al. (1977, 1982); Bryant et al. (1987).
[27]O'Brien and Sanders (1974); Ramey et al. (1985); Sanders and Stokes (1979); Klein and Sanders (1982); Kuperman (2015).
[28]Ramey and Smith (1977); Ramey and Campbell (1979); Ramey and Haskins (1981).
[29]Ramey et al. (1977).
[30]Ramey and Campbell (1979).
[31]Kuperman (2015).

below the infant nurseries. One room served children who were 1 to 2 years old and the other served children 2 to 3 years old.[32] 3-year-olds were housed in a closed classroom near the toddler rooms. On the lowest floor, 4-year-olds shared an open classroom with a public kindergarten program; the two classes were separated by a long, low bookcase. In CARE, two floors of FPGC were allocated to nurseries and classrooms. A mixed-age classroom design was implemented combining children ages 1 and 3, and children ages 2 and 4. Teacher-child ratios for these ages remained 1:5. FPGC offered two outdoor play areas for both ABC and CARE: one for children up to age 3, and the other for older children.[33]

A.5.4 Approach to Child Development

Curriculum delivery enabled a highly customized learning experience for treated subjects in both ABC and CARE. Infant caregivers recorded child observations on progress charts and collaborated with FPGC's curriculum developers and academic researchers to rotate learning activities every 2 to 3 weeks for each treated subject.[34] Preschool rooms featured intentionally organized environments to promote pre-literacy and access to a rich set of learning tools. The full-day curriculum emphasized active learning experiences, dramatic play, and pre-academics. Frequent 1:1 or 2:1 child-adult interactions prioritized language development for social competence. For ages 3 through 5, as the cohorts approached public school entry, classroom experiences were increasingly structured towards the development of pre-academic skills and "socio-linguistic and communicative competence."[35] FPGC offered a summer program before the start of kindergarten designed to target specific skills to ensure success in a kindergarten classroom (e.g., lining up when exiting the classroom). This program was available to subjects in both the center-based childcare and family education

---

[32]Ramey and Smith (1977); Ramey and Campbell (1979).

[33]Ramey and Campbell (1979); Ramey et al. (1982).

[34]Ramey et al. (1976); Campbell and Ramey (1994).

[35]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).

group and the family education group.[36]

ABC's and CARE's learning programs were influenced by the writings of major developmental theorists.[37] All four ABC cohorts and two CARE cohorts participated in curriculum developers Sparling and Lewis' "LearningGames for the First Three Years."[38] The "LearningGames" were implemented daily by infant and toddler caregivers in 1:1 child-adult interactions. Each "LearningGames" activity stated a developmentally-appropriate objective, the necessary materials, directions for teacher behavior, and expected child outcome. The activities were designed for use both indoors and outdoors, while dressing, eating, bathing, or during play.[39]

Supplemental curricula for preschool rooms varied throughout the study, and included "Cook and Learn," "Peabody Early Experiences Kit," "GOAL Math Program," and "My Friends and Me."[40]

CARE subjects randomized into the center-based childcare and family education group or the family education group also received home visits designed to transmit information on child development and skills involved with parenting including strategies for parent-child interactions based on "LearningGames" activities and problem-solving techniques.[41] Home visitors were trained to ensure they were able to form a strong relationship with the parent and successfully implement the curriculum.[42] The visits lasted about an hour, and occurred weekly until the child was 3 years old. After age 3, the home visits were less frequent and depended on the preferences of the parents. They were usually about once a month after

---

[36]Ramey et al. (1985).

[37]These include including Bowlby, Piaget, and Vygotsky. (Sparling, 1974; McGinness and Ramey, 1981; Kuperman, 2015).

[38]Sparling and Lewis (1979).

[39]Ramey and Campbell (1979); Ramey and Haskins (1981); Sparling and Lewis (1979).

[40]Greenberg and Epstein (1973); Karnes (1973); Dunn et al. (1976); Davis (1977); Wallach and Wallach (1976).

[41]Bryant et al. (1987); Wasik et al. (1990); Burchinal et al. (1997).

[42]Bryant et al. (1987).

age 3.[43]

A.5.5 Medical Care and Nutrition

ABC and CARE provided comprehensive on-site medical care because it was conducted in conjunction with a longitudinal medical research study on infectious respiratory diseases in group environments.[44] Treatment group children were monitored daily for signs of illness. All treated children received medical care while attending center-based childcare; the first ABC cohort of control-group children also received medical care during the program's first year of implementation.[45,46]

In ABC, primary pediatric care was provided by a family nurse practitioner and a licensed practical nurse, both under the supervision of one pediatrician who was on continuous duty at the center.[47] In CARE, the medical staff included two pediatricians, a family nurse practitioner, and a licensed practical nurse.[48] The medical staff provided regularly scheduled check-ups, immunizations, parental counseling, and initial assessment of illnesses.[49] The treatment group received standard check-ups when they were 2, 4, 6, 9, 12, 18, and 24 months old and annually thereafter. While in treatment, they also received the standard immunizations.[50] In ABC, a licensed practical nurse visited classrooms for up to two hours on a daily basis to monitor the subjects' health status.[51] Although this medical care was offered to the treatment-group families free of charge, it was the policy of the medical staff to refer

---

[43]Bryant et al. (1987); Wasik et al. (1990); Burchinal et al. (1997).

[44]Henderson et al. (1982).

[45]Ramey et al. (1976); Bryant et al. (1987); Ramey and Campbell (1991); Campbell and Ramey (1994).

[46]Subjects in both the treatment and control groups of the first cohort received free medical care provided by ABC. The control group of the first cohort only received medical care in the first year of the program; the treatment group of the first cohort received medical care for all years of the program. In the subsequent cohorts, only subjects in the treatment group received free medical care provided by ABC. Both CARE cohorts of treated subjects received medical care.

[47]Haskins et al. (1978).

[48]Bryant et al. (1987).

[49]Ramey et al. (1977); Bryant et al. (1987).

[50]Bryant et al. (1987); Campbell et al. (2014).

[51]Sanyal et al. (1980).

families to a community hospital for serious treatment. While ABC and CARE provided aspirin, immunizations, and basic medicines, families were responsible for purchasing any prescription medication subjects required. There are no data currently available on treatment received for serious conditions or use of prescription medication.

Infants were supplied with iron-fortified formula. Children older than 15 months of age were provided breakfast, lunch, and an afternoon snack all planned by a nutritionist.[52] Control families received diapers for up to three years and unlimited iron-fortified bottled formula through 15 months.[53]

A.5.6 School-age Treatment

The ABC subjects were randomized into a second-phase, school-age treatment (95 subjects continued to this stage of treatment). The CARE subjects in the center-based childcare and family education group and the family education group received the school-age treatment without randomization. The school-age treatment lasted for the first three years of elementary school and consisted of home visits conducted by a Home/School Resource Teacher.[54] These visits were structured to increase exposure to reading and mathematics and promote parental involvement in the academic process.

The curriculum was delivered through sets of activities that developed skills such as handwriting, phonics, and math facts.[55] Teachers worked to encourage parental involvement in the subjects' academics and provided incentives to families to comply with the treatment, such as giving gift certificates to restaurants and books for the subjects upon the completion of activity packets.

Teachers had graduate-level education, training in special education, *or* were qualified to act

---

[52]Bryant et al. (1987); Campbell et al. (2014); Kuperman (2015).
[53]Ramey et al. (1976); Ramey and Campbell (1979); Ramey et al. (1985).
[54]Burchinal et al. (1997).
[55]There were about 60 activities per year. See Campbell and Ramey (1989) for details.

as consultants for in-school teachers to address any problems that arose.[56] They met with parents at home and with teachers in the schools to deliver new activities for the parents to complete with their children and discuss the child's level of success with the previous set of activities. In addition, they helped parents with issues such as adult literacy, housing, and medical care. Thus, the teacher had a dual role as a parent educator and an advocate for the subject in their educational institution.

A.6   Control Group Substitution

In ABC, the families of 75% of the control-group subjects enrolled their children in alternative center-based childcare. In CARE, 74% of families in the control group and 62% of families in the family education group enrolled their children in alternative center-based childcare. We refer to this phenomenon as control substitution; accounting for it is fundamental when evaluating the program. In this Appendix, we describe the characteristics and costs of the childcare centers providing alternative treatment, in order to create a comparison with the treatments offered by ABC and CARE.

Most of the families in the ABC and CARE control groups enrolled their children in alternative preschool that received federal subsidies and, therefore, were regulated. Figure A.5 and Figure A.6 show the amount of enrollment into subsidized and non-subsidized care for ABC and CARE, respectively. Subsidized centers were required to have trained staff who were able to implement curricula designed to enhance cognitive, social, and linguistic competence in disadvantaged children.[57] Thus, we consider these centers to offer high-quality center-based childcare.

---

[56] Ramey and Campbell (1991).
[57] Burchinal et al. (1989).

Figure A.5: Average Number of Months in Alternative Preschool, ABC Control Group
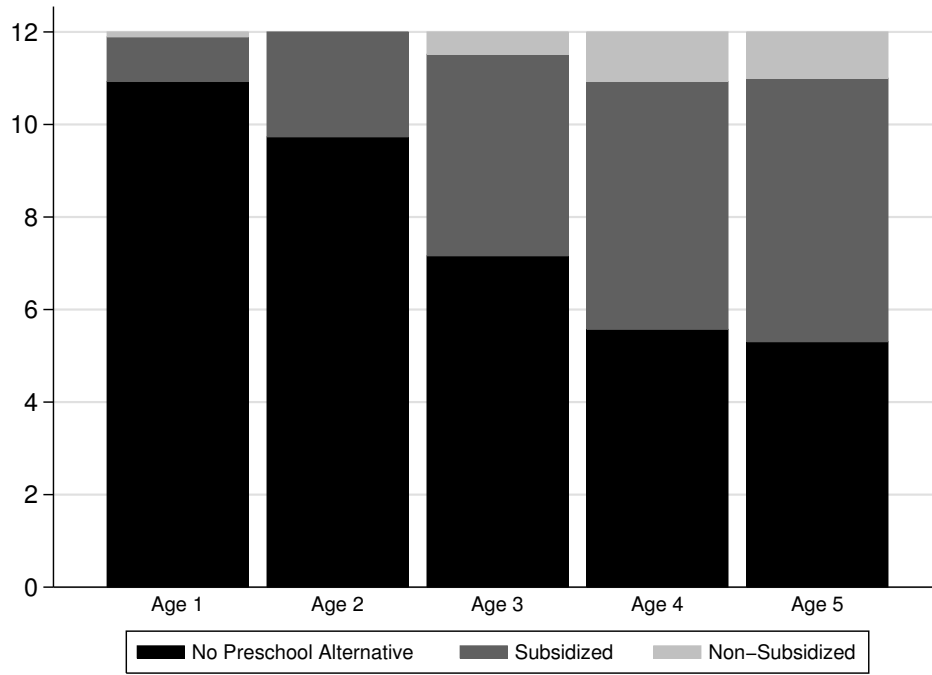
Note: This figure describes the take-up of alternative preschool by families in the ABC control group. The vertical axis represents the average number of months per year the subjects of the control group spent in alternative preschool. Subsidized centers were highly regulated and, therefore, relatively high-quality. Non-subsidized childcare services were center-based but not regulated. Other sources of childcare could have included care by parents, relatives, or non-relatives.

Figure A.6: Average Number of Months in Alternative Preschool, CARE Control and Family
Education Groups
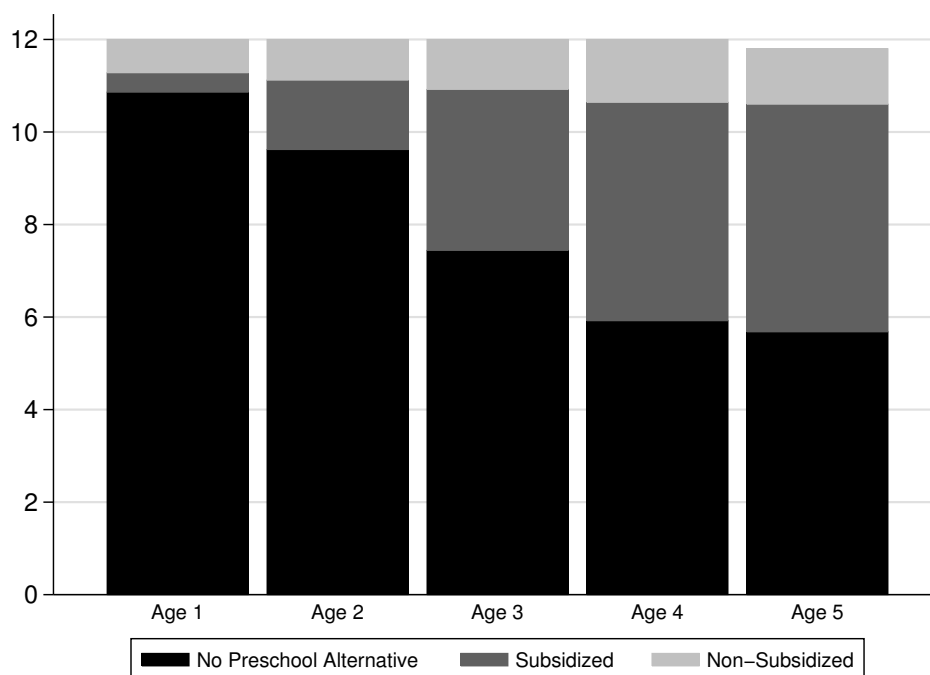
Note: This figure describes the take-up of alternative preschool by families in the CARE family education
and control groups. The vertical axis represents the average number of months per year the subjects of
the control group spent in alternative preschool. Subsidized centers were highly regulated and, therefore,
relatively high-quality. Non-subsidized childcare services were center-based but not regulated. Other sources
of childcare could have included care by parents, relatives, or non-relatives.

Table A.5 shows baseline characteristics between the control-group subjects who were en-
rolled in alternative preschool and those who stayed at home. The control-group children
who attended alternative preschool were marginally more advantaged, with the most stark
difference being maternal employment. This is seen across genders, but is only significant
for the female and pooled samples. The males who are enrolled in alternative preschool have
mothers with higher IQ scores, but lower parental income indicating lack of spousal support,
which is evident by the fewer number of fathers present in that same group. Those who were
enrolled in alternative preschools also had more siblings.

Figure A.7a shows enrollment by age and the average months of enrollment by age for the

control-group children who enrolled in program alternatives. Enrollment increases with the age of children. Figure A.7b shows the fraction of children enrolled in preschool by age. As control children age, they are more likely to enter childcare.

Figure A.7: Control Substitution Characteristics, ABC/CARE Control Group

(a) Enrollment by Age

(b) Enrollment Dynamics
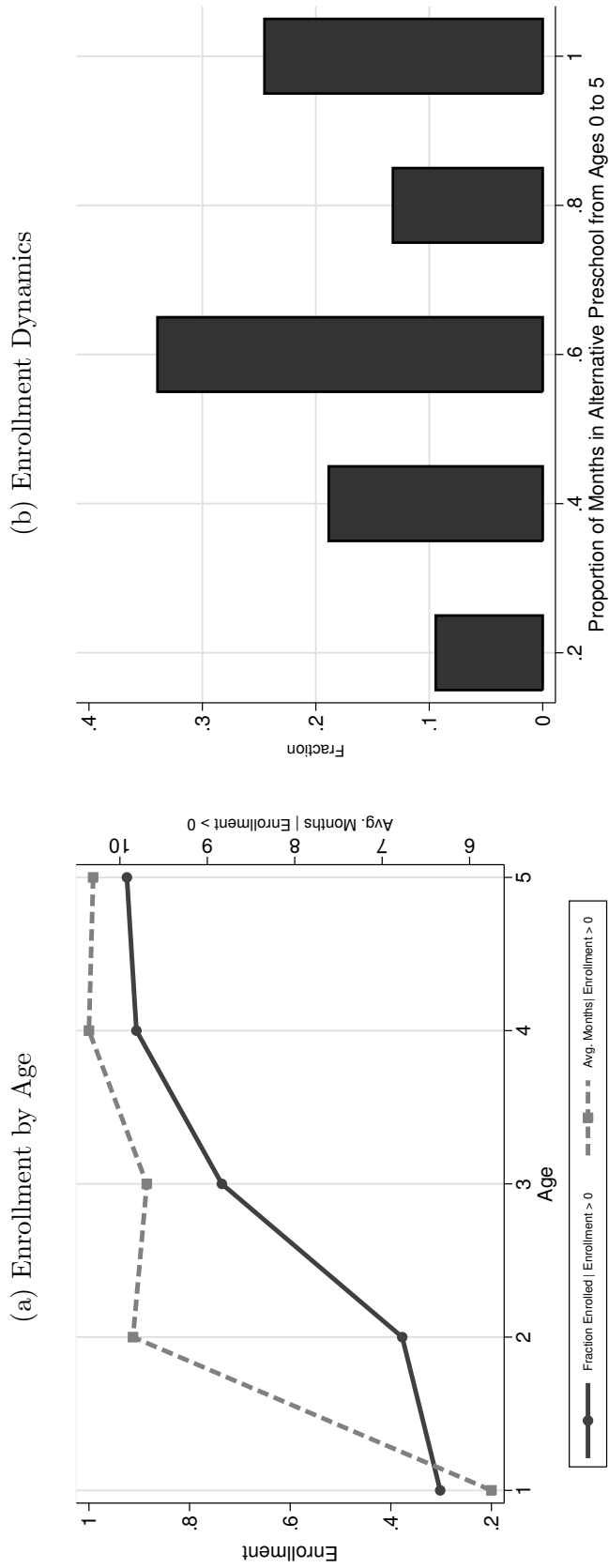
Note: Panel (a) displays the fraction of the ABC/CARE control group enrolled in alternatives by age on the left axis and average number of months in alternative preschool by age in the right axis. Panel (b) displays the fraction of children in the ABC/CARE control groups enrolled in alternatives 20%, 40%, 60%, 80%, and 100% of the time, from ages 0 to 5.

Table A.5: Baseline Characteristics and Control Group Substitution

Characteristic	Females			Males			Pooled		
	Control Substitution		p-value	Control Substitution		p-value	Control Substitution		p-value
	No	Yes		No	Yes		No	Yes	
	N = 10	N = 27		N = 9	N = 28		N = 19	N = 55	
Mother's Yrs. of Edu.	9.70	10.19	0.53	9.78	10.50	0.32	9.74	10.35	0.23
	(0.63)	(0.42)		(0.62)	(0.31)		(0.43)	(0.26)	
Mother Works	**0.00**	**0.23**	0.02	0.14	0.29	0.42	**0.08**	**0.26**	0.09
	(0.00)	(0.09)		(0.14)	(0.11)		(0.08)	(0.07)	
Mother's Age	19.40	19.89	0.80	23.67	20.64	0.39	21.42	20.27	0.55
	(1.66)	(0.91)		(3.25)	(0.89)		(1.79)	(0.63)	
Mother's IQ	81.70	84.04	0.54	82.33	87.11	0.26	82.00	85.60	0.19
	(3.15)	(1.96)		(3.62)	(1.80)		(2.32)	(1.33)	
Father Present	0.30	0.26	0.82	0.44	0.25	0.34	0.37	0.25	0.38
	(0.15)	(0.09)		(0.18)	(0.08)		(0.11)	(0.06)	
Parental Income	2,566.67	3,499.55	0.75	11,291.43	8,694.41	0.63	7,264.62	5,763.97	0.65
	(2,566.67)	(1,264.46)		(4,750.08)	(2,220.99)		(2,986.31)	(1,256.34)	
HRI Score	21.90	22.93	0.64	19.89	20.32	0.87	20.95	21.60	0.70
	(1.73)	(1.25)		(2.46)	(1.00)		(1.46)	(0.81)	
Number of Siblings	0.50	0.70	0.60	1.56	0.54	0.19	1.00	0.62	0.35
	(0.31)	(0.22)		(0.71)	(0.14)		(0.38)	(0.13)	
Male							0.47	0.51	0.80
							(0.12)	(0.07)	
Birth Year	1975.50	1975.37	0.91	1975.67	1976.07	0.71	1975.58	1975.73	0.84
	(0.99)	(0.42)		(0.97)	(0.48)		(0.68)	(0.32)	
Apgar Score, 1 min.	7.30	7.46	0.84	7.67	7.78	0.83	7.47	7.62	0.76
	(0.73)	(0.33)		(0.44)	(0.27)		(0.43)	(0.21)	
Apgar Score, 5 min.	8.40	9.04	0.21	8.89	8.92	0.91	8.63	8.98	0.24
	(0.45)	(0.17)		(0.20)	(0.21)		(0.26)	(0.14)	

Note: This table describes baseline characteristics for the children in the control group, by gender and by their alternative preschool enrollment status. The number of subjects in these groups are listed at the top of the table. Asymptotic standard errors are in parentheses. The reported p-values are from two-sided tests of difference of means. The means are bolded if the difference is significant at the 10% level.

A.6.1 Quality of Alternative Preschools

During the period when both ABC and CARE were active, North Carolina had an active, high-quality system of public childcare for vulnerable families funded by several public programs. Examples include Title IV-A of the Social Service Administration (SSA), Aid to Families with Dependent Children (AFDC), and Title IV-B of Child Welfare Services. These funding efforts were amplified in 1975 by Title XX of the SSA, Social Services Block Grant, which was the main federal source of childcare financing in the U.S. when ABC and CARE were active.[58]

Federally funded childcare services were regulated according to FIDCR standards, which defined stringent regulation for center-based programs for children between the ages of 3 and 6.[59] These requirements were enforced.[60] Additionally, North Carolina had a mandatory licensing law for childcare facilities. While FIDCR applied to centers for older children (between the ages of 3 and 6), the North Carolina regulation only applied to centers serving children below the age of 3. The relative weakness of this regulation is not very relevant for our study because treatment substitution occurred mostly after age 3 (see Figure A.5 and Figure A.6).[61] Table A.6 compares a widely-used quality standard, the child-staff ratio, between the North Carolina and FIDCR standards and the actual ABC and CARE numbers.

---

[58]Robins (1988).
[59]Department of Health, Education, and Welfare (1968).
[60]Kuperman and Hojman (2015b).
[61]North Carolina General Assembly (1971).

Table A.6: Child-Staff Ratios for North Carolina, FIDCR, and Actual ABC and CARE Ratios

Age	NC Standards Level I	FIDCR Standards	ABC and CARE Ratios
0–1	6:1*		3:1
1–2	8:1*		4-5:1
2–3	12:1*		4-5:1
3–4	15:1	5:1*	4-5:1
4–5	20:1	7:1*	5-6:1
5–6	25:1	7:1*	5-6:1

Sources: 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).

Note: The starred ratios represent the ones we believe were the most relevant for the ABC control-group subjects and the CARE control-group and family-education-group subjects.

A.6.2 Costs

Previous papers have used childcare cost rates that are not specific to North Carolina and do not account for the contemporaneous structure of the subsidies (Barnett and Masse, 2002, 2007). We use the local subsidy rates that were in place when the ABC subjects were in preschool to impute different costs of the alternative preschools. These costs depend on the specific preschool attended and the eligibility of the families to receive the subsidies.

When ABC and CARE were in operation, center-based childcare was subsidized by several federal programs (the Department of Social Services categorized these programs as Child Welfare, AFDC, and Work Incentive Programs).[62] However, our calculations of the cost of alternative preschool are simplified by the fact that the subsidies were centralized and regulated by the County Department of Social Services. Those departments used a uniform subsidy rate, regardless of the origin of the funds.[63] We collected information about the

---

[62]North Carolina State Department of Social Services (1972).

[63]Ad Hoc Committee of Professionals in Child Care Services, North Carolina (1974).

subsidy rate at the time, which approximates the price of the centers, as centers pegged their fees and services to the maximum subsidy rate. Moreover, we know which centers each ABC control subject attended. We interviewed North Carolina childcare staff and academics that study childcare to document which of those centers were subsidized and regulated at the time.[64] For subsidized centers, we impute the maximum Department of Social Services fee established at the time: $633/month in 2014 USD.[65] For non-subsidized centers, we impute the mean of costs for Level-1 centers (minimum accepted quality level) according to a 1982 North Carolina study of the cost of childcare: $298/month in 2014 USD.[66] Although the information in this survey is not ideal for assessing the cost of subsidized preschools for CARE, as the subsidies greatly changed after the end of FIDCR (1981), it provides an approximation for assessing the cost of the non-subsidized centers.

Finally, we determine if the families paid the costs themselves or if they were subsidized, in which case we also add deadweight costs. We consider if a subject was eligible for subsidies if the family lived in poverty according to the federal guidelines and all parents living at home worked. If a family is deemed eligible, then we assume the child's preschool was fully subsidized using the rates described above without additional subsidies.

A.7 Data

In Table A.7 through Table A.12, we summarize the data availability for both ABC and CARE. The data collection processes in both programs were designed to be similar. For both programs, the treatment and control groups were followed into adulthood with relatively low attrition. For ABC, subjects were followed annually through elementary school and at ages 12, 15, 21, and 30. Health and administrative crime data were collected when the subjects

---

[64]Kuperman and Hojman (2015b,a).

[65]Ad Hoc Committee of Professionals in Child Care Services, North Carolina (1974); Community Planning Services (1973).

[66]Administrative Branch, Office of Day Care Services (1982).

reached their mid-30s. For CARE, the exact same follow-ups are available with the exception of the age 15 follow-up.

Table A.7: Early Childhood Data (Part I)

Category	Sub-Category	Description	ABC Age (months)	CARE Age (months)	Measure
Demographics	Gender	Gender of subject	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54	Demographic Interview
	Race	Race/Cultural identity of subject	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54	Demographic Interview
	Birth Date	Date of birth of subject	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54	Demographic Interview
Cognitive Assessments	Language Ability	Auditory association, Verbal expression, etc.	36, 42, 48, 54	30, 42, 54	ITPA^{ABC}, GPB^{ABC}, PLP^{ABC}, MSCD
	Intelligence Levels	SBIS / WPPSI / BSID / UOSPD / RPM	24, 36, 48, 60 / 3, 6, 9, 12, 18, 24 / 15 / 60	24, 36, 48, 60 / 6, 12, 18, 24 / - / -	SBIS / WPPSI / BSID / UOSPD^{ABC} / RPM^{ABC}
	Quantitative	BSID / MSCD	3, 6, 9, 12, 18, 24 / 30, 42, 54	6, 12, 18, 24 / 30, 42, 54	BSID / MSCD
	Memory	BSID / MSCD	3, 6, 9, 12, 18, 24 / 30, 42, 54	6, 12, 18, 24 / 30, 42, 54	BSID / MSCD
	Motor Development	BSID / MSCD	3, 6, 9, 12, 18, 24 / 30, 42, 54	6, 12, 18, 24 / 30, 42, 54	BSID / MSCD
	Critical Thinking	Curiosity	30, 36, 42, 48, 54, 60, 66, 72	-	Infant Behavior Inventory^{ABC}
Non-Cognitive Assessments	Social Skills	Positive social response	30, 36, 42, 48, 54, 60, 66, 72	6, 12, 18, 24	Infant Behavior Inventory^{ABC}, Bayley Infant Inventory^{CARE}
		Creativity	30, 36, 42, 48, 54, 60, 66, 72	-	Infant Behavior Inventory^{ABC}
	Self-Control	Locus of control	3, 18	6, 18	RIES
		Distractibility, Attentiveness	30, 36, 42, 48, 54, 60, 66, 72	6, 12, 18, 24	Infant Behavior Inventory^{ABC}, Bayley Infant Inventory^{CARE}
	Emotional Health	KRT	24, 36, 48, 60	24, 30, 36, 42, 48, 60	KRT
	Self-Consciousness	Self-consciousness	30, 36, 42, 48, 54, 60, 66, 72	-	Infant Behavior Inventory^{ABC}

Sources: Authors' description.
Note: This table describes the main categories of variables that were measured for ABC and CARE subjects up to age 6. ABC and CARE ages are measured in months. This is not an exhaustive list of variables, nor does it include variables from auxiliary data. Instruments or questionnaires available for only one of the studies are in-dicated with the superscript ABC or $CARE$. **Abbreviations are as follows.** ITPA: Illinois Test of Psycholinguistic Ability. GPB: Gordon Psycholinguistic Battery. PLP: Preschool Language Performance. MSCD: McCarthy Scales of Children's Development. BSID: Bayley Scales of Infant Development and Infant Behavior. UOSPD: Uzgiris-Hunt Ordinal Scales of Psychological Development. RPM: Raven's Progressive Matrices. RIES: Rotter's Internality-Externality Scale. KRT: Kohn and Rosman Test Behavior Inventory.

Table A.8: Early Childhood Data (Part II)

Category	Sub-Category	Description	ABC Age (months)	CARE Age (months)	Measure
Family Environment	Family Members	Number of primary caretakers	Birth, 18, 30, 42, 54	18, 30, 42, 54, 60	Demographic Interview
		Relationship with family members, including father, mother, siblings, etc.	Birth, 18, 30, 42, 54	18, 30, 42, 54, 60	Demographic Interview
		Number of siblings	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54, 60	Demographic Interview
		Marital status of parents	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54, 60	Demographic Interview
		Marital conflicts between parents	6, 18	Birth, 6, 18, 36	Demographic Interview^{CARE}, Parental Attitudes Research Inventory
		Father at home	18, 30, 42, 54	18, 30, 42, 54, 60	Demographic Interview
	Family Economic Environment	Parents' occupation	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54, 60	Demographic Interview
		Mother works	18, 30, 42, 54	18, 30, 42, 54, 60	Demographic Interview
		Source of child support	Birth, 18, 30, 42, 54	18, 30, 42, 54, 60	Demographic Interview
		Family income	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54, 60	Demographic Interview
	Parents and Home Environment	Parents' authority, warmth, family conflict, etc.	6, 18, 30, 42, 54	6, 12, 18, 30, 42, 54	Parent Interview
	Family Social Status	Parents' education background	Birth, 18, 30, 42, 54	Birth, 18, 30, 42, 54, 60	Demographic Interview
		Risk taking of family members	Birth	-	Parent Interview^{ABC}
	Family Members' Physical Health	Health issues of parents	Birth	Birth	Parent Interview
		Pregnancy history	Birth	Birth	Parent Interview
Childcare	Day-care Experience	Time and location of childcare, Age when begin	Birth, 18, 30, 42, 54	18, 30, 42, 54	Demographic Interview
		Home visits	-	6, 18, 30, 42, 54, 60	Home Visit Data^{CARE}
	Parental Care	Maternal warmth, Maternal involvement with child	6, 18, 30, 42, 54	6, 12, 18, 30, 42, 54	Home Stimulation
		Provision of appropriate play materials	6, 18, 30, 42, 54	6, 12, 18, 30, 42, 54	Home Stimulation
		Avoidance of restriction and punishment	6, 18, 30, 42, 54	6, 12, 18, 30	Home Stimulation
		Authoritarian control	6, 18, 30, 42, 54	6, 12, 18, 30, 36, 42, 102	Home Stimulation, Parental Attitudes Research Inventory
		Democratic attitudes	6, 18	6, 18, 36	Parental Attitudes Research Inventory
		Hostility and rejection	6, 18	6, 18, 36	Parental Attitudes Research Inventory
		Parents' knowledge of childcare	Birth	-	Parent Interview^{ABC}
Physical Health	Growth Data	Height, Weight, Head circumference, etc.	3, 6, 9, 12, 18, 24, 36, 48, 60	Birth, 6, 12, 18, 24, 36, 48, 60	Growth Measures

Sources: Authors' description.
Note: This table describes the main categories of variables that were measured for ABC and CARE subjects up to age 6. ABC and CARE ages are measured in months. This is not an exhaustive list of variables, nor does it include variables from auxiliary data. Instruments or questionnaires available for only one of the studies are indicated with the superscript ABC or $CARE$.

Table A.9: Childhood and Adolescence Data (Part I)

Category	Sub-Category	Description	ABC Age (years)	CARE Age (years)	Measure
Cognitive Assessment	Language Ability	Adaptive Language Inventory Language Questionnaire MSCD	6, 7, 8 12 7	6, 7, 8 - -	Adaptive Language Inventory Language Questionnaire^{ABC} MSCD^{ABC}
	Intelligence Tests	SBIS WIS Kaufman^{CARE}	6 6, 7, 8, 12, 15 -	7 6, 8 6	SBIS WIS Kaufman^{CARE}
	Quantitative Skills	MSCD^{ABC}	7	-	MSCD^{ABC}
	Memory	MSCD^{ABC}	7	-	MSCD^{ABC}
	Motor Skills	MSCD^{ABC}	7	-	MSCD^{ABC}
Non-Cognitive Assessment	Interpersonal Skills	Gets along with people Relationship with the other sex	6, 8, 12, 15 15	8, 12 -	PEI, CAS, PMI^{ABC}, SAI^{ABC}, Subject Interview^{ABC}, Quality Rank^{CARE} SAI^{ABC}, Subject What I Am Like (Harter)^{ABC}
	Critical Thinking	Thinks for self, questions things Concept Attainment Kit	6, 8 6, 7, 8	8, 12 -	PEI, Harter Child^{CARE}, CBI Concept Attainment Kit^{ABC}
	Self-Control	Distracted in class Locus of control	6, 7, 8, 12, 15 15	12 -	SCAN^{ABC}, CBI, WPB^{ABC}, PMI^{ABC}, SAI^{ABC}, Self-Evaluation Inventory^{ABC} Nowicki-Strickland Data, Pearlin Mastery Scale^{ABC}
	Work Ethic	Task Orientation	6, 7, 8, 12, 15	6, 7, 8, 9, 12	SCAN^{ABC}, CBI, PMI^{ABC}
	Emotional Health	Harms self, suicidal thoughts Depression, anxiety, fear, etc.	8, 12, 15 6, 7, 8, 12, 15	8, 12 7, 8, 9, 12	Achenbach Parent, Subject Risk Taking Survey^{ABC} KRT, CAS, ETS, Achenbach Parent
	Social Activities	Athletic activities Participant of organizations, e.g. religions Reading list TV/music	8, 12, 15 8, 12, 15 12, 15 12, 15	8, 12 8, 12 12 12	Achenbach Parent, SAI^{ABC}, Subject What I Am Like (Harter)^{ABC}, PEI^{CARE} Achenbach Parent, SAI^{ABC}, Subject Interview^{ABC} CAS, SAI^{ABC}
	Self-Consciousness	Self-conscious emotions	8, 12, 15	8, 12	Achenbach Parent, Subject What I Am Like (Harter)

Sources: Authors' descriptions.

Note: This table describes the main categories of variables that were measured for ABC and CARE subjects at ages 6 to 18. ABC and CARE ages are measured in years. This is not an exhaustive list of variables, nor does it include variables from auxiliary data. Instruments or questionnaires available for only one of the studies are indicated with the superscript ABC or $CARE$. **Abbreviations are as follows.** MSCD: McCarthy Scales of Children's Development. SBIS: Stanford-Binet Intelligence Scale. WIS: Wechsler Intelligence Scale for Children. KRT: Kohn and Rosman Test Behavior Inventory. WJCA: Woodcock-Johnson Test of Cognitive Abilities. PEI: Parents as Educator Interview. CAS: Child Assessment Schedule. PMI: Psychosocial Maturity Inventory. SAI: Social Adjustment Inventory for Children and Adolescents. SCAN: Schedule of Classroom Activity Norms. CBI: Classroom Behavior Inventory. WPB: Walker Problem Behavior Checklist. ETS: Emotional/Activity/Sociability/Impulsivity Temperament Survey. FES: Family Environment Scale. PIAT: Peabody Individual Achievement Test. CAT: California Achievement Test. MARS: Mid-Adolescence Rating Scale Data.

40

Table A.10: Childhood and Adolescence Data (Part II)

Category	Sub-Category	Description	ABC Age (years)	CARE Age (years)	Measure
Family Environment	Family Members	Number of adults in house	6, 8, 12, 15	8, 12	PEI, Parent Interview, Subject Person In Household^{ABC}
		Relationship with family members, including father, mother, siblings, etc.	6, 8, 12, 15	8, 12	PEI, FES, SAI, Subject Interview^{ABC}, Adult Self Report^{ABC}, Parent Interview. Achenbach Parent
		Number of siblings	6, 8, 12, 15	7, 8, 12	PEI^{ABC}, Parent Interview
		Marital status of parents	6, 8, 12, 15	7, 8, 12	PEI^{ABC}, Parent Interview
		Father at home	18, 30, 42, 54	18, 30, 42, 54, 60	Demographic Interview
	Parents' Education Style	Role of parents in education	6, 8	8, 12	PEI, Parent Interview^{CARE}
		Parents' education beliefs & methods	6, 8	8, 12	PEI, Parent Interview^{CARE}
		Parents' aspiration & attitudes towards child	6, 8, 12, 15	8, 12	PEI, Parent Interview
	Family Economic Environment	Parents' occupation	6, 8, 12, 15	7, 8, 12	PEI^{ABC}, Parent Interview
		Mother works	9	5, 7, 8	Demographic Interview
		Source of child support	6, 8, 12, 15	7, 8, 12	PEI^{ABC}, Parent Interview
		Family income	6, 8, 12, 15	7, 8, 12	PEI^{ABC}, Parent Interview
	Parents and Home Environment	Parents' authority, warmth, family conflict, etc.	8	8	Parent Interview
	Family Social Status	Parents' education background	6, 8, 12, 15	7, 8, 12	PEI^{ABC}, Parent Interview
		Criminal history and risk taking of family members	8, 12, 15	-	Subject Taylor Life Events^{ABC}, Parent Interview^{ABC}
	Family Members' Physical Health	Health issues of adults in house	8, 12, 15	12	Parent Interview, Subject Taylor Life Events^{ABC}
Academic Achievements	Standardized Tests	Reading, mathematics, and language abilities	6, 7, 8, 12	6, 8, 9,12	CAT^{ABC}, PIAT^{ABC}, WJCA
	Performance in Schoolwork	Drop in grades	12, 15	12	CAS
		Lack of interest in school	12, 15	12	CAS
		Total years in special education	17	11	Retention and Special Services Data
		Total years retained in school	17	11	Retention and Special Services Data
Physical Health	Health Issues	Health issues of subject	8, 12, 15	8, 12	Achenbach Parent, Subject Interview^{ABC}, Adult Self Report^{ABC}, PEI^{CARE}, Parent Interview^{CARE}
	Growth	Vision, weight, height	8	8	Growth Data
	Teenage Pregnancy	Teenage Pregnancy	15	-	Subject Interview^{ABC}
Social Conduct	Law Breaking	Felony, Time spent incarcerated	15	-	MARS^{ABC}, Subject Interview^{ABC}

Sources: Authors' descriptions.
Note: This table describes the major categories of variables that were measured for ABC and CARE subjects at ages 6 to 18. ABC and CARE age are measured in years. This is not an exhaustive list of variables, nor does it include variables from auxiliary data. Instruments or questionnaires available for only one of the studies are indicated with the superscript ABC or $CARE$. **Abbreviations are as follows.** MSCD: McCarthy Scales of Children's Development. SBIS: Stanford-Binet Intelligence Scale. WIS: Wechsler Intelligence Scale for Children. KRT: Kohn and Rosman Test Behavior Inventory. WJCA: Woodcock-Johnson Test of Cognitive Abilities. PEI: Parents as Educator Interview. CAS: Child Assessment Schedule. PMI: Psychosocial Maturity Inventory. SAI: Social Adjustment Inventory for Children and Adolescents. SCAN: Schedule of Classroom Activity Norms. CBI: Classroom Behavior Inventory. WPB: Walker Problem Behavior Checklist. ETS: Emotional/Activity/Sociability/Impulsivity Temperament Survey. FES: Family Environment Scale. PIAT: Peabody Individual Achievement Test. CAT: California Achievement Test. MARS: Mid-Adolescence Rating Scale Data.

Table A.11: Adult Data (Part I)

Category	Sub-Category	Description	ABC Age (years)	CARE Age (years)	Measure
Cognitive Assessments	Intelligence Tests	WIS	21	-	WIS
Non-Cognitive Assessment	Interpersonal Skills	Gets along with people	21, 30	-	Subject Interview
	Self-Control	Locus of control	21, 30	-	Nowicki-Strickland Data^{ABC}, Pearlin Mastery Scale^{ABC}
		Proud of working, interest in working	21, 30	21, 30	Job Satisfaction Survey^{ABC}, Subject Interview
	Emotional Health	Harms self, suicidal thoughts,	21	21	Achenbach, Subject Risk Taking Survey
		depression, anxiety, fear, etc.	21, 30	21, 30	KRT, Achenbach Parent, CAS, Brief Symptom Inventory, ETS
	Social Activities	Athletic activities	21	-	Achenbach,
		Participant of organizations, e.g. religions	21, 30	21, 30	Achenbach, Subject Interview
Family Environment	Family Members	Number of adults in house	21	-	Parent Interview^{ABC}, Subject Interview
		Relationship with family members, including father, mother, siblings, etc.	21, 30	30	Parent Interview, Achenbach^{ABC}, Subject Interview, Adult Self Report
		Number of siblings	21, 30	30	Parent Interview^{ABC}, Subject Interview
		Marital status of parents	21	-	Parent Interview^{ABC}, Subject Interview
		Number of children, childcare basics	21, 30	30	Subject Interview, Childcare Questionnaire
	Family Economic Environment	Parents' occupation	21	-	Parent Interview^{ABC}, Subject Interview
		Source of child support	21	30	Parent Interview^{ABC}, Subject Interview
		Family income	21	30	Parent Interview^{ABC}, Subject Interview
	Family Members and Children	Relationship quality, family health issues, attitude toward child learning	30	30	Parent Interview, Taylor Life Events^{ABC}, Child Health Questionnaire, PEI
	Marital Status	Marital status, spouse income	21, 30	21, 30	Subject Interview
		Spouse details, marriage history	21, 30	30	Subject Interview
		Relationship with spouse	21, 30	30	Subject Interview, Adult Self Report
Achievement	Education Level	Years in school, plans for future education	21, 30	21, 30	Subject Interview, Adult Self Report
		College type, certificate earned	21, 30	21, 30	Subject Interview, Adult Self Report
	Achievement Test	WJCA	21, 30	-	WJCA

Sources: Authors' description.

Note: This table describes the major categories of variables that were measured for ABC and CARE subjects at ages 21 and 30. ABC and CARE age are measured in years. This is not an exhaustive list of variables, nor does it include variables from auxiliary data. Instruments or questionnaires available for only one of the studies are indicated with the superscript ^{ABC} or ^{CARE}. **Abbreviations are as follows.** KRT: Kohn and Rosman Test Behavior Inventory. CAS: Child Assessment Schedule. ETS: Emotional/Activity/Sociability/Impulsivity Temperament Survey. WIS: Wechsler Adult Intelligence Scale. WJCA: Woodcock-Johnson Test of Cognitive Abilities. PEI: Parents as Educator Interview.

Table A.12: Adult Data (Part II)

Category	Sub-Category	Description	ABC Age (years)	CARE Age (years)	Measure
Physical Health	Health Insurance	Covered by health insurance	21, 30	21, 30	Subject Interview
	Health Issues	Health conditions, diseases, regular checkups and tests, mental health	21, 30, mid-30's	21, 30, mid-30's	Brief Symptom Inventory, Subject Interview, Adult Self Report, Biomedical Survey
Social Conduct	Risk Taking	Smoking, drinking, carry gun, fight, drug use	21, 30	21, 30	Subject Risk Taking Survey, Tobacco, Alcohol, and Drug Survey, Adult Self Report
	Law Breaking	Felony, Time spent incarcerated	21	21, 30	Subject Interview
Economic Status	Living Circumstances	Number of rooms	21, 30	21, 30	Subject Interview
		Own or rent apartment	21, 30	21	Subject Interview
		Number living in same domicile	21, 30	21	Subject Interview
	Working Condition	Currently employed	21, 30	21, 30	Subject Interview
		Job title	21, 30	21, 30	Subject Interview, Adult Self Report
		Job category	21, 30	21, 30	Subject Interview, Adult Self Report
		Hours	21, 30	21, 30	Subject Interview, Adult Self Report
		Satisfied with current job	21, 30	21, 30	Subject Interview, Subject What I Am Like (Harter), Adult Self Report
	Transportation	Own reliable transportation	21, 30	21	Subject Interview, Adult Self Report
		Public transportation	21, 30	21	Subject Interview, Adult Self Report
	Income	Income from job	21, 30	21, 30	Subject Interview, Adult Self Report
		Income from welfare programs	21, 30	30	Subject Interview, Adult Self Report
		Income from investment	21, 30	-	Subject Interview, Adult Self Report

Sources: Authors' description.
Note: This table describes the major categories of variables that were measured for ABC and CARE subjects at ages 21, 30, and the mid-30's. ABC and CARE age are measured in years. This is not an exhaustive list of variables, nor does it include variables from auxiliary data. Instruments or questionnaires available for only one of the studies are indicated with the superscript *ABC* or *CARE*.

43

Attrition was low in ABC. Information is available on 100 subjects in the age 30 follow-up, which we call the adult follow-up. In addition, 80 subjects—40 from the control group and 40 from the treatment group—consented to the release of their criminal records. Further, 70 participants consented to the release of information regarding a full-range biomedical panel—31 from the control group and 39 from the treatment group.

Attrition was also low for CARE subjects. Information is available on 58 subjects (more than 85% of the initial sample) in the age-30 follow-up. Additionally, 40 participants (11 from the control group, 18 from the family education group, and 11 from the center-based childcare and family education group) released information on the full-range biomedical sweep. Administrative crime data are not available for CARE. We do not evaluate the second-phase of treatment in CARE because it was not randomized. Rather, those in the center-based childcare and family education group and the family education group were offered school-age treatment, and those in the control group were not.

In the following set of tables (Table A.13 through Table A.21), we compare the observed, baseline characteristics between the first-phase control and treatment groups in ABC, which are the main groups we analyze, at different stages of the data collection follow-ups. For each observed characteristic, we present the bootstrapped p-value associated with the standard t-test. We also present the bootstrapped, step-down p-value on jointly testing the difference in observed characteristics across the two blocks of variables separated by the horizontal line.[67]

First, we compare the first-phase treatment and control groups on baseline characteristics.

---

[67]Lehmann and Romano (2005).

Table A.13: First-phase Treatment vs. Control Groups, ABC

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	Multiple H_0
Male	0	57	59	0.438	0.489	(0.580)	(0.700)
Birth Weight	0	56	58	7.191	6.829	(0.130)	(0.205)
No. Siblings in Household	0	57	59	0.750	0.516	(0.245)	(0.425)
Birth Year	0	57	59	1974	1974	(0.785)	(0.865)
Mother's Education	0	57	59	9.864	10.505	**(0.050)**	(0.105)
Mother's Age	0	57	59	20.103	19.564	(0.555)	(0.695)
Mother Employed	0	57	59	0.216	0.317	(0.190)	(0.370)
Parental Income	0	57	58	6,211	7,019	(0.645)	(0.755)
Mother's IQ	0	57	59	83.419	85.393	(0.360)	(0.555)
Father at Home	0	57	59	0.346	0.223	(0.135)	(0.310)

Note: This table shows the balance in observed characteristics between the treatment and control groups in ABC at baseline. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Second, we present the same exercise for each of the four cohorts ABC served.

Table A.14: First-phase Treatment vs. Control Groups, ABC Cohort 1

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	Multiple H_0
Male	0	14	14	0.348	0.286	(0.730)	(0.738)
Birth Weight	0	14	13	6.755	6.491	(0.550)	(0.655)
No. Siblings in Household	0	14	14	1.741	0.606	**(0.035)**	**(0.085)**
Birth Year	0	14	14	1972	1972	(0.240)	(0.350)
Mother's Education	0	14	14	9.885	10.561	(0.265)	(0.480)
Mother's Age	0	14	14	23.869	19.552	**(0.050)**	(0.135)
Mother Employed	0	14	14	0.152	0.205	(0.695)	(0.895)
Parental Income	0	14	13	7,164	8,298	(0.755)	(0.910)
Mother's IQ	0	14	14	76.042	81.108	(0.270)	(0.485)
Father at Home	0	14	14	0.559	0.368	(0.340)	(0.493)

Note: This table shows the balance in observed characteristics between the treatment and control groups in ABC at baseline for cohort 1. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.15: First-phase Treatment vs. Control Groups, ABC Cohort 2

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	Multiple H_0
Male	0	13	16	0.457	0.503	(0.805)	(0.875)
Birth Weight	0	13	16	7.256	6.534	(0.160)	(0.270)
No. Siblings in Household	0	13	16	0.388	0.316	(0.755)	(0.835)
Birth Year	0	13	16	1973	1973	(0.850)	(0.925)
Mother's Education	0	13	16	10.225	10.307	(0.885)	(0.940)
Mother's Age	0	13	16	18.446	17.637	(0.380)	(0.630)
Mother Employed	0	13	16	0.307	0.248	(0.690)	(0.850)
Parental Income	0	13	16	5,398	4,427	(0.790)	(0.880)
Mother's IQ	0	13	16	86.873	85.597	(0.730)	(0.855)
Father at Home	0	13	16	0.220	0.183	(0.790)	(0.895)

Note: This table shows the balance in observed characteristics between the treatment and control groups in ABC at baseline for cohort 2. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.16: First-phase Treatment vs. Control Groups, ABC Cohort 3

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	Multiple H_0
Male	0	14	15	0.376	0.596	(0.265)	(0.320)
Birth Weight	0	14	15	7.424	7.138	(0.470)	(0.730)
No. Siblings in Household	0	14	15	0.423	0.203	(0.385)	(0.645)
Birth Year	0	14	15	1975	1975	(0.510)	(0.520)
Mother's Education	0	14	15	10.133	10.704	(0.405)	(0.595)
Mother's Age	0	14	15	18.602	19.558	(0.355)	(0.570)
Mother Employed	0	14	15	0.162	0.467	**(0.070)**	(0.155)
Parental Income	0	14	15	7,034	4,981	(0.430)	(0.675)
Mother's IQ	0	14	15	85.590	88.715	(0.435)	(0.610)
Father at Home	0	14	15	0.424	0.209	(0.265)	(0.425)

Note: This table shows the balance in observed characteristics between the treatment and control groups in ABC at baseline for cohort 3. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.17: First-phase Treatment vs. Control Groups, ABC Cohort 4

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	p-value Multiple H_0
Male	0	15	14	0.599	0.567	(0.870)	(0.905)
Birth Weight	0	15	14	7.321	7.150	(0.725)	(0.840)
No. Siblings in Household	0	15	14	0.490	0.977	(0.220)	(0.380)
Birth Year	0	15	14	1977	1977	(0.615)	(0.728)
Mother's Education	0	15	14	9.530	10.424	(0.240)	(0.410)
Mother's Age	0	15	14	19.941	21.712	(0.320)	(0.570)
Mother Employed	0	15	14	0.260	0.347	(0.650)	(0.840)
Parental Income	0	15	14	5,827	10,781	**(0.065)**	(0.135)
Mother's IQ	0	15	14	85.561	86.004	(0.920)	(0.960)
Father at Home	0	15	14	0.208	0.138	(0.570)	(0.777)

Note: This table shows the balance in observed characteristics between the treatment and control groups in ABC at baseline for cohort 4. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Third, we compare the second-phase treatment and control groups on baseline characteristics.

Table A.18: Second-phase Treatment vs. Control Groups, ABC

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	p-value Multiple H_0
Male	0	47	48	0.551	0.460	(0.420)	(0.552)
Birth Weight	0	47	48	7.084	6.929	(0.610)	(0.700)
No. Siblings in Household	0	47	48	0.748	0.504	(0.285)	(0.445)
Birth Year	0	47	48	1974	1974	(0.835)	(0.915)
Mother's Education	0	47	48	10.150	10.388	(0.480)	(0.725)
Mother's Age	0	47	48	21.122	18.884	**(0.035)**	**(0.075)**
Mother Employed	0	47	48	0.314	0.256	(0.530)	(0.725)
Parental Income	0	47	48	7,589	6,714	(0.625)	(0.825)
Mother's IQ	0	47	48	83.000	85.831	(0.185)	(0.365)
Father at Home	0	47	48	0.279	0.287	(0.920)	(0.965)

Note: This table shows the balance in observed characteristics between the school-age treatment and control groups in ABC at baseline. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Fourth, we compare the observed, baseline characteristics of attrited and non-attrited sub-

jects in the first-phase treatment assignment.

Table A.19: Observed vs. Attritted Children, ABC

| | | | | Observed | Attritted | p-value | |
Variable	Age	Obs.	Att.	Mean	Mean	Single H_0	Multiple H_0
Male	0	103	13	0.488	0.248	**(0.085)**	(0.140)
Birth Weight	0	103	11	7.014	6.948	(0.825)	(0.875)
No. Siblings in Household	0	103	13	0.609	0.829	(0.600)	(0.705)
Birth Year	0	103	13	1974	1973	**(0.045)**	**(0.095)**
Mother's Education	0	103	13	10.302	9.192	**(0.100)**	(0.165)
Mother's Age	0	103	13	20.016	18.178	**(0.080)**	(0.160)
Mother Employed	0	103	13	0.268	0.255	(0.925)	(0.955)
Parental Income	0	103	12	6,622	6,442	(0.950)	(0.960)
Mother's IQ	0	103	13	85.050	78.834	**(0.070)**	(0.135)
Father at Home	0	103	13	0.278	0.329	(0.735)	(0.835)

Note: This table shows the balance in observed characteristics between ABC subjects who were followed up to at least age 21 and ABC subjects who attrited before age 21. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Fifth, we compare the observed, baseline characteristics between the subjects in the treatment and the control groups, excluding those who did not comply to treatment.

Table A.20: First-phase Treatment vs. Control Groups, Dropping Attrited Children, ABC

| | | Control | Treated | Control | Treated | p-value | |
Variable	Age	Obs.	Obs.	Mean	Mean	Single H_0	Multiple H_0
Male	0	51	52	0.452	0.524	(0.430)	(0.600)
Birth Weight	0	51	52	7.210	6.822	(0.115)	(0.220)
No. Siblings in Household	0	51	52	0.767	0.455	(0.150)	(0.230)
Birth Year	0	51	52	1974	1974	(0.635)	(0.785)
Mother's Education	0	51	52	10.000	10.598	**(0.085)**	(0.185)
Mother's Age	0	51	52	20.412	19.635	(0.405)	(0.615)
Mother Employed	0	51	52	0.221	0.314	(0.245)	(0.455)
Parental Income	0	51	52	6,409	6,846	(0.765)	(0.870)
Mother's IQ	0	51	52	84.472	85.635	(0.560)	(0.755)
Father at Home	0	51	52	0.349	0.208	(0.115)	(0.255)

Note: This table shows the balance in observed characteristics between the treatment and control groups of ABC subjects who were followed up to at least age 21. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Finally, we compare the observed, baseline characteristics between the children in the first-phase treatment, restricting the sample to the children for whom we have information on the age-34 medical data collection.

Table A.21: First-phase Treatment vs. Control Groups, Subjects Completing the Health Follow-up, ABC

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	Multiple H_0
Male	0	31	39	0.293	0.533	**(0.050)**	**(0.055)**
Birth Weight	0	31	39	7.233	6.826	(0.190)	(0.295)
No. Siblings in Household	0	31	39	0.613	0.493	(0.580)	(0.750)
Birth Year	0	31	39	1975	1974	(0.360)	(0.510)
Mother's Education	0	31	39	10.039	10.597	(0.190)	(0.385)
Mother's Age	0	31	39	19.389	19.595	(0.825)	(0.945)
Mother Employed	0	31	39	0.195	0.349	(0.185)	(0.315)
Parental Income	0	31	39	5,509	7,520	(0.280)	(0.535)
Mother's IQ	0	31	39	83.822	84.922	(0.655)	(0.860)
Father at Home	0	31	39	0.355	0.231	(0.205)	(0.450)

Note: This table shows the balance in observed characteristics between the treatment and control groups in ABC at baseline for subjects who completed the health follow-up at age 34. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Despite some exceptions, these tables indicate balance between the treatment and control groups from the first-phase randomization, which is the primary comparison we analyze in the main paper. The balance in observed characteristics holds for the different samples we consider, which differs from the initial sample due to various instances of item non-response. For the second-phase randomization, there is also balance in observed characteristics.

Table A.22 through Table A.29 are the analogous tables for CARE. We compare the two treatment groups (center-based childcare and family education, and only family education) separately across the full sample and by cohort. The inference statistics are constructed using the same methods as for Table A.13 through Table A.21.

49

Table A.22: CARE Baseline Characteristics, Control vs. Family Education and Center-based Childcare

| | | Control | Treated | Control | Treated | p-value | |
Variable	Age	Obs.	Obs.	Mean	Mean	Single H_0	Multiple H_0
Male	0	23	17	0.611	0.524	(0.565)	(0.740)
Birth Weight	0	23	15	7.102	7.508	(0.335)	(0.515)
No. Siblings in Household	0	23	17	0.619	0.653	(0.895)	(0.945)
Birth Year	0	23	17	1979	1979	(0.890)	(0.920)
Mother's Education	0	23	17	11.195	10.693	(0.390)	(0.500)
Mother's Age	0	23	17	21.636	21.896	(0.870)	(0.915)
Mother's IQ	0	23	17	87.584	86.624	(0.725)	(0.825)
Father at Home	0	23	17	0.127	0.351	**(0.095)**	(0.175)

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.23: CARE Baseline Characteristics, Control vs. Family Education

| | | Control | Treated | Control | Treated | p-value | |
Variable	Age	Obs.	Obs.	Mean	Mean	Single H_0	Multiple H_0
Male	0	23	27	0.611	0.632	(0.895)	(0.880)
Birth Weight	0	23	26	7.102	6.963	(0.755)	(0.830)
No. Siblings in Household	0	23	27	0.619	0.758	(0.680)	(0.715)
Birth Year	0	23	27	1979	1979	(0.330)	(0.480)
Mother's Education	0	23	27	11.195	10.689	(0.245)	(0.380)
Mother's Age	0	23	27	21.636	20.257	(0.305)	(0.400)
Mother's IQ	0	23	27	87.584	87.167	(0.855)	(0.915)
Father at Home	0	23	27	0.127	0.190	(0.455)	(0.585)

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.24: CARE Baseline Characteristics, Control vs. Family Education and Center-based Childcare, Cohort 5

| | | Control | Treated | Control | Treated | p-value | |
Variable	Age	Obs.	Obs.	Mean	Mean	Single H_0	Multiple H_0
Male	0	7	6	0.560	0.655	(0.810)	(0.860)
Birth Weight	0	7	4	7.223	7.502	(0.570)	(0.730)
No. Siblings in Household	0	7	6	0.428	0.541	(0.800)	(0.870)
Birth Year	0	7	6	1978	1978	(0.425)	(0.355)
Mother's Education	0	7	6	11.035	11.164	(0.865)	(0.875)
Mother's Age	0	7	6	18.808	21.652	(0.140)	(0.220)
Mother's IQ	0	7	6	89.202	92.345	(0.620)	(0.680)
Father at Home	0	7	6	0.289	0.322	(0.935)	(0.938)

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline for cohort 5. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.25: CARE Baseline Characteristics, Control vs. Family Education, Cohort 5

| | | Control | Treated | Control | Treated | p-value | |
Variable	Age	Obs.	Obs.	Mean	Mean	Single H_0	Multiple H_0
Male	0	7	14	0.560	0.504	(0.885)	(0.885)
Birth Weight	0	7	14	7.223	6.742	(0.580)	(0.710)
No. Siblings in Household	0	7	14	0.428	1.046	(0.295)	(0.365)
Birth Year	0	7	14	1978	1978	(0.175)	(0.190)
Mother's Education	0	7	14	11.035	10.699	(0.610)	(0.735)
Mother's Age	0	7	14	18.808	20.824	(0.210)	(0.285)
Mother's IQ	0	7	14	89.202	90.710	(0.695)	(0.775)
Father at Home	0	7	14	0.289	0.219	(0.755)	(0.790)

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline for cohort 5. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.26: CARE Baseline Characteristics, Control vs. Family Education and Center-based Childcare, Cohort 6

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	Multiple H_0
Male	0	16	11	0.636	0.453	(0.395)	(0.625)
Birth Weight	0	16	11	7.041	7.509	(0.410)	(0.645)
No. Siblings in Household	0	16	11	0.703	0.720	(0.975)	(0.980)
Birth Year	0	16	11	1979	1979	(0.565)	(0.498)
Mother's Education	0	16	11	11.268	10.441	(0.250)	(0.385)
Mother's Age	0	16	11	22.884	22.039	(0.690)	(0.750)
Mother's IQ	0	16	11	86.841	83.388	(0.245)	(0.370)
Father at Home	0	16	11	0.057	0.358	**(0.045)**	**(0.095)**

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline for cohort 6. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.27: CARE Baseline Characteristics, Control vs. Family Education, Cohort 6

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	Multiple H_0
Male	0	16	12	0.636	0.747	(0.575)	(0.655)
Birth Weight	0	16	12	7.041	7.208	(0.675)	(0.745)
No. Siblings in Household	0	16	12	0.703	0.490	(0.515)	(0.600)
Birth Year	0	16	12	1979	1979	(0.420)	(0.540)
Mother's Education	0	16	12	11.268	10.668	(0.355)	(0.493)
Mother's Age	0	16	12	22.884	19.905	**(0.075)**	(0.125)
Mother's IQ	0	16	12	86.841	82.920	(0.185)	(0.295)
Father at Home	0	16	12	0.057	0.177	(0.380)	(0.420)

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline for cohort 6. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.28: CARE Baseline Characteristics, Control vs. Family Education and
Center-based Childcare
Subjects Completing the Health Follow-up

| | | Control | Treated | Control | Treated | p-value | |
Variable	Age	Obs.	Obs.	Mean	Mean	Single H_0	Multiple H_0
Male	0	11	11	0.467	0.550	(0.690)	(0.815)
Birth Weight	0	11	11	6.783	7.633	(0.110)	(0.200)
No. Siblings in Household	0	11	11	0.372	0.546	(0.665)	(0.740)
Birth Year	0	11	11	1979	1979	(0.115)	(0.193)
Mother's Education	0	11	11	11.391	11.027	(0.615)	(0.703)
Mother's Age	0	11	11	22.142	21.607	(0.865)	(0.860)
Mother's IQ	0	11	11	86.317	87.505	(0.745)	(0.825)
Father at Home	0	11	11	0.085	0.362	(0.110)	(0.185)

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline for subjects who completed the health follow-up at age 34. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Table A.29: CARE Baseline Characteristics, Control vs. Family Education
Subjects Completing the Health Follow-up

| | | Control | Treated | Control | Treated | p-value | |
Variable	Age	Obs.	Obs.	Mean	Mean	Single H_0	Multiple H_0
Male	0	11	18	0.467	0.446	(0.940)	(0.935)
Birth Weight	0	11	18	6.783	6.262	(0.325)	(0.430)
No. Siblings in Household	0	11	18	0.372	0.383	(0.960)	(0.955)
Birth Year	0	11	18	1979	1979	(0.120)	(0.240)
Mother's Education	0	11	18	11.391	11.236	(0.795)	(0.845)
Mother's Age	0	11	18	22.142	19.941	(0.230)	(0.290)
Mother's IQ	0	11	18	86.317	87.611	(0.700)	(0.745)
Father at Home	0	11	18	0.085	0.237	(0.240)	(0.320)

Note: This table shows the balance in observed characteristics between the treatment and control groups at baseline for subjects who completed the health follow-up at age 34. For each characteristic, we present the p-value from a single hypothesis test. We also present the p-values from multiple hypothesis testing, where we collectively test the baseline characteristics within the blocks separated by the horizontal line. Both p-values are two-sided and non-parametric. We construct them based on 200 re-draws of the full sample.

Overall, these tables indicate a balance between the treatment and control groups, both when considering center-based childcare and family education and only family education as treatment.

53

Tables A.30 and A.31 show the treatment effects on maternal education, which change after baseline.

Table A.30: Maternal Education, ABC/CARE, Females

	N		Mean		Std. Dev.		Two-sided
	Treat.	Control	Treat.	Control	Treat.	Control	t-test
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years of Education							
Birth	37	51	10.65	10.49	1.69	1.99	0.69
1.5 years	33	47	11.36	11.15	1.43	1.89	0.56
2.5 years	33	47	11.42	11.21	1.44	1.90	0.57
3.5 years	32	47	11.53	11.38	1.32	1.97	0.69
4.5 years	30	45	12.13	11.82	1.14	2.00	0.40
5.5 years	24	24	12.46	11.38	0.98	2.37	0.05
8 years	25	31	12.64	12.16	1.08	2.48	0.34
12 years	28	37	14.82	13.49	1.66	2.61	0.01
15 years	23	25	14.96	13.88	2.03	2.15	0.08
21 years	22	26	15.14	14.00	2.03	2.26	0.07
Education Level							
12 years	28	37	4.46	3.95	0.79	1.27	0.05
15 years	23	25	4.13	3.88	1.36	1.09	0.49
Graduated High School							
12 years	37	51	1.00	0.90	0.00	0.30	0.02
15 years	30	32	1.00	0.91	0.00	0.30	0.08

Note: This table shows raw descriptives of the education variables. Columns (1) through (6) give the sample size, means, and standard deviations of the variables by experimental group. In this table, "Treat." indicates the treatment group that received ABC/CARE center-based childcare and "Control" is the control group. Column (7) gives the two-sided p-value for a t-test of means between the two groups accounting for different variances in the groups. Education level is a categorical variable with higher values corresponding to more education. The ages in parentheses are the ages of the *subjects* when the measure was collected.

Table A.31: Maternal Education, ABC/CARE, Males

	N		Mean		Std. Dev.		Two-sided
	Treat.	Control	Treat.	Control	Treat.	Control	t-test
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years of Education							
Birth	38	56	10.53	10.25	1.74	1.64	0.44
1.5 years	35	50	11.14	10.86	1.70	1.47	0.43
2.5 years	35	50	11.14	10.94	1.70	1.45	0.57
3.5 years	35	48	11.23	11.02	1.73	1.49	0.57
4.5 years	34	47	11.44	11.19	1.76	1.58	0.51
5.5 years	24	20	11.75	11.05	1.98	1.79	0.23
8 years	31	29	12.32	11.72	2.07	1.46	0.20
12 years	28	37	14.57	13.22	2.63	1.90	0.03
15 years	24	21	14.88	14.00	2.35	2.14	0.20
21 years	25	20	14.20	14.35	2.57	2.28	0.84
Education Level							
12 years	28	37	4.36	3.81	1.22	1.05	0.06
15 years	24	21	4.46	4.00	1.25	1.00	0.18
Graduated High School							
12 years	38	56	0.92	0.93	0.27	0.26	0.89
15 years	29	25	0.97	0.92	0.19	0.28	0.49

 Note: This table shows raw descriptives of the education variables. Columns (1) through (6) give the sample size, means, and standard deviations of the variables by experimental group. In this table, "Treat." indicates the treatment group that received ABC/CARE center-based childcare and "Control" is the control group. Column (7) gives the two-sided p-value for a t-test of means between the two groups accounting for different variances in the groups. Education level is a categorical variable with higher values corresponding to more education. The ages in parentheses are the ages of the *subjects* when the measure was collected.

A.8 Details on Educare

Table A.32: Educare – ABC/CARE Comparison

Educare Model Core Feature	Abecedarian Program Practice
Research-Based Practices and Strategies[*] • Design informed by research[*] • Evaluation through FPG[*]	Research-Based Practice and Strategies[*] • Design informed by research[*] • Conducted and evaluated at FPG[*]
Small Class and High Staff/Child Ratios[†] • Infants: 8 children, 3 adults[†] • Preschool: 17 children, 3 adults[*]	Small Class and High Staff/Child Ratios[†] • Infants: 14 children, 4–5 adults • Preschool: 14–18 children, 3 adults[*]
High Staff Qualifications[‡] / Intensive Professional Development[*] • Adults: Teacher with BA, Assistant with AA, Aide with HS/GED[‡] • Training and development a priority[*] • Good wages and benefits[*]	Low Staff Qualifications[‡] / Intensive Professional Development[*] • "Good with children" (not even HS required)[‡] • VERY intensive on-site, continuous training[*] • Good wages and benefits[*]
Focus on Language and Literacy - YES[*]	Focus on Language and Literacy - YES[*]
Emphasis on Social-Emotional Development to Promote School Readiness - YES[*]	Emphasis on Social-Emotional Development to Promote School Readiness - YES[*]
Continuity of care - YES[‡] • Stay with same teaching team infancy – age three; preschool – K[‡]	Continuity of care - No[‡] • New teaching team each year as moved up to next age[‡]
On-Site Family Support Services - YES[*]	On-Site Family Support Services - YES[*]
Reflective Supervision and Practice[*] • "Reflective Practice" model implemented throughout organization[*]	Reflective Supervision and Practice[*] • Although not called such, the spirit of Reflective Practice used in teacher training[*]
Interdisciplinary Team Approach - YES[*]	Interdisciplinary Team Approach - YES[*]
Art Programs Used to Support Social-Emotional, Language, and Literacy Development - YES[*]	Art Programs Used to Support Social-Emotional, Language, and Literacy Development - YES[*]
Emphasis on Prenatal Services - YES[‡]	Emphasis on Prenatal Services - NO[‡]
Enhanced Focus on Problem-Solving & Numeracy - YES[*]	Enhanced Focus on Problem-Solving & Numeracy - YES[*]
Home Visits - YES[*]	Home Visits - YES[*] (after school begins) - NO[‡] (preschool) [For CARE - YES[*] in both preschool and in school years]

Items that match[*]
Items that are close[†]
Items that don't match[‡]

Note: This table compares ABC/CARE to Educare (and then the clarifications on the symbols currently in place). (CARE and ABC are the same, except for the home policy visit noted in the last block below.)

Table A.33: Educare Programs in the United Stated

State	Program Name	Currently Served	Program Start
Arizona	Educare Arizona	191	2011
Atlanta	Educare Atlanta	215	2012
California	Educare California at Silicon Valley	168	2015
California	Educare Los Angeles at Long Beach	190	2018
Colorado	Educare Denver	120	2007
DC	Educare Washington, DC	166	2012
Florida	Educare Miami-Dade	116	2008
Illinois	Educare Chicago	146	2000
Illinois	Educare West DuPage	150	2012
Kansas	Educare Kansas City	132	2010
Louisiana	Educare New Orleans	168	2013
Maine	Educare Central Maine	116	2010
Nebraska	Educare Lincoln	158	2013
Nebraska	Educare Omaha at Kellom	156	2003
Nebraska	Educare Omaha at Indian Hill	191	2009
Nebraska	Educare Winnebago	175	2014
Oklahoma	Educare Oklahoma City	212	2009
Oklahoma	Educare Tulsa at Kendall-Whittier	196	2006
Oklahoma	Educare Tulsa at Hawthorne	160	2010
Oklahoma	Educare Tulsa at MacArthur	164	2012
Washington	Educare Seattle	156	2010
Wisconsin	Educare Milwaukee	166	2005

Note: This table lists and details the current Educare programs in the United States.

B   Program Costs[68]

In this appendix, we document the sources informing our programs' costs calculation in Section 4.4. We use a battery of primary sources obtained in the Archives of the University of North Carolina at Chapel Hill—Records of the Office of the Vice Chancellor University of North Carolina Archives (UNC). Figure B.1 exemplifies one of these sources, in which the monthly cost of treatment per child for the year 1977 is categorized by source of funding.

---

[68]Sylvi Kuperman greatly assisted us in preparing this section of the appendix.

Figure B.1: Primary-source Document Costs, Example

Frank Porter Graham Child Development Center
Highway 54 Bypass West
Chapel Hill, N.C. 27514

THE FRANK PORTER GRAHAM DAY CARE PROGRAM:

WHO PAYS WHAT?

The following table breaks down the costs of FPG's day care program. The second column shows how much of the total costs are borne by Title XX funds from Orange County. Out of a total cost to FPG of $417 per child each month, Title XX funds account for $85 per child, or about 20% of the total. Footnote "g" at the bottom of the table shows that an increase in the Orange County appropriate to $8,000 per month would raise Title XX supplement for each child to $174 per month at most.

Cost Category	Current Monthly Cost Rate	DISTRIBUTION OF COSTS			
		Title XX Contract[g]	State Via UNC	Federal Grant[d]	To be Determined
Salaries	13,956[a]	3,438[c]	2,675	3,174	4,669
Fringe Benefits	2,379[a]	585[c]	467	540	787
Non-Personnel	2,150[b]	-0-	1,796	354	-0-
Total Direct	18,485	4,023	4,938	4,068	5,456
Indirect Cost	3,210	402	615	1,528	665
Total Cost	21,695	4,425	5,553	5,596	6,121
% of Total	100	20.4	25.6	25.8	28.2

Enrollment

Current = 52[f]
Cost/Child/Mo.= $417 | $85 | $107 | $107 | $118
Scheduled[e] = 46-56
Cost/Child/Mo.= $472-387 | $96-79 | $121-99 | $121-100 | $133-109

a - May, 1977
b - Monthly Average, FY 77
c - Monthly Average, Oct., 76 thru Apr., 77
d - Expended 5/1/77
e - In FY 78
f - DSS eligibility in May = 58.9%

g - Scheduled to increase eff. May, 1977 to approx. $8,000/mo., for 5 mos., which would = 36.8% of costs and $154/mo./child at current enrollment and $174-$143/child, at scheduled enrollment level.

Note: This figure is a photograph exemplifying the primary-source document we use in the calculation of the programs' cost. It was obtained in the University of North Carolina at Chapel Hill—Records of the Office of the Vice Chancellor

Figure B.2 is another of our sources and it is the base of our calculation of personnel costs.

Figure B.2: Primary-source Document Costs, Personnel Wages

 Note: This figure is a photograph provides the estimates for the personnel wages we use. It was obtained in the University of North Carolina at Chapel Hill—Records of the Office of the Vice Chancellor

Interviews with the programs' staff Kuperman and Cheng (2014); Kuperman (2015), inform us about additional costs of the programs. An example is the salary of a social worker, who is not part of some of the of the costs estimates reported before but was part of the staff implementing treatment. These interviews are available upon request.

Finally, a valuable source is a report written by the program staff Frank Porter Graham Child Development Center (1979). As we note in Section 4.4, this report produced an estimate of the cost in a completely independent way, although perhaps using the same or similar primary sources. Our calculation of the costs comes very close to those in Frank Porter Graham Child Development Center (1979).

We summarize the yearly program costs of ABC/CARE in Table B.1.

Table B.1: Yearly Program Costs, ABC/CARE

Item	Yearly Cost in 2014 USD
1 Program Director	60,935
1 Social Worker	35,869
3 Lead Teachers and 2 Teachers Aides (Nursery)	204,457
4 Lead Teachers and 4 Teacher Aides (Toddlers)	305,181
2 Teaching Support Staff	53,341
1 Secretary	32,973
1 Clerk	32,537
Workers' Fringe Benefits	124,935
Other	4,891
Total	962,726
Total per Subject	18,514

Note: This table summarizes the yearly costs for ABC/CARE. They are based on primary-source documentation describing ABC. We assume that the costs for ABC and CARE were the same based on conversations with programs' staff (Kuperman and Cheng, 2014; Kuperman, 2015).

C Identification and Estimation of Life-Cycle Treatment Effects

This appendix presents our approach for identifying and estimating life-cycle treatment effects. Differences in the approach used for each outcome are based on different scenarios of data availability. We proceed as follows. Appendix C.1 focuses on outcomes that are fully observed over the course of the experiment with little attrition. Appendix C.2 focuses on outcomes that are partially observed over the course of the experiment with substantial rates of attrition. Appendix C.3 focuses on outcomes that we do not observe, and thus need to forecast out of sample. Appendices C.4 and C.5 provide technical details on the computations of the internal rate of return and the benefit/cost ratios, respectively. Appendix C.7 provides the steps for constructing our statistical inference. Finally, Appendix C.8 describes our procedure for selecting the background variables that we condition on.

C.1 Complete Data

We classify a variable as having essentially complete data when we observe the data for at least 90% of the individuals in the sample Table C.1 lists the variables that are completely observed. For these outcomes, we estimate the standard errors of our estimates by resampling the ABC/CARE data with replacement. We estimate non-parametric p-values based on the bootstrap distribution.

Table C.1: Variables Estimated without IPW Adjustment

Completely Observed Outcomes	Age (years)
IQ Standard Score	2, 3, 3.5, 4, 4.5, 5, 12, 15, 21
PIAT Math Standard Score	7
Achievement Score	15, 21
HOME Total Score	0.5, 1.5, 2.5, 3.5, 4.5
Mother Works	2, 3, 4, 5, 21
Biological Mother's Education Level	2, 3, 4, 5, 9
Father is Home	2, 3, 4, 5, 8
Graduated High School	NA
Attended Vocation/Tech/Community College	NA
Years of Education	30
Ever Had Special Education	NA
Total Number of Years in Special Education	NA
Ever Retained	NA
Total Number of Retained Grades	NA
Employed	30
Labor Income	21, 30
Transfer Income	30
Total Years Incarcerated	30
Self-reported Health	30
Brief Symptom Inventory Score	21
Number of Cigarettes Smoked Per Day Last Month	30
Number of Days Drank Alcohol Last Month	30
Number of Days Binge Drank Alcohol Last Month	30
Program Costs	0–26
Control Contamination Costs	0–26
Education Costs	0–26
Medical Expenditure	8–30
Justice System Costs	0–50
Prison Costs	0–50
Victimization Costs	0–50

Note: The table above lists the variables which we observe completely for the full sample.

C.2 Partially Complete Data

When we do not observe data on an outcome for more than 10% of the individuals in the sample, we consider the outcome to be partially complete. These outcomes include: parental labor income at ages 1.5, 2.5, 3.5, 8, 12, 15, and 21, for which we observe no more than 112

subjects at any given age; and items in the health survey at age 34, for which we observe no more than 93 subjects. Table C.2 lists the variables that we classify as partially complete.

For partially complete outcomes, we correct for attrition using an inverse probability weighting scheme (IPW) as in Horvitz and Thompson (1952). For each partially observed outcome, we construct an IPW scheme. The weights are based on a set of pre-treatment variables that we observe for the complete sample. We use a set of baseline variables to estimate the propensity of observing the value of a partially complete outcome. These probabilities are estimated using a logistic regression. The set of baseline variables is chosen among many possible control sets, as documented in Appendix C.8.1. For each partially observed outcome, we list the variables that we use to produce the IPW scheme in Table C.2. To estimate the standard errors of our estimates, we resample the ABC/CARE data with replacement as in Appendix C.1. Each time we resample, we reestimate the IPW weights to account for estimation error. Our standard errors thus account for sampling uncertainty and for estimation error in the computation of the IPW scheme. We estimate non-parametric p-values based on the bootstrap distribution.

Table C.2: Variables Used to Create IPW Scheme

Partially Observed Outcomes	Age	N	Variables Used to Produce IPW		
IQ Score	6.5	126	High Risk Index (HRI)	APGAR 1 min.	Cohort
IQ Score	7	118	High Risk Index (HRI)	APGAR 5 min.	Cohort
IQ Score	8	125	High Risk Index (HRI)	APGAR 1 min.	Cohort
Achievement Score	5.5	105	High Risk Index (HRI)	APGAR 1 min.	Cohort
Achievement Score	6	124	High Risk Index (HRI)	APGAR 1 min.	Cohort
Achievement Score	6.5	89	High Risk Index (HRI)	APGAR 1 min.	Cohort
Achievement Score	7	90	High Risk Index (HRI)	APGAR 1 min.	Cohort
Achievement Score	7.5	121	High Risk Index (HRI)	APGAR 1 min.	Cohort
Achievement Score	8	123	High Risk Index (HRI)	APGAR 1 min.	Cohort
Achievement Score	8.5	122	High Risk Index (HRI)	APGAR 5 min.	Cohort
Parental Labor Income	1.5	112	Mother's Age at Baseline	APGAR 1 min.	Cohort
Parental Labor Income	2.5	112	Mother's Age at Baseline	APGAR 1 min.	Cohort
Parental Labor Income	3.5	110	Mother's Age at Baseline	APGAR 1 min.	Cohort
Parental Labor Income	8	87	High Risk Index (HRI)	APGAR 1 min.	Cohort
Parental Labor Income	12	108	High Risk Index (HRI)	APGAR 1 min.	Cohort
Parental Labor Income	15	92	APGAR 5 min.	Premature at Birth	Number of Siblings, Baseline
Parental Labor Income	21	73	High Risk Index (HRI)	APGAR 1 min.	Cohort
HOME Score	8	100	High Risk Index (HRI)	APGAR 1 min.	Cohort
Father at Home	8	116	High Risk Index (HRI)	APGAR 1 min.	Cohort
Subject Public Transfer Income	21	105	High Risk Index (HRI)	APGAR 1 min.	Cohort
Total Felony Arrests	Mid-30s	115	APGAR 1 min.	APGAR 5 min.	Cohort
Total Misdemeanor Arrests	Mid-30s	115	APGAR 1 min.	APGAR 5 min.	Cohort
Self-reported Health	Mid-30s	92	APGAR 1 min.	APGAR 5 min.	Premature at Birth
Self-reported Drug User	Mid-30s	89	APGAR 1 min.	APGAR 5 min.	Premature at Birth
Systolic Blood Pressure (mm Hg)	Mid-30s	90	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Diastolic Blood Pressure (mm Hg)	Mid-30s	90	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Prehypertension, Sys. B.P. > 120 or Dys. B.P. > 80	Mid-30s	90	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Hypertension, Sys. B.P. > 140 or Dys. B.P. > 90	Mid-30s	90	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
High-Density Lipoprotein (HDL) Cholesterol (mg/dL)	Mid-30s	93	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Dyslipidemia (HDL < 40 mg/dL)	Mid-30s	93	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Hemoglobin Level (%)	Mid-30s	92	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Prediabetes, Hemoglobin > 5.7%	Mid-30s	92	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Diabetes, Hemoglobin > 6.5%	Mid-30s	92	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Vitamin D Deficiency (< 20 ng/mL)	Mid-30s	93	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Measured BMI	Mid-30s	88	APGAR 1 min.	APGAR 5 min.	Premature at Birth
Obesity (BMI > 30)	Mid-30s	90	APGAR 1 min.	APGAR 5 min.	Premature at Birth
Severe Obesity (BMI > 35)	Mid-30s	91	APGAR 1 min.	APGAR 5 min.	Premature at Birth
Waist-hip Ratio	Mid-30s	84	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Abdominal Obesity	Mid-30s	84	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Framingham Risk Score	Mid-30s	88	APGAR 1 min.	Premature at Birth	Number of Siblings, Baseline
Brief Symptom Survey (BSI) Score	Mid-30s	92	APGAR 1 min.	APGAR 5 min.	Premature at Birth

Note: This table lists the partially observed variables and the variables that we use to construct the IPW scheme to account for attrition when calculating treatment effects. The procedure to select these variables is described in Appendix C.8.1. We construct the IPW scheme using a common model across males and females.

Partially observed outcomes can occur at any age $a \leq a^*$.[69] We construct the IPW scheme using pre-treatment variables, within the age period $a \leq a^*$.

We construct the IPW scheme using the same algorithm, independent of the age a for $a \leq a^*$ in which an outcome is partially complete. For notational simplicity, we derive the IPW scheme without indexing the outcomes by age or treatment status (we estimate IPW using a common logistic regression across treatment regimes to avoid instability in the estimates that could be a problem in our small experimental sample). We restore the notation used throughout the text in the next appendix.

We use a standard inverse probability weighting (IPW) scheme.[70] Formally, recall that $R = 1$ if the child is randomized into treatment, and $R = 0$ otherwise.[71] Similarly, let $A = 1$ denote the case where we observe a generic scalar outcome Y (which is a generic entry in the vector of outcomes that we use in the paper), and $A = 0$ otherwise. As in the main text, $\boldsymbol{B}$ represents background (pre-treatment) variables.

We assume A is independent of Y conditional on $\boldsymbol{B}$. More formally, we invoke

Assumption AA–1

$$A \perp\!\!\!\perp Y | \boldsymbol{B}, R.$$

Let Y^r represent outcome Y when R is fixed to take the value r. Based on Assumption AA–1, we use IPW to identify $\mathbb{E}[Y^r]$ as follows:

---

[69]After a^*, we have incomplete data and do not observe any outcome. We explain the details for constructing predictions in Appendix C.3.

[70]Horvitz and Thompson (1952).

[71]We are able to use R (randomization into treatment) and D (participation in treatment) exchangeably as we argue in Section 3.

$$\mathbb{E}[Y^r] = \int \int y f_{Y|\boldsymbol{B}=b,R=r,A=1}(y|\boldsymbol{B}=b,R=r,A=1)f_{\boldsymbol{B}}(b)dydb \tag{1}$$

where each component of the last expression in (1) is straightforward to recover from the data. Using Bayes' Theorem, we can write an equivalent expression to make the IPW scheme explicit. That is, we apply Bayes' Theorem to $f_{\boldsymbol{B}}(b)$ to obtain

$$f_{\boldsymbol{B}}(b) = \frac{f_{\boldsymbol{B}|R=r,A=1}(b|R=r,A=1)P(R=r,A=1)}{P(R=r,A=1|\boldsymbol{B}=b)}.$$

Substituting these expressions into (1), we obtain

$$\mathbb{E}[Y^r] = \int \int y f_{Y,\boldsymbol{B}|R=r,A=1}(y,b|R=r,A=1)\frac{P(R=r,A=1)}{P(R=r|\boldsymbol{B}=b)P(A=1|R=r,\boldsymbol{B}=b)}dydb.$$

Assumption AA–1 generalizes the matching assumption of Campbell et al. (2014). The corresponding sample estimator for $\mathbb{E}[Y^r]$ is

$$\sum_{i \in \mathcal{I}} y_i \beta_i \boldsymbol{1}(r_i = r)\boldsymbol{1}(a_i = 1)$$

where $\mathcal{I}$ indexes the individuals in the sample. As in the paper that lowercase variables r, y, and b denote realizations of the random variables R, Y, and $\boldsymbol{B}$, respectively.

$$\beta_i = \frac{1}{\pi_r(b_i)\alpha(r_i, b_i)} \frac{1}{\sum_{k \in \mathcal{I}} \frac{\mathbf{1}(r_i=r)\mathbf{1}(a_i=1)}{\pi_r(b_k)\alpha(r_k,b_k)}},$$

with $\pi_r(b) := P(R = r | \boldsymbol{B} = b)$ and $\alpha(r, b) := P(A = 1 | R = r, \boldsymbol{B} = b)$. The weight π_r corrects for selection into treatment based on pre-treatment variables $\boldsymbol{B}$. The weight α corrects for item non-response based on $R, \boldsymbol{B}$.

In Tables 4 and 5, we present estimates of the internal rate of return and the benefit cost-ratio with and without adjusting for IPW. There is little sensitivity of our estimates to these adjustments.

C.3 Incomplete Data: Forecasting and Monetizing Life-Cycle and Costs and Benefits

We do not observe certain post-a^* life-cycle profiles for outcomes that are important for estimating the lifetime benefits of ABC/CARE. Examples include parental labor income, subject labor income, public-transfer income, and health-related outcome variables. In Section 3, we discuss our strategy to to identify and estimate these profiles and forecast post-a^* outcomes. This appendix documents how we implement this strategy and provides complementary evidence supporting it, cited throughout the main text. We devote separate appendices to the discussion of crime and health forecasts in Appendix E and Appendix G, respectively.[72]

---

[72]In this section, attrition or partial observations is not an issue: the predictors that we use to construct out-of-sample forecasts are classified as complete data (see Appendices C.1 and C.2).

C.3.1 Conditions for Valid Out-of-Sample Forecasts

Forecasting does not require that we take a position on the exogeneity of $\boldsymbol{X}_{k,a}^d$ for $k \in \{e, n\}$ and $a \in \{a, \ldots \bar{A}\}$ with respect to labor income. Estimated structural models with biased parameters can still give reliable forecasts if relationships between observed and unobserved variables are the same within sample and in the forecast sample. Conditions C–1 to C–3 spell out these requirements, dropping age indexing for simplicity. In the paper, we rely on Assumptions A–1 and A–3 because it facilitates the use of economic theory to interpret treatment effects, to forecast outcomes in samples where inputs are manipulated differently than in the experimental sample, and to test the validity of the construction of our synthetic cohort.

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*

$$F_{\boldsymbol{Y}_e^d, \boldsymbol{X}_e^d | \boldsymbol{B}_e = \boldsymbol{b}} (\cdot, \cdot) = F_{\boldsymbol{Y}_n^d, \boldsymbol{X}_n^d | \boldsymbol{B}_n = \boldsymbol{b}} (\cdot, \cdot), \quad d \in \{0, 1\} \tag{2}$$

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 only use means for benefit/cost analysis, we can get by with a weaker requirement for conditional means.

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

$$\mathbb{E}\left[\boldsymbol{Y}_e^d | \boldsymbol{X}_e^d = \boldsymbol{x}, \boldsymbol{B}_e = \boldsymbol{b}\right] = \mathbb{E}\left[\boldsymbol{Y}_n^d | \boldsymbol{X}_n^d = \boldsymbol{x}, \boldsymbol{B}_n = \boldsymbol{b}\right], \quad d \in \{0, 1\} \tag{3}$$

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

This can be weakened further if we rely on treatment effects.

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

$$\mathbb{E}\left[Y_e^1 - Y_e^0 | B_e = b\right] = \mathbb{E}\left[Y_n^1 - Y_n^0 | B_n = b\right] \tag{4}$$

over $supp\left(Y_{e,a}^d, B_e\right)$.[73]

We could simply invoke these conditions and be done. Our approach is to examine and test (when possible) assumptions that justify the treatment effects, which is why we develop the theory in Section 3 instead of simply invoking these assumptions.

C.3.2 Auxiliary Datasets

We rely on the following datasets to estimate life-cycle transfer and labor income profiles.[74] We use some of these same and other complementary sources to forecast the rest of the outcomes, as we explain below.

The National Longitudinal Survey of Youth (NLSY79) is a longitudinal survey beginning in 1979 that follows individuals born between 1957 and 1964. The initial interview included 12,686 respondents aged 14 to 22. The survey was designed to include 6,111 individuals representing the non-institutionalized civilian population, a supplemental sample of 5,295 civilian Hispanics, Latinos, Blacks, non-Blacks/non-Hispanics, and economically disadvantaged youth, and a sample of 1,280 who served in the military as of September 30, 1978. When appropriately weighted, the NLSY79 is nationally representative of the youth

---

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

[74]At age 21, public-transfer income includes Aid to Families with Dependent Children (AFDC) subsidies, food stamps, survivor benefits, disability benefits, social security, rent subsidies, and fuel subsidies. At age 30, public-transfer income includes food stamps, welfare, housing assistance, workman's compensation, disability, social security, supplemental security income, unemployment benefits, worker's compensation insurance, fuel subsidies, educational and aid grants, and other forms of welfare. For all other ages, we produce a forecast. We explain and justify the variables we use to forecast below.

living in the U.S. on January 1, 1979. We include individuals from all three subsamples in our analysis.

The NLSY79 collected data on labor market participation, education, family background, family life, health, assets and income, government program participation, and measures of cognitive skills.

We restrict the NLSY79 sample to Blacks with labor incomes less than $300,000 (2014 USD) at any given year to avoid estimating our forecasts with outliers in the auxiliary sample.[75] With the mean labor income (2014 USD) in the ABC/CARE sample being $32,782 at age 30, and the maximum reported being $189,938, the cut-off we impose on the auxiliary data is high enough so that the labor income support at age 30 in ABC/CARE is contained in the support of labor income at age 30 in the NLSY79.[76]

We do not impose a restriction on the birth year for the NLSY79 as all respondents are between 47 and 55 years of age at the time of the last interview (conducted in 2012). This age range is within the 31–67 range for which we extrapolate the income of the ABC/CARE subjects.

Given the biennial nature of the NLSY79, we only observe each subject at either odd or even ages. Not only does this reduce the size of the sample on which we fit our forecasting model at each age, but it can introduce biases associated with the odd-aged and even-aged cohorts. To address this issue, we linearly interpolate the variables in the NLSY79 data that are used in our forecasting model. This allows us to estimate our model on all subjects of the NLSY79 satisfying the eligibility conditions $\boldsymbol{B}_n \in \mathcal{B}_0$ at every age.

The Children of the National Longitudinal Survey of Youth (CNLSY) is a survey

---

[75]For details on how we match the individuals in the experimental sample to the individuals in the non-experimental sample, see Appendix C.3.3.

[76]For labor income at age 21, we use the CNLSY.

of the children of the mothers from the NLSY79, beginning in 1986. At the time of the initial interview, the ages of the children surveyed ranged from 0 to 23. As of 2010, the CNLSY sample includes 11,504 children born to NLSY79 mothers. With appropriate weights, the CNLSY may be considered nationally representative of children born to women who were age 14 to 22 in 1979. Interviews were conducted annually between 1986 and 1994, and biennially thereafter.

Similar to the NLSY79, the CNLSY collected data on cognitive ability, motor and social development, home environment, health information, education, attitudes, employment, income, family decisions, and more.

As we did with the NLSY79 and for the same reasons, we restrict the CNLSY sample to Black individuals with labor income less than $300,000 (2014 USD) at any given year. In addition to this, we limit the sample to subjects born between 1978 and 1983. Because the CNLSY data extends to 2012, we use the most recent data from the CNLSY in which individuals are aged 29 to 34. Finally, given the biennial nature of the survey, we perform a linear interpolation on the variables that enter into our forecasting model. This allows us to use as much of the CNLSY data as possible at every age when interpolating subject income.

The Panel Survey of Income Dynamics (PSID) is a longitudinal household survey containing between 5,000 and 8,500 families in each wave. It began as a yearly survey in 1968 and has been fielded biennially since 1996. When appropriately weighted, the PSID is designed to be representative of U.S. households. The PSID provides extensive information concerning demographics, economic outcomes, health outcomes, marriage and fertility, and more.

We restrict the PSID to Blacks born between 1945 and 1981. Because the data extend to 2013, we use the most recent subsample of individuals aged 30 to 67. We also exclude all individuals with labor income exceeding $300,000 (2014 USD) in any given year for the same

reasons. Finally, given the biennial nature of the survey, we perform a linear interpolation on the variables that enter into our forecasting model. This allows us to use as much of the PSID data as possible at every age to interpolate subject income.

Before summarizing, note that in general: (i) **we use the CNLSY to predict from ages 21 to 29**; and (ii) we use the NLSY79 and PSID to forecast from ages 29 to 79. Whenever using the NLSY79 and PSID together, we combine them and use them as a joint sample. Put differently, we use the three data sets to obtain information across our time span of interest *without placing specific weights in any particular samples.* This allows us to satisfy support conditions (see Appendix C.3.6) over the forecasted variables and the predictors that we use. An alternative to this is to to weight each of the observations in these samples according to the inverse of their variance in a set of observed characteristics. We do not do this, and instead use the unweighted observations to then construct synthetic cohorts as we explain in Appendix C.3.3.

Summary: Initial Restrictions Placed on the Auxiliary Samples

When constructing the samples used to forecast transfer and labor income, we place the following restrictions on the admissible samples *for all ages.* Additional restrictions are placed when forecasting the other outcomes, as we explain below.

1. **NLSY79:** Black, labor income less than \$300,000 (2014 USD) at any given year; subjects born between 1957 and 1965.

2. **PSID:** Black; labor income less than \$300,000 (2014 USD) at any given year; birth year between 1945 and 1981.

3. **CNLSY:** Black, labor income less than \$300,000 (2014 USD) at any given year; subjects born between 1978 and 1983.

Summary: Auxiliary Samples Used to Forecast by Age

We use the following samples to forecast labor and transfer income. Different samples are used to forecast the rest of the outcomes, as we explain below.

1. **Labor Income:**

   (a) **Ages 21 to 29:** CNLSY.

   (b) **Ages 29 to 67 (assumed age of retirement):** NLSY79 and PSID.

2. **Transfer Income:**

   (a) **Ages 21 to 29:** CNLSY.

   (b) **Ages 29 to 79 (we assume no transfers other than health-related after age 79):** NLSY79 and PSID.

C.3.3 Constructing Synthetic Cohorts

Once the initial restrictions are placed on the auxiliary samples, we weight the observations in the non-experimental sample using the following standard algorithm (e.g., Frölich, 2005).

Algorithm 1 *For individual i in experimental sample $(k = e)$, an individual $l(i)$ in the auxiliary sample $(k = n)$ is a potential counterpart if*

$$\sqrt{(\boldsymbol{B}_{i,e} - \boldsymbol{B}_{i,l(i),n})'(\boldsymbol{\Sigma}_e)^{-1}(\boldsymbol{B}_{i,e} - \boldsymbol{B}_{i,l(i),n})} \leq \epsilon \tag{5}$$

where $\boldsymbol{B}_{i,l(i),n}$ represents the observed characteristics of the matched potential counterpart in the non-experimental sample and $\boldsymbol{\Sigma}_e$ is the covariance matrix in the experimental sample. As explained in the paper, this procedure is done based on baseline variables that do not change over the life cycle and are not affected by treatment. Thus, this procedure is not done age by age. One synthetic cohort is constructed in each of the non-experimental samples by weighting the potential counterparts according to the inverse value of the left-hand-side of Equation (6).[77]

---

[77]In our empirical analysis, we set $\epsilon = 1$, but we try a range of values between 0.5 and 3 finding little sensitivity. The full set of the results we produce in the paper for multiple values in the interval [0.5,3] is available on request.

We estimate dynamic relationships between outcomes and predictors in the auxiliary samples and use the estimated relationships to generate forecasts in the non-experimental samples. When producing these estimates, we use our synthetic cohort (i.e., our weighted auxiliary samples).

Variables Used to Match to Construct Synthetic Cohorts

When constructing synthetic cohorts, we use each of the following variables, *for all the outcomes*: year of birth, gender, number of siblings at baseline. The criterion for selecting these variables is availability across all auxiliary sources. We construct one synthetic cohort in each auxiliary sample.

C.3.4 Variables Used to Forecast Out-of-Sample Outcomes

Criteria for Candidate Predictor Variables

To be considered a predictor variable, a variable must satisfy three conditions: (i) it has to be available in the experimental and non-experimental samples; (ii) it has to have predictive power for the predicted outcome. Our criterion for inclusion is that all the coefficients of a regression of the predicted variable on *all* the prediction variables are chosen to be statistically significant at the 1% level[78]; and (iii) it needs to satisfy exogeneity Assumption A–2, which we test below in Appendix C.3.6.

For a variable to satisfy restrictions (i) and (ii), it needs to be the case not only that the variable is available in the non-experimental sample but also that the survey question for it was introduced far enough back in time for us to observe it in a range that is common to the range in which we observe the variables in the experimental sample. To illustrate this, consider the case of body-mass index in the PSID. Survey questions for this variable were

---

[78]In a few cases, we decide to keep some variables slightly above this threshold because we consider them economically relevant.

introduced in the late 1990s. The sample for which the question was introduced, however, does not span enough observations for its support to cover the support of the experimental sample. To lessen this problem we pool the auxiliary datasets to maximize the available predictor variables.

We present evidence on the predictive power of the predictor variables in the synthetic cohorts that we constructed in the PSID, NLSY79, and CNLSY. Tables C.3 and C.4 show that, in each of the auxiliary samples that we use, the prediction variables $\boldsymbol{X}_{n,a}^{d}, \boldsymbol{B}_n$ are strong predictors of labor and transfer income at age 30. We present this evidence at age 30 both for brevity.

Summary: Predictor Variables

We use the following variables to forecast labor and transfer income. Different variables are used to forecast the rest of the outcomes, as we explain below.

1. **Labor Income:**

   (a) **Ages 22 to 30:** male, mother's education at birth, average PIAT Mathematics score from ages 5 to 7, years of education at age 30, one-year lagged labor income.

   (b) **Ages 31 to 67 (assumed age of retirement):** male, years of education at age 30, labor income at age 30, one-year lagged labor income.

2. **Transfer Income**

   (a) **Ages 21 to 30:** male, mother's education at birth, average PIAT Mathematics score from ages 5 to 7, years of education at age 30, transfer income at age 21, one-year lagged transfer income.

   (b) **Ages 31 to 67 (we assume no transfers other than health-related after age 79):** male, years of education at age 30, labor income at age 30, one-year lagged transfer income.

Table C.3: Predictors of Labor Income at Age 30, Synthetic Cohorts

Auxiliary Sample	(1) CNLSY	(2) NLSY79	(3) PSID
Male	3,999.07***	2,633.17***	7,087.28***
	(1,264.39)	(401.89)	(888.77)
Black	-3,366.02***	-1,236.27***	-1,900.33***
	(1,149.44)	(330.86)	(489.68)
PIAT (5-7)	83.87		
	(60.09)		
Education (30)	2,274.99***	845.98***	1,813.90***
	(410.42)	(132.05)	(195.01)
Labor Income (21)	0.13**	0.13**	0.06*
	(0.06)	(0.07)	(0.03)
Lagged Labor Income	0.53***	0.87***	0.72***
	(0.07)	(0.06)	(0.05)
Constant	-25,579.06***	-8,339.73***	-16,127.27***
	(6,453.86)	(1,354.43)	(1,968.18)
Observations	1,215	7,036	5,544
R^2	0.51	0.65	0.61

Note: All columns display regressions of labor income at age 30 on the different variables listed in the rows. If the space for the coefficient appears empty, it was not included in the regression. All money figures are in 2014 USD. The number in parentheses next to the variable indicates the age of measurement. Education is measured as years of education. Robust standard errors are in parentheses below the estimates. We weigh the individuals in the auxiliary samples to match them on observed individual characteristics to the experimental sample using the procedure in Appendix C.3.3. ***: p-value $< .01$. **: p-value $< .05$. *: p-value $< .10$.

Table C.4: Predictors of Transfer Income at Age 30, Synthetic Cohorts

Auxiliary Sample	(1) NLSY79	(2) PSID
Male	-167.28***	-515.30***
	(36.72)	(155.70)
Black	116.95***	136.30
	(44.04)	(143.46)
Education (30)	-44.70***	-103.94**
	(6.38)	(42.35)
Transfer Income (21)	0.11***	0.11***
	(0.03)	(0.03)
Lagged Transfer Income	0.63***	0.64***
	(0.03)	(0.04)
Constant	871.56***	1,950.42***
	(101.41)	(600.85)
Observations	7,127	1,539
R^2	0.48	0.58

Note: All columns display regressions of labor income at age 30 on the different variables listed in the rows. If the space for the coefficient appears empty, it was not included in the regression. All money figures are in 2014 USD. The number in parentheses next to the variable indicates the age of measurement. Education is measured as years of education. Robust standard errors are in parentheses below the estimates. We weigh the individuals in the auxiliary samples to match them on observed individual characteristics to the experimental sample using the procedure in Appendix C.3.3. ***: p-value $< .01$. **: p-value $< .05$. *: p-value $< .10$.

C.3.5 Non-Parametric Forecasts

An alternative to the forecasting strategy that we use throughout the main text and to generate the estimates reported in the preceding appendix is the following. For each individual (i) in the experimental sample (e), we can first find a match or a set of matches in the auxiliary sample using a modified version of Algorithm 1 that includes both baseline variables and variables affected by treatment. Dropping age indices, the modified algorithm is:

Algorithm 2 *For individual i in experimental sample ($k = e$), an individual $l(i)$ in the auxiliary sample ($k = n$) is a potential counterpart if*

$$\sqrt{(\boldsymbol{Z}_{i,e}^d - \boldsymbol{Z}_{i,l(i),n}^d)'(\boldsymbol{\Sigma}_e)^{-1}(\boldsymbol{Z}_{i,e}^d - \boldsymbol{Z}_{i,l(i),n}^d)} \leq \epsilon \tag{6}$$

where $\boldsymbol{Z}_{i,k}^d := \left[\boldsymbol{X}_{i,k}^d, \boldsymbol{B}_{i,k}\right]$ and $\boldsymbol{Z}_{i,l(i),k}^d := \left[\boldsymbol{X}_{i,l(i),k}^d, \boldsymbol{B}_{i,l(i)k}\right]$ with $k \in \{e, k\}$ $d \in \{0, 1\}$, and $\boldsymbol{\Sigma}_e$ is the covariance matrix in the experimental sample. We weight the potential counterparts according to the inverse value of the left-hand-side of Equation (6) and produce our non-parametric conditional mean functions.[79]

And, second, we can use the profiles of the individual(s) in the auxiliary samples as the profiles of the individual i in the experimental sample. This is a non-parametric, one-step strategy: instead of fitting a dynamic relationship in the non-experimental sample and using it to form out-of-sample forecasts, we construct a synthetic cohort and estimate dynamic relationships to impute labor income profiles.

Table C.12 compares the results from the approach we use throughout the main text and the non-parametric approach that we introduce in this section. We present results for labor income. We use pre- and post-treatment variables to match (using the same variables as the main approach of Appendix C.3.3). We present the discounted net present value (treatment - control) for labor income in 2014 dollars, together with the internal rates of return and benefit-cost ratios associated to these net present values. Table C.12 also presents results from other specifications that we discuss below in Appendix C.6. The approaches are in close agreement.

C.3.6 Testing Assumption A–2: Support Conditions

Assumption A–2 requires that the support of the auxiliary data contains the support of the experimental data. This can be checked for $a \leq a^*$. Figure C.1 validates this assumption

---

[79]In our empirical analysis, we set $\epsilon = 1$, but we try a range of values between 0.5 and 3 finding little sensitivity. The full set of the results we produce in the paper for multiple values in the interval [0.5,3] is available on request.

by displaying the overlapping support sets of ABC/CARE and our auxiliary data (which we restrict as we explain in Appendix C.3.2) for the variables used to interpolate and extrapolate labor income.[80]
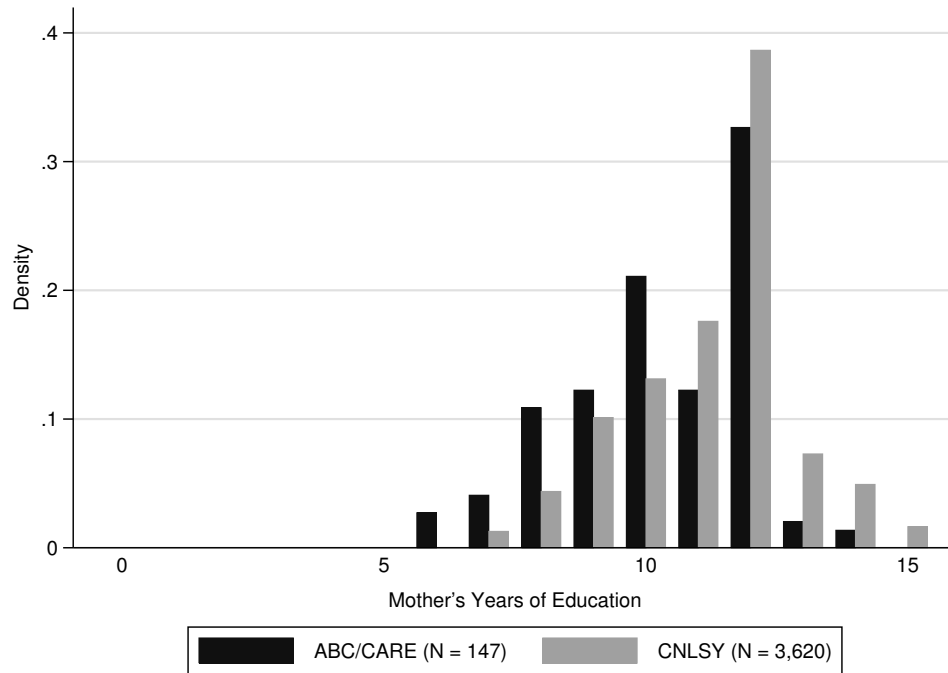
---

[80]For the male and Black indicators, we do not provide evidence of containing support. All three non-experimental samples have vast numbers of males, females, and Blacks to cover the support in the experimental samples.

Figure C.1: Support of ABC/CARE and Auxiliary Data

(a) Average PIAT Math Scores, Ages 5–7
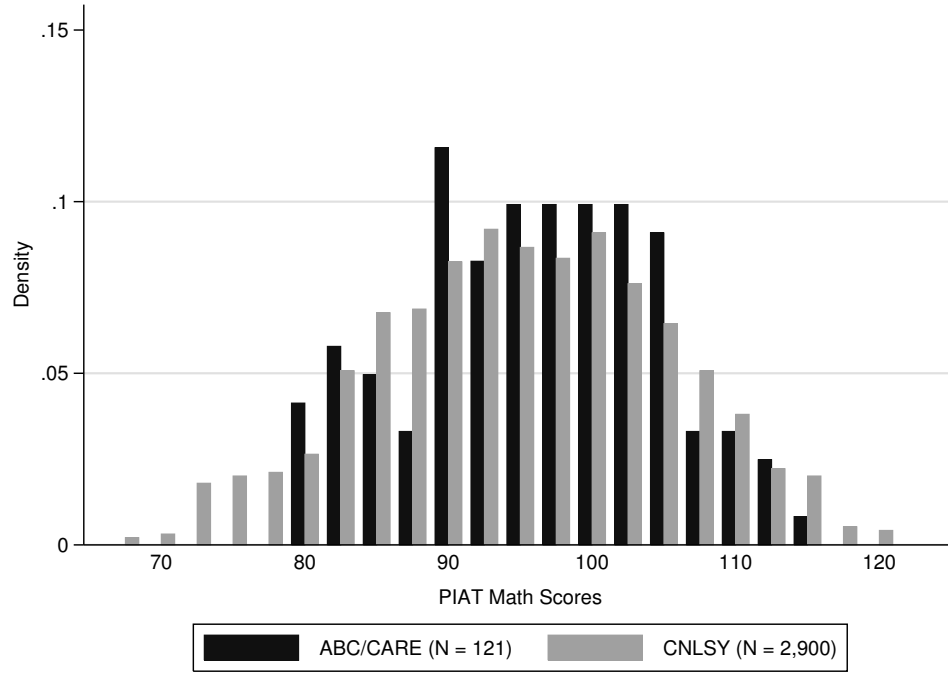
(b) Mother's Years of Education

Note: These graphs display the support of ABC, PSID, NLSY79, and CNLSY for variables we use to forecast future labor income. PIAT math scores are averaged over ages 5–7.

(c) Subject's Years of Education
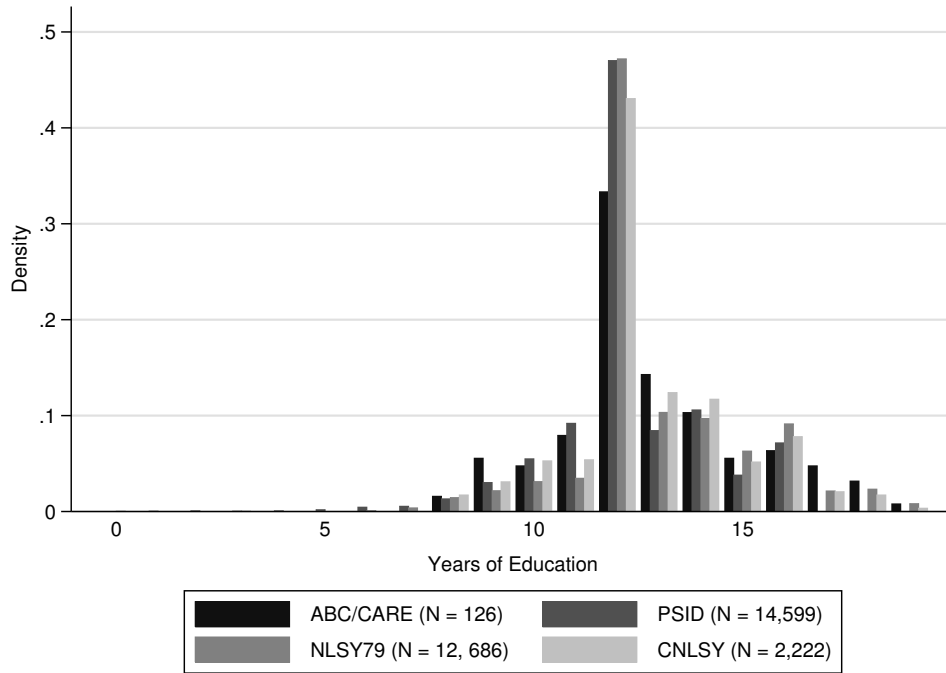
(d) Income at Age 21
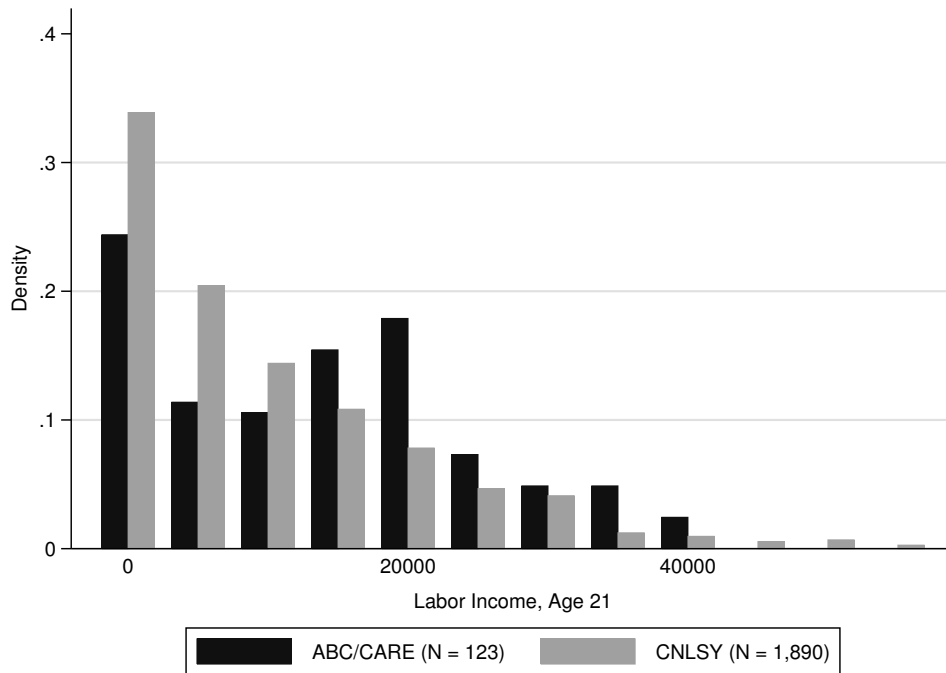
(e) Income at Age 30
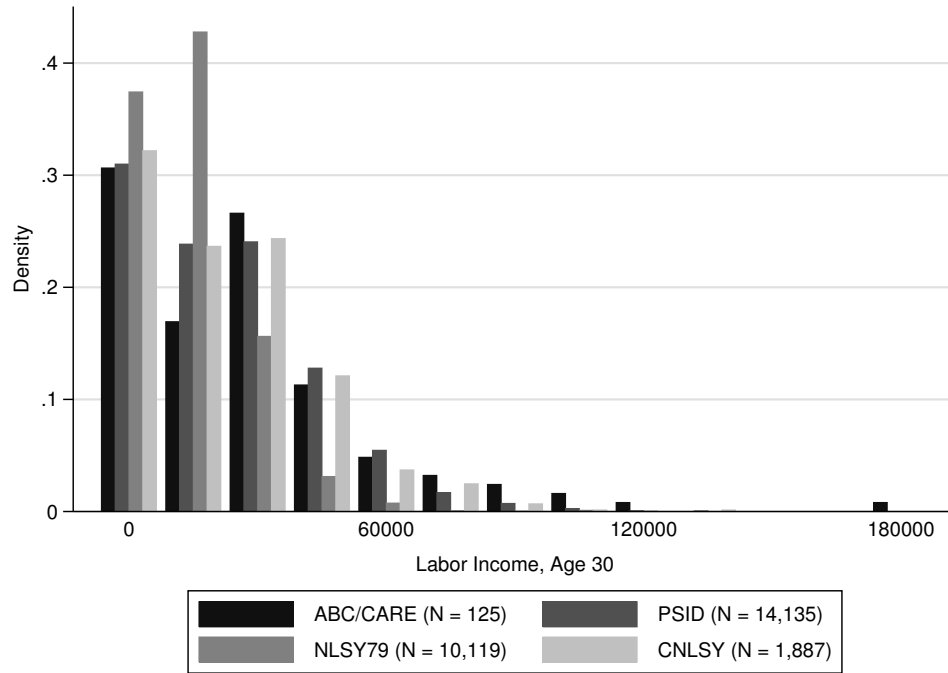
(f) Body Mass Index, Age 34
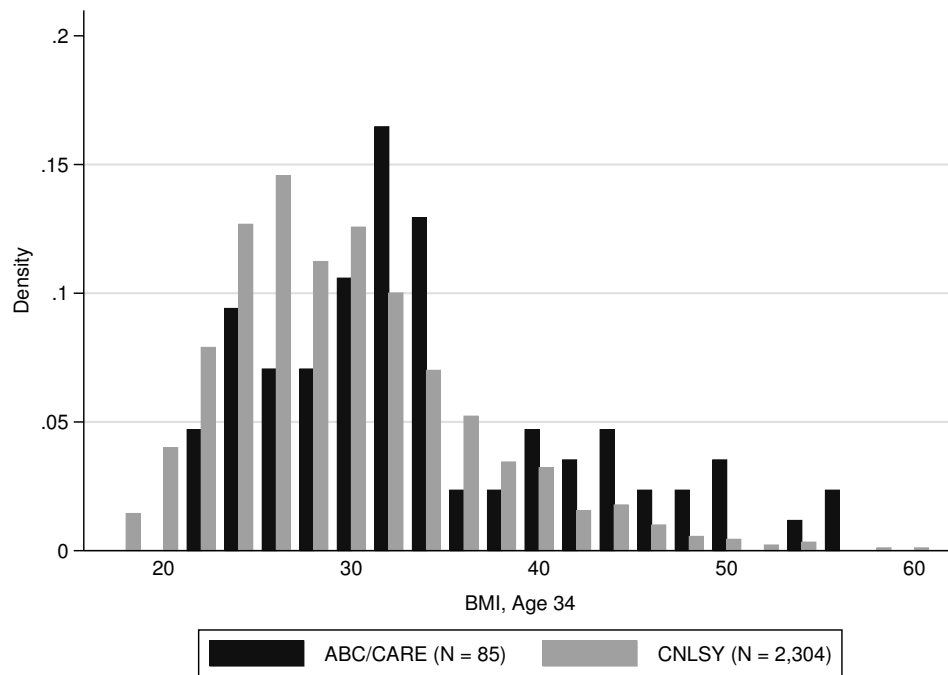
C.3.7   Testing Assumption A–3: Exogeneity

This subsection provides a framework for testing Assumption A–3. Recall from the main paper that our forecasts are based on the vector of production functions for output $\boldsymbol{Y}_{k,a}^d$:

$$\boldsymbol{Y}_{k,a}^d = \boldsymbol{\phi}_{k,a}^d(\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k) \; + \; \boldsymbol{\varepsilon}_{k,a}^d, \tag{7}$$

$d \in \{0,1\}, k \in \{e,n\}, a \in \{1,\dots,\bar{A}\}$ where $\boldsymbol{\phi}_{k,a}^d(\cdot,\cdot)$ is a vector of structural production relationships mapping inputs $\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k$ into outputs holding error term $\boldsymbol{\varepsilon}_{k,a}^d$ fixed.

For notational brevity in the definitions that follow, consider the production function for one entry (one outcome) in this vector (e.g. labor income):

$$Y_{k,a}^d = \phi_{k,a}^d(\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k) \; + \; \varepsilon_{k,a}^d, \tag{8}$$

where the variables are defined analogously to Equation (13). Note that the arguments of $\phi_{k,a}^d$ are still vectors in this case. We abuse notation to make unbolded variables generic scalars. Define an associated measurement system for the error term $\varepsilon_{k,a}^d$:

$$
\begin{aligned}
\varepsilon_{k,a}^d &= \boldsymbol{\beta}^d\boldsymbol{\theta}_a^d + \boldsymbol{\omega}_a^d && (b)\\
\boldsymbol{M}_a^d &= \boldsymbol{\lambda}^d\boldsymbol{\theta}_a^d + \boldsymbol{v}_a^d, && (c)
\end{aligned}
\tag{9}
$$

where $\boldsymbol{\theta}_a^d \perp\!\!\!\perp \boldsymbol{v}_a^d, \boldsymbol{\omega}_a^d$ and $\boldsymbol{v}_a^d \perp\!\!\!\perp \boldsymbol{\omega}_a^d$.

When the auxiliary measurement system $\boldsymbol{M}_a^d$ consists of at least three measures, we are able to identify the vectors of coefficients characterizing this system, $\boldsymbol{\lambda}^d, \boldsymbol{\beta}^d$, as well as the respective covariance matrices, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_a^d}, \boldsymbol{\Sigma}_{\boldsymbol{v}_a^d}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_a^d}$. We use the Bartlett (1938) method to obtain

an estimate of $\boldsymbol{\theta}_a^d$ (Heckman et al., 2013). Identifying and estimating the elements in system (9) enables us to use estimates of $\boldsymbol{\theta}_a^d$ as control functions when testing Assumption A–1 in the next appendix, i.e. to use these estimates to control for endogeneity.
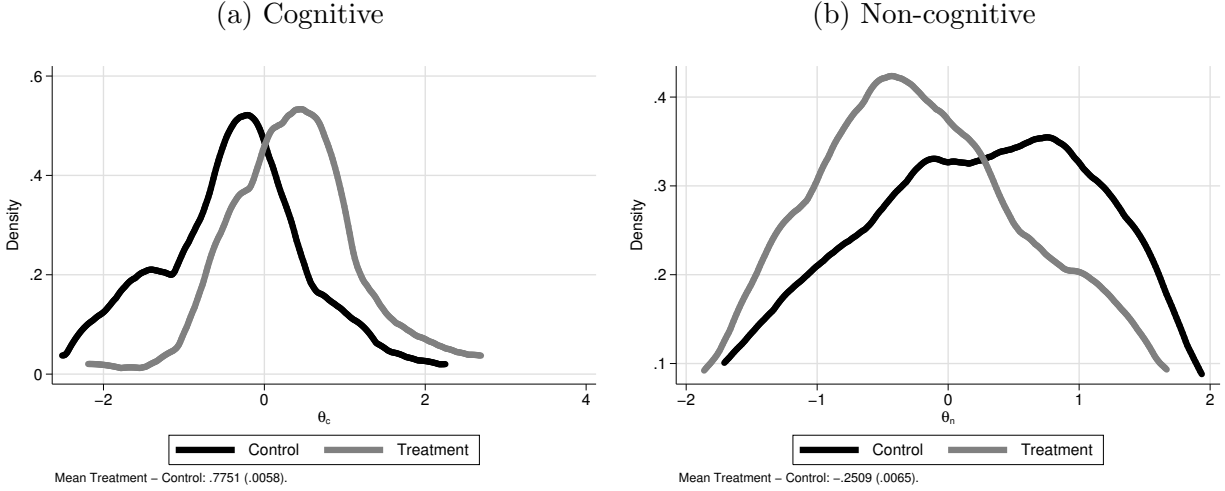
We first describe the estimates for the elements in System (9) in the experimental sample. We assume that $\boldsymbol{\theta}_a^d$ has two dimensions (one representing cognitive skill, c, and another representing non-cognitive skill, nc). We assume dedicated measures for these skills at one time period. Put simply, we have two independent systems, one to measure $\boldsymbol{\theta}_c^d$ and one to measure $\boldsymbol{\theta}_{nc}^d$, where $\boldsymbol{\theta}_a^d := \left[\boldsymbol{\theta}_c^d, \boldsymbol{\theta}_{nc}^d\right]$. Further, we assume a common measurement system for the treatment and control groups (this is a sensible assumption shown to be true in the Perry data; see Heckman et al., 2013). This assumption implies that $\boldsymbol{\lambda}^d, \boldsymbol{\beta}^d$, as well as $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_a^d}, \boldsymbol{\Sigma}_{\boldsymbol{v}_a^d}$ are the same whether $d = 0$ or $d = 1$.

At age 30 and in the experimental sample, we use a set of IQ measures from ages 2 to 8 to obtain an estimate of $\boldsymbol{\theta}_c^d$ and a set of measures of somatization, hostility, depression, and mental health all at age 21 to measure to estimate $\boldsymbol{\theta}_{nc}^d$. Figure C.2 shows our estimates by treatment status.

At age 30 and in the auxiliary sample, we can also estimate $\boldsymbol{\theta}_a^d$. For want of data to approximate $\boldsymbol{\theta}_c, \boldsymbol{\theta}_{nc}$ in PSID and NLSY79, we use the CNLSY in this appendix. Our measurement system for $\boldsymbol{\theta}_c$ consists of reading and comprehension PIAT scores as well as by the Peabody Picture Vocabulary Test (PPVT). Our measurement system for $\boldsymbol{\theta}_{nc}$ is based on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior).

Once these estimates are available, we can test Assumption A–3 in the experimental and non-experimental samples. The test consists of the following. Let $\boldsymbol{\gamma}^E$ be the parameter associated to $\boldsymbol{X}_{k,a}^d$ in Equation (14) in a linear regression when not accounting for $\boldsymbol{\theta}_a^d$. Similarly, let $\boldsymbol{\gamma}^I$ be the parameter associated with $\boldsymbol{X}_{k,a}^d$ in Equation (14) when accounting for $\boldsymbol{\theta}_a^d$. Under the null hypothesis that Assumption A–3 holds $\boldsymbol{\gamma}^E = \boldsymbol{\gamma}^I$ and $\boldsymbol{\theta}_a^d$ is an irrelevant predictor

Figure C.2: Estimates of Cognitive (θ_c^d) and Non-cognitive Skills (θ_{nc}^d)

(a) Cognitive

(b) Non-cognitive

Mean Treatment – Control: .7751 (.0058).

Mean Treatment – Control: –.2509 (.0065).

Note: Panel (a) displays a factor score estimated based on the measurement system in (9) and measures of IQ at ages 2, 3, 4, 5, 7, and 8 (cognitive skill). Panel (b) displays an analogous set of graphs for measures of somatization, hostility, depression, and mental health at age 21 (non-cognitive skill). Both measures of skills are standardized to a mean of 0 and a standard deviation of 1. "Less" in the factor measuring non-cognitive skills is "positive" given the measures we rely on to construct it. The mean difference between treatment and control is displayed below each panel, with standard error in parentheses. Standard errors are based on the empirical bootstrap distribution.

in Equation (14). We test the null hypothesis by asking if the elements in $\boldsymbol{\theta}_a^d$ are relevant predictors of labor and transfer income. We perform these tests in the experimental sample pooling the treatment and control groups (i.e., under the invariance assumption that we document in Section 3) and in the non-experimental sample using our synthetic cohort. We contrast specifications with and without including estimates of $\boldsymbol{\theta}_a^d$, and report the F-statistic corresponding to this comparison. This is a version of a Durbin-Wu-Hausman test (see Durbin, 1954; Wu, 1973; Hausman, 1978). Tables C.5 to C.8 present the results from these tests. In most cases, we are not able to reject the null hypothesis that Assumption A–3 holds.

Table C.5: Forecast of Labor Income at Age 30 Accounting for B_k and $\boldsymbol{\theta}$, $\boldsymbol{X}_{k,a}$, ABC/CARE Control and Treatment Groups

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother's Education	2,668.48	0.12	2,200.35	0.25	794.11	0.36	1,724.88	0.31
PIAT (5-7)	.	.	.	.	-126.19	0.67	-400.57	0.72
Education (30)	.	.	.	.	8,601.33	0.00	9,706.02	0.00
Labor Income (21)	.	.	.	.	0.14	0.37	0.21	0.37
Cognitive	.	.	4,260.39	0.16	.	.	1,427.18	0.44
Non Cognitive	.	.	2,899.66	0.25	.	.	7,557.01	0.05
Constant	4,443.37	0.41	9,166.30	0.38	-78,053.28	0.95	-75,621.84	0.87
F-stat	2.50		1.90		5.87		5.37	
p-value	0.29		0.31		0.29		0.01	
R^2	0.02		0.04		0.20		0.25	
Observations	132		100		130		133	
F-stat: exclude Cognitive, Non-Cognitive			2.07				2.92	
p-value			0.31				0.19	

F-stat: exclude Cognitive and Non-Cognitive: F-statistic contrasting the specifications in columns (1) and (3) and (5) and (7), respectively.

Note: Forecast of labor income at age 30 is based on the variable listed in the row. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value for the treatment and control groups as well as a test for the treatment-control difference. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (9) and measures of $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (9) and measures of IQ at ages 2, 3, 4, 5, 7, and 8 (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (9) and measures of somatization, hostility, depression, and a global mental health index at age 21 (non-cognitive skill). Both measures of skills are standardized to a mean of 0 and a standard deviation of 1. Inference is based on the empirical bootstrap distribution. If the estimates for the constant terms are in the ten or hundred thousands, we report a figure that has been rounded to the thousands.

Table C.6: Forecast of Labor Income at Age 30 Accounting for B_k and $\boldsymbol{\theta}$, $\mathbf{X}_{k,a}$, CNLSY

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother's Education	2,292.54	0.00	1,528.20	0.00	117.79	0.25	-47.06	0.50
PIAT (5-7)					262.38	0.00	447.61	0.00
Education (30)					3,722.75	0.00	4,202.69	0.00
Labor Income (21)					0.62	0.00	0.82	0.00
Cognitive			2,859.63	0.00			-4,149.19	0.88
Non Cognitive			-2,921.97	1.00			-590.26	0.75
Constant	2,840.27	0.00	11,377.06	0.00	-53,962.05	1.00	-78,072.63	1.00
F-stat	46.92		4.89		83.55		18.31	
p-value	0.00		0.04		0.00		0.00	
R^2	0.03		0.04		0.19		0.33	
Observations	1,862		350		1,860		1,862	
F-stat: exclude Cognitive, Non-Cognitive			4.18				1.77	
p-value			0.04				0.34	

F-stat: exclude Cognitive and Non-Cognitive: F-statistic contrasting the specifications in columns (1) and (3) and (5) and (7), respectively.

Note: Forecast of labor income at age 30 is based on the variable listed in the row. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value for the treatment and control groups as well as a test for the treatment-control difference. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (9) and measures of reading and comprehension of the PIAT, as weel as the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (9) and six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior) (non-cognitive skill). Both measures of skills are standardized to a mean of 0 and a standard deviation of 1. Inference is based on the empirical bootstrap distribution. If the estimates for the constant terms are in the ten or hundred thousands, we report a figure that has been rounded to the thousands.

Table C.7: Forecast of Transfer Income at Age 30 Accounting for B_k and θ, $X_{k,a}$, ABC/CARE Control and Treatment Groups

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother's Education	-299.72	0.81	-411.12	0.85	-135.23	0.62	-211.76	0.68
PIAT (5-7)	.	.	.	.	-34.90	0.81	-66.99	0.80
Education (30)	.	.	.	.	-430.88	0.96	-453.82	0.96
Labor Income (21)	.	.	.	.	-0.09	1.00	-0.08	0.96
Cognitive	.	.	-753.98	0.93	.	.	153.54	0.42
Non Cognitive	.	.	631.74	0.17	.	.	264.49	0.34
Constant	5,135.83	0.06	6,460.15	0.07	13,548.68	0.03	17,791.02	0.05
F-stat	1.78		3.04		3.86		2.75	
p-value	0.35		0.12		0.35		0.09	
R^2	0.02		0.10		0.15		0.18	
Observations	133		101		135		133	
F-stat: exclude Cognitive, Non-Cognitive			3.38				1.23	
p-value			0.17				0.44	

F-stat: exclude Cognitive and Non-Cognitive: F-statistic contrasting the specifications in columns (1) and (3) and (5) and (7), respectively.

Note: Forecast of labor income at age 30 is based on the variable listed in the row. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value for the treatment and control groups as well as a test for the treatment-control difference. $\hat{\theta}_c$: factor score estimated based on the measurement system in (9) and measures of IQ at ages 2, 3, 4, 5, 7, and 8 (cognitive skill). $\hat{\theta}_n$: factor score estimated based on the measurement system in (9) and measures of somatization, hostility, depression, and a global mental health index at age 21 (non-cognitive skill). Both measures of skills are standardized to a mean of 0 and a standard deviation of 1. Inference is based on the empirical bootstrap distribution. If the estimates for the constant terms are in the ten or hundred thousands, we report a figure that has been rounded to the thousands.
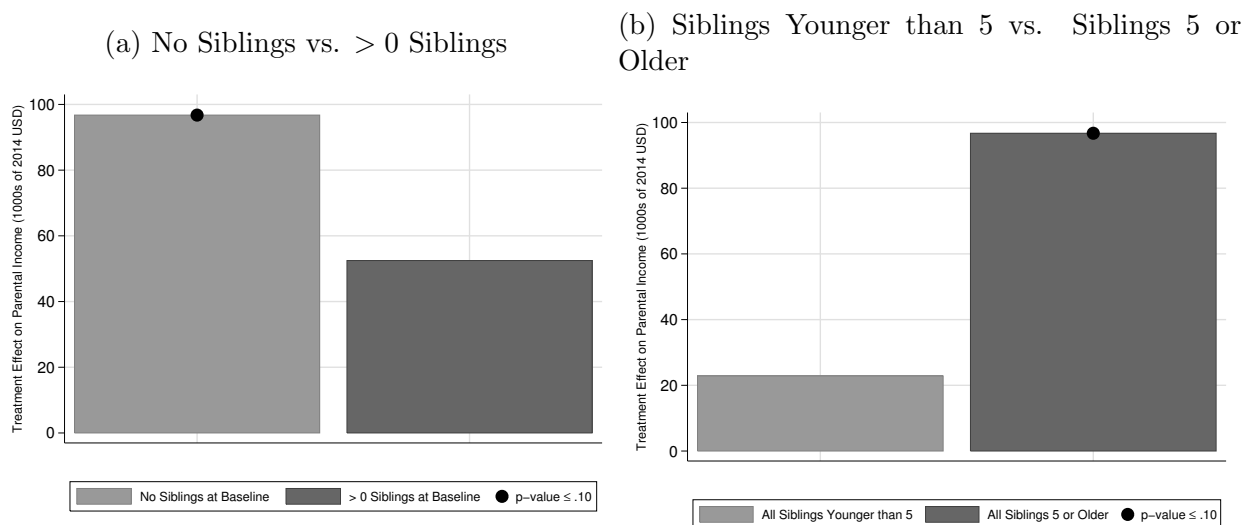
89

Table C.8: Forecast of Transfer Income at Age 30 Accounting for B_k and $\boldsymbol{\theta}$, $\boldsymbol{X}_{k,a}$, CNLSY

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother'sEducation	366.18	0.38	10,450.96	0.00	2,337.63	0.12	10,634.87	0.00
PIAT(5-7)					-872.86	0.88	-364.63	0.50
Education(30)					-8,126.93	1.00	-6,206.13	0.88
LaborIncome(21)					0.79	0.25	-0.99	1.00
Cognitive			-9,680.93	0.88			-5,092.70	0.50
NonCognitive			18,373.65	0.03			6,585.57	0.22
Constant			20,921.34	0.00			9,015.29	0.12
F-stat	0.14		1.80		1.06		1.17	
p-value	.75		0.18		0.49		0.37	
R^2	0.00		0.02		0.01		0.06	
Observations	1,101		239		1,100		1,099	
F-stat: exclude Cognitive, Non-Cognitive		1.70				0.71		
p-value		0.26				0.52		

F-stat: exclude Cognitive and Non-Cognitive: F-statistic contrasting the specifications in columns (1) and (3) and (5) and (7), respectively.
Note: Forecast of labor income at age 30 is based on the variable listed in the row. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value for the treatment and control groups as well as a test for the treatment-control difference. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (9) and measures of reading and comprehension of the PIAT, as well as the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (9) and six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior) (non-cognitive skill). Both measures of skills are standardized to a mean of 0 and a standard deviation of 1. Inference is based on the empirical bootstrap distribution. If the estimates for the constant terms are in the ten or hundred thousands, we report a figure that has been rounded to the thousands.

C.3.8 Parental Labor Income

A substantial portion of the ABC/CARE benefits come from induced increases in parental labor income. The program operated as a childcare center, as well as a child development center. The parents were usually mothers: only 27% of the mothers lived with a partner at baseline and there is little change in this status after enrollment. ABC/CARE relaxed the time constraint of the mothers, enabling them to educate themselves more and/or work more. The program had treatment effects on (i) maternal education; (ii) maternal labor supply; and (iii) maternal income. Quantifying the effect of ABC/CARE on parental labor income requires quantifying its effects beyond age 5, after the subjects entered school. The program could have shifted the labor income profile through education or work experience. To test this, we need to quantify the effect that the program had from when it started until the mothers retired.

There are three options for monetizing parental labor income. (i) A conservative approach that we follow in the main text uses the available information and calculates the treatment effect of the program on labor income from ages 0 to 21. We observe parental labor income at ages 0, 1.5, 3.5, 4.5, 8, 12, 15, and 21. We interpolate between the ages that we observe and stop at age 21. The average age of the mothers at baseline was 21. On average, then, we omit 19 years of labor income if the mothers decide to retire at 60, 24 if they retire at 67, etc. (ii) A second approach is to use the available data together with an auxiliary sample to parameterize parental labor income when the subjects are older than 21. (iii) A third approach is to follow a similar methodology as the one we use for the subjects' labor and transfer income, making some adaptations given data limitations. Option (i) is straightforward. We explain options (ii) and (iii) next.

We note that the childcare feature of the program likely benefits parents who, at baseline, did not have any other children. If they did have other children, parents would still have to take

care of the other children, weakening the childcare-driven effect on labor income (especially if there are younger siblings present). Figure C.3 shows that the net increase (treatment minus control) in discounted parental labor income is much higher in the absence of siblings (of the participant children) at baseline, using the "conservative approach" described above. The effect also weakens when comparing the outcomes of mothers whose children have siblings younger than 5 years old to the outcomes of mothers of children who have siblings 5 years old or younger.[81]

Figure C.3: Discounted Net Present Value of Parental Labor Income by Participant's Number and Age of Siblings at Baseline

<div style="display:flex">
<div>

(a) No Siblings vs. > 0 Siblings

</div>
<div>

(b) Siblings Younger than 5 vs. Siblings 5 or Older

</div>
</div>

Note: Panel (a) displays the net-present value (treatment less control) of parental labor income of parents of children with and without siblings at baseline. Panel (b) displays the average parental labor income of parents of children with young siblings (younger than 5 years old) and children with older siblings (5 years old or older) at baseline. Panel (b) drops children without siblings at baseline. Parental income is in 2014 USD discounted to child's participant age 0 using a 3% rate. We use the baseline "conservative" measure of parental labor income in Section 4.3. Results using our alternative parental labor income measures are similar (see Appendix C.3.8).

---

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

C.3.8.1 Using Mincer Equations to Forecast Parental Labor Income

This approach fits Mincer regressions for parental labor income.[82] We specify how we deal with the presence of a spouse below. The parameterization used is as follows:

$$\ln Y_a = \alpha + \beta \cdot \text{school}_a + \gamma_1 \cdot \text{experience}_a + \gamma_2 \cdot \text{experience}_a{}^2 + \boldsymbol{\psi} \mathbf{X}_a + \eta_a, \qquad (10)$$

where variables are indexed by mother's age, $\ln Y_a$ is log-labor income at age a, experience and schooling are measured in years, $\mathbf{X}_a$ are observed characteristics, and η_a is an unobserved component. $\alpha, \beta, \gamma_1, \gamma_2, \boldsymbol{\psi}$ are parameters of the labor income equation.

The parameters characterizing the profile do not differ across the treatment and control groups in ABC/CARE. This assumption is analogous to Assumption A–1.

We estimate the coefficients in (10) using the sample of mothers in ABC/CARE. We pool the longitudinal information and estimate the coefficients using ordinary least squares. We use a standard Mincer measure of experience (age − education − 6). We assign one dollar to mothers with no labor income. For mothers living with a working partner, we allocate 1/2 of total parental labor income as Y_a. To validate our estimates within ABC/CARE, we estimate the coefficients (10) using a sub-sample of disadvantaged mothers in the PSID.[83] The coefficients characterizing (10) in ABC/CARE and PSID for different combinations of control sets are in close agreement. We display them in Table C.9.

---

[82]See Mincer (1974) for the original source. Heckman et al. (2006) discuss the Mincer equation.

[83]We define disadvantaged as follows: Black, not married, labor income, education (at age 5 of child's participant), age and number of children (at age 5 of child's participant) in the same ranges as the ABC/CARE mothers, labor income below the 75th percentile in the PSID sample.

Table C.9: Mincer Equation Estimates for Mothers in ABC/CARE and the PSID

	PSID	ABC/CARE	PSID	ABC/CARE	PSID	ABC/CARE
Education	0.0762***	0.0614***	0.1155***	0.1000***	0.1109***	0.0852***
	(0.0050)	(0.0161)	(0.0057)	(0.0151)	(0.0057)	(0.0184)
Experience			0.0386***	0.0908***	0.0417***	0.0861***
			(0.0027)	(0.0086)	(0.0027)	(0.0085)
Experience^2			-0.0005***	-0.0018***	-0.0007***	-0.0015***
			(0.0001)	(0.0003)	(0.0001)	(0.0003)
Birth Year					-0.0041***	0.0104
					(0.0008)	(0.0084)
Children					-0.0803***	-0.0533
					(0.0068)	(0.0372)
Constant	8.2789***	8.8869***	7.3572***	7.8229***	15.6408***	-12.3540
	(0.0609)	(0.1892)	(0.0780)	(0.1895)	(1.5533)	(16.6026)
Observations	15,506	705	15,506	705	15,506	664
R^2	0.0145	0.0194	0.0416	0.2215	0.0514	0.2047

Note: This table presents estimates of (10) for ABC/CARE mothers and a subsample of disadvantaged mothers in the PSID. We define disadvantaged as follows: Black, not married, labor income, education (at age 5 of child's participant), age and number of children (at age 5 of child's participant) in the same ranges as the ABC/CARE mothers, labor income below the 75th percentile. Robust standard errors are in parentheses. p-value $< .01$. **: p-value $< .05$. *: p-value $< .10$.

Based on the estimates in Table C.9, we can ask two questions: (i) what is the predicted net present value of parental labor income (treatment - control) using a forecast based on the estimate of (10) and how does it differ from the method that linearly interpolates income from child's age 0 to 21?; and (ii) what would be the predicted net present value of parental labor income if we go beyond the child's age 21 data and forecast all the way up to 40 years of experience?

Table C.10 display results that answer these two questions. Precise estimates for (10) are obtained. From it we can measure (treatment - control) when the subjects are 21 years old. When using these same equations to forecast parental labor income such that mothers work for 40 years in their life times, we find that we add $30,000$ (2014 USD) to the estimate reported in the main paper.

Table C.10: Parental Labor Income, Interpolations and Prediction

	Males and Females	Male	Female
Interpolated up to Age 21	82,287	65,477	96,251
	(22,981.46)	(26,603.57)	(32,000.64)
Mincer-based up to Age 21	75,114	72,030	78,198
	(428.340)	(647.017)	(557.716)
Mincer-based up to Retirement	106,957	102,556	111,338
	(609.870)	(921.222)	(794.076)

Note: Interpolated up to Age 21: linearly interpolated parental labor income from (child's) age 0 to 21. Mincer-based up to Age 21: prediction from (child's) age 0 to 21 based on estimates coefficients of (10) (full control set). Mincer-based up to Retirement: forecast from (child's) age 0 to mother's retirement (40 years of labor force participation assumed) based on estimates coefficients of (10) (full control set). All values are in 2014 USD discounted to child's age 0. Standard errors in parentheses are based on the empirical bootstrap distribution.

C.3.8.2 Life-cycle Forecasts of Parental Labor Income

A third approach assumes that all parental labor income is earned by the mother and limits the non-experimental samples to Black females whose labor income at each age is below the in-sample 90th percentile (we calculate the percentiles for the PSID and NSLY79 separately and place the initial restrictions separately before pooling them as one sample). As for the labor incomes of program participants after age 30, we pool data from the PSID and NLSY79. For lack of data on other relevant predictors, we use only one predictor: lagged parental labor income. We initialize the forecast with the last observed measure of maternal income in the experiment and extrapolate until the mother is 67 years old. Figure C.4 displays our estimate of parental labor income in a format similar to that of Figure C.3.

Summary: Forecasts of Parental Labor Income

Any forecast of parental labor income starts at the last observation of parental labor income, which varies by individual due to sample attrition. We assume that parental labor income is equal to maternal income (only 27% of mothers in ABC/CARE report living in a two-parent home).

Figure C.4: Discounted Net-present Value of Parental Labor Income by Participant's Number and Age of Siblings at Baseline

(a) No Siblings vs. > 0 Siblings

(b) Siblings Younger than 5 vs. Siblings 5 or Older

Note: Panel (a) displays the net-present value (treatment less control) of parental labor income of parents of children with and without siblings at baseline. Panel (b) displays the average parental labor income of parents of children with young siblings (younger than 5 years old) and children with older siblings (5 years old or older) at baseline. Panel (b) drops children without siblings at baseline. Parental income is in 2014 USD discounted to child's participant age 0 using a 3% rate. We use the measure of parental labor income in Section 4.3.

1. **Mincer Model**

   (a) **Auxiliary Sample Used to Forecast:** PSID.

   (b) **Initial Restrictions Placed on the Auxiliary Sample:** Black; female; un-married; education and number of children at ages 5 in the ranges of ABC/CARE participants. Labor income at each age is below the 75th percentile.

   (c) **Variables Used to Construct Synthetic Cohorts** we pool PSID/NLSY79 restricted samples.

   (d) **Variables Used to Forecast:** education, second order polynomial in experience, birth year, number of children.

   (e) **Assumed Retirement:** after 40 years of labor force participation.

2. **Life-cycle Forecast**

(a) **Auxiliary Sample Used to Forecast:** PSID and NLSY79.

(b) **Initial Restrictions Placed on the Auxilliary Sample:** Black; female; labor income at each age is below the 90th percentile.

(c) **Variables Used to Construct Synthetic Cohorts:** we pool the PSID restricted sample, and do not construct synthetic experimental groups due to lack of data on the treatment effect predictors.

(d) **Variables Used to Forecast:** lagged labor income.

(e) **Assumed Retirement:** 67 years old.

C.4   Internal Rate of Return

To estimate the internal rate of return, we solve for ρ in the following equation:

$$\sum_{a=0}^{A} \frac{\mathbb{E}(B_a - C_a)}{(1+\rho)^a} = 0, \tag{11}$$

where we let $A = 79$. In this subsection we define B_a and C_a to be the (discounted) total benefits and costs of the program at age a, and define $\mathbb{E}(.)$ to be the sample mean.[84] That is, we estimate the internal rate of return for the *average subject* of ABC/CARE.

All outcomes of the parents and subjects affected by the program are treated as benefits. For this to make sense, we reverse the sign of the monetized effect of the program on specific outcomes. Costs of ABC/CARE consist only of the initial program costs from ages 0 to 5. Table C.11 provides a full list of the benefits and costs of ABC.

We take the sum of the treatment effects on each component of the benefits to be the total

---

[84]Note that in this subsection B is not the conditioning set for eligibility.

benefit, B_a, of the ABC/CARE program. This includes parental labor income, subject labor income, and QALYs (quality-adjusted life years). Treatment effects on costs borne by the subject or society have their signs reversed and are included as benefits. We do this for subject public-transfer income, education costs, crime costs, control substitution costs, and health costs. To account for deadweight loss, we impose a marginal welfare cost of 50% by multiplying public costs by a factor of 0.5 when they are a direct transfer from the government to the individuals.[85] When the public costs are not a direct transfer from the government to the individuals, we multiply them by a factor of 1.5.

The principle for multiplying the public costs is the following. We evaluate the social benefits of ABC/CARE and do not place a value on who receives the money. The only social cost from a direct transfer is the dead-weight loss that it generates: 50% of its total value. We do not consider education and criminal costs to be a direct transfer. Thus, we multiply them by a factor of 1.5: the value of their cost plus 50% of the value of their cost (the dead-weight loss implied in raising the public revenue to fund them). Table C.11 lists the factor we use to multiply each cost to account for its implied dead-weight loss.

After constructing our cash flow, $\mathbb{E}(B_a - C_a)$, solving for ρ reduces to an algebraic exercise. The expected life-cycle profile of net benefits need to satisfy a "single crossing property" in order to obtain a unique solution for the internal rate of return.[86] The single crossing property holds when the benefits do not go from positive to negative across the life cycle. When the single crossing property is not satisfied, the internal rate of return is not a valid summary for the efficiency of an investment. To calculate the internal rate of return, we estimate the treatment effect on each component of the benefits and costs at age a for the pooled, male, and female samples. We do this for 100 bootstrap resamples of the original

---

[85]There is no clear consensus on the marginal welfare cost of tax revenue. However, most researchers estimate the welfare cost per tax dollar to be between $0.30 and $0.50. See Feldstein (1999), Heckman and Smith (1998), and Browning (1987).

[86]See Arrow and Levhari (1969) for a formal discussion, although the discussion on multiplicity, sign, and real or complex nature of the roots of a polynomial traces back to Descartes' Rule of signs.

ABC/CARE data. In the case of health costs and subject income, for which we employ auxiliary datasets to estimate the treatment effects, we also obtain 100 bootstrap estimates from the auxiliary data for every ABC/CARE bootstrap resample, resulting in a total of 100,000 estimates. By reusing each bootstrap estimate of the treatment effect on outcomes that do not require any auxiliary data set 100 times, we obtain a total of 100,000 estimates of the cash flow. We estimate the internal rate of return on each of those cash flows, and discard those for which we find a negative internal rate of return. The remaining estimates form our empirical bootstrap distribution of the internal rate of return for the pooled, male, and female samples. We take the mean of the distributions to be the point estimates, and we take the sample standard deviations to be the standard errors. To construct the 80% confidence intervals, we take the 10th and 90th percentiles of each bootstrap distribution.

Figure C.5 reports the distributions of the internal rates of return, by gender and for each of the three parameters that we consider (treatment vs. next best, treatment staying at home, treatment vs. alternative preschool). For some parameters and genders, we discard a high percentage of the internal rate of return of the outcomes. We next discuss how we calculate the benefit/cost ratios, noting that this statistic is not subject the same caveat as the internal rate of return: we can summarize the efficiency of the investment even in the absence of the single-crossing property.

Figure C.5: Internal Rate of Return by Gender
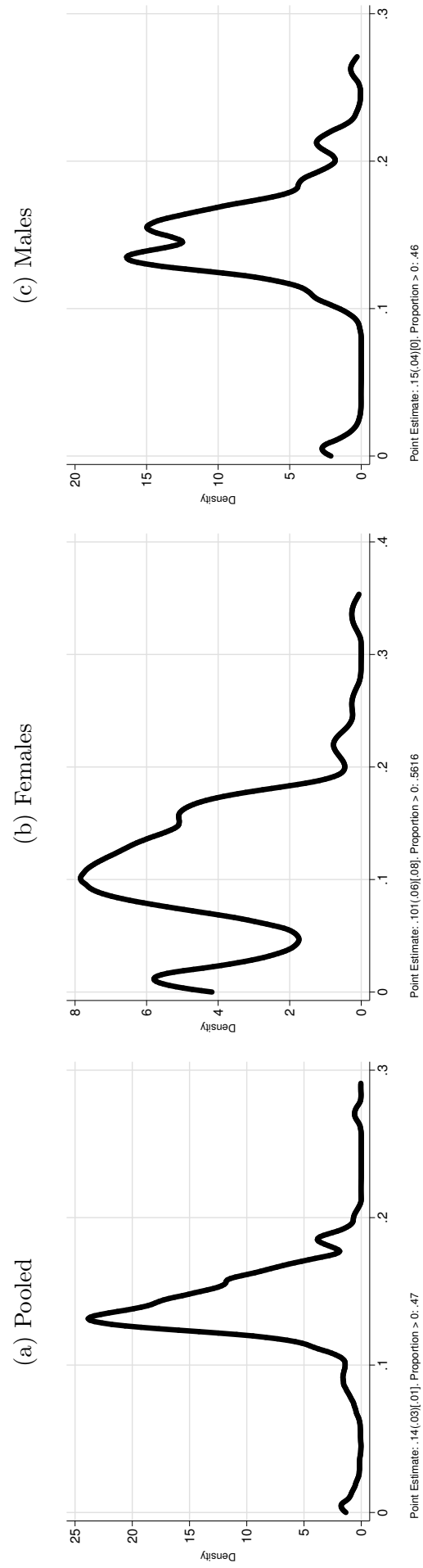
(a) Pooled

Point Estimate: .14.(.03)[.01]. Proportion > 0: .47

(b) Females

Point Estimate: .101.(.06)[.08]. Proportion > 0: .5616

(c) Males

Point Estimate: .15.(.04)[0]. Proportion > 0: .46

Note: Panel (a) displays the empirical bootstrap distribution for the estimate of the treatment vs. next best parameter in the pooled sample. The remaining panels show an analogous distribution by gender. We discard negative internal rates of returns. Each panel displays the point estimate with the standard error in parentheses, the p-value in brackets, and the percentage of positive internal rates of return out of the initial set of bootstrap resamples.

100

C.5 Computing the Benefit/Cost Ratio

The benefit/cost ratio is

$$
\mathbb{E}\left(\frac{\sum_{a=0}^{A} B_a}{\sum_{a=0}^{A} C_a}\right),
\tag{12}
$$

where we let $A = 79$, define B_a and C_a to be the benefits and costs of the program at age a, and define $\mathbb{E}(.)$ to be the sample mean. See Table C.11 for a detailed list of the components to the benefits and costs of ABC/CARE . We take the sum of the treatment effects on each component of the benefits to be the total benefits of the ABC/CARE programs.

To account for deadweight loss, we assume a marginal welfare cost of 50% by multiplying public costs components by a factor of 1.5. For the same reason, we multiply public-transfer income by a factor of 0.5. We discount each component of the benefits and costs by 3% every year to obtain their net present value at birth. We then sum up the discounted components of the benefits and find the ratio with the discounted costs.

We estimate the treatment effect for each component of the benefits and costs at age a for the pooled, male, and female samples. We do this for 100 bootstrap resamples of the original ABC/CARE data. In the case of health and subject income, for which we employ auxiliary datasets to estimate the treatment effects, we also obtain 100 bootstrap estimates from the auxiliary data for every ABC/CARE bootstrap resample, resulting in a total of 100,000 estimates. By reusing each bootstrap estimate of the treatment effect on outcomes that do not require any auxiliary data set 100 times, we obtain a total of 100,000 estimates of the costs stream and benefits stream. We estimate the benefit/cost ratio for each of those streams. This is how we form our empirical bootstrap distribution of the benefit/cost ratio for

the pooled, male, and female samples. We take the mean of the distributions to be the point estimates, and we take the standard deviations to be the standard errors. To construct the 80% confidence intervals, we take the 10th and 90th percentiles of each bootstrap distribution. Figure C.6 presents the empirical distribution of the empirical bootstrap distribution, per parameter of interest and gender.

Table C.11: Components of Benefits and Costs

Variable	Sign Reversed	Welfare Cost Factor
Benefits		
Parent Income		
Subject QALY		
Subject Labor Income		
Subject Public-transfer Income	✓	0.5
Medicare Costs	✓	1.5
Public Medicaid Costs	✓	1.5
Private Medical Costs	✓	
Miscellaneous Medical Costs	✓	
Disability Insurance Claim	✓	0.5
Social Security Claim	✓	0.5
Supplemental Security Claim	✓	0.5
Control Substitution Costs	✓	
Education Costs	✓	1.5*
Justice System Costs	✓	1.5
Prison Costs	✓	1.5
Victimization Costs	✓	
Costs		
Program Costs		

Note: The table lists the components of the costs and benefits of ABC/CARE. In order for some components to be categorized as benefits, we reversed the sign of the treatment effect. Only education costs up until age 18 are multiplied by 1.5 to account for welfare costs. This factor is drawn from Heckman et al. (2010).

Figure C.6: Benefit/Cost Ratios by Gender
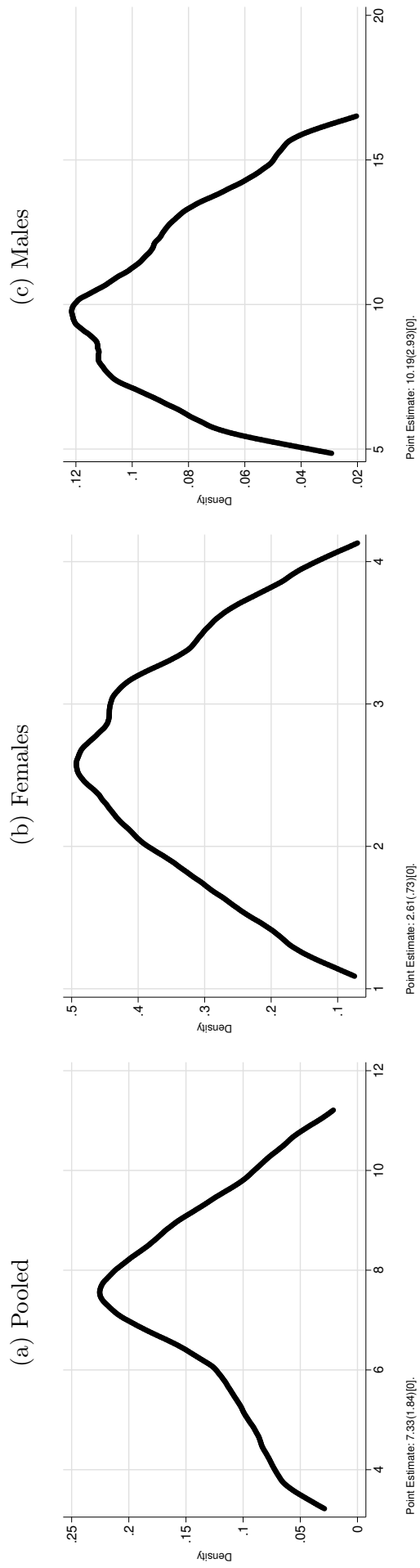
(a) Pooled

(b) Females

(c) Males

Note: Panel (a) displays the empirical bootstrap distribution for the estimate of the treatment *vs.* next best parameter in the pooled sample. The remaining panels show an analogous distribution by gender. Each panel displays the point estimate with the standard error in parentheses and the *p*-value in brackets.

C.6 Estimates from Other Forecasting Models: A Sensitivity Study

Recall from the main paper that our forecasts are based on the vector of production functions for output $\boldsymbol{Y}_{k,a}^d$:

$$\boldsymbol{Y}_{k,a}^d = \boldsymbol{\phi}_{k,a}^d(\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k) + \boldsymbol{\varepsilon}_{k,a}^d, \tag{13}$$

$d \in \{0,1\}, k \in \{e,n\}, a \in \{1,\ldots,\bar{A}\}$ where $\boldsymbol{\phi}_{k,a}^d(\cdot,\cdot)$ is a vector of structural production relationships mapping inputs $\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k$ into outputs holding error term $\boldsymbol{\varepsilon}_{k,a}^d$ fixed.

For notational brevity in the definitions that follow, consider the production function for one entry (one outcome) in this vector (e.g. labor income):

$$Y_{k,a}^d = \phi_{k,a}^d(\boldsymbol{X}_{k,a}^d, \boldsymbol{B}_k) + \varepsilon_{k,a}^d, \tag{14}$$

where the variables are defined analogously to Equation (13). Note that the arguments of $\phi_{k,a}^d$ are still vectors.

In this appendix, we explore a range of standard panel data specifications for the error terms:

$$
\begin{aligned}
\varepsilon_{k,a}^d &= f^d + \omega_{k,a}^d \\
\omega_{k,a}^d &= \rho\omega_{k,a-1}^d + U_{k,a}^d,
\end{aligned}
\tag{15}
$$

where $U_{k,a}^d \perp\!\!\!\perp \boldsymbol{X}_{k,a}^d$.

We present different structures for $\phi_{k,a}^d(\cdot,\cdot)$ and $\varepsilon_{k,a}^d$ and investigate the robustness of our estimates to different assumptions about the structure of both these elements. We do this exercise for labor income.

The system of interest is:

$$
\begin{aligned}
Y_{k,a} &= \lambda_0 + \lambda_1 Y_{k,a-1} + \lambda_2 \boldsymbol{X}_{k,a} + \varepsilon_{k,a} \\
\varepsilon_{k,a} &= \underbrace{f}_{\text{Fixed Effect}} + \underbrace{\omega_{k,a}}_{\text{Potentially Serially Correlated Component}} \\
\omega_{k,a} &= \rho \omega_{k,a-1} + \underbrace{U_{k,a}}_{\text{Independent Innovation}},
\end{aligned}
\tag{16}
$$

where $U_{k,a} \perp\!\!\!\perp \boldsymbol{X}_{k,a}$ and we drop baseline variables and d, k indices for brevity (the latter is justified by our documentation of invariance in Section 3.

Table C.12 summarizes the results from our exploration through two statistics: (i) the predicted net present value of labor income under different assumptions; and (ii) the overall benefit/cost ratio when the forecasts are done based on the different proposed alternatives. Estimates from the model that we use in our baseline estimations are not sensitive to the departures from serial independence that we analyze here.

In the auxiliary samples, we observe outcomes $Y_{n,a}$ for $a \in [a^*, \ldots, \bar{A}]$. In the experimental sample, we observe the outcome $Y_{e,a}$ for at most two ages, depending on the outcome. We initially assume that we observe the outcome at one age ($a = a^*$). We later relax this assumption. We thus use the information in the auxiliary sample at $a \in [a^*, \ldots, \bar{A}]$ to form extrapolations for the experimental sample, where we do not observe the outcome of interest during this age interval. We produce out-of-sample forecasts and calculate the net present value of labor income under different specifications that have different assumptions. Table C.12 summarizes the five specifications that we implement.

Table C.12: Net Present Value of Labor Income and Benefit/Cost Analysis Under Different Specifications for Labor Income Forecasts

| | Specification 1: ("Baseline") $\lambda_1 \neq 0$ $\rho = 0$ $f = 0$ | | | Specification 2: $\lambda_1 = 0$ $\rho \neq 0$ $f = 0$ | | | Specification 3: $\lambda_1 \neq 0$ $\rho \neq 0$ $f = 0$ | | | Specification 4: $\lambda_1 \neq 0$ $\rho = 0$ $f \neq 0$ | | | Specification 5: Non-parametric matching | | |
	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C
Pooled	133,032 (76,634)	0.14 (0.03)	7.33 (1.84)	154,547 (187,036)	0.15 (0.12)	7.31 (5.15)	268,179 (211,089)	0.26 (0.14)	12.68 (5.81)	46,953 (25,323)	0.05 (0.02)	2.22 (0.7)	132,924 (11,253)	0.13 (0.01)	6.28 (0.31)
Males	238,105 (185,375)	0.15 (0.04)	10.19 (2.93)	200,509 (160,988)	0.11 (0.05)	9.35 (5.51)	456,078 (358,534)	0.25 (0.12)	21.26 (12.28)	74,775 (54,752)	0.04 (0.02)	3.49 (1.88)	196,530 (20,210)	0.11 (0.01)	9.16 (0.69)
Females	41,908 (24,606)	0.10 (0.06)	2.61 (0.73)	79,441 (99,416)	0.19 (0.28)	4.64 (3.19)	31,303 (168,160)	0.07 (0.48)	1.83 (5.4)	19,959 (34,142)	0.05 (0.1)	1.17 (1.1)	69,317 (4,350)	0.17 (0.0 1)	4.05 (0.14)

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

C.6.1 Specification 1: Lagged Component ($\lambda_1 \neq 0$); No Serial Correlation ($\rho = 0$); and No Fixed Effect ($f = 0$)

Our baseline estimations are constructed using this framework for labor and transfer income, crime, and health. The realized values and forecasts are close, as displayed in Figure 2, to the baseline specification. We test and fail to reject the nulls of invariance across the treatment and the control groups, invariance across the experimental and auxiliary samples, and exogeneity in both the experimental and the auxiliary samples. The tests are conducted at $a = a^*$.

C.6.2 Specification 2: No Lagged Component ($\lambda_1 = 0$); Serial Correlation ($\rho \neq 0$); and No Fixed Effect ($f = 0$)

Given that $Y_{k,a-1}$ is not one of the elements in $\boldsymbol{X}_{k,a}$, it is plausible that Assumption A–3 (Exogeneity) holds even when we do not restrict ρ. In this case, it is straightforward to account for serial correlation.[87]

The predictions reported in the baseline estimates are very similar to the ones generated in this case. That is, the lag does not help in generating predictions as much as one might think it would. This is more evidence in favor of $\boldsymbol{X}_{k,a}$ summarizing the treatment effects.

C.6.3 Specification 3: Lagged Component ($\lambda_1 \neq 0$); Serial Correlation ($\rho \neq 0$); and No Fixed Effect ($f = 0$)

These estimates indicate that serial correlation is present in the data. From ages 21 to 30, we estimate the model in the CNLSY. The estimate for ρ is 0.7465. From ages 30 to 67 (assumed retirement) we estimate the model on the NLSY79/PSID. The estimate for ρ is

---

[87]Serial correlation can be accounted for in a general way using the Newey-West approach (Newey and West, 1986).

0.5426. When we restrict the sample to people who earn \$30,000 (2014) at each of these ages, the analogous estimates of ρ are 0.7370 and 0.5316. These estimates are statistically significant at the 1% level.

We can ρ-transform the system of interest to obtain consistent estimates. Drop the j index for simplicity and write:

$$Y_{k,a} = \lambda_0 \left(1 - \rho\right) + \left(\lambda_1 + \rho\right) Y_{k,a-1} - \lambda_1 \rho Y_{k,a-2} + \lambda_2 \left(\boldsymbol{X}_{k,a} - \rho \boldsymbol{X}_{k,a-1}\right) + U_{k,a}. \tag{17}$$

OLS produces consistent estimates of the coefficients in Equation (17). This enables us to construct forecasts. This model generates forecasts close to the baseline estimates.

C.6.4 Specification 4: Permanent-Transitory Decomposition of Unobserved Components $\left(\lambda_1 \neq 0; \; \rho = 0; \; f \neq 0\right)$

This framework allows for serial dependence due to the lagged dependent variable but does not allow for serial correlation in η_a. Estimates from this strategy are very similar to those from the other procedures.

C.6.5 Specification 5: Non-Parametric Predictions

An alternative to any of these scenarios is to make forecasts using a non-parametric matching algorithm. This alternative is described in Appendix C.3.5.

A sufficient condition for the validity of this procedure is Assumption A–3 (Exogeneity). Without exogeneity, the joint distributions of $\boldsymbol{X}_{k,a*}, \varepsilon_{k,a*}$ do not necessarily coincide across the experimental and the auxiliary samples. For example, in the experimental sample, individuals in the treatment group have relatively high levels of education due in part to the

exogenous boost generated by randomization into treatment. In the auxiliary sample, the usual form of ability bias may be at work: individuals with relatively high levels of education might have better motivation, better parents, etc. Thus, the empirical relationships between education and labor income may differ across experimental and auxiliary samples. Exogeneity avoids this problem, but clearly only a weaker assumption is required.

C.7 Inference

This section provides the precise steps for constructing the bootstrap distribution and for computing the standard errors for the main estimates in our paper.

C.7.1 Inference for Parametric Forecasts

We execute the following steps to compute the empirical bootstrap distribution and the standard errors when forecasting outcomes out of sample by combing experimental and auxiliary datasets.

1. Resample the experimental sample with replacement at the individual level. We resample individuals independently of their treatment status. When an individual is resampled, his/her whole life-cycle history is considered. Denote this sample by (e, s).

2. Impose the initial conditions on the auxiliary sample as described in Appendix C.3.2.

3. Resample the non-experimental sample with replacement at the individual level. When an individual is resampled, his/her whole life-cycle history is considered. Denote this sample by (n, s').

4. Construct the synthetic cohorts by matching the samples (e, s) and (n, s') using Mahalanobis Matching. Note that this matching procedure is based solely on baseline variables not affected by treatment.

5. Fit the dynamic relationships in Equation (2), using predictors as discussed in Appendix C.3.4. For estimates by gender, make different predictions by gender.

6. Use the fitted dynamic relationship to fit out of sample in (e, s). This gives us an age-by-age forecast *without forecasting error* for each individual in the experimental sample. Store the predictions at all ages for individual i in this sample in a vector $\mathcal{Y}_{i,e,s}$, where $\mathcal{Y}_{i,e,s}$ is the vector of forecasts for individual i in the experimental bootstrap sample s.

7. When fitting the dynamic relationship in sample (n, s'), individual j in the non-experimental sample has a vector containing the residuals of each of her predicted variables (for example, labor income). Call this vector of residuals $\boldsymbol{\mathcal{E}}^d_{j,n,s'}$: the vector of residuals for individual j in the auxiliary bootstrap sample s'.

8. Randomly pair individual i in sample (e, s) with individual j in sample (n, s'). This random pairing is stratified by the variables $\boldsymbol{B}_k$ and $\boldsymbol{X}^d_{k,a}$ defined in the paper. The forecast accounting for forecasting error is $\mathcal{Y}_{i,e,s} + \boldsymbol{\mathcal{E}}_{j,e,s'}$.

9. Repeat this for all pairs of samples $(n, s'), (e, s)$. Resample the experimental sample and auxiliary sample 100 times each. This gives us the empirical bootstrap distribution, with 100*100 points.

10. Compute the standard error as the sample standard deviation of the 100*100 resamples. Compute the p-value's as the proportion of times that we reject the null hypothesis, after centering the empirical bootstrap distribution according to the null hypothesis.

C.7.2 Inference for Non-Parametric Forecasts

1. Resample the experimental sample with replacement at the individual level. We resample individuals independently of their treatment status. When an individual is resampled, his/her whole life-cycle history is considered. Denote this sample by (e, s).

2. Impose the initial conditions on the auxiliary sample as described in Appendix C.3.2.

3. Resample the non-experimental sample with replacement at the individual level. When an individual is resampled, his/her whole life-cycle history is considered. Denote this sample by (n, s').

4. Perform the matching procedure to construct the synthetic cohort, estimate the dynamic relationship and impute the profiles by matching on baseline variables and on variables affected by treatment as explained in Appendix C.3.5. Store the predictions at all ages for individual i in this sample in a vector $\tilde{\mathcal{Y}}_{i,e,s}$, where $\tilde{\mathcal{Y}}_{i,e,s}$ is the vector of forecasts for individual i in the experimental bootstrap sample s.

5. Compute the residuals in step 7 of Appendix C.7.1.[88]

6. Randomly pair individual i in sample (e, s) with individual j in sample (n, s'). This random pairing is stratified by the variables $\boldsymbol{B}_k$ and $\boldsymbol{X}_{k,a}^d$ defined in the paper. The forecast accounting for forecasting error is $\tilde{\mathcal{Y}}_{i,e,s} + \boldsymbol{\mathcal{E}}_{j,e,s'}$.

7. Repeat this for all pairs of samples $(n, s'), (e, s)$. Resample the experimental sample and auxiliary sample 100 times each. This gives us the empirical bootstrap distribution, with 100*100 points.

8. Compute the standard error as the sample standard deviation of the 100*100 resamples. Compute the p-value's as the proportion of times that we reject the null hypothesis, after centering the empirical bootstrap distribution according to the null hypothesis.

---

[88]When we compute the residuals using the fully nonparametric model we get the following net present value for labor income: pooled (males and females) 137,133 (s.e. 19,989); males 204,256 (s.e. 32,212); females 64,522 (s.e. 9,987). Table C.12 shows that these results are closely aligned to the two approaches reported in the paper.

C.7.3 Benefit/cost Ratio or Internal Rate of Return

1. Use the same sampling procedure as when computing the standard error for the forecasts. In this case, compute the predictions for all outcomes.

2. Discount the forecasts to the age of birth.

3. Compute benefit/cost ratios and internal rates of return.

4. Discard internal rate of returns not satisfying the single crossing property (see Appendix C.4).

5. Compute standard errors and p-value's as before.

C.8 Procedures for Selecting Background Variables

In this appendix we first explain our method for selecting the background variables B.[89]

C.8.1 Background Variables

We select three out of fourteen candidate potential variables that best predict the relevant outcomes of interest, i.e. the outcomes we test treatment effects for. We list the fourteen variables in Table C.13 and bold the three we choose. In addition to these three variables, we account for a male indicator when computing estimates pooling males and females and a ABC/CARE indicator, to account for any difference in the programs—although we extensively document throughout the paper the similarities between them.

---

[89]This analysis is distinct from the analysis of the selection of variables to forecast life-cycle profiles of labor income and other outcomes. For that discussion see Appendix C.3.4.

Table C.13: Background Variables

Maternal IQ	Maternal education	Mother's age at birth
High Risk Index	Parent income	Premature birth
1 minute Apgar score	**5 minute Apgar score**	Mother married
Teen pregnancy	Father at home	Number of siblings
Cohort	Mother is employed	

Note: This table lists the variables we permute over when selecting the background variables we control for in our estimations. We bold the variables we choose based on the procedure explained in this section.

We briefly formalize the choice of the control sets based on most predictive models in the next lines.

Let $\mathcal{M}$ be the set of all the models we consider. In our context, $\mathcal{M}$ consists of all linear regressions of an outcome of interest on the different combinations of background variables. $m \in \mathcal{M}$ is one of such models. We choose the model minimizing the Bayesian Information Criterion (BIC) by ranking them according to their likelihood. That is, according to their posterior probability given the data. The data, in this case, are the dependent variable being predicted together with the background variables in each combination. We denote this by $\Pr(m|\text{Data})$.

Using Bayes Rule and the law of total probability,

$$
\begin{aligned}
\Pr(m|\text{Data}) &= \frac{\Pr(\text{Data}|m) \times \Pr(m)}{\Pr(\text{Data})} \\
&= \frac{\Pr(\text{Data}|m) \times \Pr(m)}{\sum_{m' \in M} \Pr(\text{Data}|m') \Pr(m')} \\
&\propto \Pr(\text{Data}|m) \times \Pr(m),
\end{aligned}
\tag{18}
$$

where $\Pr(m)$ is the prior probability of model m and $\Pr(\text{Data}|m)$ is the probability of observing Data under model m.

There are various approaches for ranking the the likelihoods of each model. Examples include rankings based on Bayesian Information Criterion (Schwarz), the Hannan-Quinn Information Criterion (HWIC), and the Akaike Information Criterion (AIC). We use the first approach because it has appealing consistency properties (Diebold, 2007). This criterion minimizes the following loss function: $2 \log[\Pr(\text{Data}|m)]$. We follow an specific approximation developed by Claeskens and Hjort (2008), which assumes uniform priors and simplifies the computation of the loss function.

This procedure allows us to choose one control set per outcomes of interest. To gain consistency across all specifications, we sum the BIC across all outcomes and choose the background variables with lower average across models. These background variables form our control set across all estimations and appear bold in Table C.13.

We perform an analogous procedure when choosing the variables to construct the IPW scheme.

C.8.2 Sensitivity Analysis

We examine the sensitivity of our estimates to the choice of background variables. To investigate this issue, we estimate treatment effects for the three counterfactuals we consider using all possible control sets for the three variables we can form with the background variables in Table C.13. We also consider all possible control sets of one and two variables in Table C.13. For brevity, we present this exercise for two outcomes, employment and education. Similar exercises for the 95 main outcomes we consider are available upon request.

Figure C.7 displays the results from this exercise. In any case, the support of the distributions is very compressed leading us to conclude that there is little sensitivity to the choice of controls sets. In any case, the support of the distributions is very compressed leading us to conclude that there is little sensitivity to the choice of controls sets. This is especially true

for the comparisons of treatment vs. staying at home and vs. alternative preschool.

Figure C.7: Sensitivity to Choice of Control Set, Treatment vs. Next Best

(a) Years of Education, Males
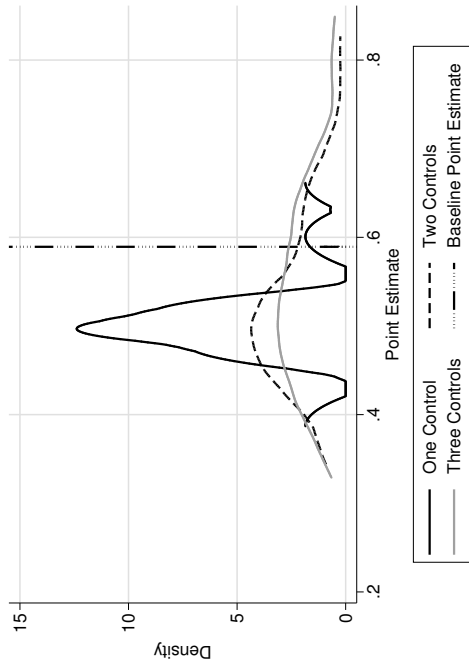
(b) Employment, Males
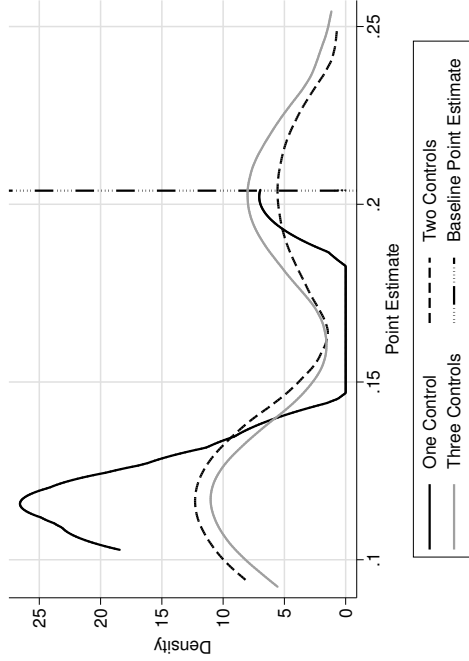
(c) Years of Education, Females
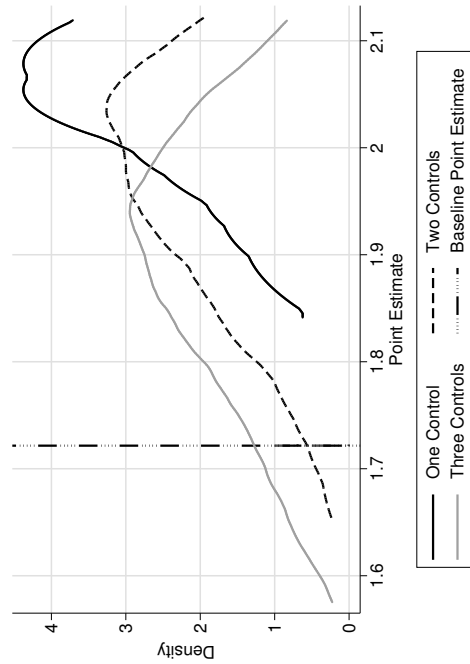
(d) Employment, Females
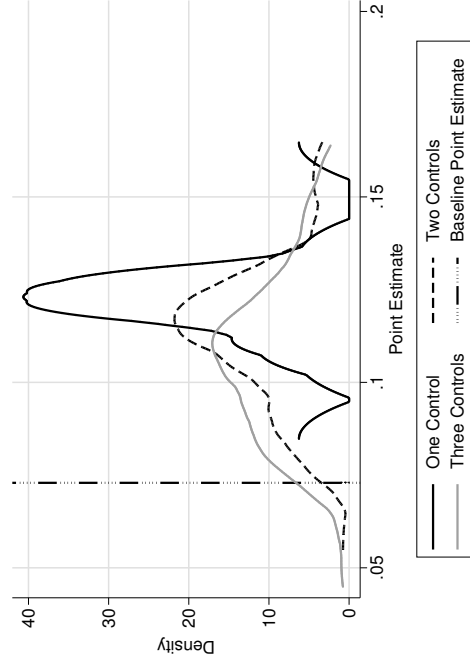
Note: Panel (a) displays the distribution of the treatment effect estimate of the treatment compared to next best counterfactual for males years of education. The distribution is obtained by using all possible combinations of one, two, and three background variables listed in Table C.13. In addition to these three variables, we account for a male indicator when computing estimates pooling males and females and a ABC/CARE indicator, to account for any difference in the programs—although we extensively document throughout the paper the similarities between them. The horizontal line marks the baseline estimate we use. The remaining panels present analogous distributions for the outcomes and genders indicated in the title.

116

D   Costs of Education

In constructing the rate of return and benefit/cost estimates, we account for short- and long-term components of educational costs. The short-term components include savings due to reductions in special education and grade retention. The long-term components include the type and level of the highest educational attainment at age 30. We do not calculate costs of education beyond age 30 because we do not have data on later educational attainment of subjects. We do not project future education costs to avoid adding further uncertainty to the model. Using the Panel Study of Income Dynamics (PSID), we document only negligible growth in educational attainment for a sample of individuals for whom $\boldsymbol{B} \in \mathcal{B}_0$, who are born between 1972 and 1982.

To estimate the costs of additional schooling, we combine various sources. Table D.1 describes the yearly cost of education at every level and the age and duration for which these costs are incurred. We apply these costs additively up to the highest level of educational attainment by age 30. Pooling males and females, the treatment groups had on average higher attainment and incurred a greater cost of education. To find the present value of the difference between the treatment and control groups, we first array educational attainment by stage as in Table D.1. We find a difference between the average educational attainment of the treatment groups (finish community college) and the average educational attainment of the control groups (start community college). This can be represented as a cost that is $12,586 higher for the treatment groups than for the control groups, as in Table D.1. The effect of the program on the educational attainment of females, however, is much greater than the effect on that of males, with corresponding higher female costs.

Table D.1: Yearly Individual Education Costs

Schooling level	Ages	Duration (Years)	Yearly Cost	Attainment & Notes
K-8	6-14	8	$9,113	All subjects. Assume dropout before Grade 9 completed up to Grade 8.
High School (Not Completed)	15-16	2	$9,113	This is from Grade 9 to Grade 10. Assume high school dropout completed up to Grade 10.
High School (Completed)	15-18	4	$9,113	This is from Grade 9 to Grade 12. 47:38 (control:treatment)
GED (Not Completed)	18	.5	$155	GED is considered a one year program. No subjects identified as having starting a GED program without finishing.
GED (Completed)	18	1	$155	1:1 (control:treatment).
Community College/Technical Training (Not Completed)	19	1	$7,001	Assume dropouts drop at the end of the first year. 19:7 (control:treatment)
Community College/Technical Training (Completed)	19-20	2	$7,001	18:19 (control:treatment)
Any College (Not Completed)	19-20	1	$11,886	Assume dropouts drop out at the end of the second year. 7:11 (control:treatment).
Any College (Completed)	19-22	4	$11,886	6:14 (control:treatment)
Graduate School (Not Completed)	23	1	$9,704	Assume dropouts drop out at the end of the first year. 3 treated individuals.
Finished Masters	23-24	2	$9,704	1 treated individual
Finished PhD	23-26	4	$9,704	1 treated individual
Grade Retention	NA	1	$9,113	
Special Ed.	NA	1	$11,705	

Sources: Snyder and Dillow (2012); Hoenack and Weiler (1975); Dhanidina and Griffith (1975); Freeman (1974).
Note: This table reports the yearly cost and duration of each type of education, as well as the ages for which we evaluate them. All amounts are inflated to 2014 USD. We show the number of subjects who identified themselves as being in each education category (total number of respondents: 101/114). To compute the total cost of education for a subject, we applied these costs additively up to the highest level of educational attainment. We only adjust the costs of K-12 education, special education, and grade retention for deadweight loss. Because it gives costs that are incurred across many years, this table does not show their present discounted value.

D.1 Measuring Lifetime Educational Attainment

Follow-up data on educational attainment were collected for ABC/CARE subjects up to age 30, on average. This may not necessarily be an accurate measure of lifetime educational attainment. We perform an exercise using data for eligible people from the Panel Study of Income Dynamics (PSID) to assess educational attainment after age 30. We find an average increase of .007 years of additional years of schooling for individuals between the ages of 40 to 60.

D.2 Cost of Education

We apply the costs described in Table D.1 to subjects' educational attainment through age 30 to calculate the public and private costs of lifetime educational attainment. Costs up to high school are assumed to be public costs with corresponding corrections required for distortions associated with taxation to finance schooling.

D.3 Non-monetary Benefits of Education

There are many social and non-monetary benefits of education that our analysis cannot capture. These benefits impact the individual's quality of life, the general well-being of society through positive peer effects as well as fewer costs and negative externalities, and even the well-being of future generations. Documenting them all may be impossible, but we briefly review some major benefits in this subsection. Vila (2000) documents private benefits with external effects, such as health (increases in longevity and better nutrition and preventative care choices). Higher education is also associated with decreased fertility rates linked with improved infant health and lower mortality rates. Moreover, higher education not only improves labor outcomes with respect to employment prospects and salary, but

also with regard to how individuals perceive work and leisure, with more education leading to increased satisfaction from leisure. Furthermore, higher education is linked with better savings behavior and higher rates of return on savings. Higher education is also connected with social stability—better education promotes good citizenship and creates communities that are less likely to experience violent social conflict.[90]

---

[90]See, e.g., Lochner (2011a,b).

E   Quantifying the Benefits in Crime Reduction[91]

We next describe how we quantify the benefits of ABC/CARE on crime reduction.

E.1   Data Description

The crime data available in ABC/CARE come from four different sources provided by the program, which we supplement with auxiliary datasets. We summarize the ABC/CARE datasets and auxiliary datasets related to crime below.

E.1.1   ABC/CARE Datasets

1. **Administrative Youth Arrests Dataset.** This dataset is only available for ABC. This dataset records all arrests up to the age at which the data were obtained (April, 1996), when ABC subjects were about 21 years old. The categories of crimes in this dataset are coarser than the ones that we use in the other datasets: it categorizes crimes into property, violent, drug, and miscellaneous crimes. We assume some equivalences, as shown in Table E.1.

2. **Administrative Adult Arrests Dataset.** Gathered when ABC and CARE subjects were around 34 years old, this dataset includes individual data on arrests, with short descriptions of the associated offenses. This dataset includes some subjects for whom the arrest history is missing. To resolve this, we use a methodology (detailed below) involving the sentences data, which does not have missing values.

3. **Administrative Sentences Dataset.** Gathered when ABC and CARE subjects were about 34 years old, this dataset includes individual data on sentences, with de-

---

[91] Andrés Hojman contributed to the analysis in this appendix.

scriptions of the crimes. It also includes total sentence length (projected and actual) and punishment type (jail, prison, parole).

4. **Self-reported Adult Crimes Dataset.** A module on crimes was included at the age-21 and age-30 interviews for both ABC and CARE. After matching all datasets, we use the information from the self-reported crimes whenever a particular crime does not appear in the other datasets.

E.1.2 Auxiliary Datasets

1. **National Crime Victimization Survey (NCVS).** The NCVS is a survey (self-reported) on crime victimization reported on the household level. The NCVS does not cover crimes to businesses, and it might under-report crimes that might not be known to all members of a family, such as rape. It also does not give statistics for murders. The data are available from 1993 to 2013. We use NCVS to estimate the total number of victims per type of crime in the U.S., which is used to construct victim-arrest ratios.

2. **Uniform Crime Reporting Statistics (UCRS).** This dataset contains all crimes that are reported to the police. It contains crimes to households, individuals, and businesses that are captured by most of the law enforcement agencies in the country. We used data from UCRS spanning the years 1996 to 2012. We use UCRS as a complement to NCVS when we estimate the total number of victims per type of crime in the U.S.

3. **National Judicial Reporting Program (NJRP).** We use the NJRP to get data for total number of sentences in the U.S. The data were taken from reports published biennially from 1986 to 2006. We use this dataset to construct arrest-sentence ratios for our crime categories.

4. **North Carolina Department of Public Safety dataset (NCDPS).** This dataset

contains information since 1972 on each individual who is convicted of a crime and enters the state prison system or community supervision in North Carolina. The data do not include arrests, or sentences involving jail or unsupervised probation. Because the North Carolina Department of Public Safety was only formed in 2015 to consolidate the state's Department of Correction, Department of Juvenile Justice, and Crime Control, among other state agencies, these data were mostly constructed by the other agencies. We use this dataset to fit a forecasting model that we use to forecast future crimes for ABC/CARE subjects.

E.1.3 Crime Categories

The administrative adult datasets have descriptions of all crimes committed by ABC/CARE subjects. We categorize these crimes to be as comparable as possible to the categories in the other data sources and in the literature on calculating the cost of crime. However, some crimes do not clearly fit into the broad categories. As shown in Table E.1, the different experimental and auxiliary datasets and the literature use different crime categories. The categorization we present is our effort to make all sources comparable.

Table E.1: Crime Categories

Categories	Youth Data	Costs of Crime	Nat. Arrests Data	Nat. Sentences Data
Arson		Arson	Arson	
Assault	Violent	Assault	Total assaults	Aggravated assaults
Burglary		Household burglary	Burglary	Burglary
Fraud		Fraud	Fraud, Forgery, Embezzlement	Fraud, Forgery
Larceny	Property	Larceny/theft	Larceny	Larceny/theft
Miscellaneous	Drug, Misc.		Drug abuse total	Drug offenses
Vehicle Theft		MV theft	MV theft	MV theft
Murder		Murder	Murder, Non-negligent manslaughter	Murder, Manslaughter
Rape		Rape/sexual assault	Forcible rape, Sex offense	Rape
Robbery		Robbery	Robbery	
Vandalism		Vandalism	Vandalism	

Note: This table shows the various measures we have for our categories of crimes from each dataset. "Costs of Crime" are from McCollister et al. (2010).

E.2 Methodology for Estimating Crime Costs

In this section we give a detailed explanation of the steps taken to calculate the treatment effects on crime and the costs associated with them. We first give a more abstract and formal summary of the process we use, and then discuss the details for each step.

1. *Count Arrests and Sentences.* We count the total number of sentences for each subject, i, and category of crime (robbery, larceny, etc.), j, up to age 34, which we denote by $S_{i,j}^{34}$. We also match the data on adult arrests, juvenile arrests, and self-reported crimes, to construct total number of arrests for each crime type up to that age, $A_{i,j}^{34}$. For some subjects, the arrest data are missing. In those cases, we impute the missing data by assuming that the national arrest-sentence ratio for crime type, j, is valid for each subject. Let $\overline{A_j}$ be the national total number of arrests for crime type, j, and let $\overline{S_j}$ be the national total number of sentences. Then, we construct $r_j = \frac{\overline{A_j}}{\overline{S_j}}$, and we impute $A_{i,j}^{34} = r_j S_{i,j}^{34}$.

2. *Construct Forecasts.* From our external sources, we have data to forecast lifetime sentences. In that data, we use sentences up to age 34 in all types of crime to forecast future sentences for that crime type, $\widehat{S_{i,j}^{35-50}}$. This gives an estimate of the lifetime sentences as $\widehat{S_{i,j}} = S_{i,j}^{34} + \widehat{S_{i,j}^{35-50}}$. Given that we do not have an analogous dataset to forecast lifetime arrests, we impute the forecasted arrests as a linear function of the forecasted number of sentences: $\widehat{A_{i,j}^{35-50}} = r_j \widehat{S_{i,j}^{35-50}}$. Then, we calculate $\widehat{A_{i,j}} = A_{i,j}^{34} + \widehat{A_{i,j}^{35-50}}$.

3. *Estimate Number of Victims.* Let the national number of victims of a given type of crime be $\overline{V_j}$. We construct a victimization inflation factor for each crime type: $f_j = \frac{\overline{V_j}}{\overline{A_j}}$. It represents the number of times someone is arrested as a fraction of the number of victims of the crimes. Then, the estimated number of victims of subject, i, for crime

124

type, j, based on arrests is estimated as $\widehat{V_{i,j}^A} = A_{i,j}f_j$. For sentences, we calculate an analogous estimate of victims based on the victimization inflation factor and the arrest-sentence ratio: $\widehat{V_{i,j}^S} = S_{i,j}f_jr_j$. Both estimates are similar, as we show below. We construct our final estimate of the lifetime victims of subject, i, for crime type, j, as the average of both estimates to achieve greater precision: $\widehat{V_{i,j}} = \left(\widehat{V_{i,j}^A} + \widehat{V_{i,j}^S}\right)/2.$

4. *Find Total Costs of Crimes.* We use estimates of the cost of crimes for victims from the literature for each crime type j, c_j^V. We impute the total victim costs of subject, i, for crime type, j, as $\widehat{C_{i,j}^V} = \widehat{V_{i,j}}c_j^V$. We also calculate different costs from the justice system (including police) associated with the different crime types, but only for the ones that included arrests or sentences (i.e. we do not consider the victimization inflation), as: $C_{i,j}^{JS} = \widehat{A_{i,j}}c_j^{JS}$. Finally, we also construct the total costs of incarceration for crime type, j, $\widehat{C_{i,j}^P}$ as the total time the subject was imprisoned for that type of crime, $P_{i,j}$, multiplied by the cost of a day in prison c_P. All of our cost estimates are discounted to birth.

E.2.1   Count Arrests and Sentences

Unlike previous studies of the crime-reducing effects of early childhood programs, we use several different datasets to construct measures of criminal activity of the ABC/CARE subject, and check for consistency across them. We generate a count of the number of arrests and the number of sentences by crime category for each ABC/CARE subject. We start by counting sentences – a straightforward task – as we have complete information on sentences for all ABC/CARE subjects in one dataset.

Unlike counting sentences, the count of arrests is more involved.[92] We now describe the procedure we follow to get a count of arrests that is as complete as possible. We begin by

---

[92]Throughout this appendix, for all the calculations involving counting arrests, if the arrest was the result of more than one *offense*, we count the number and type of offenses separately, rather than counting all of them as one.

matching, based on crime description and date, the arrests from the adult administrative data, the youth administrative data, and the self-reported dataset. It is possible that the adult data are missing youth arrests that were expunged from the criminal records or crimes committed in states other than North Carolina, which were not obtained in the collection of administrative data. Because we observe the arrest dates, we confirm that we are not duplicating any arrests. To align the youth arrests data, which are categorized more broadly, we assume that all violent crimes are assaults (the most common category of violent crime in the sample) and that all property crimes are larcenies (the most common category of property crime). We categorize both drug and miscellaneous crimes in the miscellaneous category, for which we do not compute victim costs.

The main problem we confront with these data is that there are some individuals who are missing data on arrests in the data collected at age 34. We know this is the case because there are sentences for which we do not observe the corresponding arrests. While this is expected for some crimes, such as speeding or shoplifting, it is not possible for others. This discrepancy seems to be due to the difficulty in collecting administrative information on arrests. Counting arrests using many data sources eases reconstruction of adult arrests data. We tackle this challenge by first identifying those individuals who should have arrests, because they have sentences and so necessarily were arrested. Then we impute the number of arrests for each of those sentences, based on the national arrest-sentence ratio. We estimate that 11 individuals (10% of the ABC sample) have missing arrest data. We note that the final estimates of effects on reduction in crime costs using only sentences (which do not have any missing values) are very similar to the ones using arrests.

The effect of adding the different datasets, as well as the final counts of both arrests and sentences is presented in Figure E.1. The first three pairs of bars present the number of additional crimes included in the data by the addition of the self-reported, juvenile, and sentences datasets. In the case of sentences, we first show the effect of imputing just one

arrest per sentence, and only for individuals with missing arrests. Notice that the juvenile data only adds assaults, larcenies, and miscellaneous, as discussed above. The next pair of bars shows the effect of adding more than one arrest per sentence for the individuals with missing data, using the arrest-sentence ratios. The effect is important, adding about 30% more crimes to the previous estimate. Importantly, some rape arrests are added to the control group in this step, because of one subject who presented a sentence for that crime. Finally, the total number of sentences is far lower than the total number of arrests, even if only original arrests are counted. We also note in this chart that the volume of the crimes is mostly driven by miscellaneous crimes (which are mostly drug and traffic crimes). As these crimes are counted as victimless in our methodology, their effect on the final estimates is much smaller than what this chart might suggest.

Figure E.1: Counts of Arrests and Sentences
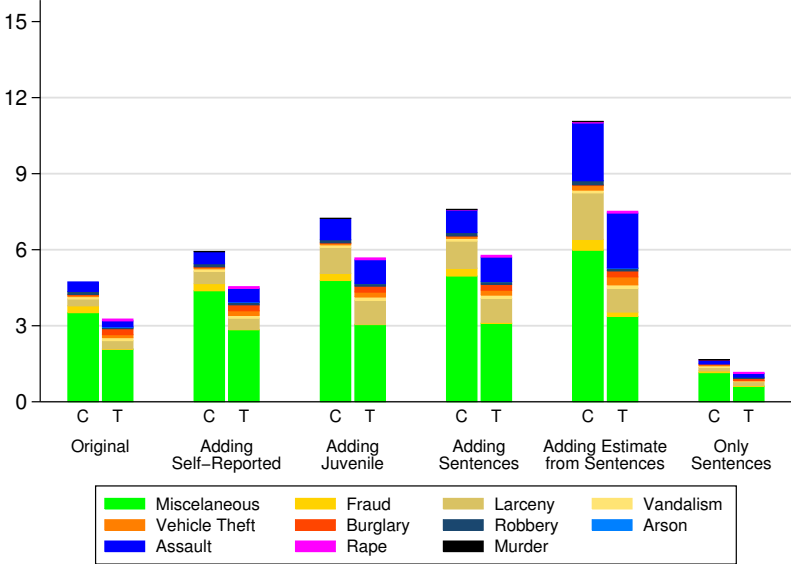
Note: This chart shows the arrests per capita for the control (C) and treatment (T) groups. The first pair of bars shows the original arrests data from the administrative adult dataset. The next pair adds the self-reported crimes that did not match with the original arrests data. The next pair adds data from the administrative juvenile dataset that did not match with the previous datasets. The next pair adds one arrest for every sentence that did not match with the previous datasets and one arrest for every sentence that had arrest data missing. The next pair adds n arrests instead of one for each sentence, where n is calculated using the arrest-sentence ratio obtained from auxiliary datasets. The final pair of bars, for reference, is the total number of sentences from the administrative sentences dataset.

E.2.2 Constructing Forecasts

The data available describe the ABC/CARE subjects' crimes committed up to age 34. However, we believe that the effect of the programs on crime does not stop at that age. To forecast the number of crimes that the study participants commit beyond age 34, we use data from the North Carolina Department of Public Safety (NCDPS). Because the crime data are obtained from the same state in which ABC/CARE were implemented, the forecasting model is appropriate for the ABC/CARE samples. To the best of our knowledge, no study of the effects of an early childhood education program has ever used microdata to estimate a predictive model for future crimes. The estimates in Heckman et al. (2010) are based on national age ratios (crimes of a certain category committed by older people over crimes of the same category committed by younger people at a specific time), rather than on microdata. However, age ratios only consider the same category of crime as an input to the estimation model and the results can reflect demographic transitions.

It would be ideal to use forecasting models estimated from the same cohorts as the ABC/CARE subjects. However, it is not possible to forecast crimes committed until later ages because the ABC/CARE subjects are currently about 46 years old. To forecast crimes at older ages, it is necessary to use outcomes for earlier cohorts. The data are available from 1972 (44 years ago as of 2016), and thus they do not contain a complete criminal history for any cohort of individuals. We assume that few crimes are committed after the age of 50. We separately estimate predictions from ages 35-40, 40-45, and 45-50. We have plentiful observations to estimate crimes in all the age ranges. We calculate our forecasts using a five-step procedure:

1. Find individuals that are at least 40 years old as of 2016 in the NCDPS dataset.

2. Regress the number of crimes of each type at ages 35–40 on those at ages 16–34.

3. Use the estimated forecasting model for ABC individuals, attributing those crimes to

age 40 for discounting purposes.

4. Repeat the three previous steps for ages 40–45 and 45–50.

5. For individuals with no criminal histories before age 34, assume that they commit no crimes after 34.

Table E.2: NCDPS Regressions of Ages 35–40 on Ages 16–35, Females

	Miscellaneous	Fraud	Larceny	Vandalism	Auto Theft	Burglary	Robbery	Arson	Assault	Rape	Murder
Miscellaneous	0.103	-0.029	0.009	0.001	0.001	0.001	0.000	-0.000	0.002	-0.000	-0.001
	(27.71)**	(5.90)**	(6.00)**	(3.88)**	(5.18)**	(2.25)*	(1.91)	(0.62)	(3.16)**	(1.27)	(2.85)**
Fraud	-0.047	0.104	0.003	-0.000	0.000	0.000	0.000	0.000	-0.001	-0.000	-0.000
	(21.71)**	(36.61)**	(3.81)**	(2.31)*	(0.60)	(2.02)*	(0.02)	(1.76)	(1.48)	(0.75)	(1.72)
Larceny	0.034	0.002	0.105	0.001	0.002	0.002	0.001	0.000	0.004	0.000	-0.000
	(8.16)**	(0.33)	(62.36)**	(2.37)*	(10.35)**	(7.80)**	(5.97)**	(2.99)**	(5.65)**	(1.30)	(1.67)
Vandal	0.000	-0.036	-0.024	0.012	0.004	0.007	0.005	0.003	0.043	0.000	-0.000
	(0.01)	(1.10)	(2.43)*	(6.40)**	(2.86)**	(3.87)**	(3.34)**	(3.40)**	(9.27)**	(0.79)	(0.23)
Auto Theft	0.059	-0.026	0.055	0.004	0.012	0.007	0.005	-0.002	-0.014	-0.000	0.002
	(1.55)	(0.52)	(3.57)**	(1.28)	(5.94)**	(2.48)*	(2.52)*	(1.30)	(2.02)*	(0.23)	(1.02)
Burglary	0.049	-0.031	0.011	0.002	0.006	0.044	0.000	-0.001	-0.003	-0.000	-0.001
	(1.77)	(0.86)	(1.01)	(0.95)	(3.86)**	(22.56)**	(0.20)	(1.26)	(0.55)	(0.25)	(0.34)
Robbery	0.030	0.054	0.040	0.002	0.005	-0.001	0.028	0.002	-0.000	-0.000	0.002
	(1.05)	(1.46)	(3.51)**	(1.03)	(3.43)**	(0.28)	(17.69)**	(2.22)*	(0.07)	(0.29)	(1.50)
Arson	0.070	-0.087	-0.035	0.008	0.011	0.006	0.025	0.010	0.017	-0.000	0.005
	(1.29)	(1.22)	(1.61)	(1.98)*	(3.98)**	(1.62)	(8.19)**	(4.95)**	(1.67)	(0.19)	(1.60)
Assault	0.064	-0.060	0.007	0.007	0.001	0.002	-0.000	0.002	0.048	-0.000	0.000
	(5.74)**	(4.11)**	(1.63)	(7.82)**	(1.83)	(1.93)	(0.14)	(4.52)**	(22.78)**	(0.36)	(0.34)
Rape	-0.015	-0.131	-0.003	-0.003	-0.003	-0.004	-0.006	-0.001	0.001	-0.000	-0.002
	(0.15)	(0.98)	(0.08)	(0.42)	(0.49)	(0.55)	(0.99)	(0.24)	(0.08)	(0.06)	(0.33)
Murder	0.006	-0.166	-0.008	0.003	-0.002	-0.002	-0.002	0.003	0.001	-0.000	-0.001
	(0.16)	(3.44)**	(0.54)	(1.19)	(0.94)	(0.75)	(1.22)	(2.24)*	(0.12)	(0.34)	(0.37)
constant	0.069	0.267	0.041	0.004	0.001	0.001	0.001	0.001	0.023	0.000	0.004
	(13.39)**	(39.01)**	(19.54)**	(10.27)**	(2.76)**	(3.17)**	(4.80)**	(3.52)**	(23.81)**	(3.38)**	(12.90)**
R^2	0.02	0.02	0.07	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00
N	63,515	63,515	63,515	63,515	63,515	63,515	63,515	63,515	63,515	63,515	63,515

Note: This table shows how the crimes at ages 35–40 (in the columns) are predicted by crimes at ages 16–35 (in the rows). The estimations are predicted using the North Carolina Department of Public Safety dataset, which contains information on all individuals that have ever been sentenced in North Carolina. We use linear regressions in all cases. The sample for this model is limited to individuals who are at least 40 years old. $* p < 0.05$; $** p < 0.01$.

Table E.3: NCDPS Regressions of Ages 35–40 on Ages 16–35, Males

	Miscellaneous	Fraud	Larceny	Vandalism	Auto Theft	Burglary	Robbery	Arson	Assault	Rape	Murder
Miscellaneous	0.095	-0.002	0.004	0.000	0.001	0.001	0.000	-0.000	0.005	-0.001	-0.001
	(73.29)**	(2.13)*	(8.97)**	(2.71)**	(6.47)**	(5.87)**	(1.50)	(0.12)	(14.73)**	(3.28)**	(5.83)**
Fraud	-0.025	0.107	0.007	0.000	0.001	0.002	0.001	-0.000	-0.000	-0.000	-0.000
	(13.76)**	(67.23)**	(11.68)**	(0.47)	(3.44)**	(5.27)**	(2.53)*	(0.61)	(0.34)	(0.35)	(1.96)*
Larceny	0.041	0.012	0.092	0.001	0.005	0.010	0.004	0.000	0.009	-0.000	-0.000
	(14.34)**	(4.83)**	(97.06)**	(5.13)**	(14.85)**	(19.72)**	(12.30)**	(1.79)	(11.37)**	(0.46)	(0.59)
Vandal	0.041	-0.017	-0.007	0.019	0.002	0.002	0.002	0.002	0.029	0.001	0.000
	(4.51)**	(2.13)*	(2.22)*	(23.49)**	(1.96)	(1.06)	(1.76)	(5.41)**	(11.26)**	(1.06)	(0.62)
Auto Theft	0.015	0.014	0.005	0.002	0.043	0.013	0.000	0.000	0.008	0.001	0.001
	(1.58)	(1.70)	(1.48)	(1.86)	(39.51)**	(7.87)**	(0.14)	(0.13)	(3.12)**	(1.28)	(1.18)
Burglary	0.005	0.011	0.010	0.002	0.004	0.040	0.003	-0.000	0.003	0.001	-0.000
	(1.26)	(3.06)**	(6.86)**	(4.33)**	(9.17)**	(52.89)**	(6.55)**	(0.30)	(2.34)*	(2.37)*	(0.73)
Robbery	-0.032	0.000	0.011	0.000	0.003	0.006	0.019	-0.000	-0.000	0.000	0.001
	(5.06)**	(0.01)	(5.42)**	(0.67)	(4.61)**	(5.67)**	(26.50)**	(1.00)	(0.21)	(0.00)	(1.68)
Arson	0.008	-0.023	-0.010	0.012	-0.004	0.008	0.004	0.007	0.031	-0.001	0.001
	(0.26)	(0.82)	(0.95)	(4.40)**	(0.98)	(1.44)	(1.22)	(6.73)**	(3.49)**	(0.21)	(0.39)
Assault	0.041	-0.010	-0.000	0.005	-0.001	0.000	0.002	0.000	0.051	0.001	0.001
	(11.14)**	(3.18)**	(0.09)	(15.94)**	(1.60)	(0.12)	(4.47)**	(2.56)*	(49.57)**	(3.43)**	(3.29)**
Rape	-0.003	-0.016	-0.010	-0.002	-0.002	-0.006	-0.002	-0.000	-0.004	0.011	-0.001
	(0.33)	(1.77)	(2.86)**	(2.32)*	(1.72)	(3.42)**	(1.66)	(0.54)	(1.21)	(8.61)**	(1.69)
Murder	-0.086	-0.053	-0.025	-0.002	-0.004	-0.011	-0.001	-0.001	-0.007	-0.003	0.003
	(6.33)**	(4.45)**	(5.51)**	(1.71)	(2.54)*	(4.31)**	(0.87)	(1.53)	(1.79)	(1.56)	(3.35)**
constant	0.213	0.124	0.031	0.005	0.004	0.011	0.006	0.001	0.048	0.009	0.006
	(69.25)**	(45.55)**	(29.74)**	(18.57)**	(10.95)**	(19.97)**	(17.78)**	(9.22)**	(55.54)**	(24.70)**	(27.76)**
R^2	0.03	0.02	0.05	0.01	0.01	0.02	0.01	0.00	0.02	0.00	0.00
N	230,706	230,706	230,706	230,706	230,706	230,706	230,706	230,706	230,706	230,706	230,706

Note: This table shows how the crimes at ages 35–40 (in the columns) are predicted by crimes at ages 16–35 (in the rows). The estimations are predicted using the North Carolina Department of Public Safety dataset, which contains information on all individuals that have ever been sentenced in North Carolina. We use linear regressions in all cases. The sample for this model is limited to individuals who are at least 40 years old. $* \ p < 0.05; \ ** \ p < 0.01$.

Table E.4: NCDPS Regressions of Ages 40–45 on Ages 16–35, Females

	Miscellaneous	Fraud	Larceny	Vandalism	Auto Theft	Burglary	Robbery	Arson	Assault	Rape	Murder
Miscellaneous	0.020	-0.006	-0.001	0.000	0.000	0.000	-0.000	0.000	0.001	0.000	-0.001
	(4.33)**	(5.18)**	(0.85)	(0.78)	(0.46)	(1.29)	(0.11)	(0.84)	(0.64)	(3.66)**	(2.65)**
Fraud	0.046	0.003	0.002	0.000	0.001	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(19.30)**	(4.41)**	(2.86)**	(1.50)	(4.86)**	(0.56)	(1.62)	(0.62)	(1.13)	(2.50)*	(0.44)
Larceny	0.025	-0.002	0.070	0.000	0.001	0.002	0.000	-0.000	0.004	-0.000	-0.000
	(5.12)**	(1.54)	(41.35)**	(0.06)	(3.61)**	(6.68)**	(0.75)	(1.27)	(5.02)**	(0.45)	(0.45)
Vandal	-0.016	-0.009	-0.018	0.002	-0.001	0.005	0.001	0.003	0.019	-0.000	-0.001
	(0.52)	(1.22)	(1.71)	(1.02)	(0.58)	(2.66)**	(0.38)	(3.06)**	(3.54)**	(0.71)	(0.49)
Auto Theft	0.012	-0.011	0.010	0.004	0.012	0.013	-0.001	-0.001	-0.017	0.007	0.001
	(0.25)	(0.87)	(0.56)	(0.95)	(4.90)**	(4.70)**	(0.36)	(0.49)	(1.98)*	(11.24)**	(0.41)
Burglary	-0.034	0.019	0.007	-0.003	0.005	0.007	0.004	0.000	-0.004	0.001	-0.001
	(1.05)	(2.33)*	(0.64)	(1.05)	(3.12)**	(3.66)**	(2.09)*	(0.22)	(0.67)	(2.93)**	(0.55)
Robbery	0.027	0.004	0.021	0.004	0.002	0.007	0.015	-0.001	0.018	-0.000	-0.001
	(0.81)	(0.47)	(1.81)	(1.52)	(1.21)	(3.98)**	(6.93)**	(0.58)	(3.19)**	(0.67)	(0.82)
Arson	-0.042	-0.016	-0.025	0.013	-0.001	-0.002	0.001	-0.001	0.001	-0.000	-0.002
	(0.65)	(0.97)	(1.09)	(2.63)**	(0.46)	(0.57)	(0.31)	(0.44)	(0.12)	(0.06)	(0.56)
Assault	-0.014	-0.008	-0.007	0.002	-0.000	-0.001	0.002	0.001	0.028	-0.000	-0.000
	(1.00)	(2.44)*	(1.41)	(1.84)	(0.20)	(1.21)	(2.10)*	(2.34)*	(11.83)**	(1.23)	(0.41)
Rape	0.169	-0.014	-0.029	-0.003	-0.002	-0.003	-0.004	-0.000	-0.016	-0.000	0.011
	(1.36)	(0.45)	(0.67)	(0.36)	(0.29)	(0.46)	(0.57)	(0.07)	(0.75)	(0.05)	(1.91)
Murder	-0.169	-0.022	-0.027	-0.004	-0.002	-0.002	0.001	-0.001	0.010	-0.000	-0.001
	(4.25)**	(2.18)*	(1.93)	(1.38)	(0.84)	(0.91)	(0.55)	(0.64)	(1.46)	(0.06)	(0.50)
constant	0.304	0.029	0.046	0.005	0.002	0.002	0.002	0.001	0.022	-0.000	0.003
	(53.09)**	(20.43)**	(22.75)**	(10.59)**	(5.51)**	(4.90)**	(4.93)**	(4.52)**	(22.64)**	(0.48)	(11.16)**
R^2	0.01	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	49,738	49,738	49,738	49,738	49,738	49,738	49,738	49,738	49,738	49,738	49,738

Note: This table shows how the crimes at ages 40–45 (in the columns) are predicted by crimes at ages 16–35 (in the rows). The estimations are predicted using the North Carolina Department of Public Safety dataset, which contains information on all individuals that have ever been sentenced in North Carolina. We use linear regressions in all cases. The sample for this model is limited to individuals who are at least 45 years old. * $p < 0.05$; ** $p < 0.01$.

Table E.5: NCDPS Regressions of Ages 40–45 on Ages 16–35, Males

	Miscellaneous	Fraud	Larceny	Vandalism	Auto Theft	Burglary	Robbery	Arson	Assault	Rape	Murder
Miscellaneous	0.065	-0.002	0.003	0.001	0.001	0.001	-0.000	0.000	0.003	-0.000	-0.001
	(40.59)**	(4.98)**	(6.29)**	(4.35)**	(4.68)**	(3.18)**	(2.61)**	(0.25)	(7.80)**	(2.60)**	(6.70)**
Fraud	0.032	0.009	0.004	-0.000	0.001	0.002	0.000	-0.000	0.000	-0.000	-0.000
	(16.63)**	(18.85)**	(6.51)**	(0.72)	(5.79)**	(5.19)**	(2.29)*	(0.83)	(0.84)	(1.28)	(0.78)
Larceny	0.029	0.001	0.055	0.002	0.003	0.005	0.002	-0.000	0.004	-0.000	0.000
	(9.25)**	(1.16)	(54.05)**	(6.64)**	(11.18)**	(9.45)**	(8.46)**	(2.55)*	(4.35)**	(0.64)	(0.70)
Vandal	0.003	-0.001	-0.009	0.007	-0.000	0.002	0.001	0.000	0.024	0.001	-0.000
	(0.26)	(0.46)	(2.68)**	(8.19)**	(0.04)	(1.05)	(1.02)	(1.29)	(8.44)**	(0.66)	(0.37)
Auto Theft	0.053	0.002	0.007	0.001	0.016	0.017	0.002	0.001	0.010	0.001	0.000
	(4.40)**	(0.76)	(1.82)	(0.57)	(14.05)**	(8.26)**	(1.85)	(1.36)	(3.12)**	(0.63)	(0.19)
Burglary	0.012	-0.002	0.015	0.001	0.003	0.025	0.003	0.000	0.003	0.000	-0.000
	(2.76)**	(1.27)	(10.20)**	(2.07)*	(6.28)**	(33.40)**	(6.14)**	(2.32)*	(2.77)**	(0.96)	(0.10)
Robbery	-0.008	-0.000	0.022	0.000	0.001	0.002	0.012	-0.000	0.003	0.000	0.000
	(1.10)	(0.07)	(9.41)**	(0.65)	(1.84)	(1.89)	(19.34)**	(0.27)	(1.68)	(0.27)	(0.70)
Arson	-0.020	-0.003	0.002	0.007	0.006	-0.000	-0.000	0.006	0.033	0.004	0.000
	(0.60)	(0.40)	(0.21)	(2.48)*	(1.85)	(0.03)	(0.14)	(6.06)**	(3.60)**	(1.04)	(0.08)
Assault	0.018	-0.004	-0.003	0.003	0.001	0.000	0.001	0.000	0.037	0.000	0.000
	(4.15)**	(3.24)**	(2.13)*	(7.43)**	(2.47)*	(0.48)	(3.43)**	(2.96)**	(31.56)**	(1.04)	(1.45)
Rape	-0.015	-0.003	-0.010	-0.000	-0.003	-0.004	0.000	-0.000	-0.001	0.009	-0.001
	(1.26)	(1.13)	(2.56)*	(0.49)	(2.38)*	(2.11)*	(0.13)	(1.36)	(0.31)	(7.53)**	(1.32)
Murder	-0.086	-0.005	-0.020	-0.002	-0.001	-0.008	-0.000	0.001	-0.003	-0.001	0.005
	(5.84)**	(1.26)	(4.17)**	(2.01)*	(0.89)	(3.38)**	(0.37)	(2.55)*	(0.64)	(0.44)	(5.35)**
constant	0.337	0.018	0.034	0.005	0.004	0.011	0.005	0.001	0.048	0.007	0.005
	(104.31)**	(21.78)**	(31.87)**	(17.23)**	(12.30)**	(20.05)**	(15.95)**	(8.16)**	(54.59)**	(18.92)**	(24.05)**
R^2	0.02	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00
N	188,556	188,556	188,556	188,556	188,556	188,556	188,556	188,556	188,556	188,556	188,556

Note: This table shows how the crimes at ages 40–45 (in the columns) are predicted by crimes at ages 16–35 (in the rows). The estimations are predicted using the North Carolina Department of Public Safety dataset, which contains information on all individuals that have ever been sentenced in North Carolina. We use linear regressions in all cases. The sample for this model is limited to individuals who are at least 45 years old. * $p < 0.05$; ** $p < 0.01$.

133

Table E.6: NCDPS Regressions of Ages 45–50 on Ages 16–35, Females

	Miscellaneous	Fraud	Larceny	Vandalism	Auto Theft	Burglary	Robbery	Arson	Assault	Rape	Murder
Miscellaneous	-0.009	0.000	-0.003	-0.001	-0.000	-0.000	-0.000	-0.000	-0.001	-0.000	-0.001
	(1.75)		(1.41)	(1.98)*	(0.48)	(0.16)	(0.58)	(1.11)	(0.70)	(0.36)	(1.71)
Fraud	0.012	0.000	0.000	0.000	-0.000	0.000	-0.000	-0.000	-0.001	-0.000	-0.000
	(5.55)**		(0.36)	(1.63)	(0.23)	(0.07)	(0.44)	(0.56)	(2.18)*	(0.13)	(0.76)
Larceny	0.006	0.000	0.037	-0.001	0.001	0.000	0.000	-0.000	0.001	-0.000	-0.000
	(1.28)		(21.43)**	(1.44)	(2.02)*	(1.11)	(2.27)*	(0.71)	(1.61)	(0.11)	(0.95)
Vandal	-0.043	0.000	-0.019	0.007	-0.001	-0.001	-0.001	0.001	0.019	-0.000	-0.002
	(1.34)		(1.52)	(2.53)*	(0.58)	(0.55)	(0.95)	(0.91)	(3.07)**	(0.05)	(0.78)
Auto Theft	0.009	0.000	-0.022	-0.002	-0.001	-0.001	-0.001	-0.000	0.010	-0.000	-0.001
	(0.16)		(1.07)	(0.51)	(0.34)	(0.33)	(0.44)	(0.13)	(0.99)	(0.02)	(0.29)
Burglary	-0.001	0.000	0.004	-0.002	-0.001	0.001	-0.001	-0.000	-0.013	-0.000	0.001
	(0.03)		(0.31)	(0.83)	(0.53)	(0.31)	(0.63)	(0.28)	(1.95)	(0.03)	(0.36)
Robbery	0.015	0.000	-0.004	0.001	-0.001	-0.001	0.004	-0.000	0.004	-0.000	0.000
	(0.48)		(0.30)	(0.27)	(0.51)	(0.50)	(2.58)**	(0.31)	(0.66)	(0.04)	(0.20)
Arson	-0.032	0.000	-0.027	-0.003	-0.001	-0.001	-0.001	-0.000	0.000	-0.000	0.008
	(0.57)		(1.23)	(0.59)	(0.22)	(0.24)	(0.25)	(0.24)	(0.01)	(0.04)	(2.29)*
Assault	0.007	0.000	-0.009	0.005	0.000	-0.000	0.001	0.000	0.018	-0.000	0.000
	(0.50)		(1.72)	(4.13)**	(0.11)	(0.36)	(2.29)*	(0.36)	(6.56)**	(0.11)	(0.12)
Rape	0.130	0.000	0.000	-0.002	-0.001	-0.001	-0.001	-0.000	0.038	-0.000	-0.002
	(1.02)		(0.01)	(0.17)	(0.08)	(0.10)	(0.12)	(0.08)	(1.54)	(0.02)	(0.24)
Murder	-0.129	0.000	-0.023	0.011	0.004	0.002	-0.001	0.002	0.014	-0.000	-0.002
	(4.07)**		(1.95)	(3.99)**	(1.94)	(0.90)	(0.52)	(2.09)*	(2.35)*	(0.12)	(1.02)
constant	0.264	0.000	0.037	0.003	0.001	0.001	0.001	0.001	0.017	0.000	0.003
	(54.62)**		(20.07)**	(7.64)**	(4.01)**	(4.65)**	(3.15)**	(4.38)**	(18.65)**	(1.12)	(8.69)**
R^2	0.00	.	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	35,432	35,432	35,432	35,432	35,432	35,432	35,432	35,432	35,432	35,432	35,432

Note: This table shows how the crimes at ages 45–50 (in the columns) are predicted by crimes at ages 16–35 (in the rows). The estimations are predicted using the North Carolina Department of Public Safety dataset, which contains information on all individuals that have ever been sentenced in North Carolina. We use linear regressions in all cases. The sample for this model is limited to individuals who are at least 50 years old. * $p < 0.05$; ** $p < 0.01$.

Table E.7: NCDPS Regressions from Ages 45–50 on Ages 16–35, Males

	Miscellaneous	Fraud	Larceny	Vandalism	Auto Theft	Burglary	Robbery	Arson	Assault	Rape	Murder
Miscellaneous	0.037	0.000	0.003	0.000	0.000	0.001	0.000	-0.000	0.001	-0.001	-0.001
	(19.83)**		(4.07)**	(2.77)**	(1.77)	(2.02)*	(0.64)	(0.23)	(2.40)*	(3.15)**	(4.97)**
Fraud	0.020	0.000	0.005	-0.000	0.000	0.002	-0.000	0.000	-0.000	0.000	-0.000
	(10.46)**		(8.12)**	(0.04)	(1.44)	(5.94)**	(0.68)	(0.97)	(0.34)	(0.05)	(1.14)
Larceny	0.023	0.000	0.038	-0.000	0.001	0.004	0.001	0.000	0.002	0.000	-0.000
	(7.04)**		(34.47)**	(0.33)	(3.21)**	(7.72)**	(2.74)**	(0.57)	(2.71)**	(0.72)	(1.91)
Vandal	-0.000	0.000	-0.008	0.007	0.000	0.002	0.002	0.000	0.016	-0.001	-0.001
	(0.01)		(2.12)*	(7.81)**	(0.00)	(0.97)	(1.87)	(0.17)	(5.00)**	(0.39)	(0.77)
Auto Theft	0.016	0.000	0.014	0.001	0.015	0.003	0.003	-0.000	0.004	-0.002	-0.000
	(1.07)		(2.76)**	(1.25)	(10.20)**	(1.12)	(2.24)*	(0.99)	(0.88)	(0.94)	(0.34)
Burglary	0.003	0.000	0.013	0.000	0.004	0.014	0.002	0.000	0.001	0.000	-0.000
	(0.57)		(8.34)**	(0.08)	(8.14)**	(19.75)**	(4.39)**	(2.81)**	(0.72)	(0.02)	(0.26)
Robbery	-0.017	0.000	0.013	0.001	0.000	0.001	0.008	-0.000	0.005	-0.001	-0.000
	(2.33)*		(5.15)**	(1.36)	(0.08)	(0.90)	(12.97)**	(1.90)	(2.27)*	(0.81)	(0.63)
Arson	-0.036	0.000	-0.012	0.012	-0.002	-0.003	-0.003	0.004	0.027	-0.004	-0.002
	(1.04)		(1.02)	(4.70)**	(0.59)	(0.52)	(0.83)	(3.73)**	(2.82)**	(0.92)	(0.98)
Assault	-0.004	0.000	-0.001	0.002	-0.001	-0.002	0.000	0.000	0.026	0.000	-0.000
	(0.88)		(0.72)	(6.45)**	(1.28)	(2.32)*	(0.06)	(0.56)	(19.80)**	(0.72)	(0.18)
Rape	-0.025	0.000	-0.001	-0.000	-0.001	-0.001	0.001	0.000	-0.001	0.010	-0.001
	(1.94)		(0.25)	(0.11)	(0.91)	(0.30)	(0.99)	(0.04)	(0.40)	(6.70)**	(1.25)
Murder	-0.068	0.000	-0.015	-0.001	-0.001	-0.004	-0.001	-0.000	-0.002	-0.000	0.000
	(4.82)**		(3.19)**	(1.38)	(0.83)	(1.81)	(1.03)	(0.93)	(0.42)	(0.15)	(0.42)
constant	0.318	0.000	0.026	0.003	0.003	0.008	0.003	0.001	0.041	0.006	0.004
	(103.46)**		(25.22)**	(14.03)**	(10.92)**	(16.47)**	(10.00)**	(7.13)**	(48.64)**	(16.51)**	(21.79)**
R^2	0.01	.	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	147,478	147,478	147,478	147,478	147,478	147,478	147,478	147,478	147,478	147,478	147,478

Note: This table shows how the crimes at ages 45–50 (in the columns) are predicted by crimes at ages 16–35 (in the rows). The estimations are predicted using the North Carolina Department of Public Safety dataset, which contains information on all individuals that have ever been sentenced in North Carolina. We use linear regressions in all cases. The sample for this model is limited to individuals who are at least 50 years old. * $p < 0.05$; ** $p < 0.01$.

135

We repeat the process just described separately for males and females. We use linear forecast models, but replace negative forecasted values of crime by zero. Table E.2 through Table E.7 show the estimated models. As expected, generally the most important prediction factor for a crime is the number of occurrences of the same crime type in a previous period. The coefficients in some cases are substantial, which implies that considering forecasted crimes is an important part of an assessment of the crime benefits of the program.

This procedure gives a forecast of the number of sentences that the ABC/CARE subjects will receive after age 34. From the forecasted number of sentences, we forecast the number of arrests up to age 50. Figure E.2 shows the additional estimates of criminal activity generated by our forecasting methodology. The effect is quantitatively much larger for arrests than for sentences. For both arrests and sentences, including forecast crimes, this effect adds 30-50% more crimes to our previous totals. The predictions are roughly proportional to the previous crimes, as discussed before.

Figure E.2: Constructed Forecasts
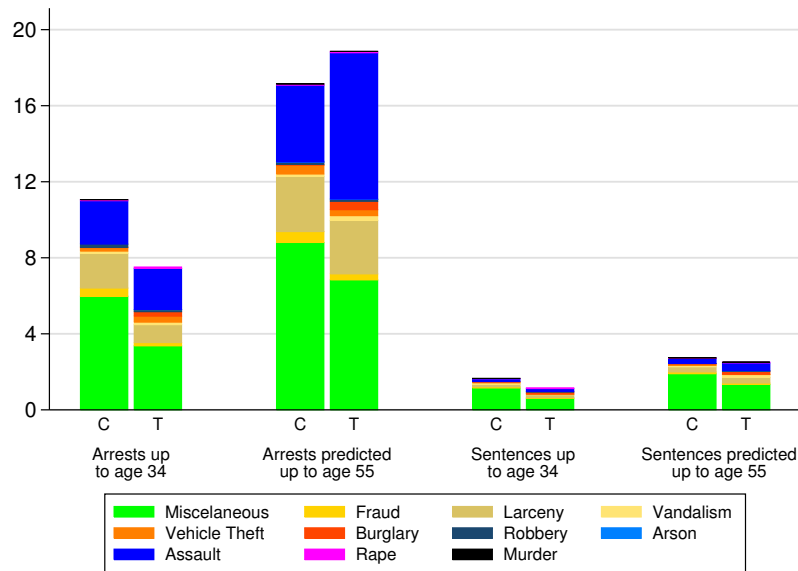
Note: This figure continues Figure E.1. It shows, for the control (C) and treatment (T) groups, the effects of adding forecasts. The first pair of columns is the same as the fifth pair of columns in Figure E.1. The second pair of columns includes the arrests that we forecast. The third and fourth pairs of columns are the analogous pairs for sentences.

136

E.3 Victimization Inflation

Even though we have administrative data on crimes, we only observe the crimes that had criminal justice system consequences (arrests or sentences). However, it is possible that the subjects committed more crimes than what we observe. Victimization Inflation (VI) is a method to capture benefits in crime reduction for crimes that did not result in justice system consequences.[93] For most types of crimes in the U.S., there are many more victims than arrests or sentences. Using arrests as an example, VI assumes that those "unpunished crimes" were committed by the same people who were arrested for crimes of the same type, and in the same proportion. The calculation of VI uses as an input the national ratios of total number of reported crimes over the number of arrests. VI assumes that those national ratios are also valid for each individual. Under those assumptions, it is possible to find the total number of crimes committed by a subject for a given type of crime as the total number of arrests for that type of crime multiplied by the estimated national ratios for that type of crime. We estimate the total number of victims using two methods, one based on arrests and one based on sentences. Given that the "unpunished" crimes are by definition unobserved, it is not straightforward to use a data-driven method to allocate them between those subjects with arrests, those with sentences, and those with neither arrests nor sentences. We calculate separate estimates for arrests and sentences and use the average of those estimates as our main estimate.

E.3.1 Construction of the Total Number of Victims in the U.S.

The numerator of the VI ratio is an estimate of the total number of crimes of a certain crime type committed in the U.S. We construct this estimate using two datasets. First, we use the National Crime and Victimization Survey (NCVS). It has self-reported data on victimization

---

[93]Belfield et al. (2006); Heckman et al. (2010).

of crimes reported on the household level. The data are available from 1995 to 2012. We also use the Uniform Crime Reporting Statistics (UCRS), which contains all crimes committed against households, individuals, and businesses that are reported to the police. These data are available from 1960 to 2013. Given that these two datasets independently underestimate the total number of crimes, but likely have significant overlap between them, we choose the largest estimate over both datasets for each type of crime. We refer to this estimate, $\overline{V_{j,t}}$, as the total number of victims in the country for type of crime j in year t.

E.3.2 Construction of the Total Number of Arrests in the U.S.

The denominator of the VI ratio is an estimate of the total number of arrests of a certain type committed in the U.S. We have data from the "National Arrests Analysis Tool" of the National Bureau of Justice Statistics. These data are available from 1980 to 2012, which spans the years of all crimes that we observe in the ABC data. There is one problem with this dataset that we consider relatively minor: not all law enforcement agencies report the number of crimes (there are dozens of agencies that can legally arrest in the U.S.). However, as a large majority of them report the numbers of crimes, and because we are using national estimates, this should not greatly affect our calculations. We use these data to create $\overline{A_{j,t}}$, the total number of arrests in the country for type of crime j in year t.

E.3.3 Victimization Inflation Factors

Figure E.3 and Figure E.4 show the VI factors calculated by year. The ratios in the charts are constructed as $r_{j,t} = \frac{\overline{V_{j,t}}}{\overline{A_{j,t}}}$. In practice, we use an average of all the yearly measures in our calculations given that this exercise imputes unobserved crimes that do not have a clearly defined date. This average of all the yearly measures is given by $r_j = \Sigma_{t=t_0}^{T} r_{j,t}/T$, and has more precise estimates of the ratio. The VI factors we use for sentences are equal to the

factors used for arrests, multiplied by the arrest-sentence ratios discussed above. Below, we discuss combining these different estimates as per our arrest-based and sentence-based methodology. For sentences, we have data from the National Judicial Reporting Program (NJRP). These data are available from 1986 onwards. Using this dataset, we construct $\overline{S_{j,t}}$, the total number of sentences in the country for type of crime j in year t.

Figure E.3: Victim-arrest Ratios by Crime
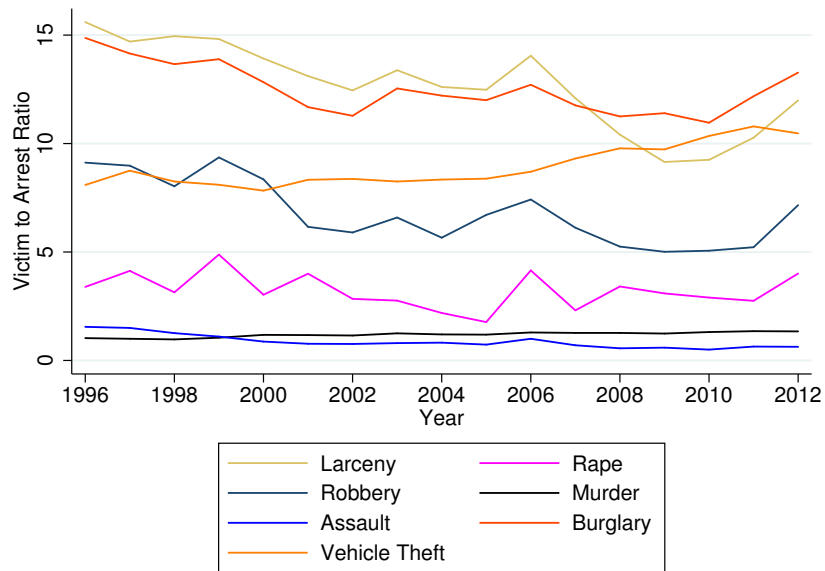
Note: This figure shows, by year and type of crime, the number of victims (estimated from the NCVS and the UCRS datasets) divided by the number of arrests (estimated from the National Arrests Analysis Tool from the NBJS). In practice, we use a single number for each type of crime, which is an average across years.

Figure E.4: Arrest-sentence Ratio by Crime
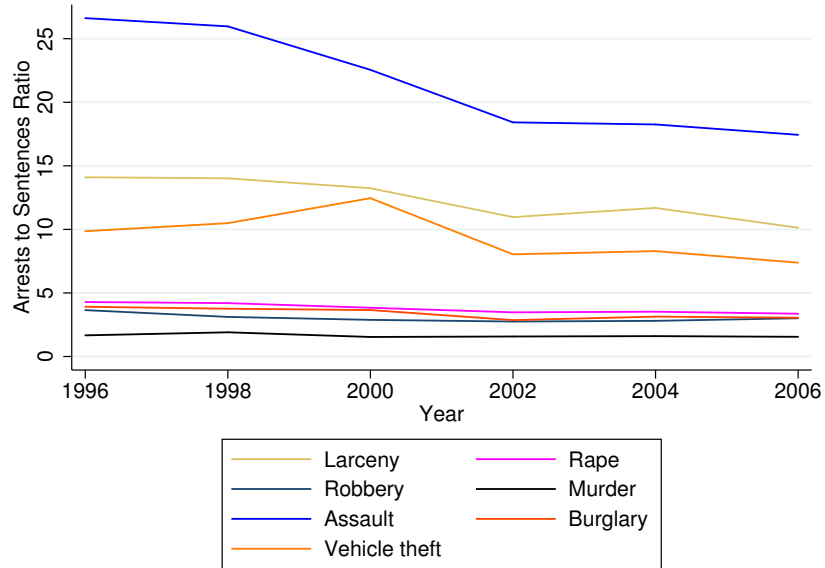
Note: This figure shows, by year and type of crime, the number of arrests (estimated from the National Arrests Analysis Tool from the NBJS) divided by the number of sentences (estimated from the National Justice Reporting Program). In practice, we use a single number for each type of crime, which is an average across years.

E.3.4 Effects on Number of Crimes, After Victimization Inflation

Figure E.5 shows the effects of VI on our estimates of the number of crimes committed. Note that the magnitudes in the axis are much larger than those of previous charts. The largest effects are for larceny, which is common in the data and has a victim-arrest factor of 12.6, the largest factor of all the categories of crime used in the paper. Given that the victim cost of larcenies is low, it affects the estimates less than what this chart suggests.

Figure E.5: Effects on Number of Crimes, After Victimization Inflation
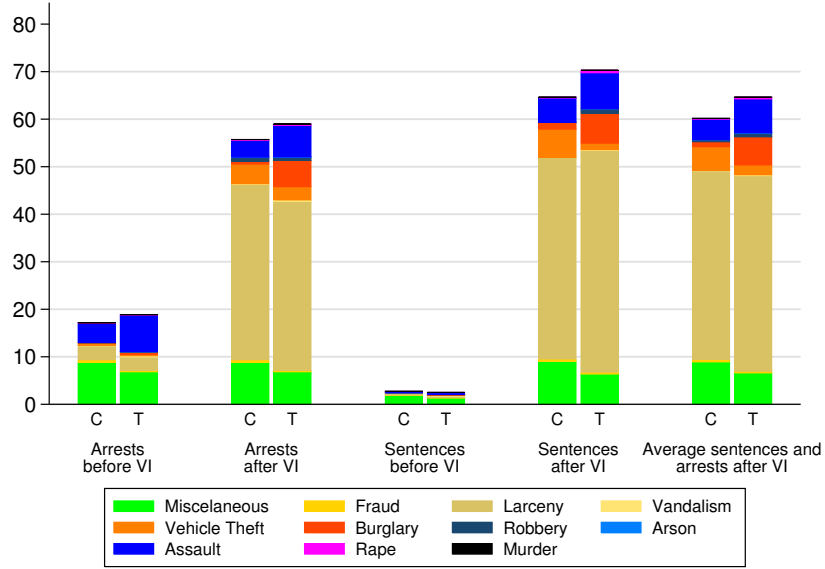
Note: This chart continues Figure E.2. It shows, for the control (C) and treatment (T) groups, the effects of adding victimization inflation (VI). The first pair of columns is the same as the second pair of columns in Figure E.2. The second pair of columns expand the arrests to account for VI. The third pair of columns is the same as the fourth pair of columns in Figure E.2. The fourth pair of columns expand the sentences to account for VI. The last pair of columns averages the second and the fourth pairs of columns in this chart.

While the assumptions required for the victimization inflation methodology are strong, we argue that this is the best approximation for a total toll of crime's costs. The highest victim-arrest ratio shown in the figures are sensible and are not for the most costly categories of crime in the data, which stabilizes the estimates.

E.4   Literature on Costs of Specific Crimes

There are many methods for estimating the unit costs of specific crimes.[94] In this document, we only review the literature related to the inputs necessary for this paper.

We start by classifying the costs of crime. We then present the two general types of method-

---

[94]Cohen and Bowles (2010) and McCollister et al. (2010) give comprehensive reviews of the state of the literature.

ologies that are used to estimate the total costs of crime: the Top-down methodologies and the Bottom-up methodologies. The former attempts to quantify the value that people put into *ex-ante* prevention of crime, while the latter attempts to gather *ex-post* all sources of costs that crime generates. The difference between these two methods can be large. Cohen and Bowles (2010) show that for the particular case of estimates of the cost of rape, the top-down approach gives a value that is twice as large as the value given by most bottom-up studies. Other studies give cost estimates between the estimates from these two approaches.

E.4.1 Classifying the Costs of Crime

Some methodologies used to estimate costs of crime are only able to capture some types of costs. Some important types of costs are:

- Costs to the victim that can be directly quantified, such as medical bills, property losses, and lost productivity.

- Costs to the victim that cannot be observed, such as pain and suffering.

- Costs to the community in terms of prevention of crime, such as alarms, avoidance behavior, and police presence.

- Costs to the community in terms of fear.

- Costs to the community in terms of the criminal justice system, especially imprisonment.

- Costs to the offender in terms of lowered productivity, such as forgone wages.

E.4.2 Bottom-up (BU) Methodologies

These approaches sum each type of cost that is imposed after the crime has been committed. The most well-known studies combine direct (also known as tangible) costs of the crimes with intangible costs. Tangible costs are everything that can be directly measured by observation, such as foregone wages, hospital costs, and police expenditure. Intangible costs are subjective (e.g., pain and suffering). One way to measure these costs is using jury awards. For example, a jury award given as a result of an arm broken at a construction site can be used as a proxy of the intangible cost of having an arm broken in an assault.[95] The problem of these approaches is that many of the costs of crime are not directly imposed on the victim and are hard to quantify, such as the "fear of crime," the increased expenditure on crime prevention, and the negative impact of imprisonment on the community.

E.4.3 Top-down (TD) Methodologies

The other way to estimate the cost of crime is using TD methods, based on eliciting willingness to pay for avoiding crimes. The main advantage of these methods is that, in principle, they consider costs that are hard or impossible to measure directly, such as the cost of fear, avoidance behavior, and expenditures in preventative measures. There are three main methodologies for this approach, which we now briefly describe.

1. **Stated Preferences.** This basic method elicits the willingness to pay for hypothetical programs that would reduce crime nationwide for a sample of people.[96] It is assumed that the costs obtained by this method include factors that affect the community, and that are hard to capture, such as fear. However, it is unclear whether people consider factors like the cost of the justice system in their answers to these questions. An

---

[95]This was first used in Cohen (1988), and has been extensively used in BU studies after that. Miller et al. (1996) improved on previous estimates by using jury awards specifically coming from criminal cases.
[96]Cohen et al. (2004).

obvious drawback of these surveys of responses to hypothetical events is that people might not answer the true amount they would be willing to pay in actual transactions.

2. **Revealed Preferences.** This method infers the value that individuals assign to crime reductions from market transactions. The most standard way to calculate these estimations is running regressions to explain the total price of houses with several factors, including the rates of crime in the area. Those parameters associated with the crime rate are considered the revealed valuation of avoiding crimes.[97] One weakness of this method is that it assumes that people are well-informed on the crime rates in an area. Another problem is that, in absence of large, detailed data on crimes and housing prices, it is not possible to separately identify the costs of different types of crimes. To the best of our knowledge, no paper has yet been able to convincingly obtain estimates per type of crime with this method.

3. **Life Satisfaction.** For this method, people are surveyed about their preferences between different life conditions, in which several different factors are considered. Some of those factors are income and rates of crime. By doing so, people implicitly associate monetary values to the levels of crime in the communities they would live in.[98]

E.4.4 Costs Used in this Study

To summarize, both approaches have strengths and weaknesses: the TD approaches are more likely to reflect costs to the community (e.g. fear and anxiety, avoidance behavior, and protective measures) and better capture the goals of a prevention program. However, in practice TD estimates rely on strong assumptions, and there are methodological issues associated with obtaining detailed values for the different types of crimes. It is also possible that when people answer the surveys used to make TD calculations they include some costs

---

[97]See, e.g., Thaler (1978).

[98]Moore and Shepard (2006); Moore (2006).

that we are including separately, such as justice system costs, and risk of death from non-murder crimes, while BU does not include them. Given those considerations, and the lack of TD costs for some categories of crime, we use BU costs for our main estimates. However, we present cost estimates using both approaches to explore the sensitivity of our estimates to the choice of the measures of the cost of crime. We choose Cohen et al. (2004) as representative of the TD approaches, and McCollister et al. (2010) as representative of the BU approaches. In terms of timing, both of these studies match well with the ABC/CARE data. The bulk of crimes in the ABC/CARE data occurred between the late 1990s and early 2000s. While Cohen et al. (2004) do not report the exact year of their survey, they use Census 2000 figures for their estimates. Even though McCollister et al. (2010) is a more recent study, many of the productivity estimates that their costs are based on are taken from papers using data from years with more crimes the late 1990s and early 2000s. The costs in those studies are presented in Table E.8. Notice that there are some strong differences in the cost of crimes, such as assault, burglary, and especially robbery.

Table E.8: Unitary Costs of Crime for Victims

Crime	Top-Down Approach Cohen et al. (2004)	Bottom-Up Approach McCollister et al. (2010)
Arson		12,093
Assault	95,200	16,132
Burglary	34,000	1,467
Fraud		0
Larceny		528
Motor Vehicle Theft		6,699
Murder	13,192,000	9,286,200
Rape	322,320	224,021
Robbery	315,520	7,273
Vandalism		0

Note: All amounts are in 2014 USD. The amounts reported in McCollister et al. (2010) for non-murder crimes have the extra cost for risk of death and the cost of a crime career removed (both were obtained from correspondence with the author). Risk-of-death costs do not apply, because we know the outcomes of the crimes. Crime-career costs do not apply, as we directly observe the income of the individuals. These costs also don't include police and legal system costs, as those are imputed separately and only for the cases for which individuals were arrested or sentenced.

E.4.5 Timing of Effects: Incidence vs. Prevalence

We would like to discount the costs of crime according to whether they were *incurred* during a particular age of ABC/CARE subjects, because those values should be discounted at a different rate than costs incurred later, even if both costs were *imposed* in the same year. Thus, the value of the imprisonment is discounted year-by-year. We have no information about the timing of costs for victims, so the value of the different crimes for the victims are discounted according to the time they were imposed (the time of the crime's occurrence).

E.4.6   Costs of Imprisonment

Unlike previous studies, we observe the sentences of the ABC/CARE subjects. This allows for a precise estimation of the costs of imprisonment. For the cost of jail and state prison, we use estimates from the U.S. Department of Justice (1988). In 2014 USD, these costs are $25,338 for a year in a state prison, and $21,939 for a year in jail. It is important to clarify that we only include costs of the justice system for the crimes that are known by the justice system, not for the crimes that we impute through victimization inflation.

E.5   Effects on Costs of Crime

E.5.1   Effects on Costs Before Victimization Inflation

Figure E.6 presents the estimated costs per type of cost before victimization inflation. There are clear positive effects for the treatment group in terms of reductions in the costs of crime. Those reductions are almost exclusively given by the large effect of the murder case we observe in the control group (note that murder also figures into the treatment group costs because of the forecasts). Comparing the bars in this figure, the costs from the justice system and from imprisonment are low compared with the victimization costs, even without victimization inflation. While the levels of the arrest-based estimates are higher than the levels of the sentence-based estimates for both the treatment and control groups, the impacts of the program are quite similar across both methods (Figure E.6).

Figure E.6: Costs of Crime Before Victimization Inflation
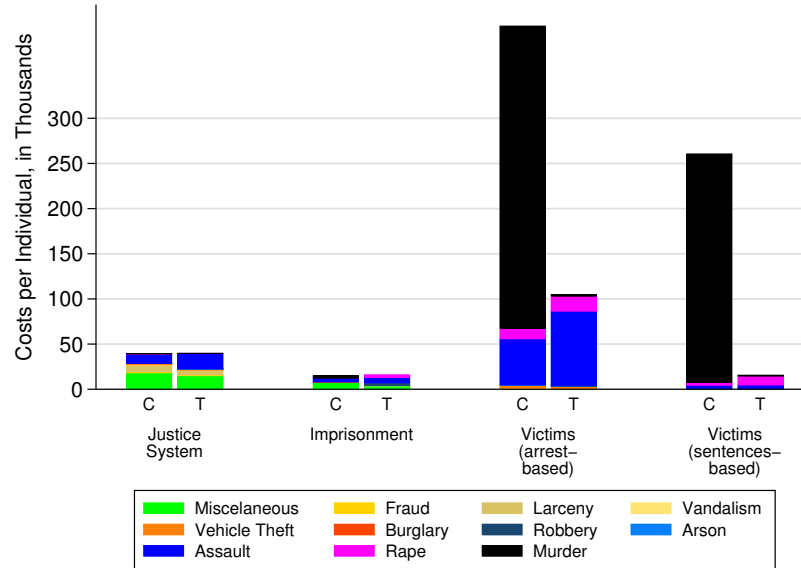
Note: This figure depicts the per capita cost for the different categories of costs and crimes we use, by control (C) and treatment (T). The first pair of columns adds up the justice system costs (including police) for all arrests inputed for each subject. The second pair of columns adds up the cost of Imprisonment. It is important to note that the costs are per capita, so there are individual cases that have much higher costs. The next two pairs of columns show the pre-victimization inflation estimates of number of crimes multiplied by the individual victim cost of the different crimes. The costs are taken from the Bottom-up approach in Table E.8. All costs are in thousands of 2014 USD.

E.5.2 Effects on Costs After Victimization Inflation

Figure E.7 presents the data after applying the victimization inflation. As shown below, the inflation allows us to include a substantial amount of crime that otherwise would not have been considered. This chart shows that the treatment effects using arrest-based estimations are not substantively different from the ones using sentence-based estimations. In order to use all available information, we use the estimates based on the averages of the two for our analysis.

Figure E.7: Costs of Crime After Victimization Inflation
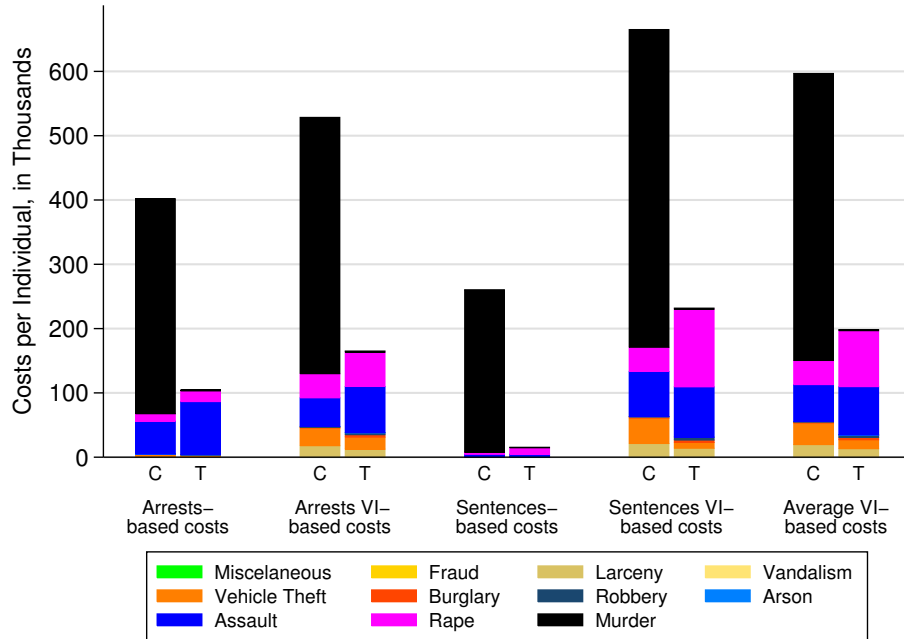
Note: This figure depicts the per capita cost for the different categories of costs and crimes we use, by control (C) and treatment (T). The first two pairs of columns adds up all of the arrest-based costs for each subject and compares the pre- and post-victimization inflation costs. The third and fourth pair of columns compare the pre- and post-victimization inflation sentence-based costs. It is important to note that the costs are per capita, so there are individual cases that have much higher costs. The next two pairs of columns show the post-victimization inflation estimates of number of crimes multiplied by the individual victim cost of the different crimes. The costs are taken from the Bottom-up approach in Table E.8. All costs are in thousands of 2014 USD.

We consider the impact on murder as a consequence of the program rather than a statistical coincidence. We use as precedent the cost-benefit analysis of Perry in which three control group individuals and one treatment group individual committed murders.[99]

Some of the sources of cost estimates, such as the more serious crimes, result in volatile estimates due to the small sample sizes. Our estimations of the standard errors associated with the objects of interest in this paper—the present value of the program and the internal rate of return—consider those sources of volatility. Ultimately, the benefit-cost analysis is a unidimensional summary of benefits of a program, and specific flows of benefits with high

---

[99]Heckman et al. (2010).

variability enter naturally into the process of aggregation.

E.6   Sensitivity Analyses Using Alternative Cost Estimates

Thus far, we have examined several ways to construct estimates of the costs of crime:

1. We show that not matching the different crime datasets could reduce the number of crimes, but that the general patterns are stable. No dataset is especially influential.

2. We show that not including forecasts up to age 50 noticeably reduces the number of crimes, but the general patterns are not modified.

3. We show that estimates using arrest-based estimations versus sentence-based estimations show large differences in terms of the number of crimes before victimization inflation. However, they are small after adjusting for this, and do not substantially change the total benefits calculations.

In Figure E.8, we present additional perturbations of our main estimates. In particular, we show how the estimates change when three different cost schedules are used: (i) Top-down costs, (ii) Bottom-up costs, and (iii) Bottom-up costs assuming that the costs of murders and rapes are identical to the cost of an assault. We also note that BU costs are a "conservative" option in the sense that the effects of the program are higher using TD costs. We can also see that with no murders and rapes, the effect of the program on crime is still positive, but much smaller than when those crimes are considered.

Figure E.8: Different Cost Schedules
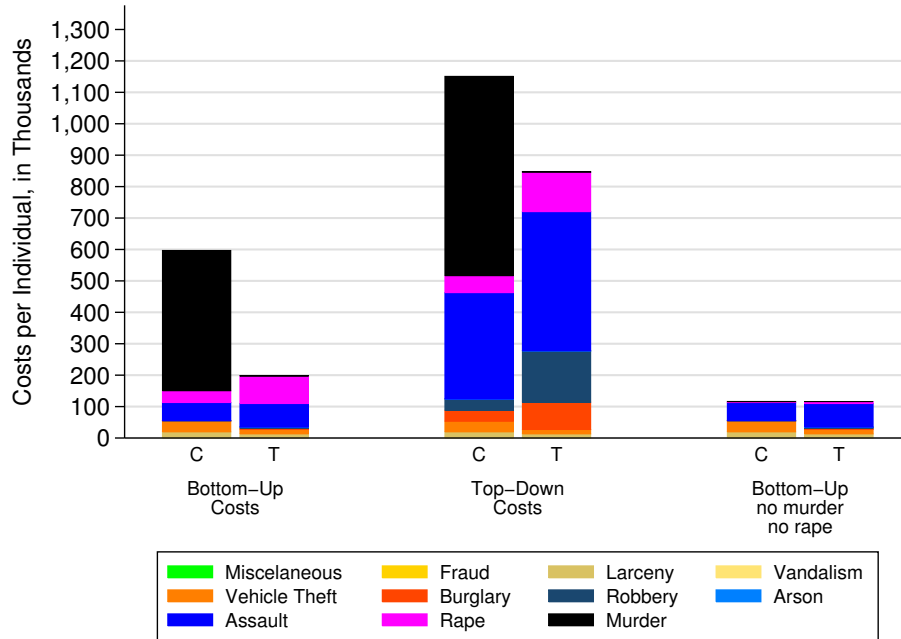
Note: This figure depicts the per capita cost for the different categories of costs and crimes we use, by control (C) and treatment (T). It presents some sensitivity analyses. The first pair of bars represents costs using the Bottom-up approach in Table E.8 to determine individual costs of crime. The second pair of bars represents costs using the Top-down approach in Table E.8 to determine individual costs of crime. The third pair of bars uses the Bottom-up approach, but replaces the values of murders and rapes with that of assaults.

Finally, in Figure E.9, we present the effect that discounting has on our estimates that have been adjusted for deadweight loss. The first pair of bars represents the deadweight loss-adjusted cost estimates that are not discounted, and the second pair of bars represents the deadweight loss-adjusted costs that have been discounted. It is clear that the effect of discounting is substantial, approximately halving the total cost estimates.

Figure E.9: Effect of Discounting Crime Costs
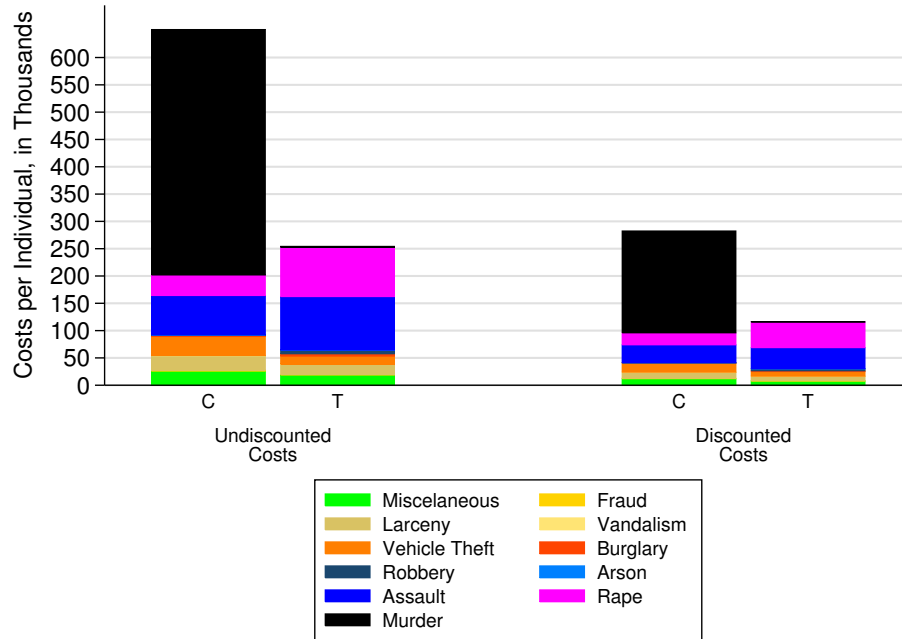
Note: This figure depicts the per capita cost for the different categories of costs and crimes we use, by control (C) and treatment (T). The discounted costs use 3% as a discount factor, and are discounted to birth. The deadweight loss (DWL) adjustments increases the costs from justice system and imprisonment by 50%. All costs are in thousands of 2014 USD.

F   Accounting for Alternatives Chosen by the Control Group

Our main analysis so far compares the treatment to the control group, irrespective of the choice of childcare chosen by the families of the control-group children. This exercise compares treatment to the "next best" alternative that the parents of the control-group children elected. As we document in Section 2 of this paper, almost 75% of the control-group children were enrolled in an alternative.

In this Appendix we disaggregate our estimates by computing social efficiency measures for ABC/CARE relative to (a) alternative (low-quality) formal childcare options and (b) staying at home. We refine our analysis to draw two further comparisons: (i) ABC/CARE compared to no treatment at all (so that the control child stayed at home); and (ii) ABC/CARE compared alternative preschool arrangements, which we describe in Section 2.[100]

All treatment-group children have the same exposure. We simplify the analysis of the controls by creating two categories. "H" indicates that the control-group child is in home care throughout the entire length of the program. "C" indicates that a control-group child is in alternative center childcare for any amount of time.[101] We thus compress a complex reality into two counterfactual outcome states at age a for control-group subjects:

$$Y_{a,H}^0 \quad : \quad \textbf{Subject received home care exclusively}$$

$$Y_{a,C}^0 \quad : \quad \textbf{Subject received some alternative childcare}.$$

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

---

[100]García et al. (2018) discuss this issue in depth.

[101]In a companion paper, García et al. (2018), we show that this assumption is innocuous.

control status is

$$\boldsymbol{Y}_a^0 := (1 - V) \, \boldsymbol{Y}_{a,H}^0 + (V) \, \boldsymbol{Y}_{a,C}^0. \tag{19}$$

It is fruitful to assess 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:

$$\boldsymbol{\Delta}_a \left(\textbf{Fix } V = 0 \right) := \mathbb{E} \left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_a^0 | \textbf{Fix } V = 0, W = 1 \right] = \mathbb{E} \left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_{a,H}^0 | \textbf{Fix } V = 0, \boldsymbol{B} \in \mathcal{B}_0 \right]. \tag{20}$$

where $\textbf{Fix } V$ refers to fixing V rather than just conditioning on it.[102]

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

$$\boldsymbol{\Delta}_a \left(\textbf{Fix } V = 1 \right) := \mathbb{E} \left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_a^0 | \textbf{Fix } V = 1, W = 1 \right] = \mathbb{E} \left[\boldsymbol{Y}_a^1 - \boldsymbol{Y}_{a,C}^0 | \textbf{Fix } V = 1, \boldsymbol{B} \in \mathcal{B}_0 \right]. \tag{21}$$

Random assignment to treatment does not directly identify (20) or (21). Econometric methods are required to identify these parameters. In this paper we rely on matching (conditioning on observables) to control for selection into the home or take-up of alternatives. In García et al. (2018), we document that alternative approaches to selection produce similar estimates. We combine the matching strategies with the methodology in Section 3 to produce forecasts and provide measures of social efficiency refining the counterfactual comparison for the control-group children.

Figure F.1 summarizes our estimates. There are substantial differences between males and females in one counterfactual: treatment vs. alternative preschools. The estimated treatment

---

[102]For the distinction between fixing and conditioning, see Heckman and Pinto (2015).

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) fathers are more likely to remain at home with the mother if the child is a boy and such families have more resources, diminishing the treatment effect for ABC/CARE compared to the home alternative, and (ii) boys are more adversely affected by low quality formal childcare environments. The formal alternative childcare centers are of generally low quality (see García et al., 2018).

For females, there is a gain of 4.93 dollars per dollar invested compared to staying at home. 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.

Figure F.1: Life-cycle Net Present Value of Main Components of the CBA

(a) Males
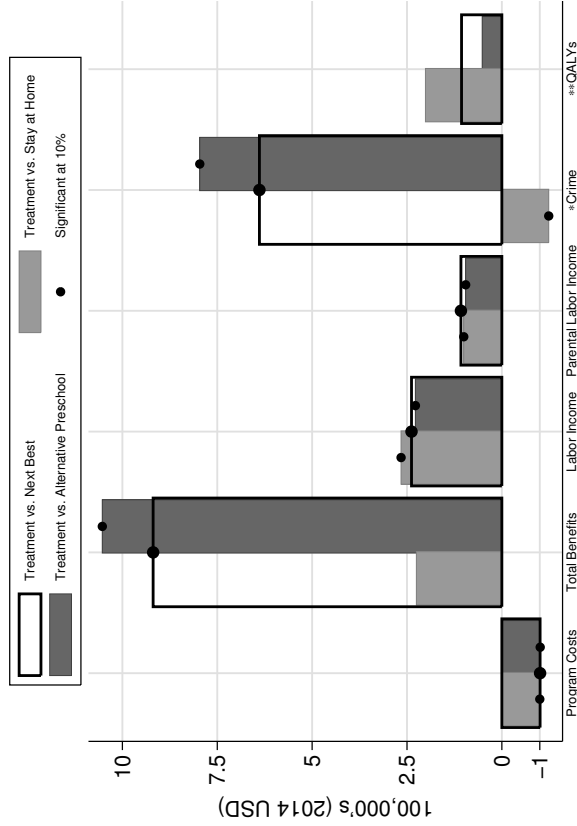
(b) Females
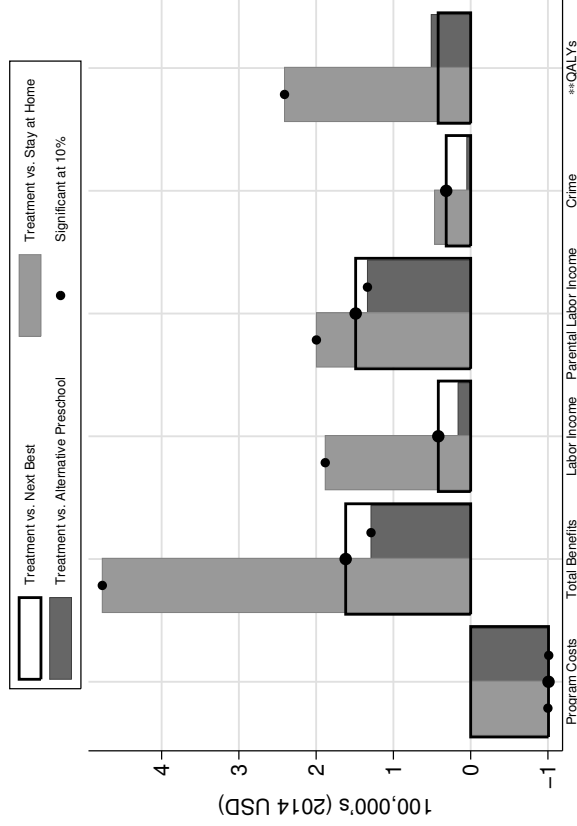
Note: This figure displays the life-cycle net present values of the main components of the cost/benefit analysis of ABC/CARE from birth to forecasted death, discounted to birth at a rate of 3%. "Treatment vs. Next Best": compares the treatment to the control group. "Treatment vs. Stay at Home": compares the treatment group to those subjects who stayed at home. "Treatment vs. Alternative Preschool": compares the treatment group to those subjects who attended alternative preschools. The latter two are based on matching estimators that account for selection on observable variables. By "net" we mean that each component represents the total value for the treatment group minus the total value for the control group. Program costs: the total cost of ABC/CARE, including the welfare cost of taxes to finance it. Total net benefits: are for *all* of the components we consider. Labor income: total individual labor income from ages 20 to the retirement of program participants (assumed to be age 67). Parental labor income: total parental labor income of the parents of the participants from when the participants were ages 1.5 to 21. Crime: the total cost of crime (judicial and victimization costs). To simplify the display, the following components are not shown in the figure: (i) cost of alternative preschool paid by the parents of control group children; (ii) the social welfare costs of transfer income from the government; (iii) disability benefits and social security claims; (iv) costs of increased individual and maternal education (including special education and grade retention); (v) total medical public and private costs. Inference is based on non-parametric, one-sided *p*-values from the empirical bootstrap distribution. We indicate point estimates significant at the 10% level.

*The treatment *vs.* stay at home net present value is sizable and negative (-$123,498.2); its standard error is $62,745.72.
**QALYs refers to the quality-adjusted life years. Any gain corresponds to better health conditions until forecasted death, with $150,000 (2014 USD) as the base value for a year of life.

156

G   Forecasting Health Outcomes Using the Future Adult Model (FAM)

In this paper, we also forecast and monetize health outcomes for treatments and controls. When studying labor income, a first-order autoregressive model suffices to make valid forecasts. When forecasting health, a richer model conditioning on multiple health conditions is required to make valid forecasts. We use a dynamic microsimulation model (the Future Adult Model - FAM) as a vehicle for forecasting health-state occupancy probabilities and associated benefits and costs. Using this tool, we forecast the health trajectories of the ABC/CARE individuals. To forecast health, we use a comprehensive ABC/CARE age-30 health interview and a health follow-up conducted when the subjects were in their mid-30s to initialize the forecast of their health trajectories. We then monetize health outcomes using quality-adjusted life years and medical costs.[103] Before providing formal details, we summarize the data sets used, as well as variable construction and imputation assumptions.

G.1   Data Sources

FAM uses data from ABC/CARE surveys to set the initial state of the cohort. The state-occupancy model parameters are estimated from the 1997 to 2013 waves of the Panel Study of Income Dynamics (PSID). We supplement the PSID with data from the Health and

---

[103]The microsimulation model that we use is an extension of the model used by Prados et al. (2015). Technical details are described in Tysinger et al. (2015). Both models are related to the Future Elderly Model (FEM), which is a microsimulation tool originally developed to examine the health and health care costs among the elderly Medicare population (Goldman et al., 2004). It has been used extensively to assess health and disease prevention scenarios: FEM has been used to assess the future costs of disease, the benefits of preventing disease among older population, the consequences of new medical technologies, trends in disability, and the fiscal consequences of worsening population health (see Goldman et al. (2004), Lakdawalla et al. (2004), Goldman et al. (2005), and Zissimopoulos et al. (2014)). The main differences of FAM with FEM are that the model we use starts with cohorts of individuals at age 30 instead of 50, and that it simulates more outcomes than FEM. These additional outcomes are useful in explaining health outcomes and medical expenditure at younger ages. For example, we analyze the evolution of partnership and marital status, work status, and family size.

Retirement Study (HRS). We use the National Health and Nutrition Examination Survey (NHANES) to account for differences between measured and self-reported BMI. To estimate medical care costs associated with health conditions, we use the Medical Expenditures Panel Survey (MEPS) and the Medicare Current Beneficiaries Survey (MCBS).

G.1.1 PSID

The Panel Study of Income Dynamics (PSID) provides extensive information concerning demographics, economic outcomes, health care access, health outcomes, and health behaviors (such as smoking history, alcohol consumption, and exercise habits). Health outcome variables include diagnosis of diabetes, heart disease, hypertension, lung disease, and cancer, among others.

We estimate state-occupancy models using waves from 1997 to 2013. We create a dataset of respondents who have formed their own households, either as single heads of households, cohabiting partners, or married partners. These heads, wives, and husbands respond to the richest set of PSID questions, including the health questions that are critical for our purposes. We use all respondents aged 25 and older.[104] The length of the PSID panel dataset is a significant advantage, because we can use measures of past health behaviors to predict current health outcomes. This dataset provides adequate sample sizes to explore health outcomes of specific groups. PSID does not follow individuals who are institutionalized in nursing homes or other long-term care facilities. To overcome this weakness, we pool the PSID sample with the HRS sample when estimating mortality models.

---

[104]While we use the full sample in our main analysis, we explored using a few different subsamples to better adapt to the demographics of the ABC/CARE subjects.

G.1.2 HRS

The Health and Retirement Study (HRS) is a longitudinal panel that surveys a nationally representative sample of individuals over the age of 50 and their spouses every two years. When appropriately weighted, the HRS in 2010 is representative of U.S. households where at least one member is 51 years old. This study collects in-depth information about income, work, health, and medical expenditures. In our model, waves from 1998 to 2012 are pooled with the PSID for estimation of mortality and widowhood models. The HRS data are harmonized with the PSID for all relevant variables. Because the PSID does not follow respondents into nursing homes, we also use the HRS to estimate the model for nursing home residency. We use all cohorts in the dataset created by RAND (RAND HRS, version O) as the basis for our analysis.

G.1.3 MCBS

The Medicare Current Beneficiary Survey (MCBS) is a nationally representative sample of aged, disabled, and institutionalized Medicare beneficiaries. The MCBS attempts to interview each respondent twelve times over three years, regardless of whether he or she resides in the community, a facility, or transitions between community and facility settings. The disabled (under 65 years of age) and very elderly (85 years of age or older) are over-sampled. The first round of interviewing was conducted in 1991. Originally, the survey was a longitudinal sample with periodic supplements and indefinite periods of participation. In 1994, the MCBS switched to a rotating panel design with limited periods of participation. Each fall, a new panel is introduced, with a target sample size of 12,000 respondents. Each summer, a panel is retired. Institutionalized respondents are interviewed by proxy. The MCBS contains comprehensive self-reported information on the health status, health care use and expenditures, health insurance coverage, and socioeconomic and demographic characteristics of the

entire spectrum of Medicare beneficiaries. Medicare claims data for beneficiaries enrolled in fee-for-service plans are also used to provide more accurate information on health care use and expenditures. MCBS data from 2007 to 2010 are used for estimating medical costs and enrollment models.

G.1.4   MEPS

The Medical Expenditure Panel Survey (MEPS), which began in 1996, is a set of large-scale surveys of families and individuals, their medical providers, and employers across the U.S. The Household Component (HC) of the MEPS provides data from individual households and their members, which is supplemented by data from their medical providers. The HC collects data from a representative subsample of households drawn from the previous year's National Health Interview Survey (NHIS). NHIS does not include the institutionalized population. Neither does MEPS. These restrictions imply that we can only use the MEPS to estimate medical costs for the non-elderly (ages 25–64) population. Information collected during household interviews include: demographic characteristics, health conditions, health status, use of medical services, sources of medical payments, and body weight and height. Each year the household survey includes approximately 12,000 households, or 34,000 individuals. Sample size for those aged 25-64 is about 15,800 in each year. MEPS has comparable measures of socioeconomic status as those in PSID, including age, race and ethnicity, educational attainment, census region, and marital status. We estimate medical expenditure and utilization using data from 2008 to 2010. We use waves from 2001 to 2003 to estimate models of quality-adjusted life years (QALYs), due to availability of EQ-5D instrument in these waves.[105]

---

[105]Section G.2.2 explains the estimation of the QALY model.

G.1.5 NHANES

The National Health and Nutrition Examination Survey (NHANES) targets a nationally representative sample of approximately 5,000 individuals in each year since 1999. The data collected includes responses to interview questions about demographics, disease conditions, height, and weight, as well as physical measurement of BMI. We use NHANES years 2002 to 2010 to estimate a model for imputing measured BMI from self-reported BMI. The methodology for doing so is described below in Section G.1.6.1.

G.1.6 ABC/CARE

FAM uses ABC/CARE data to initialize the state of each ABC/CARE subject when they enter the simulation process. These data are taken from the the parental interviews at various subject ages from birth to age 21, age-30 subject interview, and mid-30s biomedical survey. The goal is to align each subject's initial state in the simulation match with their status at the age-30 subject interview. However, because several key FAM inputs are not available at the age-30 interview, we use PSID or ABC/CARE surveys corresponding to other ages to impute missing elements. These imputations are discussed in Section G.1.6.1.

G.1.6.1 Variable Construction and Imputations

Marital status and childbearing in FAM are determined in part by the subject mother's education level. The ABC/CARE age-30 subject interview did not ask about mother's education, but the ABC age-21 parent interview did. For ABC subjects, we assume that each subject's mother had the same education level at the age-30 subject interview as what was reported in the age-21 parent interview. For CARE subjects, we impute mother's education from an ordered Probit model using race, ethnicity, education, disease conditions, employment status, presence of a health-related work limitation, and a self-report of whether

161

or not the subject was "poor" as a child. The model is estimated using age 30 to 31 PSID subjects with birth years between 1945 and 1981. Each of the model covariate values is taken from the CARE age 30 interview. At the beginning of each individual simulation, an education level is randomly drawn from the probability distribution for each CARE subject and assigned to be the mother's education level.

FAM state-occupancy models depend on a three-level measure of parents' economic status when the subject was a child. This is based on the PSID question: "Were your parents poor when you were growing up, pretty well off, or what?" The three possible responses are "poor," "average"/"it varied", or "pretty well off." This question is not included in the ABC/CARE interviews, but because preliminary eligibility for the program focused on children from high-risk backgrounds, based on socioeconomic factors, the value of this variable is set to "poor" (when growing up) for all ABC/CARE subjects.

FAM state-occupancy models depend on demographics of the subject, including whether or not the subject is Hispanic. This information is not available in the ABC/CARE data, but it is assumed that none of the ABC/CARE subjects are Hispanic.[106]

FAM state-occupancy models also depend on smoking status. Employment status affects FAM state occupancies in marital status, childbearing, claiming of disability insurance (DI) and supplemental security income (SSI), and type of health insurance. One male in the ABC control group is missing smoking status and, although known to be not working, is also missing specific employment status (unemployed or out of the labor force). We use a multinomial logit model to jointly estimate the probability of each combined smoking and employment category among 25- to 35-year-olds in the PSID who were not working. At the beginning of each simulation repetition, we use a Monte Carlo random draw generated from this distribution to assign this subject's smoking and employment statuses. This same

---

[106]Census data on Hispanics in North Carolina were not available for 1970 and 1980, but Hispanic migration into this state is more recent than in other regions, and as late as 1990, only 2% of the North Carolina poor were Hispanic (Johnson, 2003).

subject is also missing information about binge drinking. A separate binary Probit binge drinking model was estimated using the age 25–35 PSID data. A Monte Carlo random draw is taken according the Probit probability to forecast binge drinking behavior at the beginning of the simulation.

BMI is an important determinant of the state occupancies that we model. BMI is a determinants of the diseases that we model as well as functional status, employment, and smoking. There are two issues with the measurement of BMI in the experimental and non-experimental samples. First, weight and height are self reported in the PSID. Then, BMI is constructed. Thus BMI could be measured with error in the PSID. Second, BMI is not measured at the same time across experimental and non-experimental datasets. In the PSID it is measured at age 30. In ABC/CARE it is measured at age 30 for CARE subjects and at age 34 for ABC subjects. BMI could increase or decrease systematically in the years between the age-30 subject interview and the age-34 health interview. In order to estimate state-occupancy models and then forecast using BMI as an initial condition, we solve these issues as described next.

To address the first issue, we use a variation on the method of Courtemanche, Pinkston, and Stewart (2015) to impute measured BMI in the PSID. While the method in Courtemanche et al. (2015) works for imputing height and weight, we apply the following specification to directly model BMI. Using respondents aged 30 to 40 in the 2002-2010 NHANES waves, we forecast measured BMI from percentile ranks of self-reported BMI using the model specification in Courtemanche et al. (2015). Three variations on the spline interactions of Courtemanche et al. (2015) are also considered. After estimating these models using NHANES data, covariate values from the PSID age 30–34 data in years 2002–2013 are used to impute measured BMI values for PSID respondents. A Kolmogorov-Sminov (K-S) test and a visual inspection of smoothed histograms are used to compare the distribution of PSID imputed values to the distribution of observed values in the NHANES estimation sample.

The model specification used for imputation has the smallest K-S distance between the two distributions.

After imputing values of measured BMI for PSID respondents age 30–34, we turn to the second issue: accounting for systematic trends in BMI from the age 30 interview to the health interview. The goal is to have a model that maps from measured BMI at the health interview around age 34 to self-reported BMI at the age 30 interview. Employing the longitudinal structure of PSID, we match each respondent's first interview between age 30–32 with their imputed measured BMI between ages 33–40. We then estimate a model using self-reported BMI between ages 30–32 as the response variable and imputed measured BMI at ages 33–40, the age when BMI is actually measured, along with other variables observed at age 30 as explanatory variables. This imputation model is applied to any ABC/CARE subject who has their health interview at least one year after their age 30 interview.

For ABC/CARE subjects who have their health interview within one year of the age 30 interview, we assume that any systematic time trends in BMI are too small to have any practical significance. However, we still need to convert the imputed measured BMI to a self-reported value for compatibility with other state-occupancy models estimated using PSID data. This model is estimated on ages 30–32 in the PSID and uses covariates from the age 30 interview along with imputed measured BMI to forecast self-reported BMI.

At the beginning of each simulation repetition, we choose the appropriate model to impute self-reported BMI for each ABC/CARE subject based on the time between their age 30 interview and their health interview. Their expected BMI is estimated from this model. A Monte Carlo Normal random draw is generated using the subject's expected BMI and the estimated variance from the model. This Monte Carlo draw is then assigned to be the subject's initial self-reported BMI in the simulation. Using BMI from the health interview limits the ABC/CARE subjects simulated in FAM to only those who have height and weight measurements in the health interview.

Subjects' health insurance coverage affects their medical costs. FAM uses three categories of health insurance: none, public only, and some private. Five ABC subjects and three CARE subjects were missing health insurance status. Three cases were imputed by assuming that subjects have no health insurance if they do not know their insurance status and either go to an emergency room or community health clinic or do not go anywhere when they need health care. In order to impute the insurance category for the remaining five cases, we use age 25–35 PSID data to estimate a Probit model for whether or not a subject had insurance. The predictors are gender, earnings, marital status, self-reported health, employment status, and whether or not the subject had any biological children. We use this model to compute the probability of having insurance at the start of the simulation (at the age-30 interview). Then, we generate a Monte Carlo binary random variate according to this probability. If the outcome is positive, the subject is assigned to have some private insurance.

FAM uses six Activities of Daily Living (ADLs) about which there is data in PSID: walking, dressing, eating, bathing or showering, getting in and out of bed or a chair, and using the toilet, including getting to the toilet. FAM simulates the number of these ADLs in which the subject has difficulty. ADL difficulties forecast FAM state occupancies in benefits claiming, mortality, employment status, insurance category, and nursing home residency. FAM also models the count of difficulties among six Instrumental Activities of Daily Living (IADLs) from PSID: preparing one's own meals; shopping for personal toilet items or medicines; managing one's own money, such as keeping track of expenses or paying bills; using the phone; doing heavy housework, like scrubbing floors or washing windows; and doing light housework, like doing dishes, straightening up, or light housecleaning. Both ADLs and IADLs are components of FAM's model for quality-adjusted life years (QALYs). The ABC/CARE age-30 subject interview does not ask about ADLs or IADLs, but it does ask if the subject has a physical or nervous condition that keeps them from working. PSID respondents are also asked this question. We create an imputation model for each of these two measures using an ordered Probit model estimated on PSID respondents aged 25 to 35. We use these

models to compute the probabilities for each number of ADLs and IADLs. To start the simulation, we generate Monte Carlo random draws according to these probabilities and use them to assign the corresponding counts.

When subjects claim DI benefits, it affects their employment status, insurance category, and Medicare enrollment. DI claiming also affects medical costs. SSI claiming affects FAM employment status. Lastly, claiming Social Security retirement benefits affects employment status and insurance category. The ABC age-30 subject interview has a single yes/no question about claiming which asks: "Currently are you receiving income from workman's compensation, disability, or Social Security benefits including Supplemental Security Income?" CARE asks a similar question. The PSID has separate questions for each benefit type. We use a multinomial logit model to estimate the joint probability of each combination of DI and SSI claiming. The estimation uses PSID respondents aged 25 to 35 who were claiming at least one of the following benefits: workman's compensation, DI, or SSI. A Monte Carlo random draw generated from this distribution is used to assign each ABC/CARE subject's DI- and SSI-claiming status at the start of the simulation. One ABC subject is missing data about whether or not they were claiming and was assumed to not be claiming any benefits.

As discussed in Section G.2.2, FAM uses different models to estimate medical costs depending on whether or not a subject is Medicare-eligible. Subjects can enroll in Medicare before the age of 65 if they are claiming DI. The cost estimates for Medicare-eligible subjects depend on the subjects' current disease status at the age-30 interview and their disease status two years prior to the interview. Unfortunately, ABC/CARE does not have disease data two years before the age-30 interview. We initialize the process assuming that all subjects do not have their disease conditions in the previous period.

G.2 Methods and Analysis for Forecasting Health

This subsection presents our simulation methodology.

G.2.1 Formalizing the FAM

Let $\mathcal{M}$ denote a set of possible health states, some of which can be absorbing. Transition times are ages of measurement. Let $\mathcal{A} := [0, \ldots, \bar{A}]$ index ages, where $\bar{A}$ is the last age that we forecast. We define $h_{a,m,m'}$ as the probability of transitioning from state m to state m' at age $a \in \mathcal{A}$, where $m, m' \in \mathcal{M}$. We drop individual subscripts to avoid notational clutter.

We denote a transition from m to m' at age a by $D_{a,m,m'} = 1$. If this transition does not occur, $D_{a,m,m'} = 0$. We let $\tilde{D}_{a,m}$ be the indicator of occupancy of state m at age a. $\tilde{D}_{a,m}$ is a generic entry in the the vector of age-a state occupancy indicators denoted by $\tilde{\boldsymbol{D}}_a$. $\tilde{\boldsymbol{D}}_0$ denotes the vector of initial state-occupancy conditions.

Instead of estimating a full competing risks model, we simplify a complex reality of multiple health states and general state dependence. We model individual health states including heart disease, hypertension, and stroke.

The probability of occupying state $m \in \mathcal{M}$ at age $a \in A$ is assumed to be generated by an index threshold-crossing model:

$$I_{a,m} \;=\; \boldsymbol{1}\left(\tilde{\boldsymbol{D}}_0 \geq 0\right)\boldsymbol{\Omega}_m + \boldsymbol{1}\left(\tilde{\boldsymbol{D}}_{a-1} \geq 0\right)\boldsymbol{\Lambda}_m + \boldsymbol{W_a}\boldsymbol{\beta}_m + \boldsymbol{B}\boldsymbol{\alpha}_m + \tau_{a,m} + \varepsilon_{a,m}, \quad (22)$$

where $\tilde{D}_{a,m} = \boldsymbol{1}\left(I_{a,m} \geq 0\right)$, $\boldsymbol{1}\left(\tilde{\boldsymbol{D}}_0\right)$ is a vector of indicators of mutually exclusive initial conditions with associated coefficients $\boldsymbol{\Omega}_m$, $\boldsymbol{1}\left(\tilde{\boldsymbol{D}}_{a-1} \geq 0\right)$ is a vector of mutually exclusive health conditions in the previous period with associated coefficients $\boldsymbol{\Lambda}_m$, $\boldsymbol{W}_a$ denotes variables that can be affected by treatment with associated coefficients $\boldsymbol{\beta}_m$, $\boldsymbol{B}$ is the vector of

eligibility conditions with associated coefficients $\boldsymbol{\alpha}_m$, $\tau_{a,m}$ is a function of age,[107] and $\varepsilon_{a,m}$ is a serially uncorrelated shock which is assumed to be uncorrelated with all of the right-hand side observables. In practice, when analyzing discrete outcomes, we assume that it is a unit normal random variable.

Instead of modeling initial conditions, we directly condition on them in Equation (22). The assumption of uncorrelatedess of $\varepsilon_{a,m}$ across ages, health states, and subjects together with normality allows us to perform separate estimation of the coefficients characterizing Equation (22) for each $m \in \mathcal{M}$ using maximum likelihood. We estimate $[\boldsymbol{\Omega}_m, \boldsymbol{\Lambda}_m, \boldsymbol{\beta}_m, \boldsymbol{\alpha}_m, \tau_{a,m}]$ for each $m \in \mathcal{M}$. Despite these simplifications, our model performs well in fitting population means of the outcomes that we analyze (see Tysinger et al., 2015).

The probability of occupying discrete states is generated in FAM using Equation (22). When simulating the model to forecast the health outcomes for the ABC/CARE subjects, we use their observed age-30 conditions as initial conditions $\tilde{\boldsymbol{D}}_0$. Most of these initial conditions relate to health, but we also include initial conditions related to economic status. The vector $\tilde{\boldsymbol{D}}_0$ can be affected by treatment. Campbell et al. (2014) estimate the wide variety of effects on health outcomes that ABC/CARE had. This motivates the inclusion of health outcomes into our analysis. We follow the simulation algorithm described in Section 3 modified to account for discrete states.

This model easily accommodates absorbing states. An absorbing state is a state $m \in \mathcal{M}$ such that once a subject occupies that state he/she never leaves it. Thus, $m \in \mathcal{M}$ is an absorbing state if $\tilde{D}_{a,m} = 1$ implies that $\tilde{D}_{a',m} = 1$ $\forall a' \in \mathcal{A}$ with $a' \geq a$. Once an individual reaches an absorbing stage his/her model in Equation (22) is no longer used to generate transitions. An obvious example of an absorbing state is death.

Ordered and unordered outcomes are also easily accommodated using standard methods in

---

[107]In our empirical analysis, we approximate the coefficients on $\tau_{a,m}$ using splines with knots at ages 35, 45, 55, 65, and 75.

discrete choice (e.g. Train, 2009). An example of the former is the level of psychological distress (e.g., low, medium, high). An example of the latter is labor force status (e.g., labor force status is categorized as out of labor force, unemployed, working part time, or working full time).

We can also model continuous outcomes so that $I_{a,m}$ is observed and the variables in $\tilde{\boldsymbol{D}}_a$ are replaced by observed counterparts. An example of a continuous outcome is body-mass index (BMI). We can also modify the model to accommodate mixed discrete/continuous variables.

Estimation of the inputs affecting health outcomes is performed separately. The economic outcomes that we include are labor income, for which we use the forecast discussed in Section 3, as well as labor force participation, relationship status, and childbearing models generated in FAM to account for the disadvantage of participants.

Tables G.1 to G.3 list the variables determining each of the states and health and economic outcomes that we consider.[108] For each of the outcomes we list: (1) the outcome itself; (2) the variable type—e.g., absorbing state, binary outcome, continuous outcome; (3) initial health state occupancies and other other outcomes—$\tilde{\boldsymbol{D}}_0$ in Equation (22); (4) lagged health-state occupancies—$\tilde{\boldsymbol{D}}_{a-1}$ in Equation (22); (5) and (6) other health and economic outcomes used to determine the outcome of interest—$\boldsymbol{W}_a$ in Equation (22); and (7) background variables—$\boldsymbol{B}$ in Equation (22). From this information, the reader can read off the type of outcome being model, its classification, as well as determinants and their categorization according to Equation (22). The outcomes that help predict each other are based on research and advice of clinicians and other medical professionals, as explained and justified in Goldman et al. (2015).

---

[108]For a precise construction of each of the variables listed in this tables, the reader can request the set of codes to replicate the computations in this paper.

Table G.1: Determinants of Equation (22) for Different Outcomes

(1) Outcome	(2) Variable Type	(3) $\tilde{D}_0$ Initial Conditions	(4) $\tilde{D}_{a-1}$ Other States	(5) W_a Health Outcomes	(6) W_a Economic Outcomes	(7) B Demographics
Heart Disease	Absorbing	Childhood Economic Environment Education Asthma	Hypertension Diabetes	Smoking BMI Physical Activity		Race Ethnicity Age Gender
Hypertension	Absorbing	Childhood Economic Environment Education	Diabetes	Smoking BMI Physical Activity		Race Ethnicity Age Gender
Stroke	Absorbing	Childhood Economic Environment Education	Heart Disease Hypertension Diabetes Cancer	Smoking BMI Physical Activity		Race Ethnicity Age Gender
Lung Disease	Absorbing	Childhood Economic Environment Education Asthma		Smoking BMI Physical Activity		Race Ethnicity Age Gender
Diabetes	Absorbing	Childhood Economic Environment Education		Smoking BMI Physical Activity		Race Ethnicity Age Gender
Cancer	Absorbing	Childhood Economic Environment Education		Smoking BMI Physical Activity		Race Ethnicity Age Gender
Mortality	Absorbing	Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer Functional Status	Smoking Binge Drinking		Race Ethnicity Age Gender
Functional Status	Ordered	Childhood Economic Environment Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking BMI Physical Activity Functional Status		Race Ethnicity Age Gender
Smoking	Binary	Childhood Economic Environment Education	Heart Disease Lung Disease Diabetes	BMI Binge Drinking Physical Activity Psychological Distress		Race Ethnicity Age Gender

Table G.2: Determinants of Equation (22) for Different Outcomes, Continued

(1) Outcome	(2) Variable Type	(3) $\tilde{D}_0$ Initial Conditions	(4) $\tilde{D}_{a-1}$ Other States	(5) W_a Health Outcomes	(6) W_a Economic Outcomes	(7) B Demographics
BMI	Continuous	Childhood Economic Environment, Education		BMI	Marital Status	Race, Ethnicity, Age, Gender
Binge Drinking	Binary	Childhood Economic Environment, Education	Marital Status	Binge Drinking		Race, Ethnicity, Age, Gender
Physical Activity	Binary	Childhood Economic Environment, Education	Marital Status	Physical Activity		Race, Ethnicity, Age, Gender
Psychological Distress	Ordered	Childhood Economic Environment, Education	Heart Disease, Hypertension, Stroke, Lung Disease, Diabetes, Cancer	Smoking, BMI, Physical Activity, Psychological Distress, Functional Status		Race, Ethnicity, Age, Gender
Childbearing	Ordered	Mother's Education, Education	Cancer	Marital Status	Labor Force Participation, Number of Children	Race, Ethnicity, Age, Gender
Paternity	Ordered	Mother's Education, Education			Labor Force Participation, Marital Status, Number of Children	Race, Ethnicity, Age
Marital Status	Binary	Mother's Education, Education			Labor Force Participation, Earnings, Marital Status, Number of Children	Race, Ethnicity, Age, Gender
Partner Mortality	Binary	Education				Race, Ethnicity, Age, Gender

Table G.3: Determinants of Equation (22) for Different Outcomes, Continued

(1) Outcome	(2) Variable Type	(3) $\tilde{D}_0$ Initial Conditions	(4) $\tilde{D}_{a-1}$ Other States	(5) W_a Health Outcomes	(6) W_a Economic Outcomes	(7) B Demographics
Labor Force Participation	Unordered Categorical	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking BMI Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Supplemental Security Income Claiming Earnings Marital Status	Race Ethnicity Age Gender
Full-time Employment	Binary	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking BMI Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Supplemental Security Income Claiming Earnings	Race Ethnicity Age Gender Marital Status
Disability Insurance Claiming	Binary	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Earnings	Race Ethnicity Age Gender
Social Security Claiming	Absorbing	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Earnings Marital Status	Race Ethnicity Age Gender
Supplemental Security Income Claiming	Binary	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Supplemental Security Income Claiming Earnings Marital Status	Race Ethnicity Age Gender
Health Insurance Type	Unordered Categorical	Childhood Economic Environment, Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Smoking Functional Status	Labor Force Participation Disability Insurance Claiming Social Security Claiming Earnings Marital Status Health Insurance Type	Race Ethnicity Age Gender
Nursing Home Residency	Binary	Education	Heart Disease Hypertension Stroke Lung Disease Diabetes Cancer	Functional Status	Nursing Home Residency Widowhood	Race Ethnicity Gender Age

G.2.2 Quality-Adjusted Life Years and Medical Costs

The health states and outcomes *per se* do not readily translate into monetary values of the treatment effects on health. With the health outcomes in hand, we use predicted QALYs, medical expenditure, and Social Security participation to monetize benefits and costs associated with health.

We predict QALYs using the EQ-5D instrument, a widely-used, health-related quality-of-life (HRQoL) measure. The scoring system for EQ-5D was first developed using a U.K. sample[109] Later, a scoring system based on a U.S. sample was generated.[110] The PSID does not ask the appropriate questions for computing EQ-5D, but MEPS does. We forecast EQ-5D scores from the MEPS onto the PSID data using variables common across MEPS and PSID. We map EQ-5D scores from the MEPS onto PSID data using common variables between the MEPS and PSID. We then run a linear regression of EQ-5D on PSID variables that are predicted in FAM (including ADL counts, IADL counts, and diseases).[111] The microsimulation uses this linear regression to compute QALYs.

Also based on a version of Equation (22), we predict and then forecast medical cost models. For individuals who are not Medicare-eligible, the cost models are estimated from MEPS data. Once an individual becomes Medicare-eligible, their costs are estimated from MCBS data. Both sets of models include the following covariates: age, gender, race and ethnicity, education level, relationship status, disease conditions, and labor income. The MEPS models also include type of health insurance as a covariate. Because MCBS follows respondents for more than two years (the time step length in the FAM simulation), the FAM cost models for the Medicare-eligible population and includes covariates for the stage of each disease. In the initial stage, a patient has a diagnosis in the current two-year period, but did not have the

---

[109]Dolan (1997).

[110]Shaw et al. (2005).

[111]The main variables in this prediction are self-reported health and requiring help with ADLs.

diagnosis in the previous two-year period. Then, in the maintenance stage, a patient had a diagnosis in the previous two-year period and survives with the diagnosis in the current two-year period (all disease states are absorbing—it is impossible to transition out of a diagnosis). Finally, in the terminal stage, a patient has a diagnosis and dies in the current two-year period. The medical costs models underestimate health care spending reported in the National Healthcare Expenditures Account (NHEA) data, due in part to underreporting of Medical costs in MEPS.

G.2.3 FAM simulation

We estimate the models in Equation (22). With these models in hand, we simulate the health trajectories for the ABC/CARE subjects, by initializing it their with the initial conditions, such as we do for labor income in Section 3. Missing values are imputed with the imputation models described in section G.1.6.1.

To match the biennial structure of the PSID data used to estimate the models, the simulation proceeds in two-year increments.[112] Once the new states have been determined, the cross-sectional models for medical costs and QALYs are applied. That is, for each subject and at each age, we map the health conditions into medical costs and QALYs. This enables us to monetize health outcomes. The simulation ends when all simulated ABC/CARE subjects are deceased (in the simulation).[113]

Among the ABC/CARE subjects simulated in FAM, the years of completion of the age-30 interview range from 2003 to 2009. FAM's two-year time step only allows the simulation of even or odd years. For this reason, we run the simulation twice—once for the ABC/CARE subjects entering in odd years and again for the ABC/CARE subjects entering in even years.

---

[112]The end of each two-year step is designed to occur on July 1st to allow for easier matching with population forecasts from Social Security Administration (SSA).

[113]Less than half of the simulated subjects (48%) survive to age 80.

Figure G.1: Year-over-Year Excess Real Growth in Medical Costs
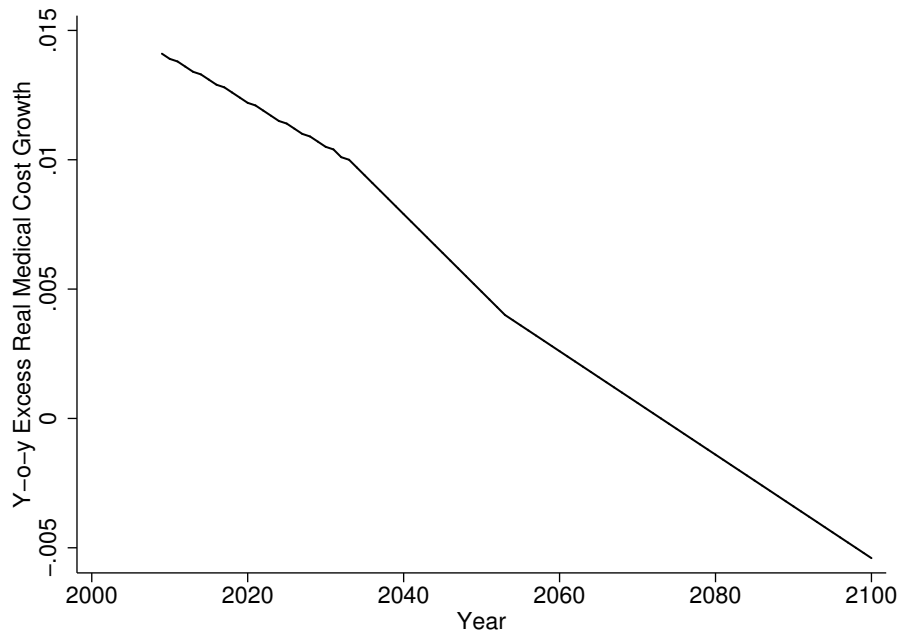
Sources: Congressional Budget Office, Social Security Administration.
Note: The year-over-year excess real medical cost growth over GDP is used to model medical cost growth in FAM.

The simulation model takes as inputs assumptions regarding the normal retirement age, future improvements in mortality, and real medical cost growth. The normal retirement age is assumed to be 67 for all ABC/CARE subjects, as in Section 3.

The FAM mortality model represents mortality rates in 2009. The estimated mortality probabilities are reduced in simulated future years to represent improvements in mortality from sources such as medical innovation that are not included in the model. There are different adjustment factors for the populations under and over the age of 65. The mortality reduction factors are taken from the intermediate cost mortality projections in the 2013 Social Security Trustee's Report.

Medical cost growth assumptions are derived from several underlying assumptions about growth in GDP and the labor force. The real medical cost growth factor in each year is calculated by first finding the minimum of (i) the year-over-year GDP growth plus year-

over-year excess medical cost growth or (ii) the Affordable Care Act cap on year-over-year medical cost growth. In order to obtain the medical cost adjustment factor for the current year of the simulation, FAM takes the cumulative product of the yearly growth factors since 2004 and then divides it by the relative growth in the labor force since 2004.[114]

G.2.4 Medical Costs Before the Age-30 Interview

Data on utilization of medical services is sparse before the age 30 interview. There are questions about utilization in the age 12, 15, and 21 interviews along with records of births for female subjects. We combine this with information about demographics, family structure, and parents' utilization of public services to estimate medical costs at each age from 8 to 32. Models were estimated separately for males and females. All imputation and cost models are estimated using MEPS data.

Medical costs for ages 8 to 11 were estimated in three stages using age 12 interview data. First, we impute whether or not a subject spent a week in the hospital for those subjects who are missing this information in their age 12 interview. The imputation model forecasts utilization based on race and whether or not the subject was ever diagnosed with asthma between ages 8–11. Next we separate the ABC/CARE subjects into the group that spent a week in the hospital in this age range and the group that did not. For the group that did not spend any time in the hospital, we forecast medical costs as a two stage model. The first stage predicts whether there were any medical costs at all. Then, the second stage forecasts the amount of medical costs for those subjects who were predicted to have some costs. We assume the group that spent time in the hospital had some medical costs, so we skip the first stage and go directly to predicting the amount. The cost models use race, asthma diagnosis, whether or not the father was absent from the home, family use of food stamps, and number

---

[114]The medical cost growth assumptions come from Congressional Budget Office and SSA assumptions. The year-over-year growth assumptions for medical costs are shown in Figure G.1. The 2010-2019 GDP assumptions are based on CBO's analysis of the President's Budget, March 2009. GDP assumptions for 2020-2100 are based on the 2008 OASDI Trustee's Report long-term projection of 2.1% real GDP growth.

of siblings as predictors.

Estimating medical costs for ages 12 to 14 follow a strategy similar to that used to forecast age 8–11 costs. First, we impute whether or not a subject had any hospitalization for those subjects who do not report this in their age 15 interview. Imputations are based on race and presence or absence of an asthma diagnosis between ages 12–14. Again, we separated the ABC/CARE subjects into a group that had a hospitalization between age 8–11 and a group that did not. A two-stage model was used to forecast medical costs for those with no hospitalization. Medical costs for the group that had a hospitalization were estimated directly from a single-stage model. These cost models use race, asthma diagnosis, whether the mother, father, or both parents were absent from the home, family use of food stamps, and number of siblings.

To estimate medical costs for ages 15 to 20, we first impute whether or not the subject spent time in the hospital for those who are missing this information in the age 21 interview. The imputation model was based on race, asthma diagnosis between ages 15–20, and, for females, the birth of any children. The age 21 interview asks about the number of days spent in the hospital. However, it does not record the ages at which these hospital stays occurred. Considering the difficulty of assigning the hospital days to specific ages in the absence of other information, we decided to use only the indicator of whether or not there were any days spent in the hospital. Next, we separated subjects into a group that spent some time in the hospital between ages 15–20 and those who did not. As before, we used the direct model to forecast costs for those who had been to the hospital and used a two-stage model for those who had not. The cost models forecast costs based on race, asthma diagnosis, any births (female model only), use of food stamps, whether or not the subject was of working age, work status, living at college, and living with parents, and marital status.

Unlike the interviews at younger ages, the age 30 interview does not ask about utilization of medical services. To estimate costs for ages 21–31, we skipped the utilization imputation

step and moved directly to cost models. We used two-stage cost models. The first stage predicts whether or not there were any costs based on race, asthma diagnosis between ages 21-31, education, use of food stamps, any births (female model only), whether or not the subject was working age, living at college, living with parents, and marital status.

Table G.4 summarizes individual and family characteristics used to forecast medical expenditure models for each age.

Table G.4: Health Expenditure Models by Age Group, before Age 30

Explanatory variable	Age Group			
	8-11	12-14	15-20	21-30
Race/ethnicity	✓	✓	✓	✓
Education	✗	✗	✗	✓
Asthma Diagnoses	✓	✓	✓	✓
Hospital stays	if ≥ 1 week	any stay	any stay	✗
Births	✗	✗	✓	✓
Mother present	✗	✓	✗	✗
Father present	✓	✓	✗	✗
Number of siblings	✓	✓	✗	✗
Foodstamps	✓	✓	✓	✓
Living arrangements	✗	✗	✓	✓
Working, if working age	✗	✗	✓	✓

Note: This table summarizes the explanatory variables included in the models we use to forecast medical expenditure for each age group. Possible living arrangements are: living with parents, away at college, married, or other.

G.2.4.1   Specification Tests for the First-order Markov Assumption in FAM

The FAM model assumes a vector first-order Markov process for its states with a variable state space depending on the outcome analyzed. In Equation (22), only $\tilde{\boldsymbol{D}}_{a-1}$ and the initial state value enter the model. For heart disease, hypertension, and stroke, we compare the fit of Equation (22) with the following second-order model:

$$I_{a,m} = \mathbf{1}\left(\tilde{\boldsymbol{D}}_0 \geq 0\right)\boldsymbol{\Omega}_m + \mathbf{1}\left(\tilde{\boldsymbol{D}}_{a-1} \geq 0\right)\boldsymbol{\Lambda}_m + \left(\tilde{\boldsymbol{D}}_{a-2} \geq 0\right)\tilde{\boldsymbol{\Lambda}}_m$$
$$+ \quad \boldsymbol{W_a}\boldsymbol{\beta}_m + \boldsymbol{B}\boldsymbol{\alpha}_m + \tau_{a,m} + \varepsilon_{a,m}. \tag{23}$$

We use a likelihood ratio test to test the null hypothesis $H_0 : \tilde{\boldsymbol{\Lambda}}_m = \mathbf{0}$ in Equation (23). Table G.5 show the results from these tests. For the health states considered, we do not reject the null hypothesis.

Table G.5: Tests Comparing First-Order and Second Markov Processes for Disease State-Occupancy Specifications

Statistic	LR Statistic	Degrees of Freedom	p-value
Desease			
Heart Disease	2.18	2	0.71
Hypertension	0.05	1	0.83
Stroke	3.94	4	0.14

Note: This table presents likelihood ratios contrasting the models in Equations (22) and (23) by testing the null hypothesis $H_0 : \tilde{\boldsymbol{\Lambda}}_{a,m} = 0$. The variables included in the right-hand-side of each model are in Tables G.1 to G.3.

An alternative test for a first-order Markov process is based on the following version of a Lagrange multiplier test: (i) use a linear probability model approximation to a first-order Markov process and the variables in Tables G.1 to G.3 for each outcome of interest; (ii) calculate the correlation of the residuals with higher order lagged variables. Tables G.6 to G.8 show these correlations. The results of these tests strongly support a first-order Markov assumption.

To illustrate how to read these tests consider Table G.6. This table reports simple first-order Pearson correlations with the indicated variable. The row presents the residuals from forecasting heart disease at age a using different orders of lags. The estimated correlations

are small—and jointly statistically insignificant.[115]

---

[115]To perform the joint test, we proceed as follows. Let $\hat{D}_{a,m}$ be the residual of the linear-probability approximation of Equation (22). For each disease in Tables G.6 to G.8, we regress $\hat{D}_{a,m}$ on the diseases listed in each table's columns. Let Ξ_m denote the coefficients of this regression. We use an F-test to test the null hypothesis $H_0 : \Xi_m = \mathbf{0}$. The p-values associated with the F-test for Tables G.6 to G.8 are $p > 0.85$ in the three cases.

Table G.6: Tests for Linear Probability Forecasts of Heart Disease at a

Correlations with	Residuals of Forecast of Heart Disease at age a									
	Diabetes at $a-2$	Hypertension at $a-2$	Diabetes at $a-3$	Hypertension at $a-3$	Diabetes at $a-4$	Hypertension at $a-4$	Diabetes at $a-5$	Hypertension at $a-5$	Diabetes at $a-6$	Hypertension at $a-6$
	-0.002	-0.006	-0.004	-0.009	-0.017	-0.020	-0.015	-0.017	-0.009	0.004

Note: This table presents the correlation between the residuals from forecasting heart disease at age a and the predictors indicated in each column. The residuals are obtained from a linear-probability approximation of Equation (22). The right-hand side variables used are listed in Tables G.1 to G.3.

Table G.7: Tests for Linear Probability Forecasts of Hypertension at a

Correlations with	Residuals of Forecast of Hypertension at age a				
	Diabetes at $a-2$	Diabetes at $a-3$	Diabetes at $a-4$	Diabetes at $a-5$	Diabetes at $a-6$
	0.003	0.009	0.012	0.008	0.001

Note: This table presents the correlation between the residuals from forecasting hypertension at age a and the predictors indicated in each column. The residuals are obtained from a linear-probability approximation of Equation (22). The right-hand side variables used are listed in Tables G.1

Table G.8: Tests for Linear Probability Forecasts of Stroke at a

Correlations with	Residuals of Forecast of Stroke at age a							
	Cancer at $a-2$	Diabetes at $a-2$	Heart Disease at $a-2$	Hypertension at $a-2$	Cancer at $a-3$	Diabetes at $a-3$	Heart Disease at $a-3$	Hypertension at $a-3$
	0.004	0.008	0.001	0.001	-0.002	0.017	-0.004	-0.001

Note: This table presents the correlation between the residuals from forecasting stroke at age a and the predictors indicated in each column. The residuals are obtained from a linear-probability approximation of Equation (22). The right-hand side variables used are listed in Tables G.1

H Sensitivity Analyses

Figure H.1 summarizes the range of all of the estimates that we generate in this paper. The overall benefit/cost ratio (internal rate of return) ranges from 1.52 to 17.40 (8 to 18.3) for the full sample. They range from 2.23 to 25.45 (6 to 19.4) for the males, and from 1.12 to 5.79 (4 to 18) for the females. Benefits from reductions in criminality and increased labor income are pronounced for males, contributing to their larger estimates relative to the females estimates. However, when we omit crime from our analysis, we still estimate substantial returns for males.

Figure H.1: Distribution of Benefit/Cost and Internal Rate of Return Estimates

(a) Benefit/Cost Ratio

(b) Internal Rate of Return

The vertical line represents the baseline estimate.
Average. Pooled: 6.90. Females: 2.63. Males: 9.67.
Median. Pooled: 6.75. Females: 2.49. Males: 9.85.

The vertical line represents the baseline estimate.
Average. Pooled: 10.07. Females: 13.34. Males: 13.34.
Median. Pooled: 13.00. Females: 10.00. Males: 13.70.
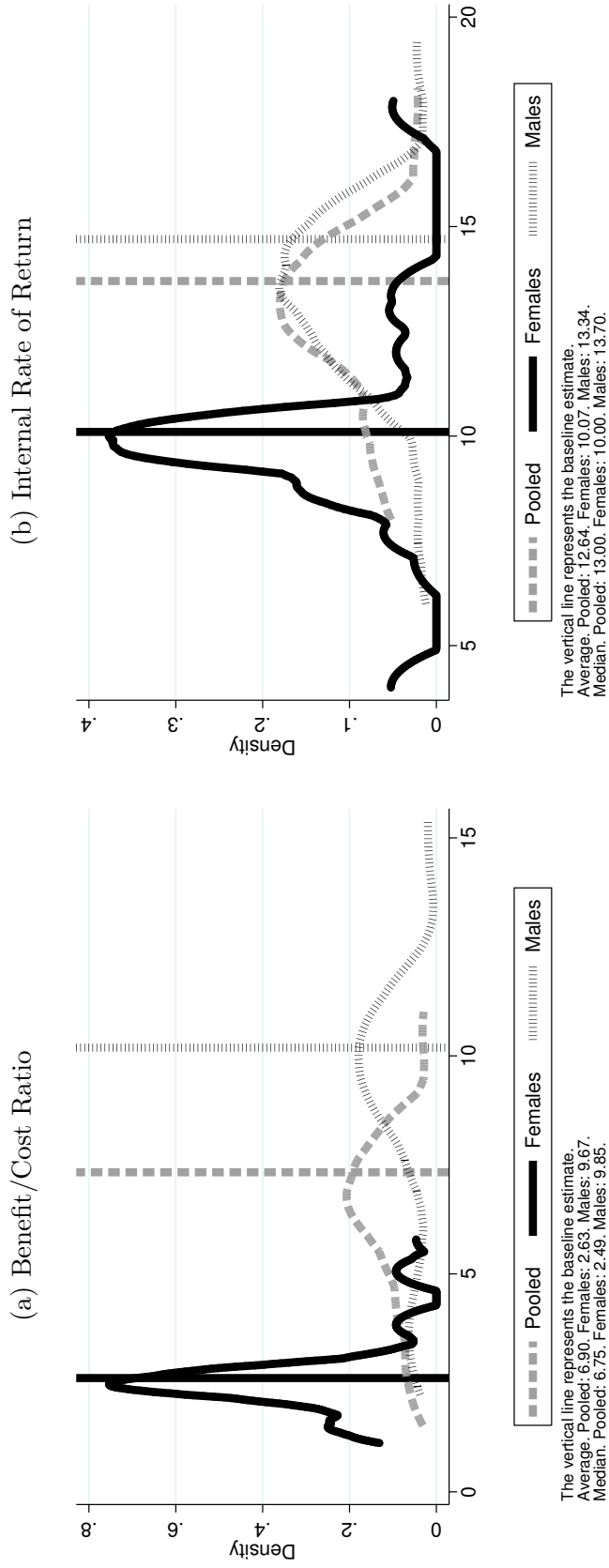
Note: Panel (a) displays the distribution of estimates of the benefit/cost ratio that we estimate throughout the paper. Vertical lines indicate the baseline estimates, presented in Figure 3. Panel (b) presents the analogous figure for the internal rate of return. See Figure H.2 in Appendix H for the analogous plots of net present values.

183

This appendix evaluates how the estimates in the main paper vary as we alter certain sample selections from all of our data sources and other parameters. We first present plot of all net present value estimations across specifications in Figure H.2. This is analogous to Figure H.1.

Figure H.2: Distribution of Net Present Value Estimates
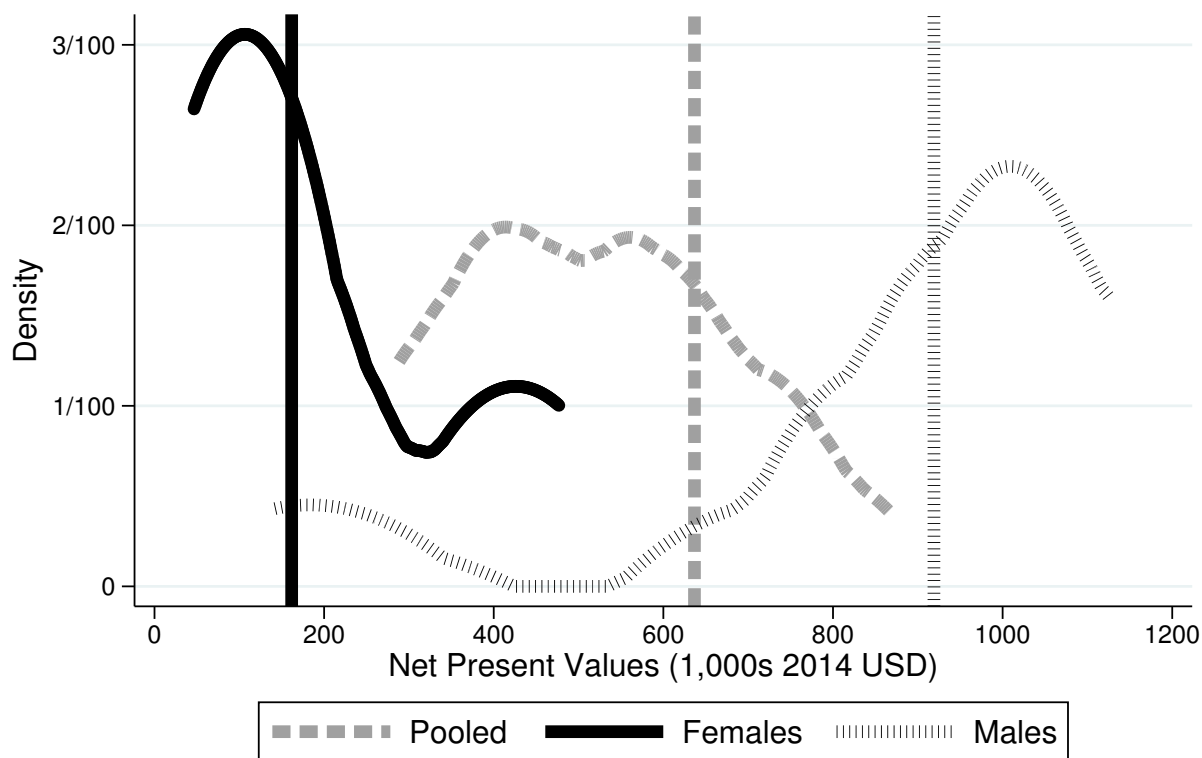
The vertical line represents the baseline estimate.
Average. Pooled: 517,608. Females: 190,627. Males: 872,593.
Median. Pooled: 509,466. Females: 129,260. Males: 966,735.

Note: This figure displays the distribution of estimates of the net present value that we estimate throughout the paper. Vertical lines indicate the baseline estimates, presented in Figure 3.

We next analyze sensitivity due to (i) assumptions on the values of the discount rate; (ii) assumptions on the deadweight loss to society coming from taxes raised to fund public programs; and (iii) the magnitude of the components contributing to the benefit/cost ratio and internal rate of return. Table H.1 summarizes our sensitivity analyses. We conduct a rather drastic sensitivity analysis by removing components of the cost/benefit analysis entirely and examining its impact on the resulting estimates. Even when completely removing the gain

184

associated with crime for males, the program is socially efficient—both the internal rate of return and the benefit/cost ratio are substantial. Parental labor income and crime are the components for which the internal rate of return and the benefit/cost ratio are the most sensitive. The reason for this sensitivity to parental labor income is that the amount is substantial and it is not heavily discounted because it accumulates during the first 21 years of the subjects' lives. Crime occurs later in life and its benefits are discounted accordingly. The amount due to savings in crime is large, so removing it diminishes both the internal rate of return and the benefit/cost ratio (but they remain statistically significant).

Table H.1: Cost/Benefit Analysis of ABC/CARE, Summary

Removed Component	Females			Males			Pooled		
	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C
None	161,759	**10.1%**	**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 other rows present estimates for scenarios in which we remove the NPV estimates of the component listed in the first column. The category "Alternative Preschools" refers to the money spent in alternatives to treatment from the control-group children parents. QALYs refers to the quality-adjusted life years. Any gain corresponds to better health conditions through the age of death. The quantity listed in the NPV columns is the component we remove from NPV when computing the calculation in each row. All the money figures are in 2014 USD and are discounted to each child's birth, unless otherwise specified. For the B/C ratio we use a discount rate of 3%, unless otherwise specified. We test the null hypotheses IRR = 3% and B/C = 1—we select 3% as the benchmark null because that is the discount rate we use. Inference is based on non-parametric, one-sided p-values from the empirical bootstrap distribution. We highlight point estimates significant at the 10% level. Total cost of the program per child is $92,570 (2014).

H.1   Varying the Discount Rate

Below we examine how the benefit/cost ratio is impacted by our choice of the discount rate.

Figure H.3 displays how the benefit/cost ratio changes as we adjust the rate at which we discount the cash flows. We find that for males, the benefits of the programs exceed the costs for discount rates as high as 15%. We also find that the benefit/cost ratios for discount rates of 2–12% remain within the 80% confidence interval of our actual point estimate. The case is different for females, for whom the ratio falls below 1 at a discount rate closer to 10%. The benefit/cost ratio for females remain within the 80% confidence interval of our estimate for discount rates of 2–15%. Although the alternate estimates generally remain within the 80% confidence intervals for both males and females, the slope of the curves in Figure H.3 indicate that our estimates are sensitive to our choice of the discount rate, especially for males.

Figure H.3: Benefit/cost Ratio vs. Discount Rate

(a) Females
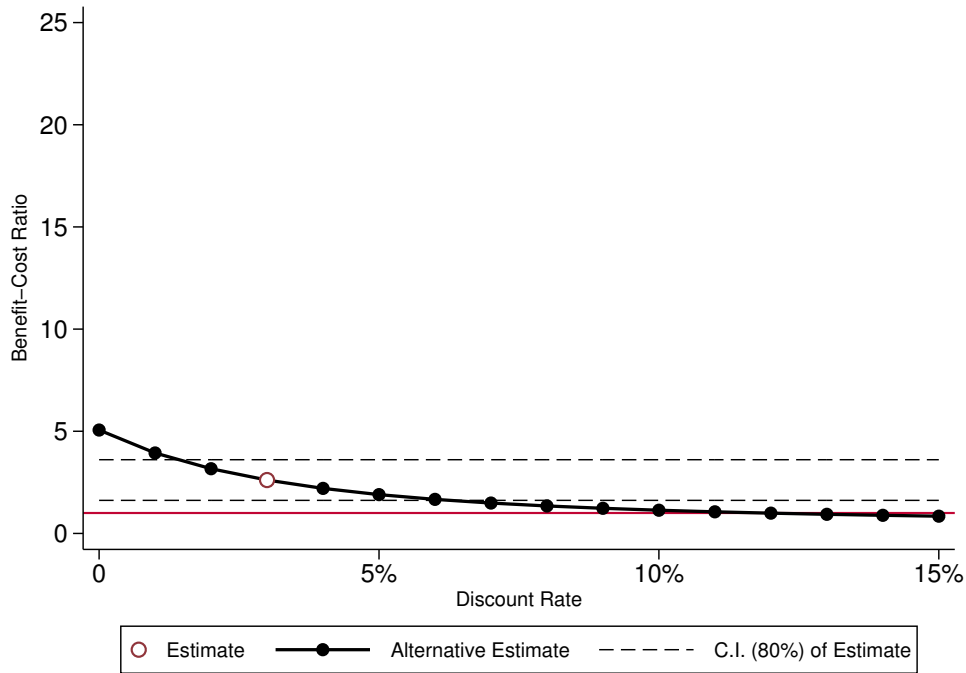
(b) Males
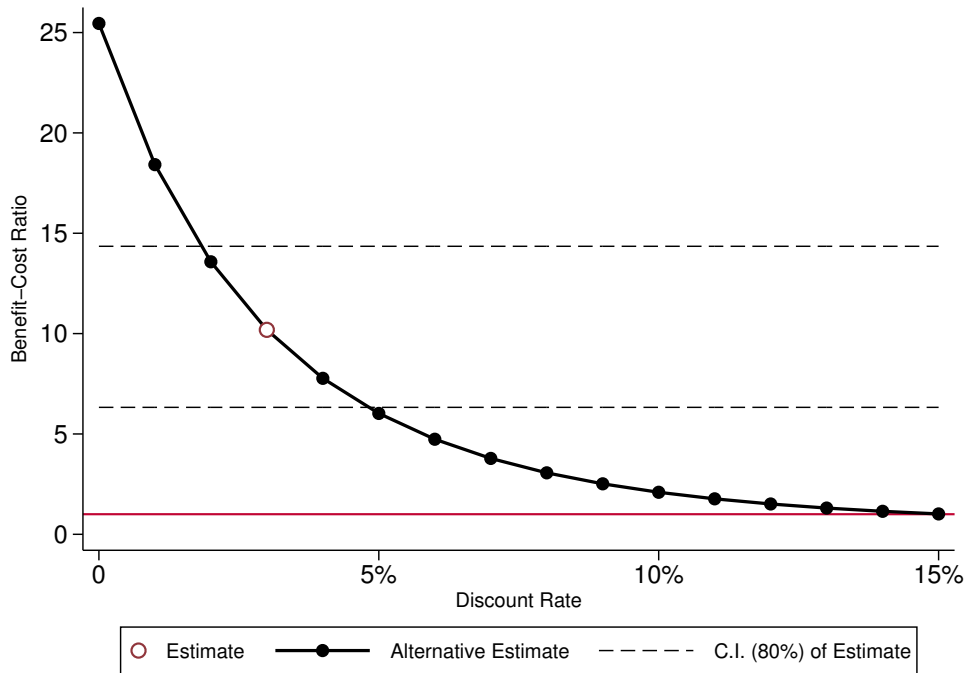
Note: These graphs display how the benefit/cost ratios change for females and males as we vary the rate at which we discount to obtain the present value. The red line indicates a benefit/cost ratio of 1. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the discount rate. The estimates presented in the paper assume that the marginal cost of welfare is $0.50 for every dollar of tax revenue. The estimates 187 means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10$^\text{th}$ and 90$^\text{th}$ quantiles of the bootstrap distribution.

H.2   Varying Deadweight Loss

Below we examine how our estimates of the internal rate of return (IRR) and benefit/cost ratios move with respect to changes in the marginal cost of welfare.

Figure H.4 shows how the IRR changes as we adjust the marginal cost of welfare. For both males and females, we find the IRRs to be most sensitive at the lower marginal costs. The IRRs for both sexes steadily decline as we increase the marginal cost of welfare to $3 for every dollar of tax revenue. This is likely due to the fact that both females and males in treatment live longer, and are expected to receive more Medicare and Medicaid benefits in their later life. Also, we treat the costs of implementing ABC/CARE as a public cost. Thus, the steady increase in deadweight loss results in a steady decline in the IRR. Nonetheless, we see that the IRRs remain within the 80% confidence interval of our original estimate for both females and males and are not particularly sensitive to changes within the neighborhood of our assumed marginal cost of welfare (note each point on the x-axis represents a $0.25 increment in the cost of welfare per dollar of tax revenue).

Figure H.4: Internal Rate of Return vs. Deadweight Loss
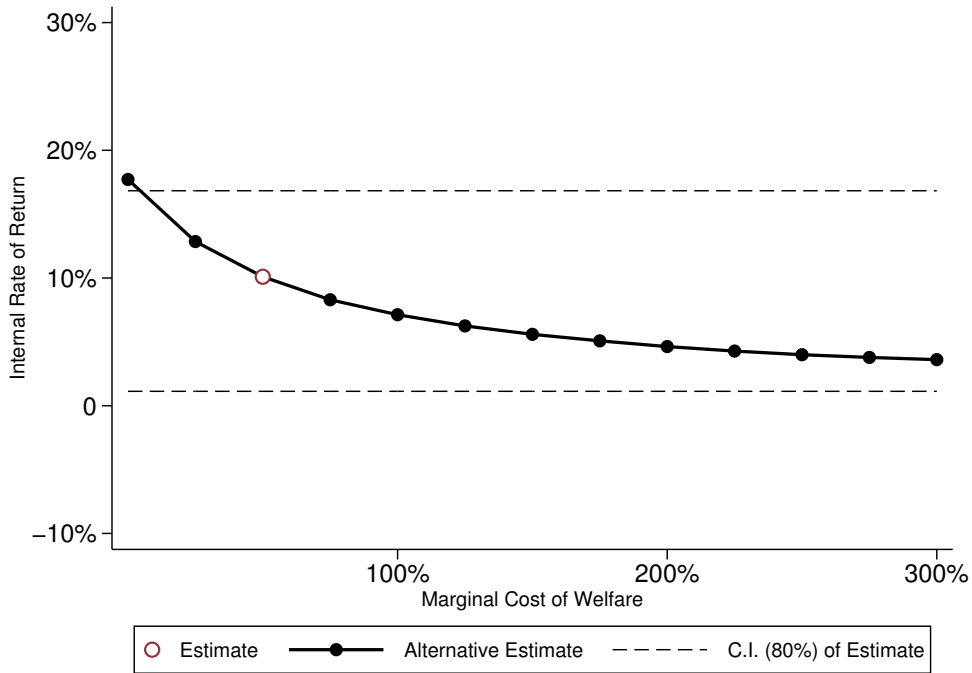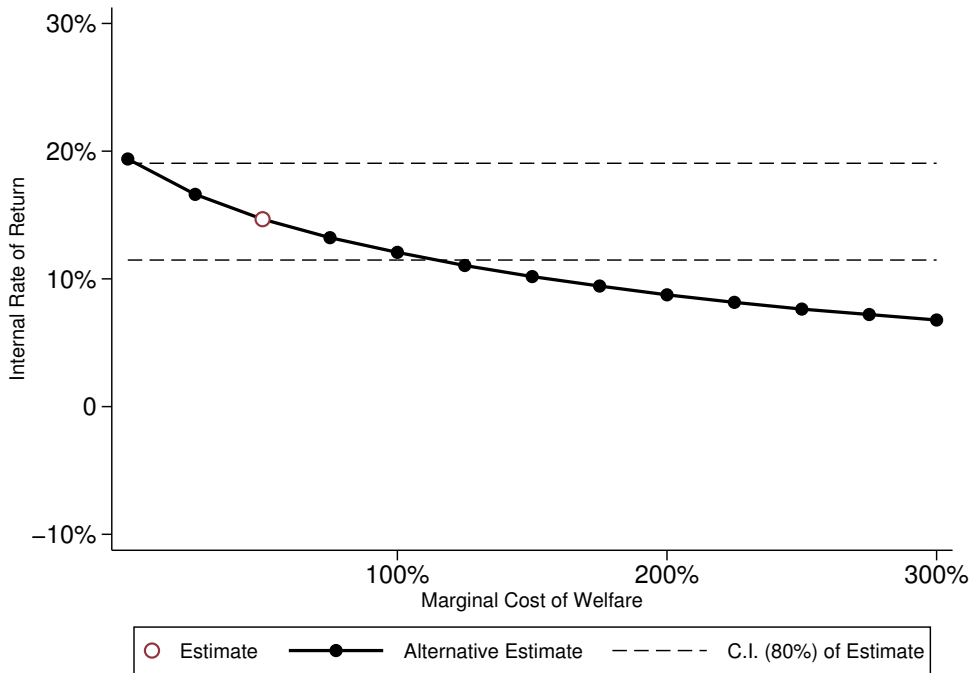
(a) Females

(b) Males

Note: These graphs display how the internal rate of return changes for females and males as we vary the marginal cost of welfare. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the marginal cost of welfare. The estimates presented in the paper assume that the marginal cost of welfare is \$0.50 for every dollar of tax revenue. The estimates are means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10th and 90th quantiles of the bootstrap distribution.

Figure H.5 illustrates how the benefit/cost ratio changes as we vary the marginal cost of welfare. For females we see that the benefits exceed the costs (both are discounted at a rate of 4%) even when the marginal cost of welfare is assumed to equal \$3 for every dollar of tax revenue. We see a similar relationship hold for men, however, the ratio is significantly higher at every marginal cost.

Figure H.5: Benefit/cost Ratio vs. Deadweight Loss
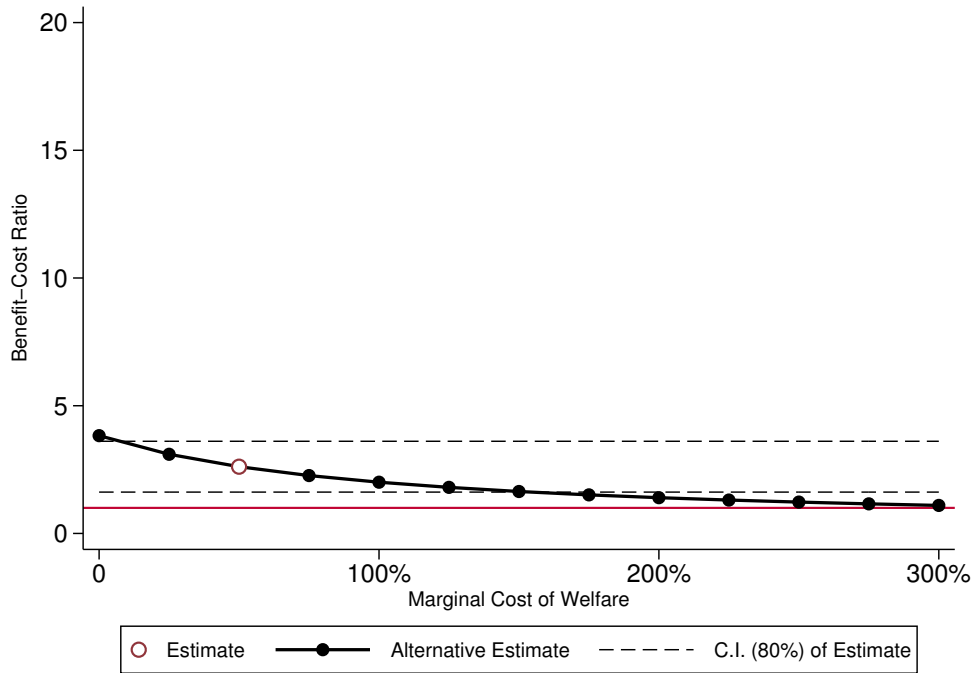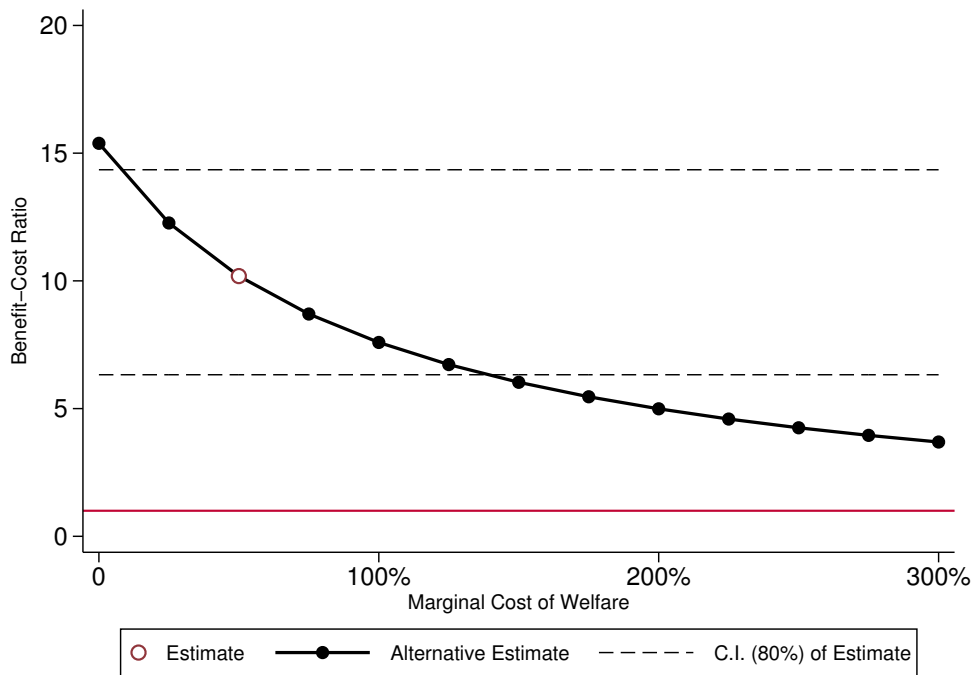
(a) Females

(b) Males

Note: These graphs display how the benefit/cost ratio changes for females and males as we vary the marginal cost of welfare. The red line indicates a benefit/cost ratio of 1. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the marginal cost of welfare. The estimates presented in the paper assume that the marginal cost of welfare is \$0.50 for every dollar of tax revenue. The estimates are means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10th and 90th quantiles of the bootstrap distribution.

H.3 Varying Magnitudes of Components

In this section, we explore how the internal rate of return (IRR) and benefit/cost ratio change as we increase and decrease the value of each component of the benefit and cost streams. This entails multiplying each component by factors ranging from 0 to 3. A factor of 0 is equivalent to removing the component entirely from our analysis. In the particular case of QALYs, the multiplicative factors correspond to different valuations of a year of perfect health, i.e., a QALY equal to 1. For instance, as our current estimates assume a QALY of 1 to be worth \$150,000, a factor of 0.5 corresponds to a year of perfect health being worth \$75,000, and a factor of 3 corresponds to a year of perfect health being worth \$450,000 (all values are in 2014 USD).

Figure H.6 displays how the IRR for females changes as we multiply each component by a factor between 0 and 3. We find that the IRR is stable across different levels of public-transfer income, QALYs, health costs, and criminal costs. Parental income has the biggest effect on the IRR, and is one of the only components for which the alternative IRR falls outside of the 80% confidence interval of the original estimate. This occurs when we double the parental income component in the flow of benefits. This sensitivity is due to both the magnitude of the treatment effect on parental income, as well as how the treatment effect took place earlier in the ABC/CARE subjects' lives. The sensitivity of the IRR to the program costs is also a result of the timing of the costs in the subjects lives. As females in the treatment groups attained higher levels of education than females in the control groups, we observe that the IRR decreases as we multiply the expenditure on education by increasingly large factors. On the other hand, this translates to additional labor income for females, which we observe to have a positive effect on the IRR. However, the IRR appears to be more sensitive to the costs of education relative to labor income, as schooling costs are borne at earlier stages of each subject's life.

Figure H.6: Internal Rate of Return vs. Components, Females
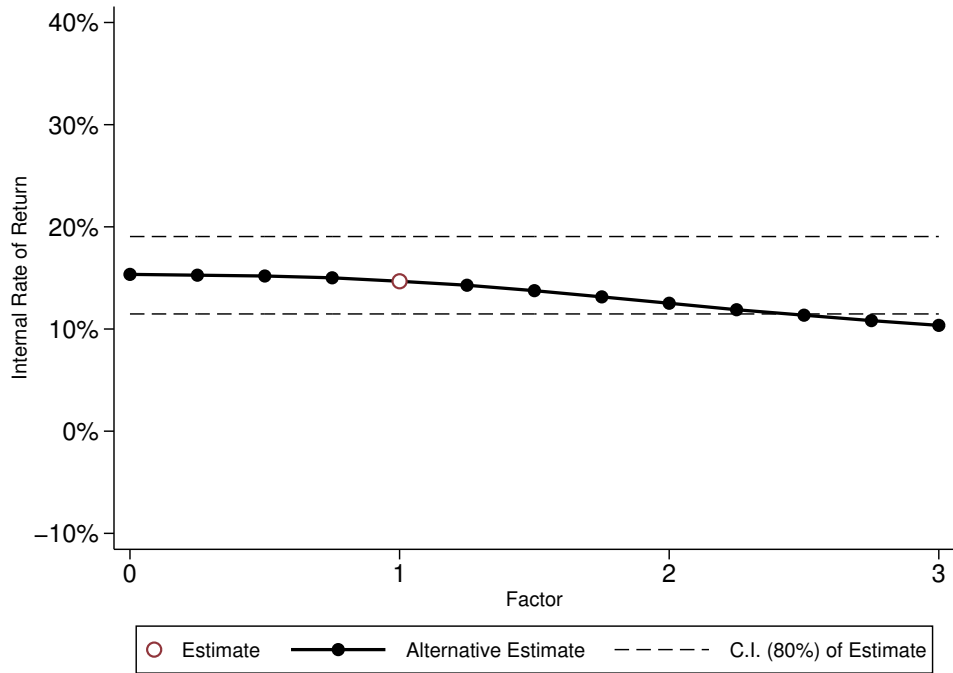
(a) Labor Income
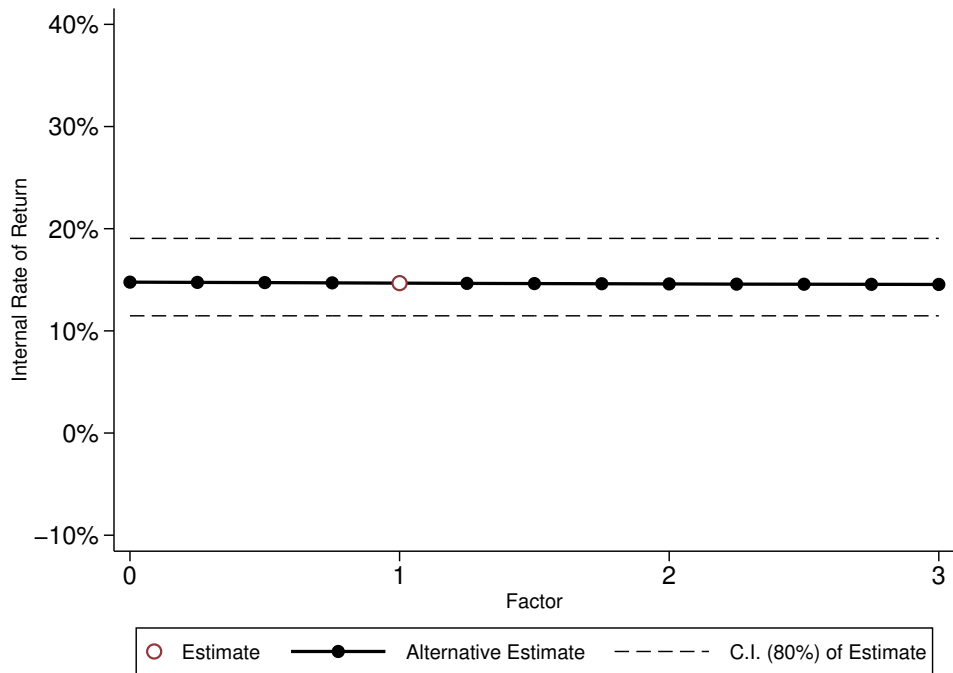

(b) Public-Transfer Income

(c) Parental Income
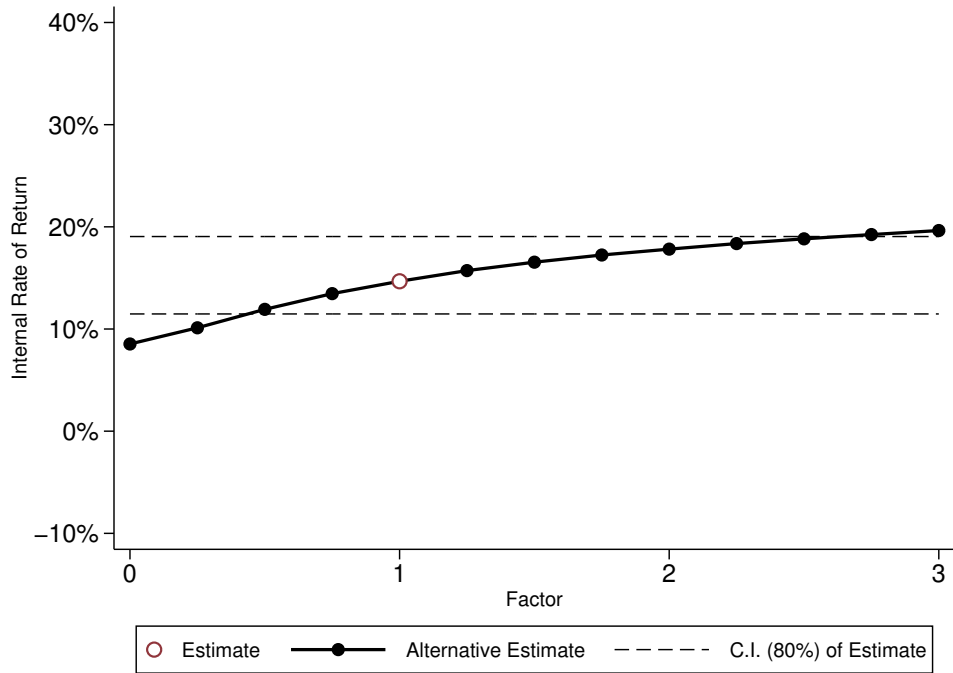

(d) Quality-Adjusted Life Years
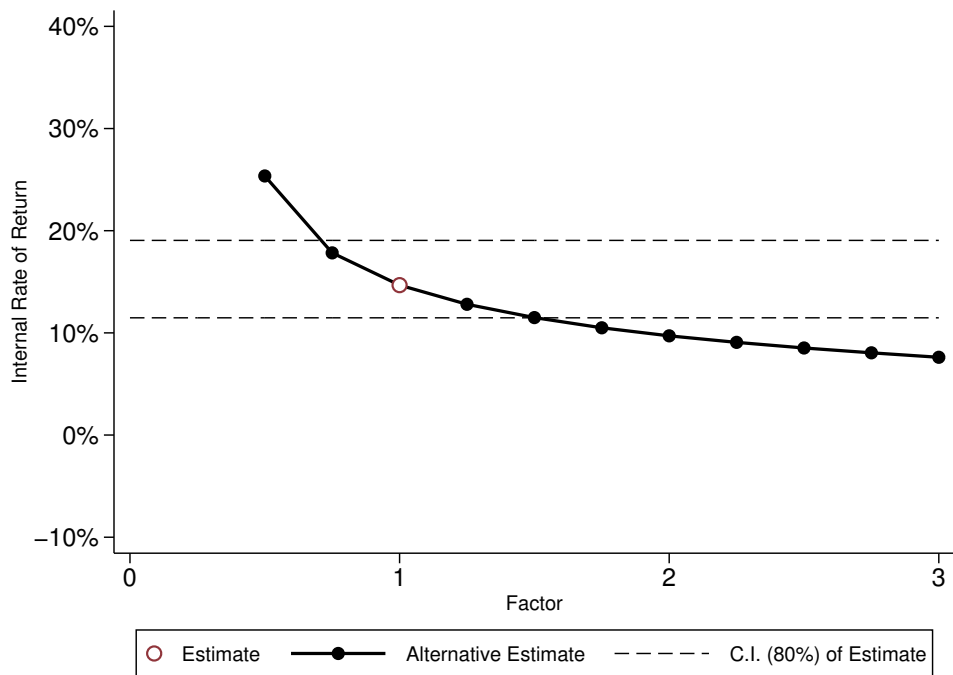
(e) Health Costs
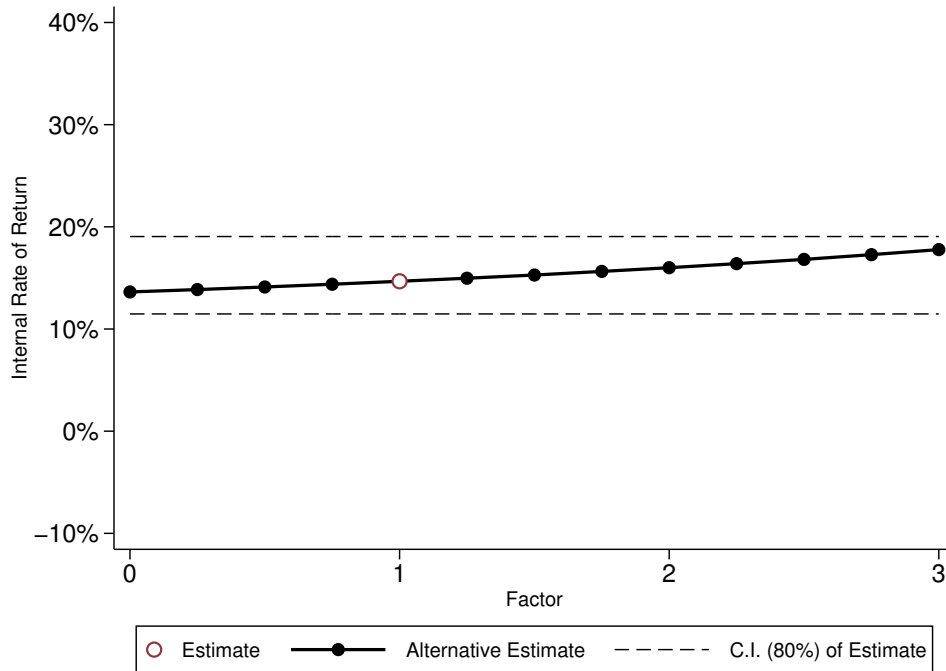

(f) Education Costs

(g) Crime Costs

(h) Program Costs
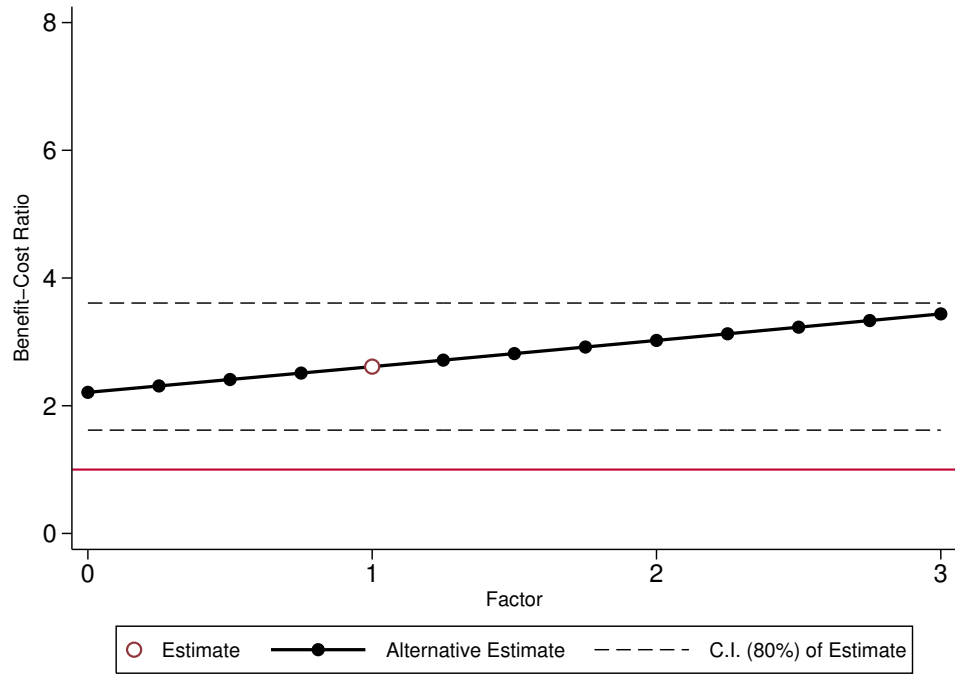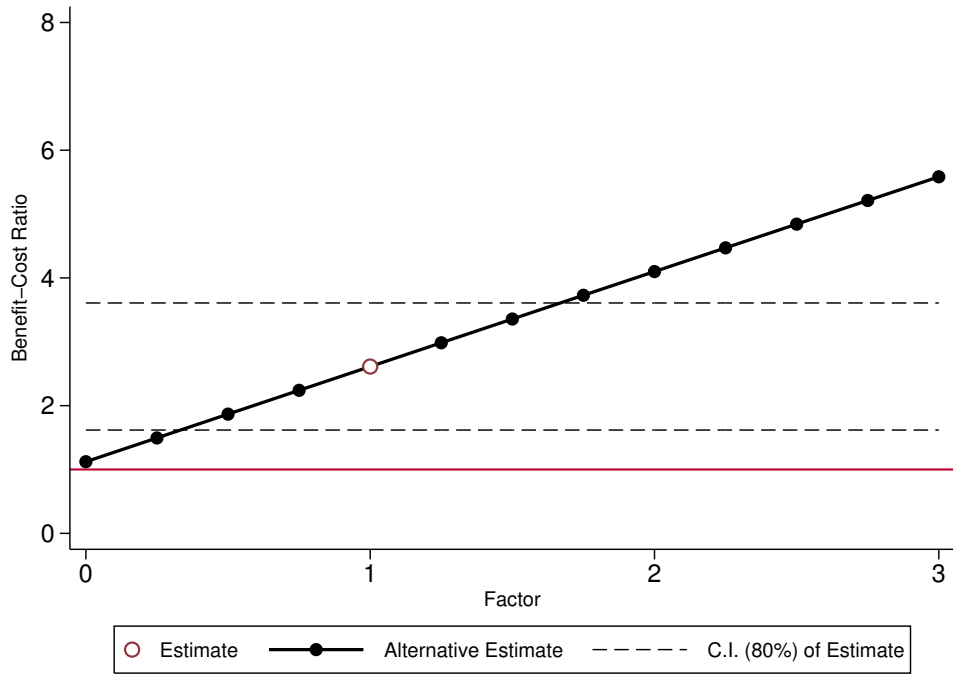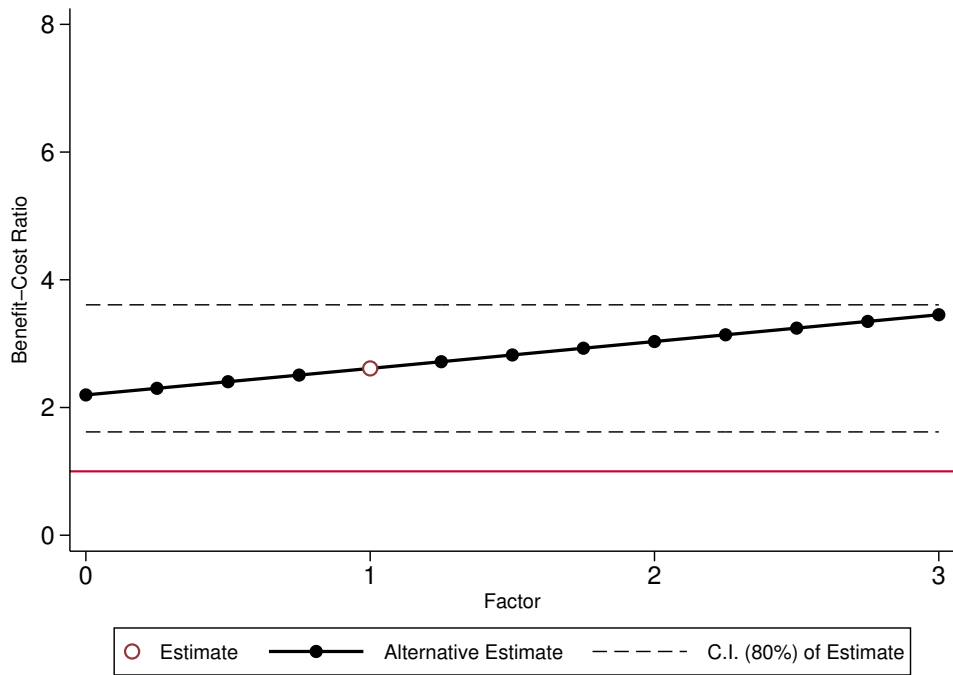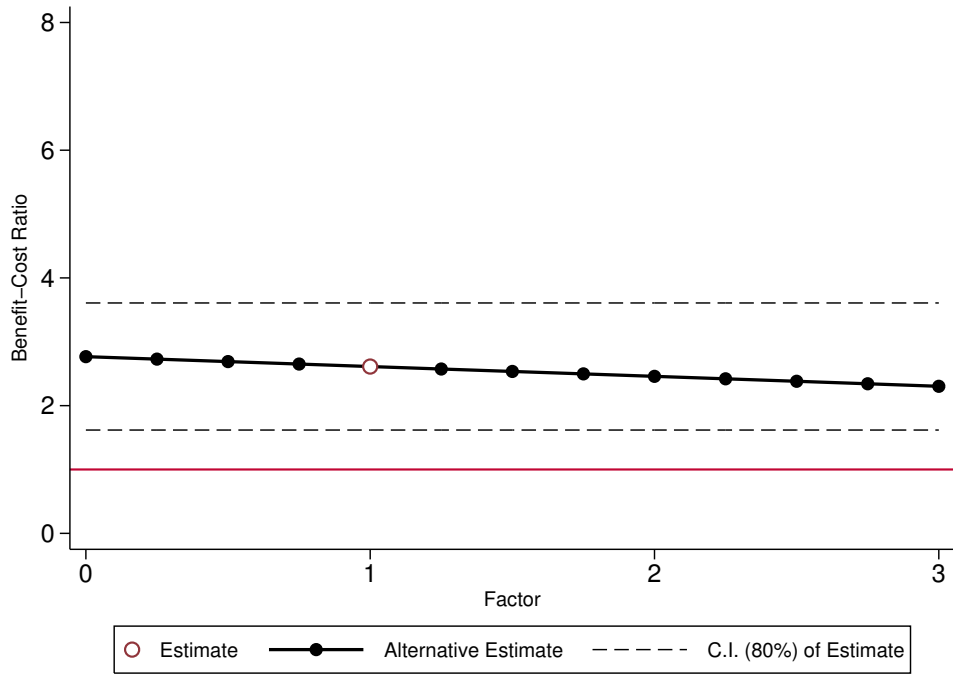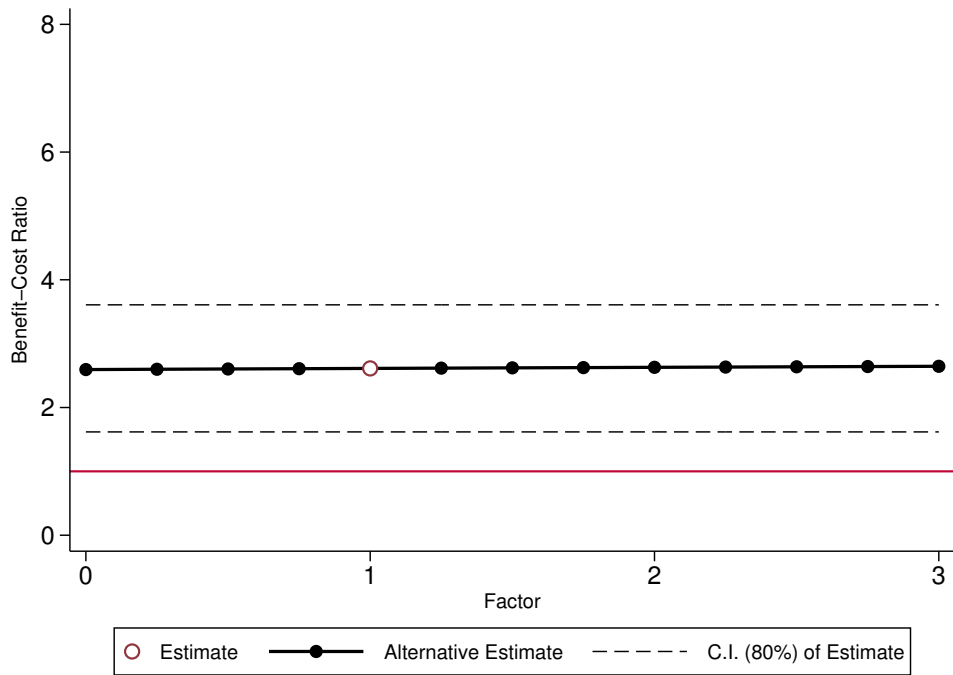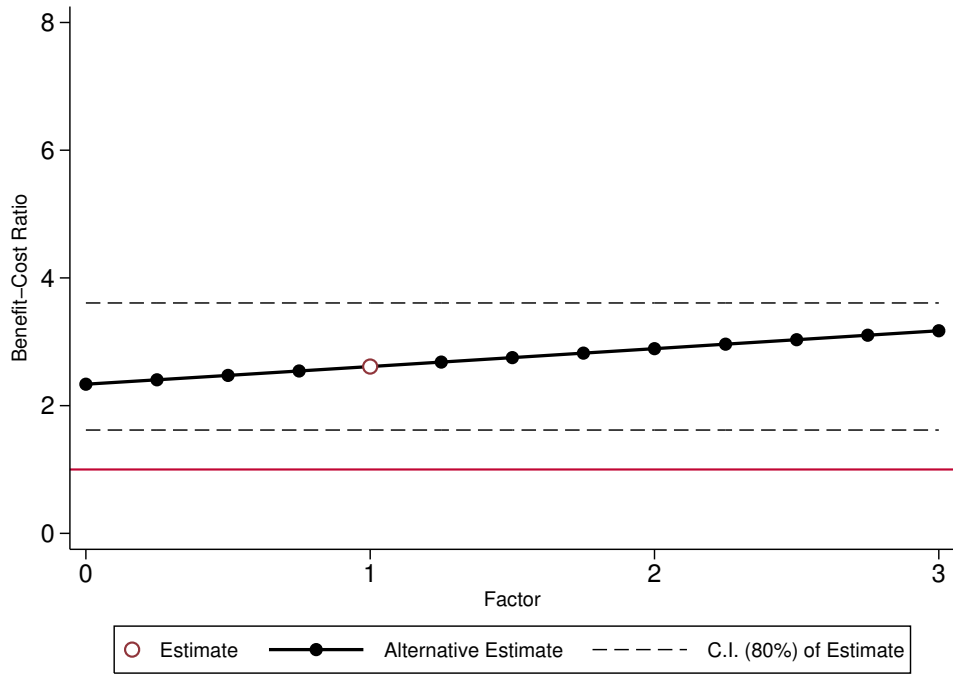
(i) Control Substitution Costs

Note: These graphs display how the internal rate of return changes for females as we multiply each component by a factor from 0 to 3. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the magnitude of each component. The estimates presented in the paper are equal to the IRRs presented above when the multiplicative factor is equal to 1. The estimates are means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10th and 90th quantiles of the bootstrap distribution.

Figure H.7 displays how the IRR for males changes as we vary the magnitude of each component of the benefits and costs. Our findings for males are similar to those for females: the IRR is insensitive to changes in public-transfer income, QALYs, and health costs. As the parental income and the program cost components for the female subsample is the same as those of the male subsample, we also observe that the IRR for males is sensitive to changes in both of these components. The IRR for males is not sensitive to increasing the weight of the cost of control substitution and responds to increases in costs at a smaller rate than females. This is a condequence of the different populations of males and females who were enrolled in alternative preschools. The IRR for the male subsample is a little less sensitive to changes in labor income than for the female subsample, but we still observe in Figure H.7

that the IRR for males rises as we multiply the benefits stemming from this component by increasingly large factors. Finally, the IRR increases for males as we increase the magnitude of the criminal cost component. This is not surprising because the reduction in the costs of crimes is the largest benefit of ABC/CARE for males.

Figure H.7: Internal Rate of Return vs. Components, Males

(a) Labor Income

(b) Public-Transfer Income

(c) Parental Income
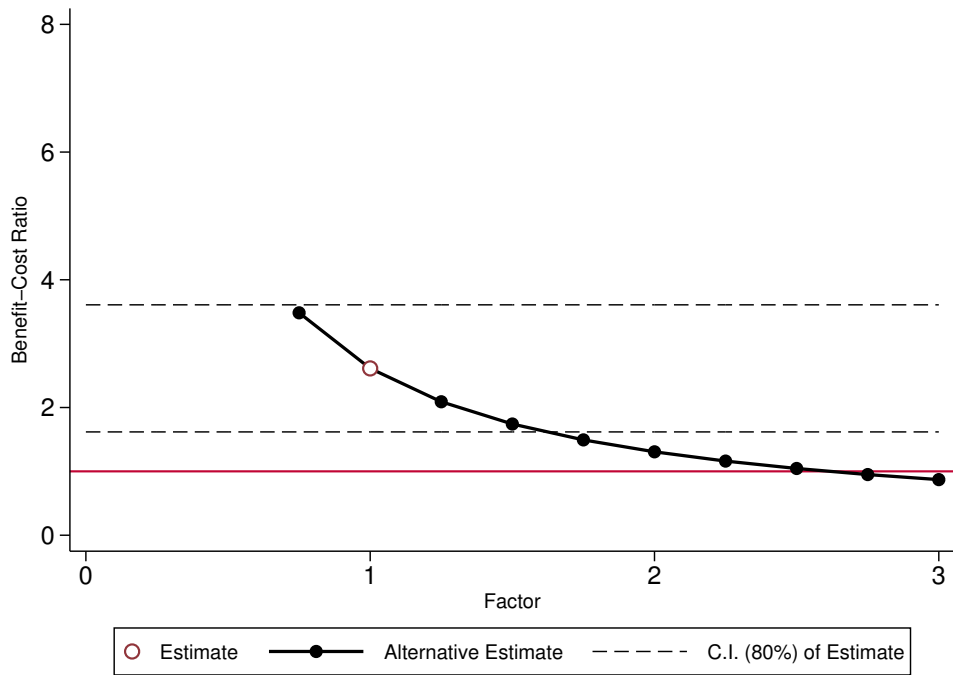

(d) Quality-Adjusted Life Years
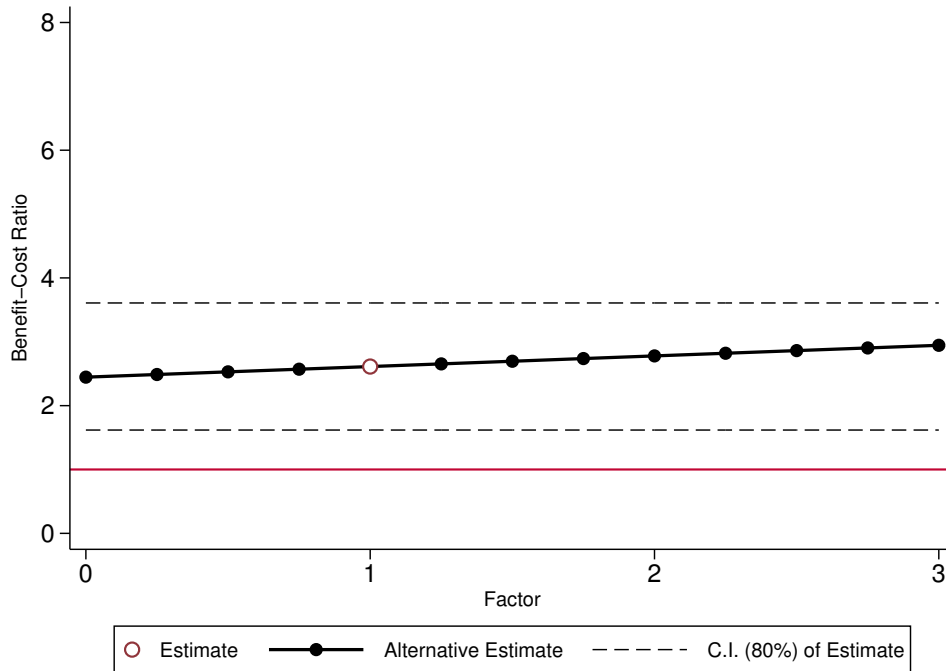
(e) Health Costs

(f) Education Costs

(g) Crime Costs

(h) Program Costs

(i) Control Substitution Costs

Note: These graphs display how the internal rate of return changes for males as we multiply each component by a factor from 0 to 3. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the magnitude of each component. The estimates presented in the paper are equal to the IRRs presented above when the multiplicative factor is equal to 1. The estimates are means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10th and 90th quantiles of the bootstrap distribution.
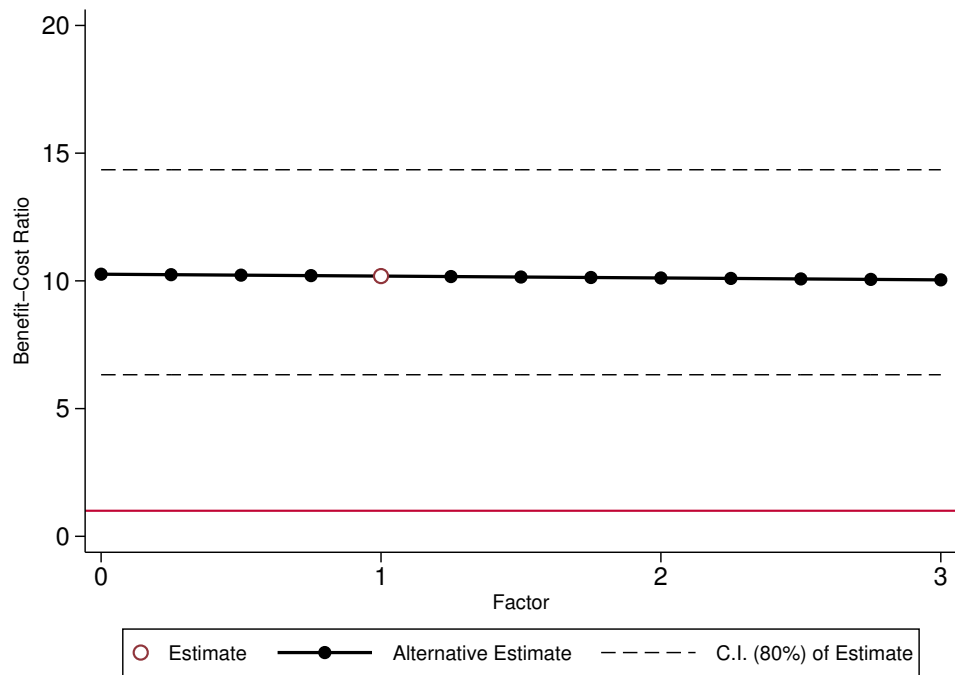
Figure H.8 shows how the benefit/cost ratio changes for females as we multiply each component of the benefits and costs by a factor between 0 and 3. We observe that the estimated cost-benefit ratio is generally insensitive to changes in all the components, except parental income, labor income, education costs, crime costs, and program costs. The sensitivity to program costs is due to the fact that it is the denominator of the benefit/cost ratio. In the case of the other components, when discounted, they exhibit the largest present values. The sensitivity of the benefit/cost ratio to changes in these components therefore indicates the magnitude of those components relative to the rest, with parental income having the largest magnitude in terms of discounted treatment effect, followed by labor income, and then education costs.
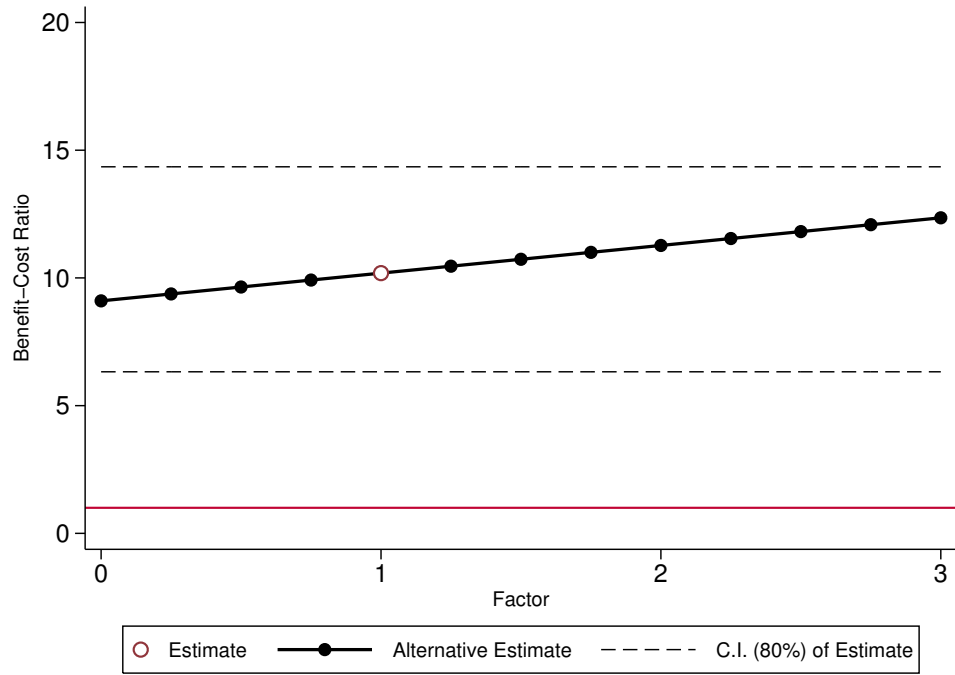
Figure H.8: Benefit/cost Ratio vs. Components, Females
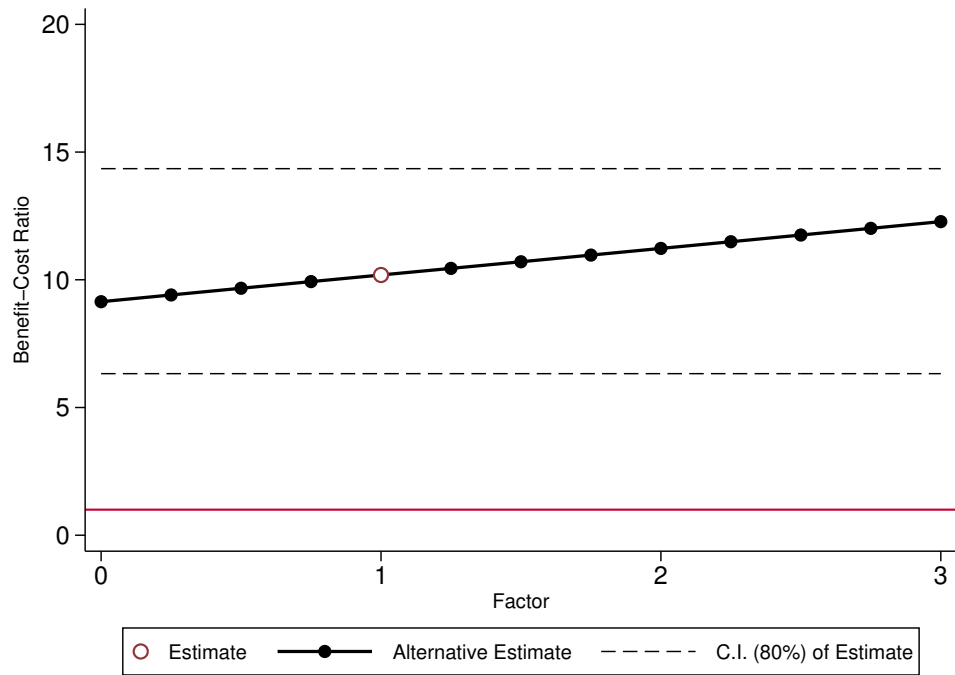
(a) Labor Income

(b) Public-Transfer Income

(c) Parental Income

(d) Quality-Adjusted Life Years

(e) Health Costs

(f) Education Costs

206

(g) Crime Costs

(h) Program Costs

207

(i) Control Substitution Costs

Note: These graphs display how the benefit/cost ratio changes for females as we multiply each component by a factor from 0 to 3. The red line indicates a benefit/cost ratio of 1. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the magnitude of each component. The benefit/cost ratio presented in the paper is equal to those presented above when the multiplicative factor is equal to 1. The estimates are means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10th and 90th quantiles of the bootstrap distribution.

Figure H.9 displays how the benefit/cost ratio for males varies as we multiply each component of the benefits and costs by a factor between 0 and 3. We observe that the ratio is insensitive to the scaling of public-transfer income, health costs, education costs, and control substitution costs. Except for program costs, the components that vary the benefit/cost ratio the most are crime costs, parental income, labor income, health expenditure, and QALYs, in that order. This is simply a result of the relevant magnitude of each of those components in the benefits stream after discounting.

Figure H.9: Benefit/cost Ratio vs. Components, Males

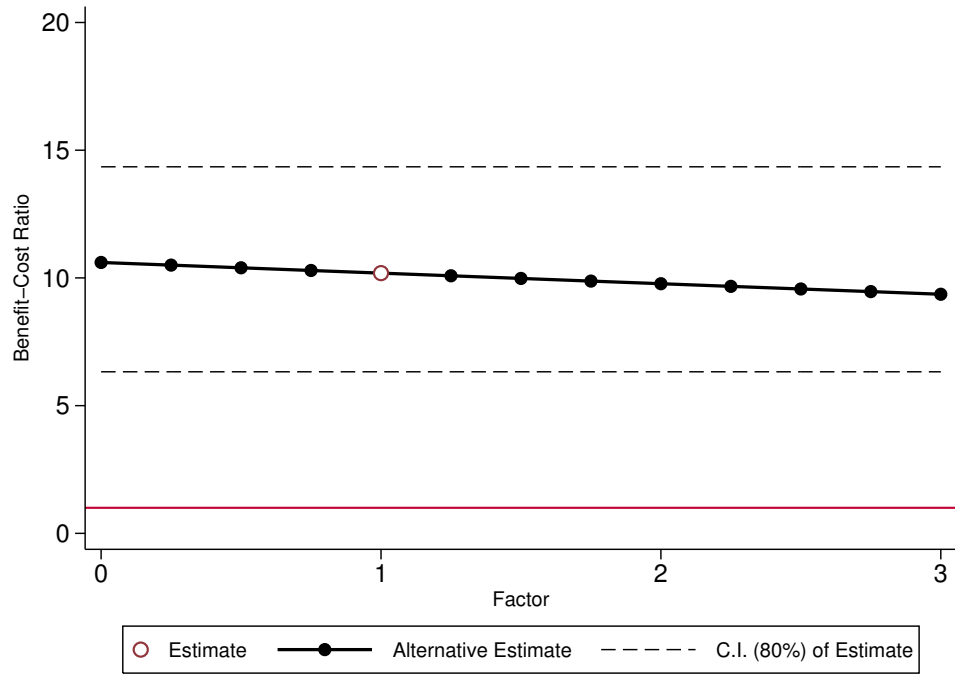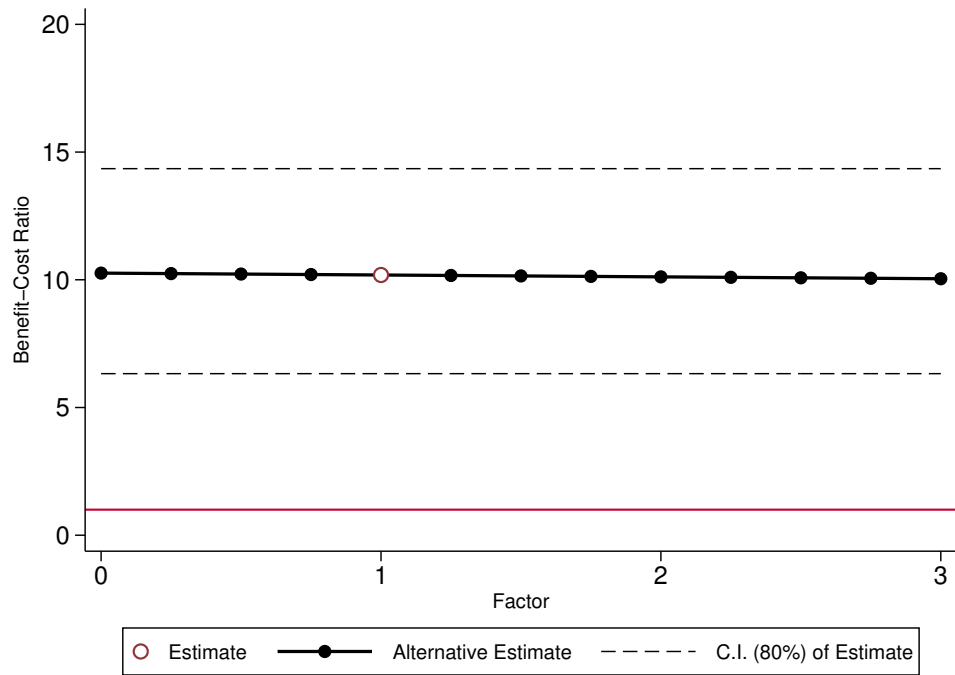(a) Labor Income

(b) Public-Transfer Income

(c) Parental Income
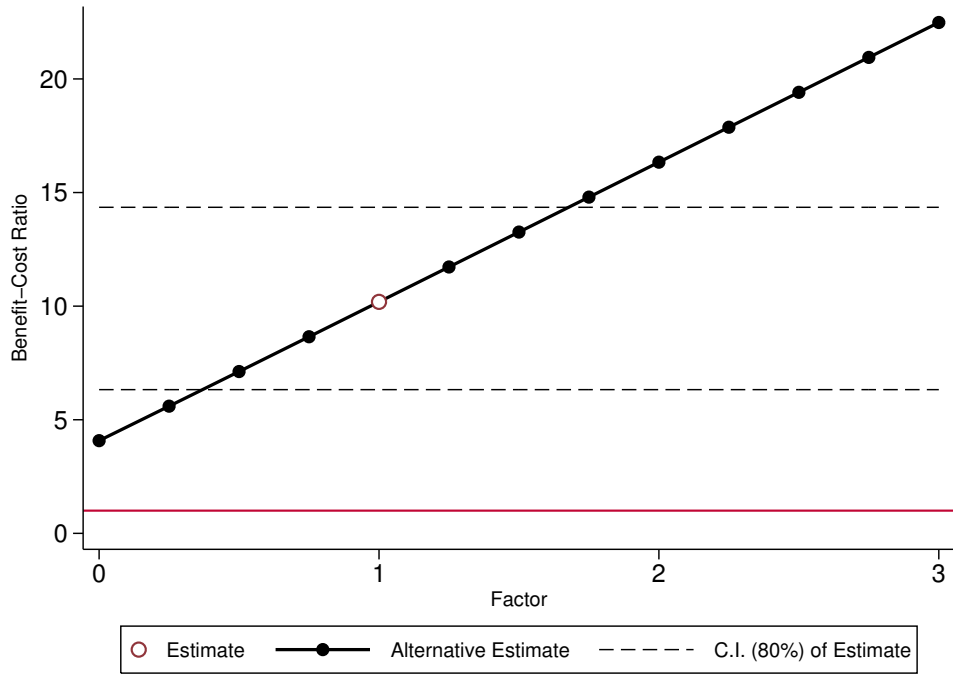

(d) Quality-Adjusted Life Years
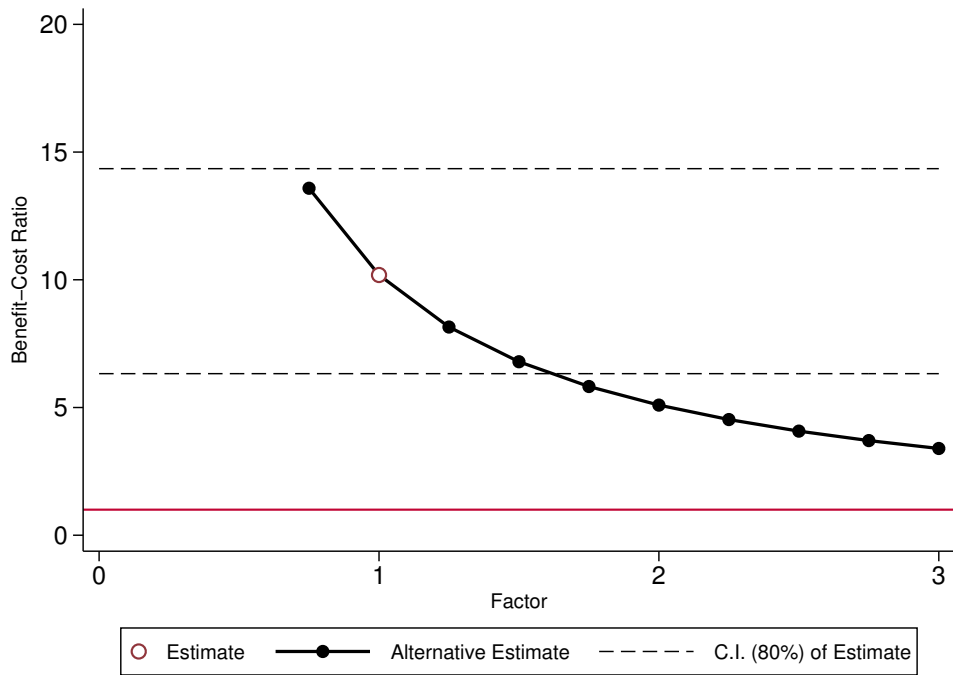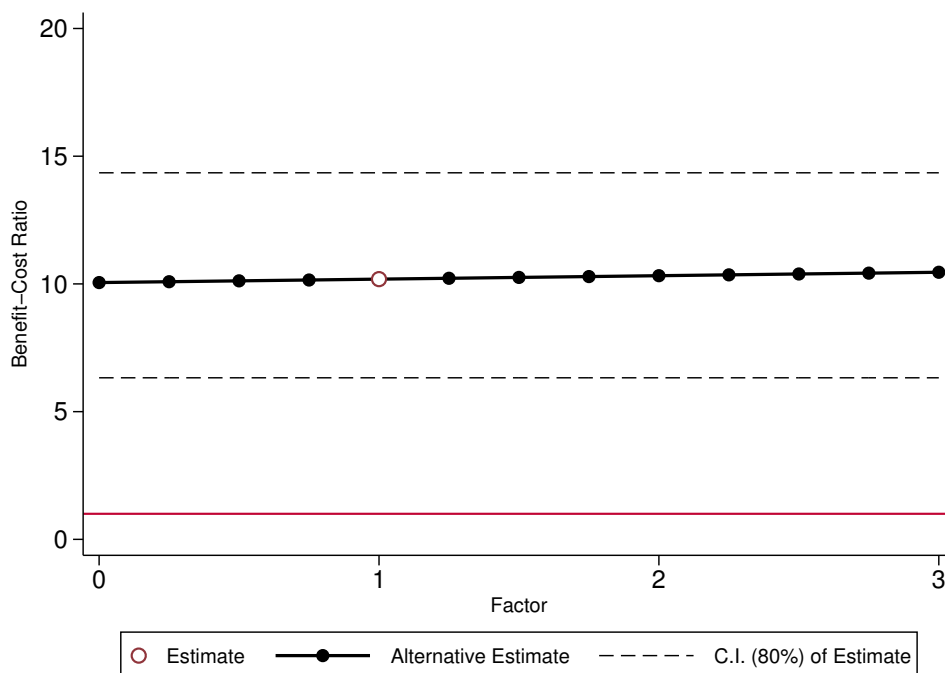
(e) Health Costs

(f) Education Costs

(g) Crime Costs

(h) Program Costs

(i) Control Substitution Costs

Note: These graphs display how the benefit/cost ratio changes for males as we multiply each component by a factor from 0 to 3. The red line indicates a benefit/cost ratio of 1. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the magnitude of each component. The benefit/cost ratio presented in the paper is equal to those presented above when the multiplicative factor is equal to 1. The estimates are means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10th and 90th quantiles of the bootstrap distribution.

Overall, we find that the benefit/cost ratio is stable across changes in the data, as well as changes in our assumptions regarding the discount rate and the marginal cost of welfare. The IRR tends to be more sensitive to changes in the data and assumptions.

Appendix References

Health affairs: Records of the Office of the Vice Chancellor. Subgroup 2: Centers, institutes and programs. Series 2: Child Development Institute. Day care center funding, 1978; 1981.

Ad Hoc Committee of Professionals in Child Care Services, North Carolina (1974, July). Letter to Governor Holshouser.

Administrative Branch, Office of Day Care Services (1982, December). Day care cost study. Technical report, Administrative Branch, Office of Day Care Services, Raleigh, NC.

Arrow, K. J. and D. Levhari (1969, September). Uniqueness of the internal rate of return with variable life of investment. *Economic Journal 79*(315), 560–566.

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.

Bartlett, M. S. (1938, February). Methods of estimating mental factors. *Nature 141*, 609–610.

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.

Browning, E. K. (1987, March). On the marginal welfare cost of taxation. *American Economic Review 77*(1), 11–23.

Bryant, D. M., C. T. Ramey, J. J. Sparling, and B. H. Wasik (1987). The carolina approach to responsive education a model for day care. *Topics in Early Childhood Special Education 7*(1), 48–60.

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.

Burchinal, M. R., M. Lee, and C. T. Ramey (1989). Type of day-care and preschool intellectual development in disadvantaged children. *Child Development 60*(1), 128–137.

Campbell, F. A., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. P. Pungello, and Y. Pan (2014). Early childhood investments substantially boost adult health. *Science 343*(6178), 1478–1485.

Campbell, F. A. and C. T. Ramey (1989, April). Preschool vs. school-age intervention for disadvantaged children: Where should we put our efforts? Report, North Carolina University, Frank Porter Graham Center, Chapel Hill. Paper presented at the Biennial Meeting of the Society for Research in Child Development (Kansas City, MO, April 27–29, 1989).

Campbell, F. A. and C. T. Ramey (1994, April). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development 65*(2), 684–698.

Claeskens, G. and N. L. Hjort (2008). *Model Selection and Model Averaging.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press.

Cohen, M. A. (1988). Pain, suffering, and jury awards: A study of the cost of crime to victims. *Law and Society Review 22*(3), 537–556.

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.

Cohen, M. A., R. T. Rust, S. Steen, and S. T. Tidd (2004, February). Willingness-to-pay for crime control programs. *Criminology 42*(1), 89–110.

Community Planning Services (1973, August). Durham's share in day care: A task force study. Summary report, Community Planning Services, Durham, NC.

Courtemanche, C., J. C. Pinkston, and J. Stewart (2015, December). Adjusting body mass for measurement error with invalid validation data. *Economics & Human Biology 19*, 275–293.

Davis, D. E. (1977). *My friends and me.* Circle Pines, MN: American Guidance Service.

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.

Dhanidina, L. and W. S. Griffith (1975). Costs and benefits of delayed high school completion. *Adult Education Quarterly 25*(4), 217–230.

Diebold, F. X. (2007). *Elements of Forecasting* (4 ed.). Mason, OH: Thomson South-Western.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care 35*(11), 1095–1108.

Dunn, L. M., L. T. Chun, D. C. Crowell, L. G. Dunn, L. G. Avery, and E. R. Yachel (1976). *Peabody early education kit.* Circle Pines, MN: American Guidance Service.

Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute 22*, 23–32.

Feagans, L. V. (1996). *Children's Talk in Communities and Classrooms.* Cambridge, MA: Blackwell.

Feldstein, M. (1999, November). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics 81*(4), 674–680.

Finkelstein, N. W. (1982). Aggression: Is it stimulated by day care? *Young Children 37*(6), 3–9.

Frank Porter Graham Child Development Center (1979). The Frank Porter Graham Child Development Center progress report 1979. Technical report, Child Development Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Freeman, R. B. (1974). Occupational training in proprietary schools and technical institutes. *The Review of Economics and Statistics 56*(3), 310–318.

Frölich, M. (2005). Matching estimators and optimal bandwidth choice. *Statistics and Computing 15*(3), 197–215.

García, J. L., J. J. Heckman, and A. L. Ziff (2018). Gender differences in the effects of early childhood education. *European Economics Review, Forthcoming.* Unpublished.

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.

Goldman, D. P., B. Shang, J. Bhattacharya, A. M. Garber, M. Hurd, D. N. Joyce, Geoffrey F. Lakdawalla, C. Panis, and P. G. Shekelle (2005). Consequences of health trends and medical innovation for the future elderly. *Health Affairs 24*, W5–R17.

Goldman, D. P., P. G. Shekelle, J. Bhattacharya, M. Hurd, and G. F. Joyce (2004). Health status and medical treatment of the future elderly. Technical report, RAND Corporation.

Greenberg, P. and B. Epstein (1973). *Bridges to reading.* Morristown, NJ: General Learning Corporation.

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

Haskins, R., A. M. Collier, C. T. Ramey, and P. O. Hirschbiel (1978). The effect of mild illness on habituation in the first year of life. *Journal of Pediatric Psychology 3*(3), 150–155.

Hausman, J. A. (1978, November). Specification tests in econometrics. *Econometrica 46*(6), 1251–1272.

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 (2010, February). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1–2), 114–128.

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 J. A. Smith (1998). Evaluating the welfare state. In S. Strom (Ed.), *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, pp. 241–318. New York: Cambridge University Press.

Henderson, F. W., A. M. Collier, M. A. Sanyal, J. M. Watkins, D. L. Fairclough, W. A. Clyde, Jr., and F. W. Denny (1982, June). A longitudinal study of respiratory viruses and bacteria in the etiology of acute otitis media with effusion. *New England Journal of Medicine 306*(23), 1377–1383.

Hoenack, S. A. and W. C. Weiler (1975). Cost-related tuition policies and university enrollments. *Journal of Human Resources 10*(3), 332–360.

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*(260), 663–685.

Johnson, J. H. (2003). The Changing Face of Poverty in North Carolina. *Popular Government 68*(3), 14–24.

Karnes, M. B. (1973). *GOAL program: Mathematical Concepts.* Springfield, MA: Milton-Bradley.

Klein, C. and J. Sanders (1982). Status report on programs and projects (as of January 1, 1982). Report, North Carolina University, Frank Porter Graham Center, Chapel Hill.

Kuperman, S. (2014-2015). Interviews of Frances Campbell, Carrie Bynum, Phyllis Royster, Gael McGinness, Joseph Sparling, Albert Collier, Barbara Wasik, Lynne Vernon-Feagans, Tom Richey, Margaret Burchinal, Thelma Harms, and Richard Clifford. University of Chicago's Center for the Economics of Human Development, Chicago.

Kuperman, S. and C. Cheng (2014). Interviews of Frances Campbell, Carrie Bynum, Phyllis Royster, Gael McGinness, Joseph Sparling, Albert Collier, Barbara Wasik, Lynne Vernon-Feagans, Tom Richey, Margaret Burchinal, Thelma Harms, Richard clifford, Ron Haskins, and Susann Hutaff Haskins. University of Chicago's Center for the Economics of Human Development, Chicago.

Kuperman, S. and A. Hojman (2015a). Interview with mary hodgers.

Kuperman, S. and A. Hojman (2015b). Interview with Richard Clifford and Sue Russell.

Lakdawalla, D. N., D. P. Goldman, and B. Shang (2004). The Health and Cost Consequences of Obesity Among the Future Elderly. *Health Affairs (Project Hope) 24*, W5R30–41.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (3 ed.). New York: Springer-Verlag.

Lochner, L. (2011a, January). Non-production benefits of education: Crime, health, and good citizenship. Working Paper 16722, National Bureau of Economic Research.

Lochner, L. (2011b). Nonproduction benefits of education: Crime, health, and good citizenship. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, Chapter 2, pp. 183–282. Amsterdam: Elsevier B. V.

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.

McGinness, G. D. (1982). The language of the poverty child: Implications from center-based intervention and evaluation programs. In L. Feagans and D. Farran (Eds.), *The Language of Children Reared in Poverty: Implications for Evaluation and Intervention*, pp. 219–240. New York: Academic Press.

McGinness, G. D. and C. T. Ramey (1981). Developing sociolinguistic competence in children. *Canadian Journal of Early Childhood Education 1*(2), 22–43.

Miller, T. R., M. A. Cohen, and B. Wiersema (1996). *Victim Costs and Consequences: A New Look*. Washington, DC: U.S. Dept. of Justice, Office of Justice Programs, National Institute of Justice.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press for National Bureau of Economic Research.

Moore, S. C. (2006). The Value of Reducing Fear: An Analysis Using the European Social Survey. *Applied Economics 38*(1), 115–117.

Moore, S. C. and J. Shepard (2006). The Cost of Fear: Shadow Pricing the Intangible Costs of Crime. *Applied Economics 38*(3), 293–300.

Newey, W. K. and K. D. West (1986, April). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Technical Report 55, National Bureau of Economic Research.

Noll, S. and J. Trent (Eds.) (2004). *Mental Retardation in America: A Historical Reader (The History of Disability)*. New York: NYU 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.

North Carolina State Department of Social Services (1972). Licensed day care facilities in North Carolina as of february 1, 1971. Raleigh, NC.

O'Brien, C. and J. Sanders (1974). The Carolina Abecedarian Project. Frank Porter Graham Child Development Center.

Prados, M. J., J. Zissimopoulos, and D. P. Goldman (2015). How Much Can an Education Intervention Reduce Lifecycle Health Disparities?

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. and F. A. Campbell (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency 88*(5), 515–523.

Ramey, C. T. and F. A. Campbell (1991). Poverty, early childhood education and academic competence: The Abecedarian experiment. In A. C. Houston (Ed.), *Children in Poverty: Child Development and Public Policy*, Chapter 8, pp. 190–221. New York: Cambridge University Press.

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., P. Mills, F. A. Campbell, and C. O'Brien (1975, July). Infants' home environments: A comparison of high-risk families and families from the general population. *American Journal of Mental Deficiency 80*(1), 40–42.

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.

Robins, P. (1988). Federal support for child care: Current policies and a proposed new system. *Focus 11*(2), 1–9.

Sanders, J. and D. B. Stokes (1979, March). Status report on programs and projects as of March 1, 1979. Report, North Carolina University, Frank Porter Graham Center, Chapel Hill.

Sanyal, M. A., F. W. Henderson, E. C. Stempel, A. M. Collier, and F. W. Denny (1980). Effect of upper respiratory tract infection on Eustachian tube ventilatory function in the preschool child. *Journal of Pediatrics 97*(1), 11–15.

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.

Snyder, T. D. and S. A. Dillow (2012). *Digest of Education Statistics 2011.* National Center for Education Statistics.

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.

Thaler, R. (1978, January). A Note on the Value of Crime Control: Evidence from the Property Market. *Journal of Urban Economics 5*(1), 137–145.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

Tysinger, B., D. Ermini Leaf, and D. P. Goldman (2015). The Future Adult Model: Technical Documentation. Technical report, Working Document.

U.S. Department of Justice (1988). Report to the Nation on Crime and Justice.

Vila, L. E. (2000). The Non-Monetary Benefits of Education. *European Journal of Education*, 21–32.

Wallach, M. A. and L. Wallach (1976). *Teaching All Children to Read.* Chicago: University of Chicago Press.

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

Wu, D. (1973, July). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica 41*(4), 733–750.

Zissimopoulos, J., E. Crimmins, and P. St. Clair (2014). The Value of Delaying Alzheimer's Disease Onset. In *Forum for Health Economics and Policy.*