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

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

May 31, 2017

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 with 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

[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	54	23	
Treatment Given	Diapers from birth to age 3, unlimited formula from birth to 15 months	Diapers from birth to age 3, unlimited formula from birth to 15 months	✓
Control Substitution	75%	74%	
Treatment	Center-based childcare	Center-based childcare and family education	
Center-based Childcare			
N	53 (participated)	17	
Intensity	6.5–9.75 hours a day for 50 weeks per year	6.5–9.75 hours a day for 50 weeks per year	✓
Components	Stimulation, medical care, nutrition, social services	Stimulation, medical care, nutrition, social services	✓
Staff-to-child Ratio	1:3 during ages 0–1	1:3 during ages 0–1	✓
	1:4–5 during age 1–4	1:4–5 during age 1–4	✓
	1:5–6 during ages 4–5	1:5–6 during ages 4–5	✓
Staff Qualifications	Range of degrees beyond high school; experience in early childcare	Range of degrees beyond high school; experience in early childcare	✓
Home Visitation			
N	(not part of the program)	27	
Intensity		Home visits lasting 1 hour. 2–3 per month during ages 0–3. 1–2 per month during ages 4–5	
Curriculum		Social and mental stimulation; parent-child interaction	
Staff-to-child Ratio		1:1	
Staff Qualifications		Home visitor training	
School-age Treatment			
N	46	39	
Intensity	Every other week	Every other week	✓
Components	Parent-teacher meetings	Parent-teacher meetings	✓
Curriculum	Reading and math	Reading and math	✓
Staff Qualifications	Range of degrees beyond high school; experience in early childcare	Range of degrees beyond high school; experience in early childcare	✓

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

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.

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).

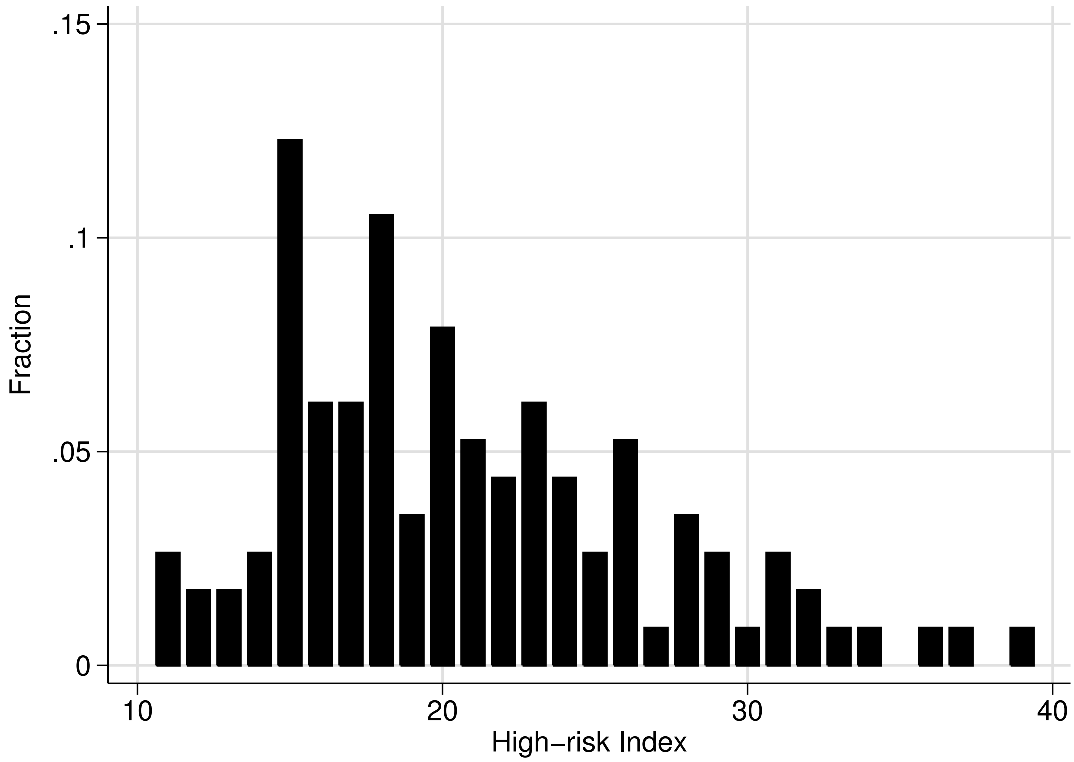
The HRI for CARE was similar to that of ABC—it also contained 13 weighted variables and a score of 11 or above was required to be considered eligible. The items for maternal and paternal education levels have the same categories and weights as the ABC HRI. The other identical items are having an absent 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 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 asked if payments were received from welfare agencies in the past 5 years instead of the past 3 years. Similarly, it asks 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).

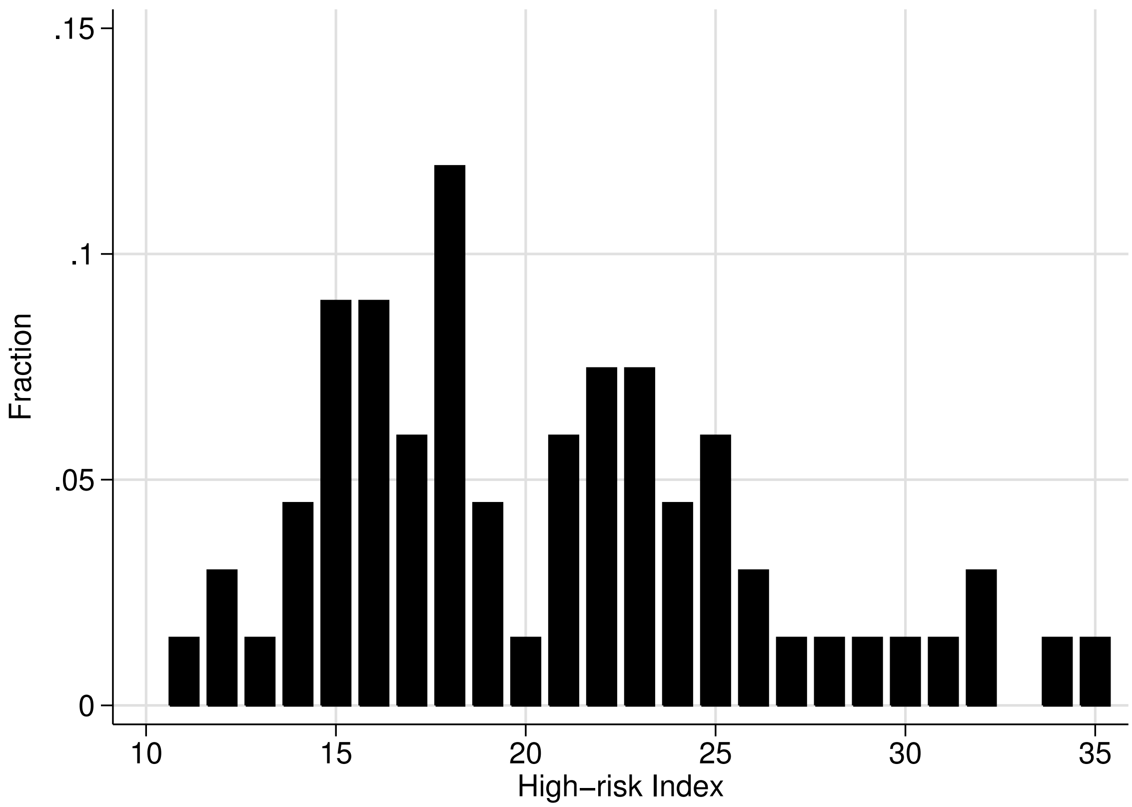
[3] Ramey and Campbell (1991); Campbell and Ramey (1994).

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 when 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, this 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.6, 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

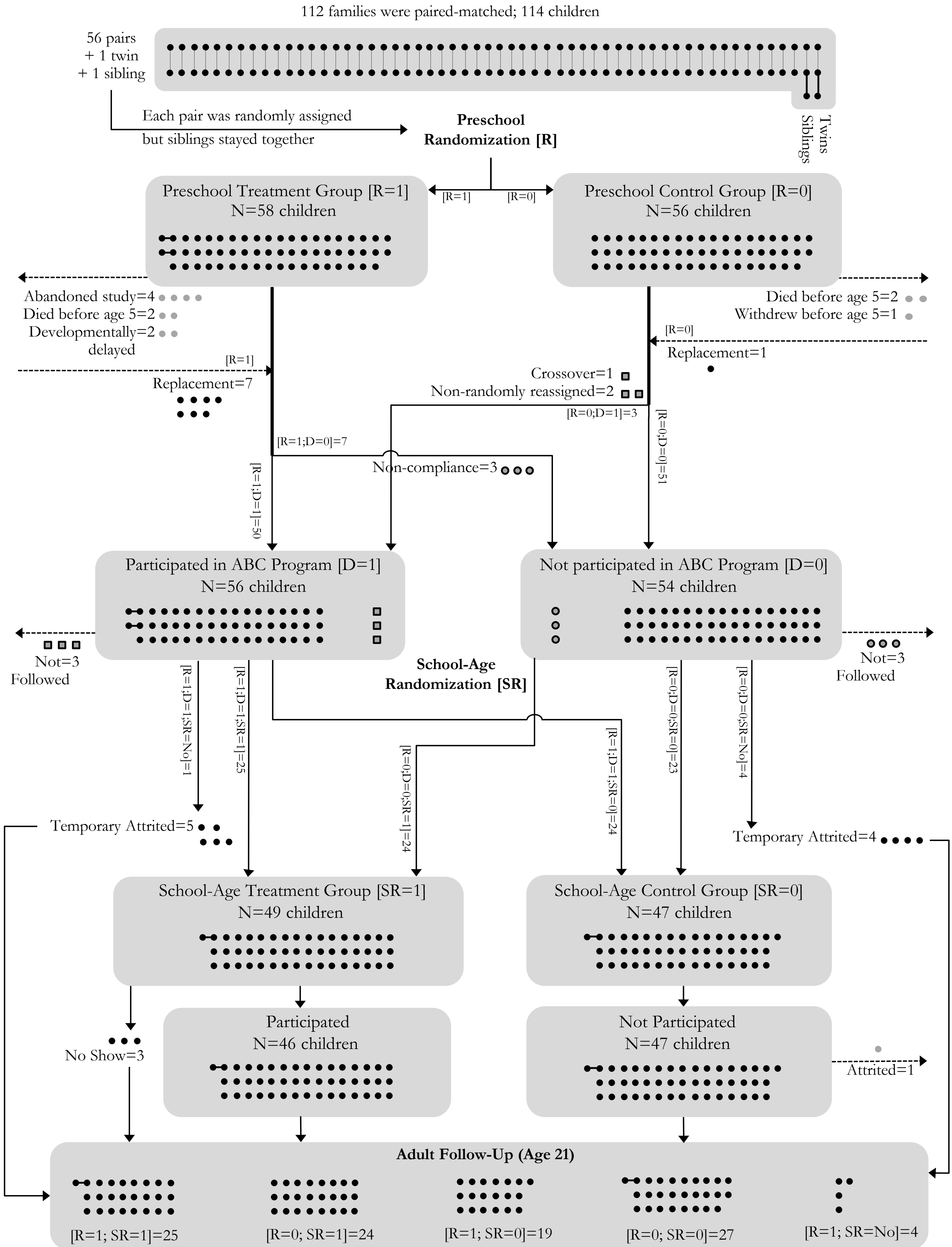
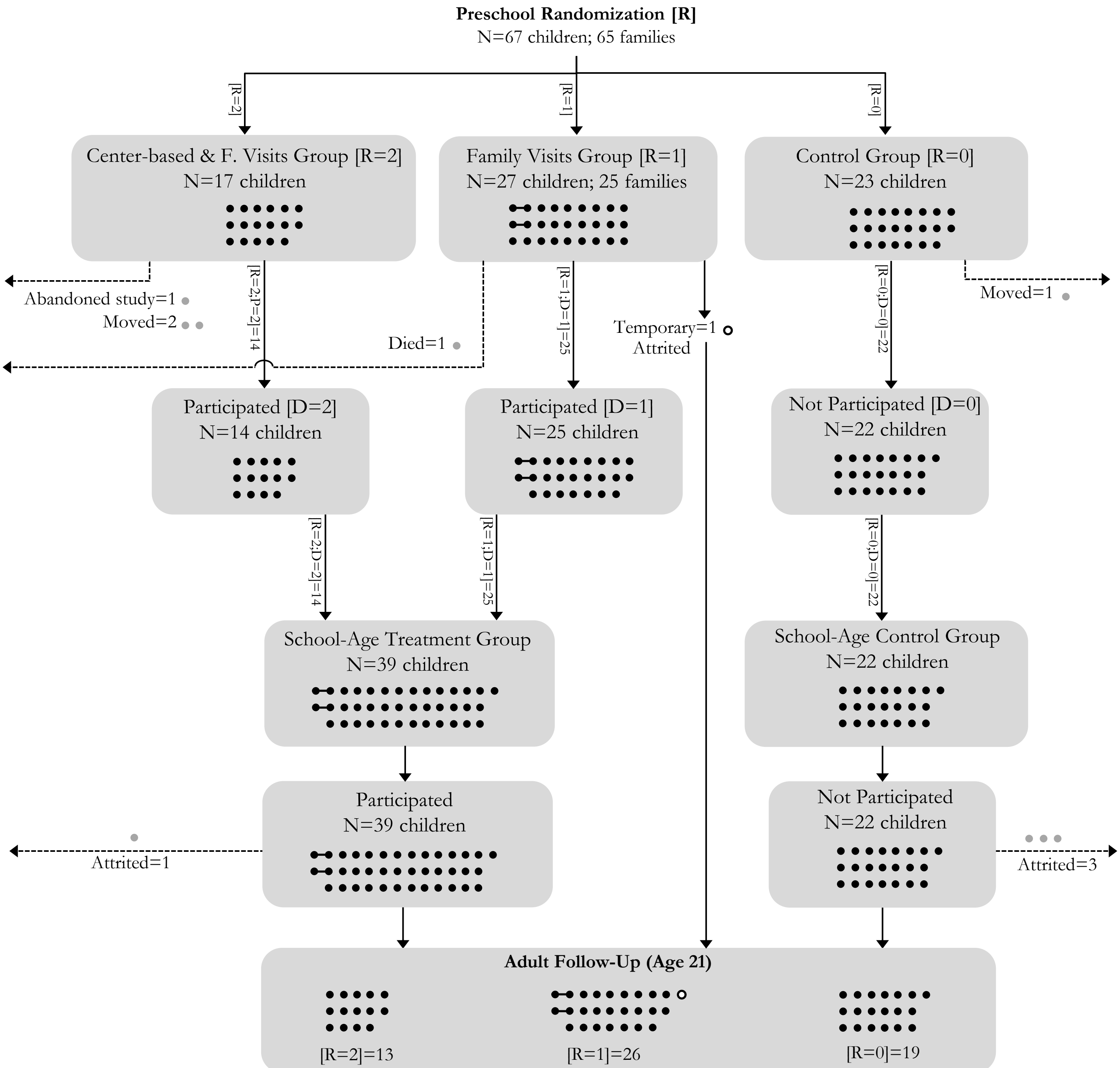


Figure A.4: Randomization Protocol and Treatment Compliance, CARE



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.

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]

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

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. Twenty-two subjects did not complete the first-phase of treatment as initially assigned by the randomization (see Table A.3).[5]

Of these cases, 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 them anymore.

[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.

Table A.3: Randomization Compromises, ABC

Case	Initial Assignment	Compromise Description	Data Availability	Methodology Assumption
1	Treatment	Left the study	None	Missing at random
2	Treatment	Left the study	None	Missing at random
3	Treatment	Left the study	None	Missing at random
4	Treatment	Left the study	None	Missing at random
5	Control	Death (age 0), heart disease	Baseline; before dead	Attrition after death
6	Control	Death (age 0), heart disease	Baseline; before dead	Attrition after death
7	Treatment	Death (age 0), SIDS	Baseline; before dead	Attrition after death
8	Treatment	Death (age 4), pedestrian accident	Baseline; before dead	Attrition after death
9	Treatment	Non-compliance	Baseline; before age 8	Attrition after age 8
10	Treatment	Non-compliance	Baseline; before age 8	Attrition after age 8
11	Treatment	Non-compliance	Baseline; before age 8	Attrition after age 8
12	Control	Crossover from control to treatment	Baseline; before age 8	Attrition after age 8
13	Treatment	3 months of treatment	Baseline; after age 2	Same as treatment group
14	Treatment	10 months of treatment	Baseline; after age 2	Same as treatment group
15	Treatment	6 months of treatment	Baseline; after age 2	Same as treatment group
16	Treatment	9 months of treatment	Baseline; after age 2	Same as treatment group
17	Control	Left study at 54 months	Baseline; before 54 months	Attrition after 54 months
18	Treatment	Developmentally delayed at 6 months	No data after diagnosis	Dropped (non-eligible)
19	Treatment	Developmentally delayed at 36 months	No data after diagnosis	Dropped (non-eligible)
20	Control	Crossover from control to treatment	Baseline, before age 8	Attrition after age 8
21	Control	Crossover from control to treatment	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 stance 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.

Finally, 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

[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.

after age 8.

Analysis of each of these cases leads to the following conclusions. For four subjects, we 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 to 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 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 111 subjects: 53 in the treatment group and 58 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 yielded a sample of 96 subjects in the second phase: 49 in treatment and

[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.

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.

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	Data Availability	Methodology Assumption
1	Family education	Death (age 0), unknown causes	Baseline	Attrition after dead
2	Center-based Childcare and Family Education	Left study at age 5	Baseline; before age 5	Attrition after age 5
3	Control	Move at 11 months old	Baseline; before 11 months	Attrition after 11 months
4	Center-based Childcare and Family Education	Move at 5 months old	Baseline; before 5 months	Attrition after 5 months
5	Center-based Childcare and Family Education	Move at age 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.

A.4 Program Description and Content

A.4.1 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 evolved each successive year as program staff evaluated ongoing outcome data.[15]

A.4.2 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 ap-

[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).

[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).

proximately 15 months of age.[21]

A.4.3 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 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]

[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).
[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).

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 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.4.4 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

[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).
[32] Ramey and Smith (1977); Ramey and Campbell (1979).
[33] Ramey and Campbell (1979); Ramey et al. (1982).

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 group and the family education group.[36]

ABC's and CARE's learning programs were influenced by key 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

[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).

[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).

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 age 3.[43]

A.4.5 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

[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).

[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.

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 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.4.6 School-age Treatment

The ABC subjects were randomization 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

[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).
[52] Bryant et al. (1987); Campbell et al. (2014); Kuperman (2015).
[53] Ramey et al. (1976); Ramey and Campbell (1979); Ramey et al. (1985).

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 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.5 Control Substitution

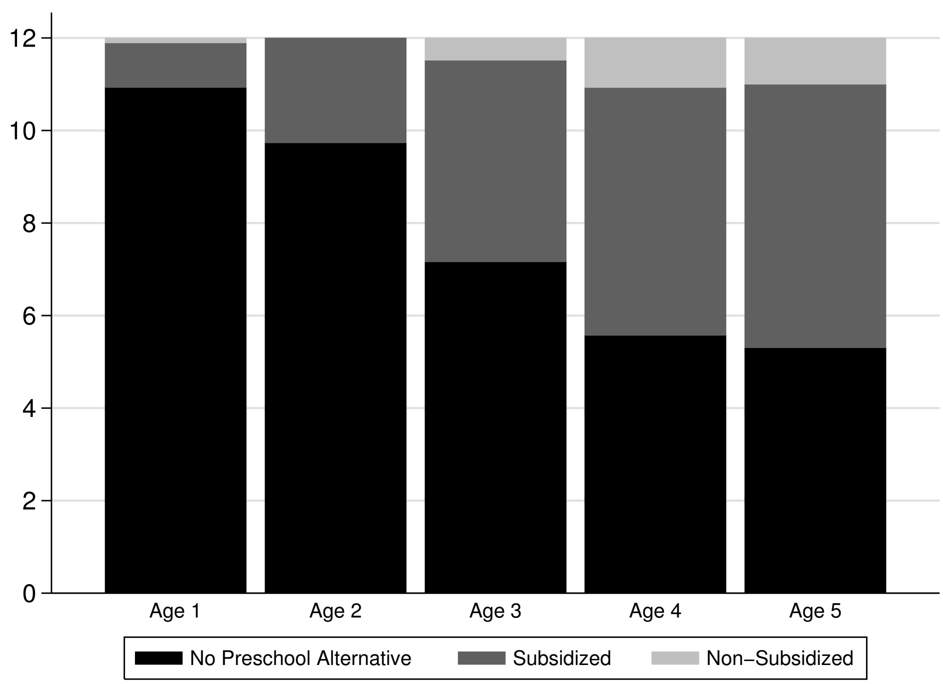
In ABC, the families of 75% of the control-group subjects enrolled their children in alternative center-based childcare. In CARE, 74% of families in the control group and 62% of families in the family education group enrolled their children in alternative center-based childcare. We refer to this phenomenon as control substitution; accounting for it is fundamental when evaluating the program, as we argue in Section 3.1. In this Appendix, we thoroughly de-

[54] Burchinal et al. (1997).
[55] There were about 60 activities per year. See Campbell and Ramey (1989) for details.
[56] Ramey and Campbell (1991).

scribe 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.

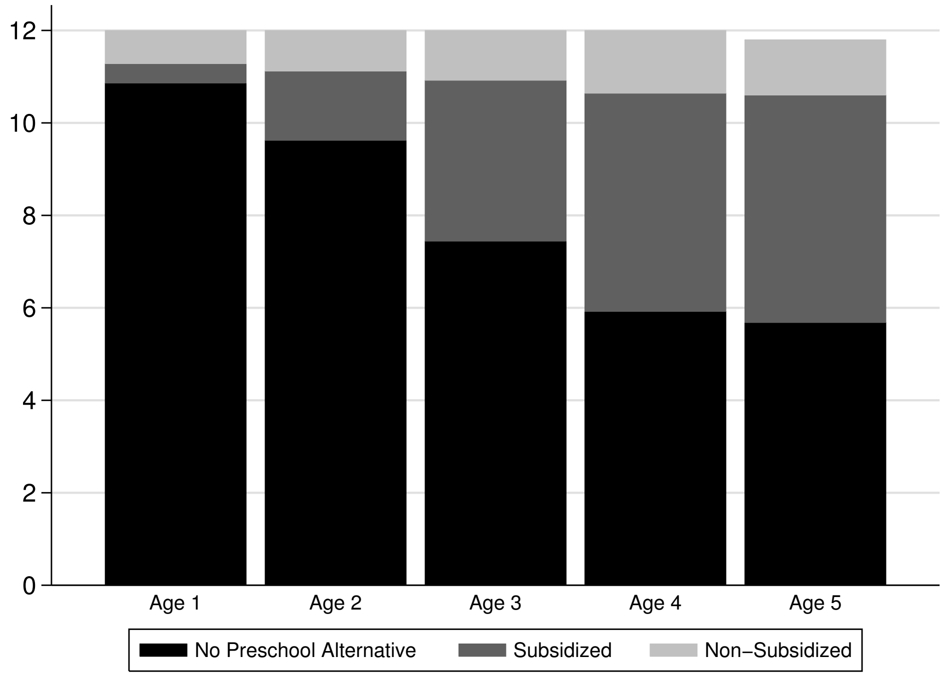
Figure A.5: Average Number of Months in Alternative Preschool, ABC Control Group



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.

[57] Burchinal et al. (1989).

Figure A.6: Average Number of Months in Alternative Preschool, CARE Control and Family Education Groups



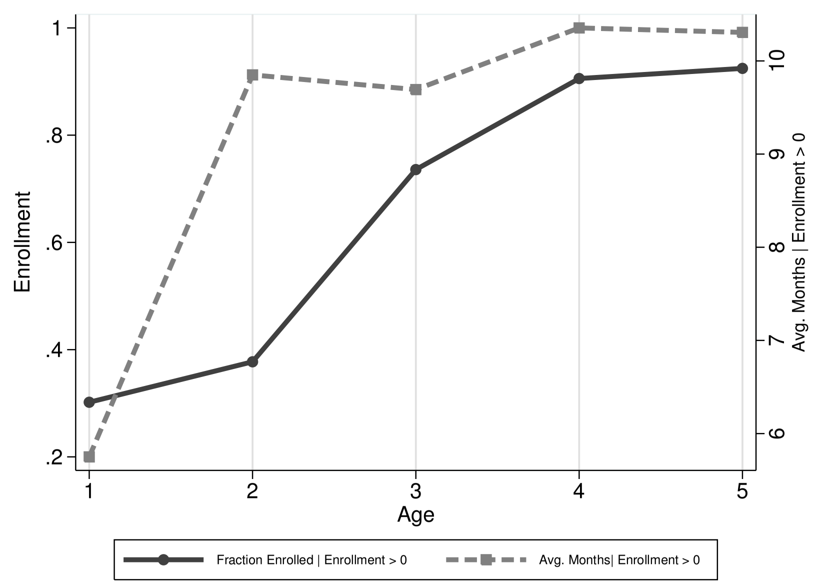
Note: This figure describes the take-up of alternative preschool by families in the CARE family education and control groups. The vertical axis represents the average number of months per year the subjects of the control group spent in alternative preschool. Subsidized centers were highly regulated and, 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 enrolled 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.

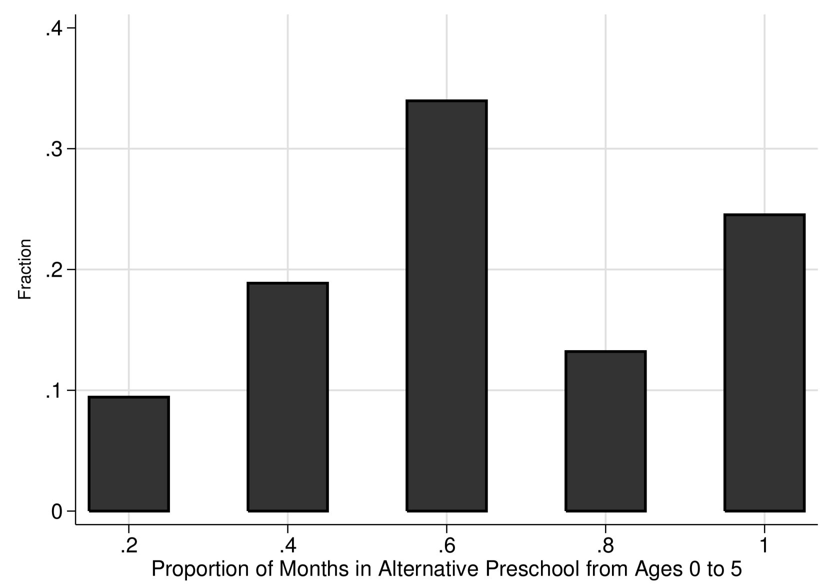
Figure A.7a shows enrollment by age and the average months of enrollment by age for the control-group children who enrolled in program alternatives. Enrollment increases with the age of children. Figure A.7b shows the fraction of children enrolled in preschool by age. As control children age, they are 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 Substitution

Characteristic	Females			Males			Pooled		
	Control Substitution No $N = 10$	Control Substitution Yes $N = 27$	p -value	Control Substitution No $N = 9$	Control Substitution Yes $N = 28$	p -value	Control Substitution No $N = 19$	Control Substitution Yes $N = 55$	p -value
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.5.1 Regulation

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.5.2 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. 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 subsidy rate at the time, which approximates the price of the centers, as centers pegged

[62] North Carolina State Department of Social Services (1972).
[63] Ad Hoc Committee of Professionals in Child Care Services, North Carolina (1974).

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.6 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 analogous by design. 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 reached their mid-30s. For CARE, the exact same follow-ups are available with the exception

[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).

of the age 15 follow-up.

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

Category	Sub-Category	Description	ABC Age	CARE Age	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	24, 36, 48, 60	24, 36, 48, 60	SBIS
		WPPSI	60	60	WPPSI
		BSID	3, 6, 9, 12, 18, 24	6, 12, 18, 24	BSID
		UOSPD	15	-	UOSPD^{ABC}
		RPM	60	-	RPM^{ABC}
	Quantitative	BSID	3, 6, 9, 12, 18, 24	6, 12, 18, 24	BSID
		MSCD	30, 42, 54	30, 42, 54	MSCD
	Memory	BSID	3, 6, 9, 12, 18, 24	6, 12, 18, 24	BSID
		MSCD	30, 42, 54	30, 42, 54	MSCD
	Motor Development	BSID	3, 6, 9, 12, 18, 24	6, 12, 18, 24	BSID
		MSCD	30, 42, 54	30, 42, 54	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 indicated 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	CARE Age	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	CARE Age	Measure
Cognitive Assessment	Language Ability	Adaptive Language Inventory	6, 7, 8	6, 7, 8	Adaptive Language Inventory
		Language Questionnaire	12	-	Language Questionnaire^{ABC}
		MSCD	7	-	MSCD^{ABC}
	Intelligence Tests	SBIS	6	7	SBIS
		WIS	6, 7, 8, 12, 15	6, 8	WIS
		Kaufman^{CARE}	-	6	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	6, 8, 12, 15	8, 12	PEI, CAS, PMI^{ABC}, SAI^{ABC}, Subject Interview^{ABC}, Quality Rank^{CARE}
		Relationship with the other sex	15	-	SAI^{ABC}, Subject What I Am Like (Harter)^{ABC}
	Critical Thinking	Thinks for self, questions things	6, 8	8, 12	PEI, Harter Child^{CARE}, CBI
		Concept Attainment Kit	6, 7, 8	-	Concept Attainment Kit^{ABC}
	Self-Control	Distracted in class	6, 7, 8, 12, 15	12	SCAN^{ABC}, CBI, WPB^{ABC}, PMI^{ABC}, SAI^{ABC}, Self-Evaluation Inventory^{ABC}
		Locus of control	15	-	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	8, 12, 15	8, 12	Achenbach Parent, Subject Risk Taking Survey^{ABC}
		Depression, anxiety, fear, etc.	6, 7, 8, 12, 15	7, 8, 9, 12	KRT, CAS, ETS, Achenbach Parent
	Social Activities	Athletic activities	8, 12, 15	8, 12	Achenbach Parent, SAI^{ABC}, Subject What I Am Like (Harter)^{ABC}, PEI^{CARE}
		Participant of organizations, e.g. religions	8, 12, 15	8, 12	Achenbach Parent, SAI^{ABC}, Subject Interview^{ABC}
		Reading list	12, 15	12	CAS, SAI^{ABC}
		TV/music	12, 15	12	CAS, SAI^{ABC} , Television Checklist^{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.

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

Category	Sub-Category	Description	ABC Age	CARE Age	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 Interveiw^{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	CARE Age	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	CARE Age	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}.

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	p-value 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	p-value 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	p-value 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	p-value 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 subjects in the first-phase treatment assignment.

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

Variable	Age	Obs.	Att.	Observed Mean	Attritted Mean	p-value Single H_0	p-value 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

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	p-value 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.

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

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	p-value 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

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	p-value 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

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	p-value 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

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value Single H_0	p-value 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	p-value 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	p-value 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

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value 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

Variable	Age	Control Obs.	Treated Obs.	Control Mean	Treated Mean	p-value 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.

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		σ		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		σ		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.7 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

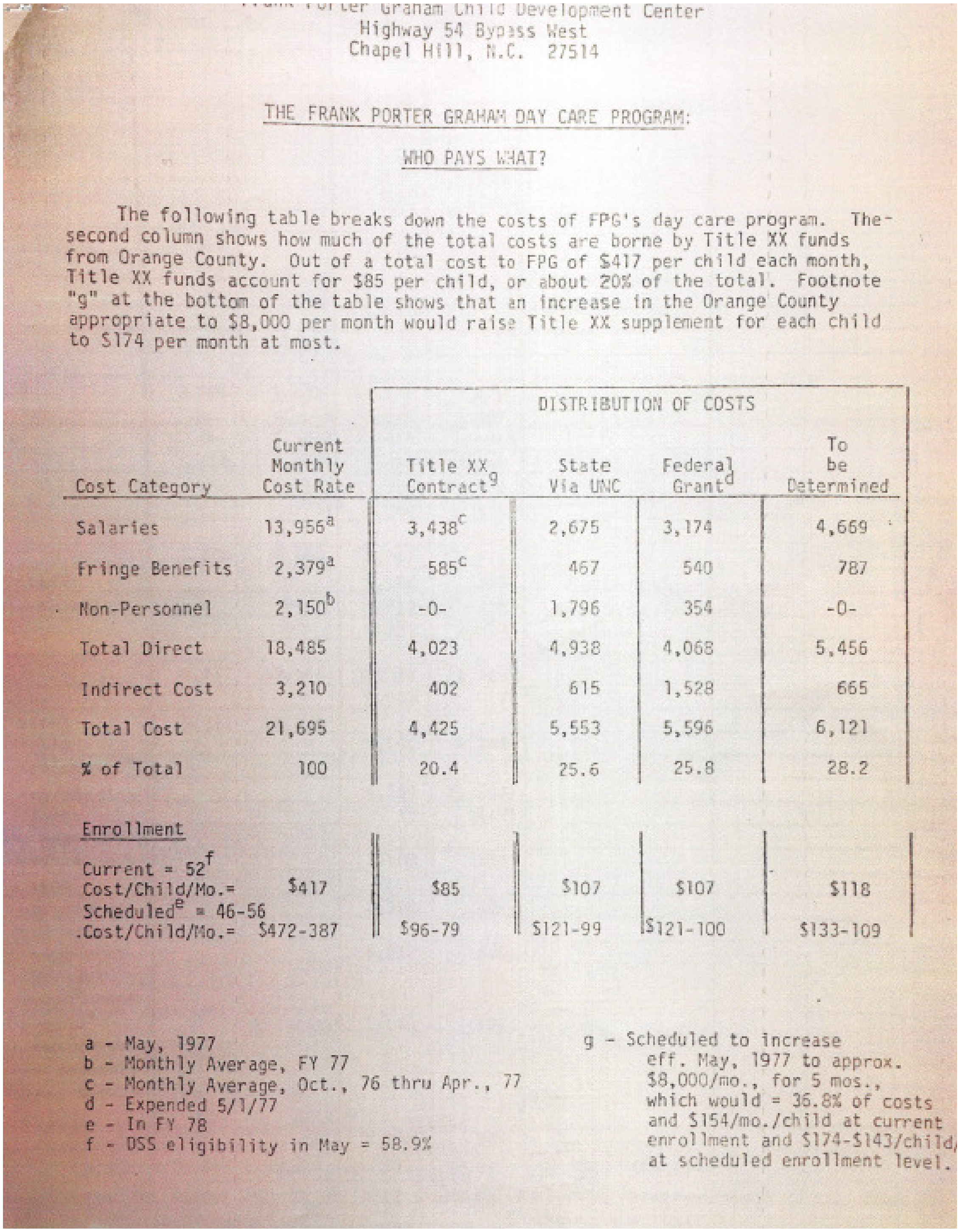
Note: This table lists and details the current Educare programs in the United States.

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

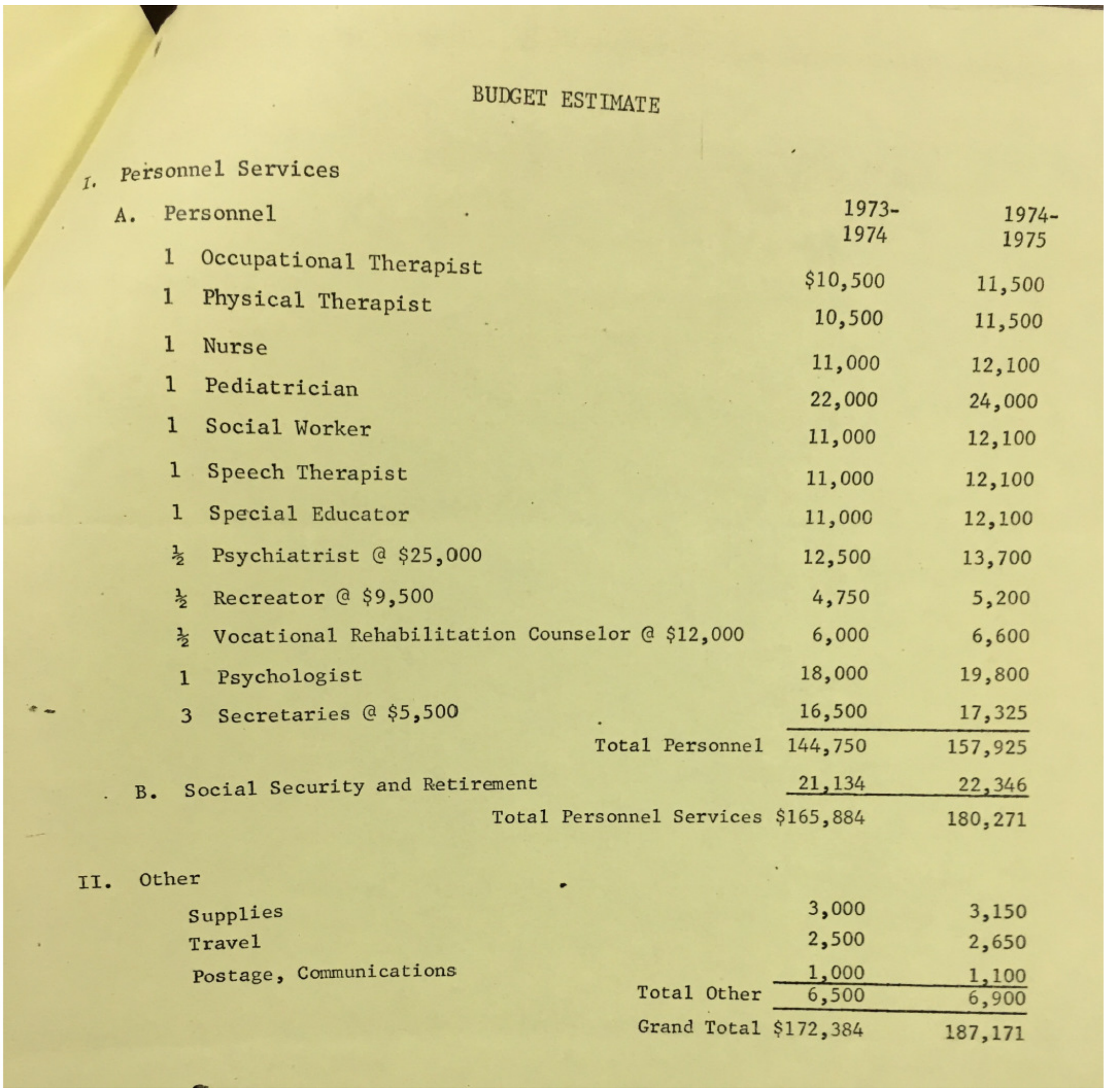
a - May, 1977
b - Monthly Average, FY 77
c - Monthly Average, Oct., 76 thru Apr., 77
d - Expended 5/1/77
e - In FY 78
f - DSS eligibility in May = 58.9%

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



BUDGET ESTIMATE

		1973-1974	1974-1975
I. Personnel Services			
A. Personnel			
1	Occupational Therapist	$10,500	11,500
1	Physical Therapist	10,500	11,500
1	Nurse	11,000	12,100
1	Pediatrician	22,000	24,000
1	Social Worker	11,000	12,100
1	Speech Therapist	11,000	12,100
1	Special Educator	11,000	12,100
½	Psychiatrist @ $25,000	12,500	13,700
½	Recreator @ $9,500	4,750	5,200
½	Vocational Rehabilitation Counselor @ $12,000	6,000	6,600
1	Psychologist	18,000	19,800
3	Secretaries @ $5,500	16,500	17,325
	Total Personnel	144,750	157,925
B.	Social Security and Retirement	21,134	22,346
	Total Personnel Services	$165,884	180,271
II. Other			
	Supplies	3,000	3,150
	Travel	2,500	2,650
	Postage, Communications	1,000	1,100
	Total Other	6,500	6,900
	Grand Total	$172,384	187,171

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 to identifying and estimating life-cycle treatment effects. Differences in the approach 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 a substantial rate 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 some technical details on the computations of the internal rate of return and the benefit/cost ratios, respectively. Appendix C.7 frames our data combination problem in the Generalized Method of Moments framework. Finally, Appendix C.8 provides the precise steps for constructing our statistical inferences.

C.1 Complete Data

We classify a variable as complete data when we observe the data for at least 85% 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. We estimate non-parametric p-values based on the bootstrap distribution. We perform inference in this same way throughout the paper.

Table C.1: Variables Estimated without IPW Adjustment

Completely Observed Outcomes	Age
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 within the experiment 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 of the partially observed outcomes, we construct an IPW scheme. The scheme is based on a set of variables that we observe for the complete sample. We use this set of complete variables to estimate the propensity of an outcome to be classified as partially complete. That is, the scheme is based on a logistic regression of "being partially complete" on a set of variables that we do observe for the full sample. The control set of variables is chosen among many possible control sets, as documented in Appendix C.9.0.1. For each of the outcomes that we partially observe, we list the variables that we use to produce the IPW scheme in Table C.2.

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 provides a list of the variables that we partially observe and the variables that we use to construct the IPW scheme to account for attrition when calculating treatment effects pooling females and males. The procedure to select these variables is described in Appendix C.9.0.1. We construct the IPW using a common model across males and females.

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

We construct the IPW using the same algorithm, independently of the age within $a \leq a^*$ in which an outcome is partially complete. For notational simplicity, we derive the IPW scheme without indexing the outcomes by age. 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 to treatment, and $R = 0$ otherwise.[71] Similarly, let $A = 1$ denote the case where we observe a generic scalar outcome Y, and $A = 0$ otherwise. As in the main text, $\boldsymbol{B}$ represents background (pre-treatment) variables and $\boldsymbol{X}$ variables that could be affected by treatment and that predict Y.

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

Assumption AA–1

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

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.

$$
\begin{aligned}
\mathbb{E}[Y^r] &= \int\int y f_{Y_r|\boldsymbol{B}}(y) f_{\boldsymbol{B}}(b) dy db \\
&= \int\int y f_{Y|\boldsymbol{B},R=r}(y) f_{\boldsymbol{B}}(b) dy db \\
&= \int\int\int y f_{Y|R=r,\boldsymbol{X},\boldsymbol{B}}(y) f_{\boldsymbol{X}|R=r}(x) f_{\boldsymbol{B}}(b) dy dx db \\
&= \int\int\int y f_{Y|R=r,\boldsymbol{X},\boldsymbol{B},A=1}(y) f_{\boldsymbol{X}|R=r,\boldsymbol{B}}(x) f_{\boldsymbol{B}}(b) dy dx db
\end{aligned} \tag{1}
$$

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 $f_{\boldsymbol{X}|R=r,\boldsymbol{B}}(x)$ and $f_{\boldsymbol{B}}(b)$ to obtain

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

and

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

Substituting these expressions into (1), we obtain

$$
\begin{aligned}
\mathbb{E}[Y_r] &= \int\int\int y f_{Y,\boldsymbol{X},\boldsymbol{B}|R=r,A=1}(y,x,b) \frac{P(R=r,A=1) P(A=1|R=r,\boldsymbol{B})}{P(R=r,A=1|\boldsymbol{B}) P(A=1|R=r,\boldsymbol{X},\boldsymbol{B})} dy dx db \\
&= \int\int\int y f_{Y,\boldsymbol{X},\boldsymbol{B}|R=r,A=1}(y,x,b) \frac{P(R=r,A=1)}{P(R=r|\boldsymbol{B}) P(A=1|R=r,\boldsymbol{X},\boldsymbol{B})} dy dx db.
\end{aligned}
$$

Assumption AA–1 generalizes the matching assumption of Campbell et al. (2014). It conditions not only on pre-program variables but also on fully observed post-treatment variables, $\boldsymbol{X}$, that predict Y. The corresponding sample estimator for $\mathbb{E}[Y^r]$ is

$$\sum_{i \in \mathcal{I}} y \alpha_i \beta_{i,r} \mathbf{1}(r_i = r)$$

where $\mathcal{I}$ indexes the individuals in the sample, α_i indicates whether we observe Y for individual i, and

$$\beta_{i,r} = \frac{1}{\pi_r(x_i)\alpha(r_i, x_i, b_i)} \frac{1}{\sum_k \frac{\mathbf{1}(r_i=r)\mathbf{1}(\alpha_i=1)}{\pi_r(x_k)\alpha(r_k,x_k,b_k)}},$$

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

For each of the estimates presented in the paper, we allow the reader to assess the sensitivity of the estimate to adjusting by the IPW, in Tables 3 and 4.[72] In Tables 3 and 4 and, we also present estimates of the internal rate of return and the benefit cost-ratio with and without adjusting by 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. The main examples are parental labor income, subject labor income, public-transfer income, and health-related outcome variables. ABC/CARE provided full-day childcare. It relaxed the time constraints of the mothers of the treated children, who were able to work more. Potentially, it might shift the life-cycle

[72] We only account for IPW for the list of variables listed here, or any calculation involving them.

profiles of the mothers, either by allowing them to take more schooling or obtain more work experience. We estimate these profiles. Similarly, a comprehensive evaluation accounts for the effect on the life-cycle profiles of subject labor income and public-transfer income. We follow the strategy outlined in Section 3 and implemented in Section 5 to identify and estimate these profiles and forecast post-a^* outcomes.[73].

This appendix documents how we implement this strategy and provides complementary evidence supporting it, cited throughout the main text. We proceed as follows. Appendix C.3.2 describes the auxiliary datasets that we use to forecast out of sample. Appendix C.3.3 provides details on the matching strategy used to construct the synthetic or virtual treatment and controls groups. Appendix C.3.4 documents the variables that we use to make forecasts. The next three appendices provide tests for the key assumptions listed in Section 3: Appendix C.3.5 provides tests of Assumption A–2, Appendix C.3.6 provides tests for Assumption A–3, and Appendix C.3.7 provides tests for Assumption A–4.[74] As in the main text, our forecasting strategy treats a subject's transfer and labor income jointly—although we provide some separate details when necessary. For parental labor income we provide three different forecasting strategies and document them in Appendix C.3.8.

We devote separate appendices to the discussion of crime and health forecasts in Appendix E and Appendix F, respectively.

C.3.1 Overview of Our Approach and A Summary of Findings

As noted in Section 3.2, the principal empirical strategy used in this paper is developed and implemented in two stages. In Stage I, we find comparison groups that are "comparable." This can be thought of as a type of coarse matching on $B \in \mathcal{B}_0$. Stage II uses the constructed

[73] 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).

[74] We provide tests for Assumption A–1 in Section 3.3.1.

samples to build models to make out of (experimental) sample forecasts, assuming structural invariance (Assumption A–4).

We rely primarily on a parametric model based approach that estimates Equation (12) in the main text in the experimental and comparison samples and extrapolates using the comparison samples. As noted in the text, under exogeneity and structural invariance, which we test for and do not reject, we can combine the two stages in a single stage matching procedure that finds counterparts to the experimental samples and controls in our auxiliary samples. We report agreement between the main estimates reported in the paper and the matching estimates. We find a surprising robustness across approaches in our estimates that inspires confidence in the benefit/cost estimates reported in the text.

C.3.2 Auxiliary Datasets

We rely on the following datasets to estimate life-cycle transfer and labor income profiles.[75] 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,

[75] 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.

1978. When appropriately weighted, the NLSY79 is nationally representative of the youth 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.[76] 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.[77]

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

[76] 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.

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

the NLSY79 satisfying the eligibility conditions $\boldsymbol{B} \in \mathcal{B}_0$ at every age.

The Children of the National Longitudinal Survey of Youth (CNLSY) is a survey 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.3.5) 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 treatment and controls groups 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; 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 Treatment and Control Groups

The first principle for constructing non-experimental synthetic treatment or control groups is that potential counterparts be eligible for the program, i.e., $\boldsymbol{B}_n \in \mathcal{B}_0$. To implement this condition, we require that the $\boldsymbol{B}$ variables be present in both experimental and control samples. An additional, harder challenge is to use these variables to find counterparts to the treated and the controlled in the non-experimental samples where no one directly receives treatment.

Our principal approach constructs counterpart experimental and control groups matching on $\mathcal{B}_n \in B_0$. However, under exogeneity and structural invariance, we also construct individual matches for members of the experimental group and use them to construct out-of-sample treatment effects. Notice that these are two different matching procedures. We summarize both in this subsection in a succinct notation, but the two should be carefully distinguished by the reader. Since the second set of matches is more inclusive and traditional, we develop that case here.

Consider outcome $Y^d_{k,j,a}$ generated by $\boldsymbol{B}_k, \boldsymbol{X}^d_{k,j,a}$ (defined as the relevant predictor variables for $Y_{k,j,a}$) and also generated by $\varepsilon^d_{k,j,a}$. Under exogeneity (Assumption A–3), we can match on both $\boldsymbol{B}_k$ and $\boldsymbol{X}^d_{k,j,a}$. This assumption allows us to use variables $\boldsymbol{X}^d_{k,j,a}$ that are caused by treatment and that forecast outcome j and are invariant after a^* (e.g., ability). If exogeneity is not satisfied, matching on $\boldsymbol{X}^d_{k,j,a}$ becomes problematic (see Heckman and Navarro, 2004). Note that we can match on $\boldsymbol{B}_k$ in the set of individuals in the auxiliary samples with $\boldsymbol{B}_k \in \mathcal{B}_0$. Note that $\boldsymbol{B}_k$ need not be exogenous. Our analysis is conditional on $\boldsymbol{B}_k \in \mathcal{B}, k \in \{e, n\}$. Below, we show that our estimates are not sensitive to whether we match on $\boldsymbol{B}_k$, match on $\boldsymbol{X}^d_{k,j,a}$ alone, or match on both $\boldsymbol{B}_k$ and $\boldsymbol{X}^d_{k,j,a}$.

For the moment, simplify notation and assume that $\boldsymbol{B}_k$ is absorbed into $\boldsymbol{X}^d_{k,j,a}$. We relax this assumption below and stress that in the main analysis reported in this paper, we only match on $\boldsymbol{B}$. For each treatment group member i in treatment status d, we find counterparts following the analysis of Heckman et al. (1998). We do not construct different synthetic treatment and control groups for each age and for each outcome that we forecast. We find one synthetic treatment and one synthetic control group in each auxiliary sample and use these samples to forecast each outcome at each age. We explain in Appendix C.3.2 what auxiliary samples are used to forecast outcomes at each age. That is, we use the same synthetic treatment and control groups to forecast all of the outcomes. For this reason, we drop the age and outcome subindices for the rest of this section.

Matched samples can be constructed in many ways using the various criteria listed at the end of Section C.3.2. One method that combines Stage I (sample construction) and Stage II (estimation) is the following. This provides a non-parametric approach to forecasting. It is conditional on a common value of $\boldsymbol{B}$ in both samples. Match an individual $l(i)$ in the auxiliary sample to person i in the treatment samples to find synthetic treatment and control groups by following Algorithm 1.

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*

$$\sqrt{(\boldsymbol{X}^d_{i,e} - \boldsymbol{X}^d_{i,l(i),n})'(\boldsymbol{\Sigma}^d_e)^{-1}(\boldsymbol{X}^d_{i,e} - \boldsymbol{X}^d_{i,l(i),n})} \leq \epsilon \tag{2}$$

where $\boldsymbol{X}^d_{i,l(i),n}$ represents the observed characteristics of the matched potential counterpart in the non-experimental sample for $d \in \{0,1\}$, where $\boldsymbol{\Sigma}^d_e$ is the covariance matrix in the experimental sample for fixed to treatment status d. We construct a synthetic control group ($d = 0$) and a synthetic treatment group ($d = 1$) by weighting the potential counterparts according to the inverse value of the left-hand-side of (2).[78]

We primarily use synthetic control and treatment groups matching solely on $\boldsymbol{B}$ to estimate dynamic relationships between outcomes and predictors in the auxiliary samples and use the estimated relationships to generate forecasts in the non-experimental samples. In this model-based approach, we construct treatment effects for each outcome at each age. In our main approach to estimation, we match on $\boldsymbol{B}_k$. We next examine the sensitivity of our estimates to the use of different matching variables. We find little sensitivity to choices of matching variables and to using parametric or non-parametric approaches. The lack of sensitivity suggests that the dynamic relationship that we fit in the auxiliary sample is invariant to the sample used to estimate it, which supports exogeneity Assumption A–4.

Summary: Variables Used to Match to Construct Synthetic Control and Treatment Groups

When constructing synthetic control and treatment groups we use each of the following variables, *for all the outcomes at all ages*: year of birth, gender, number of siblings at baseline.

[78] In practice, 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.

The criterion for selecting these variables is availability across all auxiliary sources. We construct one synthetic control group and one synthetic treatment group in each auxiliary sample (not one group for each age). We explain how we use the synthetic groups for forecasting each outcome in Appendix C.3.2.

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

We base our model-based forecasts for a generic outcome, $Y_{k,j,a}$ on background (pre-treatment) variables, $\boldsymbol{B}_k$, and variables that could have been affected by treatment $\boldsymbol{X}_{k,j,a}$. $\boldsymbol{X}_{k,j,a}$ can contain lagged values of $Y_{k,j,a}$.

Our forecasts are based on identifying and estimating relationships between the outcomes we seek to predict at age a, $Y_{k,j,a}$, in the experimental samples and using the fitted relationships in the auxiliary samples. We produce forecasts for each outcome at each age.

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.[79]; and (iii) it needs to satisfy Assumption A–2, which we document below in Appendix C.3.5.[80]

For a variable to satisfy restrictions (i) and (ii), it needs to be the case not only that the

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

[80] A more logical way to proceed would be to use a sub-sample of predictors satisfying this criteria and also being affected by treatment, so that we are able to predict treatment effects. This is impossible due to data limitations. Most of the post-treatment predictors we use, however, display sizable treatment effects. See Appendix C.9.

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 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 both the control and treatment synthetic groups we construct 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 $\boldsymbol{X}_a, \boldsymbol{B}_k$ are strong predictors of labor and transfer income at age 30. We present this evidence at age 30 both for brevity and compare the predictive power of $\boldsymbol{X}_a, \boldsymbol{B}_k$ on the outcomes we consider in the ABC/CARE sample.

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, Auxiliary Sources

	(1)	(2)	(3)	(4)	(5)	(6)
Group Matched	Control	Treatment	Control	Treatment	Control	Treatment
Auxiliary Sample	CNLSY		NLSY79		PSID	
Male	3,999.07***	3,911.59***	2,633.17***	2,616.23***	7,087.28***	7,096.48***
	(1,264.39)	(1,211.81)	(401.89)	(378.45)	(888.77)	(891.09)
Black	-3,366.02***	-3,175.47***	-1,236.27***	-1,236.89***	-1,900.33***	-1,926.12***
	(1,149.44)	(1,128.24)	(330.86)	(307.09)	(489.68)	(493.72)
PIAT (5-7)	83.87	79.94				
	(60.09)	(57.83)				
Education (30)	2,274.99***	2,224.96***	845.98***	833.57***	1,813.90***	1,824.82***
	(410.42)	(394.57)	(132.05)	(121.78)	(195.01)	(196.46)
Labor Income (21)	0.13**	0.13**	0.13**	0.13**	0.06*	0.06*
	(0.06)	(0.06)	(0.07)	(0.06)	(0.03)	(0.03)
Lagged Labor Income	0.53***	0.55***	0.87***	0.88***	0.72***	0.72***
	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)
Constant	-25,579.06***	-25,100.97***	-8,339.73***	-8,190.57***	-16,127.27***	-16,235.51***
	(6,453.86)	(6,262.12)	(1,354.43)	(1,249.59)	(1,968.18)	(1,982.72)
Observations	1,215	1,215	7,036	8,219	5,544	5,544
R^2	0.51	0.52	0.65	0.65	0.61	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 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, Auxiliary Sources

	(1)	(2)	(3)	(4)
Group Matched	Control	Treatment	Control	Treatment
Auxiliary Sample	NLSY79		PSID	
Male	-167.28***	-166.06***	-515.30***	-515.46***
	(36.72)	(35.08)	(155.70)	(155.49)
Black	116.95***	123.03***	136.30	135.77
	(44.04)	(42.49)	(143.46)	(143.13)
Education (30)	-44.70***	-43.84***	-103.94**	-103.93**
	(6.38)	(6.15)	(42.35)	(42.23)
Transfer Income (21)	0.11***	0.11***	0.11***	0.11***
	(0.03)	(0.03)	(0.03)	(0.03)
Lagged Transfer Income	0.63***	0.64***	0.64***	0.64***
	(0.03)	(0.03)	(0.04)	(0.04)
Constant	871.56***	855.82***	1,950.42***	1,950.78***
	(101.41)	(97.62)	(600.85)	(599.32)
Observations	7,127	8,306	1,539	1,539
R^2	0.48	0.48	0.58	0.57

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 using the procedure in Appendix C.3.3. ***: p-value < .01. **: p-value < .05. *: p-value < .10.

C.3.4.1 Non-Parametric Forecasts

An alternative to the forecasting strategy that we use throughout the main text and the preceding appendix, is the following. For each individual (i) in the experimental sample (e), we can: (i) find a match or a set of matches in the auxiliary sample (n) using Algorithm 1; and (ii) 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 strategy: instead of fitting a dynamic relationship in the non-experimental sample and using it to form out-of-sample forecasts, we simply match individuals using Algorithm 1 to impute labor income profiles.

Table C.5 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). An individual $l(i)$ in the auxiliary sample is a match for individual i in the experimental sample if it is in the neighborhood defined by the left-hand side of (2) across all individuals in the auxiliary sample. We present the discounted net present value (treatment - control) for labor income in 2014 dollars. The approaches are in close agreement. (Analogous results for other variables are available upon request from the authors.)

Table C.5: Net Present Value of Labor Income: Parametric and Non-Parametric Approaches

	Labor Income	
Pooled	Male	Female
a. Parametric (Main Paper)		
133,032	238,105	41,908
(76,634)	(185,375)	(24,606)
b. Non-Parametric (This Section)		
132,924	195,530	69,317
(11,253)	(20,210)	(4,350)

Note: this table compares the net-present value of labor income (treatment - control) using the parametric approach of the main text and the approach that we use in this section. All values are discounted to birth and reported in 2014 dollars.

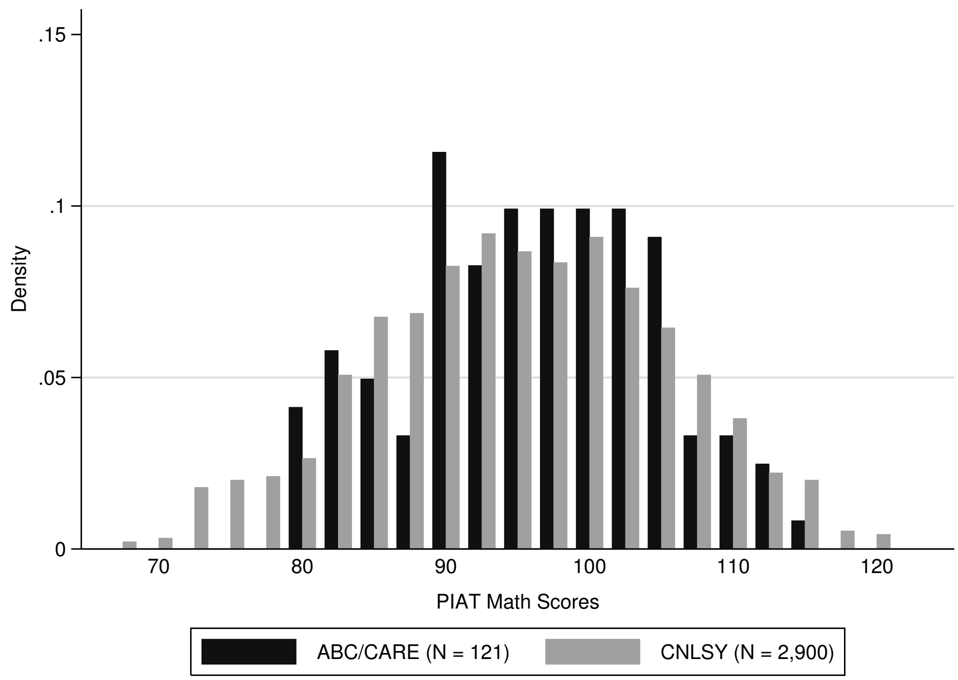
C.3.5 Testing Assumption A–2: Support Conditions

Assumption A–2 requires that the support of the auxiliary data contain the support of the experimental data. This can be checked for $a \leq a^*$. Figure C.1 validates this assumption 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.[81]

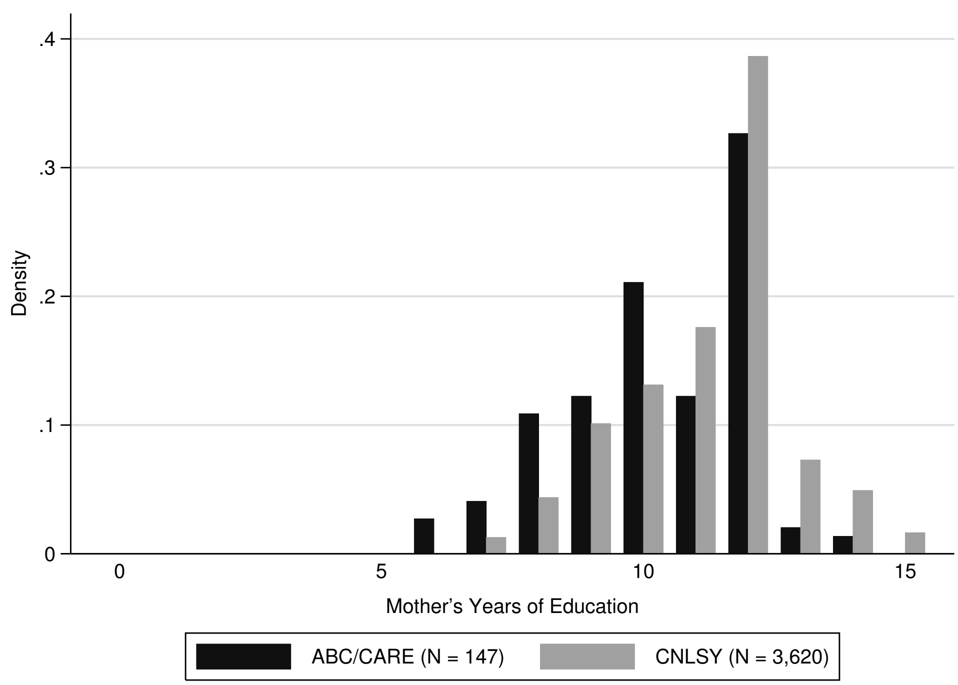
[81]For the male and Black indicators, we do not provide evidence of containing support. All three non-experimental samples have vast numbers of males, females, and Blacks to cover the support in the experimental samples.

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

(a) Average PIAT Math Scores, Ages 5–7

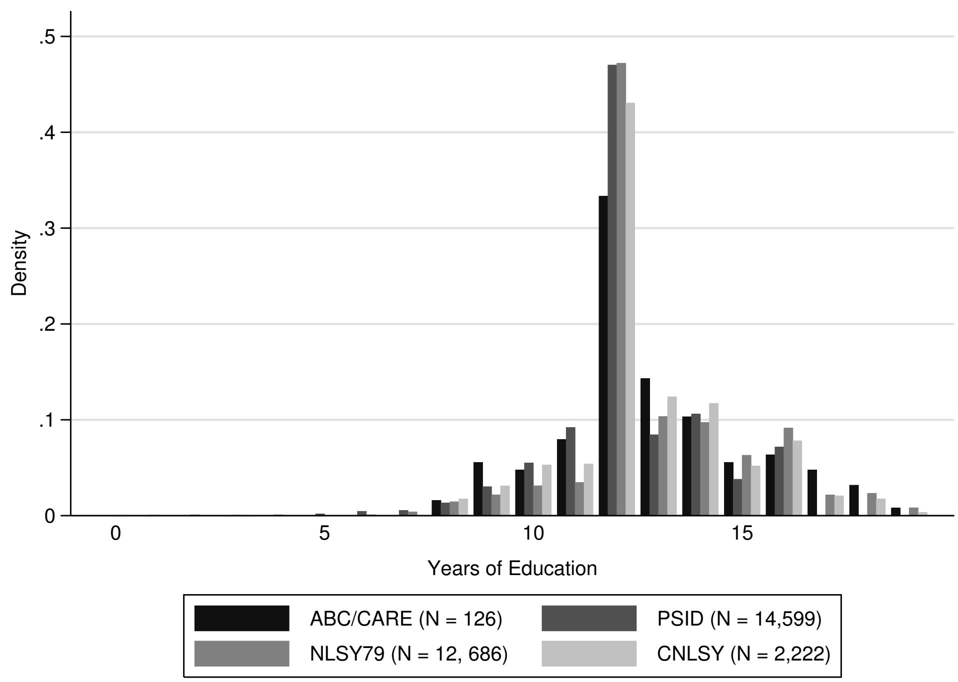


(b) Mother's Years of Education

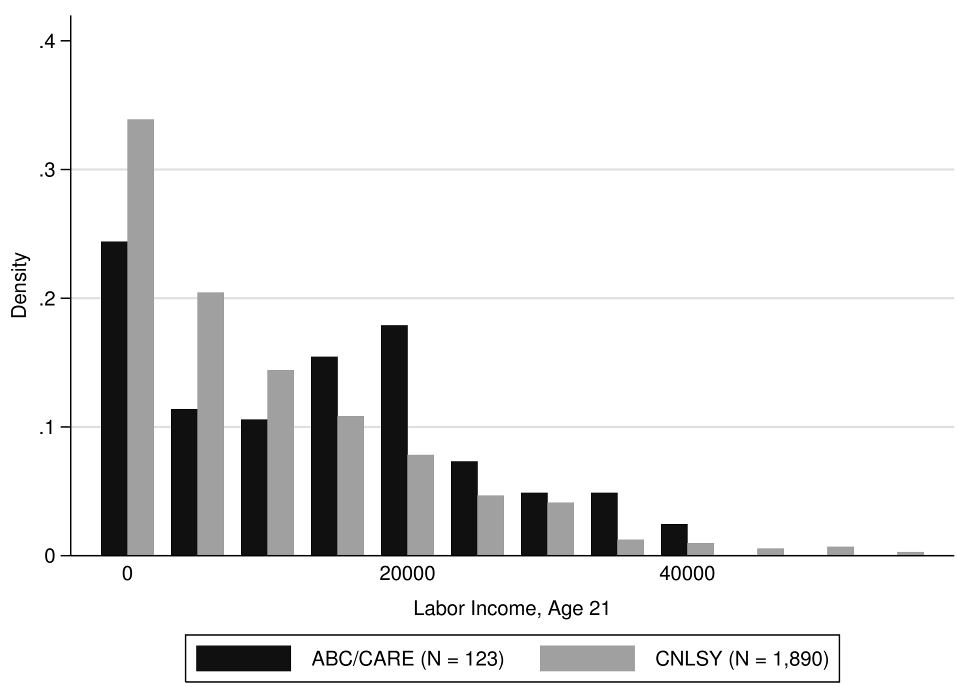


Note: These graphs display the support of ABC, PSID, NLSY79, and CNLSY for variables we use to forecast future labor income. PIAT math scores are averaged over ages 5–7.

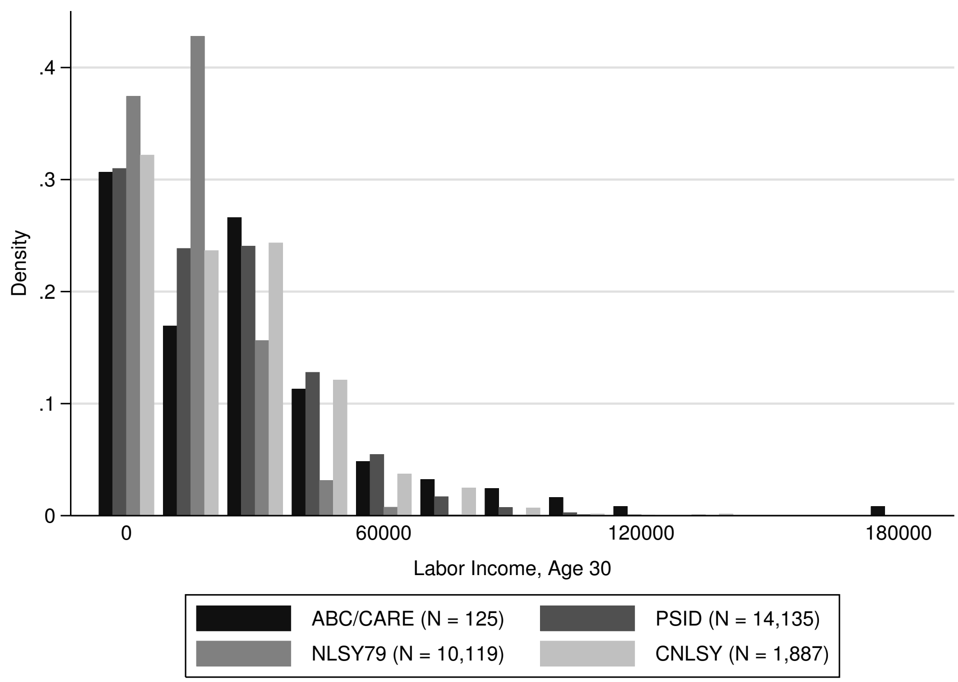
(c) Subject's Years of Education



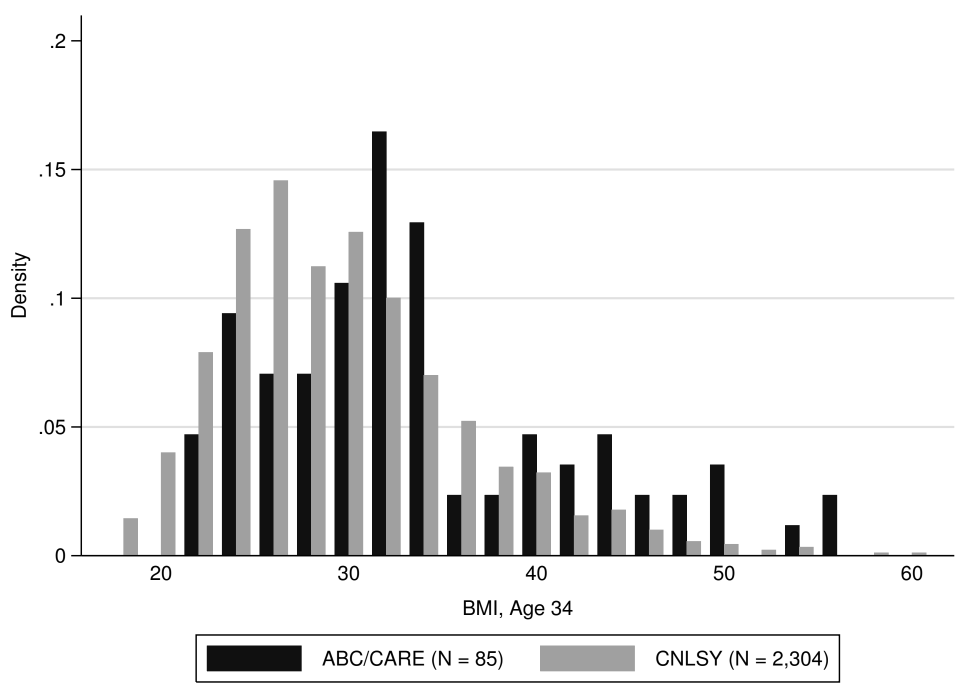
(d) Income at Age 21



(e) Income at Age 30



(f) Body Mass Index, Age 34



C.3.6 Testing Assumption A–3: Exogeneity

The following framework helps us to test both Assumptions A–3 and A–4. We discuss this framework and test Assumption A–3 in this section of the appendix. We test Assumption A–4 in the next section.

Define an outcome vector for treatment status d, in sample k at age a as

$$\boldsymbol{Y}_{k,a}^{d} = \boldsymbol{X}_{k,a}^{d}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_{a}^{d} \qquad (a)$$

with an associated measurement system

$$\begin{aligned} \boldsymbol{\varepsilon}_{a}^{d} &= \boldsymbol{\beta}^{d}\boldsymbol{\theta}_{a}^{d} + \boldsymbol{\omega}_{a}^{d} && (b) \\ \boldsymbol{M}_{a}^{d} &= \boldsymbol{\lambda}^{d}\boldsymbol{\theta}_{a}^{d} + \boldsymbol{\upsilon}_{a}^{d}, && (c) \end{aligned} \qquad (3)$$

where $\boldsymbol{\theta}^{d} \perp\!\!\!\perp \boldsymbol{\upsilon}_{a}^{d}, \boldsymbol{\omega}_{a}^{d}$ and $\boldsymbol{\upsilon}_{a}^{d} \perp\!\!\!\perp \boldsymbol{\omega}_{a}^{d}$. We use additional conditioning variables in these equations. To simplify the notation, we keep them implicit.

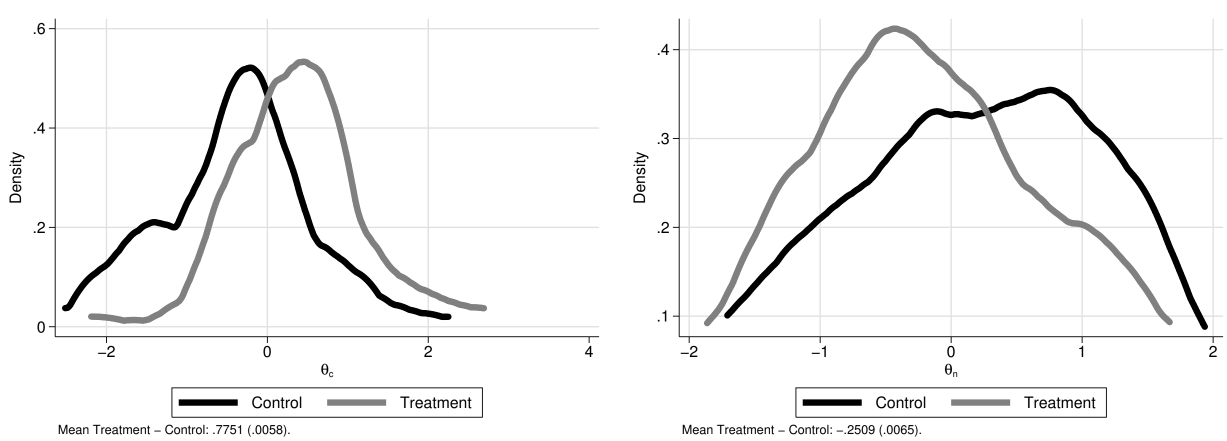
When the auxiliary measurement system $\boldsymbol{M}_{a}^{d}$ consists of at least three measures, we are able to identify the vectors of coefficients characterizing this system, $\boldsymbol{\lambda}^{d}, \boldsymbol{\beta}^{d}$, as well as the respective covariance matrices, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_{a}^{d}}, \boldsymbol{\Sigma}_{\boldsymbol{\upsilon}_{a}^{d}}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_{a}^{d}}$, we use the method from Bartlett (1938) to obtain an estimate of $\boldsymbol{\theta}_{a}^{d}$ (Heckman et al., 2013). Identifying and estimating the elements in system (3) serves two purposes: (i) it facilitates a test of Assumption A–3; and (ii) it enables us to use estimates of $\boldsymbol{\theta}_{a}^{d}$ as control functions when testing Assumption A–4 in the next appendix, i.e. to use these estimates to control for endogeneity.

We first describe the estimates for the elements in System (3) 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 $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_a^d}, \boldsymbol{\Sigma}_{\boldsymbol{\upsilon}_a^d}$ are the same whether $d = 0$ or $d = 1$.

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 $\boldsymbol{\theta}_{nc}^d$.[82] Figure C.2 shows our estimates by treatment status.

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

(a) Cognitive (b) Non-cognitive



Note: Panel (a) displays a factor score estimated based on the measurement system in (3) 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.

[82]For definitions and treatment effects on these variables see Appendix C.9.

We can also estimate $\boldsymbol{\theta}_a^d$ in the auxiliary sample. 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 auxiliary samples. The test consists of the following. Let $\boldsymbol{\gamma}^E$ be the parameter associated to $\boldsymbol{X}_{k,a}^d$ in Equation (a) in System (3) 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 (a) in System (3) when accounting for $\boldsymbol{\theta}_a^d$. Under the null hypotheses, Assumption A–3 holds and $\boldsymbol{\theta}_a^d$ is an irrelevant predictor in Equation (a) in System (3). This makes the OLS estimate of $\boldsymbol{\gamma}^E$ inconsistent. If the null hypotheses are false, $\boldsymbol{X}_{k,a}^d$ and ε_a^d are not independent, $\boldsymbol{\gamma}^I$ is consistent and $\boldsymbol{\gamma}^E$ is not. We test the null hypothesis by asking if the elements in $\boldsymbol{\theta}_a^d$ are relevant predictors of a set of outcomes at age 30, so that we can perform the tests on both the experimental and the auxiliary samples. 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.10 to C.13 present the results. In most cases, we are not able to reject the null hypothesis that Assumption A–3 holds.

Table C.6: Forecast of Labor Income at Age 30 Accounting for $\boldsymbol{B}_k$ and $\boldsymbol{\theta}, \boldsymbol{X}_{k,a}$, ABC/CARE Control Group

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother's Education	1,599.57	0.17	867.41	0.34	-769.20	0.68	-580.88	0.62
PIAT (5-7)	.	.	.	.	45.98	0.41	423.44	0.20
Education (30)	.	.	.	.	3,415.53	0.03	4,505.94	0.04
Labor Income (21)	.	.	.	.	0.69	0.02	0.97	0.03
Cognitive	.	.	758.28	0.43	.	.	-8,009.28	0.93
Non Cognitive	.	.	-342.62	0.52	.	.	7,275.49	0.09
Constant	10,239.82	0.28	16,530.50	0.22	-23,140.28	0.80	-80,679.09	0.96
F-stat	2.27		1.80		11.89		7.91	
p-value	0.42		0.41		0.42		0.01	
R^2	0.03		0.07		0.30		0.40	
Observations	66		51		65		63	
F-stat: exclude Cognitive, Non-Cognitive			1.70				4.14	
p-value			0.45				0.09	

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 (3) 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 (3) 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.7: Forecast of Labor Income at Age 30 Accounting for $\boldsymbol{B}_k$ and $\boldsymbol{\theta}$, $\boldsymbol{X}_{k,a}$, ABC/CARE Treatment Group

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother's Education	3,134.16	0.23	2,600.34	0.35	2,913.44	0.28	5,835.67	0.22
PIAT (5-7)	.	.	.	.	-263.29	0.66	-871.06	0.76
Education (30)	.	.	.	.	11,600.24	0.00	13,069.48	0.00
Labor Income (21)	.	.	.	.	-0.18	0.64	-0.62	0.75
Cognitive	.	.	2,766.35	0.40	.	.	4,828.93	0.34
Non Cognitive	.	.	7,600.33	0.18	.	.	6,223.32	0.19
Constant	3,900.73	0.47	10,553.93	0.42	-122,709.85	0.91	-109,410.81	0.76
F-stat	1.72		2.45		4.59		4.95	
p-value	0.38		0.21		0.38		0.06	
R^2	0.02		0.10		0.26		0.33	
Observations	64		49		65		63	
F-stat: exclude Cognitive, Non-Cognitive	2.49				2.03			
p-value	0.21				0.31			

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 (3) 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 (3) 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.8: Forecast of Labor Income at Age 30 Accounting for $\boldsymbol{B}_k$ 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 (3) 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 (3) 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.9: Forecast of Labor Income at Age 30 Accounting for $\boldsymbol{B}_k$ 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'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 (3) 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 (3) 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.10: Forecast of Transfer Income at Age 30 Accounting for $\boldsymbol{B}_k$ and $\boldsymbol{\theta}, \boldsymbol{X}_{k,a}$, ABC/CARE Control Group

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother's Education	-413.76	0.78	-406.14	0.69	48.75	0.49	51.97	0.47
PIAT (5-7)	.	.	.	.	27.10	0.29	-101.93	0.77
Education (30)	.	.	.	.	-684.75	0.99	-693.44	0.91
Labor Income (21)	.	.	.	.	-0.14	0.99	-0.15	0.93
Cognitive	.	.	-348.53	0.69	.	.	1,696.96	0.13
Non Cognitive	.	.	1,622.92	0.05	.	.	887.17	0.19
Constant	6,664.39	0.11	6,614.39	0.18	9,942.56	0.18	22,736.59	0.10
F-stat	1.93		2.96		3.53		2.68	
p-value	0.34		0.15		0.21		0.27	
R^2	0.04		0.15		0.21		0.27	
Observations	68		52		70		70	
F-stat: exclude Cognitive, Non-Cognitive	3.38				2.42			
p-value	0.19				0.27			

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 (3) 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 (3) 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.11: Forecast of Transfer Income at Age 30 Accounting for $\boldsymbol{B}_k$ and $\boldsymbol{\theta}, \boldsymbol{X}_{k,a}$, ABC/CARE Treatment Group

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
Mother'sEducation	-212.39	0.75	-336.44	0.81	-199.39	0.74	-302.60	0.79
PIAT(5-7)	.	.	.	.	-46.36	0.86	-22.41	0.65
Education(30)	.	.	.	.	-35.62	0.56	-72.66	0.59
LaborIncome(21)	.	.	.	.	-0.05	0.94	-0.05	0.90
Cognitive	.	.	-421.59	0.75	.	.	-273.48	0.62
Non Cognitive	.	.	-825.26	0.95	.	.	-987.11	0.98
Constant	3,348.22	0.16	4,937.75	0.14	9,041.98	0.09	8,432.47	0.18
F-stat	1.23		2.59		1.81		2.27	
p-value	0.45		0.18		0.45		0.25	
R^2	0.03		0.15		0.13		0.25	
Observations	63		49		65		63	
F-stat: exclude Cognitive, Non-Cognitive	3.08				2.79			
p-value	0.20				0.18			

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 (3) 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 (3) 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.12: Forecast of Transfer Income at Age 30 Accounting for $\boldsymbol{B}_k$ 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	-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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) 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 (3) 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.13: Forecast of Transfer Income at Age 30 Accounting for $\boldsymbol{B}_k$ 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 (3) 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 (3) 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.3.7 Testing Assumption A–4: Structural Invariance

This appendix uses the framework presented in Appendix C.3.6 to test Assumption A–4. The tests are made under the null hypothesis that Assumption A–3 holds. In Section 3.3.5 we show that Assumption A–4, together with Assumption A–3, implies:

$$\mathbb{E}\left[Y^{d}_{k,j,a}|\boldsymbol{X}^{d}_{k,a}=\boldsymbol{x},\boldsymbol{B}_{k}=\boldsymbol{b},D=d\right]=\mathbb{E}\left[Y_{k,j,a}|\boldsymbol{X}^{d}_{k,a}=\boldsymbol{x},\boldsymbol{B}_{k}=\boldsymbol{b}\right], \tag{4}$$

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

A direct test of this hypothesis uses the experimental sample and asks if, once we account for a set of the variables in $\boldsymbol{X}_{k,a}$, R (randomization to treatment assignment in ABC/CARE, which, as discussed in text, is equivalent to D) predicts the outcome of interest, conditional on $\boldsymbol{B}_k$. This test will check whether this specific set of $\boldsymbol{X}_{k,a}$ suffices to summarize the treatment generated by R. Under the null hypothesis, the coefficient associated with R when forecasting based on $\boldsymbol{B}_k$ and $\boldsymbol{X}_{k,a}$ is zero. We test this using the following predictors. Average Mathematics PIAT scores at ages 5 to 7, education at age 30, and labor income at age 21. That is, children who are offered treatment attend it. For a set of outcomes of interest, even beyond those that we forecast, once we condition on $\boldsymbol{B}_k, \boldsymbol{X}_{k,a}, \boldsymbol{\theta}$, the coefficient associated with R is not significant. We test this with and without accounting for endogeneity, as explained in Appendix C.3.6.

Table C.14: Forecast of High School Graduation at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Pooled Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	0.130	0.040	0.106	0.170	0.019	0.415	0.019	0.425
Mother's Education	0.093	0.000	0.081	0.000	0.051	0.000	0.043	0.050
PIAT (5-7)	.	.	.	.	-0.005	0.910	-0.004	0.765
Education (30)	.	.	.	.	0.119	0.000	0.124	0.000
Labor Income (21)	.	.	.	.	0.000	0.005	0.000	0.010
Cognitive	.	.	0.020	0.360	.	.	-0.038	0.740
Non Cognitive	.	.	-0.027	0.720	.	.	0.011	0.395
Constant	-0.410	0.985	-0.283	0.865	-1.082	1.000	-1.148	0.985
F-stat	14.497		5.819		40.509		25.147	
R^2	0.151		0.143		0.440		0.434	
Observations	134		102		135		133	

Note: Forecast of high school graduation at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.15: Forecast of High School Graduation at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Female Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	0.196	0.010	0.091	0.250	0.031	0.385	0.037	0.425
Mother's Education	0.077	0.000	0.062	0.050	0.036	0.055	0.023	0.170
PIAT (5-7)	.	.	.	.	-0.008	0.860	-0.004	0.655
Education (30)	.	.	.	.	0.093	0.000	0.102	0.005
Labor Income (21)	.	.	.	.	0.000	0.000	0.000	0.000
Cognitive	.	.	0.051	0.285	.	.	-0.073	0.775
Non Cognitive	.	.	-0.076	0.895	.	.	0.051	0.200
Constant	-0.266	0.800	-0.065	0.540	-0.444	0.790	-0.872	0.870
F-stat	10.180		5.545		35.887		31.753	
R^2	0.172		0.197		0.556		0.612	
Observations	68		53		70		70	

Note: Forecast of high school graduation at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.16: Forecast of High School Graduation at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Male Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	0.084	0.220	0.146	0.140	0.088	0.190	0.067	0.305
Mother's Education	0.116	0.000	0.131	0.000	0.057	0.045	0.072	0.095
PIAT (5-7)	.	.	.	.	0.000	0.470	-0.000	0.500
Education (30)	.	.	.	.	0.152	0.000	0.156	0.000
Labor Income (21)	.	.	.	.	0.000	0.170	0.000	0.440
Cognitive	.	.	-0.023	0.645	.	.	-0.009	0.525
Non Cognitive	.	.	0.051	0.195	.	.	0.004	0.475
Constant	-0.636	0.970	-0.824	0.950	-2.092	1.000	-2.227	0.955
F-stat	11.144		6.555		17.009		10.294	
R^2	0.190		0.215		0.467		0.460	
Observations	67		49		65		70	

Note: Forecast of high school graduation at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.17: Forecast of Employment at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Pooled Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	0.123	0.040	0.121	0.115	-0.008	0.550	0.057	0.265
Mother's Education	0.033	0.060	0.017	0.230	0.031	0.150	0.029	0.155
PIAT (5-7)	.	.	.	.	0.008	0.020	0.012	0.060
Education (30)	.	.	.	.	0.046	0.005	0.026	0.080
Labor Income (21)	.	.	.	.	-0.000	0.850	-0.000	0.875
Cognitive	.	.	0.077	0.060	.	.	-0.016	0.595
Non Cognitive	.	.	0.034	0.285	.	.	0.060	0.170
Constant	0.361	0.075	0.530	0.020	-0.877	0.975	-0.966	0.895
F-stat	3.903		4.073		5.239		3.979	
R^2	0.057		0.124		0.177		0.229	
Observations	133		101		135		133	

Note: Forecast of employment at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.18: Forecast of Employment at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Female Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	0.124	0.135	0.003	0.495	-0.078	0.705	-0.092	0.745
Mother's Education	-0.000	0.520	-0.000	0.500	-0.012	0.670	-0.000	0.500
PIAT (5-7)	.	.	.	.	0.010	0.030	0.008	0.145
Education (30)	.	.	.	.	0.040	0.035	0.030	0.085
Labor Income (21)	.	.	.	.	0.000	0.260	-0.000	0.520
Cognitive	.	.	0.151	0.005	.	.	0.065	0.240
Non Cognitive	.	.	-0.027	0.655	.	.	0.019	0.425
Constant	0.702	0.000	0.754	0.000	-0.624	0.865	-0.359	0.655
F-stat	1.873		5.089		3.432		3.918	
R^2	0.048		0.207		0.229		0.289	
Observations	67		52		65		70	

Note: Forecast of employment at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.19: Forecast of Employment at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Male Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	0.132	0.065	0.228	0.035	0.083	0.220	0.271	0.005
Mother's Education	0.066	0.020	0.045	0.150	0.094	0.020	0.116	0.000
PIAT (5-7)	.	.	.	.	0.008	0.090	0.022	0.015
Education (30)	.	.	.	.	0.023	0.235	-0.009	0.635
LaborIncome (21)	.	.	.	.	-0.000	0.990	-0.000	0.940
Cognitive	.	.	-0.030	0.665	.	.	-0.180	0.970
Non Cognitive	.	.	0.110	0.020	.	.	0.138	0.030
Constant	-0.002	0.500	0.203	0.350	-1.202	0.940	-2.416	0.975
F-stat	4.050		3.140		3.899		5.322	
R^2	0.114		0.192		0.240		0.443	
Observations	66		49		65		63	

Note: Forecast of employment at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.20: Forecast of Labor Income at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Pooled Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	10576.303	0.065	11,165.829	0.125	283.356	0.490	1,836.270	0.410
Mother's Education	1,851.130	0.205	1,131.843	0.375	496.581	0.430	1,052.668	0.365
PIAT (5-7)					-81.009	0.595	-320.784	0.705
Education (30)					8,097.138	0.000	9,141.309	0.000
Labor Income (21)					0.130	0.330	0.192	0.325
Cognitive			2,308.860	0.305			785.891	0.465
Non Cognitive			2,665.092	0.190			6,876.181	0.065
Constant	7,067.552	0.405	14,188.359	0.340	-73,300.00	0.965	-70,500.00	0.920
F-stat	1.965		1.522		5.746		4.742	
R^2	0.031		0.056		0.210		0.251	
Observations	132.000		101.000		130.000		133.000	

Note: Forecast of employment at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.21: Forecast of Labor Income at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Female Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	3,401.892	0.305	1,194.706	0.410	-6,899.006	0.915	-5,862.320	0.840
Mother's Education	-852.061	0.675	-1,688.467	0.835	-2,581.049	0.975	-2,473.902	0.965
PIAT (5-7)					260.764	0.165	347.907	0.170
Education (30)					3,580.642	0.000	3,916.084	0.005
LaborIncome (21)					0.329	0.175	0.392	0.160
Cognitive			3,828.286	0.130			-2,905.637	0.785
Non Cognitive			-1,663.392	0.655			2,051.882	0.300
Constant	32,117.510	0.055	39,943.031	0.025	-21,500.00	0.800	-36,600.00	0.840
F-stat	1.234		2.812		9.052		8.916	
R^2	0.039		0.143		0.354		0.393	
Observations	67		52		65		70	

Note: Forecast of labor income at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.22: Forecast of Labor Income at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Male Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	18169.158	0.075	21,891.223	0.150	15,649.704	0.115	18,835.850	0.185
Mother's Education	5,722.000	0.090	6,064.495	0.260	4,618.608	0.155	8,200.867	0.160
PIAT (5-7)					459.787	0.180	1,828.085	0.110
Education (30)					15,803.528	0.000	22,139.904	0.015
Labor Income (21)					0.107	0.410	0.193	0.365
Cognitive			-896.956	0.525			-13,700	0.815
Non Cognitive			10,273.761	0.105			7,533.493	0.175
Constant	-31,600.00	0.780	-34,800.00	0.630	-272,000.00	0.985	-526,000.00	0.965
F-stat	2.327		1.963		4.833		7.182	
R^2	0.068		0.128		0.343		0.465	
Observations	66		48		65		63	

Note: Forecast of labor income at age 30 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.23: Forecast of Body-Mass Index at Age 34 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Pooled Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	1.027	0.270	2.864	0.150	1.213	0.250	3.367	0.090
Mother's Education	-0.130	0.615	-0.116	0.560	0.003	0.500	-0.273	0.665
PIAT (5-7)					0.076	0.260	0.277	0.060
Education (30)					-0.116	0.575	-0.295	0.610
Labor Income (21)					0.000	0.290	0.000	0.095
Cognitive			-1.675	0.935			-3.431	0.960
Non Cognitive			1.615	0.195			2.392	0.100
Constant	34.913	0.000	33.909	0.000	26.682	0.070	9.604	0.330
F-stat	1.366		2.612		1.663		2.830	
R^2	0.027		0.110		0.090		0.209	
Observations	87		66		85		84	

Note: Forecast of body-mass index at age 34 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.24: Forecast of Body-Mass Index at Age 34 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Female Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	3.675	0.110	7.167	0.035	4.623	0.090	6.526	0.020
Mother's Education	-0.148	0.580	-0.654	0.820	-0.492	0.715	-0.909	0.835
PIAT (5-7)					-0.119	0.775	0.040	0.440
Education (30)					0.238	0.445	0.269	0.435
Labor Income (21)					0.000	0.340	0.000	0.385
Cognitive			-2.171	0.925			-2.366	0.815
Non Cognitive			2.285	0.155			2.536	0.145
Constant	36.244	0.000	39.310	0.000	46.750	0.020	33.957	0.075
F-stat	1.837		3.151		2.442		3.206	
R^2	0.065		0.206		0.191		0.285	
Observations	51		41		50		49	

Note: Forecast of body-mass index at age 34 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.25: Forecast of Body-Mass Index at Age 34 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Male Sample, ABC/CARE

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
R	-0.189	0.510	1.262	0.370	-0.397	0.545	1.150	0.400
Mother's Education	-0.513	0.805	0.448	0.380	-0.091	0.510	1.074	0.215
PIAT (5-7)					0.224	0.050	0.651	0.075
Education (30)					0.445	0.250	1.482	0.220
Labor Income (21)					0.000	0.165	0.000	0.100
Cognitive			-1.677	0.800			-4.854	0.920
Non Cognitive			0.119	0.475			0.563	0.380
Constant	36.443	0.000	26.285	0.030	3.330	0.455	-64.561	0.885
F-stat	1.835		2.330		5.387		31.866	
R^2	0.076		0.180		0.230		0.504	
Observations	37		25		35		35	

Note: Forecast of body-mass index at age 34 is based on the variables 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. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior). 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.

In Section 3.3.5 we show that Assumption A–4, together with Assumption A–3, implies equality of conditional expectations in the experimental and auxiliary samples.

$$\mathbb{E}\left[Y_{e,j,a}|\boldsymbol{X}^d_{e,a}=\boldsymbol{x},\boldsymbol{B}_e=\boldsymbol{b}\right]=\mathbb{E}\left[Y_{n,j,a}|\boldsymbol{X}^d_{n,a}=\boldsymbol{x},\boldsymbol{B}_e=\boldsymbol{b}\right], \quad d\in\{0,1\}, \quad j\in\mathcal{J}_a. \tag{5}$$

We test this hypothesis at $a = a^*$, where $a^* = 30$. Our non-experimental data source at $a = a^*$ is the CNLSY. The test is analogous to the one that we perform before. Under the null hypothesis, an indicator of sample membership (experimental or auxiliary) is statistically equal to zero. Results are reported in Tables C.26 to C.29. Using the full set of conditioning variables in $\boldsymbol{B}_k, \boldsymbol{X}_{k,a}$, we do not reject the null of equality in Equation (5).

Table C.26: Forecast of Labor Income at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Female Sample, ABC/CARE and CNLSY

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
K^*	4,396.848	0.060	2,292.980	0.250	456.565	0.410	539.899	0.445
Mother's Education	289.800	0.400	-1253.548	0.800	-1878.064	0.985	-2,126.096	0.960
PIAT (5-7)					207.361	0.090	221.599	0.215
Education (30)					3381.137	0.000	3652.225	0.000
Labor Income (21)					0.345	0.020	0.366	0.050
Cognitive			4,078.844	0.055			-,1479.220	0.670
Non Cognitive			-1370.089	0.640			2229.399	0.195
Constant	1,7358.422	0.100	33,633.047	0.030	-25,100.00	0.960	-27,400.00	0.840
F-stat	1.924		2.882		13.153		9.163	
R^2	0.022		0.106		0.279		0.312	
Observations	382		128		380		385	

* $K = 1$ if $k = e$; $K = 0$ if $k = n$.

Note: Forecast of labor income at age 30 is based on the variable listed in the row and uses the ABC/CARE and the CNLSY sample constructed according to the procedure in Appendix C.3.3. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) in CNLSY and IQ at ages 2, 3, 4, 5, 7, and 8 in ABC/CARE (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior) in CNLSY and measures of somatization, hostility, depression, and a global mental health index at age 21 in ABC/CARE (non-cognitive skill). 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.27: Forecast of Labor Income at Age 30 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Male Sample, ABC/CARE and CNLSY

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
K^*	17984.393	0.010	21406.389	0.015	4864.750	0.260	3301.140	0.305
Mother's Education	4,182.211	0.035	2,885.837	0.295	1,991.183	0.150	3,960.881	0.210
PIAT (5-7)					13.463	0.480	608.659	0.210
Education (30)					11,855.479	0.000	18,995.199	0.010
Labor Income (21)					0.289	0.165	0.243	0.260
Cognitive			5,012.976	0.205			-1498.498	0.560
Non Cognitive			6,902.538	0.115			6,335.481	0.070
Constant	-23,300.00	0.805	-1.13e+04	0.575	-1.50e+05	0.985	-318,000.00	0.965
F-stat	4.333		2.187		9.588		8.790	
R^2	0.059		0.087		0.283		0.403	
Observations	312		102		310		315	

* $K = 1$ if $k = e$; $K = 0$ if $k = n$.

Note: Note: Forecast of labor income at age 30 is based on the variable listed in the row and uses the ABC/CARE and the CNLSY sample constructed according to the procedure in Appendix C.3.3. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) in CNLSY and IQ at ages 2, 3, 4, 5, 7, and 8 in ABC/CARE (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior) in CNLSY and measures of somatization, hostility, depression, and a global mental health index at age 21 in ABC/CARE (non-cognitive skill). 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.28: Forecast of Body-Mass Index at Age 34 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Female Sample, ABC/CARE and CNLSY

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
K^*	4.380	0.000	3.620	0.015	4.538	0.020	3.731	0.115
Mother's Education	-0.110	0.585	-0.273	0.675	-0.225	0.705	-0.433	0.735
PIAT (5-7)					-0.006	0.530	0.076	0.285
Education (30)					0.001	0.500	0.337	0.420
Labor Income (21)					0.000	0.315	-0.000	0.525
Cognitive			-0.480	0.680			-0.773	0.705
Non Cognitive			0.858	0.255			0.805	0.275
Constant	32.921	0.000	34.948	0.000	34.288	0.000	25.174	0.085
F-stat	6.255		3.312		3.929		2.370	
R^2	0.075		0.110		0.122		0.167	
Observations	366		117		365		364	

* $K = 1$ if $k = e$; $K = 0$ if $k = n$.

Note: Forecast of body-mass index at age 34 is based on the variable listed in the row and uses the ABC/CARE and the CNLSY sample constructed according to the procedure in Appendix C.3.3. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) in CNLSY and IQ at ages 2, 3, 4, 5, 7, and 8 in ABC/CARE (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior) in CNLSY and measures of somatization, hostility, depression, and a global mental health index at age 21 in ABC/CARE (non-cognitive skill). 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.29: Forecast of Body-Mass Index at Age 34 Accounting for $R, \boldsymbol{B}_k, \boldsymbol{\theta}$, and $\boldsymbol{X}_{k,a}$ Male Sample, ABC/CARE and CNLSY

	(1) Estimate	(2) p-value	(3) Estimate	(4) p-value	(5) Estimate	(6) p-value	(7) Estimate	(8) p-value
K^*	2.327	0.025	2.812	0.065	1.822	0.100	3.135	0.060
Mother's Education	-0.180	0.700	0.347	0.320	-0.029	0.515	0.518	0.180
PIAT (5-7)					0.080	0.085	0.236	0.050
Education (30)					0.161	0.300	0.399	0.270
Labor Income (21)					0.000	0.080	0.000	0.020
Cognitive			-1.270	0.835			-2.362	0.970
Non Cognitive			0.188	0.385			0.482	0.265
Constant	30.350	0.000	25.170	0.000	18.323	0.010	-6.799	0.600
F-stat	2.828		3.562		3.171		3.867	
R^2	0.050		0.142		0.096		0.260	
Observations	283		79		285		280	

* $K = 1$ if $k = e$; $K = 0$ if $k = n$.

Note: Forecast of body-mass index at age 34 is based on the variable listed in the row and uses the ABC/CARE and the CNLSY sample constructed according to the procedure in Appendix C.3.3. Empty cells indicate that the variable was not used in the prediction. For each coefficient we provide the point estimate and p-value. $\hat{\boldsymbol{\theta}}_c$: factor score estimated based on the measurement system in (3) and PIAT sections that we do not use to forecast (reading and comprehension) as well as by the Peabody Picture Vocabulary Test (PPVT) in CNLSY and IQ at ages 2, 3, 4, 5, 7, and 8 in ABC/CARE (cognitive skill). $\hat{\boldsymbol{\theta}}_n$: factor score estimated based on the measurement system in (3) and on six scales of the Behavior Problems Index (e.g., anxiety, dependency, social behavior) in CNLSY and measures of somatization, hostility, depression, and a global mental health index at age 21 in ABC/CARE (non-cognitive skill). 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 fraction of the ABC/CARE benefits come from 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.[83] 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.

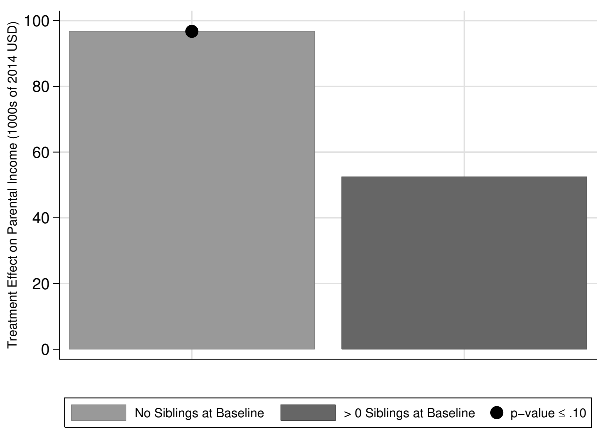
We note that the childcare feature of the program likely benefits parents who, at baseline, did not have any other children. If they did have other children, parents would still have to take

[83] See Appendix C.9.

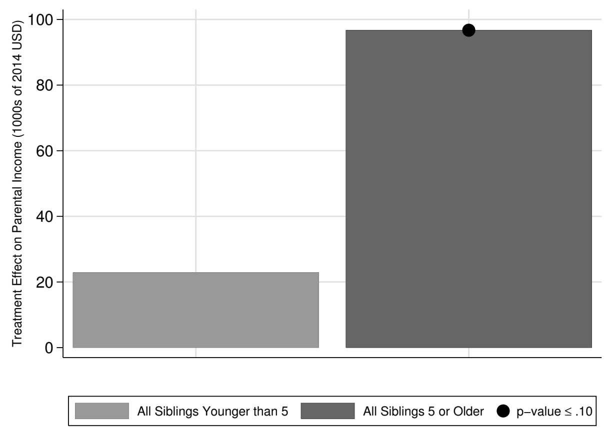
care of the other children, weakening the childcare-driven effect on labor income (especially if there are younger siblings present). Figure C.3 shows that the net increase (treatment minus control) in discounted parental labor income is much higher in the absence of siblings (of the participant children) 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.[84]

Figure C.3: 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 baseline "conservative" measure of parental labor income in Section 4.2. Results using our alternative parental labor income measures are similar (see Appendix C.3.8).

[84] 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.[85] We specify how we deal with the presence of a spouse below. The parameterization used is as follows:

$$\ln Y_a = \alpha + \beta \cdot \text{school}_a + \gamma_1 \cdot \text{experience}_a + \gamma_2 \cdot {\text{experience}_a}^2 + \boldsymbol{\psi}\mathbf{X}_a + \eta_a, \tag{6}$$

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–4, which we test and fail to reject in Appendix C.3.7.

We estimate the coefficients in (6) 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 (6) using a sub-sample of disadvantaged mothers in the PSID.[86] The coefficients characterizing (6) in ABC/CARE and PSID for different combinations of control sets are in close agreement. We display them in Table C.30.

[85] See Mincer (1974) for the original source and Heckman et al. (2006) for a extended discussion of the Mincer equation.

[86] 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.30: 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 (6) 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.30, 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 (6) 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.31 display results that answer these two questions. Precise estimates for (6) 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.31: 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 (6) (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 (6) (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 this for the PSID and NLSY79 separately before using them jointly 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

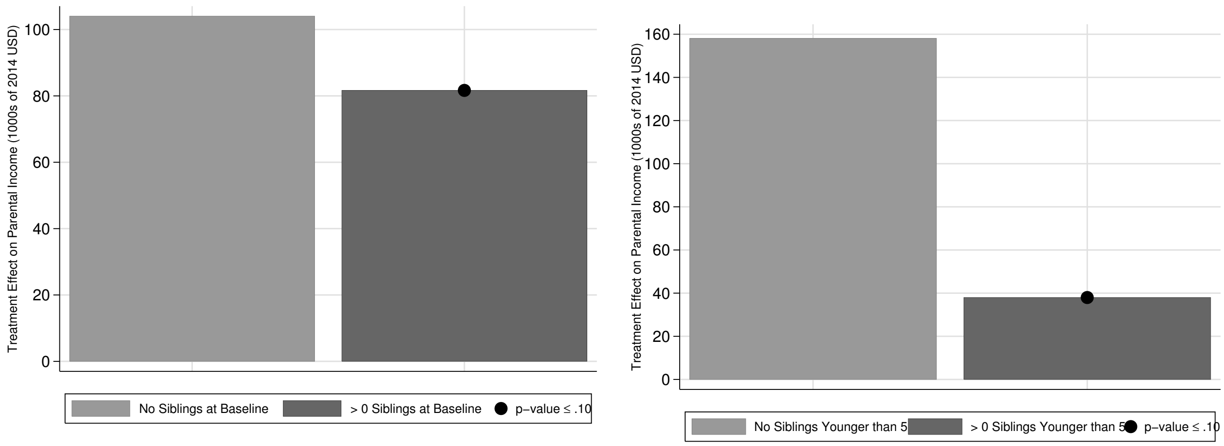
Any forecast of parental labor income starts at the last observation of parental labor income, which varies by individual due to sample attrition. We assume that parental labor income is equal to maternal income (only 27% of mothers in ABC/CARE report living in a two-parent home).

1. **Mincer Model**

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 4.2.

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

(b) **Initial Restrictions Placed on the Auxiliary Sample:** Black; female; unmarried; 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 Control and Treatment Groups:** 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 Control and Treatment Groups:** 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:

$$\sum_{a=0}^{A} \frac{\mathbb{E}(B_a - C_a)}{(1+\rho)^a} = 0, \tag{7}$$

where we let $A = 79$, 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.[87] 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.32 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 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

[87] This is an abuse of notation given that B_a and C_a are not discounted in Appendix C.4.

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.[88] 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.32 lists the factor we use to multiply each cost to account for its implied dead-weight loss.

Having constructed 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.[89] 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 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

[88] 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).

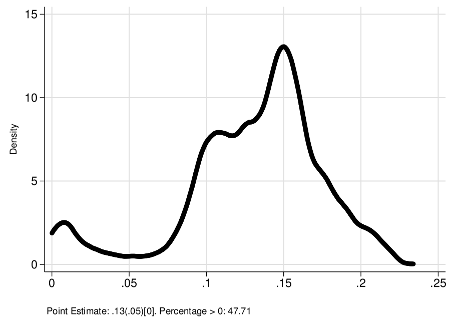
[89] See Arrow and Levhari (1969) for a formal discussion, although the discussion on multiplicity, sign, and real or complex nature of the roots of a polynomial traces back to Descartes' Rule.

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 10^{th} and 90^{th} percentiles of each bootstrap distribution.

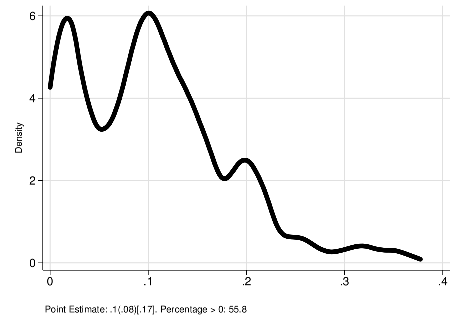
Figure C.5 reports the distributions of the internal rates of return, by gender and for each of the three parameters that we consider (treatment vs. next best, treatment staying at home, treatment vs. alternative preschool). For some parameters and genders, we discard a high percentage of the internal 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 and by Parameter

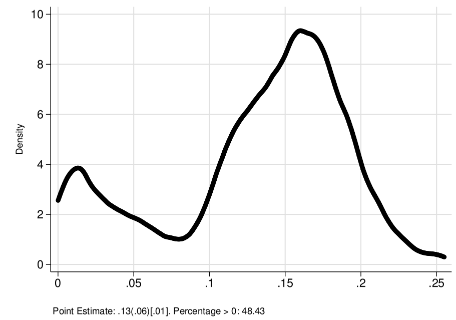
(a) Treatment vs. Next Best, Pooled



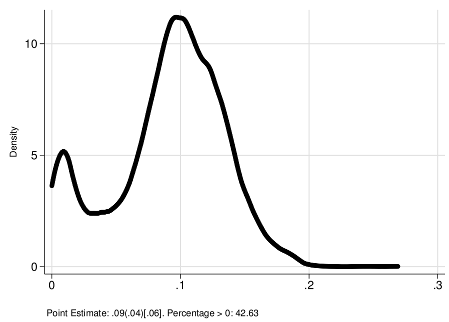
(b) Treatment vs. Next Best, Females



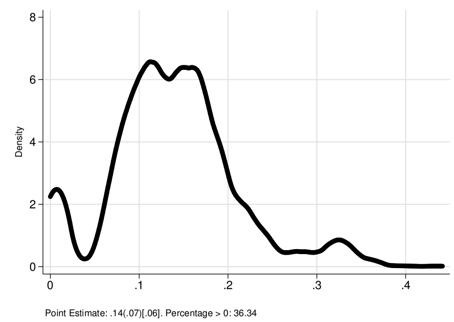
(c) Treatment vs. Next Best, Males



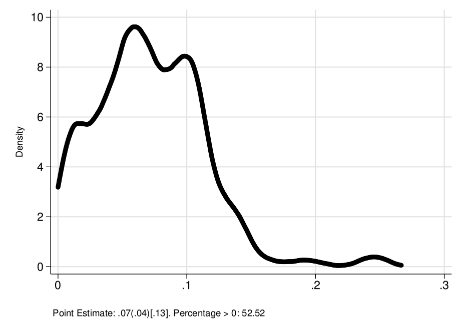
(d) Treatment vs. Staying at Home, Pooled



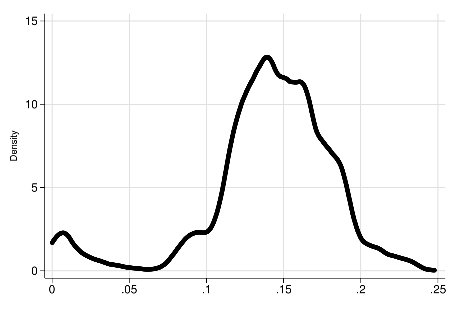
(e) Treatment vs. Staying at Home, Females



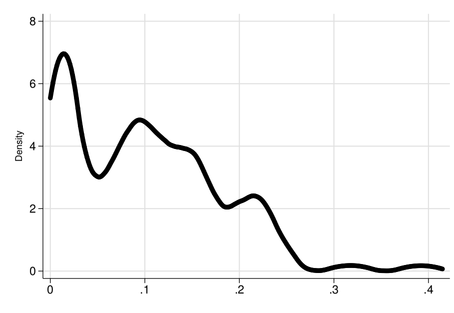
(f) Treatment vs. Staying at Home, Males



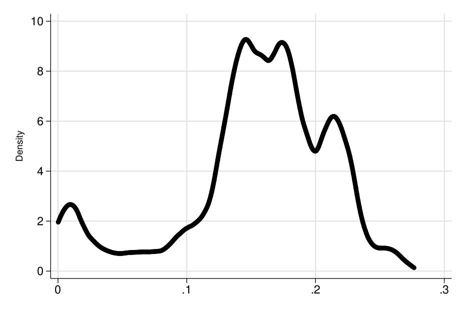
(g) Treatment vs. Alternative Preschool, Pooled



(h) Treatment vs. Alternative Preschool, Females



(i) Treatment vs. Alternative Preschool, Males



Note: Panel (a) displays the empirical bootstrap distribution for the estimate of the treatment vs. next parameter in the pooled sample. The remaining panels show an analogous distribution when varying the parameter and the gender. See Section 3.1 for the definition of the parameters. 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.

C.5 Computing the Benefit/Cost Ratio

The benefit/cost ratio is

$$\mathbb{E}\left(\frac{\sum_{a=0}^{A} B_a}{\sum_{a=0}^{A} C_a}\right), \tag{8}$$

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.32 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.

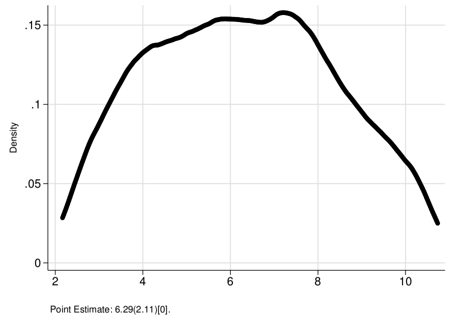
Table C.32: Components of Benefits and Costs

Variable	Sign Reversed	Welfare Cost Factor
Benefits		
Parent Income		
Subject QALY		
Subject Labor Income		
Subject Public-transfer Income	✓	0.5
Medicare Costs	✓	1.5
Medicaid Costs	✓	1.5
Out-of-pocket 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		

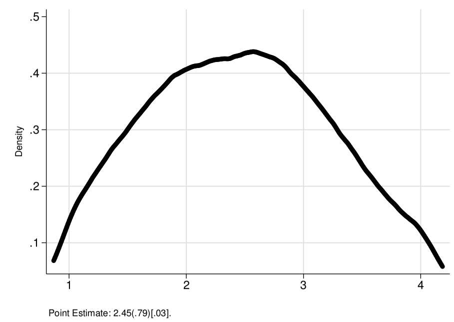
Note: The table lists the components of the costs and benefits of ABC/CARE. In order for some components to be categorized as benefits, we reversed the sign of the treatment effect. Only education costs up until age 18 are multiplied by 1.5 to account for welfare costs. This factor is drawn from Heckman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

Figure C.6: Benefit/Cost Ratios, by Gender and by Parameter

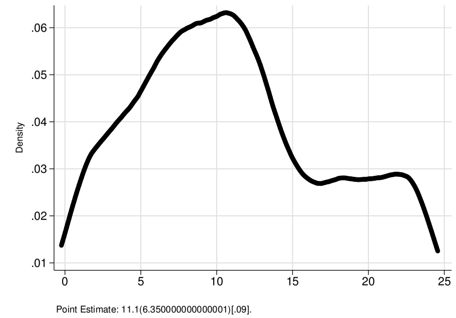
(a) Treatment vs. Next Best, Pooled



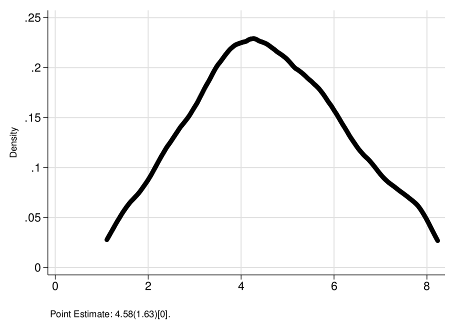
(b) Treatment vs. Next Best, Females



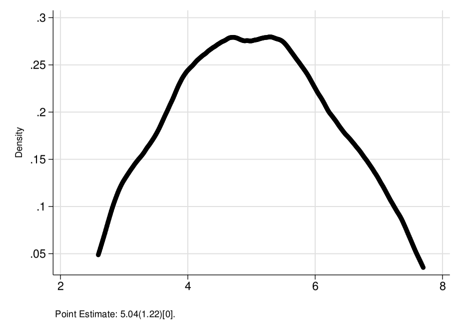
(c) Treatment vs. Next Best, Males



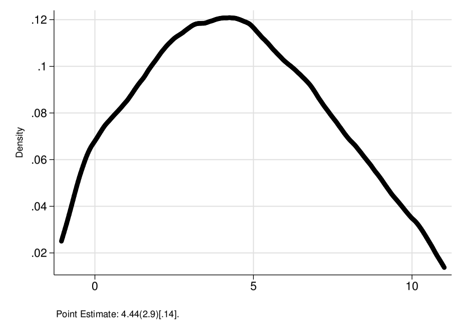
(d) Treatment vs. Staying at Home, Pooled



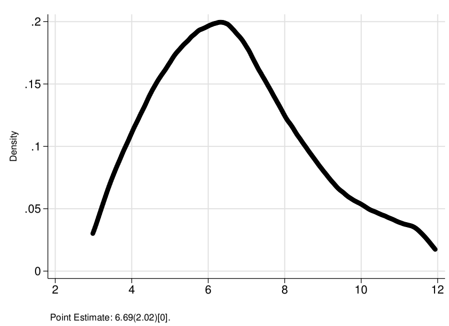
(e) Treatment vs. Staying at Home, Females



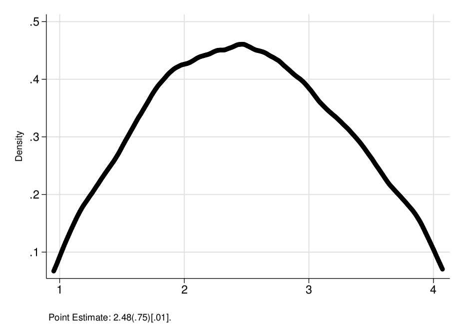
(f) Treatment vs. Staying at Home, Males



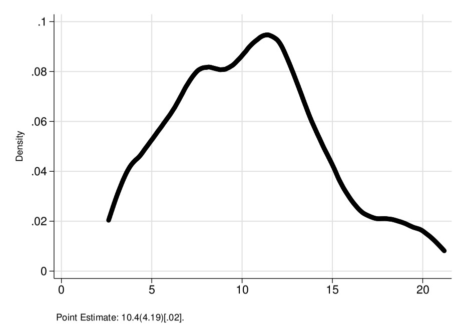
(g) Treatment vs. Alternative Preschool, Pooled



(h) Treatment vs. Alternative Preschool, Females



(i) Treatment vs. Alternative Preschool, Males



Note: Panel (a) displays the empirical bootstrap distribution for the estimate of the treatment vs. next parameter in the pooled sample. The reminder panels show an analogous distribution varying the parameter and the gender. See Section 3.1 for the definition of the parameters. Each panel displays the point estimate with the standard error in parentheses and the p-value in brackets.

C.6 Exploring the Impact of Other Forecasting Models

Our analysis is based on a causal model for treatment ($d = 1$) and control ($d = 0$) outcomes for measure j at age a in sample $k \in \{e, n\}$. We explore a range of standard panel data specifications for the error terms:

$$\begin{aligned} \varepsilon^d_{k,j,a} &= f^d + \omega^d_{k,j,a} \\ \omega^d_{k,j,a} &= \rho\omega^d_{k,j,a-1} + U^d_{k,j,a}, \end{aligned} \tag{9}$$

where $U^d_{k,j,a} \perp\!\!\!\perp \boldsymbol{X}^d_{k,a}$.

Here, we present different structures for $\phi^d_{k,j,a}(\cdot,\cdot)$ and $\varepsilon^d_{k,j,a}$ and investigate the robustness of our estimates to different assumptions about the structure of both these elements. We do this exercise for labor income.[90]

Note that Assumption A–4 (Invariance) implies that $\phi^d_{k,j,a}(\cdot,\cdot) = \phi_{k,j,a}(\cdot,\cdot) = \phi_{j,a}(\cdot,\cdot)$. That is, invariance holds across the treatment and the control groups and invariance holds across the experimental and the auxiliary samples. It is important to note that invariance across the treatment and the control groups implies that the variables $\boldsymbol{X}^d_{k,a}$ summarize the effect of the treatment on the outcome. Given this and Assumption A–3 (Exogeneity), the distribution of $\varepsilon^d_{k,j,a}$ is the same across the treatment and the control groups. We then drop the superscript in $\varepsilon^d_{k,j,a}$.[91]

In Appendix C.3.5, we also document that the support of $Y^d_{n,j,a}, \boldsymbol{X}^d_{n,a}, \boldsymbol{B}_n$ covers the support of $Y^d_{e,j,a}, \boldsymbol{X}_{e,a}, \boldsymbol{B}_e$ for $d \in \{0, 1\}$. This allows us to drop the d superscript in $Y^d_{n,j,a}, \boldsymbol{X}^d_{n,a}$ given that we estimate an invariant model.

We use linear specifications of $\phi_{j,a}$. We explore different alternatives. The system of

[90] In Appendix C.7, we describe the precise steps that we follow to construct out-of-sample forecasts based on these different structures and frame our estimation strategy in a GMM framework (Hansen, 1982).

[91] We test invariance across the treatment and the control groups and invariance across the experimental and the auxiliary samples in Appendix C.3.7.

interest is:

$$
\begin{aligned}
Y_{k,j,a} &= \lambda_0 + \lambda_1 Y_{k,j,a-1} + \lambda_2 \boldsymbol{X}_{k,a} + \varepsilon_{k,j,a} \\
\varepsilon_{k,j,a} &= \underbrace{f}_{\text{Fixed Effect}} + \underbrace{\omega_{k,j,a}}_{\text{Potentially Serially Correlated Component}} \\
\omega_{k,j,a} &= \rho \omega_{k,j,a-1} + \underbrace{U_{k,j,a}}_{\text{Independent Innovation}},
\end{aligned} \tag{10}
$$

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

Table C.33 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 $Y_{n,j,a}$ for $a \in [a^*, \ldots, A]$. In the experimental sample, we observe the outcome $Y_{e,j,a}$ 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 $a \in [a^*, \ldots, A]$ 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.33 summarizes the five specifications that we implement.

C.6.0.3 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 4, 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 $a = a^*$ (see Appendix C.3.7).

C.6.0.4 Specification 2: No Lagged Component ($\lambda_1 = 0$); Serial Correlation ($\rho \neq 0$); and No Fixed Effect ($f = 0$)

Given that $Y_{k,j,a-1}$ 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.[92]

The predictions reported in the baseline estimates are extremely 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.6.0.5 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 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:

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

$$Y_{k,a} = \lambda_0 (1 - \rho) + (\lambda_1 + \rho) Y_{k,a-1} - \lambda_1 \rho Y_{k,a-2} + \lambda_2 (\boldsymbol{X}_{k,a} - \rho \boldsymbol{X}_{k,a-1}) + U_{k,a}. \tag{11}$$

OLS produces consistent estimates of the coefficients in Equation (24). This enables us to construct forecasts, as explained in Appendix C.7. This model generates forecasts close to the baseline estimates.

C.6.0.6 Specification 4: Permanent-Transitory Decomposition of Unobserved Components ($\lambda_1 \neq 0$; $\rho = 0$; $f \neq 0$)

This framework allows for serial dependence due to the lagged dependent variable but does not allow for serial correlation in η_a. We explain our estimation strategy for this specification in Appendix C.7. Estimates from it are very similar to those from the other procedures.

C.6.0.7 Specification 5: Non-Parametric Predictions

An alternative to any of these scenarios is to make forecasts using a non-parametric matching algorithm. Specifically, (i) for each individual i in the experimental sample, e, find individual(s) $l(i)$ in the non-experimental sample, n, using Algorithm 1 in Appendix C.3.3; (ii) we impute the post-a^* trajectory of $Y_{k,j,a}$ of individual(s) $l(i)$ in the non-experimental sample, n, to individual i in the experimental sample, e.

A sufficient condition for the validity of this procedure is Assumption A–3 (Exogeneity). Without exogeneity, the joint distributions of $\boldsymbol{X}_{j,a*}, \varepsilon_{j,a^*}$ 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.

Table C.33: Net Present Value of Labor Income and Cost/Benefit Analysis Under Different Specifications for Labor Income Forecasts

	Specification 1: ("Baseline") $\boldsymbol{\lambda_1} \neq 0$ $\boldsymbol{\rho} = 0$ $\boldsymbol{f} = 0$			**Specification 2:** $\boldsymbol{\lambda_1} = 0$ $\boldsymbol{\rho} \neq 0$ $\boldsymbol{f} = 0$			**Specification 3:** $\boldsymbol{\lambda_1} \neq 0$ $\boldsymbol{\rho} \neq 0$ $\boldsymbol{f} = 0$			**Specification 4:** $\boldsymbol{\lambda_1} \neq 0$ $\boldsymbol{\rho} = 0$ $\boldsymbol{f} \neq 0$			**Specification 5:** Non-parametric matching		
	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C	NPV	IRR	B/C
Pooled	636,674	0.14	7.33	154,547	0.15	7.31	268,179	0.26	12.68	46,953	0.05	2.22	132,924	0.13	6.28
	(183,224)	(0.03)	(1.84)	(187,036)	(0.12)	(5.15)	(211,089)	(0.14)	(5.81)	(25,323)	(0.02)	(0.7)	(11,253)	(0.01)	(0.31)
Males	919,049	0.15	10.19	200,509	0.11	9.35	456,078	0.25	21.26	74,775	0.04	3.49	196,530	0.11	9.16
	(287,442)	(0.04)	(2.93)	(160,988)	(0.05)	(5.51)	(358,534)	(0.12)	(12.28)	(54,752)	(0.02)	(1.88)	(20,210)	(0.01)	(0.69)
Females	161,759	0.10	2.61	79,441	0.19	4.64	31,303	0.07	1.83	19,959	0.05	1.17	69,317	0.17	4.05
	(72,355)	(0.06)	(0.73)	(99,416)	(0.28)	(3.19)	(168,160)	(0.48)	(5.4)	(34,142)	(0.1)	(1.1)	(4,350)	(0.0 1)	(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.7 Estimation Procedure and Data Combination Estimator in the GMM Framework

Our analysis is based on a causal model for treatment ($d = 1$) and control ($d = 0$) outcomes for measure j at age a in sample $k \in \{e, n\}$ where e denotes membership in the experimental sample and n denotes membership in the auxiliary sample:

$$Y^d_{k,j,a} = \phi^d_{k,j,a}(\boldsymbol{X}^d_{k,a}, \boldsymbol{B}_k) + \varepsilon^d_{k,j,a}, \quad k \in \{n, e\}, \quad j \in \mathcal{J}_a, \quad d \in \{0, 1\}. \tag{12}$$

$\phi^d_{k,j,a}(\cdot, \cdot)$ is an invariant structural production relationship mapping inputs $\boldsymbol{X}^d_{k,a}, \boldsymbol{B}_k$ into output $Y^d_{k,j,a}$ holding error term $\varepsilon^d_{k,j,a}$ fixed. For simplicity, we initially assume A–3 (Exogeneity) holds. We relax this below.

In this section, we: (i) explain the procedure that we follow to form out-of-sample forecasts; and (ii) formulate the estimation procedure in a generalized method of moments (GMM) framework (Hansen, 1982).

In the auxiliary sample, we observe the outcome $Y_{n,j,a}$ for $a \in [a^*, \ldots, A]$. In the experimental sample, we observe the outcome $Y_{e,j,a}$ for at most two ages, depending on the outcome. We initially assume that we observe the outcomes at one age (a^*). We relax this assumption below.

Before explaining our estimation procedure, note that Assumption A–4 (Invariance) implies that $\phi^d_{k,j,a}(\cdot, \cdot) = \phi_{k,j,a}(\cdot, \cdot) = \phi_{j,a}(\cdot, \cdot)$. That is, invariance holds across the treatment and the control groups and invariance holds across the experimental and the auxiliary samples. Invariance across the treatment and the control groups implies that the variables $\boldsymbol{X}^d_{k,a}$ summarize the effect of the treatment on the outcome. Given Assumption A–4 (Invariance) and

Assumption A–3 (Exogeneity), the distribution of $\varepsilon^d_{k,j,a}$ is the same across the treatment and the control groups. This allows us to drop the superscript in $\varepsilon^d_{k,j,a}$.

We test and do not reject invariance across the treatment and the control groups and invariance across the experimental and the auxiliary samples in Appendix C.3.7.

In Appendix C.3.5, we document that the support of $Y^d_{n,j,a}, \boldsymbol{X}^d_{n,a}, \boldsymbol{B}_n$ covers the support of $Y^d_{e,j,a}, \boldsymbol{X}_{e,a}, \boldsymbol{B}_e$ for $d \in \{0,1\}$. So we drop the d superscript in $Y^d_{n,j,a}, \boldsymbol{X}^d_{n,a}$ given that we estimate an invariant model. We write:

$$Y_{k,j,a} = \phi_{j,a}(\boldsymbol{X}_{k,a}, \boldsymbol{B}_k) + \varepsilon_{k,j,a}, \quad k \in \{n, e\}, \quad j \in \mathcal{J}_a. \tag{13}$$

As we note in Appendix C.6, we work with a linear specification of $\phi_{j,a}$ in our empirical analysis. We explain our estimation procedure and the GMM framework for a general specification of $\phi_{j,a}$.

First, we explain our estimation and forecasting procedure using **Specification 1** in Appendix C.6. This is the specification that we follow in the main text. It is as follows.

1. Use the auxiliary sample (n) to estimate the the coefficients characterizing $\phi_{j,a}(\cdot,\cdot)$.[93]

   We denote these coefficients by $\boldsymbol{\theta}_{j,a}$ and the estimate of this function as $\hat{\phi}_{j,a}(\cdot,\cdot)$. At each age, we are able to compute the residuals from this estimation procedure as follows:

$$Y_{n,j,a} - \hat{\phi}_{j,a}(\boldsymbol{X}_{k,a}, \boldsymbol{B}_k) := \hat{\varepsilon}_{n,j,a}. \tag{14}$$

   For outcome j, we form the vector of residuals $\hat{\boldsymbol{\varepsilon}}_{n,j} := [\varepsilon_{n,\hat{j,a^*}+1}, \ldots, \varepsilon_{n,\hat{j,A}}]$.

[93] In practice, we use a weighted version of the auxiliary samples. The weights give relatively high importance to the individuals in the auxiliary sample whose characteristics $\boldsymbol{B}_k$ are close to the those of the individuals in the experimental sample. See Appendix C.3.3.

2. At age $a^* + 1$, we construct the forecasted outcome for the experimental sample (e) for each individual as follows:

$$\hat{Y}_{e,j,a^*+1} = \hat{\phi}_{j,a^*+1} \left(\boldsymbol{X}_{e,a^*+1}, \boldsymbol{B}_e \right). \tag{15}$$

We are able to evaluate $\hat{\phi}_{j,a^*+1}$ at $\boldsymbol{X}_{e,a^*+1}, \boldsymbol{B}_e$ even when $\boldsymbol{X}_{e,a^*+1}$ contains a one-period lag of Y_{e,j,a^*+1} because we observe Y_{e,j,a^*}. This prediction does not account for estimation error. We discuss estimation error below.

3. At age $a^* + 2$, we construct the forecasted outcome in the experimental sample (e) as follows:

$$\hat{Y}_{e,j,a^*+2} = \hat{\phi}_{j,a^*+1} \left(\boldsymbol{X}_{e,a^*+1}, \boldsymbol{B}_e \right). \tag{16}$$

We are able to evaluate $\hat{\phi}_{j,a^*+2}$ at $\boldsymbol{X}_{e,a^*+2}, \boldsymbol{B}_e$ even when $\boldsymbol{X}_{e,a^*+2}$ contains a one-period lag of Y_{e,j,a^*+2} because we can forecast Y_{e,j,a^*+1} from the previous step.

4. We iterate this procedure up to age A. For outcome j, we form the vector of forecasts $\hat{\boldsymbol{Y}}_{e,j} := \left[\hat{Y}_{e,j,a^*+1}, \ldots, \hat{Y}_{e,j,A} \right]$.

5. Under Assumption A–4 (Invariance), the distribution of $\hat{\boldsymbol{\varepsilon}}_{n,j}$ is a consistent estimator of the distribution of $\hat{\boldsymbol{\varepsilon}}_{e,j}$. We form a forecast that accounts for forecasting error as follows:

$$\tilde{\boldsymbol{Y}}_{e,j} = \hat{\boldsymbol{Y}}_{e,j} + \hat{\boldsymbol{\varepsilon}}_{n,j}. \tag{17}$$

We randomly sample a vector of residuals from an individual j in the auxiliary sample (n) and pair it with the vector $\hat{\boldsymbol{Y}}_{e,j}$ of individual i in the experimental sample (e) to form the forecast $\tilde{\boldsymbol{Y}}_{e,j}$ for individual i in the experimental sample. That is, the pairing of individual j in the auxiliary sample (n) with individual i in the experimental sample

(e) is random. Random pairing is valid under invariance and exogeneity, i.e. under this assumption the vector of residuals from any individual j in the auxiliary sample is a valid estimate for the vector of residuals of any individual i in the experimental sample. We form the pairing one time for the main point estimates, and then bootstrap this pairing when producing inference. See Appendix C.8 for more details on our inference procedures.

Second, we formulate our estimation strategy in terms of GMM (Hansen, 1982). To this end, note that Assumption A–3 (Exogeneity) and Assumption A–4 (Invariance) imply the following moment condition:

$$\mathbb{E}\left[\boldsymbol{m}_{j,a}\left(\boldsymbol{X}_{n,a}^{d}, \boldsymbol{B}_{n}; \boldsymbol{\theta}_{j,a}\right)\right] = 0, \quad k \in \{n, e\}, \quad j \in \mathcal{J}_{a} \tag{18}$$

where $\boldsymbol{m}_{j,a}\left(\boldsymbol{X}_{n,a}, \boldsymbol{B}_{n}; \boldsymbol{\theta}_{j,a}\right) := \boldsymbol{X_{n,a}}' \left(Y_{n,j,a}^{d} - \phi_{j,a}\left(\boldsymbol{X}_{n,a}^{d}, \boldsymbol{B}_{n}\right)\right)$ for $a \in [0, \ldots A]$.

We use the auxiliary sample (n) to estimate the vector of coefficients. Let $\boldsymbol{m}(\cdot, \boldsymbol{\theta})$, stack the function $\boldsymbol{m}_{j,a}\left(\boldsymbol{X}_{n,a}, \boldsymbol{B}_{n}; \boldsymbol{\theta}_{j,a}\right)$ for all $j \in \mathcal{J}_{a}$, all $a \in [0, \ldots A]$, and $k = n$.

Observing the outcomes at age a^{*} provides us with additional moment conditions. To see this, note that, in our analysis, $\boldsymbol{X}_{k,a^{*}+1}$ contains a lagged variable of the outcome to forecast and define the moment: $h_{j,a^{*}+1}\left(\boldsymbol{X}_{e,a^{*}+1}, \boldsymbol{B}_{n}; \boldsymbol{\theta}_{j,a^{*}+1}\right) =: \boldsymbol{X}_{e,a^{*}+1}' \left(\hat{Y}_{e,j,a^{*}+1} - \phi_{j,a^{*}+1}\left(\boldsymbol{X}_{e,a^{*}+1}, \boldsymbol{B}_{e}\right)\right)$, where $\hat{Y}_{e,j,a^{*}+1}$ is defined as before. Although this moment uses information in the auxilliary sample (through the construction of $\hat{Y}_{e,j,a^{*}+1}$), it provides additional information (not in (18)) through $\boldsymbol{X}_{e,a^{*}+1}$. It is a key moment, because it initializes the out-of-sample forecasts.

For some outcomes, there are gaps in the experimental sample. For example, we observe labor and transfer income at ages 21 and 30. In this case, we have two additional moments, not only one. Stack these set of additional moments and denote them by $\boldsymbol{h}(\cdot, \boldsymbol{\theta})$ (and helps us initialize the out-of-sample forecasts). These additional set of moments overidentify the parameter vector of interest, $\boldsymbol{\theta}$. Standard procedures allow us to use these set of additional moments to improve efficiency.

Let $\boldsymbol{W}$ be a positive definite matrix. We estimate $\boldsymbol{\theta}$ by minimizing

$$Q := \begin{bmatrix} \bar{\boldsymbol{m}}\left(\cdot;\boldsymbol{\theta}\right) \\ \bar{\boldsymbol{h}}\left(\cdot;\boldsymbol{\theta}\right) \end{bmatrix}' \boldsymbol{W}^{-1} \begin{bmatrix} \bar{\boldsymbol{m}}\left(\cdot;\boldsymbol{\theta}\right) \\ \bar{\boldsymbol{h}}\left(\cdot;\boldsymbol{\theta}\right) \end{bmatrix}, \tag{19}$$

where $\bar{u}$ denotes the empirical counterpart of u.

$\boldsymbol{W}$ is not restricted to be diagonal so that these moments are allowed to correlate. Iterated, feasible procedures to obtain an estimate of $\boldsymbol{W}$ jointly with the parameters of interest guarantee efficiency and are straightforward to implement (Hansen, 1982; Amemiya, 1985).[94]

We explain the samples used to construct each empirical counterpart and the procedure to obtain standard errors on the predictions in Appendix C.3 and Appendix C.8, respectively.

We next adapt the procedure and the GMM framework to the rest of the specifications. **Specification 2** is simpler because it does not depend on lagged outcomes. The steps are the following:

1. Use the auxiliary sample (n) to estimate the the coefficients characterizing $\phi_{j,a}\left(\cdot,\cdot\right)$.[95]

   We denote these coefficients by $\boldsymbol{\theta}_{j,a}$ and the estimate of this function as $\hat{\phi}_{j,a}\left(\cdot,\cdot\right)$. At

[94] Altonji and Segal (1996) show that GMM presents downwards bias in absolute value in small-sample size setting, which is a concern in our setting.

[95] In practice, we use a weighted version of the auxiliary samples. The weights give relatively high importance to the individuals in the auxiliary sample whose characteristics $\boldsymbol{B}_k$ are close to the those of the individuals in the experimental sample. See Appendix C.3.3.

each age, we are able to compute the residuals from this estimation procedure as follows:

$$Y_{n,j,a} - \hat{\phi}_{j,a}(\boldsymbol{X}_{k,a}, \boldsymbol{B}_k) := \hat{\varepsilon}_{n,j,a}. \tag{20}$$

For outcome j, we form the vector of residuals $\hat{\boldsymbol{\varepsilon}}_{n,j} := [\hat{\varepsilon}_{n,j,a^*+1}, \ldots, \hat{\varepsilon}_{n,j,A}]$.

2. At age $a \geq a^* + 1$, we construct the forecasted outcome for the experimental sample (e) for each individual as follows:

$$\hat{Y}_{e,j,a} = \hat{\phi}_{j,a}\left(\boldsymbol{X}_{e,a}, \boldsymbol{B}_e\right). \tag{21}$$

We are able to evaluate $\hat{\phi}_{j,a^*+1}$ at $\boldsymbol{X}_{e,a^*+1}, \boldsymbol{B}_e$ because $\boldsymbol{X}_{e,a^*+1}$ is fully observed in the experimental data. We stack the forecasts across ages in the following vector $\hat{\boldsymbol{Y}}_{e,j} := \left[\hat{Y}_{e,j,a^*+1}, \ldots, \hat{Y}_{e,j,A}\right]$. These forecasts do not account for estimation error. We discuss estimation error below.

3. Under Assumption A–4 (Invariance), the distribution of $\hat{\boldsymbol{\varepsilon}}_{n,j}$ is a consistent estimator of the distribution of $\hat{\boldsymbol{\varepsilon}}_{e,j}$. We form a forecast that accounts for forecasting error as follows:

$$\tilde{\boldsymbol{Y}}_{e,j} = \hat{\boldsymbol{Y}}_{e,j} + \hat{\boldsymbol{\varepsilon}}_{n,j}. \tag{22}$$

In practice, we randomly sample a vector of residuals from an individual j in the auxiliary sample (n) and pair it with the vector $\hat{\boldsymbol{Y}}_{e,j}$ of individual i in the experimental sample (e) to form the forecast $\tilde{\boldsymbol{Y}}_{e,j}$ for individual i in the experimental sample. That is, the pairing of individual j in the auxiliary sample (n) with individual i in the experimental sample (e) is random. Random pairing is valid under invariance and exogeneity, i.e. under this assumption the distribution of the residuals for individuals

j in the auxiliary sample is a valid estimate for the vector of residuals for individuals i in the experimental sample. We form the pairing one time to obtain our estimates, and then bootstrap this pairing when producing inference. See Appendix C.8 for more details on our inference procedures.

In this specification, there is no "initialization" of the forecast out of sample. Thus, the GMM estimate consists of minimizing

$$Q := \Big[\bar{\boldsymbol{m}}\,(\cdot;\boldsymbol{\theta})\Big]'\boldsymbol{W}^{-1}\Big[\bar{\boldsymbol{m}}\,(\cdot;\boldsymbol{\theta})\Big], \tag{23}$$

where $\boldsymbol{m}_{j,a}\left(\boldsymbol{X}_{n,a},\boldsymbol{B}_n;\boldsymbol{\theta}_{j,a}\right) := \boldsymbol{X_{n,a}}'\left(Y^d_{n,j,a} - \phi_{j,a}\left(\boldsymbol{X}^d_{n,a},\boldsymbol{B}_n\right)\right)$ for $a \in [0,\dots A]$ and $\boldsymbol{X}_{n,a}$ contains no lags of $Y^d_{n,j,a}$.

To explain the forecasting steps for **Specification 3**, recall that we ρ-transform the model and write:

$$Y_{k,a} = \lambda_0\left(1-\rho\right) + \left(\lambda_1+\rho\right)Y_{k,a-1} - \lambda_1\rho Y_{k,a-2} + \lambda_2\left(\boldsymbol{X}_{k,a} - \rho\boldsymbol{X}_{k,a-1}\right) + U_{k,a} \tag{24}$$

This is a model with two lags and no serial correlation. The estimation procedure and the GMM framework are analogous to those of **Specification 1**. The two lags are not an issue for estimation in the auxiliary sample because we observe labor income for the full range of relevant ages, thus we estimate the prediction function. To initialize the procedure in the experimental sample, however, we face an issue: we do not observe labor income at a^*-1. We assume that labor income at age a^* is the same as at age a^*-1 and then proceed in a similar way as in **Specification 1**, the estimation procedure and the GMM framework remain the same.

Now, we explore **Specification 4**. We drop the exogenous regressors for expostional simplicity, as they do not bring in interesting features to the problem. We write:

$$
\begin{aligned}
Y_{k,a} &= \lambda_0 + \lambda_1 Y_{k,a-1} + \varepsilon_a & (25)\\
\varepsilon_{k,a} &= f + U_{k,a}, & (26)
\end{aligned}
$$

where $\mathbb{E}[U_a] = \mathbb{E}[U_a, U_{a'}] = 0$. We follow Arellano and Bond (1991) and note that two-lagged age values of $Y_{k,a}$ are valid instruments in the first-difference version of Equation (26). This allows us to obtain consistent estimates of λ_0, λ_1 by minimizing a weighted function (as in the previous specifications) of the empirical counterparts of the following set of moments:

$$
\mathbb{E}\left[\left(\Delta Y_{k,a} - \lambda_1 \Delta Y_{k,a-1}\right) Y_{k,a-j}\right] \quad j = 2, \dots, a-1; a = a^* + 2, \dots, A. \tag{27}
$$

Using estimates from this procedure, we form the forecast in the following steps:

1. Use the auxiliary sample (n) to estimate the coefficients in Equation (25) based on the set of moments in (27).

2. At age $a^* + 1$, use these coefficients to form the (out-of-sample) forecast in the experimental sample (e):

$$
\hat{Y}_{e,a^*+1} = \hat{\lambda}_0 + \hat{\lambda}_1 Y_{e,a^*}, \tag{28}
$$

   noting that we observe Y_{k,a^*}.

3. At age $a^* + 2$, use the same coefficients to form the (out-of-sample) forecast, based on the $a^* + 1$ forecast. That is:

$$\hat{Y}_{e,a^*+2} = \hat{\lambda}_0 + \hat{\lambda}_1 \hat{Y}_{e,a^*+1}. \tag{29}$$

4. Iterate this procedure of to age A and stack the vector of forecasts (without accounting for forecasting error) as $\hat{\boldsymbol{Y}}_e := \left[\hat{Y}_{e,a^*+1}, \ldots, \hat{Y}_{e,A}\right]$.

5. To account for forecasting error we need an individual level estimate of $f + u_a$. We proceed as follows: (i) we observe labor income at two ages, 21 and 30. We use the estimates of the coefficients characterizing Equation (25) from the auxiliary sample (n) to forecast labor income from ages 22 to 29. Then, we estimate the coefficients in Equation (25) in the experimental sample (e). This allows us to recover an estimate for $f + u_a$. In fact, we recover one estimate of $f + u_a$ for each $a \in [22, \ldots, 30]$. Each of these estimates is a valid estimate for $f + u_a$ because u_a is i.i.d. To form our forecasting error, at each age, we take the average of these available estimates. We add it to $\hat{Y}_{e,a}$ for $a \geq a^* + 1$ to form a prediction that accounts for error.

C.8 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.8.1 Forecasts

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

1. Resample the experimental sample with replacement at the individual level. This gives us a new (re-sampled) panel dataset. Information on the entire history of each individual is obtained in each re-sample.[96] Call this resampled sample (e, s). Separate

[96] We re-sample individuals independently of their treatment status.

this sample by treatment and control group into $(e, s, 1)$ and $(e, s, 0)$, respectively.

2. Perform the same resampling procedure on the auxiliary sample. Call this sample (n, s').

3. Form synthetic treatment and control groups by using Algorithm 1 to weight the individuals in sample (n, s'). We do not do this by age due to problems of data availability. We use the algorithm once to match (e, s) to the CNLSY and once to match (e, s) to the PSID and NLSY79. We use the synthetic groups obtained from each of these samples to form predictions at different ages, as we explain in Appendix C.3.2. We identify synthetic control and treatment groups $(n, s', 0)$ and $(n, s', 1)$, respectively. That is, (n, s', d) for $d = 0, 1$.

4. Fit the dynamic relationship in Equation (12), using predictors as detailed in Appendix C.3.4. We fit two parameterizations of the dynamic relationships. One for the synthetic treatment, and one for the synthetic control. When providing estimates by gender, we also produce different predictions by gender.

5. Use the parameterization in Step 4. to fit out of sample in $(e, s, 1)$ and $(e, s, 0)$, respectively. This gives us an age-by-age forecast *without forecasting error* for our treatment and control groups. Store the predictions at all ages for individual i in this sample in a vector $\boldsymbol{Y}^d_{i,e,s}$, where $\boldsymbol{Y}^d_{i,e,s}$ is the vector of forecasts for individual i in the experimental bootstrap sample s, experimental group d.

6. In step 4., we compute an individual-level vector of residuals in each of the samples $(n, s', 0)$ and $(n, s', 1)$. That is, each individual has a vector containing the residuals of each of her predicted variable (for example, labor income). Call this vector of residuals $\boldsymbol{\mathcal{E}}^d_{i',n,s'}$: the vector of residuals for individual i' in the auxiliary bootstrap sample s', in the synthetic group d.

7. Randomly pair individual i' in s' with individual i in s. The forecast accounting for

forecasting error is $\boldsymbol{Y}^d_{i,e,s} + \boldsymbol{\mathcal{E}}^d_{i',e,s'}$. As described in Appendix C.7, this step changes. We estimate the forecasting error from the experimental sample (and we account for this when bootstrapping as well).

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

9. 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.8.2 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.9 Procedures for Selecting Background Variables, Estimated Treatment Effects, and Estimated Combining Functions

In this appendix we first explain our method for selecting the background variables that we control for when estimating treatment effects.[97] Then, we present the treatment effects of the center-based treatment in ABC/CARE estimates for the 95 main outcomes we consider. For each set of estimates, we first present a summary of the effect of the program using a combining function counting the number of socially positive treatment effects. We then present tables of treatment effect estimates for each outcome. Finally, we test for statistically significant treatment effects using the step-down procedure to test multiple hypotheses.

C.9.0.1 Background Variables

We select three out of fourteen 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.34 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.

Table C.34: 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.

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

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 application, $\mathcal{M}$ 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,

$$\begin{aligned}
\Pr(m|\text{Data}) &= \frac{\Pr(\text{Data}|m) \times \Pr(m)}{\Pr(\text{Data})} \\
&= \frac{\Pr(\text{Data}|m) \times \Pr(m)}{\sum\limits_{m' \in M} \Pr(\text{Data}|m') \Pr(m')} \\
&\propto \Pr(\text{Data}|m) \times \Pr(m),
\end{aligned} \tag{30}$$

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 to rank the the likelihood 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.34.

C.9.0.2 Matching Variables

We use matching estimators for different versions of the “treatment vs. stay at home” and “treatment vs. alternative preschools” parameters. For treatment vs. stay at home, we construct the Mahalanobis distance between the individuals in the treatment group and the control group who stay at home and use an Epanechnikov metric to construct an individual-level weight—giving a relatively high weight to individuals in the treatment group who would have been likely to stay at home if randomized to the control group. We proceed analogously when estimating the treatment vs. alternative preschool parameters. We use the same variables to "match" and to "control".

Other forms of matching estimates such as propensity score matching and nearest neighbor(s) give very similar results and are available upon request. We analyze sensitivity to the choice of controls and matching variables next.

C.9.0.3 Sensitivity Analysis

An immediate route of inquiry has to do with the sensitivity of our estimates to the choice of background variables. Especially in the context of our small sample, in which estimates can vary to different model specifications. To investigate this, 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.34. We also consider all possible control sets of one and two variables in Table C.34. 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 to Figure C.9 display 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. This is especially true for the comparisons of treatment vs. staying at home and vs. alternative preschool.

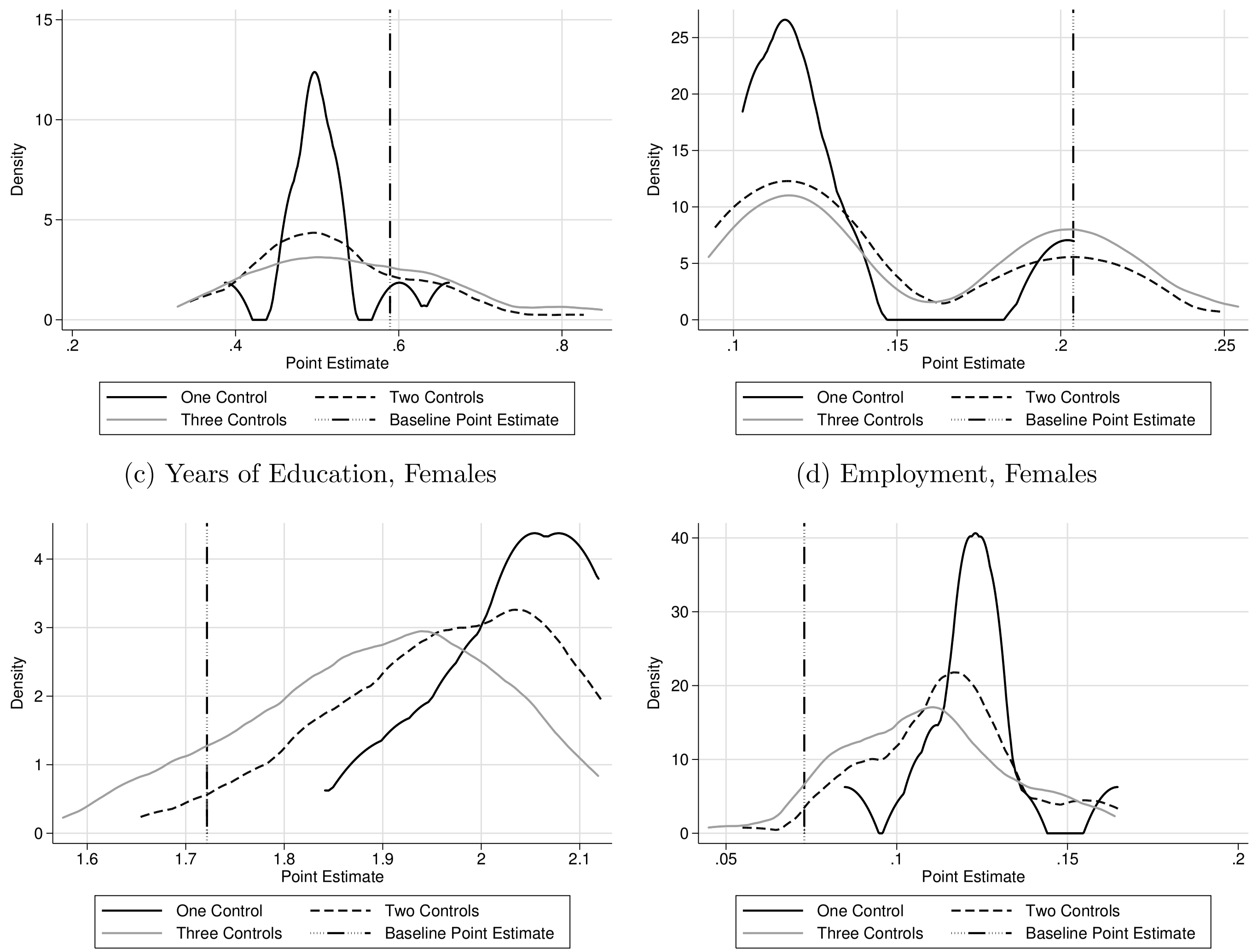
Figure C.7: Sensitiviy to Choice of Control Set, Treatment vs. Next Best

(a) Years of Education, Males

(b) Employment, Males

(c) Years of Education, Females

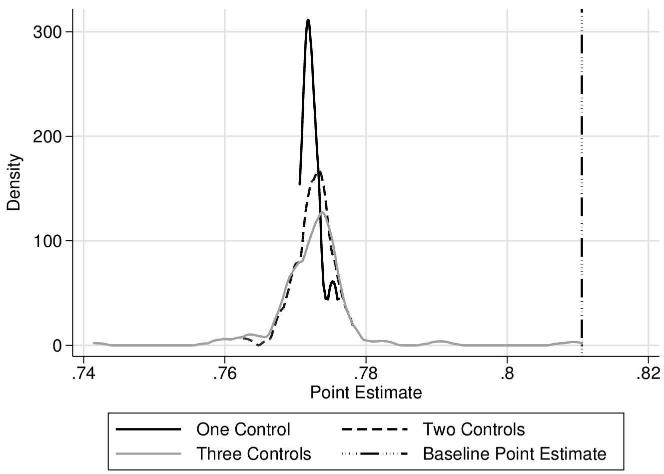
(d) Employment, Females



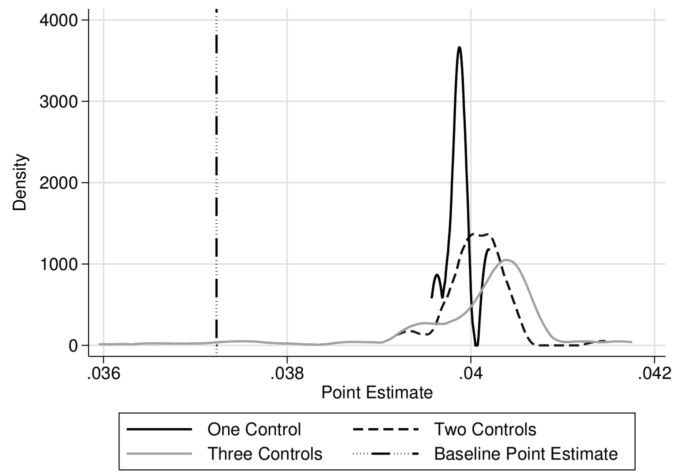
Note: Panel (a) displays the distribution of the treatment effect estimate of the treatment compared to next best counterfactual for males years of education. The distribution is obtained by using all possible combinations of one, two, and three background variables listed in Table C.34. In addition 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 reminder panels present analogous distributions for the outcomes and genders indicated in the title.

Figure C.8: Sensitiviy to Choice of Control Set, Treatment vs. Stay at Home

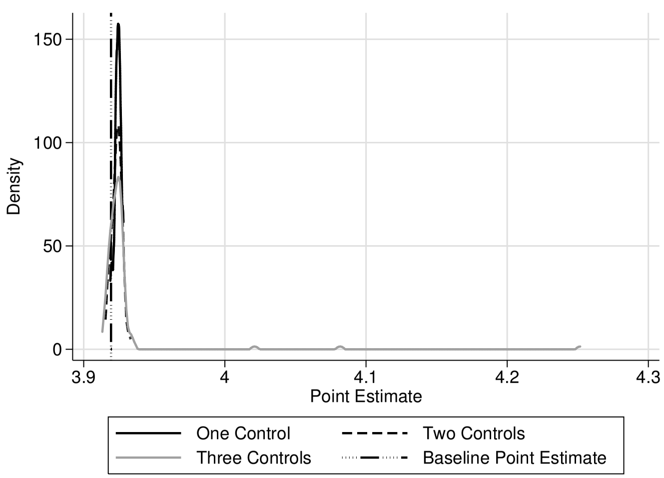
(a) Years of Education, Males



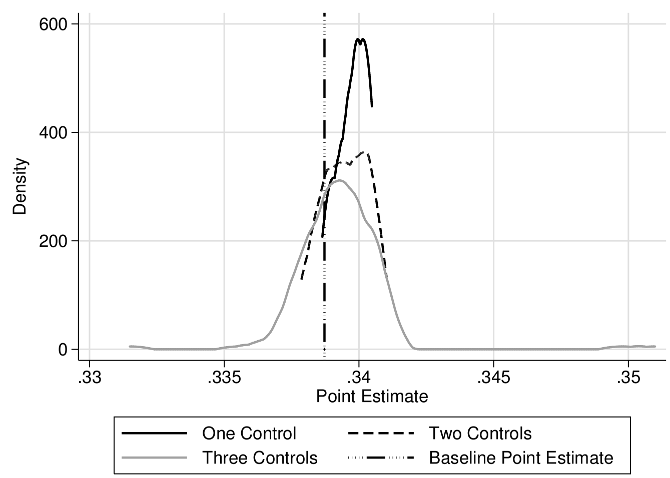
(b) Employment, Males



(c) Years of Education, Females



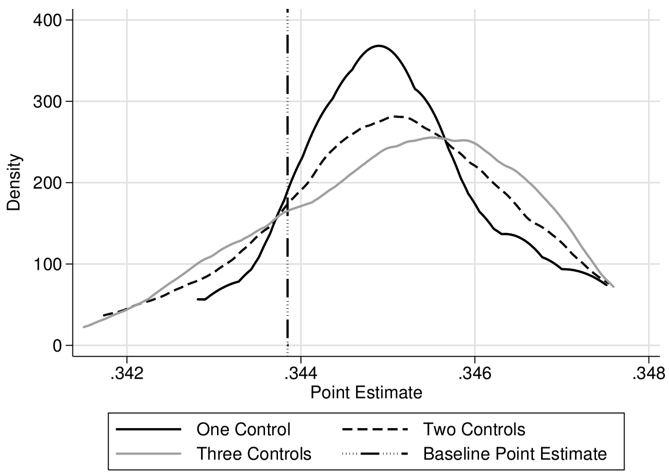
(d) Employment, Females



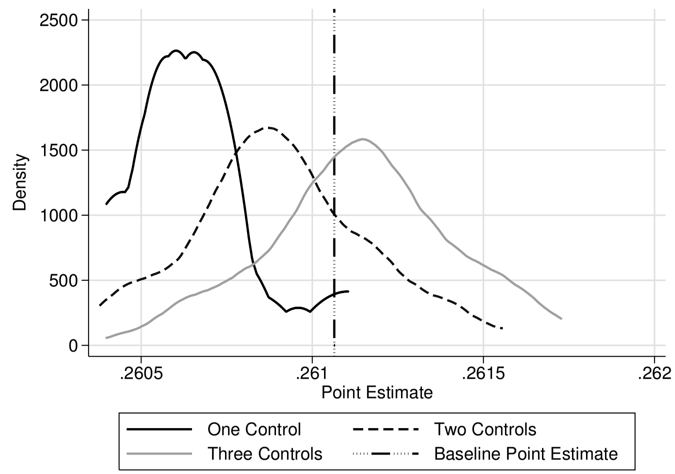
Note: Panel (a) displays the distribution of the treatment effect estimate of the treatment compared to stay at home counterfactual for males years of education. The distribution is obtained by using all possible combinations of one, two, and three background variables listed in Table C.34. In addition 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. We “match” and “control” using the same set of variables. The horizontal line marks the baseline estimate we use. The reminder panels present analogous distributions for the outcomes and genders indicated in the title.

Figure C.9: Sensitiviy to Choice of Control Set, Treatment vs. Alternative Preschool

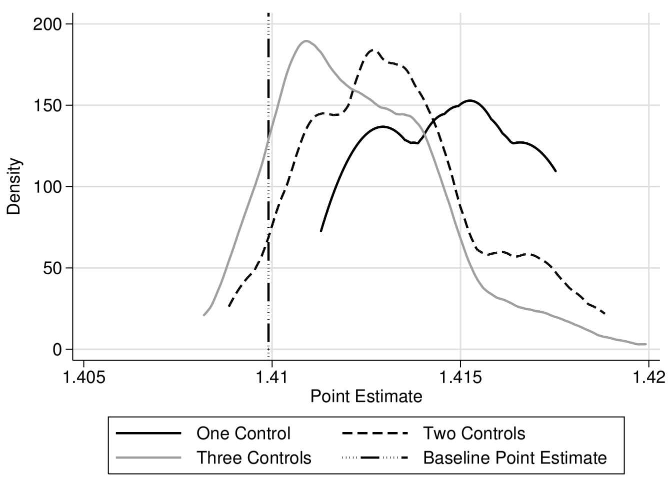
(a) Years of Education, Males



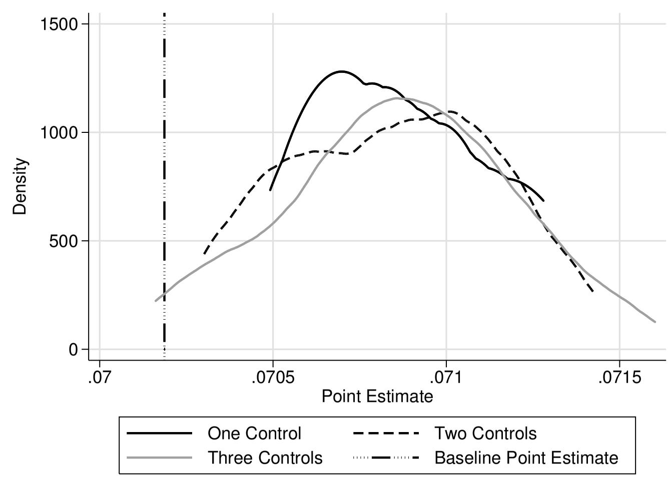
(b) Employment, Males



(c) Years of Education, Females



(d) Employment, Females



Note: Panel (a) displays the distribution of the treatment effect estimate of the treatment compared to alternative preschool 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.34. 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. We "match" and "control" using the same set of variables. The horizontal line marks the baseline estimate we use. The reminder panels present analogous distributions for the outcomes and genders indicated in the title.

D Costs of Education

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 the subjects' later educational attainment. Instead of forming a projection of future education costs, we do not add further modeling uncertainty through this component and document that, at the national level, education beyond age 30 increases marginally. To document this, we use the Panel Study of Income Dynamics (PSID) for a representative sample of individuals 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.

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. Thus, we perform an exercise using nationally representative data from the Panel Study of Income Dynamics (PSID) to assess educational attainment after age 30. We find only one additional year of schooling for individuals between the ages of 40 and 60.

D.2 Cost of Education

We apply the costs described in Table D.1 to subjects' educational attainment at age 30 to calculate the public and private costs of lifetime educational attainment. Costs up to high school are assumed to be public 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. Only K-12 education, special education, and grade retention costs account for deadweight loss. Because it gives costs that are applied across many years, this table does not show their present discounted value.

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 section. 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.[98]

E Quantifying the Benefits in 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.

[98] Lochner (2011b) or Lochner (2011a).

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 descriptions 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 created 3 years ago 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, it is inevitable to have some crimes that 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 total treatment effect on crime and the costs associated with that effect. We first give a more abstract and formal summary of the process, 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 data, we have a dataset to forecast lifetime sentences. In that dataset, 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 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: $\widehat{V_{i,j}^S} = S_{i,j} f_j r_j$. 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, we use several datasets to construct measures of criminal activity of the ABC/CARE subject. We generate a count of the number of arrests and of the number of sentences by crime category for each ABC/CARE subject. We start by counting sentences, which is trivial, 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.[99] We now describe the procedure we follow to get a count of arrests that is as complete as possible. We begin by 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.

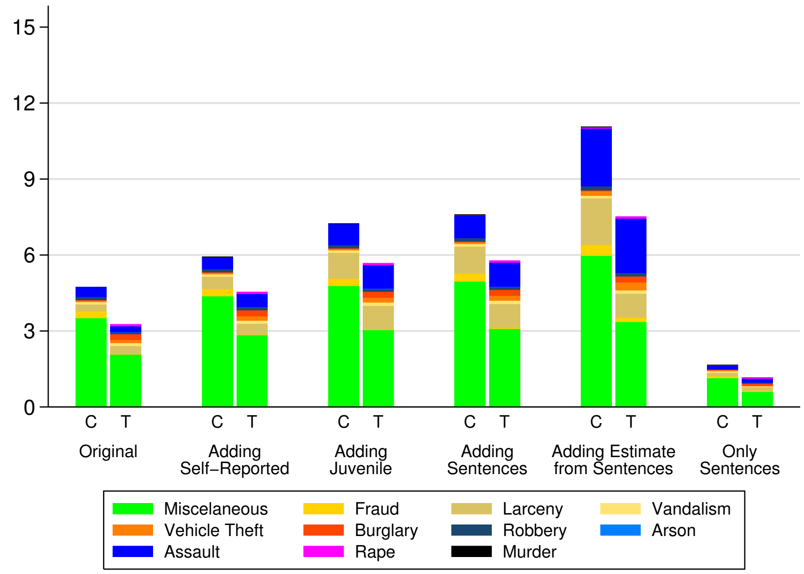
The main problem we confront with these data is that there are some individuals who are missing data on arrests in the data collected at age 34. We know this is the case because there are sentences for which we do not observe the corresponding arrests. While this is

[99] 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. An offense is the precise definition for the unit that we work with.

expected for some crimes, such as speeding or shoplifting, it is not possible for others. This 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 that necessarily imply they were arrested at some point of time. 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 significant, 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 the 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 Construct 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 estimations 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 older ages because the ABC/CARE subjects are currently about 46 years old. To forecast crimes at older ages, it is necessary to use 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. * $p < 0.05$; ** $p < 0.01$.

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

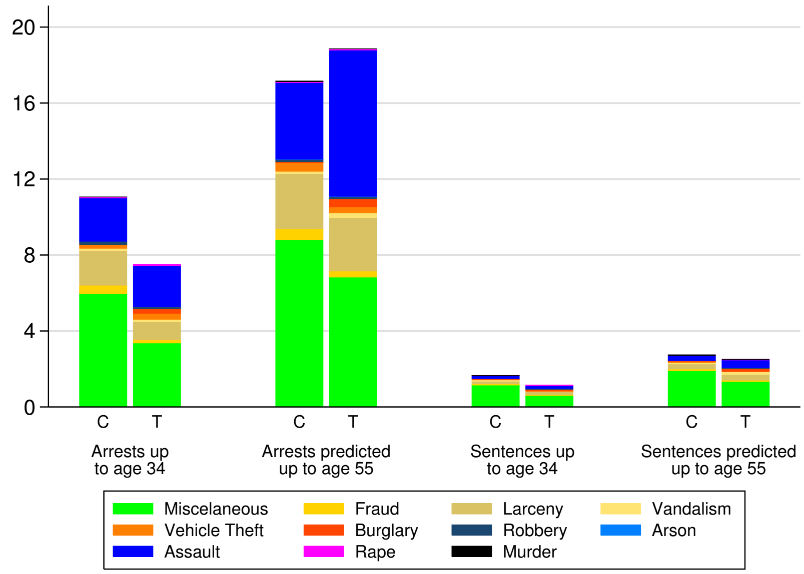
	Miscellaneous	Fraud	Larceny	Vandalism	Auto Theft	Burglary	Robbery	Arson	Assault	Rape	Murder
Miscellaneous	0.037	0.000	0.003	0.000	0.000	0.001	0.000	-0.000	0.001	-0.001	-0.001
	(19.83)**		(4.07)**	(2.77)**	(1.77)	(2.02)*	(0.64)	(0.23)	(2.40)*	(3.15)**	(4.97)**
Fraud	0.020	0.000	0.005	-0.000	0.000	0.002	-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. * $p < 0.05$; ** $p < 0.01$.

We perform the previous process separately for males and females. We use linear forecasts, 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 effect of our forecasting methodology. The effect is quantitatively much larger for arrests than for sentences. For both arrests and sentences, including the forecasted 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.

E.3 Victimization Inflation

Even though we have administrative data on crimes, we only observe the crimes that had 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.[100] 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

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

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 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 highest estimate among 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.

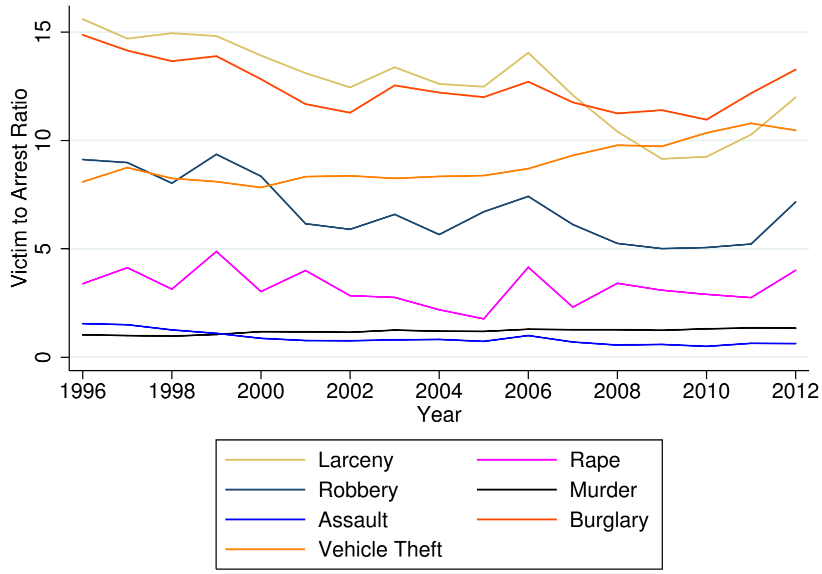
E.3.2 Construction of the Total Number of Arrests in the U.S.

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

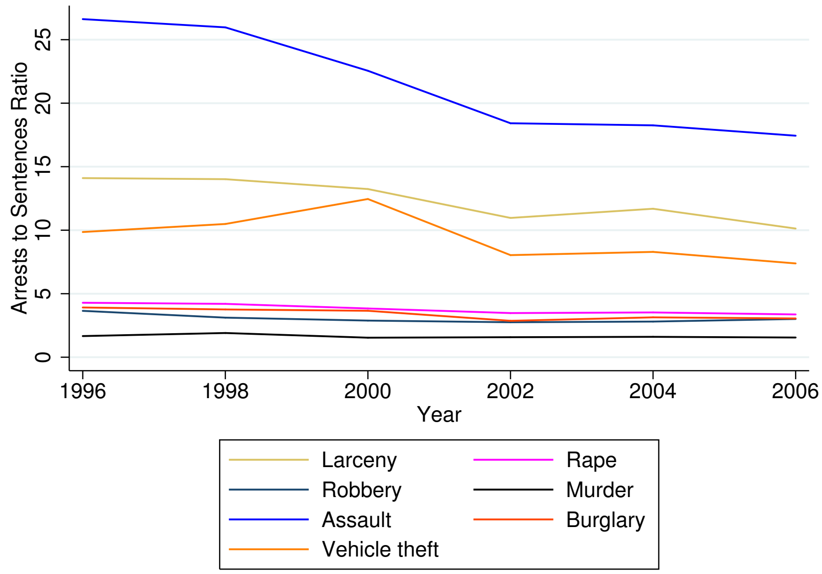
Figure E.3 and Figure E.4 show the VI factors calculated by year. The ratios in the charts are constructed as $r_{j,t} = \frac{\overline{V_{j,t}}}{\overline{A_{j,t}}}$. In practice, we use an average of all the yearly measures in our calculations given that this exercise imputes unobserved crimes that 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 will 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

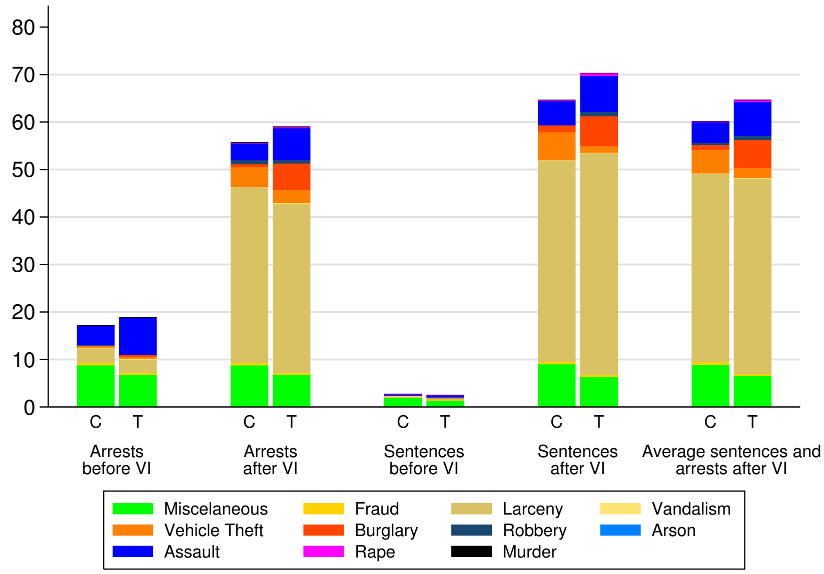


Note: This figure shows, by year and type of crime, the number of arrests (estimated from the National Arrests Analysis Tool from the NBJS) divided by the number of sentences (estimated from the National Justice Reporting Program). In practice, we use a single number for each type of crime, which is 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 to estimate unit costs of representative crimes, and many studies presenting estimates.[101] In this document, we only review the literature related to the inputs necessary for this paper.

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

We start by classifying the costs of crime, which is necessary to later discuss the methods to estimate the costs. Then, we present the two general types of methodologies 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 that are more homogeneous between 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, and it might not even be clear what other methodologies are capturing. 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, like 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.[102] 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.[103] Being an example of a TD methodology, it is expected that the costs obtained by this method

[102] 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.
[103] Cohen et al. (2004).

would 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 obvious caveat of this method is that people might not answer the real amount they would be willing to pay in these surveys.

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.[104] 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 extremely large and rich 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.[105]

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 spirit 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

[104] Thaler (1978).

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

that when people answer the survey used for TD calculations they include some costs 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. For completeness, we present cost estimates using both approaches. 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

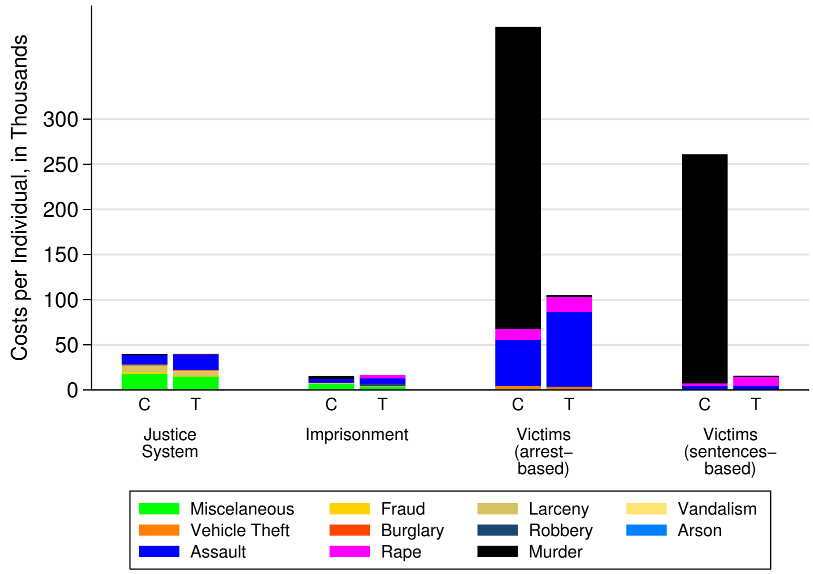
Unlike previous studies, we observe the sentences of the ABC/CARE subjects. This allows for a precise estimation of the costs of imprisonment. For the cost of jail and state prison, we use estimates from the U.S. Department of Justice (1988). In 2014 USD, these costs are \$25,338 for a year in a state 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 appears in 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

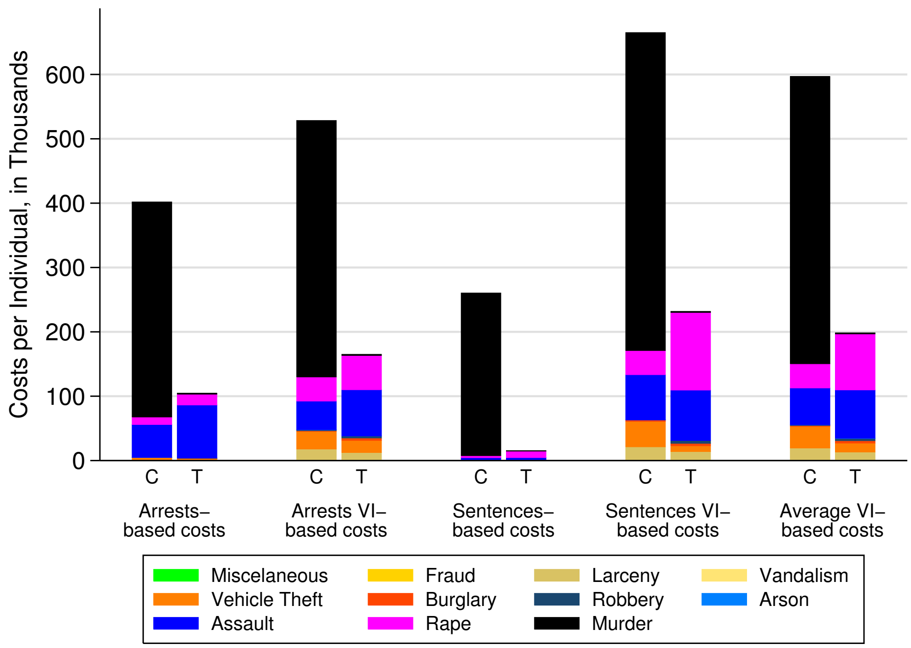


Note: This figure depicts the per capita cost for the different categories of costs and crimes we use, by control (C) and treatment (T). The first pair of columns adds up the justice system costs (including police) for all arrests inputed for each subject. The second pair of columns adds up the cost 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. Thus, 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.[106]

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

[106] Heckman et al. (2010).

unidimensional summary of benefits of a program, and specific flows of benefits with high variability enter naturally into the process of aggregation.

E.6 Sensitivity Analyses Using Alternative Cost Estimates

So far, we study several ways to construct our estimates:

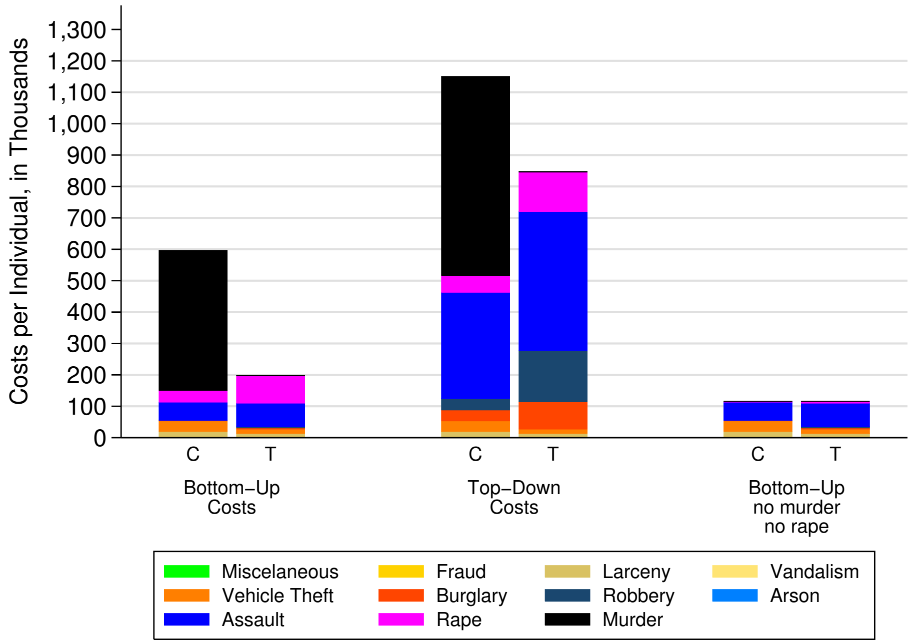
1. We show that not matching the different crime datasets could reduce the number of crimes, but that the general patterns are stable, and no dataset is especially influential.

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 estimates using arrest-based estimations versus sentence-based estimations, and find that the differences are large in terms of the number of crimes before victimization inflation, but small after it, and do not substantially change the total benefits calculations.

In Figure E.8, we present additional deviations from our main estimates. In particular, we show how the estimations 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.

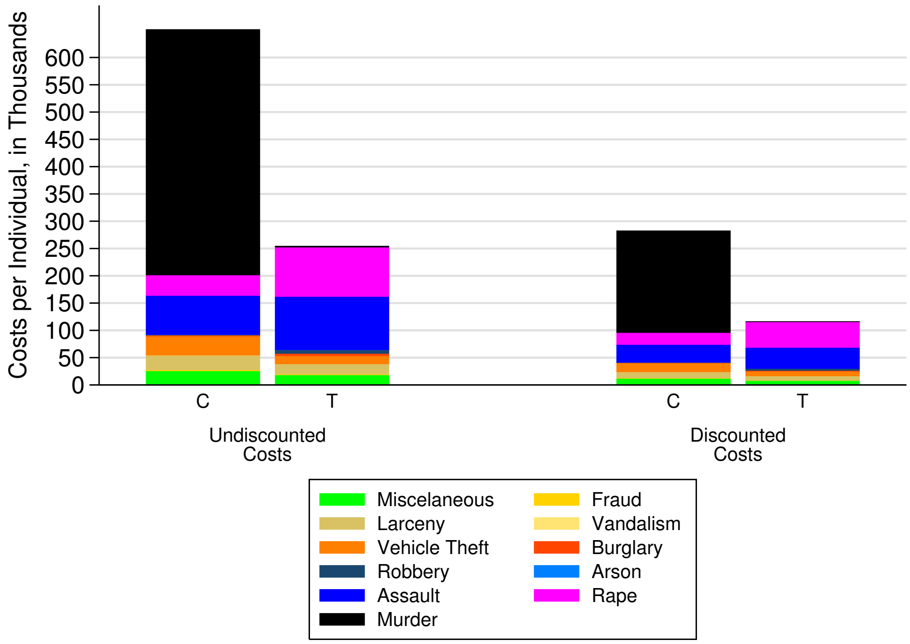
Figure E.8: Different Cost Schedules



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 Health Outcomes in the Future America Model (FAM)

F.1 Background and Description of FAM

In this appendix, we explain our methodology to measure the projected differences in health outcomes and medical expenditure over the adult life for the treated and control groups in ABC/CARE. Health outcomes and behaviors of ABC/CARE subjects are measured at the age-30 interview and in a health follow-up conducted when the subjects were in their mid-30s. To project the life-cycle path of health outcomes and medical expenditures, we use a dynamic microsimulation model to track the treatment and control cohorts from age 30 until death.[107]

[107]This microsimulation model is an extension of the model used by Prados et al. (2015); the technical details are described in Tysinger et al. (2015). Both models are related to the Future Elderly Model (FEM),

The defining characteristic of this approach is the modeling of real rather than synthetic cohorts, all of whom are followed at the individual level. This allows for more heterogeneity in behavior than would be allowed by a cell-based approach. The core of the model can be described as follows: first, the cohort starter module contains the health outcomes of the ABC/CARE subjects at age 30 and other variables that are used as input for the simulation of individual trajectories. This module uses ABC/CARE data. Missing variables in the ABC/CARE data are imputed probabilistically from models estimated using PSID data. Next, the transition module calculates the individual probabilities of transiting across various health states and other outcomes relevant to health. The transition probabilities are estimated from the longitudinal data in the PSID, taking as inputs risk factors such as smoking, weight, alcohol consumption, gender, race and ethnicity, age and education, along with lagged health, personal, and economic states.[108] This scheme allows for a great deal of heterogeneity and fairly general feedback effects. Finally, the outcomes module aggregates projections of individual-level outcomes into outcomes such as QALY and medical expenditures.

The cohort starter module includes the following variables for each ABC/CARE subject:

- Individual characteristics: year of birth, gender, treatment status, education of the mother, self-reported "poor" economic condition as a child, race, and education at age 30.

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 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, because they are important to explain health outcomes and medical expenditure at younger ages, like evolution of partnership and marital status, work status, and family size.

[108]Section F.2 provides details about the data sources used in the estimation.

- Economic outcomes at age 30: working status, earnings, and health insurance status.
- Health outcomes and health behaviors at age 30: body mass index (BMI), smoking, binge drinking, physical activity, psychological distress, asthma, high blood pressure, heart disease, cancer, lung disease, diabetes, and stroke.

Health conditions and health behaviors are derived from survey questions about doctor-diagnosed conditions and self-reported health behaviors. For details about how we deal with missing variables and assumptions, see Section F.3.1.

The core of the microsimulation is a set of models of disease conditions designed to forecast future health and functional status of each individual from his or her current health state at age 30. To forecast health and economic outcomes over time, the model calculates the transition probabilities between various health states and other outcomes. Health states include diabetes, heart disease, stroke, cancer, hypertension, lung disease, and number of difficulties in physical and instrumental activities of daily living (ADLs and IADLs); health risks include BMI (defined as weight in kilograms divided by height in meters squared, which is used to measure incidence of obesity), binge drinking (defined as binge drinking at least three times per month[109]), smoking behavior and (lack of) physical activity; other outcomes include health insurance status, changes in family structure (partnership or marriage, child-bearing), labor market participation, working status, receipt of Social Security, participation in public programs, and medical expenditures. We also estimate quality-adjusted life years (QALYs), a measure of the quality of life that adjusts for the burden of disease.[110]

The likelihood of developing a health condition depends on key risk factors including age, gender, education, race and ethnicity, obesity (BMI greater than 30 kg/m^2), smoking status,

[109] Where binge drinking behavior is defined as drinking more than five alcoholic drinks in an instance for males and more than four for females.

[110] A QALY equals one year of life in the absence of disease. This measure has been widely used in the literature to evaluate the value of medical interventions and improvements.

physical activity, age of asthma diagnosis, and lagged health outcomes. By incorporating lagged variables into the model, we account for the likelihood that past behaviors may influence risks far into the future. This capacity is important because prior research indicates that past health behaviors, such as recency of prior smoking and a history of obesity, can influence current health outcomes.[111] Furthermore, because transition probabilities vary depending on demographic characteristics, such as race and ethnicity, education, gender, and age, the model tracks outcomes by socioeconomic subgroups, and it allows for responses to policies to be subgroup-specific.

Like the previous literature that uses FEM and FAM, we model transitions of all health conditions, risk factors, disability, and mortality with a first-order Markov process. From a practical point of view, there are two main reasons why we prefer the assumption that risk factors and health conditions only from the prior period determine health transitions, instead of allowing for a higher order process. The first reason has to do with the ABC/CARE data: the available health follow-up data lacks multiple consecutive observations of health conditions for adults. Therefore, it is only possible to implement the simulation for the entering cohort as long as the transition matrix only depends on the previous period state vector (which corresponds to the health data in the ABC/CARE interviews). The second reason concerns the estimation: restricting the PSID sample to individuals present in three consecutive waves could introduce bias by leaving out those who have a higher probability of dropping out, such as individuals in poor health.[112]

Health conditions are treated as absorbing states, i.e., once a person has a disease she is assumed to have it forever. But this is not the case for risk factors: a person can transition

[111] Tong et al. (1996); Moore et al. (2008).

[112] For the transition of ADLs, the PSID data favors a specification with a higher order Markov assumption. However, the ABC/CARE data lack lagged values for variables related to ADLs. To implement a higher order Markov model in the simulation of ADLs for the ABC/CARE cohort we need to further develop a strategy to impute the lacking lagged values.

out of an obese state and back into it, a person can quit smoking and resume smoking. To discipline the rich dynamics of the model and based on evidence from the medical literature, a number of restrictions are placed on the way a disease or condition is associated with the transitions of other conditions (see Section F.3.3 for details).

Family formation models estimate transition probabilities between the following relationship statuses: single, cohabiting, married, separated/divorced, and widowed. We use multivariate regression models to estimate the number of children born separately for women and men. Economic models are developed to estimate labor force participation and employment status (possible states allowed by the model are: unemployed, out of the labor force, working part-time, or working full-time), and the take-up of government social insurance programs such as disability insurance. Transitions of labor earnings are projected outside of the simulation (Section C.3 describes the methodology). Because there is no information about assets in the age 30 ABC or CARE data, we do not simulate wealth transitions.[113]

To evaluate the performance of the estimated model, we validate it using various techniques, including comparing model results from early years with actual data available for later years.[114] Using these estimated transitions, we simulate outcomes for cohorts that have the initial characteristics of the ABC/CARE treatment and control groups at age 30. In each year, we use the health, family, and economic transition models to forecast obesity, smoking behavior, health status, economic status, family characteristics, disability, and mortality. We then use the models of health care spending to calculate medical costs for Medicare, other public sources excluding Medicare, and medical costs for private sources. We repeat the simulation each year until everyone in the cohort would have died.

[113]Additional details of the transition models are provided in Tysinger et al. (2015).
[114]Tysinger et al. (2015).

F.2 Data Sources

FAM uses data from ABC/CARE follow-up surveys to build the initial state of the cohort. The transition 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 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).

F.2.1 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 the transition 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.[115] The length of the PSID is a significant advantage, because we can include past health behaviors as explanatory variables for 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

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

or other long-term care facilities. To overcome this weakness, we pool the PSID sample with the HRS sample when estimating mortality models.

F.2.2 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 at least 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 to 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.

F.2.3 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.

F.2.4 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). Since NHIS does not include the institutionalized population, neither does MEPS; this implies 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.[116]

F.2.5 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 is described in Section F.3.1.

F.2.6 ABC/CARE

FAM uses ABC/CARE data to initialize the state of each ABC/CARE subject when they enter into the simulation. 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 have each subject's initial state in the simulation match 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 F.3.1.

F.3 Methods and Analysis

F.3.1 ABC/CARE Data Assumptions and Imputations

Marital status transitions and childbearing in FAM are affected by the subject's 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

[116] Section F.3.2 explains the estimation of the QALY model.

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

Many FAM transition 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.

All FAM transition 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.[117]

Most FAM models depend on smoking status. Employment status affects FAM transitions 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

[117] 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).

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 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 affects FAM transitions in health, functional status, employment, and smoking. The FAM transition models are estimated with BMI computed from self-reported height and weight in the PSID. The only BMI data in ABC/CARE come from height and weight measured during the health interview. This interview took place at roughly age 30 for CARE subjects, and at age 34 for ABC subjects. This poses two challenges. First, self-reported BMI can be biased by factors such as actual height and weight, gender, and race.[118] Second, it is possible that BMI could increase or decrease systematically in the years between the age-30 subject interview and the age-34 health interview.

To address the first BMI imputation challenge, 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 with height and weight, we apply the 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

[118] Cawley (2004).

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. The smoothed histogram of the distributions for the entire samples and the black subgroup in each data set appears reasonably close.

After imputing values of measured BMI for PSID respondents age 30–34, we turn to the second problem: 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 transition models estimated in PSID. 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 logically 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 were 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 transitions in benefits claiming, mortality, employment status, insurance category, and nursing home residency. FAM also transitions 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. At the start of the simulation, we generate Monte Carlo random draws according to these probabilities and use them to assign the corresponding counts.

When a subject claims DI benefits, it affects their FAM transitions in employment status, insurance category, and Medicare enrollment. DI claiming also affects medical costs. SSI claiming affects FAM transitions in employment status. Lastly, claiming Social Security retirement benefits affects FAM transitions in 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 F.3.2, 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. It is assumed that all subjects did not have their disease conditions in the previous period. In other words, for any subjects who reported a disease condition in the age-30 interview, their costs in the first simulation time step is estimated as if it were their incident year of the disease. Section F.3.2 describes the implications of this assumption.

F.3.2 FAM Models and Estimation

We develop models to estimate the determinants of transitions between health outcomes, labor market outcomes, educational attainment, and family formation, for individuals aged 25 and older. Additionally, we estimate transition probabilities by gender, race and ethnicity, and educational attainment as a function of individual characteristics (see below). Each transition model includes a subset of variables and relevant interactions from the following list: age, gender, race and ethnicity, education, parents' education, self-reported body mass index (BMI), smoking history, physical activity, binge drinking, lagged health conditions, asthma diagnosis before age 30, number of biological children, past earnings and work status, partnership status (single, cohabiting, married, separated/divorced, or widowed), disability status, and health insurance status. We consider three racial and ethnic groups (black non-Hispanic, white non-Hispanic, and Hispanic), and four educational groups (less than high school degree; high school graduate, including some college or associate's degree; college; and more than college).

The health transition models estimate the probability that a person transitions between health states, e.g. obesity or heart disease, as a function of current health status, demographic characteristics (including race, gender, age, and education), and risk factors (including weight, smoking status, physical activity, asthma, and number of births if female for BMI transitions), enabling us to age the cohorts. This mechanism to model health transitions accounts for the fact that certain health conditions increase the likelihood of comorbidities. We estimate a transition model for each of the following health conditions: heart disease, blood pressure, stroke, lung disease, diabetes, and cancer. Each disease model includes gender, race and ethnicity, and educational group as covariates. We select conditions that are prevalent in the U.S. and are characterized by significant disparities in outcomes across education, race and ethnicity, and income. The reason for this is that the incidence and progression of these conditions can potentially be reduced by preventive services, education policies, and modifications in health behaviors. These chronic conditions are treated as absorbing states, i.e., once the individual transitions into a chronic condition, the condition persists until death.

Additionally, we allow individuals to transition in and out of risk factors, such as smoking, binge drinking, and BMI. These transitions are estimated as a function of demographics, past health, and risk factors. In the estimation, changes in risk behaviors alter future health outcomes and risk factors (e.g., smoking cessation may impact changes in BMI, and continuing smoking may affect incidence of lung disease). We also transition mental health, approximated by the Kessler mental distress scale, which is one of the predictors of the medical costs models.

Because the PSID sample covers a broad age range, it is smaller than the HRS sample at older ages where mortality becomes more likely. Also, the PSID does not follow respondents into nursing homes. Therefore, the FAM mortality model is estimated using a pooled PSID and HRS sample. The mortality model includes these covariates: age, gender, race and ethnicity, education, disease conditions, ADL count, binge drinking and current smoking status. Similarly, a partner mortality model is estimated from the pooled PSID and HRS data for transitions into widowhood. The covariates in the partner mortality model are age, gender, race and ethnicity, and education. These covariates are characteristics of the respondents, not the partners who are facing mortality (FAM does not simulate the characteristics of partners).

Since the PSID does not follow respondents into nursing homes, the model for nursing home residency is estimated using only HRS data. It includes these covariates: age, gender, race and ethnicity, education, disease conditions, ADL and IADL counts, and widowhood.

The marriage transition model estimates transitions between partnership status (we distinguish between single, cohabiting, and married). The model is a function of demographics, past employment status, earnings, mother's education and number of children. There are also childbearing models that estimate new births for each gender. Childbearing is modeled as an ordered probit model that is a function of past health and birth history, demographics, education, past work status, and past partnership status.

The employment status model estimates the probability that a person transitions into different employment states (unemployed, out of the labor force, working part-time, or working full-time). This transition is a function of demographics, marital or partnership status, education, health status and behaviors, past earnings and benefits claiming.

The combination of transition models allows us to address the aspects of the life-cycle that are most relevant for the proposed analysis. To complete the analysis, there are also models to estimate QALYs, medical expenditure, and Social Security participation.

We compute a QALY model based on 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.[119] Later, a scoring system based on a U.S. sample was generated.[120] The PSID does not ask the appropriate questions for computing EQ-5D, but the MEPS does.

We forecast EQ-5D scores from the MEPS onto the PSID data using common measures between the MEPS and PSID.[121] We then forecast the EQ-5D scores for all PSID members running a linear regression using the variables that are transitioned in FAM (including ADL counts, IADL counts, and diseases). The microsimulation uses this linear regression to compute QALYs.

FAM has two versions of each medical cost model. 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 earnings. 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 include 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 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 tend to underestimate health care spending reported in the National Healthcare Expendi-

[119] Dolan (1997).
[120] Shaw et al. (2005).
[121] The main variables in this prediction are self-reported health and requiring help with ADLs.

tures Account (NHEA) data, due in part to underreporting of Medical costs in MEPS.

F.3.3 Transition Models

We denote $j_{i,0}$ to be the first age at which subject i is observed and j_{i,T_i} the last age at which he is observed. Hence, we observe outcomes at ages $j_i = j_{i,0}, \ldots, j_{i,T_i}$. We first start with discrete outcomes which are absorbing states (e.g. disease diagnostic, mortality, and benefit claiming). We record the hazard as $h_{i,j_i,m} = 1$ if the individual outcome m has occurred as of age j_i. We assume the individual-specific component of the hazard can be decomposed into time-invariant and time-variant parts. The time-invariant part is composed of the effect of observed characteristics, x_i, that are constant over the entire life course and initial conditions, $h_{i,j_0,-m}$, (where $-m$ denotes outcomes other than the outcome m) that are determined before the first age in which each subject is observed.

The time-variant part is the effect of previously diagnosed outcomes, $h_{i,j_i-1,-m}$, on the hazard for m.[122] We assume an index of the form $z_{j_i,m} = x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + h_{i,j_0,-m}\psi_m$. Hence, the latent component of the hazard is modeled as

$$h^*_{i,j_i,m} = x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + h_{i,j_0,-m}\psi_m + a_{m,j_i} + \varepsilon_{i,j_i,m}, \tag{31}$$

where $m = 1, \ldots, M$, $j_i = j_{i0}, \ldots, j_{i,T_i}$, and $i = 1, \ldots, N$. The term $\varepsilon_{i,j_i,m}$ is a time-variant shock specific to age j_i. We assume that this last shock is normally distributed and uncorrelated across diseases. We approximate a_{m,j_i} with an age spline with knots at ages 35, 45, 55, 65, and 75. This simplification is made for computational reasons since the joint estimation with unrestricted age fixed effects for each condition would imply a large number

[122]With some abuse of notation, $j_i - 1$ denotes the previous age at which the subject was observed.

of parameters. The absorbing outcome, conditional on being at risk, is defined as

$$h_{i,j_i,m} = \max\{I(h^*_{i,j_i,m} > 0), h_{i,j_i-1,m}\}.$$

The occurrence of mortality censors observation of other outcomes in a current year.

A number of restrictions are placed on the way feedback is allowed in the model. Our microsimulation model starts the health forecasts at age 30, with the information on observed characteristics available at this age. We restrict it to the individuals for whom we have information from the health follow-up. This allows us to account for components that are crucial for forecasting health outcomes, such as the body mass index (BMI). In sum, the models forecast the probability of being in any of the states in the horizontal axis of Table F.1 at age $a+1$ based on the state at age a, which is described by the vertical axis of the table. The crosses indicate if the estimation of the probability of being in a state at age $a+1$ considers the relevant state at age a. Absorbing states are an exception. For example, heart disease at age a does not enter in the estimation of transitions for heart disease at age $a+1$ because it is an absorbing state: once a person has heart disease, she carries it through the rest of her life. The same is true for chronic or permanent conditions such as hypertension, having a stroke, etc.

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

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

Note: This table illustrates how health outcomes at age a forecast health outcomes at age $a+1$. The crosses indicate if we use the age a outcome to forecast the age $a+1$ outcome. DI Claim: claims of benefits to individuals and their families if "insured"—i.e., worked long enough and paid Social Security taxes; SS claim: claim of social security retirement benefits; SSI Claim: claims of stipends for low-income people who are older than 65 years old, blind, or disabled.

We have four other types of outcomes:

1. Binary outcomes that are not absorbing states, such as starting to smoke. We specify latent indices as in Equation (31) for these outcomes as well but where the lag-dependent outcome also appears as an independent variable. This allows for state-dependence.

2. Ordered outcomes, which are also modeled as in Equation (31) recognizing that the observation rule is a function of unknown thresholds ς_m. Similar to binary outcomes, we allow for state-dependence by including the lagged outcome on the right-hand side.

3. Continuous outcomes are modeled with linear models. An example continuous outcome is the transitions in log(BMI). We allow for state-dependence by including the lagged outcome on the right-hand side.

4. Categorical models, but without an ordering, are considered. For example, an individual can transition to being unemployed, out of the labor force, or working (either part- or full-time). In situations like this, we utilize a multinomial logit model, including the lagged outcome on the right-hand side.

In total, we have M outcomes. The parameters $\theta_1 = \left(\{\beta_m, \gamma_m, \psi_m, \varsigma_m\}_{m=1}^{M}\right)$, can be estimated by maximum likelihood. Given the normality distribution assumption on the time-variant unobservable, the joint probability of all time-intervals until failure, right-censoring, or death conditional on the initial conditions, $h_{i,j_0,-m}$, is the product of normal univariate probabilities. Since these sequences, conditional on initial conditions, are also independent across diseases, the joint probability over all disease-specific sequences is simply the product of those probabilities.

For a given subject observed from the initial age, j_{i0}, to the last age, j_{T_i}, the probability of the observed health history is (omitting the conditioning on covariates for notational simplicity)

$$l_i^{-0}(\theta; h_{i,j_{i0}}) = \left[\prod_{m=1}^{M-1} \prod_{j=j_{i1}}^{j_{T_i}} P_{ij,m}(\theta)^{(1-h_{ij-1,m})(1-h_{ij,M})}\right] \times \left[\prod_{j=j_{i1}}^{j_{T_i}} P_{ij,M}(\theta)\right]$$

We use the -0 superscript to make explicit the conditioning on $\mathbf{h}_{i,j_{i0}} = (h_{i,j_{i0},0}, \ldots, h_{i,j_{i0},M})'$. We have limited information on outcomes prior to this age. The likelihood is a product of M terms with the mth term containing only $(\beta_m, \gamma_m, \psi_m, \varsigma_m)$. This allows the estimation to be done separately for each outcome.

F.3.3.1 Specification Tests for the First-order Markov Assumption in FAM

The FAM model assumes a vector first-order Markov process. To make this clear, take the example of heart disease. As previously explained and stated in Table F.1, we forecast "heart disease" at age $a+1$ based on hypertension and diabetes at age a, as well as other risk factors and health behaviors (smoking, BMI, and physical activity).[123] In this model, heart disease is assumed to be an absorbing state. That is, if an individual suffers heart disease at age a, her probability of having heart disease at age $a+1$ equals one.

In our empirical analysis, we estimate the transition probabilities for each disease using a Probit model, and the variables indicated in Table F.1 as well as background characteristics not affected by treatment. We test the first-order Markov process assumption using a likelihood test ratio comparing a model based on first-order lags (a first-order Markov process) to forecast the disease of interest (null) and a model based on second-order lags (second-order Markov process). Table F.2 show the results from these tests.

[123] The diseases that help predicting each other are based on research and advice of clinicians, as explained and justified in Goldman et al. (2015).

Table F.2: Tests Comparing First-Order and Second Markov Processes for Disease Transition 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 comparing first- and second-order Markov processes to forecast diseases at age $a+1$, based on diseases and other health and demographic conditions at age a. The null hypotheses is the first-order Markov process, and the alternative is the second order.

The interpretation of the results in Table F.2 is the following. The heart disease of Column "1st vs. 2nd" tests the null that first-order lags suffice to construct the transition of heart disease from age a to age $a+1$. We test the first-order Markov assumption with respect to other diseases. Limited support on smoking, BMI, and physical activity does not allow us to test the first-order Markov assumption with respect to these variables.[124] We cannot reject that a 1st-order Markov model is enough to forecast heart disease at age $a+1$, if compared to a second-order Markov model. This test has two degrees of freedom because we add second-order lags for hypertension and diabetes in the alternative model (see Table F.1).[125] Unfortunately, limited support for the other components of the transition models do not allow us to make credible tests.[126]

An alternative test for a first-order Markov process is the following: (i) use a linear probability model approximation to a first-order Markov process and the variables in Table F.1 to forecast the disease of interest using multiple versions of the forecasting models; (ii) calculate the correlation of the residuals with higher order lagged variables. Tables F.3 to F.5

[124] We lack observations in the auxiliary sample for which information is available for this conditions and the diseases of interest for multiple lags.

[125] In work available on request, we find that higher order Markov processes better predictors of heart disease at age $a+1$, relative to a first-order Markov model, for most outcomes. However, for stroke, there is some evidence of higher order dependence.

[126] We would need to drop thousands of observations in the auxiliary samples and perform the tests using very selected samples.

show these correlations. The results of these tests very strongly support a first-order Markov assumption.

To illustrate how to read these tests consider Table F.3. This table reports simple first-order Pearson correlations with the indicated variable. The row presents the residuals from forecasting heart disease at age $a + 1$ using different orders of lags, for the diseases listed in Table F.1 (for heart disease, these are hypertension and diabetes). The estimated correlations are low.

Table F.3: Tests for Linear Probability Forecasts of Heart Disease at $a+1$

	Residuals of Forecast of Heart Disease at age $a+1$									
Correlations with	Predictor 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 $a+1$ and the predictors indicated in each column.

Table F.4: Tests for Linear Probability Forecasts of Hypertension at $a+1$

	Residuals of Forecast of Hypertension at age $a+1$				
Correlations with	Predictor 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 $a+1$ and the predictors indicated in each column.

Table F.5: Tests for Linear Probability Forecasts of Stroke at $a+1$

	Residuals of Forecast of Stroke at age $a+1$							
Correlations with	Predictor 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 $a+1$ and the predictors indicated in each column.

F.3.3.2 Further Details on Specific Transition Models

This section describes the modeling strategy for particular outcomes.

Employment Status

Ultimately, we aim to simulate whether an individual is unemployed, out of the labor force, working part-time, or working full-time at time t. We treat the estimation of this as a two-stage process. In the first stage, we forecast whether the individual is unemployed, out of the labor force, or working for pay using a multinomial logit model. Then, conditional on working for pay, we estimate if the individual is working part- or full-time using a probit model.

Relationship Status

We are interested in three relationship statuses: single, cohabiting, and married. In each case, we treat the transition from time t to time $t + 1$ as a two-stage process. In the first stage, we estimate if the individual will remain in his current status. In the second stage, we estimate which of the two other states the individual will transition to, conditional on leaving his current state.

Childbearing

We estimate the number of children born in two-year periods separately for females and males. We model this using an ordered probit with three categories: no new births, one birth, and two births. Based on the PSID data, we found the exclusion of three or more births in a two-year period to be appropriate.

F.3.4 FAM simulation

A simulation of the model starts by loading the entering cohort, generated from the ABC/CARE data. Missing values are imputed with the imputation models described in section F.3.1. To this entering cohort, the model applies the transition models for mortality, health, working status, family structure, wealth, and benefit claiming, estimated from PSID, with Monte Carlo decisions to calculate the new states of the population. The simulated financial outcomes are in 2014 USD.

To match the biennial structure of the PSID data used to estimate the transition models, the simulation proceeds in two-year increments.[127] Once the new states have been determined, the cross-sectional models for medical costs and QALYs are applied. Computation of medical costs includes the people who died to account for end-of-life costs. The simulation ends when all simulated ABC/CARE subjects are deceased.[128]

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 ran the simulation twice—once for the ABC/CARE subjects entering in odd years and again for the ABC/CARE subjects entering in even years.

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.

The FAM mortality model is assumed to represent mortality in 2009. The estimated mortality probabilities are reduced in simulated future years to represent improvements in mortality

[127] 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).

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

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.[129]

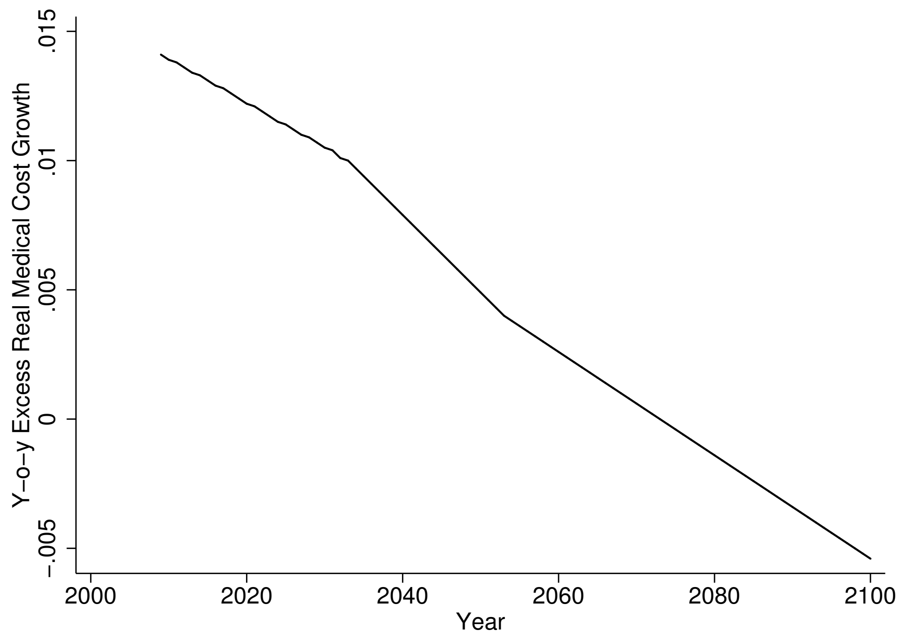
F.3.5 Medical Costs Before Age 30 Interview

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 combined 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

[129] 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 F.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.

Figure F.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.

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 forecasted 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 of siblings as predictors.

Medical costs for ages 12 to 14 follow a strategy similar to the 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 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 F.6 summarizes individual and family characteristics used to forecast medical expenditure models for each age.

Table F.6: 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 $\geq$ 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.

F.4 Validation

To evaluate the performance of the full version of the FAM model, we validate it using various techniques.

F.4.1 Cross-validation

The cross-validation exercise randomly samples half of the PSID respondent IDs for use in estimating the transition models. The respondents not used for estimation, but who were present in the PSID sample in 1999, are then simulated from 1999 through 2013. Demographic, health, and economic outcomes are compared between the simulated (FAM) and actual (PSID) populations.

It is worth noting how the composition of the population changes in this exercise: In 1999, the sample represents those 25 and older. Since we follow a fixed cohort, the age of the population will increase to 39 and older in 2013. This has consequences for some measures in later years where the eligible population shrinks.

Demographics Mortality and demographic measures are presented in Tables F.7 and F.8. Mortality incidence is comparable between the simulated and observed populations. Demographic characteristics do not differ between the two.

Health Outcomes Binary health outcomes are presented in Table F.9. FAM underestimates the prevalence of ADL and IADL limitations compared to the cross-validation sample. Binary outcomes, like cancer, diabetes, heart disease, and stroke do not differ. FAM underforecasts hypertension and lung disease compared to the cross-validation sample.

Health Risk Factors Risk factors are presented in Table F.10. BMI is not statistically different between the two samples. Current smoking is not statistically different, but more individuals in the cross-validation sample report being former smokers.

On the whole, the cross-validation exercise is reassuring. There are aspects that will be explored and improved upon in the future.

Table F.7: Crossvalidation of simulated 1999 cohort: Mortality in 2001, 2007, and 2013

	2001			2007			2013		
Outcome	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value
Died	0.014	0.018	0.013	0.019	0.023	0.133	0.027	0.025	0.514

Note: This cross-validation exercise randomly samples half of the PSID respondent IDs for use in estimating the transition models. The respondents not used for estimation, but who were present in the PSID sample in 1999, are then simulated from 1999 through 2013. This table compares outcomes between the simulated (FAM) and actual (PSID) populations.

Table F.8: Crossvalidation of simulated 1999 cohort: Demographic outcomes in 2001, 2007, and 2013

	2001			2007			2013		
Outcome	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value
Age on July 1st	49.120	49.022	0.668	53.079	53.379	0.197	56.759	57.959	0.000
Black	0.094	0.093	0.742	0.093	0.088	0.228	0.094	0.092	0.710
Hispanic	0.075	0.077	0.604	0.080	0.084	0.342	0.084	0.093	0.058
Male	0.457	0.460	0.643	0.455	0.463	0.296	0.451	0.458	0.443

Note: This cross-validation exercise randomly samples half of the PSID respondent IDs for use in estimating the transition models. The respondents not used for estimation, but who were present in the PSID sample in 1999, are then simulated from 1999 through 2013. This table compares outcomes between the simulated (FAM) and actual (PSID) populations.

Table F.9: Crossvalidation of simulated 1999 cohort: Binary health outcomes in 2001, 2007, and 2013

	2001			2007			2013		
Outcome	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value
Any ADLs	0.079	0.064	0.000	0.102	0.126	0.000	0.124	0.142	0.004
Any IADLs	0.098	0.113	0.001	0.111	0.130	0.000	0.129	0.170	0.000
Cancer	0.041	0.036	0.034	0.071	0.059	0.002	0.101	0.103	0.667
Diabetes	0.066	0.062	0.233	0.100	0.092	0.090	0.136	0.145	0.108
Heart Disease	0.098	0.106	0.095	0.135	0.152	0.002	0.175	0.173	0.792
Hypertension	0.185	0.174	0.067	0.287	0.272	0.030	0.385	0.410	0.004
Lung Disease	0.038	0.039	0.606	0.062	0.058	0.190	0.084	0.091	0.174
Stroke	0.019	0.021	0.412	0.027	0.034	0.015	0.037	0.049	0.001

Note: This cross-validation exercise randomly samples half of the PSID respondent IDs for use in estimating the transition models. The respondents not used for estimation, but who were present in the PSID sample in 1999, are then simulated from 1999 through 2013. This table compares outcomes between the simulated (FAM) and actual (PSID) populations.

Table F.10: Crossvalidation of simulated 1999 cohort: Risk factor outcomes in 2001, 2007, and 2013

	2001			2007			2013		
Outcome	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value	FAM mean	PSID mean	p-value
BMI	26.690	26.723	0.671	27.379	27.397	0.850	27.848	27.639	0.039
Current smoker	0.197	0.200	0.521	0.162	0.167	0.461	0.136	0.146	0.108
Ever smoked	0.481	0.513	0.000	0.479	0.525	0.000	0.472	0.531	0.000

Note: This cross-validation exercise randomly samples half of the PSID respondent IDs for use in estimating the transition models. The respondents not used for estimation, but who were present in the PSID sample in 1999, are then simulated from 1999 through 2013. This table compares outcomes between the simulated (FAM) and actual (PSID) populations.

F.4.2 External Corroboration

Finally, we compare FAM population forecasts to Census forecasts of the US population. Here, we focus on the full PSID population (25 and older) and those 65 and older. For this exercise, we begin the simulation in 2009 and simulate the full population through 2049. Population projections are compared to the 2012 Census projections for years 2012 through 2049. See results in Table F.11. By 2049, FAM forecasts for 25 and older remain within 2% of Census forecasts.

Table F.11: Population forecasts: Census compared to FAM

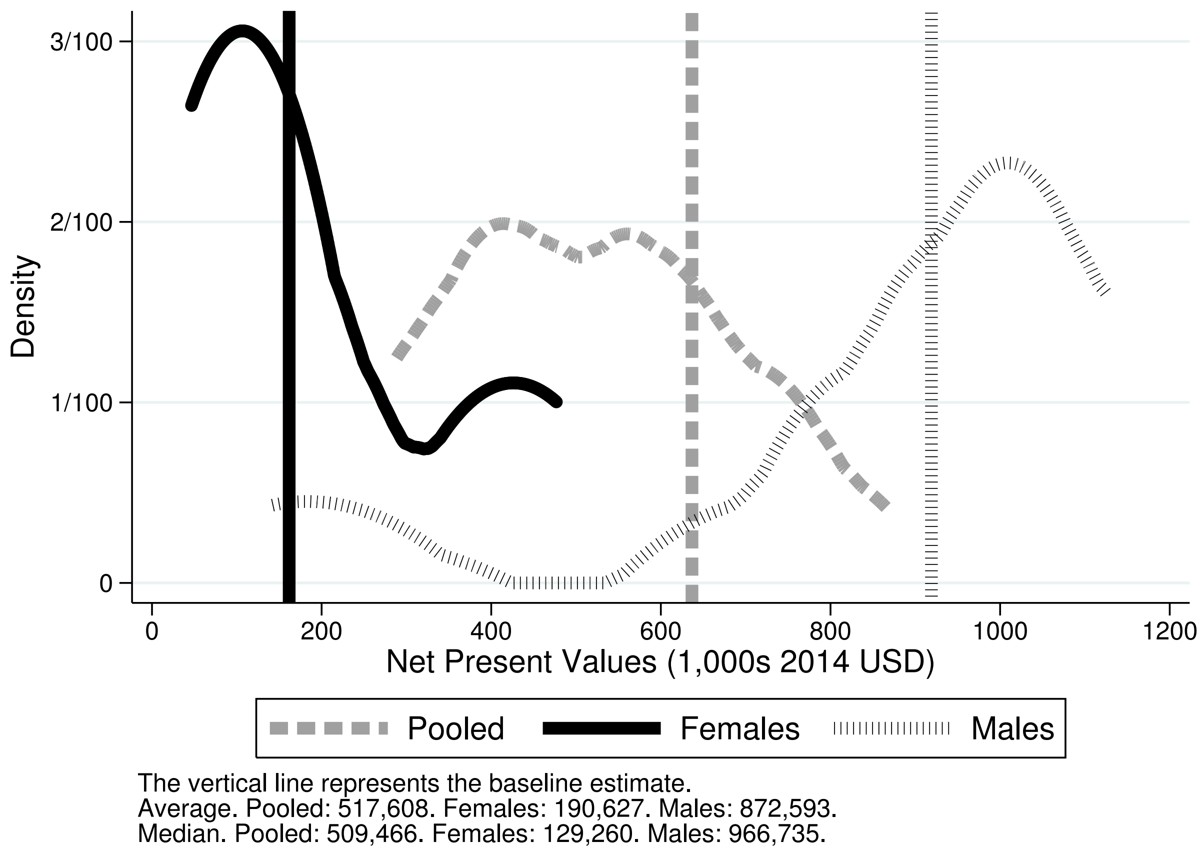
Year	Census 25+	FAM 25+	Census 65+	FAM 65+
2009	202.1	202.0	39.6	39.4
2011	206.6	206.5	41.4	41.0
2013	211.0	210.5	44.7	43.9
2015	215.9	215.1	47.7	47.1
2017	220.9	219.7	50.8	50.1
2019	225.5	224.1	54.2	52.6
2021	229.8	227.8	57.7	55.5
2023	233.9	231.6	61.4	57.9
2025	238.0	235.7	65.1	61.6
2027	241.9	239.6	68.4	65.2
2029	245.7	243.5	71.4	68.8
2031	249.3	247.2	73.8	71.6
2033	252.9	250.5	75.5	72.8
2035	256.0	253.4	77.3	75.1
2037	259.2	256.2	78.8	75.6
2039	262.6	259.3	79.4	76.1
2041	265.8	262.6	79.9	76.1
2043	269.0	265.8	80.4	77.7
2045	272.2	269.1	81.3	78.9
2047	275.3	272.2	82.2	79.7
2049	278.4	275.2	83.2	80.3

Note: Comparison between Census population projections and a FAM simulation of a full population starting in 2009 through 2049.

G Sensitivity Analysis

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 a plot of all net present value estimations across specifications in Figure G.1. This is analogous to Figure 3. We 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.

Figure G.1: Distribution of Net Present Value Estimates



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 1.

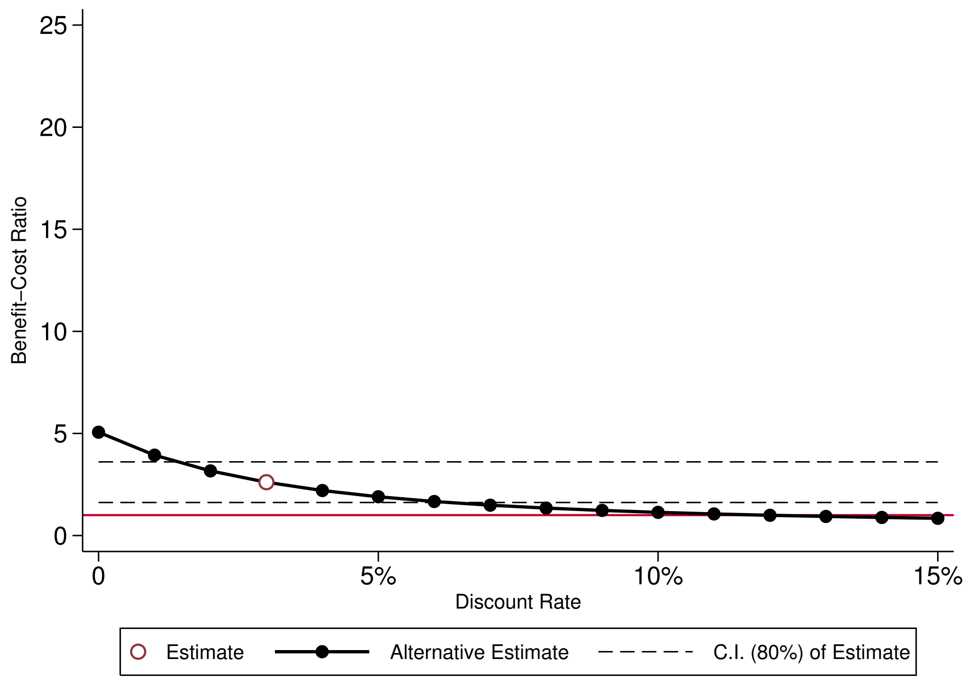
G.1 Varying the Discount Rate

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

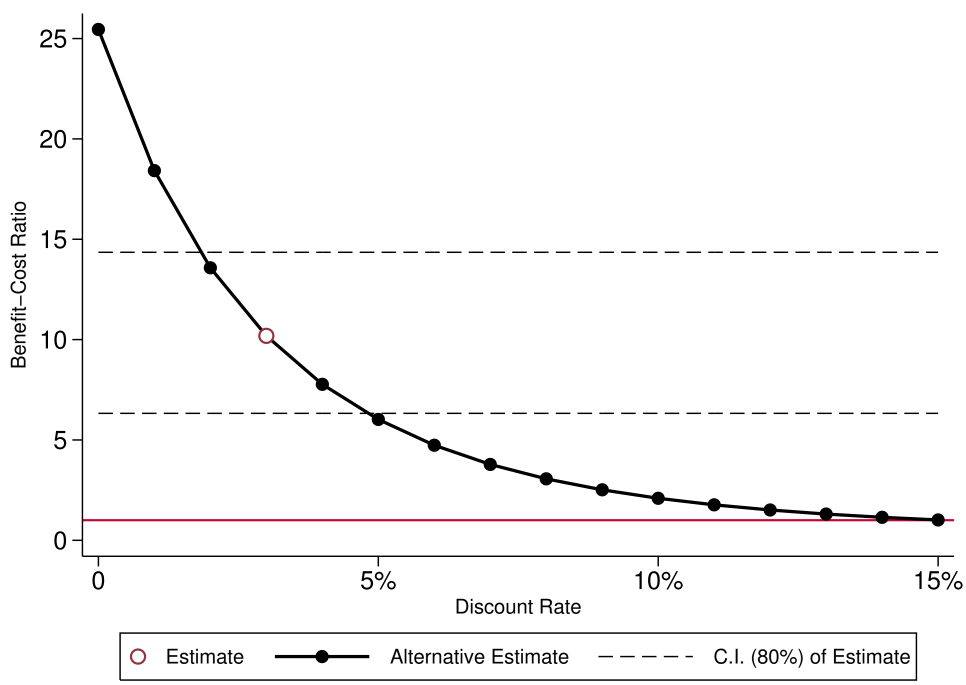
Figure G.2 displays how the benefit/cost ratio changes as we adjust the rate at which we discount the cash flows. We find that for males, the benefits of the programs exceed the costs for discount rates as high as 15%. We also find that the benefit/cost ratios for discount rates of 2–12% remain within 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 G.2 indicate that our estimates are sensitive to our choice of the discount rate, especially for males.

Figure G.2: 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 are means of the empirical bootstrap distribution. The 80% confidence intervals are obtained by taking the 10^{th} and 90^{th} quantiles of the bootstrap distribution.

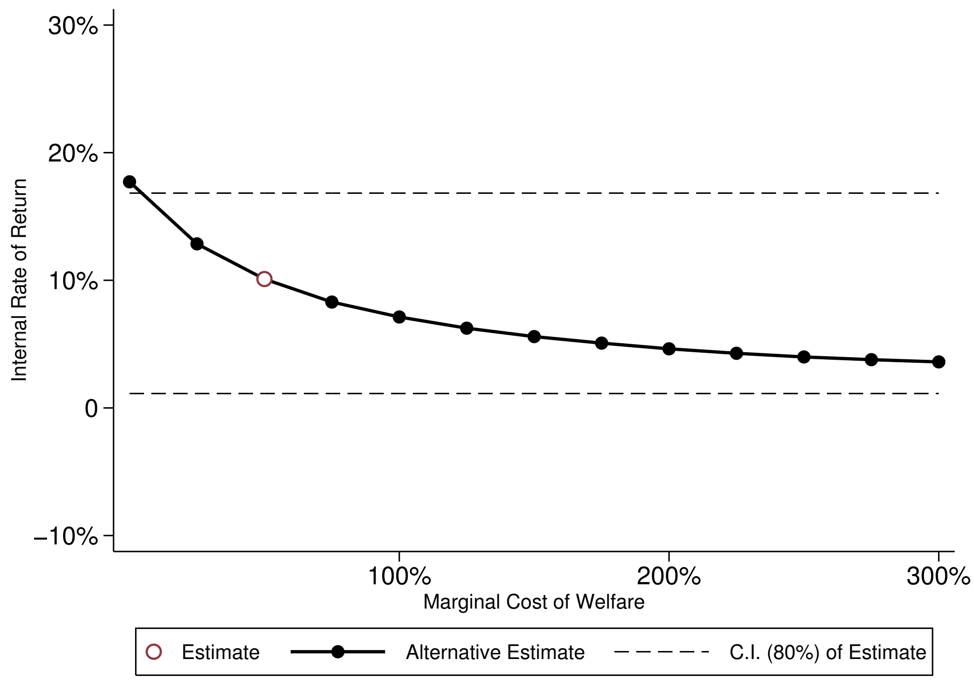
G.2 Varying Deadweight Loss

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.

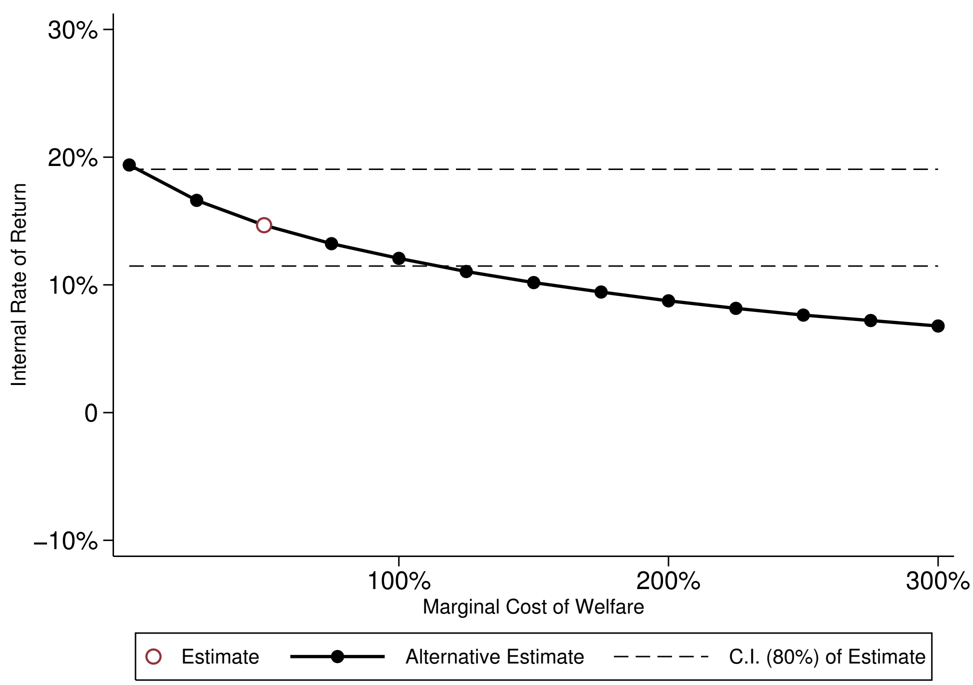
Figure G.3 shows how the IRR changes as we adjust the marginal cost of welfare. For both males and females, we find the IRRs to be most sensitive at the lower marginal costs. The IRRs for both sexes steadily decline as we increase the marginal cost of welfare to \$3 for every dollar of tax revenue. This 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 G.3: Internal Rate of Return vs. Deadweight Loss

(a) Females



(b) Males

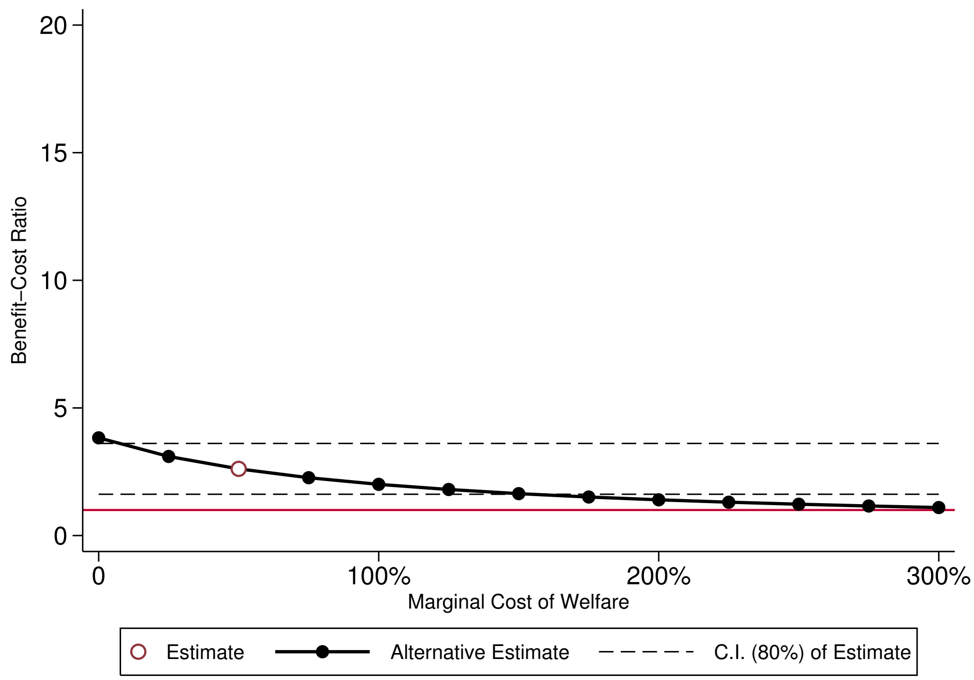


Note: These graphs display how the internal rate of return changes for females and males as we vary the marginal cost of welfare. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the marginal cost of welfare. The estimates 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 10^{th} and 90^{th} quantiles of the bootstrap distribution.

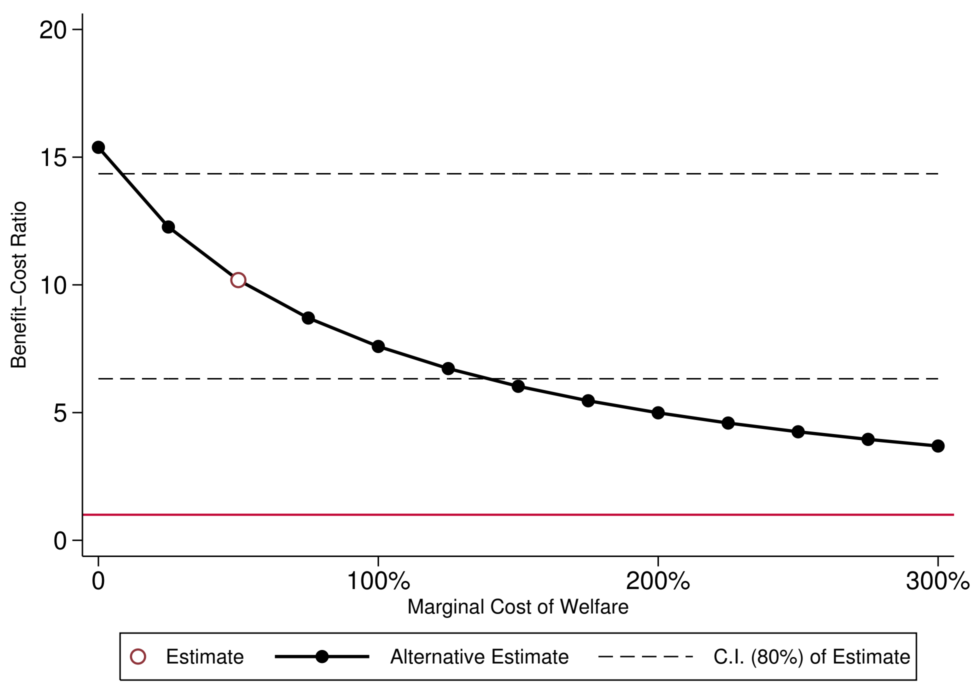
Figure G.4 illustrates how the benefit/cost ratio changes as we vary the marginal cost of welfare. For females we see that the benefits exceed the costs (both are discounted at a rate of 4%) even when the marginal cost of welfare is assumed to equal $3 for every dollar of tax revenue. We see a similar relationship hold for men, however, the ratio is significantly higher at every marginal cost.

Figure G.4: 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 10^{th} and 90^{th} quantiles of the bootstrap distribution.

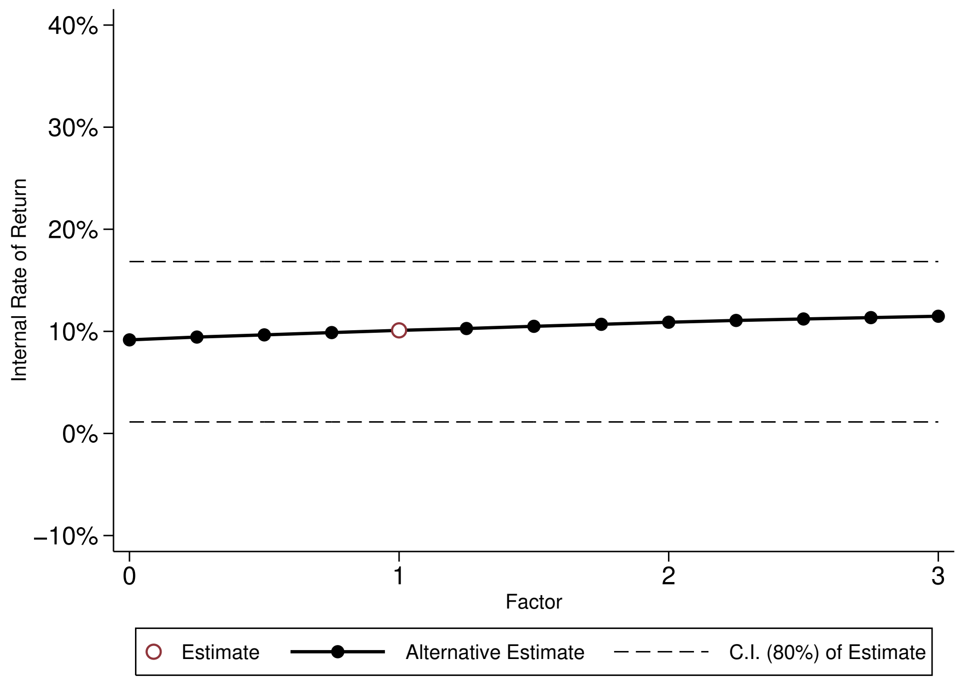
G.3 Varying Component Magnitudes

Below 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, e.g., 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).

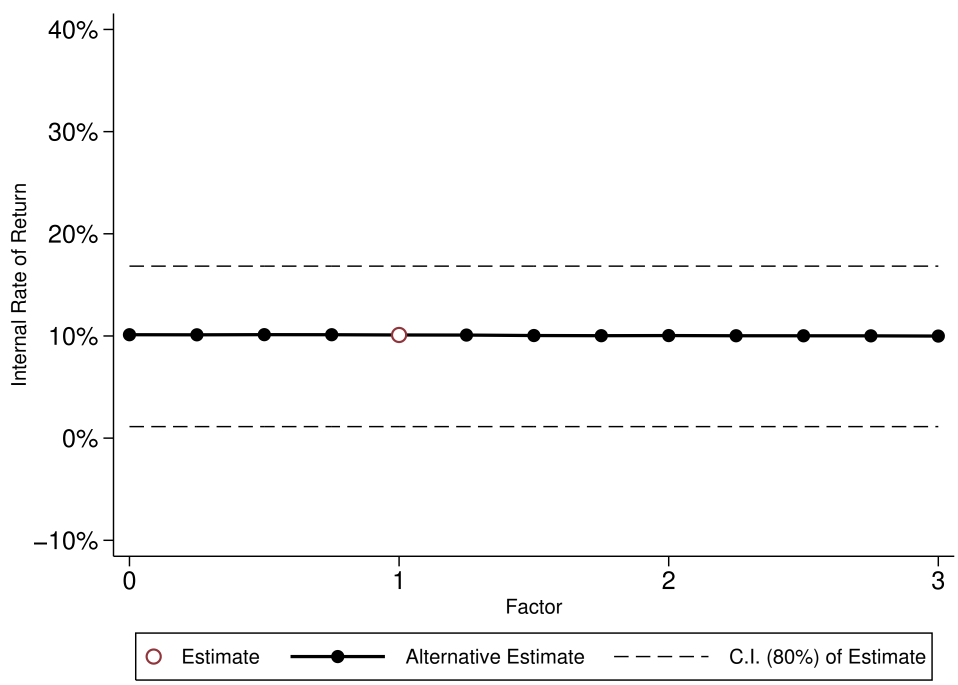
Figure G.5 displays how the IRR for females changes as we multiply each component by a factor between 0 and 3. We find that the IRR is stable across different levels of public-transfer income, QALYs, health costs, and criminal costs. Parental income has the biggest effect on the IRR, and is one of the 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 G.5: Internal Rate of Return vs. Components, Females

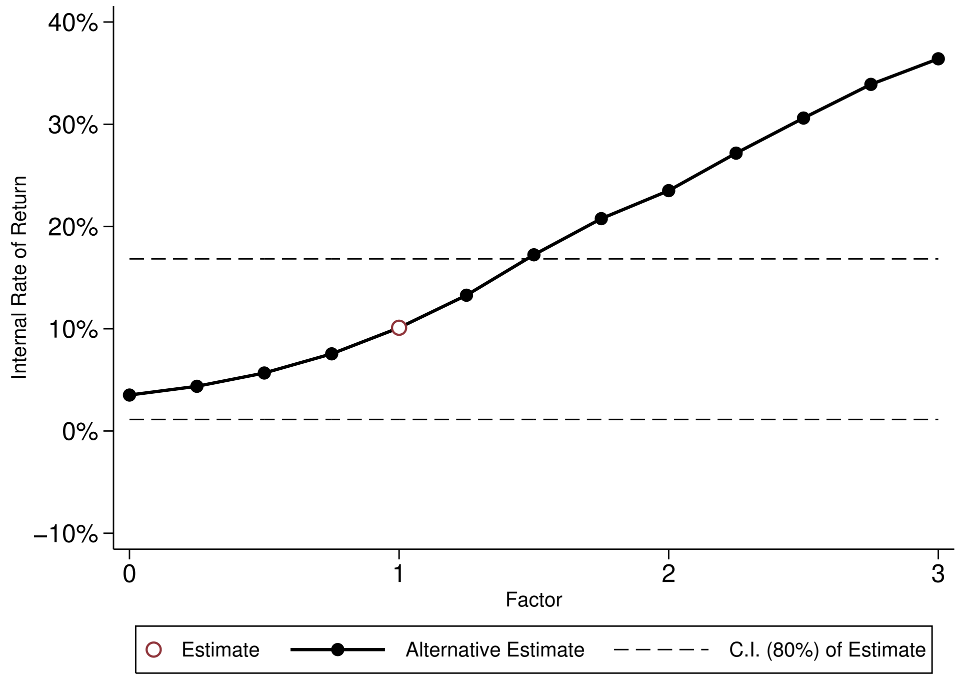
(a) Labor Income



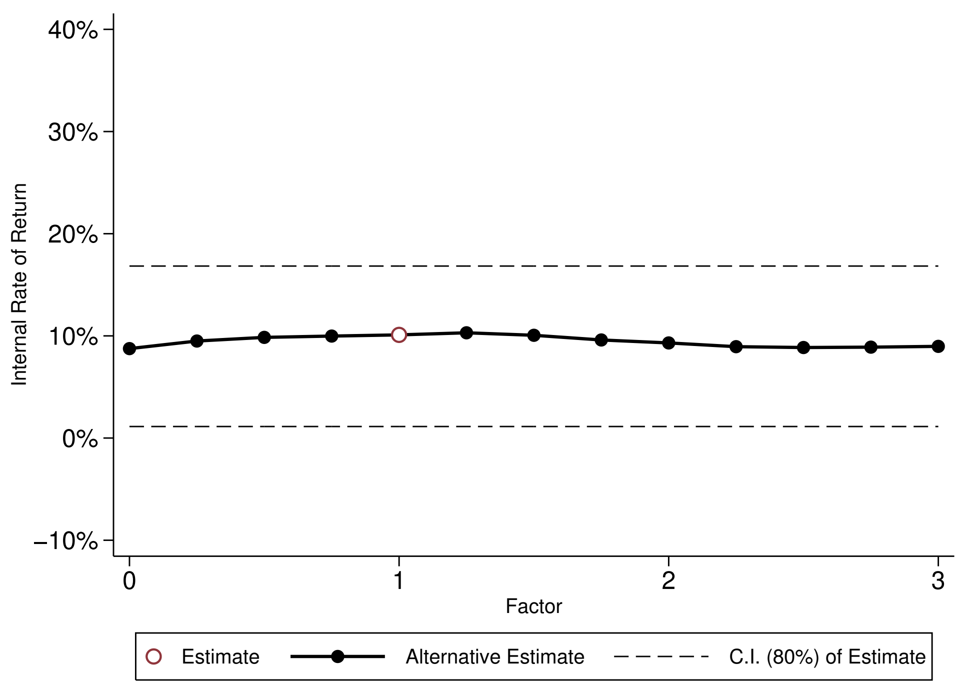
(b) Public-Transfer Income



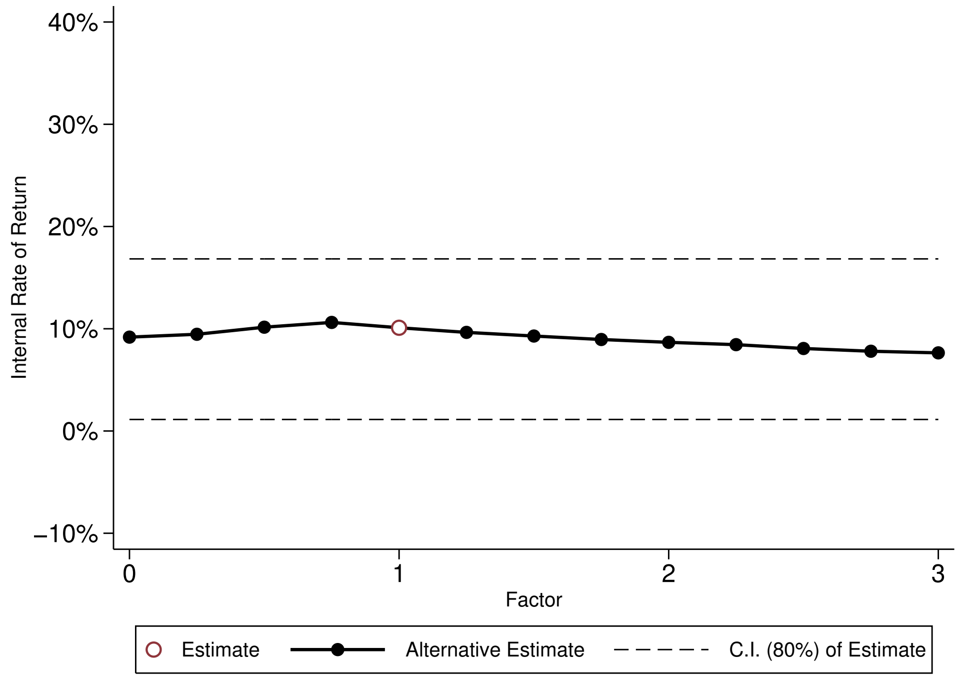
(c) Parental Income



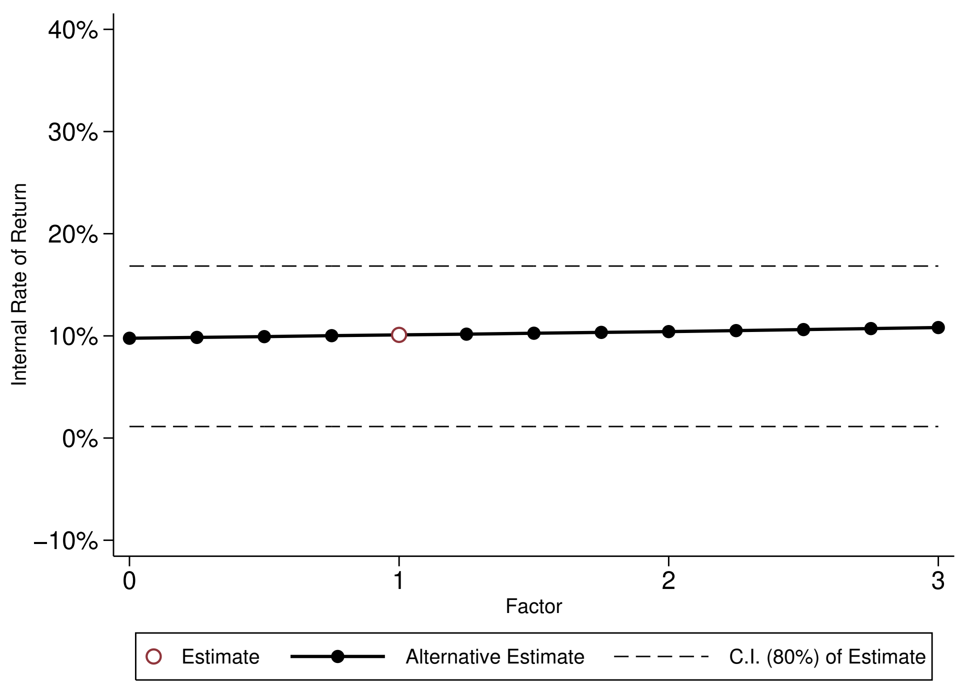
(d) Quality-Adjusted Life Years



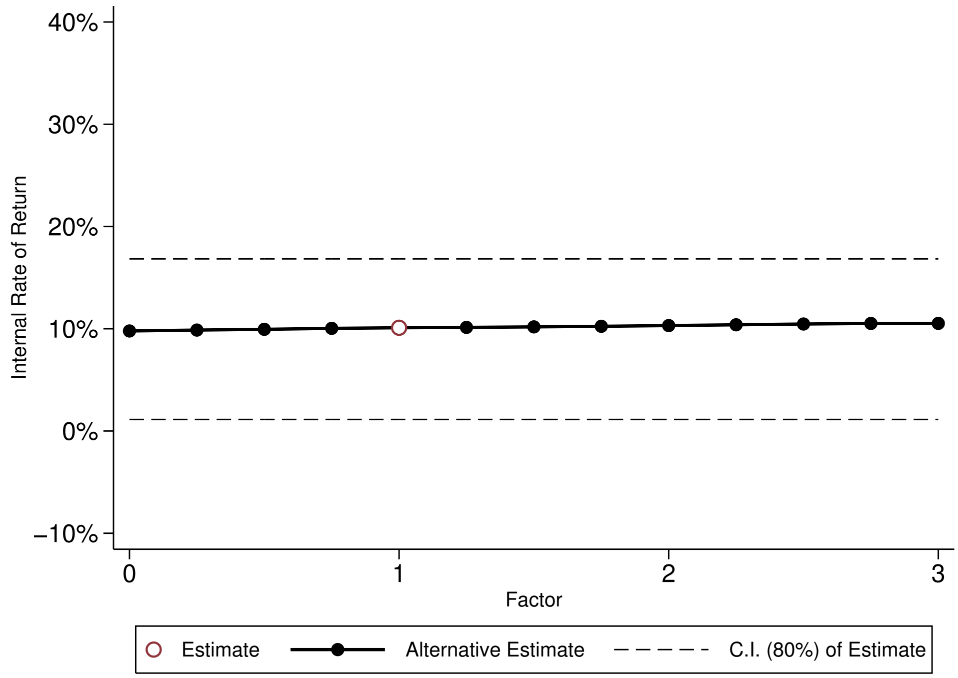
(e) Health Costs



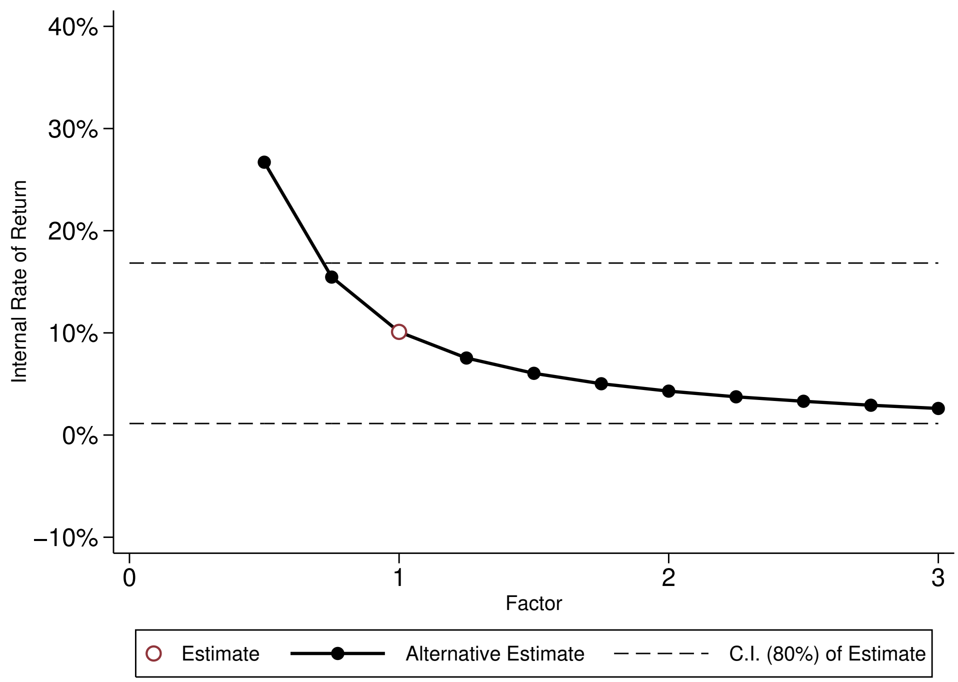
(f) Education Costs



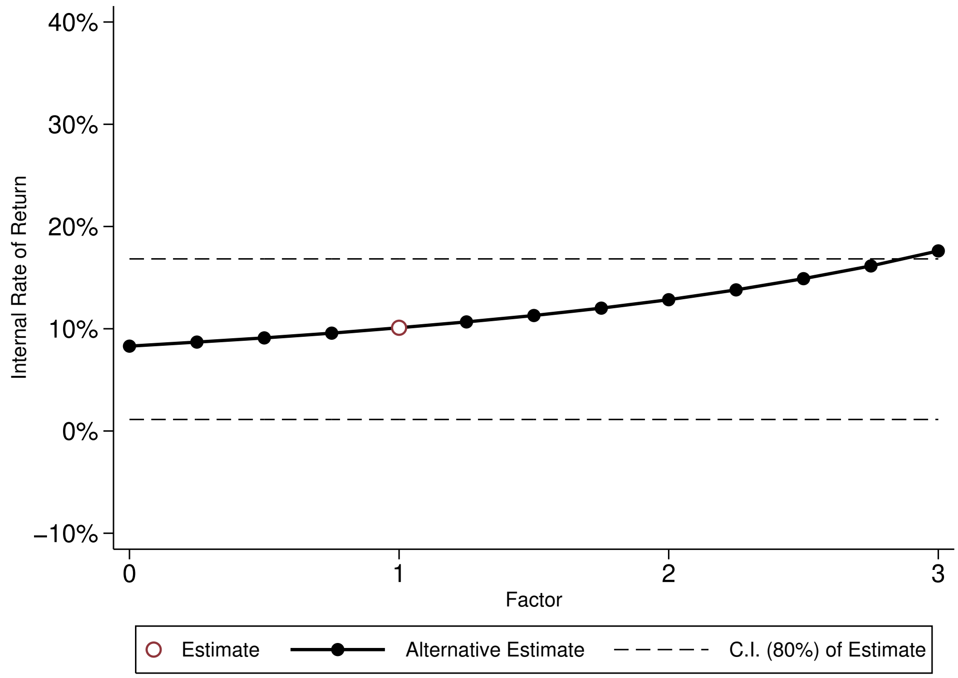
(g) Crime Costs



(h) Program Costs



(i) Control Substitution Costs



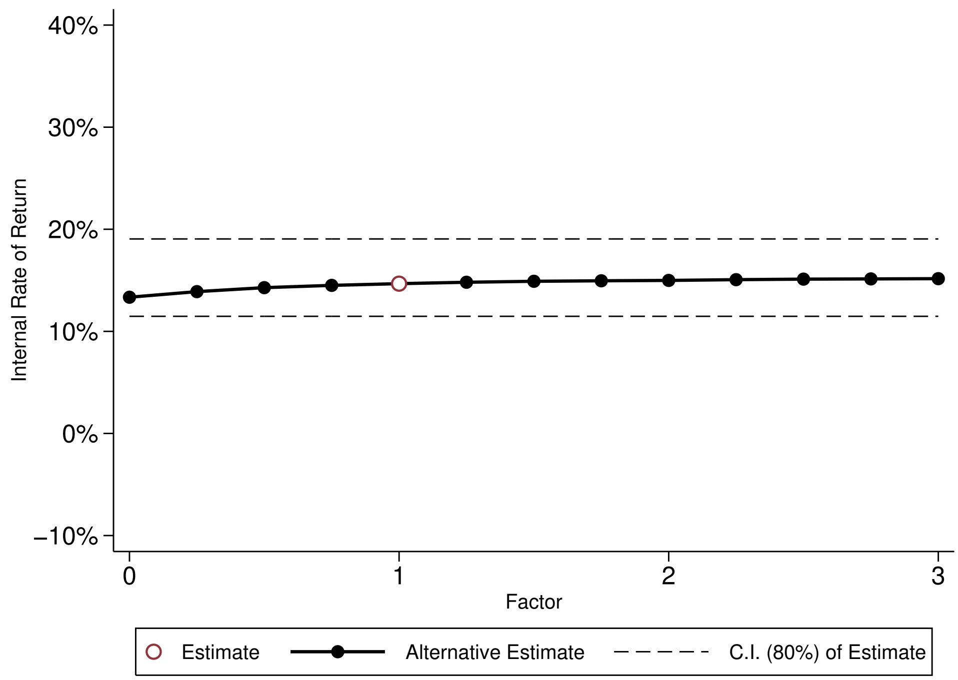
Note: These graphs display how the internal rate of return changes for females as we multiply each component by a factor from 0 to 3. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the magnitude of each component. The 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 10^{th} and 90^{th} quantiles of the bootstrap distribution.

Figure G.6 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 product 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 G.6

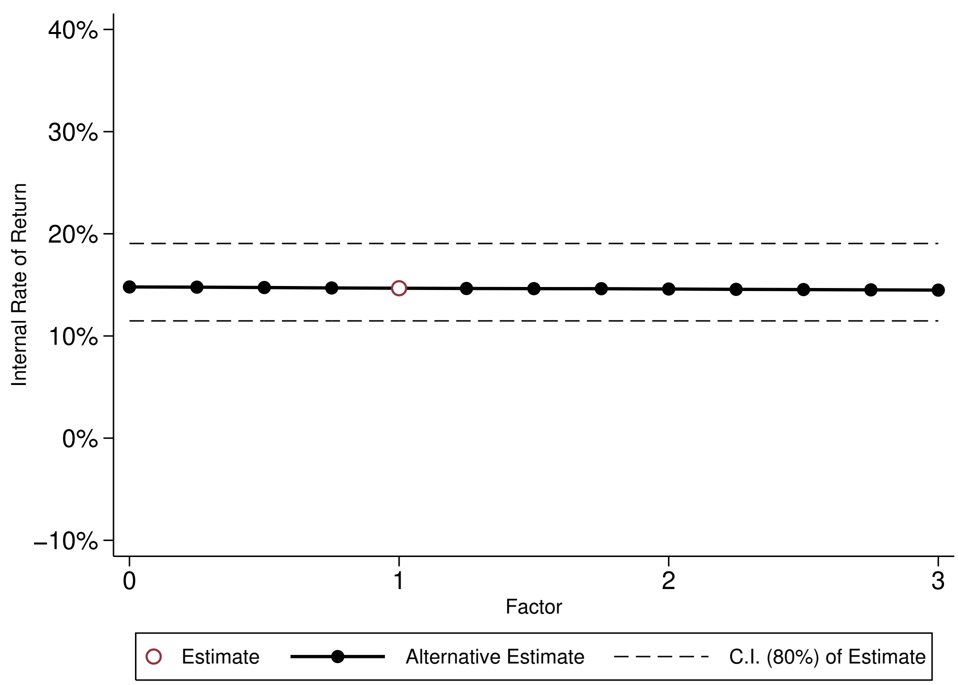
that the IRR for males rises as we multiply the benefits stemming from this component by increasingly large factors. Finally, the IRR increases for males as we increase the magnitude of the criminal cost component. This is not surprising because the reduction in the costs of crimes is the largest benefit of ABC/CARE for males.

Figure G.6: Internal Rate of Return vs. Components, Males

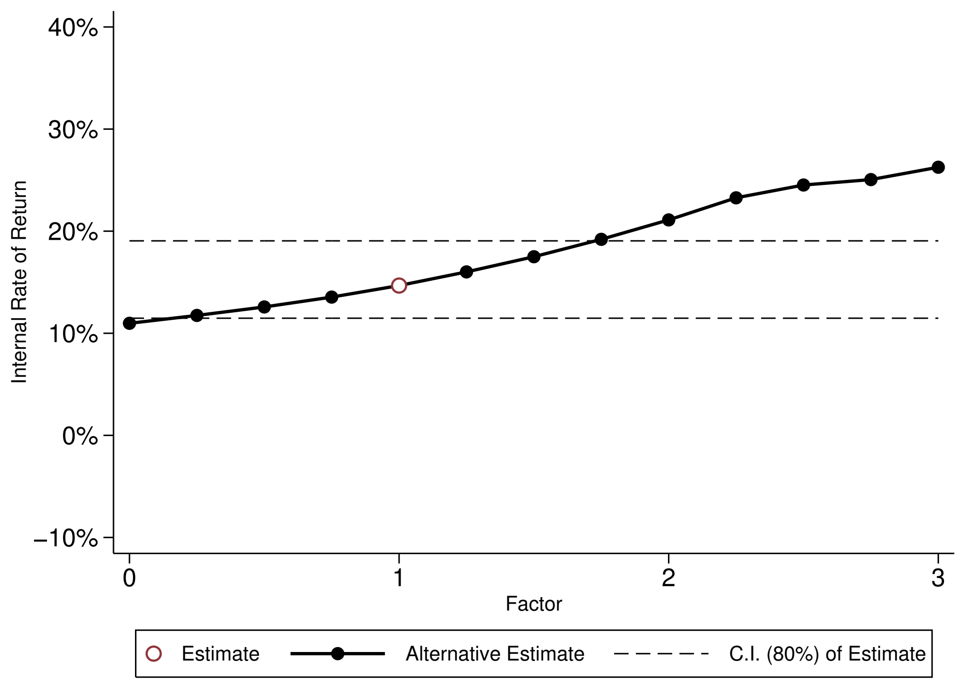
(a) Labor Income



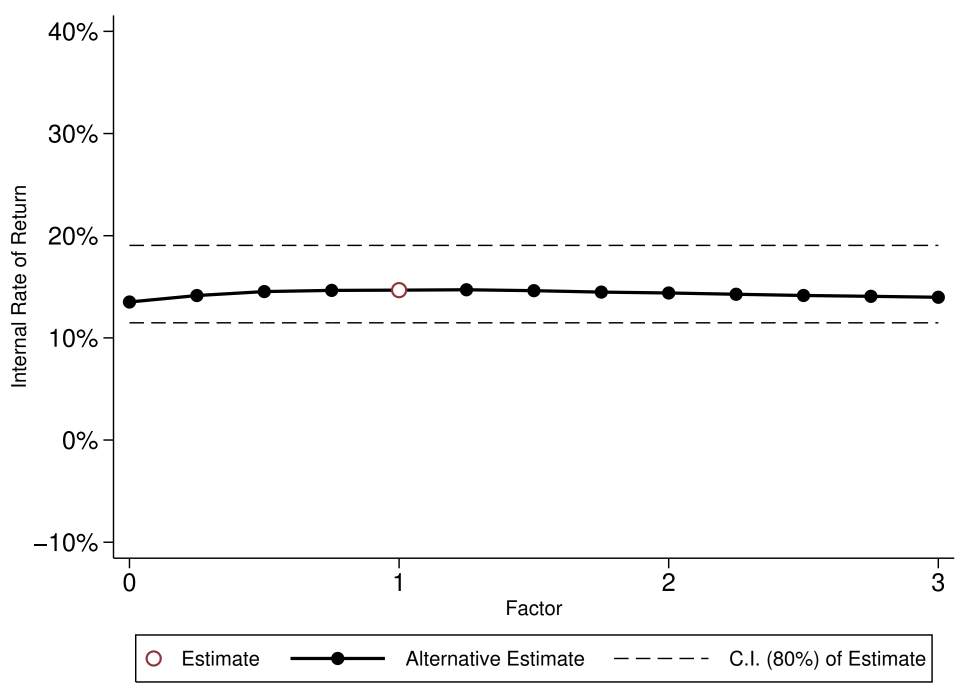
(b) Public-Transfer Income



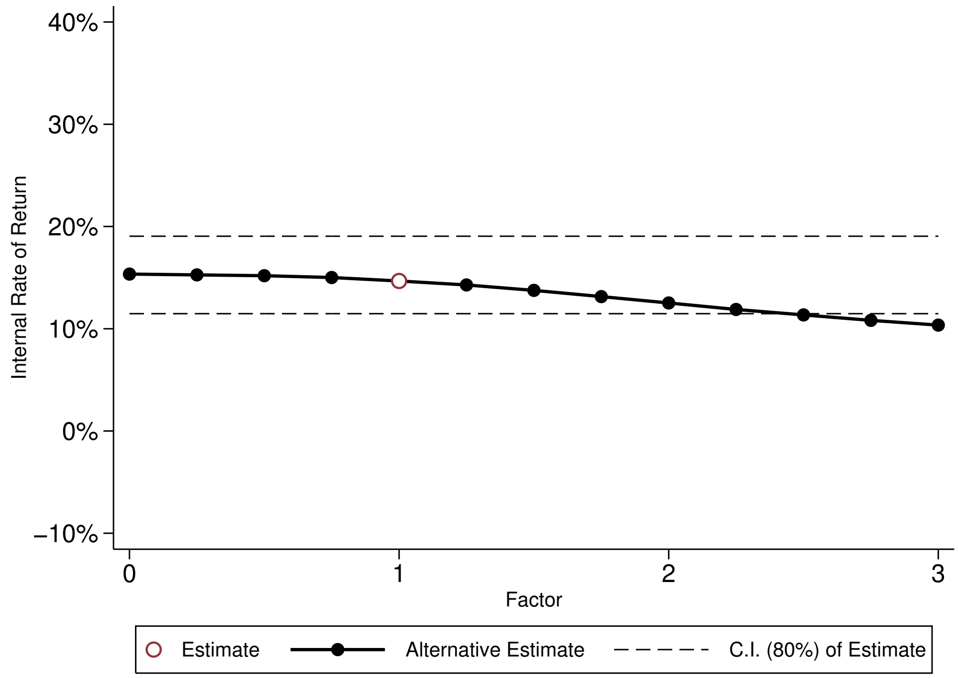
(c) Parental Income



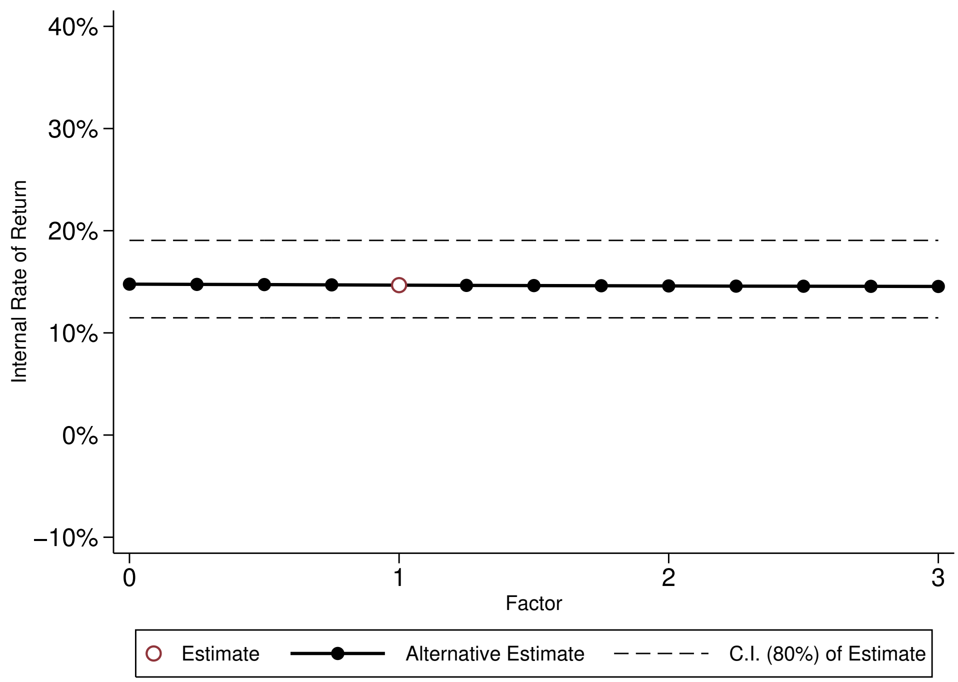
(d) Quality-Adjusted Life Years



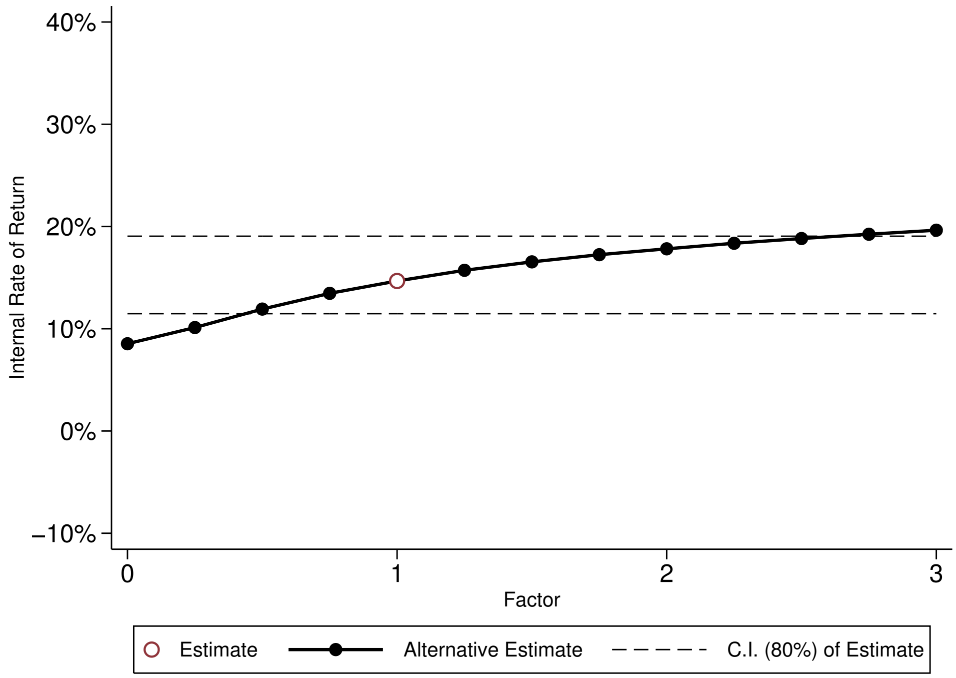
(e) Health Costs



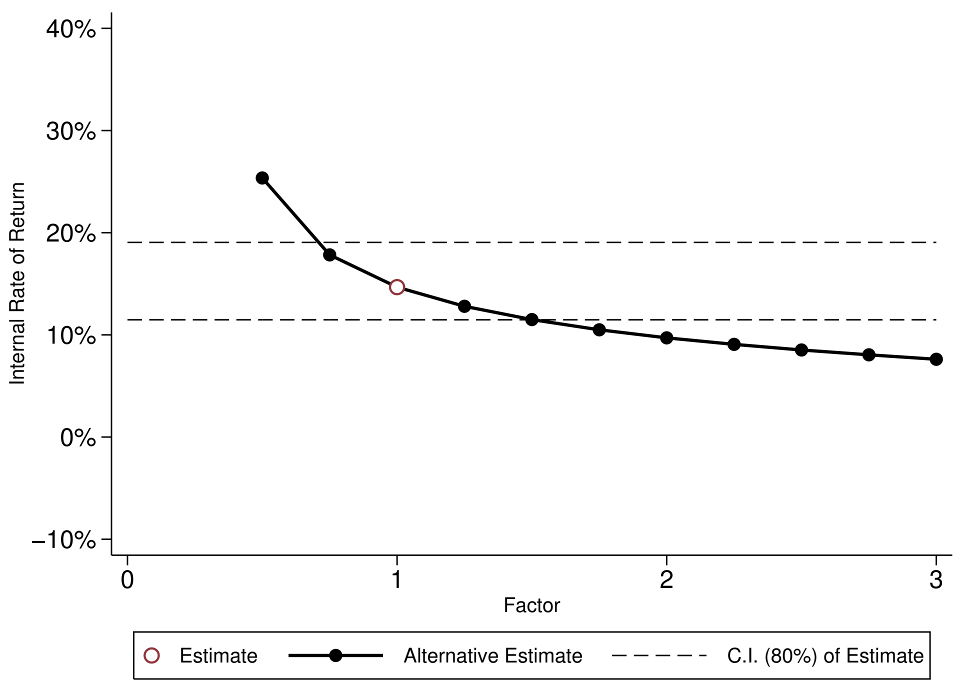
(f) Education Costs



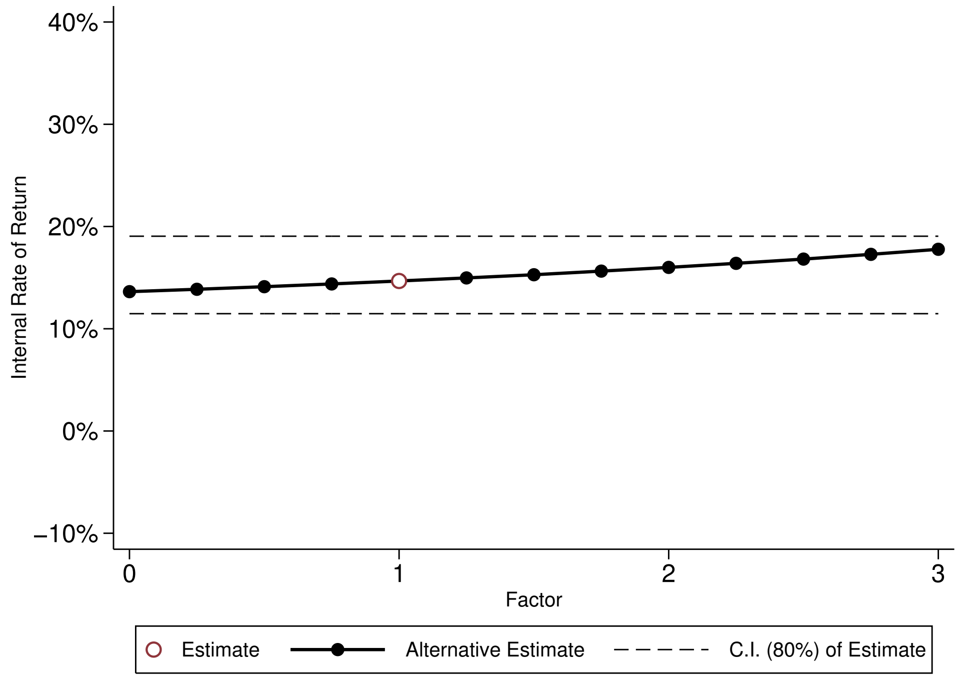
(g) Crime Costs



(h) Program Costs



(i) Control Substitution Costs

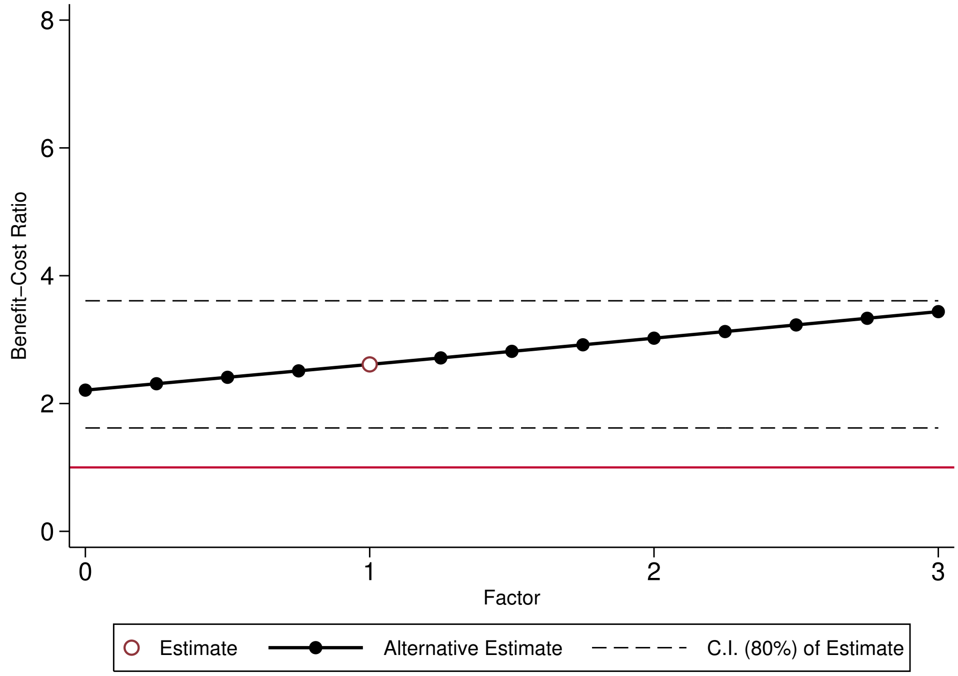


Note: These graphs display how the internal rate of return changes for males as we multiply each component by a factor from 0 to 3. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by varying the magnitude of each component. The estimates 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 10^{th} and 90^{th} quantiles of the bootstrap distribution.

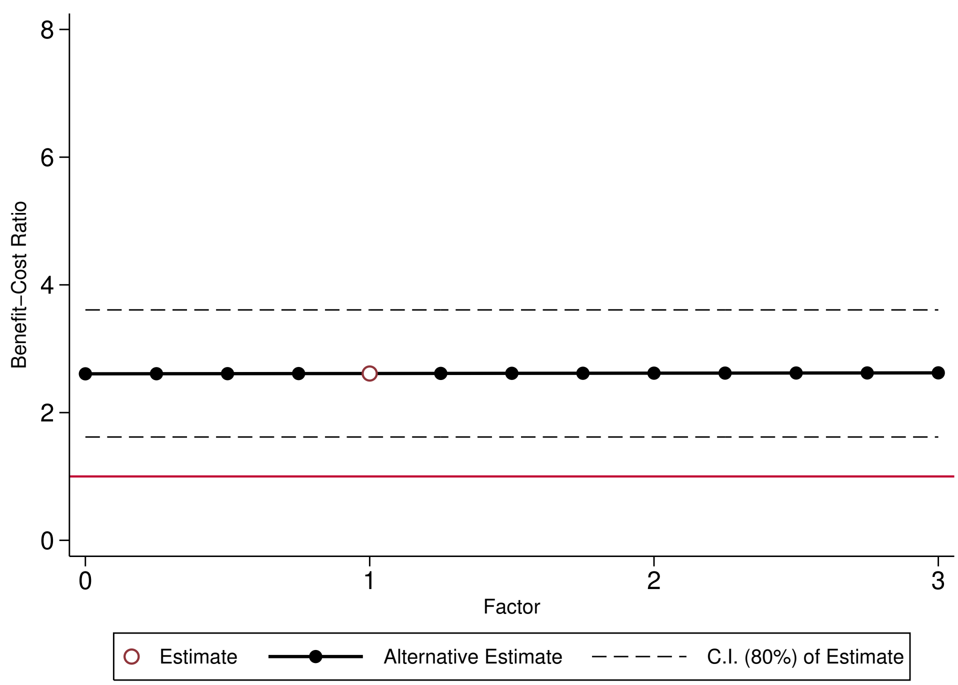
Figure G.7 shows how the benefit/cost ratio changes for females as we multiply each component of the benefits and costs by a factor between 0 and 3. We observe that the ratio is generally insensitive to changes in all the components, except parental income, labor income, education costs, crime costs, 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 G.7: Benefit/cost Ratio vs. Components, Females

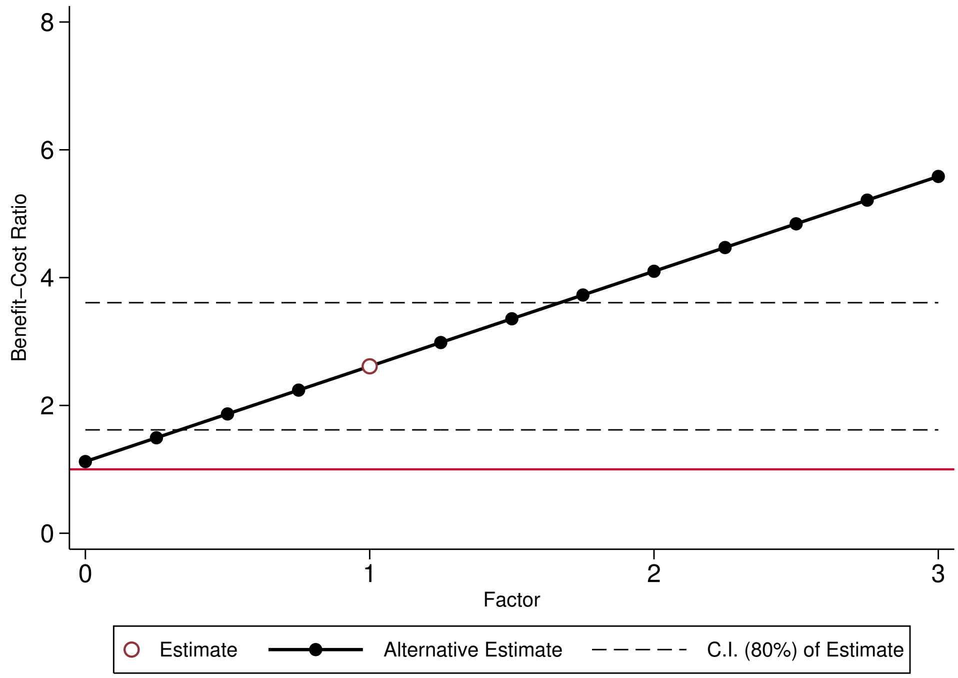
(a) Labor Income



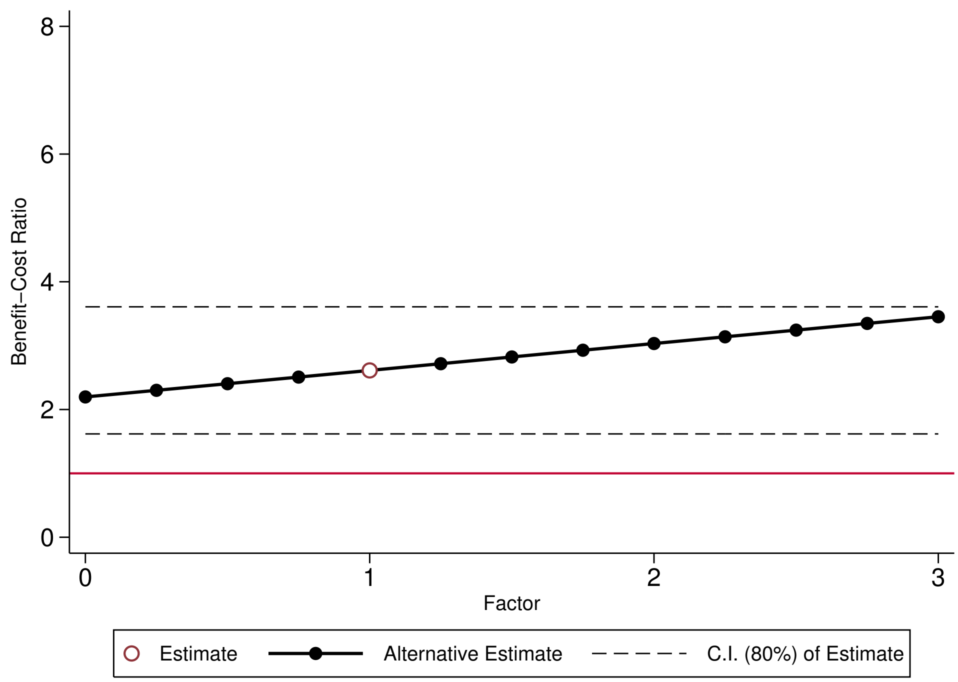
(b) Public-Transfer Income



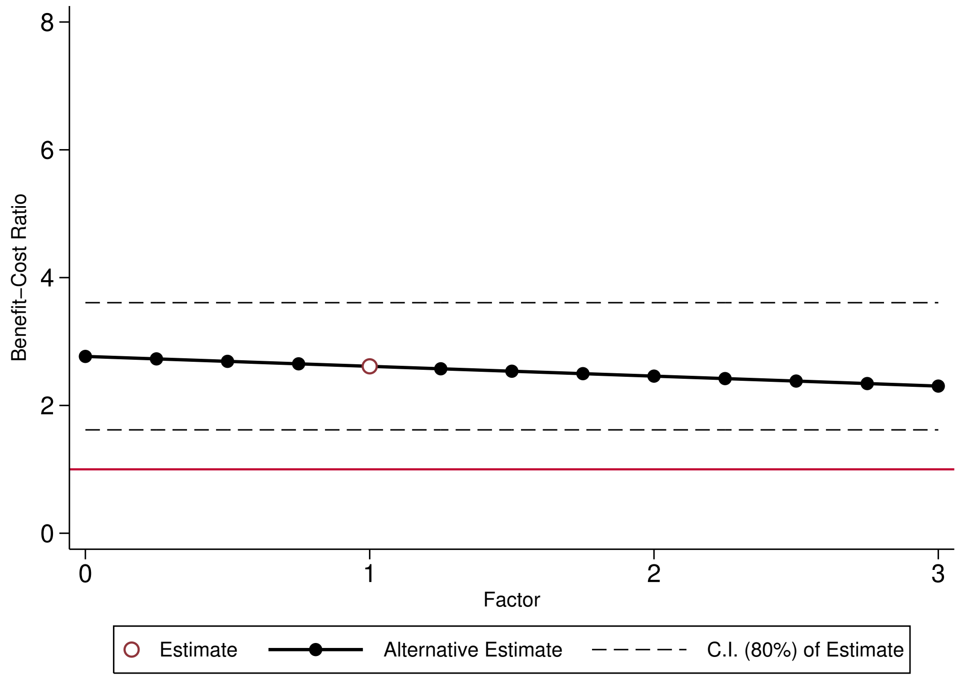
(c) Parental Income



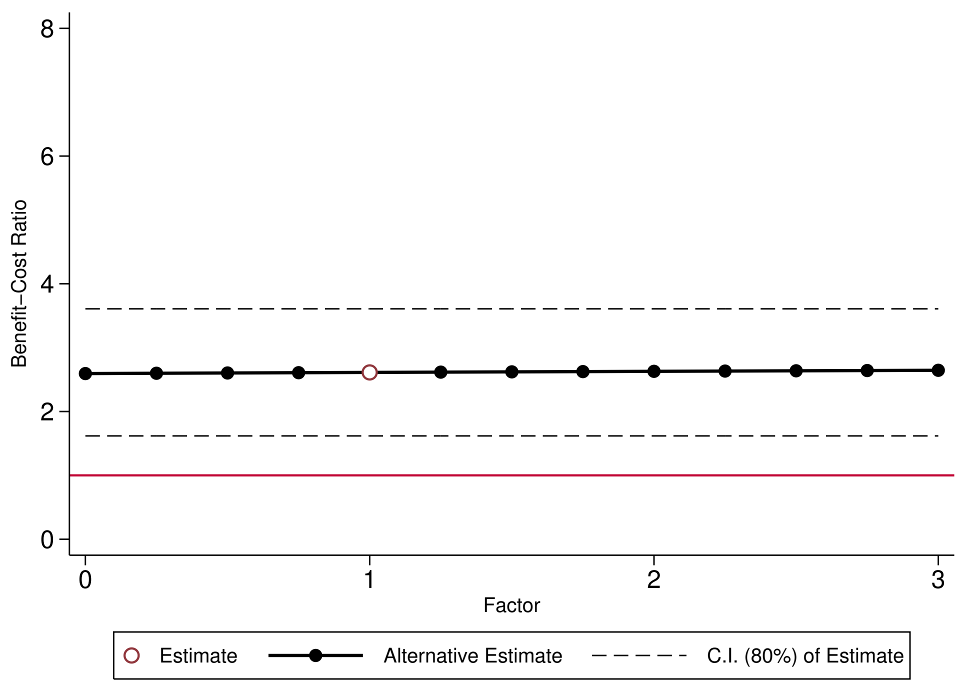
(d) Quality-Adjusted Life Years



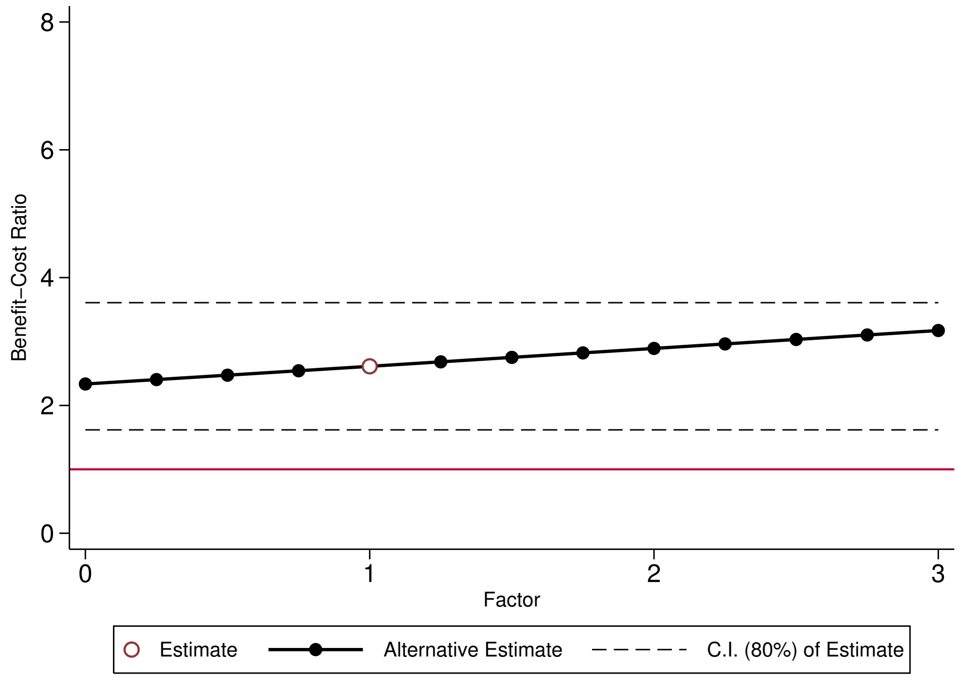
(e) Health Costs



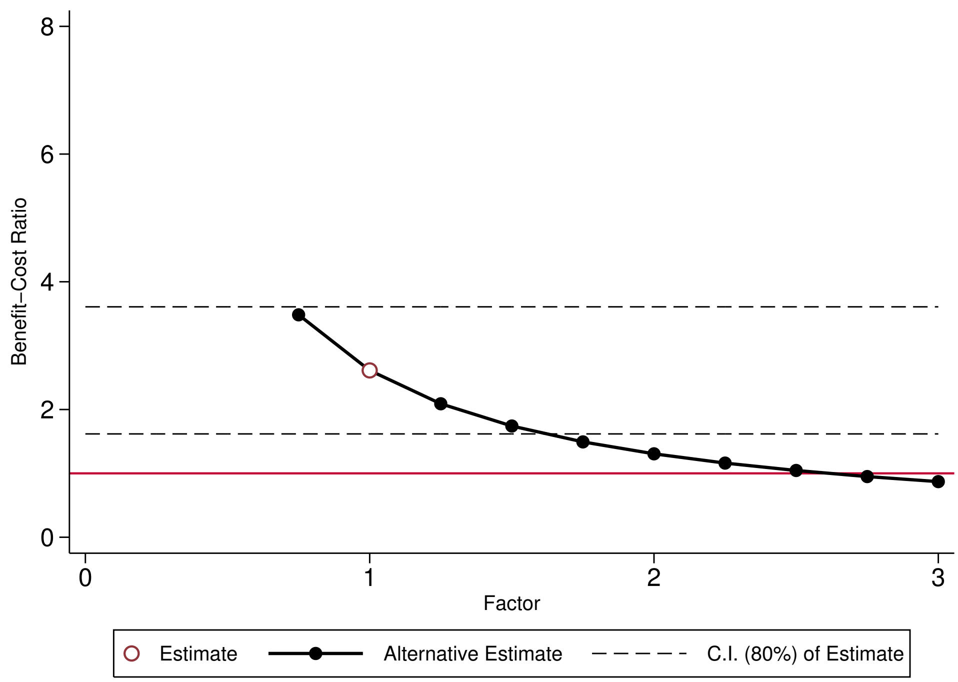
(f) Education Costs



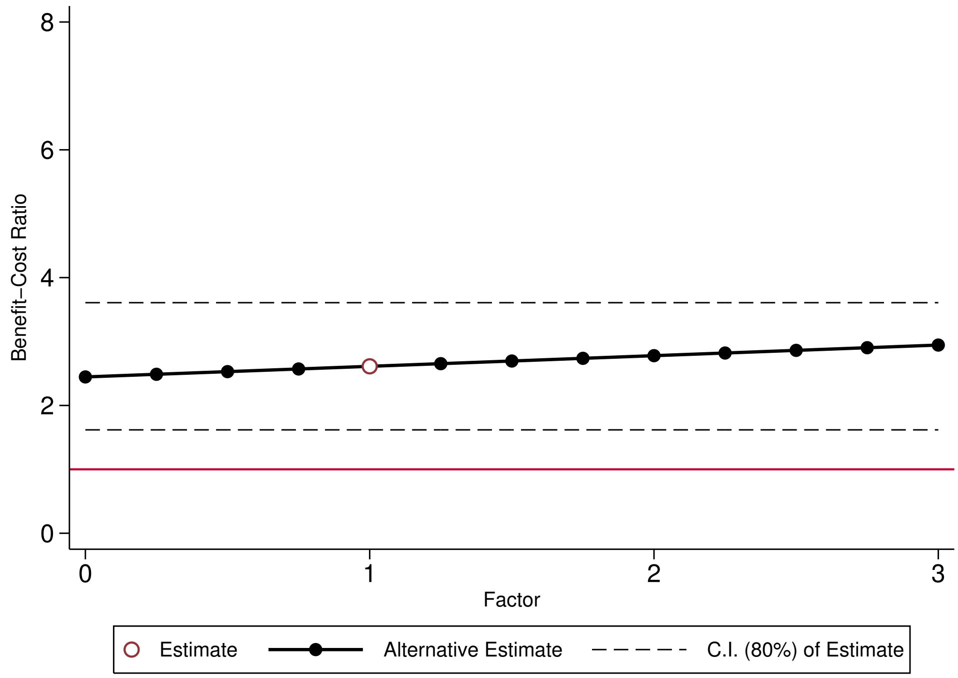
(g) Crime Costs



(h) Program Costs



(i) Control Substitution Costs

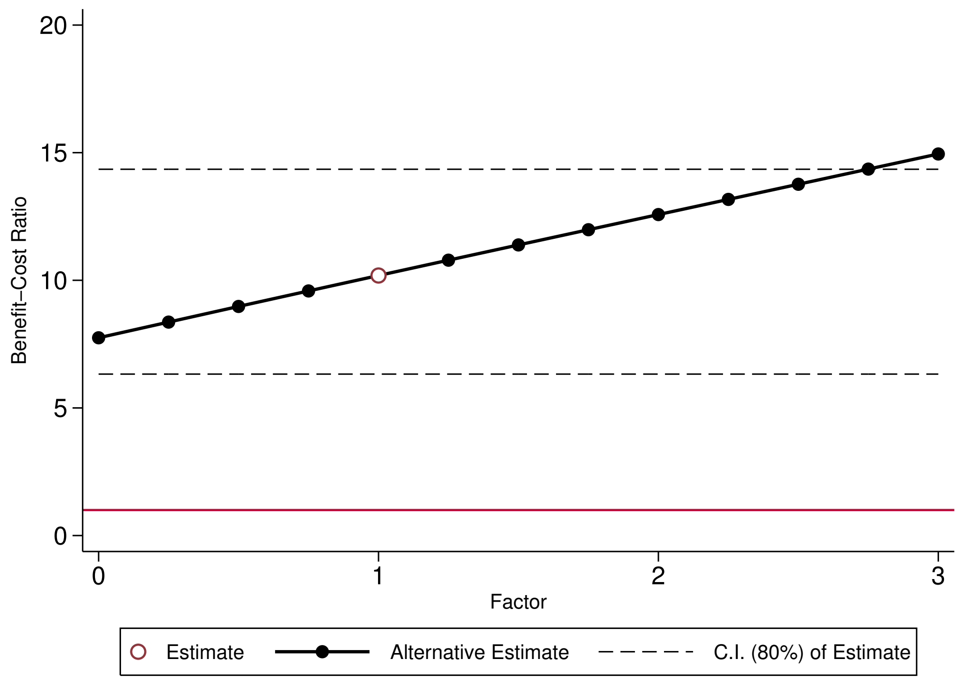


Note: These graphs display how the benefit/cost ratio changes for females as we multiply each component by a factor from 0 to 3. The red line indicates a benefit/cost ratio of 1. The hollow circle represents our actual estimates, whereas the solid dots represent the alternative estimates we obtain by 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 10^{th} and 90^{th} quantiles of the bootstrap distribution.

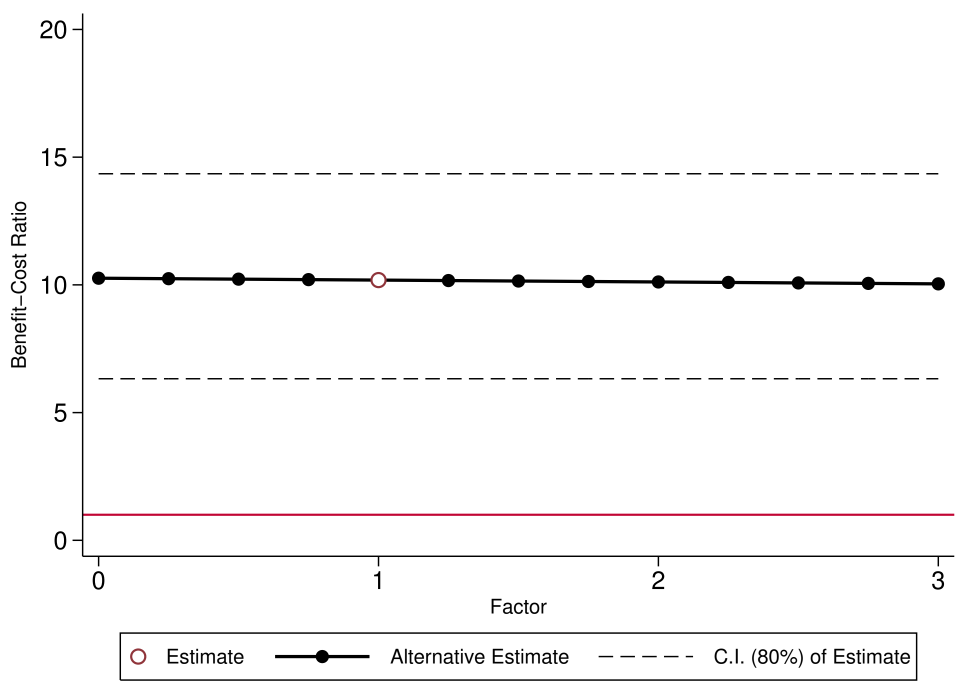
Figure G.8 displays how the benefit/cost ratio for males varies as we multiply each component of the benefits and costs by a factor between 0 and 3. We observe that the ratio is insensitive to the scaling of public-transfer income, health costs, education costs, and control substitution costs. Barring 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 G.8: Benefit/cost Ratio vs. Components, Males

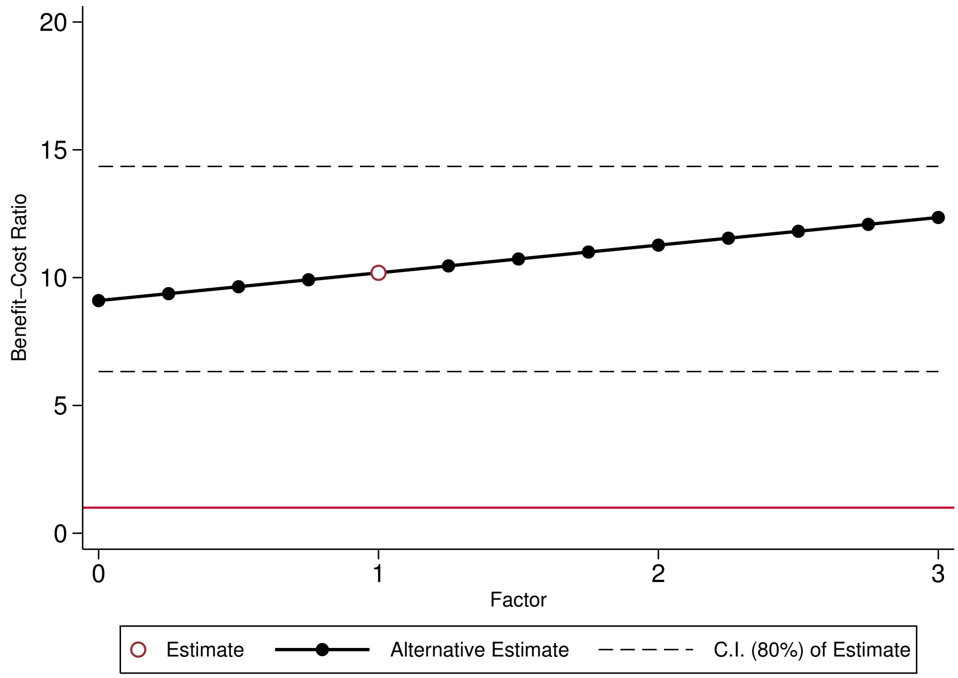
(a) Labor Income



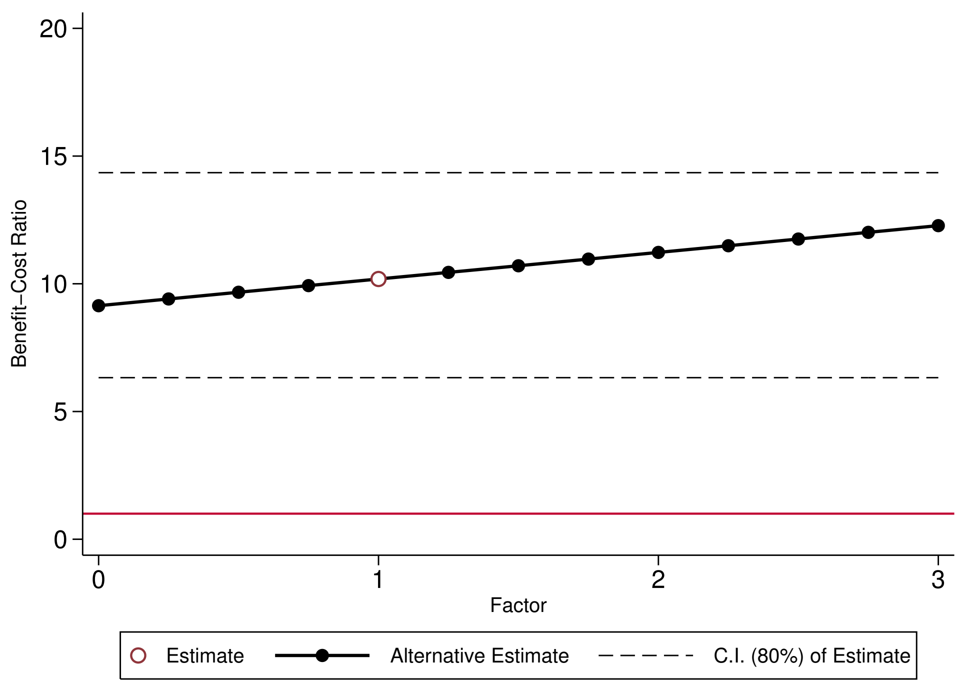
(b) Public-Transfer Income



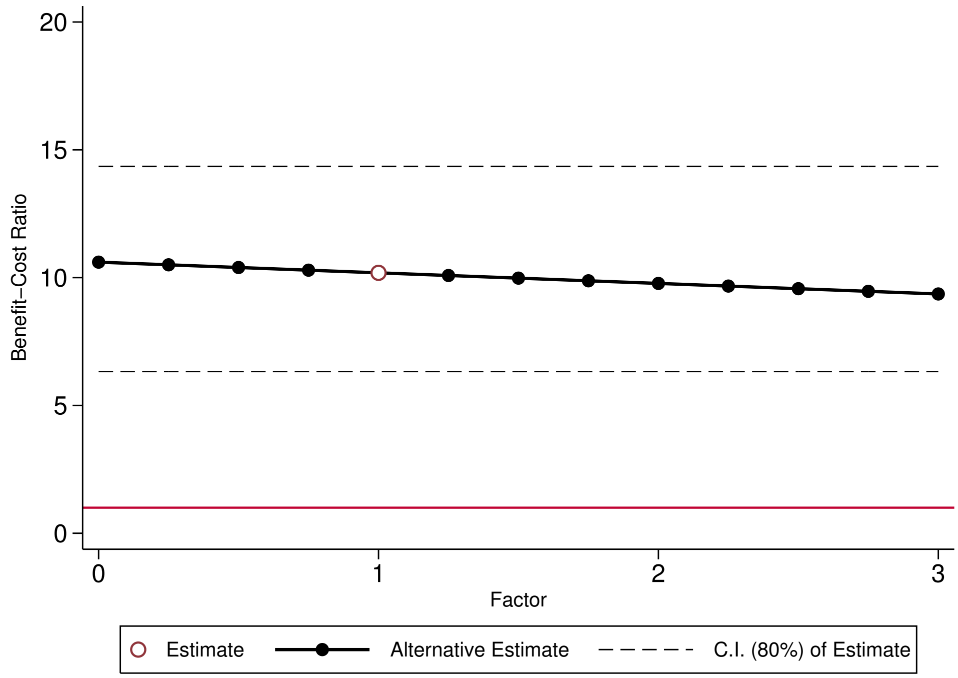
(c) Parental Income



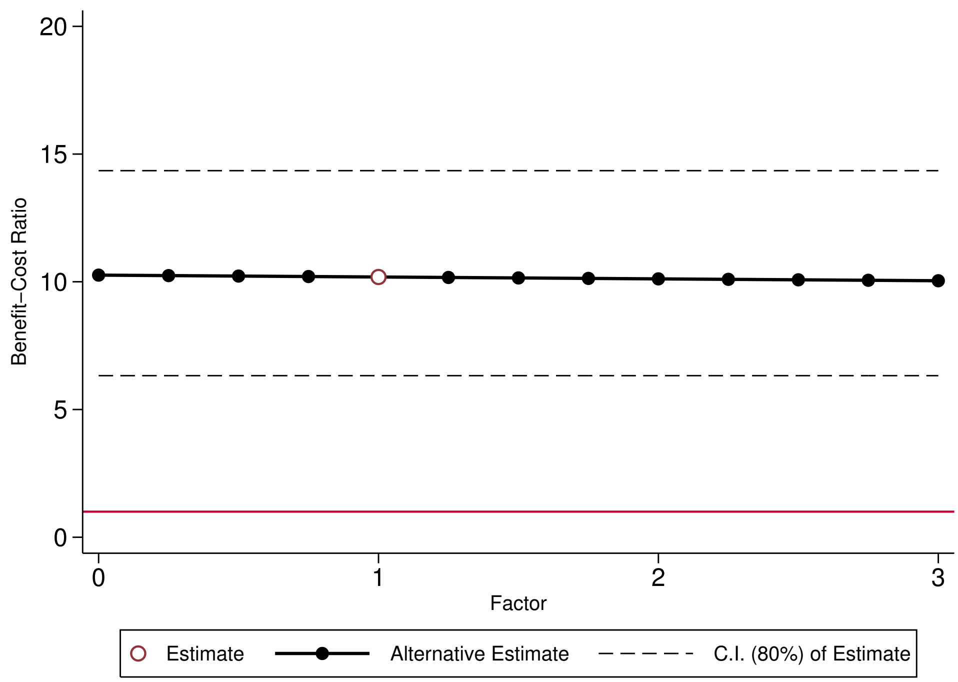
(d) Quality-Adjusted Life Years



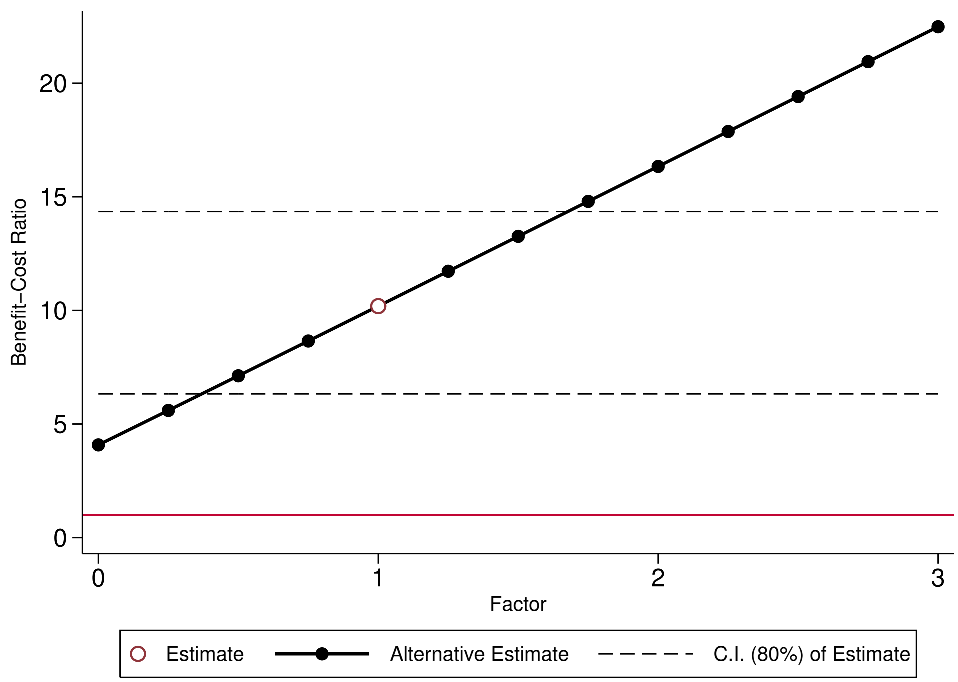
(e) Health Costs



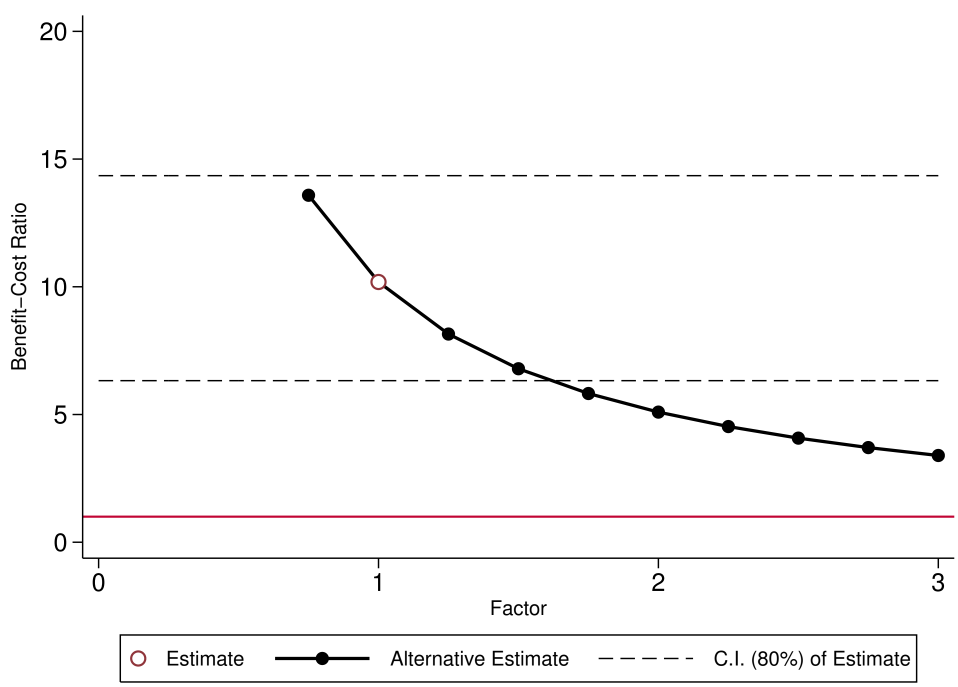
(f) Education Costs



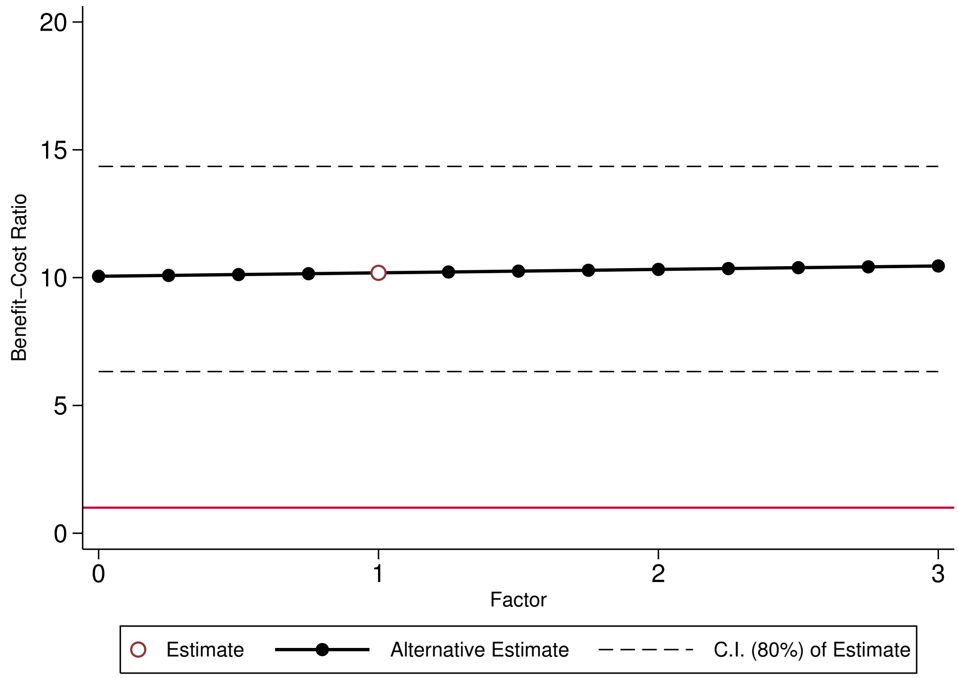
(g) Crime Costs



(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 10^{th} and 90^{th} 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.

Altonji, J. G. and L. M. Segal (1996, July). Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics 14*(3), 353–366.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 227–297. Oxford University Press.

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.

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.

Cawley, J. (2004, Spring). The impact of obesity on wages. *Journal of Human Resources 39*(2), 451–474.

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.

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.

Hansen, L. P. (1982, July). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

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., H. Ichimura, and P. E. Todd (1998, April). Matching as an econometric evaluation estimator. *Review of Economic Studies 65*(2), 261–294.

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 S. Navarro (2004, February). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics 86*(1), 30–57.

Heckman, J. J., 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., S. T. Mayne, B. I. Graubard, A. Schatzkin, D. Albanes, C. Schairer, R. N. Hoover, and M. F. Leitzmann (2008). Past Body Mass Index and Risk of Mortality Among Women. *International Journal of Obesity 32*(5), 730–739.

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.

Tong, L., M. R. Spitz, J. J. Fueger, and C. I. Amos (1996). Lung Carcinoma in Former Smokers. *Cancer 78*(5), 1004–1010.

Tysinger, B., D. Ermini Leaf, and D. P. Goldman (2015). The Future America 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*.