The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviours
Short Title: Health Effects of Two Early Interventions

Gabriella Conti, James J. Heckman, and Rodrigo Pinto

October 20, 2016

Contents

A Eligibility Criteria and Randomisation Protocol in the ABC and PPP

PPP The Perry Program admitted five entry cohorts in the early 1960s, drawn from the population surrounding the Perry Elementary school. Candidate families for the study were identified from a survey of the families of the students attending the elementary school, by neighborhood group referrals, and through door-to-door canvassing. The eligibility rules for participation were that the participants should: (1) be African-American; (2) have a low Stanford-Binet IQ (between 70 and 85) at study entry; and[1] (3) be disadvantaged as measured by the cultural deprivation scale, which included parental employment level, parental education, and housing density (persons per room). The Perry study targeted families who were more disadvantaged than other African-American families in the U.S. but were representative of a large segment of the disadvantaged African-American population. Heckman *et al.* (2010) show that if PPP were applied nationwide, the bottom 17% (for males) and 15% (for females) of the African-American population of the same cohort would be eligible for it. PPP families were particularly disadvantaged among families in the Perry Elementary School neighborhood. PPP families had lower levels of parental education, had fewer working mothers, had larger number of family members, and greater participation in welfare when compared to families in the Perry Elementary School neighborhood. See Heckman, Moon, Pinto, Savelyev and Yavitz (2010) for a comparison between PPP families and other families in the Perry School catchment area.

According to Weikart *et al.* (1978), families were assigned to treatment and control groups according to a stepwise randomisation protocol on the basis of some known pre-programme variables, namely: gender, wave cohort, IQ at entry, socioeconomic status as measured by the cultural deprivation scale and maternal employment at the onset of the programme. The randomisation protocol is complex and only partially known. According to Weikart *et al.* (1978, p. 16), for each designated eligible entry cohort, children were assigned to treatment and control groups in the following way:

1. Younger siblings of any entering cohort were assigned the same treatment status of older siblings previously enrolled.[2]

[1] In practice, IQs ranged from 61-88. See Heckman *et al.* (2010).

[2] The rationale for excluding younger siblings from the randomisation process was that enrolling children from the same family in different treatment groups would weaken the observed treatment effect due to within-family spillovers.

2. Remaining participants were ranked by their entry IQ scores. Odd- and even-ranked subjects were assigned to two separate unlabelled groups.

3. Some individuals initially assigned to one group were swapped between the unlabelled groups to balance gender and mean socioeconomic (SES) status, "with Stanford-Binet scores held more or less constant".

4. A flip of a coin (a single toss) labelled one group as "treatment" and the other as "control."

5. Some individuals provisionally assigned to treatment, whose mothers were employed at the time of the assignment, were swapped with control individuals whose mothers were not employed.

The rationale for these swaps associated with maternal employment was due to the difficulty for working mothers to participate in the home visits as part of the treatment and to the lack of funds for transportation.[3] We refer to Heckman *et al.* (2010) and Heckman *et al.* (2011) for a more detailed description of the Perry randomisation protocol. We apply their method in this paper.

ABC The Abecedarian study included four cohorts of children born between 1972 and 1977 and living in or near Chapel Hill, North Carolina. The children were recruited in four separate waves by eliciting referrals from local community organisations such as antenatal clinics, hospitals, and social services (Breitmayer and Ramey, 1986). Admission to the programme took place over a 5-year period resulting in four cohorts of approximately 28 children each.

The ABC sample size is also small. The final sample consists of 111 children. More detailed information is available on the randomisation protocol for ABC than for the Perry intervention. First, candidate eligible families were defined as those whose children appeared healthy and free from biological conditions that could be associated with mental, sensory, or motor disabilities (Ramey and Campbell, 1984). Once a candidate was identified, a High Risk Index (HRI) was computed from 13 socioeconomic factors capturing disadvantage (see Table 1 in the paper for the

[3]The following quotation from an early monograph on Perry summarises the logic of the study planners: "*Occasional exchanges of children between groups also had to be made because of the inconvenience of half-day preschool for working mothers and the transportation difficulties of some families. No funds were available for transportation or full-day care, and special arrangements could not always be made".* (Weikart, Bond and McNeil, 1978, p. 17)

complete list), as described in Ramey and Smith (1977), in order to determine eligibility.[4] The selection process took place either before or shortly after the birth of the subject child, and at the end of it a total of 122 eligible families were invited to enroll in the programme. Among those, 121 agreed to participate and one mother experienced a miscarriage, so that at the end a total of 122 children out of 120 families were randomised. According to Ramey and Campbell (1984, p. 517), the initial 122 children were matched in pairs by "identifying most similar pairs in the High-Risk Index". Published reports of ABC's randomisation protocol by FPGC investigators vary over time. Protocols reported include: pair-matching of infants by maternal IQ, sex of infant, and family income (Ramey *et al.*, 1975); pair-matching of families on sex of child, maternal IQ, number of siblings, and total HRI score (Breitmayer and Ramey, 1986; Ramey *et al.*, 1976, 1977); pair-matching of mother-infant dyads by maternal IQ, number of siblings, and sex of the infant (Ramey and Smith, 1977); pair-matching of mothers by maternal age, IQ, education, and parity (Ramey and Campbell, 1979); pair-matching on the HRI and maternal IQ (Ramey *et al.*, 1983); and pair-matching of families and children on the HRI (Ramey and Campbell, 1984, 1991). However, the precise matched pairs are unknown. Subsequently, the matched children were randomly assigned to treatment or control status within each pair. Treatment status is known for 118 out of 122 children; among these, 61 were assigned to the treatment, and 57 to the control group. At entry, infants were between 6 to 21 weeks of age, averaging about 8.8 weeks old.

Unlike the Perry sample, which had a high retention rate, the ABC intervention suffered losses from attrition at multiple stages.[5] First, 11 children out of 122 (from 120 families) were lost before the start of the treatment (Campbell and Ramey, 1995) because eight families declined participation after learning their treatment assignment.[6] Two participants previously assigned to the control group were swapped to the treatment group at the request of the local authorities, as their lives were threatened due to poor health conditions. Finally, one child was excluded after being diagnosed with idiopathic moderate retardation associated with a seizure disorder (Campbell and Ramey, 1995). The base sample who underwent the treatment includes 111 children (57 treated,

[4]Each family's value was compared to a cutoff value to determine eligibility, defined as scoring more than 11 on the index. Ramey and Smith (1977) indicate a score of "more than 11", however, Ramey and Campbell (1991) state "11 points or higher". The final eligibility status was then confirmed after an interview with the mothers.

[5]Ramey and Campbell (1979) include this statement: "New children were admitted to the study to replace three children who either died or moved away before 6 months of age."

[6]Of these families, seven belonged to the treatment group and only one to the control group.

54 control) born from 109 families (it included one pair of twins and one pair of siblings, both assigned to the treatment group). However, soon after the start of the treatment, four of the 111 children were lost for the following reasons.[7] One child was diagnosed to be biologically retarded, two died at very early ages (3 and 4 months), and a fourth child was withdrawn from the sample before 6 months of age. Subsequent losses due to attrition were encountered in the preschool phase, for a total of 12 children out of 107. Two children died at 12 and 50 months of age, and the remaining 10 children dropped out by 54 months (although one of them returned to the school-age intervention).[8] Of these 95 children who completed the preschool intervention, 49 belong to the treatment group, and 46 to the control group, so that retention rates are 86% for the treatment and 85% for the control group, respectively. Lastly, while few participants who dropped out in the first phase were subsequently successfully traced in the follow-ups, attrition was particularly severe at the last (biomedical) data collection round, carried out in the mid 30s. For a detailed study of the nature and the extent of attrition in the biomedical sweep, we refer the reader to Campbell *et al.* (2014).

[7]The report by Ramey and Campbell (1984) indeed includes 107 children.
[8]The report by Breitmayer and Ramey (1986), indeed, includes 95 children.

B Moderators

Although both ABC and PPP are early childhood intervention programmes that share many similar features, they also differ in various ways as mentioned in the paper. First, ABC provided full-day and full-year services for the first 5 years of life while PPP was a half-day and half-year programme that lasted only 2 years from age 3. Second, PPP was implemented in the 1960s in the Midwest of the U.S. (Ypsilanti, MI), followed by ABC about 10 years later in a Southeastern state (Chapel Hill, NC). Third, ABC included a nutrition and health care component while PPP did not. Fourth, background characteristics are slightly different between two programmes as shown in Section 1 of the paper.

In principle, to directly compare the two programmes, we would need to fully control for all of these differences. However, this is a challenging task, given data limitations and lack of full information on the ecology of each programme. In this section, we instead partially address this issue by examining the sensitivity of the treatment effects to a set of selected moderators, as shown in Table 2 of the paper. This set includes the Stanford-Binet IQ at age 3 for both programmes: by doing this, we aim to remove all cognitive effects arising from the first 3 years of treatment for the ABC sample. On the other hand, we also include several family characteristics such as father's presence, mother's age, schooling, and working status, and the number of siblings in the household. Our aim, by conditioning on these variables, is to place the two samples on the same basis at the same age for comparison purposes.

Table A1 presents the estimated treatment effects (γ) for the two programmes based on the following simple linear regression estimated by least squares, $Y_i = \alpha + \beta X_i + \gamma D_i + \epsilon_i$ where X_i and D_i denote moderators and treatment indicator for participant i, respectively. The top panel does not account for moderators but the bottom panel does. The top and the bottom panels together show whether the treatment effect estimates (γ) change by conditioning on age-3 moderators (X). The outcome measures (Y) listed in the first column of Table A1 are directly comparable between the two programmes (see Table 2 in the Web Appendix C). The results indicate that conditioning on age 3 moderators does not change substantially the inference about estimated treatment effects.

Table A1: Treatment Effects with and without moderators: Abecedarian and Perry

	ABC							PPP						
	Age	Female			Male			Age	Female			Male		
		Diff.	s.e.	p-val	Diff.	s.e.	p-val		Diff.	s.e.	p-val	Diff.	s.e.	p-val
Without moderators														
Weight	35	3.721	8.669	0.350	-6.850	9.739	0.244	40	-3.997	4.071	0.166	2.840	5.408	0.699
Height	35	-0.007	0.020	0.371	0.050	0.029	**0.047**	40	-0.023	0.022	0.151	-0.014	0.019	0.762
BMI	35	1.785	3.144	0.286	-4.075	2.809	**0.080**	40	-0.546	1.595	0.367	1.294	1.351	0.829
Overweight (BMI≥25)	35	-0.066	0.085	0.223	-0.028	0.196	0.444	40	-0.039	0.161	0.405	0.086	0.108	0.784
Obese (BMI≥30)	35	-0.061	0.149	0.343	-0.069	0.218	0.376	40	-0.033	0.129	0.399	0.076	0.114	0.746
Not a frequent drinker	30	0.023	0.095	0.405	-0.072	0.118	0.272	27	0.107	0.111	0.170	0.082	0.106	0.780
Alcohol consumption	30	-0.356	1.793	0.422	2.609	2.492	0.150	27	-0.618	1.268	0.314	-2.969	2.029	0.926
With moderators														
Weight	35	9.950	10.069	0.165	-6.047	12.723	0.320	40	0.393	4.739	0.467	2.218	6.004	0.644
Height	35	-0.014	0.023	0.271	0.049	0.037	**0.100**	40	0.001	0.025	0.478	-0.026	0.022	0.882
BMI	35	4.559	3.636	0.109	-4.472	3.725	0.124	40	0.286	1.818	0.438	1.331	1.522	0.807
Overweight (BMI≥25)	35	-0.077	0.094	0.210	-0.260	0.233	0.140	40	0.078	0.189	0.340	0.079	0.124	0.737
Obese (BMI≥30)	35	0.125	0.168	0.231	-0.340	0.293	0.132	40	0.014	0.138	0.459	0.068	0.130	0.698
Not a frequent drinker	30	0.020	0.115	0.432	-0.115	0.182	0.267	27	0.035	0.125	0.391	0.074	0.120	0.731
Alcohol consumption	30	-0.276	2.251	0.451	0.464	3.909	0.453	27	0.334	1.439	0.409	-2.868	2.313	0.890

Notes: p-values ≤ 0.10 are printed in boldface. All outcome measures listed in the first column are directly comparable across programmes. Moderators include father's presence at home, mother's age, schooling, and working status, the number of siblings, and Stanford-Binet IQ at age 3. The columns labelled as "Diff." present the estimated treatment effects (γ) obtained from $Y_i = \alpha + \gamma D_i + \epsilon_i$ for the top panel and from $Y_i = \alpha + \beta X_i + \gamma D_i + \epsilon_i$ for the bottom panel where X_i denotes moderators for participant i. One-sided p-values are reported.

C Description of the variables used in the paper

In the following we report how the variables used in the paper have been constructed, for both ABC and PPP. Notice that the questions have only been asked in the sweep in which they are used: the majority of them has only been asked in one sweep, with the exception of questions related to smoking and drinking, which in PPP have been asked at ages 27 and 40. Table A2 (at the end of this section) presents an overview of the comparability of measures across the two studies.

C.1 Physical Health

1. **Excellent or very good health**

   PPP, age 40 questionnaire: "Would you say that in general your health is: 1) Excellent 2) Very good 3) Good 4) Fair 5) Poor".

   ABC, age 30 questionnaire: "How would you rate your health? a) Excellent b) Very good c) Good d) Fair e) Poor".

   The answers "excellent" and "very good" have been coded as 1, the answers "good", "fair", and "poor" have been coded as 0.

2. **Health stopped from working**

   PPP, age 40 questionnaire: "In the past 15 years, has your health ever stopped you from working for a week or more, including holding a job or being able to do housework? 1) No 2) Yes".

   ABC, age 30 questionnaire: "Have you ever had a physical or nervous condition that kept you from working (For women: not including pregnancy or child birth)? 1) No 2) Yes".

   The answer "yes" has been coded as 1, the answer "no" has been coded as 0.

3. **Weight**

   PPP, age 40 questionnaire: "About how much do you weigh without shoes? Pounds".

   ABC, biomedical sweep (mid 30s): Measured by the physician during the physical exam.

In both cases, the measures have been collected in pounds, and they have been converted in kilos.

4. **Height**

   PPP, age 40 questionnaire: "About how tall you are without shoes? Feet Inches".

   ABC, biomedical sweep (mid 30s): Measured by the physician during the physical exam.

   In both cases, the measures have been collected in feet and inches, and they have been converted in metres.

5. **BMI**

   It is derived according to the standard formula: $\frac{weight(kg)}{height(m)^2}$.

6. **Overweight**

   A subject is defined overweight if BMI $\geq$25.

7. **Obese**

   A subject is defined obese if BMI $\geq$30.

8. **Diastolic and systolic blood pressure (ABC only)**

   ABC, biomedical sweep (mid 30s): Measured by the physician during the physical exam.

9. **Hypertension I and II (ABC only)**

   Hypertension I is defined as systolic blood pressure $>$ 140 and diastolic blood pressure $>$ 90.

   Hypertension II is defined as systolic blood pressure $>$ 140 or diastolic blood pressure $>$ 90.

C.2 Health Insurance

1. **Health care coverage**

   PPP, age 40 questionnaire: "Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs (Health Maintenance Organisations), or government plans such as Medicare or Medicaid? 1) No 2) Yes". The possible options are then stated in the following question: "How is most of your medical care paid for? Is the coverage through: Your employer; Someone else's employer; A plan that you or someone

else buys on your own; Medicaid or Medical Assistance [or substitute state programme name]; The military, CHAMPUS, TriCare, or the VA [or CHAMP-VA]; Medicare; Some other source (specify); Out of pocket".

ABC, age 30 questionnaire: "Which of the following best describes your current health insurance situation? a) No health insurance b) Covered by husband's or wife's insurance policy c) Get own insurance through work d) Get insurance through a union e) Get insurance because of attending school f) Covered because of being on active duty in the military g) Buy private insurance on your own h) Covered by Medicaid i) You don't know what your health insurance situation is j) Other (Please specify)".

For PPP, the answer "yes" has been coded as 1, the answer "no" has been coded as 0.[9] For ABC, the answers "No health insurance" and "You don't know what your health insurance situation is" have been coded as 0, all the others have been coded as 1.

2. **Employer-provided or bought**

PPP, age 40 questionnaire: "How is most of your medical care paid for? Is the coverage through: Your employer? 1) No 2) Yes. A plan that you or someone else buys on your own? 1) No 2) Yes."

ABC, age 30 questionnaire: 'Which of the following best describes your current health insurance situation? a) No health insurance b) Covered by husband's or wife's insurance policy c) Get own insurance through work d) Get insurance through a union e) Get insurance because of attending school f) Covered because of being on active duty in the military g) Buy private insurance on your own h) Covered by Medicaid i) You don't know what your health insurance situation is j) Other (Please specify)".

For PPP, the answer "yes" to either of the questions has been coded as 1, the answer "no" has been coded as 0. For ABC, the answers "Get own insurance through work" and "Buy private insurance on your own" have been coded as 1, all the others have been coded as 0.

3. **Provided in prison (PPP only[10])**

[9] Individuals who are in prison and subsequently report that they have health care coverage in prison are also coded as "yes".

[10] This option is not available in ABC.

PPP, age 40 questionnaire: "How is most of your medical care paid for? Is the coverage through: Some other source: state prison? 1) No 2) Yes".

C.3 Demand for Health Care

1. **Hospitalised**

   PPP, age 40 questionnaire: "During the past 12 months, have you been admitted as a patient in the hospital for any reason? 1) No 2) Yes".

   ABC, biomedical sweep (mid 30s): "Past Hospitalisations 0) No 1) Yes 2) DK". (Recorded at the time of the physician exam).

   For both, the answer "yes" has been coded 1, the answer "no" and "DK" (ABC only) has been coded 0.

2. **Scheduled treatment or exam**

   PPP, age 40 questionnaire: "Now I have some questions about the health services that you use. Please indicate how many times you saw each of the following doctors in the past 12 months about your physical health. Include only visits regarding your own physical health, not visits when you took someone else to be examined. A doctor, emergency room, or clinic for scheduled treatment or exam. 0) None X) Number of times".

   ABC, age 30 questionnaire: "When did you last have an exam for an illness or an injury? a) Within the past 3 months b) 4-6 months ago c) Within the past year d) Longer than one but less than 2 years ago e) More than two years ago".

   For PPP, any "number of times" was coded as 1, "none" as 0. For ABC, the answers "Within the past 3 months", "4-6 months ago", and "Within the past year" were coded as 1, all the others were coded as 0.

C.4 Lifestyles

1. **Physical Activity**

PPP, age 40 questionnaire: "During the past month, did you participate regularly in any physical activities such as walking for exercise, running, sports or other fitness exercise? 1) No 2) Yes". "What type of physical activity or exercise did you spend the most time doing during the past month? 1) Walking/treadmill 2) Jogging 3) Lifting weights 4) Biking 5) Rollerskating/rollerblading/ice skating 6) Dancing 7) Aerobics 8) Swimming 9) Organised sports/basketball/soccer/hockey/volleyball/etc. 10) Go to fitness centre/home gym 11) Isometrics 12) Horse shoes 13) Stair master/stepper 14) Yard work 15) Personal training 16) Stretching 17) T-bow 18) Shadow boxing". "How often did you take part in this activity during the past month? 1) Two or more times per week 2) Once per week 3) Two or three times per month 4) Once a month 5) Rarely or never".

ABC, age 21 *Risk Taking Survey*:[11] "On how many of the past 7 days did you exercise or participate in sports activities for at least 20 minutes that made you sweat and breathe hard, such as basketball, jogging, fast dancing, swimming laps, tennis, fast bicycling, or similar aerobic activities? A) 0 days B) 1 day C) 2 days D) 3 days E) 4 days F) 5 days G) 6 days H) 7 days".

In both cases we constructed variables consistent with the CDC guidelines for adults.[12] In PPP there is no information on the time spent in physical activities, so we followed the guideline "muscle-strengthening activities on 2 or more days a week". Hence, the variable was coded 1 if the individual reported having participated in any non-walking physical activity in the past month for two or more times per week. In the ABC, there is information on the time spent, so we followed the guideline "1 hour and 15 minutes (75 minutes) of vigorous-intensity aerobic activity (i.e., jogging or running) every week". Hence, the variable was coded 1 if the individual exercised or participated in sports activities four or more days per week (= 80 minutes at least).

2. **Diet**

PPP, age 40 questionnaire: "In the past 15 years have you made any changes in your

[11]The question on physical activity was not asked at the age 30 interview.
[12]http://www.cdc.gov/physicalactivity/everyone/guidelines/adults.html

diet for health reasons? 1) No 2) Yes".

ABC, age 21 *Risk Taking Survey*:[13] "Yesterday, did you eat fruit?" "A) No B) Yes, once only C) Yes, twice or more".

For PPP, the answer "yes" has been coded as 1, the answer "no" has been coded as 0. For ABC, for all the variables, answers "no" have been coded as 0, answers "yes, once only" have been coded as 1, and answers "yes, twice or more" have been coded as 2.

C.4.1 Smoking

3. **Never smoker**

   PPP, age 40 questionnaire: "Have you smoked at least 100 cigarettes (= 5 packs) in your entire life? 1) No 2) Yes".

   ABC, age 30 *Tobacco, Alcohol and Drugs (TAD) Survey*: "How old were you when you first started smoking cigarettes regularly (at least one cigarette every day for 30 days)? a) I have never smoked regularly b) Less then 9 years old c) 9-12 years old d) 13-16 years old e) 17-20 years old f) 21-24 years g) 25 or older".

   For PPP, the answer "no" has been coded 1, the answer "yes" has been coded 0. For ABC, the answer "I have never smoked regularly" has been coded 1, the others have been coded 0.

4. **Not a daily smoker**

   PPP, age 27 questionnaire: "Do you smoke cigarettes or use other forms of tobacco daily? 1) No 2) Yes 9) dk/nr".

   PPP, age 40 questionnaire: "Do you now smoke cigarettes or use other forms of tobacco? 1) No 2) Yes". "If yes, do you now smoke cigarettes every day or just some days? 1) Every day 2) Some days".

   ABC, age 30 *Tobacco, Alcohol and Drugs (TAD) Survey*: "During the past 30 days, on the days that you smoked, how many cigarettes did you smoke per day? a) I did not smoke during the past 30 days b) Fewer than one per day c) 1 cigarette per day d) 2-5

[13] The questions on food consumption were not asked at the age 30 interview.

cigarettes per day e) 6-10 cigarettes per day f) 11-20 cigarettes per day g) Two or more packs per day".

For PPP, answers "yes" (age 27 & 40) and "every day" (age 40) have been coded 0, answers "no" and "some days" have been coded as 1. For ABC, answers "I did not smoke during the past 30 days" and "Fewer than one per day" have been coded 1, all the others have been coded 0.

5. **Not a heavy smoker**

   PPP, age 27 questionnaire: "How many cigarettes do you usually smoke or how much tobacco do you use each day? Cigarettes; dk/nr; na, do not smoke".

   PPP, age 40 questionnaire: "If daily, how many cigarettes do you usually smoke each day? Cigarettes".

   ABC, age 30 *Tobacco, Alcohol and Drugs (TAD) Survey*: "During the past 30 days, on the days that you smoked, how many cigarettes did you smoke per day? a) I did not smoke during the past 30 days b) Fewer than one per day c) 1 cigarette per day d) 2-5 cigarettes per day e) 6-10 cigarettes per day f) 11-20 cigarettes per day g) Two or more packs per day".

   Individuals have been coded as "not heavy smokers" if they report smoking 10 or less cigarettes per day (the non-smokers are coded as smoking 0 cigarettes).

6. **Cigarettes per day**

   PPP, age 27 questionnaire: "How many cigarettes do you usually smoke or how much tobacco do you use each day? Cigarettes; dk/nr; na, do not smoke".

   PPP, age 40 questionnaire: "If daily, how many cigarettes do you usually smoke each day? Cigarettes".

   ABC, age 30 *Tobacco, Alcohol and Drugs (TAD) Survey*: "During the past 30 days, on the days that you smoked, how many cigarettes did you smoke per day? a) I did not smoke during the past 30 days b) Fewer than one per day c) 1 cigarette per day d) 2-5

cigarettes per day e) 6-10 cigarettes per day f) 11-20 cigarettes per day g) Two or more packs per day".

For PPP, the variable has been used as reported. For ABC, the midpoints have been used: I did not smoke during the past 30 days has been coded as 0; Fewer than one per day has been coded 0.5; 1 cigarette per day has been coded 1; 2-5 cigarettes per day has been coded 3.5; 6-10 cigarettes per day has been coded 8; 11-20 cigarettes per day has been coded 15.5; Two or more packs per day has been coded 40.

7. **Age of onset of smoking (ABC only)**

   ABC, age 21 *Risk Taking Survey*: "How old were you when you first started smoking cigarettes regularly (at least one cigarette every day for 30 days)? a) I have never smoked cigarettes regularly b) Less then 9 years old c) 9 or 10 years old d) 11 or 12 years old e) 13 or 14 years old f) 15 or 16 years g) 17 or more years old"; age 30 *Tobacco, Alcohol and Drugs (TAD) Survey*:[14] "How old were you when you first started smoking cigarettes regularly (at least one cigarette every day for 30 days)? a) I have never smoked regularly b) Less then 9 years old c) 9-12 years old d) 13-16 years old e) 17-20 years old f) 21-24 years g) 25 or older".

   This variable is only defined for smokers. The midpoints of the age bands have been used: "less than 9 years old" and "9 or 10 years old" have not been answered by anyone; "11 or 12 years old" (age 21) has been coded 11.5; "9-12 years old" (age 30) has been coded 10.5; "13 or 14 years old" (age 21) has been coded 13.5; "15 or 16 years" (age 21) has been coded 15.5; "13-16 years old" (age 30) has been coded 14.5; "17 or more years old" (age 21) has been coded as the midpoint between 17 and the age at the interview, resulting in values ranging from 19 to 21; "17-20 years old" (age 30) has been coded 18.5, "21-24 years" (age 30) has been coded 22.5, and "25 or older" (age 30) has been coded as the midpoint between 25 and the age at the interview, resulting in values ranging from 27 to 28. The response given at age 21 has been used to minimise recall bias, unless the individual did not respond at age 21 or had not started smoking yet, in which cases the response given at age 30 was used.

[14] The age at smoking onset has not been asked for PPP.

C.4.2 Drinking

8. **Not a frequent drinker**

   PPP, age 27 questionnaire: "How often do you drink alcoholic beverages? 1) never 2) once in a while 3) several times a week 4) daily 8) DK 9) NR".

   PPP, age 40 questionnaire: "During the past month, how many days did you drink any alcoholic beverages? Times".

   ABC, age 30 *Tobacco, Alcohol and Drugs (TAD) Survey*: "During the past 30 days, on how many days did you have at least one drink of alcohol? a) 0 days b) 1-2 days c) 3-5 days d) 6-9 days e) 10-19 days f) 20-29 days g) All 30 days".

   Individuals drinking never or once in a while (PPP age 27) or less than 10 days per month (PPP age 40 and ABC) have been coded 1, those drinking several times a week or daily (PPP age 27) or 10 days or more (PPP age 40 and ABC) have been coded 0.

9. **Alcohol consumption per month**

   PPP, age 27 questionnaire: "How often do you drink alcoholic beverages? 1) never 2) once in a while 3) several times a week 4) daily 8) DK 9) NR".

   PPP, age 40 questionnaire: "During the past month, how many days did you drink any alcoholic beverages? Times".

   ABC, age 30 *Tobacco, Alcohol and Drugs (TAD) Survey*: "During the past 30 days, on how many days did you have at least one drink of alcohol? a) 0 days b) 1-2 days c) 3-5 days d) 6-9 days e) 10-19 days f) 20-29 days g) All 30 days". For PPP at 27, the variable has been recoded as follows: 'never' has been coded 0, 'once in a while' has been coded 4 (once a week), 'several times a week' has been coded 12 (assuming 3 times a week), 'daily' has been coded 30. For PPP at 40, the variable has been used as reported. For ABC, the midpoints of the day bands have been used: "1-2 days" has been coded 1.5, "3-5 days" has been coded 4, "6-9 days" has been coded 7.5, "10-19 days" has been coded 14.5, "20-29 days" has been coded 24.5, and "all 30 days" has been coded 30.

10. **Underage drinker (ABC only)**

ABC, age 21 *Risk Taking Survey*:[15] "How old were you when you had your first drink of alcohol other than a few sips? a) I have never had a drink of alcohol other than a few sips b) Less then 9 years old c) 9 or 10 years old d) 11 or 12 years old e) 13 or 14 years old f) 15 or 16 years g) 17 or more years old".

An individual is defined as "underage drinker" if he reported started drinking at age less than 17: "Less then 9 years old", "9 or 10 years old", "11 or 12 years old", "13 or 14 years old", and "15 or 16 years" have been coded as 1, "I have never had a drink of alcohol other than a few sips" and "17 or more years old" have been coded as 0.

[15]This question has also been asked at age 30, however, we use the age 21 interview to minimise the possibility of recall bias.

Table A2: Comparability of Data on Outcomes in the ABC and PPP Studies

Outcome	Age PPP	Age ABC	Compa-rability	Notes
				Physical Health
Excellent or very good health	40	30	High	Same 5-point scale.
Health stopped from working	40	30	Medium	Different recall period ('past 15 yrs' in PPP, 'ever' in ABC) and reference period ('a week or more' in PPP, not specified in ABC).
Weight	40	30	High	Self-reported in PPP, measured in ABC.
Height	40	30	High	Self-reported in PPP, measured in ABC.
BMI	40	30	High	Height and weight are self-reported in PPP, measured in ABC.
Overweight (BMI≥25)	40	30	High	Height and weight are self-reported in PPP, measured in ABC.
Obese (BMI≥30)	40	30	High	Height and weight are self-reported in PPP, measured in ABC.
Diastolic blood pressure	n/a	mid 30s	None	Only measured in ABC.
Systolic blood pressure	n/a	mid 30s	None	Only measured in ABC.
				Health Insurance
Health care coverage	40	30	Medium	PPP refers to 'health care coverage', ABC to 'health insurance'; the options are comparable.
Employer-provided or bought	40	30	Medium	PPP refers to 'medical care coverage', ABC to 'health insurance'; the options are comparable.
Provided in prison	40	n/a	None	Answer not available in ABC.
				Demand for Health Care
Hospitalised	40	mid 30s	Medium	Different recall period ('past 12 mths' in PPP, not specified in ABC).
Scheduled treatment or exam	40	30	Medium	PPP refers to 'doctor, emergency room, or clinic for scheduled treatment or exam'; ABC refers to 'exam for illness or injury'; the recall period is the same (12 months).
				Lifestyles - Diet and Physical Activity
Physical activity	40	21	Low	Different recall period ('past month' in PPP, 'past 7 days' in ABC); different reference time ('regularly' in PPP, 'at least 20 minutes' in ABC); different wording ('any physical activities' in PPP, 'sports activities that made you sweat and breathe hard' in ABC); the variables are constructed following CDC guidelines.
Diet & nutrition	40	21	Low	Different recall period ('past 15 years' in PPP, 'yesterday' in ABC); different wording ('changes in diet for health reason' in PPP, 'number of times ate fruit' in ABC).
				Lifestyles - Smoking
Never smoker	40	30	Medium	Different wording ('at least 100 cigarettes (=5 packs)' in PPP, 'regular smoking' in ABC).
Not a daily smoker	27 & 40	30	Medium	Different recall period ('now' in PPP, 'during the past 30 days' in ABC).
Not a heavy smoker	27 & 40	30	Medium	Different recall period ('now' in PPP, 'during the past 30 days' in ABC); different reference sample ('daily smoker' in PPP, 'on the days that you smoked' in ABC).
Number of cigarettes per day	27 & 40	30	Medium	Different recall period ('now' in PPP, 'during the past 30 days' in ABC); different reference sample ('daily smoker' in PPP, 'on the days that you smoked' in ABC). Note: the variable is continuous in PPP, banded in ABC.
Age of onset of smoking	n/a	21-30	None	Only asked in ABC.
				Lifestyles - Drinking
Not a frequent drinker	27	30	Medium	Different recall period (unspecified in PPP, 'during the past 30 days' in ABC); different wording (frequency in PPP, number of days in ABC).
Not a frequent drinker	40	30	High	Same recall period (past 30 days), same wording (number of days).
Alcohol consumption	27	30	Medium	Different recall period (unspecified in PPP, 'during the past 30 days' in ABC); different wording (frequency in PPP, number of days in ABC).
Alcohol consumption	40	30	High	Same recall period (past 30 days), same wording (number of days).
Age of onset of drinking < 17	n/a	21	None	Only asked in ABC.

Notes: This table reports the level of comparability of the outcomes in ABC and PPP. The description of how the variables have been constructed on the basis of the questions asked in the data collections carried out at the different sweeps is reported above. We define comparability as 'high' when the questions are the same, and at most there is a difference in the collection method (e.g., weight). We define comparability as 'medium' when there is a difference in the questions asked, either in the recall/reference period, or in the wording; however, the questions refer otherwise to the same construct (e.g., daily smoking). We define comparability as 'low' when only the domains the variables belong to are the same (e.g., physical activity), but there are significant differences in the question's wording. We define comparability as 'none' when the variables have only been collected in one of the two interventions.

D Methodology

The standard model of programme evaluation describes the observed outcome Y_i of participant $i \in \mathcal{I}$ by:

$$Y_i = D_i Y_i(1) + (1 - D) Y_i(0), \tag{D.1}$$

where $\mathcal{I} = \{1, \ldots, N\}$ denotes the sample space, D_i denotes the treatment assignment for participant i ($D_i = 1$ if treatment occurs, $D_i = 0$ otherwise) and $(Y_i(0), Y_i(1))$ are potential outcomes for participant i when treatment is *fixed* at control and treatment status, respectively. By "fixed" we mean the value the outcome takes when treatment is exogenously set at some treatment status (see Heckman and Pinto, 2015 for a discussion on fixing and causality).

Heckman *et al.* (2010) discuss how randomised experiments solve potential problems of selection bias by inducing independence between counterfactual outcomes $(Y_i(0), Y_i(1))$ and treatment status D_i when conditioned on the pre-programme variables X used in the randomisation protocol. In our notation, we write:

Assumption A-1. $(Y(1), Y(0)) \perp\!\!\!\perp D \mid X$

where variables X, D, and $Y(d)$; $d \in \{0, 1\}$ are N-dimensional vectors whose elements are associated with participants $i \in \mathcal{I}$, e.g., $X = (X_i; i \in \mathcal{I})$, $D = (D_i; i \in \mathcal{I})$, and $Y(d) = (Y_i(d); i \in \mathcal{I}); d \in \{0, 1\}$. In the same fashion, we represent the vector of observed outcomes of Equation (D.1) by $Y = (Y_i; i \in \mathcal{I})$.

In the Perry study, variables X used in the randomisation protocol are cohort, gender, children IQ, Socioeconomic Status (SES), and maternal employment. In the ABC study, variables X are cohort, gender, maternal IQ, HRI, and number of siblings. In the case of Perry, treatment assignment was randomised for each family on the basis of strata defined by variables X of the eldest sibling of each family. In the case of Abecedarian, participants were matched in pairs on the basis of strata defined by variables X.

Our aim is to test the null hypothesis of no treatment effect. This hypothesis is equivalent to the statement that the conditional counterfactual outcome vectors share the same distribution:

Hypothesis H-1. $Y(1) \overset{dist}{=} Y(0) \mid X$,

where $\overset{dist}{=}$ denotes equality in distribution. The no treatment effect hypothesis **H-1** can be restated in more tractable form:

Hypothesis H-1′. Under Assumption **A-1** and Hypothesis **H-1**, we have that $Y \perp\!\!\!\perp D \mid X$.

Testing Hypothesis **H-1′** poses some statistical challenges. First, Perry and Abecedarian have small sample sizes, which cast doubt on inference that relies on the asymptotic behaviour of test statistics. We address the problem of small sample size by using exact permutation tests which are tailored to the randomisation protocol implemented in each intervention. Second, the presence of multiple outcomes allows for the arbitrary selection of statistically significant ones. The selective reporting of statistically significant outcomes is often termed *cherry picking* and generates a downward biased inference with smaller p-values. We solve the problem of cherry picking by implementing a multiple-hypothesis testing correction based on the stepdown procedure of Romano and Wolf (2005). Third, non-random attrition can create bias in the estimation and inference of treatment effects. We address the problem of attrition by using an inverse probability weighting estimator that controls for missing observations by estimating the probability of dropping out of the sample as function of pre-programme variables.

The rest of the exposition is organised as follows. Subsection D.1 discusses small sample permutation test. Subsection D.2 presents our multiple-hypothesis inference. Subsection D.3 describes our correction for attrition.

D.1 Small Sample Inference

We test Hypothesis **H-1′** through a permutation test that is valid for the small sample sizes of the Perry and Abecedarian interventions. To do so, we explore the invariance of the joint distribution of (Y, D) under permutations that swap the elements of the vector of treatment status D.

The invariance of joint distribution (Y, D) stems from two statistical properties. First, the randomisation guarantees that D is exchangeable for a set of selected permutations. Otherwise stated, the distribution of D remains the same for selected swaps of elements in D. Second, the *joint* invariance comes from the assumption of no treatment effect, which implies that $Y \perp\!\!\!\perp D \mid X$ according to Hypothesis **H-1′**.

The exchangeability property of D comes from the fact that scrambling the order of the partic-

ipants sharing the same values on X would not change the underlying distribution of the vector of treatment assignments D (see Heckman *et al.*, 2010 for a discussion). More notation is needed to formalise this property.

Let $\mathscr{G}_X$ be the set of all permutations that permute elements only within each stratum of X. Formally:

$$\mathscr{G}_X = \{\pi_g : \mathcal{I} \to \mathcal{I} : \text{ such that } \pi_g \text{ is a bijection and } (\pi_g(i) = j) \Rightarrow (X_i = X_j)\, \forall\, i \in \mathcal{I}\}.$$

The exchangeability property of D can be written as:

$$D \overset{d}{=} gD \quad \forall\, g \in \mathscr{G}_X, \tag{D.2}$$

where $gD = (D_{\pi_g(i)} : i \in \mathcal{I})$.

An important feature of the exchangeability property stated in Equation (D.2) is that it relies on limited information of the randomisation protocol. It does not require a full specification of the distribution D nor of the assignment mechanism, but only the knowledge of which variables are used as X in the randomisation protocol. Moreover, exchangeability (D.2) remains valid under compromises of the randomisation that is based on the information contained in X.

We use the the exchangeability property (D.2) to generate a permutation-based inference of the hypothesis of no treatment effects. The rationale for permutation-based inference relies on both the exchangeability of treatment assignments and the no-treatment hypothesis. Exchangeability assures that the distribution of the vector of treatment assignments is invariant under valid permutations. Moreover, the no-treatment Hypothesis **H-1** states that the vector of outcomes is independent of the vector of treatment assignments. Thus, the joint distribution of outcomes and treatment status is invariant under permutations in $\mathscr{G}_X$:

Theorem T-1. Under Hypothesis **H-1**, the joint distribution of outcomes Y and treatment assignments D are invariant under permutations $\mathscr{G}_X$ of treatment assignments within strata formed by values of covariates X, that is:

$$(Y, D) \overset{d}{=} (Y, gD)\ \forall\ g \in \mathscr{G}_X.$$

Proof. By Equation (D.2), $D \overset{d}{=} gD \; \forall \; g \in \mathscr{G}_X$. But $Y \perp\!\!\!\perp D \mid X$ by Hypothesis **H-1**. Thus, $(Y, D) \overset{d}{=} (Y, gD) \; \forall \; g \in \mathscr{G}_X$. □

Theorem **T-1** states what is often called the randomisation hypothesis.

A consequence of Theorem **T-1** is that a statistic based on assignments D and outcomes Y is distribution-invariant under permutations $g \in \mathscr{G}_X$. Moreover, under the null hypothesis, the exact distribution of a statistic is given by the collection of its values generated by all permutations in $\mathscr{G}_X$ (Lehmann and Romano, 2005).

We can use Theorem **T-1** to construct a permutation test. Let larger values of a statistic $T(Y, D)$ provide evidence against Hypothesis **H-1**. Next, let a critical value c be such that we reject the null hypothesis of treatment effect if $T(Y, D) \geq c$. Thus, if our inference aims to control for a Type-I error at significance level α, then the following equation must hold:

$$\begin{aligned} &\Pr(\text{Reject Hypothesis } \textbf{H-1} \mid \text{Hypothesis } \textbf{H-1} \text{ is true}) \\ &\qquad = \Pr(T(Y, D) \geq c \mid \text{ Hypothesis } \textbf{H-1} \text{ is true}) \leq \alpha. \end{aligned} \tag{D.3}$$

The critical value can be computed by taking the α-quantile of the set $\{T(Y, gD) : \; g \in \mathscr{G}_X\}$. In practice, permutation tests compare a test statistic computed on the original (unpermuted) data with a distribution of test statistics computed on resamplings of that data according to permutations in $\mathscr{G}_X$. The measure of evidence against the randomisation hypothesis, the p-value, is computed as the fraction of resampled data which yields a test statistic greater than that yielded by the original data.

Permutation-based inference is often termed data-dependent because the computed p-values are conditional on the observed data. These tests are also *distribution-free* because they do not rely on assumptions about the parametric distribution from which the data have been sampled. Because permutation tests give accurate p-values even when the sampling distribution is skewed, they are often used when sample sizes are small and sample statistics are unlikely to be normal.

We adopt the pre-pivoted t-statistics between treatment and control groups for the choice of test statistics. We show the permutation mid-p-value for inference. See details of those statistics in Heckman *et al.* (2010).

D.2 Correcting for Multiple-Hypothesis Testing

The availability of multiple outcomes creates the danger of selectively reporting statistically significant results that might occur just by chance. To make this statement more precise, suppose that a single-hypothesis test statistic rejects a true null hypothesis at significance level α. Thus, the probability of rejecting a single hypothesis out of K true hypotheses is given by $1-(1-\alpha)^K$. As the number of outcomes K increases, the likelihood of rejecting a true null hypothesis departs from α.

We correct for the possibility of arbitrary selection of statistically significant outcomes by adjusting for multiple hypothesis testing. We adopt the *familywise error rate* (FWER) as the Type-I error for multiple hypotheses. FWER is the probability of rejecting any true null hypothesis in a joint test of a set of hypotheses. The stepdown algorithm of Lehmann and Romano (2005) exhibits *strong FWER control*, that is to say that FWER is held at or below a specified level regardless of which single hypotheses are true within a set of hypotheses.

The Lehmann and Romano (2005) stepdown method has better statistical properties than traditional Bonferroni and Holm methods by exploiting the statistical dependence of the distributions of test statistics. By accounting for the correlation among single hypothesis p-values, we are able to perform a less conservative multiple-hypothesis test. In addition, the stepdown method generates as many adjusted p-values as there are hypotheses, which facilitates the examination of which sets of hypotheses are rejected.

In this framework, a set of necessary conditions for strong FWER control can be stated as follows:

(i) Permutations g belong to $\mathscr{G}_X$, that is, permutations are such that, under the null hypotheses, the joint distribution (Y, gD) is invariant for each permutation.

(ii) The same draw of permutation is used to compute the test statistics associated with all hypotheses we ought to test.

(iii) The joint-hypothesis test statistic at each stepdown stage is chosen to be the maximum of the individual-hypothesis test statistics.

There is some arbitrariness in defining the blocks of hypotheses that are jointly tested in a multiple-

hypothesis testing procedure. To avoid arbitrariness, we define blocks of independent interest that are selected on *a priori* grounds.

D.3 Correcting for Attrition

Numerous outcomes Y are not observed for some participants due to attrition, whereas the treatment D and pre-programme variables X are observed for all participants. As a consequence, the restricted sample may flaw the validity of the randomisation property Assumption D.1 **A-1** when conditioned to the non-missing data.

We correct for attrition using statistical models that adjust missing data using observed covariates. Specifically, we retrieve statistics for the full outcome distribution through reweighing non-missing observations according to their likelihood of compliance. This is usually termed Inverse Probability Weighting (IPW) and goes back to Horvitz and Thompson (1952). For a recent review, see Huber (2012).

The key assumption underlying IPW methods is that controlling for a set of observed variables, we are able to retrieve the full distribution of an outcome of interest. In that sense, IPW methods rely on matching on observed variables and can be stated as:

Assumption A-2. $Y \perp\!\!\!\perp A|(D, X, Z)$,

where Z are pre-programme variables other than X, and A is an attrition indicator $A = (A_i : i \in \mathcal{I})$ which takes value $A_i = 1$ if participant i has non-missing data on outcome Y and $A_i = 0$ otherwise. If Assumption **A-2** holds, then $E(Y|D, X, Z, A) = E(Y|D, X, Z)$. Moreover, $D \perp\!\!\!\perp Y(d)|X, Z$ holds due to randomisation, and the Average Treatment Effect (ATE) can be evaluated by:

$$\begin{aligned} E(Y_i(1) - Y_i(0)) &= \int E(Y_i|D_i = 1, X_i = x, Z_i = z) - E(Y_i|D_i = 0, X_i = x, Z_i = z) dF_{X,Z}(x, z), \\ &= \int E(Y_i|D = 1, A_i = 1, X_i = x, Z_i = z) - E(Y_i|D_i = 0, A_i = 1, X_i = x, Z_i = z) dF_{X,Z}(x, z), \end{aligned} \tag{D.4}$$

where the last equation holds by Assumption **A-2**. ATE is identified as $E(Y|D, A = \mathbf{1}, X, Z)$ and the distribution of (X, Z) are observed from data. The standard IPW formula for ATE can be

obtained by applying Bayes' theorem to Equation D.4:

$$E\left(\frac{Y_i 1[A_i=1, D_i=1]}{\Pr(A_i=1|D_i=1, X_i, Z_i)\Pr(D_i=1|X_i, Z_i)} - \frac{Y_i 1[A_i=1, D_i=0]}{\Pr(A_i=1|D_i=0, X_i, Z_i)\Pr(D_i=0|X_i, Z_i)}\right), \tag{D.5}$$

where $\Pr(\cdot)$ means probability. The implementation of expression (D.5) as developed by Johnston and DiNardo (1997) is given by:

$$\widehat{ATE} = \sum_{i=1}^{N} \frac{Y_i \cdot 1[D_i=1] \cdot 1[A_i=1] \cdot \omega_{i,1}}{N_1} - \sum_{i=1}^{N} \frac{Y_i \cdot 1[D_i=0] \cdot 1[A_i=1] \cdot \omega_{i,0}}{N_0} \tag{D.6}$$

$$\text{where: } \omega_{i,d} = \frac{1}{\hat{p}_{i,d}} \Bigg/ \left(\frac{1}{N_d}\sum_{j=1}^{N} \frac{1[D_i=d] \cdot 1[A_i=1]}{\hat{p}_{j,d}}\right) \qquad d \in \{0,1\},$$

$$\text{and: } p_{i,d} = \Pr(A=1|D=d, X_i, Z_i)\Pr(D=d|X_i, Z_i) \qquad d \in \{0,1\},$$

$$\text{where: } N_d = \sum_{i=1}^{N} 1[D_i=d] \cdot 1[A_i=1] \qquad d \in \{0,1\},$$

where N is the total sample size and $\hat{p}_{i,d}$ is an estimate for $p_{i,d}$. Weights $\omega_{i,d}$ are set such that their sum adds to the available sample size of the respective treatment, that is, $\sum_{i=1}^{N} \omega_{i,d} \cdot 1[D_i = d] \cdot 1[A_i = 1] = N_d$.

For the ABC intervention, probabilities $\hat{p}_{i,d}$ are estimated using the logit model based on gender- and wave-specific covariates, as in Campbell *et al.* (2014). For the Perry intervention, probabilities $\hat{p}_{i,d}$ are estimated using the following five variables: child Stanford-Binet IQ at entry; father present at home; number of siblings; family social economic status (SES) as measured by the cultural deprivation scale; and gender.[16] We use the weighted pre-pivoted t-statistics between treatment and control groups as test statistics. Small sample IPW inference is done by recalculating these probabilities for each permutation draw.

[16]Even though the matching assumption is not testable, we can do inference whether the choice of model used in the implementation of the IPW method generates reliable results. These tests are called balancing tests (Lee, 2013). We performed these tests on the set of variables used in our analysis.

E Results

In this section we report estimates of the treatment effects and associated p-values for health and healthy behaviours of both interventions, using the methodology described in the paper. We use the methodology previously applied by Heckman *et al.* (2010) to the Perry Preschool Project and by Campbell *et al.* (2014) to the Abecedarian Project. Table A3 shows estimates for the males in PPP, and Table A4 shows the corresponding results for females. Table A5 shows estimates and p-values for the males in ABC, and Table A6 shows the corresponding results for females. For each of the four tables, the outcomes are grouped according to the following categories: physical health; health insurance; demand for health care; diet and physical activity; smoking; and drinking. Table A7 presents tests for the equality of the treatment effects in comparable outcomes across ABC and PPP (see Table A2 in Section C of this Appendix for a list of measures that are comparable).

Table A3: Perry Preschool Intervention - Males

Variable	# C	# T	Ctr. M.	Treat. M.	Diff. Ms.	Asy. p-val.	Naive p-val.	Blk. Per. p-val.	Blk. Per. S.D.	Blk. IPW p. p-val.	Blk. IPW p. S.D.	Gen. Diff. p-val.	Gen. Diff. S.D.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Physical Health at 40 y.o.												
Excellent or very good health	36	30	0.639	0.700	0.061	0.303	0.311	0.235	0.235	0.166	0.166	0.752	0.752
Health stopped from working	36	30	0.528	0.433	0.094	0.226	0.226	0.387	0.450	0.276	0.319	0.669	0.888
Self-reported weight	35	30	88.619	91.459	2.840	0.699	0.698	0.723	0.993	0.672	0.988	0.420	0.617
Self-reported height	35	30	1.792	1.778	0.014	0.762	0.761	0.849	0.994	0.831	0.995	0.826	0.826
BMI	35	30	27.426	28.720	1.294	0.829	0.828	0.905	0.905	0.860	0.860	0.375	0.632
Overweight (BMI≥25)	35	30	0.714	0.800	0.086	0.784	0.798	0.896	0.896	0.859	0.859	0.484	0.722
Obese (BMI≥30)	35	30	0.257	0.333	0.076	0.746	0.762	0.923	0.981	0.871	0.960	0.537	0.537
	Health Insurance at 40 y.o.												
Health care coverage	35	30	0.743	0.800	0.057	0.296	0.313	0.367	0.367	0.376	0.376	0.541	0.771
Employer-provided or bought	36	30	0.361	0.567	0.206	**0.049**	**0.056**	0.102	0.145	0.103	0.145	0.688	0.688
Provided in prison	36	30	0.222	0.100	0.122	**0.095**	0.105	**0.078**	0.174	**0.082**	0.187	0.281	0.593
	Demand for Health Care at 40 y.o.												
Hospitalised	35	30	0.200	0.133	0.067	0.241	0.246	0.119	0.203	0.136	0.237	0.149	0.272
Scheduled treatment or exam	35	30	0.171	0.167	0.005	0.480	0.492	0.515	0.515	0.543	0.543	0.171	0.171
	Lifestyles - Diet and Physical Activity												
Physical activity at 40 y.o	35	30	0.457	0.367	0.090	0.766	0.779	0.584	0.584	0.545	0.545	**0.024**	**0.048**
Healthy Diet at 40 y.o.	35	29	0.229	0.379	0.151	**0.097**	0.113	**0.015**	**0.033**	**0.020**	**0.072**	0.982	0.982
	Lifestyles - Smoking												
Not a daily smoker at 27 y.o.	39	31	0.462	0.581	0.119	0.164	0.160	**0.092**	0.092	**0.089**	**0.089**	0.977	0.977
Not a heavy smoker at 27 y.o.	39	31	0.615	0.903	0.288	**0.003**	**0.002**	**0.004**	**0.005**	**0.004**	**0.005**	**0.031**	**0.066**
No. of cigarettes at 27 y.o.	39	31	8.744	4.291	4.453	**0.011**	**0.010**	**0.008**	**0.009**	**0.006**	**0.011**	0.189	0.272
Never smoker at 40 y.o.	36	30	0.444	0.600	0.156	0.107	0.109	**0.042**	**0.042**	**0.040**	**0.040**	0.589	0.589
Not a daily smoker at 40 y.o.	36	30	0.472	0.667	0.194	**0.058**	**0.063**	**0.014**	**0.042**	**0.010**	**0.035**	0.500	0.833
Not a heavy smoker at 40 y.o.	35	28	0.743	0.929	0.186	**0.027**	**0.027**	**0.013**	**0.023**	**0.011**	**0.021**	0.543	0.838
No. of cigarettes at 40 y.o.	35	28	6.543	3.714	2.829	**0.080**	**0.082**	**0.043**	**0.057**	**0.035**	**0.049**	0.557	0.766
	Lifestyles - Drinking												
Not a frequent drinker at 27 y.o.	39	30	0.718	0.800	0.082	0.220	0.223	0.138	0.138	0.120	0.120	0.869	0.869
Alcohol consumption at 27 y.o.	39	30	7.436	4.467	2.969	**0.074**	**0.064**	**0.026**	**0.036**	**0.024**	**0.040**	0.374	0.498
Not a frequent drinker at 40 y.o.	35	29	0.943	0.897	0.046	0.750	0.785	0.442	0.442	0.528	0.528	0.954	0.954
Alcohol consumption at 40 y.o.	35	29	2.600	3.034	0.434	0.618	0.620	0.339	0.387	0.366	0.385	0.460	0.603

Notes: This table presents the inference results for selected outcomes of the Perry intervention, male sample. p-values ≤ 0.10 are printed in boldface. The columns present the following information: (1) describes the variable of interest; (2) displays the sample size for the control group; (3) displays the sample size for the treatment group; (4) displays the control mean; (5) displays the treatment mean; (6) displays the unconditional difference in means between treatment and control groups (absolute value); (7) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the unconditional difference in means. The remaining columns present permutation p-values based on 30,000 draws. (8) displays the single hypothesis one-sided naive permutation p-value (by naive we mean based on an unconstrained permutation scheme); (9) displays the one-sided single hypothesis constrained permutation p-value based on the t-statistic associated with the difference in means between treatment groups (by constrained permutation we mean that permutations are done within strata defined by the pre-programme variables used in the randomisation protocol: gender, cohort indicator, the median of the cultural deprivation scale, child IQ at entry, and mother employment status. More specifically, we simulate the pairwise matching defined in the randomisation protocol using these variables and permute the treatment status within matched participants). (10) displays the multiple-hypothesis testing (stepdown) p-values associated with (9). The multiple-hypothesis testing is applied to blocks of outcomes indicated by italicised headings. (11) displays the one-sided single-hypothesis constrained permutation p-value based on the Inverse Probability Weighting (IPW) t-statistic associated with the difference in means between treatment groups. Probabilities of IPW are estimated using the following variables: gender, presence of the father in the home at entry, cultural deprivation scale, child IQ at entry (Stanford-Binet), number of siblings, and maternal employment status. (12) displays the multiple-hypothesis testing (stepdown) p-values associated with (11). The multiple-hypothesis testing is applied to blocks of outcomes indicated by italicised headings. (13) displays the double-sided single hypothesis p-value for the test of gender differences in the treatment effects. (14) displays the double-sided multiple hypothesis testing (stepdown) p-value associated with (13). Ctr. or C=Control; Treat. or T=Treatment; M.=Mean; Ms.=Means; Diff.=Difference; Gen.=Gender; Asy.=Asymptotic; Blk.=Block; Per.=Permutation; p-val.=p-value; S.D.=Stepdown; y.o.=years old.

Table A4: Perry Preschool Intervention - Females

Variable	# C	# T	Ctr. M.	Treat. M.	Diff. Ms.	Asy. p-val.	Naive p-val.	Blk. Per. p-val.	Blk. Per. S.D.	Blk. IPW p. p-val.	Blk. IPW p. S.D.	Gen. Diff. p-val.	Gen. Diff. S.D.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Physical Health at 40 y.o.												
Excellent or very good health	22	24	0.500	0.500	0.000	0.500	0.555	0.459	0.459	0.342	0.342	0.752	0.752
Health stopped from working	22	24	0.591	0.417	0.174	0.123	0.155	0.225	0.249	0.198	0.210	0.669	0.888
Self-reported weight	20	18	72.665	68.669	3.997	0.165	0.161	0.223	0.929	0.248	0.880	0.420	0.617
Self-reported height	22	24	1.648	1.625	0.023	0.849	0.844	1.000	1.000	1.000	1.000	0.826	0.826
BMI	20	18	26.829	26.283	0.546	0.367	0.366	0.886	0.991	0.886	0.989	0.375	0.632
Overweight (BMI≥25)	20	18	0.650	0.611	0.039	0.405	0.416	0.781	0.781	0.753	0.936	0.484	0.722
Obese (BMI≥30)	20	18	0.200	0.167	0.033	0.399	0.420	0.903	0.948	0.884	0.884	0.537	0.537
	Health Insurance at 40 y.o.												
Health care coverage	22	24	0.909	0.875	0.034	0.641	0.639	0.113	0.113	0.133	0.133	0.541	0.771
Employer-provided or bought	22	24	0.455	0.583	0.129	0.197	0.219	**0.052**	**0.052**	**0.055**	**0.055**	0.688	0.688
Provided in prison	22	24	0.000	0.000	0.000	-	-	-	-	-	-	0.281	0.593
	Demand for Health Care at 40 y.o.												
Hospitalised	22	24	0.136	0.292	0.155	0.895	0.898	0.804	0.913	0.785	0.904	0.149	0.272
Scheduled treatment or exam	22	24	0.091	0.292	0.201	0.955	0.963	0.819	0.819	0.796	0.796	0.171	0.171
	Lifestyles - Diet and Physical Activity												
Physical activity at 40 y.o.	22	24	0.045	0.375	0.330	**0.003**	**0.003**	**0.002**	**0.005**	**0.002**	**0.012**	**0.024**	**0.048**
Healthy Diet at 40 y.o.	22	24	0.227	0.375	0.148	0.143	0.144	0.238	0.238	0.283	0.283	0.982	0.982
	Lifestyles - Smoking												
Not a daily smoker at 27 y.o.	22	25	0.409	0.520	0.111	0.229	0.221	**0.091**	0.201	0.110	0.277	0.977	0.977
Not a heavy smoker at 27 y.o.	22	25	0.818	0.760	0.058	0.682	0.692	0.673	0.673	0.662	0.662	**0.031**	**0.066**
No. of cigarettes at 27 y.o.	22	25	7.682	7.600	0.082	0.489	0.496	0.281	0.456	0.297	0.482	0.189	0.272
Never smoker at 40 y.o.	22	24	0.409	0.458	0.049	0.372	0.379	0.103	0.103	0.137	0.504	0.589	0.589
Not a daily smoker at 40 y.o.	22	23	0.455	0.522	0.067	0.330	0.317	0.156	0.416	0.206	0.472	0.500	0.833
Not a heavy smoker at 40 y.o.	22	23	0.773	0.870	0.097	0.203	0.225	0.356	0.397	0.387	0.436	0.543	0.838
No. of cigarettes at 40 y.o.	22	23	6.818	5.870	0.949	0.360	0.370	0.427	0.440	0.486	0.486	0.557	0.766
	Lifestyles - Drinking												
Not a frequent drinker at 27 y.o.	22	25	0.773	0.880	0.107	0.169	0.193	**0.004**	**0.019**	**0.015**	**0.028**	0.869	0.869
Alcohol consumption at 27 y.o.	22	25	3.818	3.200	0.618	0.314	0.320	**0.085**	**0.085**	**0.094**	**0.094**	0.374	0.498
Not a frequent drinker at 40 y.o.	22	23	0.909	0.870	0.040	0.659	0.663	0.600	0.600	0.698	0.698	0.954	0.954
Alcohol consumption at 40 y.o.	22	23	4.227	2.826	1.401	0.248	0.256	0.406	0.406	0.467	0.469	0.460	0.603

Notes: This table presents the inference results for selected outcomes of the Perry intervention, female sample. p-values $\leq$ 0.10 are printed in boldface. The columns present the following information: (1) describes the variable of interest; (2) displays the sample size for the control group; (3) displays the sample size for the treatment group; (4) displays the control mean; (5) displays the treatment mean; (6) displays the unconditional difference in means between treatment and control groups (absolute value); (7) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the unconditional difference in means. The remaining columns present permutation p-values based on 30,000 draws. (8) displays the single hypothesis one-sided naive permutation p-value (by naive we mean based on an unconstrained permutation scheme); (9) displays the one-sided single hypothesis constrained permutation p-value based on the t-statistic associated with the difference in means between treatment groups (by constrained permutation we mean that permutations are done within strata defined by the pre-programme variables used in the randomisation protocol: gender, cohort indicator, the median of the cultural deprivation scale, child IQ at entry, and mother employment status. More specifically, we simulate the pairwise matching defined in the randomisation protocol using these variables and permute the treatment status within matched participants). (10) displays the multiple-hypothesis testing (stepdown) p-values associated with (9). The multiple-hypothesis testing is applied to blocks of outcomes indicated by horizontal lines. (11) displays the one-sided single hypothesis constrained permutation p-value based on the Inverse Probability Weighting (IPW) t-statistic associated with the difference in means between treatment groups. Probabilities of IPW are estimated using the following variables: gender, presence of the father in the home at entry, cultural deprivation scale, child IQ at entry (Stanford-Binet), number of siblings, and maternal employment status. (12) displays the multiple-hypothesis testing (stepdown) p-values associated with (11). The multiple-hypothesis testing is applied to blocks of outcomes indicated by italicised headings. (13) displays the double-sided single-hypothesis p-value for the test of gender differences in the treatment effects. (14) displays the double-sided multiple-hypothesis testing (stepdown) p-value associated with (13). Ctr. or C=Control; Treat. or T=Treatment; M.=Mean; Ms.=Means; Diff.=Difference; Gen.=Gender; Asy.=Asymptotic; Blk.=Block; Per.=Permutation; p-val.=p-value; S.D.=Stepdown; y.o.=years old.

Table A5: Abecedarian Intervention - Males

Variable	# C	# T	Ctr. M.	Treat. M.	Diff. Ms.	Asy. p-val.	Naive p-val.	Blk. Per. p-val.	Blk. Per. S.D.	Blk. IPW p. p-val.	Blk. IPW p. S.D.	Gen. Diff. p-val.	Gen. Diff. S.D.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Physical Health in the 30s												
Excellent or very good health	21	27	0.667	0.815	0.148	0.124	0.128	0.137	0.240	0.113	0.205	0.530	0.530
Health stopped from working	21	27	0.048	0.111	0.063	0.780	0.731	0.547	0.547	0.545	0.545	0.184	0.333
Measured weight	9	18	100.647	93.797	6.850	0.242	0.274	0.320	0.320	0.154	0.303	0.412	0.412
Measured height	9	18	1.739	1.790	0.050	**0.044**	**0.061**	**0.083**	0.187	0.215	0.215	0.199	0.351
BMI	8	18	33.293	29.217	4.075	**0.076**	0.108	0.141	0.175	**0.093**	0.204	0.187	0.373
Overweight (BMI $\geq$ 25)	8	18	0.750	0.722	0.028	0.444	0.455	0.391	0.466	0.234	0.234	0.834	0.969
Obese (BMI $\geq$ 30)	8	18	0.625	0.556	0.069	0.376	0.378	0.448	0.448	0.227	0.335	0.972	0.972
Diastolic blood pressure	9	19	92.000	78.526	13.474	**0.017**	**0.046**	**0.075**	**0.075**	**0.025**	**0.025**	0.245	0.333
Systolic blood pressure	9	19	143.333	125.789	17.544	**0.022**	**0.059**	**0.057**	**0.085**	**0.019**	**0.031**	0.280	0.280
Hypertension I	9	19	0.444	0.105	0.339	**0.019**	**0.043**	**0.063**	**0.063**	**0.010**	**0.018**	0.268	0.268
Hypertension II	9	19	0.556	0.211	0.345	**0.033**	**0.049**	**0.061**	**0.095**	**0.037**	**0.037**	**0.078**	0.123
	Health Insurance at 30 y.o.												
Health care coverage	21	27	0.476	0.704	0.228	**0.057**	**0.062**	**0.080**	**0.080**	**0.040**	**0.040**	**0.075**	0.131
Employer-provided or bought	21	27	0.333	0.444	0.296	**0.021**	**0.018**	**0.034**	**0.048**	**0.035**	**0.055**	0.200	0.200
	Demand for Health Care in the 30s												
Hospitalised	9	19	0.556	0.211	0.345	**0.033**	**0.039**	**0.042**	**0.042**	**0.100**	**0.100**	**0.043**	**0.087**
Scheduled treatment or exam	21	27	0.476	0.222	0.254	**0.033**	**0.040**	**0.026**	**0.051**	**0.043**	**0.080**	0.127	0.127
	Lifestyles - Diet and Physical Activity at 21 y.o.												
Physical activity	23	26	0.391	0.308	0.084	0.726	0.733	0.811	0.811	0.866	0.866	**0.057**	0.110
# Fruit servings	23	26	0.826	0.846	0.020	0.470	0.469	0.437	0.618	0.524	0.745	0.137	0.137
	Lifestyles - Smoking at 30 y.o.												
Never a regular smoker	20	27	0.500	0.444	0.056	0.644	0.632	0.589	0.702	0.563	0.691	0.185	0.424
Not a daily smoker	20	27	0.650	0.556	0.094	0.738	0.732	0.664	0.664	0.564	0.564	0.493	0.753
Not a heavy smoker	20	27	0.900	0.889	0.011	0.547	0.525	0.508	0.771	0.419	0.677	0.707	0.707
No. of cigarettes	20	27	3.125	3.611	0.486	0.625	0.626	0.596	0.753	0.450	0.642	0.664	0.817
Age of onset of smoking	14	18	16.893	19.722	2.829	**0.017**	**0.019**	**0.033**	0.101	**0.055**	0.185	**0.045**	0.164
	Lifestyles - Drinking at 30 y.o.												
Not a frequent drinker	20	27	0.850	0.778	0.072	0.728	0.735	0.595	0.697	0.532	0.612	0.522	0.522
Alcohol consumption	20	27	4.150	6.759	2.609	0.850	0.855	0.781	0.781	0.663	0.663	0.332	0.430
Age of onset of drinking $<$ 17	23	26	0.609	0.538	0.070	0.314	0.317	0.302	0.546	0.329	0.567	0.263	0.511

Notes: This table presents the inference results for selected outcomes of the Abecedarian intervention, male sample. p-values $\leq$ 0.10 are printed in boldface. The columns present the following information: (1) describes the variable of interest; (2) displays the sample size for the control group; (3) displays the sample size for the treatment group; (4) displays the control mean; (5) displays the treatment mean; (6) displays the unconditional difference in means between treatment and control groups (absolute value); (7) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the unconditional difference in means. The remaining columns present permutation p-values based on 30,000 draws. (8) displays the single hypothesis one-sided naive permutation p-value (by naive we mean based on an unconstrained permutation scheme); (9) displays the one-sided single-hypothesis constrained permutation p-value based on the t-statistic associated with the difference in means between treatment groups (by constrained permutation we mean that permutations are done within strata defined by the pre-programme variables used in the randomisation protocol: gender, cohort indicator, number of siblings, high risk index at birth, and mother WAIS full IQ score. More specifically, we simulate the pairwise matching defined in the randomisation protocol using these variables and permute the treatment status within matched participants). (10) displays the multiple-hypothesis testing (stepdown) p-values associated with (9). The multiple-hypothesis testing is applied to blocks of outcomes indicated by italicised headings. (11) displays the one-sided single-hypothesis constrained permutation p-value based on the Inverse Probability Weighting (IPW) t-statistic associated with the difference in means between treatment groups. Probabilities of IPW are estimated using gender- and wave-specific covariates. See Campbell *et al.* (2014) for details. (12) displays the multiple-hypothesis testing (stepdown) p-values associated with (11). The multiple-hypothesis testing is applied to blocks of outcomes indicated by italicised headings. (13) displays the double-sided single hypothesis p-value for the test of gender differences in the treatment effects. (14) displays the double-sided multiple-hypothesis testing (stepdown) p-value associated with (13). Ctr. or C=Control; Treat. or T=Treatment; M.=Mean; Ms.=Means; Diff.=Difference; Gen.=Gender; Asy.=Asymptotic; Blk.=Block; Per.=Permutation; p-val.=p-value; S.D.=Stepdown; y.o.=years old.

Table A6: Abecedarian Intervention - Females

Variable	# C	# T	Ctr. M.	Treat. M.	Diff. Ms.	Asy. p-val.	Naive p-val.	Blk. Per. p-val.	Blk. Per. S.D.	Blk. IPW p. p-val.	Blk. IPW p. S.D.	Gen. Diff. p-val.	Gen. Diff. S.D.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Physical Health in the 30s												
Excellent or very good health	28	25	0.536	0.560	0.024	0.431	0.427	0.477	0.477	0.355	0.355	0.530	0.530
Health stopped from working	28	25	0.071	0.000	0.071	**0.089**	0.174	0.186	0.354	0.193	0.332	0.184	0.333
Measured weight	22	18	92.143	95.864	3.721	0.665	0.649	0.889	0.912	0.899	0.918	0.412	0.412
Measured height	22	18	1.642	1.635	0.007	0.629	0.631	0.542	0.819	0.691	0.924	0.199	0.351
BMI	22	18	34.081	35.866	1.785	0.714	0.692	0.902	0.902	0.923	0.923	0.187	0.373
Overweight (BMI $\geq$ 25)	22	18	0.955	0.889	0.066	0.222	0.179	0.388	0.600	0.482	0.685	0.834	0.969
Obese (BMI $\geq$ 30)	22	18	0.727	0.667	0.061	0.343	0.343	0.715	0.715	0.788	0.788	0.972	0.972
Diastolic blood pressure	22	18	89.227	85.333	3.894	0.233	0.243	0.338	0.338	0.452	0.452	0.245	0.333
Systolic blood pressure	22	18	135.636	129.667	5.970	0.180	0.187	0.218	0.285	0.307	0.385	0.280	0.280
Hypertension I	22	18	0.318	0.222	0.096	0.255	0.263	0.263	0.364	0.380	0.499	0.268	0.268
Hypertension II	22	18	0.409	0.500	0.091	0.712	0.709	0.628	0.628	0.721	0.721	**0.078**	0.123
	Health Insurance at 30 y.o.												
Health care coverage	28	25	0.857	0.760	0.097	0.812	0.813	0.928	0.928	0.945	0.945	**0.075**	0.131
Employer-provided or bought	28	25	0.357	0.400	0.043	0.377	0.386	0.499	0.691	0.512	0.706	0.200	0.200
	Demand for Health Care in the 30s												
Hospitalised	22	18	0.136	0.222	0.086	0.756	0.746	0.695	0.695	0.422	0.692	**0.043**	**0.087**
Scheduled treatment or exam	28	25	0.393	0.440	0.047	0.633	0.638	0.619	0.888	0.538	0.538	0.127	0.127
	Lifestyles - Diet and Physical Activity at 21 y.o.												
Physical activity	28	25	0.071	0.320	0.249	**0.010**	**0.013**	**0.009**	**0.009**	**0.004**	**0.004**	**0.057**	0.110
# Fruit servings	28	25	0.286	0.800	0.514	**0.005**	**0.009**	**0.002**	**0.004**	**0.003**	**0.006**	0.137	0.137
	Lifestyles - Smoking at 30 y.o.												
Never a regular smoker	28	25	0.429	0.640	0.211	**0.064**	**0.056**	**0.082**	0.272	**0.077**	0.245	0.185	0.424
Not a daily smoker	28	25	0.679	0.720	0.041	0.374	0.358	0.365	0.583	0.394	0.717	0.493	0.753
Not a heavy smoker	28	25	0.929	0.960	0.031	0.314	0.397	0.293	0.627	0.447	0.704	0.707	0.707
No. of cigarettes	28	25	2.179	1.860	0.319	0.387	0.388	0.334	0.631	0.477	0.691	0.664	0.817
Age of onset of smoking	18	10	17.861	17.050	0.811	0.755	0.771	0.850	0.850	0.845	0.845	**0.045**	0.164
	Lifestyles - Drinking at 30 y.o.												
Not a frequent drinker	28	25	0.857	0.880	0.023	0.405	0.414	0.493	0.586	0.547	0.547	0.522	0.522
Alcohol consumption	28	25	3.536	3.180	0.356	0.422	0.430	0.536	0.536	0.516	0.586	0.332	0.430
Age of onset of drinking $<$ 17	28	25	0.571	0.280	0.291	**0.016**	**0.018**	**0.023**	**0.061**	**0.009**	**0.023**	0.263	0.511

Notes: This table presents the inference results for selected outcomes of the Abecedarian intervention, female sample. p-values $\leq$ 0.10 are printed in boldface. The columns present the following information: (1) describes the variable of interest; (2) displays the sample size for the control group; (3) displays the sample size for the treatment group; (4) displays the control mean; (5) displays the treatment mean; (6) displays the unconditional difference in means between treatment and control groups (absolute value); (7) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the unconditional difference in means. The remaining columns present permutation p-values based on 30,000 draws. (8) displays the single hypothesis one-sided naive permutation p-value (by naive we mean based on an unconstrained permutation scheme); (9) displays the one-sided single hypothesis constrained permutation p-value based on the t-statistic associated with the difference in means between treatment groups (by constrained permutation we mean that permutations are done within strata defined by the pre-programme variables used in the randomisation protocol: gender, cohort indicator, number of siblings, high risk index at birth, and mother WAIS full IQ score. More specifically, we simulate the pairwise matching defined in the randomisation protocol using these variables and permute the treatment status within matched participants). (10) displays the multiple-hypothesis testing (stepdown) p-values associated with (9). The multiple-hypothesis testing is applied to blocks of outcomes indicated by italicised headings. (11) displays the one-sided single hypothesis constrained permutation p-value based on the Inverse Probability Weighting (IPW) t-statistic associated with the difference in means between treatment groups. Probabilities of IPW are estimated using gender- and wave-specific covariates. See Campbell *et al.* (2014) for details. (12) displays the multiple-hypothesis testing (stepdown) p-values associated with (11). The multiple-hypothesis testing is applied to blocks of outcomes indicated by italicised headings. (13) displays the double-sided single hypothesis p-value for the test of gender differences in the treatment effects. (14) displays the double-sided multiple-hypothesis testing (stepdown) p-value associated with (13). Ctr. or C=Control; Treat. or T=Treatment; M.=Mean; Ms.=Means; Diff.=Difference; Gen.=Gender; Asy.=Asymptotic; Blk.=Block; Per.=Permutation; p-val.=p-value; S.D.=Stepdown; y.o.=years old.

Table A7: Equality of the treatment effects between ABC and PPP

	PPP				ABC				Difference ABC-PPP			
Variable	**Unc. TE**	***p*-val.**	**Cond. TE**	***p*-val.**	**Unc. TE**	***p*-val.**	**Cond. TE**	***p*-val.**	**Unc. TE**	***p*-val.**	**Cond. TE**	***p*-val.**
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Males												
Excellent or very good health	0.061	0.296	0.066	0.297	0.148	0.135	0.139	0.159	0.087	0.622	0.073	0.696
Weight	2.840	0.695	4.475	0.772	-6.850	0.227	-8.237	0.198	-9.691	0.367	-12.712	0.266
Height	-0.014	0.767	-0.015	0.761	0.050	**0.053**	0.052	**0.062**	0.064	**0.080**	0.066	**0.094**
BMI	1.294	0.814	1.714	0.865	-4.075	**0.080**	-4.459	**0.043**	-5.369	**0.063**	-6.174	**0.040**
Overweight	0.086	0.781	0.094	0.787	-0.028	0.441	-0.120	0.270	-0.113	0.604	-0.214	0.351
Obese	0.076	0.740	0.071	0.709	-0.069	0.366	-0.072	0.369	-0.146	0.536	-0.144	0.569
Not a frequent drinker	-0.046	0.710	-0.074	0.792	-0.072	0.769	-0.060	0.721	-0.026	0.841	0.014	0.917
Alcohol consumption	0.434	0.597	0.374	0.578	2.609	0.895	2.455	0.875	2.175	0.426	2.080	0.467
Females												
Excellent or very good health	0.000	0.500	0.130	0.210	0.024	0.431	0.063	0.333	0.024	0.906	-0.067	0.759
Weight	-3.997	0.283	-2.750	0.365	3.721	0.707	5.623	0.769	7.717	0.430	8.373	0.449
Height	-0.023	0.867	-0.023	0.854	-0.007	0.618	-0.005	0.586	0.016	0.590	0.018	0.571
BMI	-0.546	0.415	-0.100	0.486	1.785	0.761	2.466	0.812	2.331	0.517	2.566	0.525
Overweight	-0.039	0.381	0.055	0.657	-0.066	0.301	-0.061	0.322	-0.027	0.882	-0.116	0.542
Obese	-0.033	0.407	-0.054	0.358	-0.061	0.331	0.060	0.666	-0.027	0.891	0.114	0.576
Not a frequent drinker	-0.039	0.653	-0.130	0.888	0.023	0.402	-0.059	0.730	0.062	0.647	0.071	0.622
Alcohol consumption	-1.401	0.241	0.074	0.513	-0.356	0.423	0.687	0.637	1.045	0.700	0.613	0.833

Notes: This table presents tests for the equality of the treatment effects for the outcomes with high degree of comparability (see Table A2), PPP and ABC samples. p-values ≤ 0.10 are printed in boldface. The upper panel presents the results for males, the lower panel presents the results for females. The columns present the following information: (1) describes the variable of interest; (2) displays the unconditional treatment effect for PPP; (3) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the unconditional treatment effect in (2); (4) displays the conditional treatment effect for PPP, where the conditioning set is the same as in Tables A3 and A4; (5) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the conditional treatment effect in (4); (6) displays the unconditional treatment effect for ABC; (7) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the unconditional treatment effect in (6); (8) displays the conditional treatment effect for ABC, where the conditioning set is the same as in Table A5; (9) displays the asymptotic p-value for the one-sided single hypothesis based on the t-statistic associated with the conditional treatment effect in (8); (10) displays the difference in the unconditional treatment effects between ABC and PPP; (11) displays the asymptotic p-value for the two-sided single hypothesis based on the t-statistic associated with the difference in the unconditional treatment effects in (10); (12) displays the difference in the conditional treatment effects between ABC and PPP; (13) displays the asymptotic p-value for the two-sided single hypothesis based on the t-statistic associated with the difference in the conditional treatment effects in (12). The outcomes 'Not a frequent drinker' and 'Alcohol consumption' for PPP refer to age 40 (see Table A2). Unc.=Unconditional; Cond.=Conditional; TE=Treatment Effect.

F Mediation Analyses

This section presents specification tests and the complete results for the dynamic and static mediation analyses described in Section 2.4 and reported in Section 3.2 of the paper. Table A8 shows the specification tests for the Perry Preschool intervention, following Heckman *et al.* (2013) (in particular, Appendix L). Table A9 shows the results of the dynamic mediation analysis of the Perry Preschool intervention for males. Table A10 shows the results of the dynamic mediation analysis of the Perry Preschool intervention for females. Table A11 shows the results of the static mediation analysis of the Perry Preschool intervention for males. Table A12 shows the specification tests for the Abecedarian intervention, following the procedures in Heckman *et al.* (2013). Table A13 shows the results of the dynamic mediation analysis of the Abecedarian intervention for males. Table A14 shows the results of the static mediation analysis of the Abecedarian intervention for males and females separately.

Table A8: Perry Preschool Intervention - Specification Tests

Outcome	**Test**						
	Dynamic Mediation			*Static Mediation*			
				Early Inputs		*Late Inputs*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Males						
Not a daily smoker at 27 y.o.							
Test statistic	0.42	1.96	0.28	1.00	0.15	3.50	0.88
p-value	(0.742)	(0.133)	(0.758)	(0.399)	(0.865)	**(0.021)**	(0.422)
Not a heavy smoker at 27 y.o.							
Test statistic	0.90	0.41	0.05	0.62	0.17	0.09	0.13
p-value	(0.450)	(0.747)	(0.949)	(0.604)	(0.844)	(0.965)	(0.875)
No. of cigarettes at 27 y.o.							
Test statistic	0.66	0.69	0.18	0.42	0.14	0.83	0.38
p-value	(0.580)	(0.561)	(0.832)	(0.742)	(0.871)	(0.485)	(0.685)
Never smoker at 40 y.o.							
Test statistic	1.04	0.97	0.93	0.81	0.09	0.83	0.58
p-value	(0.384)	(0.418)	(0.402)	(0.496)	(0.914)	(0.484)	(0.565)
Not a daily smoker at 40 y.o.							
Test statistic	0.29	2.37	1.94	0.56	0.38	3.06	2.56
p-value	(0.830)	(0.085)	(0.157)	(0.646)	(0.687)	**(0.036)**	(0.087)
Not a heavy smoker at 40 y.o.							
Test statistic	0.48	0.33	0.11	0.81	0.38	0.12	0.19
p-value	(0.696)	(0.807)	(0.900)	(0.496)	(0.685)	(0.948)	(0.825)
No. of cigarettes at 40 y.o.							
Test statistic	0.22	0.68	0.58	0.25	0.36	0.77	1.06
p-value	(0.881)	(0.568)	(0.567)	(0.863)	(0.701)	(0.514)	(0.355)
	Females						
Physical activity at 40 y.o.							
Test statistic	0.07	1.06	0.54	0.37	0.58	0.63	0.11
p-value	(0.975)	(0.385)	(0.588)	(0.778)	(0.565)	(0.604)	(0.892)
Not a frequent drinker at 27 y.o.							
Test statistic	0.28	1.25	1.46	0.36	1.34	2.24	2.26
p-value	(0.838)	(0.311)	(0.252)	(0.782)	(0.277)	**(0.100)**	(0.119)
Alcohol consumption at 27 y.o.							
Test statistic	1.10	1.42	2.43	1.07	2.79	1.81	2.19
p-value	(0.366)	(0.259)	(0.108)	(0.374)	(**0.076**)	(0.163)	(0.127)

Notes: This table presents Wald test statistics with p-values in parentheses for a number of specification tests of whether the regression coefficients in the outcome equations are the same for the treatment and the control group. p-values $\leq$ are printed in boldface. Specifically, we present tests for the following hypotheses: $H_0 : \boldsymbol{\alpha}_{C,1} = \boldsymbol{\alpha}_{C,0}$ in equation 9 (col. (1)), $H_0 : \boldsymbol{\alpha}_1^A = \boldsymbol{\alpha}_0^C$ in equation 9 (col. (2)), $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$ in equation 9 (col. (3)), $H_0 : \boldsymbol{\alpha}_{C,1} = \boldsymbol{\alpha}_{C,0}$ in equation 13 (col. (4)), $H_0 : \boldsymbol{\beta}_{C,1} = \boldsymbol{\beta}_{C,0}$ in equation 13 (col. (5)), $H_0 : \boldsymbol{\alpha}_1^A = \boldsymbol{\alpha}_0^C$ in equation 14 (col. (6)), $H_0 : \boldsymbol{\beta}_{A,1} = \boldsymbol{\beta}_{A,0}$ in equation 14 (col. (7)).

Table A9: Perry Preschool Intervention - Dynamic Mediation Results

Outcome	**Mediator**														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Males														
Not a daily smoker at 27 y.o.															
Share	0.018	0.481	-0.034	-0.033	-0.151	0.267	0.016	-0.017	0.008	0.011	-0.032	0.045	0.001	0.003	0.005
Inference on share (p-value)	0.341	**0.084**	0.732	0.623	0.825	0.207	0.244	0.752	0.232	0.170	0.843	0.186	0.375	0.238	0.235
Not a heavy smoker at 27 y.o.															
Share	0.013	0.078	0.011	0.009	0.021	0.006	-0.004	0.002	0.000	-0.003	0.004	0.001	-0.000	-0.000	0.000
Inference on share (p-value)	0.311	0.153	0.275	0.392	0.260	0.467	0.688	0.278	0.441	0.857	0.226	0.414	0.604	0.685	0.453
No. of cigarettes at 27 y.o.															
Share	0.025	0.167	-0.009	-0.006	-0.014	0.055	0.003	-0.002	0.002	0.002	-0.003	0.009	0.000	0.000	0.001
Inference on share (p-value)	0.271	**0.075**	0.698	0.603	0.660	0.317	0.265	0.630	0.260	0.243	0.731	0.255	0.385	0.320	0.271
Never smoker at 40 y.o.															
Share	0.038	0.349	-0.008	-0.102	-0.301	0.117	0.049	-0.034	0.003	0.034	-0.064	0.020	0.002	0.007	0.002
Inference on share (p-value)	0.304	**0.075**	0.589	0.639	0.877	0.295	0.213	0.797	0.234	0.134	0.886	0.225	0.399	0.212	0.271
Not a daily smoker at 40 y.o.															
Share	0.036	0.293	-0.045	-0.083	-0.270	0.151	0.040	-0.030	0.005	0.028	-0.057	0.026	0.002	0.006	0.003
Inference on share (p-value)	0.311	**0.066**	0.809	0.638	0.887	0.257	0.192	0.804	0.207	0.113	0.900	0.191	0.394	0.210	0.234
Not a heavy smoker at 40 y.o.															
Share	0.011	0.200	-0.006	-0.003	0.157	-0.186	0.002	0.018	-0.006	0.001	0.033	-0.031	0.000	-0.004	-0.004
Inference on share (p-value)	0.369	**0.046**	0.660	0.559	**0.092**	0.888	0.302	0.171	0.789	0.342	0.094	0.887	0.447	0.803	0.777
No. of cigarettes at 40 y.o.															
Share	0.087	0.435	-0.075	-0.042	0.008	-0.082	0.020	0.001	-0.002	0.014	0.002	-0.014	0.001	-0.000	-0.002
Inference on share (p-value)	0.247	**0.052**	0.845	0.653	0.393	0.789	0.222	0.414	0.686	0.143	0.372	0.810	0.369	0.578	0.667

Notes: This table presents conditional decomposition results for the dynamic mediation analysis of the statistically significant outcomes for the Perry Preschool intervention. p-values ≤ 0.10 are printed in boldface. We use six mediators, three early childhood mediators: the Stanford-Binet IQ score for ages 7-9, the externalizing behaviour factor for ages 7-9, and the academic motivation factor for ages 7-9 (as in Heckman *et al.*, 2013); and three adult mediators: being a high school graduate by age 19, number of months unemployed in the last 2 years at age 27, and monthly income at age 27 adjusted to 2006 prices (as in Heckman *et al.*, 2010). The conditioning variables used are: father at home at entry and mother employed at entry. For each outcome we present two lines of results. The first line presents the decomposition share. Columns (1)-(3) display the shares which can be attributed to the direct health effects of experimentally-induced changes in childhood factors: $\alpha_j^C E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 9), where col. (1) refers to cognition, col. (2) to externalizing behaviour, and col. (3) to academic motivation. Columns (4)-(6) display the shares which can be attributed to the direct health effect of experimentally-induced changes in adulthood factors: $\alpha_j^A E\left(I_{i,j}^A(1) - I_{i,j}^A(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 9), where col. (4) refers to education, col. (5) to unemployment, and col. (6) to income. Columns (7)-(15) display the shares which can be attributed to the indirect effect of experimentally-induced changes in childhood factors affecting health through the adulthood factors: $\alpha_j^A \gamma^j E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equations 9 and 10). In particular, columns (7)-(9) refer to the effect of experimentally enhanced cognition through education (col. (7)), unemployment (col. (8)), and income (col. (9)); columns (10)-(12) refer to the effect of experimentally induced reductions in externalizing behaviour through education (col. (10)), unemployment (col. (11)), and income (col. (12)); and columns (13)-(15) refer to the effects of experimentally-increased academic motivation through education (col. (13)), unemployment (col. (14)), and income (col. (15)). The second line presents the one-sided p-value that tests if the share is statistically significantly different from zero, computed using the bootstrap method (1,000 replications).

Table A10: Perry Preschool Intervention - Dynamic Mediation Results (ctd.)

Outcome	**Mediator**														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Females														
Physical activity at 40 y.o.															
Share	0.286	-0.454	-0.034	-0.340	0.058	0.059	-0.029	-0.008	-0.007	-0.018	0.000	0.011	-0.002	0.014	0.024
Inference on share (p-value)	**0.095**	0.851	0.754	0.868	0.114	0.290	0.877	0.771	0.730	0.780	0.446	0.363	0.714	0.190	0.285
Not a frequent drinker at 27 y.o.															
Share	-0.099	0.163	0.320	-0.779	0.018	-0.019	-0.065	-0.003	0.002	-0.041	0.000	-0.004	-0.005	0.004	-0.008
Inference on share (p-value)	0.734	0.291	0.215	0.791	0.461	0.615	0.850	0.608	0.428	0.719	0.520	0.546	0.591	0.450	0.598
Alcohol consumption at 27 y.o.															
Share	0.027	0.376	0.212	-1.736	0.129	0.124	-0.146	-0.018	-0.016	-0.092	0.001	0.023	-0.011	0.031	0.052
Inference on share (p-value)	0.352	0.273	0.226	0.851	0.260	0.247	0.900	0.706	0.765	0.792	0.488	0.296	0.666	0.266	0.246

Notes: This table presents conditional decomposition results for the dynamic mediation analysis of the statistically significant outcomes for the Perry Preschool intervention. p-values ≤ 0.10 are printed in boldface. We use six mediators, three early childhood mediators: the Stanford-Binet IQ score for ages 7-9, the externalizing behaviour factor for ages 7-9, and the academic motivation factor for ages 7-9 (as in Heckman *et al.*, 2013); and three adult mediators: being a high school graduate by age 19, number of months unemployed in the last 2 years at age 27, and monthly income at age 27 adjusted to 2006 prices (as in Heckman *et al.*, 2010). The conditioning variables used are: father at home at entry and mother employed at entry. For each outcome we present two lines of results. The first line presents the decomposition share. Columns (1)-(3) display the shares which can be attributed to the direct health effects of experimentally induced changes in childhood factors: $\alpha_j^C E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 9), where col. (1) refers to cognition, col. (2) to externalizing behaviour, and col. (3) to academic motivation. Columns (4)-(6) display the shares which can be attributed to the direct health effect of experimentally-induced changes in adulthood factors: $\alpha_j^A E\left(I_{i,j}^A(1) - I_{i,j}^A(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 9), where col. (4) refers to education, col. (5) to unemployment, and col. (6) to income. Columns (7)-(15) display the shares which can be attributed to the indirect effect of experimentally induced changes in childhood factors affecting health through the adulthood factors: $\alpha_j^A \gamma^j E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equations 9 and 10). In particular, columns (7)-(9) refer to the effect of experimentally-enhanced cognition through education (col. (7)), unemployment (col. (8)), and income (col. (9)); columns (10)-(12) refer to the effect of experimentally induced reductions in externalizing behaviour through education (col. (10)), unemployment (col. (11)), and income (col. (12)); and columns (13)-(15) refer to the effects of experimentally-increased academic motivation through education (col. (13)), unemployment (col. (14)), and income (col. (15)). The second line presents the one-sided p-value that tests if the share is statistically significantly different from zero, computed using the bootstrap method (1,000 replications).

Table A11: Perry Preschool Intervention - Static Mediation Results

Outcome	**Mediator**					
	(1)	(2)	(3)	(4)	(5)	(6)
	Males					
Not a daily smoker at 27 y.o.						
Share	0.033	0.567	-0.030	0.036	0.844	-0.369
Inference on share (p-value)	0.306	**0.099**	0.720	0.226	0.118	0.863
Not a heavy smoker at 27 y.o.						
Share	0.017	0.089	0.007	-0.002	0.073	0.002
Inference on share (p-value)	0.276	0.151	0.325	0.581	0.259	0.433
No. of cigarettes at 27 y.o.						
Share	0.035	0.197	-0.012	0.008	0.184	-0.043
Inference on share (p-value)	0.256	**0.061**	0.707	0.232	0.158	0.722
Never smoker at 40 y.o.						
Share	0.049	0.318	0.007	0.135	0.780	-0.917
Inference on share (p-value)	0.281	**0.081**	0.314	0.227	**0.090**	0.929
Not a daily smoker at 40 y.o.						
Share	0.048	0.279	-0.029	0.084	0.581	-0.652
Inference on share (p-value)	0.267	**0.061**	0.762	0.211	**0.106**	0.924
Not a heavy smoker at 40 y.o.						
Share	0.039	0.249	-0.024	0.009	-0.068	0.139
Inference on share (p-value)	0.280	**0.051**	0.795	0.228	0.769	0.111
No. of cigarettes at 40 y.o.						
Share	0.122	0.485	-0.083	0.043	0.122	0.023
Inference on share (p-value)	0.237	**0.052**	0.858	0.223	0.276	0.345
	Females					
Physical activity at 40 y.o.						
Share	0.235	-0.438	0.006	-0.123	0.037	0.037
Inference on share (p-value)	**0.096**	0.862	0.431	0.705	0.350	0.198
Not a frequent drinker at 27 y.o.						
Share	-0.184	0.166	0.325	-1.591	0.157	0.134
Inference on share (p-value)	0.843	0.260	0.197	0.835	0.293	0.306
Alcohol consumption at 27 y.o.						
Share	-0.195	0.423	0.314	-4.748	0.587	0.497
Inference on share (p-value)	0.846	0.235	0.173	0.942	0.144	**0.098**

Notes: This table presents conditional decomposition results for the two multiple static mediation analyses of the statistically significant outcomes for the Perry Preschool intervention. p-values ≤ 0.10 are printed in boldface. Columns (1)-(3) report the results for the static mediation analysis, which only includes the early childhood mediators: the Stanford-Binet IQ score for ages 7-9, the externalizing behaviour factor for ages 7-9, the academic motivation factor for ages 7-9 (as in Heckman *et al.*, 2013). Columns (4)-(6) report the results for the static mediation analysis, which only includes the adult mediators: being a high school graduate by age 19, monthly income at age 27 adjusted to 2006 prices, and number of months unemployed in the last 2 years at age 27 (as in Heckman *et al.*, 2010). The conditioning variables used are father at home at entry and mother employed at entry. For each outcome we present two lines of results. The first line presents the decomposition share, i.e. $\alpha_j^C E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 13) in columns (1)-(3), where col. (1) refers to cognition, col. (2) to externalizing behaviour, and col. (3) to academic motivation; and $\alpha_j^A E\left(I_{i,j}^A(1) - I_{i,j}^A(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 14) in columns (4)-(6), where col. (4) refers to education, col. (5) to unemployment, and col. (6) to income. The second line presents the one-sided p-value that tests if the share is statistically significantly different from zero, computed using the bootstrap method (1,000 replications).

Table A12: Carolina Abecedarian Intervention - Specification Tests

Outcome	**Test**						
	Dynamic Mediation			*Static Mediation*			
				Early Inputs		*Late Inputs*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Males						
Diastolic blood pressure at mid 30s							
Test Statistic	1.11	3.26	0.56	8.92	6.28	9.04	0.80
p-value	(0.417)	(0.110)	(0.596)	**(0.003)**	**(0.007)**	**(0.005)**	(0.550)
Systolic blood pressure at mid 30s							
Test Statistic	1.02	8.81	4.75	5.82	6.32	16.80	5.31
p-value	(0.447)	(0.016)	(0.058)	**(0.012)**	**(0.007)**	**(0.000)**	**(0.012)**
Hypertension I at mid 30s							
Test Statistic	2.05	3.21	0.23	1.26	3.19	4.99	2.57
p-value	(0.209)	(0.113)	(0.800)	(0.335)	(0.057)	**(0.029)**	(0.097)
Hypertension II at mid 30s							
Test Statistic	5.01	1.36	2.67	2.84	1.30	1.66	0.82
p-value	(0.045)	(0.326)	(0.148)	(0.087)	(0.330)	(0.235)	(0.541)
Health care coverage at 30 y.o.							
Test Statistic	0.01	1.34	1.80	1.07	2.16	3.72	3.17
p-value	(0.998)	(0.282)	(0.165)	(0.379)	(0.098)	**(0.022)**	**(0.028)**
Health coverage provided by the employer or bought at 30 y.o.							
Test Statistic	1.23	0.17	1.07	2.80	1.49	1.86	0.51
p-value	(0.322)	(0.847)	(0.397)	(0.057)	(0.230)	(0.157)	(0.727)
	Females						
Physical activity at 21 y.o.							
Test Statistic	n/a	n/a	n/a	0.08	1.10	n/a	n/a
p-value	(-)	(-)	(-)	(0.972)	(0.374)	(-)	(-)
# Fruit servings at 21 y.o.							
Test Statistic	n/a	n/a	n/a	0.50	0.77	n/a	n/a
p-value	(-)	(-)	(-)	(0.685)	(0.551)	(-)	(-)
Age of onset of drinking < 17							
Test Statistic	n/a	n/a	n/a	1.81	0.25	n/a	n/a
p-value	(-)	(-)	(-)	(0.167)	(0.905)	(-)	(-)

Notes: This table presents Wald test statistics with p-values in parentheses for a number of specification tests of whether the regression coefficients in the outcome equations are the same for the treatment and the control group. p-values ≤ 0.10 are printed in boldface. Notice that the tests cannot be performed for the dynamic mediation analysis and for the static mediation analysis with late inputs for the female sample, since the outcomes are measured at age 21 and the adult mediators at age 30. Specifically, we present tests for the following hypotheses: $H_0 : \boldsymbol{\alpha}_{C,1} = \boldsymbol{\alpha}_{C,0}$ in equation 9 (col. (1)), $H_0 : \boldsymbol{\alpha}_1^A = \boldsymbol{\alpha}_0^C$ in equation 9 (col. (2)), $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$ in equation 9 (col. (3)), $H_0 : \boldsymbol{\alpha}_{C,1} = \boldsymbol{\alpha}_{C,0}$ in equation 13 (col.(4)), $H_0 : \boldsymbol{\beta}_{C,1} = \boldsymbol{\beta}_{C,0}$ in equation 13 (col. (5)), $H_0 : \boldsymbol{\alpha}_1^A = \boldsymbol{\alpha}_0^C$ in equation 14 (col. (6)), $H_0 : \boldsymbol{\beta}_{A,1} = \boldsymbol{\beta}_{A,0}$ in equation 14 (col. (7)).

Table A13: Carolina Abecedarian Intervention - Dynamic Mediation Results

Outcome	**Mediator**														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Males														
Diastolic blood pressure at mid 30s															
Share	-0.263	0.217	0.292	-0.068	-0.022	0.004	-0.018	0.001	-0.001	-0.002	-0.001	0.002	-0.006	-0.011	0.004
Inference on share (p-value)	0.904	**0.092**	**0.095**	0.820	0.562	0.536	0.765	0.452	0.622	0.689	0.501	0.410	0.781	0.482	0.501
Systolic blood pressure at mid 30s															
Share	-0.224	0.276	0.296	-0.130	-0.087	0.026	-0.035	0.004	-0.008	-0.004	-0.002	0.009	-0.012	-0.044	0.022
Inference on share (p-value)	0.901	**0.081**	**0.049**	0.811	0.706	0.531	0.724	0.368	0.684	0.595	0.528	0.336	0.830	0.623	0.412
Hypertension I at mid 30s															
Share	-0.318	0.175	0.306	-0.029	-0.097	0.001	-0.008	0.004	-0.000	-0.001	-0.002	0.000	-0.003	-0.050	0.001
Inference on share (p-value)	0.908	**0.076**	**0.078**	0.863	0.818	0.738	0.824	0.324	0.384	0.684	0.577	0.648	0.690	0.738	0.747
Hypertension II at mid 30s															
Share	-0.147	0.179	0.200	-0.097	0.097	0.023	-0.026	-0.004	-0.007	-0.003	0.002	0.008	-0.009	0.050	0.019
Inference on share (p-value)	0.912	**0.083**	0.147	0.817	0.211	0.506	0.732	0.665	0.625	0.640	0.452	0.339	0.769	0.265	0.327
Health care coverage at 30 y.o.															
Share	-0.365	0.046	0.019	0.077	0.394	0.047	0.020	-0.017	-0.015	0.003	0.009	0.017	0.007	0.200	0.039
Inference on share (p-value)	0.729	0.392	0.321	0.263	0.163	0.289	0.288	0.603	0.816	0.417	0.546	0.197	0.162	0.203	0.254
Health coverage provided by the employer or bought at 30 y.o.															
Share	-0.068	-0.074	0.038	0.040	0.260	0.066	0.011	-0.011	-0.021	0.001	0.006	0.024	0.004	0.132	0.005
Inference on share (p-value)	0.550	0.769	0.233	0.322	**0.100**	0.236	0.299	0.641	0.798	0.480	0.532	0.199	0.168	0.179	0.212

Notes: This table presents conditional decomposition results for the dynamic mediation analysis of the statistically significant outcomes for the Abecedarian intervention, male sample. p-values ≤ 0.10 are printed in boldface. This analysis cannot be performed on the female sample since the statistically significant outcomes are surveyed at age 21, while the adult mediators are recorded at age 30. We use six mediators, three early childhood mediators: the average of the standardised Bayley Mental Development Index and of the Stanford-Binet IQ score for ages 1-2, the average of the standardised Infant Behavior Record (IBR) Task Orientation scales for ages 1-2 (Burchinal *et al.*, 1997), and the average of the Body Mass Index for-age at ages 1-2 (Campbell *et al.*, 2014); and three adult mediators: college education, employment, and earnings (as in García *et al.*, 2014). The conditioning variables used are: mother's WAIS IQ score, High Risk Index (HRI), presence of the father at home, and mother employed at entry. For each outcome we present two lines of results. The first line presents the decomposition share. Columns (1)-(3) display the shares which can be attributed to the direct health effects of experimentally induced changes in childhood factors: $\alpha_j^C E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 9), where col. (1) refers to cognition, col. (2) to task orientation, and col. (3) to BMI. Columns (4)-(6) display the shares which can be attributed to the direct health effect of experimentally induced changes in adulthood factors: $\dfrac{\alpha_j^A E\left(I_{i,j}^A(1) - I_{i,j}^A(0)\right)}{E\left(Y_i(1) - Y_i(0)\right)}$ (see equation 9), where col. (4) refers to education, col. (5) refers to employment, and col. (6) to earnings. Columns (7)-(15) display the shares which can be attributed to the indirect effect of experimentally-induced changes in childhood factors affecting health through the adulthood factors: $\alpha_j^A \gamma^j E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equations 9 and 10). In particular, columns (7)-(9) refer to the effect of experimentally enhanced cognition through education (col. (7)), employment, (col. (8)) and earnings (col. (9)); columns (10)-(12) refer to the effect of experimentally enhanced task orientation through education (col. (10)), employment (col. (11)), and earnings (col. (12)); and columns (13)-(15) refer to the effect of experimentally reduced BMI through education (col. (13)), employment (col. (14)), and earnings (col. (15)). The second line presents the one-sided p-value that tests if the share is statistically significantly different from zero, computed using the bootstrap method (1,000 replications). Hypertension I is defined as systolic blood pressure > 140 and diastolic blood pressure > 90. Hypertension II is defined as systolic blood pressure > 140 or diastolic blood pressure > 90.

Table A14: Carolina Abecedarian Intervention - Static Mediation Results

Outcome	**Mediator**					
	(1)	(2)	(3)	(4)	(5)	(6)
	Males					
Diastolic blood pressure at mid 30s						
Share	-0.264	0.172	0.262	-0.010	-0.039	0.069
Inference on share (p-value)	0.917	**0.038**	**0.080**	0.589	0.596	0.322
Systolic blood pressure at mid 30s						
Share	-0.190	0.207	0.268	-0.086	-0.256	0.123
Inference on share (p-value)	0.874	**0.061**	**0.066**	0.627	0.788	0.376
Hypertension I at mid 30s						
Share	-0.258	0.130	0.267	0.042	-0.160	0.070
Inference on share (p-value)	0.928	**0.041**	**0.053**	0.328	0.753	0.327
Hypertension II at mid 30s						
Share	-0.143	0.038	0.252	-0.050	0.134	0.096
Inference on share (p-value)	0.932	**0.079**	**0.107**	0.686	0.423	0.275
Health care coverage at 30 y.o.						
Share	-0.462	0.198	0.265	0.076	0.705	0.113
Inference on share (p-value)	0.739	0.292	0.210	0.213	**0.076**	0.338
Health coverage provided by the employer or bought at 30 y.o.						
Share	-0.156	0.098	0.265	0.107	0.467	0.126
Inference on share (p-value)	0.684	0.341	0.152	0.170	**0.051**	0.252
	Females					
Physical activity at 21 y.o.						
Share	0.424	0.070	0.044	n/a	n/a	n/a
Inference on share (p-value)	0.131	0.205	0.569	-	-	-
# Fruit servings at 21 y.o.						
Share	-0.275	0.123	-0.012	n/a	n/a	n/a
Inference on share (p-value)	0.696	0.320	0.420	-	-	-
Age of onset of drinking < 17						
Share	-0.015	-0.056	0.038	n/a	n/a	n/a
Inference on share (p-value)	0.409	0.687	0.487	-	-	-

Notes: This table presents conditional decomposition results for the two multiple static mediation analyses of the statistically significant outcomes for the Abecedarian intervention. p-values ≤ 0.10 are printed in boldface. Columns (1)-(3) report the results for the static mediation analysis, which only includes the early childhood mediators: the IQ score for ages 1-2, the IBR task orientation scale for ages 1-2, the BMI for ages 1-2 (as in Burchinal *et al.*, 1997; Campbell *et al.*, 2014). Columns (4)-(6) report the results for the static mediation analysis, which only includes the adult mediators: being a college graduate, employment, and earnings at age 30 (as in García *et al.*, 2014). This latter analysis cannot be performed on the female sample since the statistically significant outcomes are surveyed at age 21, while the adult mediators are recorded at age 30. The conditioning variables used are mother's WAIS IQ score, High Risk Index (HRI), presence of the father at home, and mother employed at entry. For each outcome we present two lines of results. The first line presents the decomposition share, i.e., $\alpha_j^C E\left(I_{i,j}^C(1) - I_{i,j}^C(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 13) in columns (1)-(3), where col. (1) refers to cognition, col. (2) to task orientation, and col. (3) to BMI; and $\alpha_j^A E\left(I_{i,j}^A(1) - I_{i,j}^A(0)\right) / E\left(Y_i(1) - Y_i(0)\right)$ (see equation 14) in columns (4)-(6), where col. (4) refers to education, col. (5) refers to employment and col. (6) to earnings. The second line presents the one-sided p-value that tests if the share is statistically significantly different from zero, computed using the bootstrap method (1,000 replications). Hypertension I is defined as systolic blood pressure > 140 and diastolic blood pressure > 90. Hypertension II is defined as systolic blood pressure > 140 or diastolic blood pressure > 90.

References

Breitmayer, B.J. and Ramey, C.T. (1986). 'Biological nonoptimality and quality of postnatal environment as codeterminants of intellectual development', *Child Development*, vol. 57(5), pp. 1151–1165.

Burchinal, M.R., Campbell, F.A., Bryant, D.M., Wasik, B.H. and Ramey, C.T. (1997). 'Early intervention and mediating processes in cognitive performance of children of low-income African American families', *Child Development*, vol. 68(5), pp. 935–954.

Campbell, F.A., Conti, G., Heckman, J.J., Moon, S.H., Pinto, R., Pungello, E.P. and Pan, Y. (2014). 'Early childhood investments substantially boost adult health', *Science*, vol. 343(6178), pp. 1478–1485.

Campbell, F.A. and Ramey, C.T. (1995). 'Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention', *American Educational Research Journal*, vol. 32(4), pp. 743–772.

García, J., Heckman, J.J. and Pinto, R. (2014). 'Reassessing Carolina Abecedarian Project: Attrition, control contamination, and multiple hypotheses testing', University of Chicago, Department of Economics.

Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A. and Yavitz, A.Q. (2010). 'Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program', *Quantitative Economics*, vol. 1(1), pp. 1–46.

Heckman, J.J. and Pinto, R. (2015). 'A unified approach to examine treatment effects: Causality and identification', Forthcoming, *American Economic Review*.

Heckman, J.J., Pinto, R. and Savelyev, P.A. (2013). 'Understanding the mechanisms through which an influential early childhood program boosted adult outcomes', *American Economic Review*, vol. 103(6), pp. 2052–2086.

Heckman, J.J., Pinto, R., Shaikh, A.M. and Yavitz, A.Q. (2011). 'Inference with imperfect randomization: The case of the Perry Preschool Program', National Bureau of Economic Research.

Horvitz, D.G. and Thompson, D.J. (1952). 'A generalization of sampling without replacement from a finite universe', *Journal of the American Statistical Association*, vol. 47(260), pp. 663–685.

Huber, M. (2012). 'Identifying causal mechanisms in experiments (primarily) based on inverse probability weighting', University of St. Gallen, School of Economics and Political Science, Department of Economics.

Johnston, J. and DiNardo, J.E. (1997). *Econometric Methods*, New York, NY: McGraw-Hill, fourth edn.

Lee, W.S. (2013). 'Propensity score matching and variations on the balancing test', *Empirical Economics*, vol. 44(1), pp. 47–80.

Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, New York: Springer-Verlag, 3 edn., doi:10.1007/0-387-27605-X.

Ramey, C.T. and Campbell, F.A. (1979). 'Compensatory education for disadvantaged children', *The School Review*, vol. 87(2), pp. 171–189.

Ramey, C.T. and Campbell, F.A. (1984). 'Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project', *American Journal of Mental Deficiency*, vol. 88(5), pp. 515–523.

Ramey, C.T. and Campbell, F.A. (1991). 'Poverty, early childhood education and academic competence: The Abecedarian experiment', in (A. C. Houston, ed.), *Children in Poverty: Child Development and Public Policy*, pp. 190–221, chap. 8, New York: Cambridge University Press.

Ramey, C.T., Collier, A.M., Sparling, J.J., Loda, F.A., Campbell, F.A., Ingram, D.A. and Finkelstein, N.W. (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., Dorval, B. and Baker-Ward, L. (1983). 'Group day care and socially disadvantaged families: Effects on the child and the family', in (S. Kilmer, ed.), *Advances in Early Education and Day Care*, pp. 69–106, vol. 3, JAI Press.

Ramey, C.T., Holmberg, M.C., Sparling, J.H. and Collier, A.M. (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*, pp. 101–121, chap. 7, New York: Walker and Company.

Ramey, C.T., Mills, P., Campbell, F.A. and O'Brien, C. (1975). 'Infants' home environments: A comparison of high-risk families and families from the general population', *American Journal of Mental Deficiency*, vol. 80(1), pp. 40–42.

Ramey, C.T. and Smith, B.J. (1977). 'Assessing the intellectual consequences of early intervention with high-risk infants', *American Journal of Mental Deficiency*, vol. 81(4), pp. 318–324.

Romano, J.P. and Wolf, M. (2005). 'Exact and approximate stepdown methods for multiple hypothesis testing', *Journal of the American Statistical Association*, vol. 100(469), pp. 94–108, doi:10.1198/016214504000000539.

Weikart, D.P., Bond, J.T. and McNeil, J.T. (1978). *The Ypsilanti Perry Preschool Project: Preschool Years and Longitudinal Results Through Fourth Grade*, Ypsilanti, MI: HighScope Press.