

Supplementary Materials

Table of Contents

A	The Carolina Abecedarian Project (ABC): Overview of Program Design and Main Features	32
A.1	Recruitment and Randomization	34
A.2	Preschool Treatment	35
B	Definition of Variables	38
C	Methodology	47
D	Attrition	56
E	Estimation results	66
F	Inference about the Treatment Effect for the School-Age Component	73
G	A Simple Mediation Analysis	81
H	Methodology for Permutation-based Inference	86
I	Bootstrap Analysis of Test Statistics and an Analysis of Confidence Intervals	93

Tables and Figures

Table 1. Abecedarian Intervention, Males: Main Health Results, Biomedical Sweep	20
Table 2. Abecedarian Intervention, Females: Main Health Results, Biomedical Sweep	21
Table 3. Abecedarian Intervention, Males: Health Care at age 30; Physical Development in Childhood	22
Table 4. Abecedarian Intervention, Females: Health Care at age 30; Physical Development in Childhood	23
Fig. 1: Body Mass Index (BMI) ages 0–5 by Treatment Status, Males	24
Fig. 2: Body Mass Index ages 0–4 (figures a, c, e) and 2–8 (figures b, d, f), by Treatment and Obesity Status at Mid-30s, Males	25
Fig. 3: Body Mass Index (BMI) ages 0–5 by Treatment Status, Females	26
Fig. 4: Body Mass Index ages 0–4 (figures a, c, e) and 2–8 (figures b, d, f), by Treatment and Obesity Status at Mid-30s, Females	27
Table S1: Selection of IPW Covariates, Males	52
Table S2: Selection of IPW Covariates, Females	54
Table S3: Attrition in the Abecedarian Intervention	59
Table S4: Differences between Attrited and Non-Attrited: (Treated, Females)	60
Table S5: Differences between Attrited and Non-Attrited: (Control, Females)	61
Table S6: Differences between Attrited and Non-Attrited: (Treated, PE, Males)	62
Table S7: Differences between Attrited and Non-Attrited: (Control, PE, Males)	63
Table S8: Differences between Attrited and Non-Attrited: (Treated, Lab, Males)	64
Table S9: Differences between Attrited and Non-Attrited: (Control, Lab, Males)	65
Table S10: Inference for the Abecedarian Intervention	68
Table S11: Inference for the Abecedarian Intervention: (Males)	69

Table S12: Inference for the Abecedarian Intervention: (Females)	70
Table S13: Inference for the Abecedarian Intervention: (Males)	71
Table S14: Inference for the Abecedarian Intervention: (Females)	72
Table S15(a): Second-Stage (School-Age) Component of the Abecedarian Intervention: (Males)	75
Table S15(b): Second-Stage (School-Age) Component of the Abecedarian Intervention: (Females)	76
Table S16(a): Second-Stage (School-Age) Component of the Abecedarian Intervention: (Males)	77
Table S16(b): Second-Stage (School-Age) Component of the Abecedarian Intervention: (Females)	78
Table S17: Tests of H1, H2 and H3: (Males)	79
Table S18: Tests of H1, H2 and H3: (Females)	80
Table S19: Decompositions of Treatment Effects on Health Outcomes (Biomedical Sweep), Males	84
Table S20: Unrestricted Decompositions of Treatment Effects on Health Outcomes (Biomedical Sweep), Males	85
Table S21(a): Unconditional Bootstrap Inference for the Abecedarian Intervention (Males)	98
Table S21(b): Unconditional Bootstrap Inference for the Abecedarian Intervention (Males)	99
Table S21(c): Unconditional Bootstrap Inference for the Abecedarian Intervention (Males)	100
Table S21(d): Unconditional Bootstrap Inference for the Abecedarian Intervention (Males)	101
Table S22(a): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)	102
Table S22(b): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)	103
Table S22(c): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)	104
Table S22(d): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)	105
Table S23(a): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)	106
Table S23(b): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)	107
Table S23(c): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)	108

Table S23(d): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)	109
Table S24(a): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Females)	110
Table S24(b): Conditional IPW Bootstrap Inference for the Abecedarian Daycare Intervention (Females)	111
Table S24(c): Conditional IPW Bootstrap Inference for the Abecedarian Daycare Intervention (Females)	112
Table S24(d): Conditional IPW Bootstrap Inference for the Abecedarian Daycare Intervention (Females)	113
Table S25(a): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Males)	114
Table S25(b): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Males, cont'd)	115
Table S26(a): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Females)	116
Table S26(b): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Females, cont'd)	117

A The Carolina Abecedarian Project (ABC): Overview of Program Design and Main Features

The Carolina Abecedarian Project (ABC) is a prospective longitudinal randomized controlled trial of an early educational intervention implemented to prevent the development of mild mental retardation for children at high risk for poor cognitive and academic outcomes. The ABC project consists of four consecutive cohorts of children from low-income families (96% African American, 4% White) from the same community in a semi-rural county in North Carolina who were enrolled in a randomized trial of early childhood education in the 1970s. The intervention included both a preschool and a school-age component. There are two stages of randomization. Some children in the first-stage control group were randomized into the second stage treatment group and some children in the first-stage treatment group are randomized out of treatment in the second stage. The main paper only analyzes the outcomes for those who participated in the preschool (0–5) phase irrespective of whether they participated in the school-age component. We test and do not reject the null hypothesis that those who received the second-stage treatment have the same outcomes as those who were randomized out of the treatment in the second stage. (See the Supplementary Material, Section F.)

Eligibility was based on a High-Risk Index that included maternal and paternal educational levels, family income, father's presence, and 9 other indicators of family status and functioning: "Absence of maternal relatives in the area"; "Siblings of school age one or more grades behind age-appropriate level or with equivalently low scores on school-administered achievement tests"; "Payments received from welfare agencies within past 3 years"; "Record of father's work indicates unstable or unskilled and semiskilled labor"; "Record of mother's or father's IQ indicate scores of 90 or below"; "Record of sibling's IQ indicates scores of 90 or below"; "Relevant social

agencies in the community indicate the family is in need of assistance”; “One or more members of the family has sought counseling or professional help in the past 3 years”; and “Special circumstances not included in any of the above are likely contributors to cultural or social disadvantage” (51).

A.1 Recruitment and Randomization

Eligibility Children were recruited in four separate waves by eliciting referrals from local community organizations such as prenatal clinics, hospitals, and social services (52). Families whose children appeared healthy and free from biological conditions that could be associated with mental, sensory, or motor disabilities were contacted (53). Once a potential candidate was identified, a High-Risk Index (HRI) was computed on the basis of 13 socio-economic factors as described in (54) to determine preliminary eligibility: a family was included if the index was greater than 11. Final eligibility was then determined after an interview with the mothers of the potential participants. On this occasion, demographic background data were collected, and maternal intelligence was assessed using the Wechsler Adult Intelligence Scale (WAIS). The selection process took place either before or shortly after the birth of the subject child. A total of 122 eligible families (combining all waves) were invited to enroll in the program. However, one family declined to participate and one mother miscarried her baby, so that the final randomized set consists of 120 families associated with 122 children (18, 55).

Randomization The randomization was performed immediately after each cohort was formed (56). For each cohort, the initial 120 families were matched in pairs on the basis of the sex of the child, the IQ of the mother, the number of siblings, and the High-Risk Index score (18, 52, 53). Unfortunately, the precise matched pairs are unknown. The matched children were randomly assigned either to treatment or to control status within each pair. Out of 122 children, 65 were assigned to the treatment, and 57 to the control group.

Compliance and non-random re-assignment Among the 120 families (or 122 children) who were randomized, only 112 families accepted their actual assignments: 8 families (7 in the treatment and 1 in the control group) refused to participate after learning it (57). Among the 112 families (or 114 children) who complied, one infant assigned to the treatment group proved to have

biological retardation, so he was excluded from the study (although he attended the day care center as a regular member and underwent the regular series of assessments). In addition, two infants in the control group were non-randomly re-assigned to the treatment group at the request of protective service officials. The final base sample included 109 families (111 children). Of these 111 children, 57 were assigned to the treatment status, and 54 to the control. Among them there were one pair of twins and one pair of siblings. Both were assigned to the treatment group (18). Average age at entry for the treated was 8.7 weeks, and it ranged between 3 and 21 weeks.

A.2 Preschool Treatment

The preschool treatment consisted of three components: an educational component, a health care component, and a nutritional component.

Educational Component The educational component consisted of full-day child care for five days per week, 50 weeks per year, for the first 5 years of life, starting at two months. Child care was characterized by a low teacher-child ratio (1:3 for infants). It used a curriculum developed by Sparling and Lewis, consisting of a series of “educational games,” which emphasized language, emotional development, and cognitive skills (15, 16).

The ABC program shares many important features with the HighScope Perry Preschool program in the curriculum for 3- to 4-year-olds. At that age, both programs have intensive teacher-child interactions with children in small group settings that engage children in a similar set of activities. ABC starts earlier (first two months of life) and continues later (to age eight). Unlike Perry, the ABC intervention provided free primary pediatric care and nutrition to the treated children. This included routine screenings (yearly for vision, monthly for hearing), immunizations, visits by the member of the pediatric care staff, and laboratory tests and appropriate cultures. Medicines were prescribed when the children were sick, but they had to be paid for by the families (although the pediatric staff made sure children took them while at the center). In case specialist

visits were needed, the children were referred to UNC doctors, and the parents were responsible for paying the bills. The control group had access to community clinics (crowded and with rotating doctors/nurses), the local office of the health department (for well-baby checkups and immunizations), and the hospital ER. Conditional on satisfying the eligibility criteria, Medicaid was also available in North Carolina in those years. Additionally, in the ABC preschool intervention, there was no form of parental tutoring in the form of home visits, apart from the health counseling parents received during the well-child care assessments (58).

Families in both treatment and control groups received social service support as needed (52, 59). Families in the control group also received diapers until the child was toilet-trained as an incentive for their participation (60). The average annual cost of the intervention was about \$13,900 per child in the year 2002 dollars (61).

Health Care Component The Abecedarian intervention also included a health care component, which provided complete medical care for the children who attended the FPG center (62)(63). The medical staff consisted of three pediatricians, a family nurse practitioner, and a licensed practical nurse. Active research on respiratory tract infections in children was also ongoing (64, 65). The *well-child care* component included assessments at ages 2, 4, 6, 9, 12, 18, and 24 months, and yearly thereafter, in which a complete physical exam was performed and parents were counseled about child health care, nutrition, growth, and development. The *ill-child care* component included daily surveillance of all children in the treatment group for illness (66). When ill, the children were examined by a member of the health care staff, laboratory tests were performed, the appropriate treatment was given, and the child was followed until recovery (63). Only the treated children had daily surveillance while at the child care center. The licensed practical nurse visited the classroom daily to review the health status of the children and receive reports from the parents.

Nutritional Component In addition to receiving their primary pediatric care, children in the treatment group also received two meals and a snack at childcare the center. The food was catered by kitchens approved by the local health department. A nutritionist who planned the local public school menus consulted with the kitchen service to plan menus for breakfast, lunch, and daily snacks. Infants were given iron-fortified formula until the doctors advised adding solids, which was commercial baby food chosen by the staff. These nutritional supplements were not offered to the control group. However, control families were offered free iron-fortified formula for the first 15 months of life in order to ensure the quality of nutrition during a period of rapid brain growth.

B Definition of Variables

This section provides details on the main outcome variables.

Health The main health outcomes analyzed come from the information collected in the biomedical sweep, both from the physical exam in the MD office and from the lab tests. Those residing out of the state were brought in for the visit. Three measurements for blood pressure were taken at 60-second intervals, after the subject was seated for 5 minutes, and were averaged.

- The blood pressure categories as defined by (27) are as follows: normal is < 120 for systolic and < 80 for diastolic; pre-hypertension is 120–139 for systolic or 80–89 for diastolic; high blood pressure (hypertension) stage 1 is 140–159 for systolic or 90–99 for diastolic; high blood pressure (hypertension) stage 2 is 160 or higher for systolic or 100 or higher for diastolic. We use two alternative definitions for **pre-hypertension**: systolic bp ≥ 120 & diastolic bp ≥ 80 and systolic bp ≥ 120 or diastolic bp ≥ 80 ; and two alternative definitions for **hypertension**: systolic bp ≥ 140 & diastolic bp ≥ 90 and systolic bp ≥ 140 or diastolic bp ≥ 90 . While the second definition is the one adopted in (27), the first one is also widely used; see, e.g., (28).
- **Dyslipidemia** is defined as HDL-C < 40 mg/dL for males, and as HDL-C < 50 mg/dL for females (30), (67).
- **Pre-Diabetes** is defined as Glycosylated Hemoglobin $\geq 5.7\%$ (31).
- **Vitamin D Deficiency** is defined as Total Vitamin D < 20 ng/mL (32).
- **Overweight, Obese, and Severely Obese** are defined as having a Body Mass Index (BMI) ≥ 25 , ≥ 30 , and ≥ 35 , respectively.
- **Abdominal Obesity** is defined as WHR > 0.9 for males, WHR > 0.85 for females (33), (68).
- **Multiple Risk Factors**: here the outcome “Hypertension” used refers to the second definition (systolic bp ≥ 140 mmHg or diastolic bp ≥ 90 mmHg).

- **The Metabolic Syndrome** (30) is defined as the presence of the following three conditions: (1) Central obesity, defined as waist circumference > 102 cm or 40 inches (male), > 88 cm or 35 inches (female) (34); (2) Dyslipidemia, defined as HDL-C < 40 mg/dL (male), HDL-C < 50 mg/dL (female); (3) High blood pressure $\geq 130/85$ mmHg. We do not have the information required to construct the WHO index (33).
- **The Framingham Risk Score** is calculated according to the equation derived in (35). The equation is derived from a Cox proportional hazard model estimated on the Framingham Study population. The equation uses age, sex, total cholesterol, HDL cholesterol, systolic blood pressure, diastolic blood pressure, diabetes status, and smoking status at the time of the biomedical sweep (69) to estimate the subject's risk of experiencing "total CHD," defined as stable angina, unstable angina, myocardial infarction, or CHD death, within 10 years.

Health Care

- The variable "**Covered by health insurance at age 30**" is constructed on the basis of information collected at the age 30 interview. It is based on the subject's response to the question: "Which of the following best describes your current health insurance situation?" The responses A="No health insurance" and I="You don't know what your health insurance situation is" were coded as 0 and the responses B="Covered by husband 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", and J="Other" were coded as 1.
- The variable "**Buys insurance through work or purchases it at age 30**" is constructed on the basis of information collected at the age 30 interview. It is based on the subject's response to the question: "Which of the following best describes your current health insurance situation?" The responses C="Get own insurance through work" and G="Buy private insurance on your own"

were coded as 1 and responses A="No health insurance", B="Covered by husband's or wife's insurance policy", D="Get insurance through a union", E="Get insurance because of attending school", F="Covered because of being on active duty in the military", H="Covered by Medicaid", I="You don't know what your health insurance situation is" and J="Other" were coded as 0.

- The variable "**Has hospital or doctor for care when sick at age 30**" is constructed on the basis of information collected at the age 30 interview. It is based on the subject's response to the question: "Where do you usually go when you are sick and need health care?" Responses of C="Hospital-based clinic", D="Emergency room", and F="Private doctor's office" were coded as 1 and responses of A="Never get sick or need care", B="Get sick but don't go anywhere for health care", E="Community health department or clinic", G="Infirmary at school", H="Military hospital or clinic", I="Employee health clinic at my work", and J="Other" were coded as 0.

Behavioral Risk Factors: Smoking

- The variable "**Never a Regular Smoker**" takes the value 1 if the subject has never reported being a regular smoker by age 30. It is constructed by combining information from the age 21 and 30 sweeps as follows. The variable is coded as 1 if the subject reported having never smoked regularly both at the age 21 and at the age 30 sweeps, in case both have nonmissing observations, or at either of those, in case the other is missing. Specifically, the same question was asked at the age 21 sweep in the *Risk Taking Survey*, and at the age 30 sweep in the *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)?" Response A="I have never smoked cigarettes regularly" was coded as 1, all the others as 0.

- The variable "**Age of Onset of Regular Smoking**" takes a value corresponding to the age at which the subject started smoking regularly. It has been constructed on the basis of the following question asked in the *Risk Taking Survey* administered to the subject at the age 21 and in the

Tobacco, Alcohol and Drugs (TAD) survey administered to the subject at the age 30 data collection sweeps, respectively: “How old were you when you first started smoking cigarettes regularly (at least one cigarette every day for 30 days)?”. The responses given at age 21 were used, unless the subject did not respond at age 21 or had not started smoking regularly yet, in which cases the responses given at age 30 were used. The responses were given in age ranges and were converted into ages by taking the midpoints of the intervals as follows. At age 21, responses of D=“11 or 12 years old” were coded as 11.5, responses of E=“13 or 14 years old” as 13.5, responses of F=“15 or 16 years old” as 15.5, and responses of G=“17 or more years old” were coded as the average of age 17 and the age of the subject at the administration of the age 21 interview. At age 30, responses of C=“9–12 years old” were coded as 10.5, responses of D=“13–16 years old” were coded as 14.5, responses of E=“17–20 years old” were coded as 18.5, responses of F=“21–24 years old” were coded as 22.5, and responses of G=“25 years or older” were coded as the average of 25 and the age of the subject at the administration of the age 30 interview (70).

- The variable **“Cigarettes Smoked per Day Past 30 Days at ages 21&30 (Smokers Only)”** is constructed by summing the responses at ages 21 and 30. The number of cigarettes the subject smoked per day at age 21(30) was measured by his/her response to the question “During the past 30 days, on the days you smoked, how many cigarettes did you smoke per day?” Responses of A=“I did not smoke during the past 30 days” were coded as missing, responses of B=“Fewer than one per day” were coded as 0.5, responses of C=“1 cigarette per day” were coded as 1, responses of D=“2–5 cigarettes per day” were coded as 3.5, responses of E=“6–10 cigarettes per day” were coded as 8, responses of F=“11–20 cigarettes per day” were coded as 15.5, and responses of G=“More than 20 cigarettes per day” were coded as 20.

Behavioral Risk Factors: Alcohol Use

- The variable **“Early Onset Drinker”** takes the value 1 if the subject reported at the age 21

interview having started drinking before age 17, and the value 0 if the subject either did not start, or if he/she started after 17. It is constructed on the basis of the following question asked in the *Risk Taking Survey* administered to the subject at the age 21 data collection sweep: “How old were you when you had your first drink of alcohol other than a few sips?” (71) Responses of A=“I have never had a drink of alcohol other than a few sips” and G=“17 or more years old” were coded as 0, and all the other responses (C=“9 or 10 years old”, D=“11 or 12 years old”, E=“13 or 14 years old”, and F=“15 or 16 years old”) (72) were coded as 1.

- The variable “**Drunk Driving at age 21**” takes the value 1 if the subject reported at the age 21 that he/she has drunk and driven at least once in the past 30 days, according to the answer to the following question: “During the past 30 days, how many times did you drive a car or other vehicle when you had been drinking alcohol?” Responses of A=“0 times” were coded as 0 and responses of B=“1 time”, C=“2 or 3 times”, D=“4 or 5 times”, and E=“6 or more times” were coded as 1.

- The variable “**Age of Onset of Alcohol Use**” takes a value corresponding to the age at which the subject started drinking alcohol. It has been constructed on the basis of the following question asked in the *Risk Taking Survey* administered to the subject at the age 21 (73) and in the *Tobacco, Alcohol and Drugs (TAD)* survey administered to the subject at the age 30 (74) data collection sweeps, respectively: “How old were you when you had your first drink of alcohol other than a few sips?” (75). The responses given at age 21 were used, unless the subject did not respond at age 21 or had not had a drink yet, in which cases the responses given at age 30 were used. The responses were given in age ranges and were converted into ages by taking the midpoints of the intervals as follows. At age 21, responses of C=“9 or 10 years old” were coded as 9.5, responses of D=“11 or 12 years old” as 11.5, responses of E=“13 or 14 years old” as 13.5, responses of F=“15 or 16 years old” as 15.5, and responses of G=“17 or more years old” were coded as the average of age 17 and the age of the subject at the administration of the age 21 interview. At age 30, responses of

C="9–12 years old" were coded as 10.5, responses of D="13–16 years old" were coded as 14.5, responses of E="17–20 years old" were coded as 18.5, responses of F="21–24 years old" were coded as 22.5, and responses of G="25 years or older" were coded as the average of 25 and the age of the subject at the administration of the age 30 interview (76).

- The variable **"Days At Least 1 Drink Past 30 Days at ages 21&30"** is constructed by summing the responses at ages 21 and 30. Frequency of alcohol use in the past 30 days at age 21(30) was measured by the subject's response to the following question: "During the past 30 days, on how many days did you have at least one drink of alcohol"? Responses of A="0 days" were coded as 0, responses of B="1–2 days" were coded as 1.5, responses of C="3–5 days" were coded as 4, responses of D="6–9 days" were coded as 7.5, responses of E="10–19 days" were coded as 14.5, responses of F="20–29 days" were coded as 24.5, and responses of G="All 30 days" were coded as 30.

- The variable **"Days 5+ Drinks Past 30 Days at ages 21&30"** is constructed by summing the responses at ages 21 and 30. Frequency of binge drinking in the past 30 days at age 21(30) was measured by the subject's response to the following question: "During the past 30 days, on how many days did you have 5 or more drinks of alcohol in a row, that is, within a couple of hours?" Responses of A="0 days" were coded as 0, responses of B="1 day" were coded as 1, responses of C="2 days" were coded as 2, responses of D="3–5 days" were coded as 4, responses of E="6–9 days" were coded as 7.5, responses of F="10–19 days" were coded as 14.5, and responses of G="20 or more days" were coded as 20.

Behavioral Risk Factors: Drug Use

- The variable **"Never Used Marijuana"** takes the value 1 if the subject has never reported using marijuana by age 30. It is constructed by combining information from the age 21 and 30 sweeps as follows. First, the variable is coded as 1 if the subject reported having never tried marijuana both at

the age 21 and at the age 30 sweeps, in case both have nonmissing observations, or at either of those, in case the other is missing. Specifically, the same question was asked at the age 21 sweep in the *Risk Taking Survey*, and at the age 30 sweep in the *Tobacco, Alcohol and Drugs (TAD)* survey: “How old were you when you tried marijuana for the first time?”. Responses of A=“I have never tried marijuana” were coded as 1, all the others as 0.

- The variable “**Age of Onset of Marijuana Use**” takes a value corresponding to the age at which the subject started using marijuana. It has been constructed on the basis of the following question asked in the *Risk Taking Survey* administered to the subject at the age 21 and in the *Tobacco, Alcohol and Drugs (TAD)* survey administered to the subject at the age 30 data collection sweeps, respectively: “How old were you when you tried marijuana for the first time?”. The responses given at age 21 were used, unless the subject did not respond at age 21 or had not tried marijuana yet, in which cases the responses given at age 30 were used. The responses were given in age ranges and were converted into ages by taking the midpoints of the intervals as follows. At age 21, responses of C=“9 or 10 years old” were coded as 9.5, responses of D=“11 or 12 years old” as 11.5, responses of E=“13 or 14 years old” as 13.5, responses of F=“15 or 16 years old” as 15.5, and responses of G=“17 or more years old” were coded as the average of age 17 and the age of the subject at the administration of the age 21 interview. At age 30, responses of C=“9–12 years old” were coded as 10.5, responses of D=“13–16 years old” were coded as 14.5, responses of E=“17–20 years old” were coded as 18.5, responses of F=“21–24 years old” were coded as 22.5, and responses of G=“25 years or older” were coded as the average of 25 and the age of the subject at the administration of the age 30 interview (77).

- The variable “**Times Used Marijuana Past 30 Days at ages 21&30**” is constructed by summing the responses at ages 21 and 30. The frequency of marijuana use in the past 30 days was measured by the subject’s response to the following question at age 21(30): “During the past 30

days, how many times did you use marijuana?” Responses of A=“0 times” were coded as 0, responses of B=“1 or 2 times” were coded as 1.5, responses of C=“3 to 9 times” were coded as 6, responses of D=“10 to 19 times” were coded as 14.5, responses of 20 to 39 times were coded as 29.5 times, and responses of 40 or more times were coded as 40.

Behavioral Risk Factors: Diet and Physical Exercise

- The variable “**Physical Activity**” takes the value 1 if the subject reported at the age 21 interview engaging in physical activity 4 or more days per week, and 0 if the subject engaged 3 or fewer days per week. It has been constructed on the basis of the following question asked in the *Risk Taking Survey* administered to the subject at the age 21 data collection sweep: “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?” Responses of A=“0 days”, B=“1 day”, C=“2 days”, and D=“3 days” were coded as 0; responses of E=“4 days”, F=“5 days”, G=“6 days”, and H=“7 days” were coded as 1.

- The variable “**Number of Fruit Servings per Day**” takes a value equal to the number of times the subject reported having eaten fruit the day before the age 21 interview. It has been constructed on the basis of the following question asked in the *Risk Taking Survey* administered to the subject at the age 21 data collection sweep: “Yesterday, did you eat fruit?” The question was preceded by the prompt: “The next seven questions ask about food you ate yesterday. Think about all meals and snacks you ate yesterday from the time you got up until you went to bed. Be sure to include food you ate at home, at school or work, at restaurants, or anywhere else.” Responses of A=“No” were coded as 0, responses of B=“Yes, once only” were coded as 1, and responses of C=“Yes, twice or more” were coded as 2.

Physical Development

- We define “**at-risk overweight**” under 24 months as weight-for-length $\geq 85^{th}$ percentile, and **overweight** for 24 months and older as BMI-for-age $\geq 85^{th}$ percentile (46).
- **Weight-for-Length z-score change** between 0 and 24 months is defined as the change in weight-for-length standardized using the Center for Disease Control (CDC) or the World Health Organization (WHO) growth charts, respectively.

C Methodology

The standard model of program evaluation describes the observed outcome Y_i of participant i by

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0), \quad (1)$$

where D_i denotes the assignment indicator ($D_i = 1$ if treatment occurs, $D_i = 0$ otherwise) and $(Y_i(0), Y_i(1))$ are potential outcomes for agent i when treatment is *fixed* at control and treatment status, respectively. The major benefit of a randomized experiment is that it induces independence between counterfactual outcomes $(Y_i(0), Y_i(1))$ and treatment status D_i conditional on pre-program variables X used in the randomization protocol. Notationally we write $(Y(0), Y(1)) \perp\!\!\!\perp D \mid X$ where $Y(0), Y(1), D$, and X respectively denote the vectors of counterfactual outcomes, treatment status, and pre-program variables used in the randomization protocol. Where “ $A \perp\!\!\!\perp B \mid C$ ” denotes statistical independence, given C . Our goal is to estimate average treatment effects and test the null hypothesis of no treatment.

Our inference procedure accounts for three features of the Abecedarian data: (1) its small sample size; (2) the large number of outcomes that would allow for an arbitrary selection of statistically significant treatment effects; i.e., “cherry picking”; and (3) attrition.

We account for the small sample size by implementing a block permutation test which is based on exchangeability properties of the data arising from the randomization protocol. Blocks of permutations are based on the strata of the variables X used in the randomization protocol of the Abecedarian intervention. We refer to Appendix H for a detailed explanation of the method. The benefits of our inferential method are: (1) it is a fully non-parametric approach; (2) it only relies on the choice of the test statistics; and (3) it is valid for small samples, and allow us to test the equality

of the parameters of interest between treatment and control groups.

We correct for “cherry picking” by implementing a multiple hypothesis testing based on the step-down algorithm described in (24). We refer to Appendix H for a detailed explanation on how to implement the step-down algorithm for block permutation tests in randomized trials.

Supplementary Material, Section D, demonstrates that attrition is a potentially serious problem in this sample. We correct for attrition using statistical models that account for missing data by reweighting observations according to the inverse of their conditional probability of compliance. This is usually termed Inverse Probability Weighting (IPW) and traces back to (26).

The key assumption of IPW methods is that by conditioning on a set of observed covariates we are able to retrieve the full distribution of an outcome of interest. IPW methods rely on a matching assumption that can be stated as $Y \perp\!\!\!\perp A \mid (D, X, Z)$, where Z are pre-program variables (other than the ones used in the randomization protocol; i.e., X) and A is a vector of attrition indicators $A = (A_i : i \in J)$, $J = \{1, \dots, N\}$ which takes the value $A_i = 1$ if participant i has non-missing data on outcome Y , and $A_i = 0$ otherwise.

Under the matching assumption, $E(Y \mid D, X, A) = E(Y \mid D, X)$. Moreover, $D \perp\!\!\!\perp (Y(0), Y(1)) \mid X, Z$ holds due to randomization. Suppressing i subscripts, we can use these two conditional statistical relationships to express the Average Treatment Effect (ATE) as

$$\begin{aligned}
 E(Y(1) - Y(0)) &= \int E(Y \mid D = 1, X = x, Z = z) - E(Y \mid D = 0, X = x, Z = z) dF_{X,Z}(x, z), \\
 &= \int E(Y \mid D = 1, A = 1, X = x, Z = z) - E(Y \mid D = 0, A = 1, X = x, Z = z) dF_{X,Z}(x, z). \tag{8}
 \end{aligned}$$

The standard IPW formula for ATE can be obtained by applying Bayes’ theorem to Equation 8:

$$E \left(\frac{Y_i I[A_i = 1, D_i = 1]}{\Pr(A_i = 1 \mid D_i = 1, X_i, Z_i) \Pr(D_i = 1 \mid X_i, Z_i)} - \frac{Y_i I[A_i = 1, D_i = 0]}{\Pr(A_i = 1 \mid D_i = 0, X_i, Z_i) \Pr(D_i = 0 \mid X_i, Z_i)} \right), \tag{9}$$

where “Pr(·)” means probability. Making the i subscripts explicit, the implementation of expression (9) suggested in (25) is given by:

$$\widehat{ATE} = \sum_{i=1}^N \frac{Y_i \cdot I[D_i = 1] \cdot I[A_i = 1] \cdot \omega_{i,1}}{N_1} - \sum_{i=1}^N \frac{Y_i \cdot I[D_i = 0] \cdot I[A_i = 1] \cdot \omega_{i,0}}{N_0} \quad (10)$$

$$\text{where } \omega_{i,d} = \frac{1}{\hat{p}_{i,d}} / \left(\frac{1}{N_d} \sum_{j=1}^N \frac{I[D_j = d] \cdot I[A_j = 1]}{\hat{p}_{j,d}} \right) \quad d \in \{0,1\}$$

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

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

where N is the total sample size and $\hat{p}_{i,d}$ is an estimate of $p_{i,d}$ described below. Weights $\omega_{i,d}$ are set such that their sum adds to the available sample size of the respective treatment; that is,

$$\sum_{i=1}^N \omega_{i,d} \cdot I[D_i = d] \cdot I[A_i = 1] = N_d.$$

Probabilities $\hat{p}_{i,d}$ are estimated by gender using a logit model. The dependent variable is an age-specific attrition indicator. Regressors are: (1) a constant term, (2) the treatment status, and (3) a set of pre-program variables termed IPW covariates.

As shown in Tables S4–S9 in Supplement D, we find statistically significant differences between attrited and non-attrited (by treatment status and gender) for a range of covariates at baseline and at age 30. All of these variables are potential candidates as covariates for the estimation of the attrition probability. Our small sample size, however, prohibits the use of a large number of conditioning covariates. An excessive number of covariates, indeed, will invalidate the assumptions required to estimate the model by generating an exact forecast of the attrition indicator.

The problem of selection of eligible covariates is an aspect of model uncertainty. An extensive literature in statistics examines model uncertainty through information criteria, which

are used as measures of model fitness. A standard approach in this literature is to use the lowest value of the information criteria among competing models for the basis of model selection. See (78) (and the references therein) for a review. As an example, (79) compares logistic models for non-response that differ in the number and types of conditioning covariates. Its statistic of choice for comparing the predictive power of the various models is a deviance statistic, which is an information criterion often used to examine nested models. (80) estimates different nested attrition probits. Its statistic of choice is the R-squared goodness-of-fit measure for binary choice models. In this paper we adopt the Akaike Information Criterion (AIC) (81) as a more general criterion for model comparison among non-nested models. The AIC penalizes models for the number of predictor variables used. Specifically, we choose the set of covariates that minimize AIC among all combinations of pre-program variables where the difference in means between attrited and non-attrited groups is statistically significant as determined from block permutation tests. As in (79), the probability of participation in the biomedical sweep in our case depends both on baseline characteristics and past history.

IPW covariate selection is made using the following steps:

1. For each gender and for a fixed number of covariates k , we run logit models on all possible combinations of k pre-program variables in addition to the treatment status and a constant term.
2. We then select the set of covariates associated with the lowest value of the Akaike Information Criteria among all these combinations.
3. We reiterate on steps 1 and 2 for different numbers of covariates k .
4. We select the maximum number k that does not generate perfect forecasts of the attrition indicator among the logit regressions that generate the lowest value of AIC.

We apply a special method for selecting the IPW covariates for the equation predicting

attrition at age 35. In this case, we search for IPW covariates not only among the pre-program variables, but also among the age 30 variables. Table S1 presents the selection of IPW covariates for females. Table S2 presents the selection of IPW covariates for males. It turns out that the maximum number k that does not generate perfect forecasts of the attrition indicator for the logit regressions is 4.

Table S1: Selection of IPW Covariates, Males

Age	3 Covariates	4 Covariates	5 Covariates
3 Months	Low Birthweight	High-Risk Index component 4	High-Risk Index component 5
	High-Risk Index component 10	Mother WAIS Comprehension Scale	High-Risk Index component 10
	High-Risk Index	Mother WAIS Digit Symbol	Father's Age at Entry
	-	High-Risk Index component 10	Mother WAIS IQ
	-	-	Low Birthweight
6 months	Low Birthweight	High-Risk Index component 4	High-Risk Index component 5
	High-Risk Index component 10	Mother WAIS Comprehension Scale	High-Risk Index component 10
	High-Risk Index	Mother WAIS Digit Symbol	Father's Age at Entry
	-	High-Risk Index component 10	Mother WAIS IQ
	-	-	Low Birthweight
9 months	Mother WAIS Block Design	Mother WAIS Performance IQ Score	High-Risk Index component 10
	Prematurity	Mother WAIS Digit Symbol	Mother WAIS IQ
	High-Risk Index component 10	Mother WAIS Block Design	Mother WAIS Performance IQ
	-	High-Risk Index component 10	Mother WAIS Comprehension Scale
	-	-	Mother WAIS Digit Symbol
12 months	Mother WAIS Block Design	Mother WAIS Performance IQ Score	High-Risk Index component 10
	Prematurity	Mother WAIS Digit Symbol	Mother WAIS IQ
	High-Risk Index component 10	Mother WAIS Block Design	Mother WAIS Performance IQ
	-	High-Risk Index component 10	Mother WAIS Comprehension Scale
	-	-	Mother WAIS Digit Symbol
18 months	Mother WAIS Digit Symbol	High-Risk Index component 10	Mother WAIS Digit Symbol
	High-Risk Index component 10	Mother WAIS Digit Symbol	High-Risk Index component 10
	Firstborn	High-Risk Index	Prematurity
	-	Firstborn	Birthweight
	-	-	Low Birthweight
24 months	Mother WAIS Digit Symbol	High-Risk Index component 10	Mother WAIS Digit Symbol
	High-Risk Index component 10	Mother WAIS Digit Symbol	High-Risk Index component 10
	Firstborn	High-Risk Index	Prematurity
	-	Firstborn	Birthweight
	-	-	Low Birthweight
36 months	Mother WAIS Digit Symbol	High-Risk Index component 10	High-Risk Index component 4
	High-Risk Index component 10	Mother WAIS Digit Symbol	Mother WAIS Comprehension Scale
	Firstborn	High-Risk Index	Mother WAIS Digit Symbol
	-	Firstborn	High-Risk Index component 10
	-	-	Firstborn

This table reports the selection of IPW covariates for males at different ages. Our selection of IPW covariates is based on a method that minimizes the Akaike Information Criteria among the various estimated logit models. Our selection is based on the following four basic steps. First we fix the number of covariates at k . Second, for each fixed number of covariates, we run logit regressions of the attrition indicator on the treatment status, a constant term and all possible combinations of k covariates that are statistically imbalanced between the attrited and non-attrited groups for each age. We then select the set of variables that generates the lowest information criteria.

Age	3 Covariates	4 Covariates	5 Covariates
48 months	Mother WAIS Digit Symbol	High-Risk Index component 10	High-Risk Index component 4
	High-Risk Index component 10	Mother WAIS Digit Symbol	Mother WAIS Comprehension Scale
	Firstborn	High-Risk Index	Mother WAIS Digit Symbol
	-	Firstborn	High-Risk Index component 10
	-	-	Firstborn
60 months	Mother WAIS Digit Symbol	High-Risk Index component 10	High-Risk Index component 4
	High-Risk Index component 10	Mother WAIS Digit Symbol	Mother WAIS Comprehension Scale
	Firstborn	High-Risk Index	Mother WAIS Digit Symbol
	-	Firstborn	High-Risk Index component 10
	-	-	Firstborn
96 months	High-Risk Index component 5	High-Risk Index component 5	High-Risk Index component 5
	Mother WAIS Block Design	High-Risk Index component 10	Mother WAIS Comprehension Scale
	Firstborn	Mother WAIS Block Design	Firstborn
	-	Firstborn	Birthlength
	-	-	High-Risk Index component 10
21 years	High-Risk Index	High-Risk Index component 10	High-Risk Index component 5
	High-Risk Index component 10	High-Risk Index component 4	Mother WAIS IQ
	Low Birthweight	Mother WAIS Comprehension Scale	Father's Age at Entry
	-	Mother WAIS Digit Symbol	Low Birthweight
	-	-	High-Risk Index component 10
30 years	Low Birthweight	High-Risk Index component 10	High-Risk Index component 4
	High-Risk Index component 10	High-Risk Index component 13	Mother WAIS Digit Symbol
	Firstborn	Birthlength	Number of siblings
	-	Firstborn	High-Risk Index component 10
	-	-	Firstborn
35 years	Prematurity	Prematurity	High-Risk Index component 5
	Substance abuse score at age 30	Substance abuse score at age 30	Mother WAIS Comprehension
	Exam for illness at age 30	Exam for illness at age 30	Prematurity
	-	ADH Problems Score at age 30	Marijuana Usage at age 30
	-	-	Substance abuse score at age 30

This table is a continuation of Table S1 and reports the selection of IPW covariates for males at different ages. Our selection of IPW covariates is based on a method that minimizes the Akaike Information Criteria among the various estimated logit models. Our selection is based on the following four basic steps. First we fix the number of covariates at k . Second, for each fixed number of covariates, we run logit regressions of the attrition indicator on the treatment status, a constant term and all possible combinations of k covariates that are statistically imbalanced between the attrited and non-attrited groups for each age. We then select the set of variables that generates the lowest information criteria.

Table S2: Selection of IPW Covariates, Females

Age	3 Covariates	4 Covariates	5 Covariates
3 Months	High-Risk Index component 10	High-Risk Index component 4	High-Risk Index component 4
	High-Risk Index component 13	High-Risk Index component 13	Mother WAIS Digit Symbol
	Firstborn	High-Risk Index component 10	Mother married
	-	Prematurity	Prematurity
	-	-	High-Risk Index component 10
6 months	High-Risk Index component 10	High-Risk Index component 4	High-Risk Index component 4
	High-Risk Index component 13	High-Risk Index component 13	Mother WAIS Digit Symbol
	Firstborn	High-Risk Index component 10	Mother married
	-	Prematurity	Prematurity
	-	-	High-Risk Index component 10
9 months	High-Risk Index component 4	High-Risk Index component 4	Mother WAIS Performance IQ
	Mother married	Mother married	Mother WAIS Block Design
	Prematurity	Prematurity	Mother working
	-	High-Risk Index component 10	High-Risk Index component 10
	-	-	Prematurity
12 months	High-Risk Index component 10	High-Risk Index component 4	Mother WAIS Digit Symbol
	High-Risk Index component 13	High-Risk Index component 13	High-Risk Index component 4
	Firstborn	High-Risk Index component 10	Mother married
	-	Prematurity	Prematurity
	-	-	High-Risk Index component 10
18 months	High-Risk Index component 10	High-Risk Index component 4	High-Risk Index component 4
	High-Risk Index component 13	High-Risk Index component 13	Mother WAIS Digit Symbol
	Firstborn	High-Risk Index component 10	Mother married
	-	Prematurity	Prematurity
	-	-	High-Risk Index component 10
24 months	High-Risk Index component 10	High-Risk Index component 4	High-Risk Index component 4
	High-Risk Index component 13	High-Risk Index component 13	Mother WAIS Digit Symbol
	Firstborn	High-Risk Index component 10	Mother married
	-	Prematurity	Prematurity
	-	-	High-Risk Index component 10
36 months	High-Risk Index component 10	High-Risk Index component 4	High-Risk Index component 4
	High-Risk Index component 13	High-Risk Index component 13	Mother WAIS Digit Symbol
	Firstborn	High-Risk Index component 10	Mother married
	-	Prematurity	Prematurity
	-	-	High-Risk Index component 10

This table reports the selection of IPW covariates for females at different ages. Our selection of IPW covariates is based on a method that minimizes the Akaike Information Criteria among the various estimated logit models. It is also age and gender specific. Our selection is based on the following four basic steps. First we fix the number of covariates at k . Second, for each fixed number of covariates, we run logit regressions of the attrition indicator on the treatment status, a constant term and all possible combinations of k covariates that are statistically imbalanced between the attrited and non-attrited groups for each age. We then select the set of variables that generates the lowest information criteria.

48 months	High-Risk Index component 10	High-Risk Index component 4	High-Risk Index component 4
	High-Risk Index component 13	High-Risk Index component 13	Mother WAIS Digit Symbol
	Firstborn	High-Risk Index component 10	Mother married
	-	Prematurity	Prematurity
	-	-	High-Risk Index component 10
60 months	High-Risk Index component 5	Mother WAIS Digit Symbol	Mother WAIS Block Design
	Birthlength	Birthlength	Father working
	Firstborn	Prematurity	Birthlength
	-	Birthweight	Birthweight
	-	-	Firstborn
96 months	Birthweight	Mother's Age at Entry	Mother WAIS Block Design
	Birthlength	Father's Age at Entry	Father working
	Firstborn	Low Birthweight	Mother working
	-	Firstborn	Birthlength
	-	-	Firstborn
21 years	High-Risk Index component 5	High-Risk Index component 10	Mother WAIS Performance IQ
	Prematurity	Mother WAIS full scale IQ	Prematurity
	High-Risk Index component 10	Firstborn	High-Risk Index component 10
	-	High-Risk Index component 2	Mother WAIS full scale IQ
	-	-	High-Risk Index component 2
30 years	High-Risk Index component 5	High-Risk Index component 10	Mother WAIS Performance IQ
	Prematurity	Mother WAIS full scale IQ	Prematurity
	High-Risk Index component 10	Firstborn	High-Risk Index component 10
	-	High-Risk Index component 2	Mother WAIS full scale IQ
	-	-	High-Risk Index component 2
35 years	Prematurity	Mother WAIS Digit Symbol	Father working
	Substance abuse score at age 30	Prematurity	Birthlength
	Rule Breaking Score at age 30	Substance abuse score at age 30	Low Birthweight
	-	Rule Breaking Score at age 30	Prematurity
	-	-	Substance abuse score at age 30

This table is a continuation of Table S2 and reports the selection of IPW covariates for females at different ages. Our selection of IPW covariates is based on a method that minimizes the Akaike Information Criteria among the various estimated logit models. Our selection is based on the following four basic steps. First we fix the number of covariates at k . Second, for each fixed number of covariates, we run logit regressions of the attrition indicator on the treatment status, a constant term and all possible combinations of k covariates that are statistically imbalanced between the attrited and non-attrited groups for each age. We then select the set of variables that generates the lowest information criteria.

D Attrition

Attrition Among the 111 children who enrolled in the intervention, 16 attrited from the study sample for various reasons before the end of the preschool treatment (82). The remaining set of 95 participants consists of 49 treated and 46 controls, resulting in overall retention rates of 86% since the start of enrollment or 78% for the randomized sample. After age 5, a new randomization reassigned treatment and control status among participants for the second phase of the ABC project. After the second randomization, three children who were assigned to be controls in the first phase and treated in the second did not receive the assigned intervention, so that the sample size was further reduced. Attrition patterns across treatment assignment and gender at all waves included in our analysis are shown in Table S3.

A severe loss in the sample was experienced in the biomedical sweep. In addition to the general reasons which lead to a lower response rate in biomedical surveys, there was one specific to this data collection: the fact that the medical visits had to be scheduled in the physician's office at very specific times, with little ability to reschedule and many missed appointments.

Among the 101 individuals who participated in the age 30 sweep, only 72 took part in the biomedical sweep. Table S3 shows that the worst loss was experienced for males belonging to the control group, where only 12 out of the 21 who were surveyed at age 30 took part in the biomedical sweep; on the other hand, the highest retention rate of 78% was obtained for the females in the control group. Additionally, out of these 72 individuals, only 68 of them underwent the physical exam: 4 of them only participated in the blood data collection (all males, 3 belonging to the control group).

Given the extent of the attrition in the biomedical sweep, and the relevance of the information collected there for our analysis, here we provide a detailed description of the

characteristics of the attriters. We compare the attriters vs. the non-attriters, by gender and treatment status (83), with respect to four sets of baseline variables: components of the High-Risk Index, maternal IQ, socioeconomic characteristics at intake and birth-related information; and with respect to different sets of covariates observed in the previous sweep: lifestyles (smoking, drinking, marijuana use), health care availability, physical and mental health. Tables S4–S9 report all the variables that are statistically significantly different between attrited and non-attrited at a 10% level (or lower) after having performed permutation.

We find a consistent pattern of selection into the biomedical sweep, which does not vary significantly by treatment status and gender. First, those who participated in the biomedical sweep were in general positively selected with respect to baseline characteristics: treated females who participated were less likely to have an absent father or no maternal relatives in the area (see Table S4) (84); control males who participated in either the physical exam or the lab tests were born to younger mothers with higher IQ (Tables S7 and S9); non-attrited males of both treatment groups had on average fewer siblings at the baseline (Tables S6–S9); and treated males who took part in the physical exam were significantly longer at birth, less likely to be low birthweight and less likely to be premature than those who attrited (Table S6). This pattern is typical. Individuals with more favorable traits and circumstances are more likely to be retained in the sample. Second, those who participated in the biomedical sweep were negatively selected with respect to age 30 characteristics: the females were more likely to be smokers and to use substances, and to have high blood pressure (Tables S4–S5) (85); non-attrited males were also more likely to engage in unhealthy behaviors, were less likely to have availability of health care at age 30, and were significantly more likely than the non-attrited to have poorer mental health, as reflected in the Achenbach DSM and problem scales (Tables S6–S9) (86). This suggests that some individuals might have participated in the biomedical sweep to obtain free health care services.

Hence, on the basis of this analysis, the full set of variables which has been selected to perform our procedure for IPW variable selection for the biomedical sweep (see Section C and Tables S1–S2) is: HRI component 1 (father absent), HRI component 2 (no maternal relatives in the area), HRI component 10 (other special circumstances), mother's WAIS performance IQ, mother's WAIS comprehension scale score, mother's WAIS digit symbol, mother's WAIS block design, mother's and father's age at entry in the program, father working at intake, mother married at intake, number of siblings, birth length, low birth weight, prematurity, ever smoker by age 30, number of cigarettes smoked in the past 30 days at age 30, Achenbach substance use scale tobacco (*t*-score) at age 30, number of times used marijuana during life at age 30, Achenbach substance use scale mean substance abuse (*t*-score) at age 30, Achenbach Adult Self-Report "Drink too much" at age 30, Achenbach Adult Self-Report indicator for having any illness at age 30, Adult Self-Report indicator for having a headache at age 30, high blood pressure (self-reported) at age 30, availability of health insurance at age 30, indicator for not having an exam for illness or injury in the past 2 years at age 30, indicator for not having a routine physical exam in the past 2 years at age 30, Achenbach DSM scale Attention Deficit/Hyperactivity problem (*t*-score) at age 30, Achenbach DSM scale Antisocial Personality problems (*t*-score) at age 30, Achenbach ASR problem scale Attention Problems (*t*-score) at age 30, and Achenbach ASR problem scale Rule Breaking (*t*-score) at age 30 (87).

Table S3: Attrition in the Abecedarian Intervention

Sweep	Control		Treated		Control		Treated	
	Males		Males		Females		Females	
	Followed	Attrited	Followed	Attrited	Followed	Attrited	Followed	Attrited
Start	23	0	29	0	31	0	28	0
3 months	22	1	28	1	26	5	23	5
6 months	22	1	25	4	27	4	24	4
9 months	20	3	16	13	26	5	15	13
12 months	21	2	27	2	25	6	24	4
18 months	20	3	26	3	27	4	22	6
24 months	15	8	27	2	17	14	23	5
36 months	19	4	25	4	23	8	21	7
48 months	20	3	24	5	26	5	22	6
60 months	20	3	24	5	23	8	22	6
96 months	19	4	26	3	23	8	21	7
Age 21	23	0	26	3	28	3	25	3
Age 30	21	3	27	2	28	3	25	3
Age 35 Lab	12	11	20	9	22	9	18	10
Age 35 PE	9	14	19	10	22	9	18	10

Notes: This table presents the sample sizes available in all the waves used in the analysis for the Abecedarian intervention. The first column shows the sweep of interest. The following eight columns present, by gender and treatment status, the number of individuals who have been followed-up in the respective sweep, and the number of the attriters. The sample sizes for the sweeps since 3 months up to 96 months refer to those who participated in the growth measurements. The sample size for the age 21 sweep refers to the *Risk Taking Survey*. The sample size for the age 30 sweep refers to the *Tobacco, Alcohol and Drugs (TAD) Survey*. Age 35 Lab refers to the sample of those who took part in the lab blood tests in the biomedical sweep. Age 35 PE refers to the sample of those who took part in the physical exam.

Table S4: Differences between Attrited and Non-Attrited: (Treated, Females)

Variable	Age	Sample Sizes		Attrited Mean	Difference in Means	Asy. <i>p</i> -value	Permutation <i>p</i> -value
		# Attrited	# Non-Attrited				
Baseline							
HRI component 1: Father absent	0	10	17	1.000	-0.235	0.096	0.070
HRI component 2: No maternal relatives in the area	0	10	17	0.200	-0.200	0.052	0.036
HRI component 10: Other special circumstances	0	10	17	0.100	0.312	0.086	0.072
Mother WAIS performance IQ score	0	10	18	87.00	-0.833	0.880	0.894
Mother's WAIS comprehension scale score	0	10	17	8.30	-0.535	0.618	0.619
Mother's WAIS digit symbol score	0	10	17	9.10	-0.806	0.457	0.497
Mother's WAIS block design score	0	10	17	7.30	-0.771	0.411	0.435
Mother's age at intake	0	10	18	19.60	-0.267	0.843	0.854
Father's age at intake	0	8	18	22.62	-0.681	0.749	0.747
Father working at intake	0	8	18	0.875	-0.264	0.185	0.142
Mother married at intake	0	10	18	0.000	0.222	0.108	0.075
Number of siblings	0	10	18	0.500	-0.056	0.853	0.871
Length at birth	0	10	18	49.80	-1.022	0.359	0.361
Low birth weight	0	10	18	0.100	0.011	0.930	0.905
Prematurity (born before <37 weeks of gestation)	0	6	18	0.167	-0.111	0.410	0.540
Age 30							
Ever tried smoking	30	7	18	0.714	-0.048	0.827	0.788
Number of cigarettes smoked per day in the past 30 days	30	7	18	0.000	2.583	0.112	0.007
Achenbach substance use scale tobacco score	30	7	18	50.00	2.556	0.089	0.035
Number of times used marijuana during life	30	7	18	2.286	26.80	0.129	0.066
Achenbach substance use scale substance abuse score	30	7	17	50.71	2.756	0.392	0.190
Achenbach adult self-report "Drink too much"	30	7	18	0.000	0.167	0.261	0.236
Achenbach adult self-report indicator "Do you have any illness"	30	7	17	0.143	-0.025	0.871	0.814
Achenbach adult self-report indicator for having a headache	30	7	18	0.429	0.016	0.952	0.937
High blood pressure (self-reported)	30	7	18	0.000	0.278	0.121	0.064
No health insurance	30	7	18	0.143	0.135	0.494	0.448
No routine physical exam in the last 2 years	30	7	18	0.000	0.056	0.540	0.923
No exam for injury or illness in the last 2 years	30	7	18	0.571	-0.183	0.425	0.429
Achenbach DSM scale Attention Deficit/Hyperactivity problems score	30	7	18	53.71	-1.825	0.401	0.585
Achenbach DSM scale Antisocial Personality problems score	30	7	18	52.86	0.087	0.962	0.967
Achenbach ASR problem scale Attention Problems score	30	7	18	54.57	-1.627	0.500	0.547
Achenbach ASR problem scale Rule Breaking score	30	7	18	51.57	-1.540	0.318	0.214

Notes: This table shows the differences in baseline and age 30 characteristics between attrited and non-attrited at the biomedical sweep for the treatment group of the Abecedarian females. For females there is no difference in the sample size between those who participated in the physical exam and in the blood test collection. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** age of the participant when the data was collected; **Col.3:** attrited sample size; **Col.4:** non-attrited sample size; **Col.5:** attrited arithmetic mean; **Col.6:** unconditional difference in means across attrited and non-attrited; **Col.7:** the asymptotic *p*-value of the two-sided single hypothesis based on the *t*-statistic associated with the unconditional difference in means; **Col.8:** single hypothesis two-sided permutation *p*-value based on an unconstrained permutation scheme (30,000 permutation draws).

Table S5: Differences between Attrited and Non-Attrited: (Control, Females)

Variable	Age	Sample Sizes		Attrited Mean	Difference in Means	Asy. <i>p</i> -value	Permutation <i>p</i> -value
		# Attrited	# Non-Attrited				
Baseline							
HRI component 1: Father absent	0	9	22	0.778	0.086	0.568	0.615
HRI component 2: No maternal relatives in the area	0	9	22	0.111	-0.111	0.114	0.025
HRI component 10: Other special circumstances	0	9	22	0.111	0.253	0.165	0.119
Mother WAIS performance IQ score	0	9	22	84.89	-1.662	0.730	0.741
Mother's WAIS comprehension scale score	0	9	22	6.444	1.237	0.178	0.078
Mother's WAIS digit symbol score	0	9	22	9.333	-0.515	0.543	0.497
Mother's WAIS block design score	0	9	22	6.889	-0.753	0.379	0.394
Mother's age at intake	0	9	22	18.44	1.237	0.459	0.444
Father's age at intake	0	9	22	20.56	2.763	0.205	0.127
Father working at intake	0	9	21	0.778	-0.159	0.412	0.380
Mother married at intake	0	9	22	0.111	0.071	0.639	0.646
Number of siblings	0	9	22	0.556	0.172	0.701	0.667
Length at birth	0	7	21	48.29	2.571	0.149	0.465
Low birth weight	0	8	22	0.250	-0.205	0.099	0.288
Prematurity (born before <37 weeks of gestation)	0	6	22	0.167	-0.167	0.048	0.003
Age 30							
Ever tried smoking	30	6	22	1.000	-0.227	0.205	0.084
Number of cigarettes smoked per day in the past 30 days	30	6	22	2.667	-0.621	0.756	0.820
Achenbach substance use scale tobacco score	30	5	22	54.20	-1.109	0.640	0.751
Number of times used marijuana during life	30	6	22	33.58	-17.31	0.304	0.432
Achenbach substance use scale substance abuse score	30	4	22	50.25	1.977	0.331	0.009
Achenbach adult self-report "Drink too much"	30	6	22	0.167	-0.121	0.319	0.533
Achenbach adult self-report indicator "Do you have any illness"	30	6	22	0.000	0.227	0.205	0.085
Achenbach adult self-report indicator for having a headache	30	6	22	0.000	0.591	0.018	0.002
High blood pressure (self-reported)	30	6	22	0.000	0.227	0.205	0.088
No health insurance	30	6	22	0.167	-0.030	0.857	0.854
No routine physical exam in the last 2 years	30	6	22	0.000	0.136	0.352	0.280
No exam for injury or illness in the last 2 years	30	6	22	0.667	-0.303	0.190	0.212
Achenbach DSM scale Attention Deficit/Hyperactivity problems score	30	6	22	51.83	3.348	0.367	0.195
Achenbach DSM scale Antisocial Personality problems score	30	6	22	54.33	2.258	0.442	0.317
Achenbach ASR problem scale Attention Problems score	30	6	22	51.17	3.697	0.166	0.051
Achenbach ASR problem scale Rule Breaking score	30	6	22	53.83	2.258	0.473	0.420

Notes: This table shows the differences in baseline and age 30 characteristics between attrited and non-attrited at the biomedical sweep for the control group of the Abecedarian females. For females there is no difference in the sample size between those who participated in the physical exam and in the blood test collection. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** age of the participant when the data was collected; **Col.3:** attrited sample size; **Col.4:** non-attrited sample size; **Col.5:** attrited arithmetic mean; **Col.6:** unconditional difference in means across attrited and non-attrited; **Col.7:** the asymptotic *p*-value of the two-sided single hypothesis based on the *t*-statistic associated with the unconditional difference in means; **Col.8:** single hypothesis two-sided permutation *p*-value based on an unconstrained permutation scheme (30,000 permutation draws).

Table S6: Differences between Attrited and Non-Attrited: (Treated, PE, Males)

Variable	Age	Sample Sizes		Attrited Mean	Difference in Means	Asy. <i>p</i> -value	Permutation <i>p</i> -value
		# Attrited	# Non-Attrited				
Baseline							
HRI component 1: Father absent	0	9	19	0.667	0.175	0.303	0.348
HRI component 2: No maternal relatives in the area	0	9	19	0.000	0.000	-	-
HRI component 10: Other special circumstances	0	9	19	0.333	0.035	0.862	0.877
Mother WAIS performance IQ score	0	10	19	88.50	-1.079	0.807	0.810
Mother's WAIS comprehension scale score	0	10	19	8.800	-1.379	0.228	0.305
Mother's WAIS digit symbol score	0	10	19	9.800	0.095	0.939	0.943
Mother's WAIS block design score	0	10	19	7.600	-1.179	0.213	0.200
Mother's age at intake	0	10	19	21.20	-2.305	0.166	0.276
Father's age at intake	0	9	19	23.33	-0.333	0.850	0.871
Father working at intake	0	9	19	0.444	0.398	0.025	0.057
Mother married at intake	0	10	19	0.300	-0.195	0.195	0.245
Number of siblings	0	10	19	1.100	-0.837	0.042	0.130
Length at birth	0	9	18	46.78	3.500	0.015	0.054
Low birth weight	0	9	19	0.222	-0.222	0.029	0.009
Prematurity (born before <37 weeks of gestation)	0	8	19	0.375	-0.322	0.027	0.141
Age 30							
Ever tried smoking	30	8	19	0.750	0.197	0.140	0.269
Number of cigarettes smoked per day in the past 30 days	30	8	19	1.500	3.000	0.164	0.092
Achenbach substance use scale tobacco score	30	7	18	52.57	1.095	0.499	0.433
Number of times used marijuana during life	30	8	19	26.81	32.66	0.068	0.056
Achenbach substance use scale substance abuse score	30	7	18	51.57	2.706	0.132	0.065
Achenbach adult self-report "Drink too much"	30	8	19	0.125	0.243	0.348	0.247
Achenbach adult self-report indicator "Do you have any illness"	30	8	18	0.125	0.125	0.129	0.026
Achenbach adult self-report indicator for having a headache	30	8	19	0.000	0.263	0.110	0.059
High blood pressure (self-reported)	30	8	19	0.000	0.105	0.355	0.464
No health insurance	30	8	19	0.250	0.066	0.743	0.754
No routine physical exam in the last 2 years	30	8	19	0.000	0.263	0.110	0.054
No exam for injury or illness in the last 2 years	30	8	19	0.875	-0.401	0.049	0.031
Achenbach DSM scale Attention Deficit/Hyperactivity problems score	30	8	19	51.75	3.776	0.085	0.021
Achenbach DSM scale Antisocial Personality problems score	30	8	19	53.00	4.474	0.136	0.064
Achenbach ASR problem scale Attention Problems score	30	8	19	51.00	5.105	0.059	0.004
Achenbach ASR problem scale Rule Breaking score	30	8	19	52.62	4.796	0.108	0.045

Notes: This table shows the differences in baseline and age 30 characteristics between attrited and non-attrited at the biomedical sweep for the treatment group of the Abecedarian males who participated in the physical exam (PE). For males there is a difference in the sample size between those who participated in the physical exam and in the blood test collection. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** age of the participant when the data was collected; **Col.3:** attrited sample size; **Col.4:** non-attrited sample size; **Col.5:** attrited arithmetic mean; **Col.6:** unconditional difference in means across attrited and non-attrited; **Col.7:** the asymptotic *p*-value of the two-sided single hypothesis based on the *t*-statistic associated with the unconditional difference in means; **Col.8:** single hypothesis two-sided permutation *p*-value based on an unconstrained permutation scheme (30,000 permutation draws).

Table S7: Differences between Attrited and Non-Attrited: (Control, PE, Males)

Variable	Age	Sample Sizes		Attrited Mean	Difference in Means	Asy. <i>p</i> -value	Permutation <i>p</i> -value
		# Attrited	# Non-Attrited				
Baseline							
HRI component 1: Father absent	0	14	9	0.643	0.024	0.911	0.950
HRI component 2: No maternal relatives in the area	0	14	9	0.000	0.222	0.061	0.042
HRI component 10: Other special circumstances	0	14	9	0.143	0.302	0.109	0.153
Mother WAIS performance IQ score	0	14	9	84.86	8.143	0.025	0.025
Mother's WAIS comprehension scale score	0	14	8	9.214	-1.839	0.096	0.101
Mother's WAIS digit symbol score	0	14	9	8.571	1.984	0.021	0.027
Mother's WAIS block design score	0	14	9	6.643	1.468	0.110	0.081
Mother's age at intake	0	14	9	23.43	-4.762	0.121	0.089
Father's age at intake	0	14	9	26.57	-1.794	0.623	0.619
Father working at intake	0	14	9	0.929	-0.151	0.309	0.354
Mother married at intake	0	14	9	0.357	-0.024	0.911	0.949
Number of siblings	0	14	9	1.071	-0.738	0.186	0.141
Length at birth	0	14	9	49.36	0.310	0.818	0.812
Low birth weight	0	14	9	0.071	0.151	0.309	0.350
Prematurity (born before <37 weeks of gestation)	0	14	9	0.143	-0.032	0.834	0.845
Age 30							
Ever tried smoking	30	12	8	0.667	0.208	0.308	0.282
Number of cigarettes smoked per day in the past 30 days	30	12	8	1.000	5.312	0.013	0.054
Achenbach substance use scale tobacco score	30	9	6	50.78	3.889	0.016	0.088
Number of times used marijuana during life	30	12	8	32.29	36.02	0.061	0.079
Achenbach substance use scale substance abuse score	30	9	5	51.67	4.533	0.007	0.050
Achenbach adult self-report "Drink too much"	30	12	9	0.000	0.222	0.084	0.070
Achenbach adult self-report indicator "Do you have any illness"	30	11	9	0.182	0.040	0.832	0.863
Achenbach adult self-report indicator for having a headache	30	12	9	0.083	0.139	0.386	0.414
High blood pressure (self-reported)	30	12	9	0.167	0.056	0.761	0.781
No health insurance	30	12	9	0.333	0.444	0.037	0.044
No routine physical exam in the last 2 years	30	11	9	0.273	0.283	0.209	0.216
No exam for injury or illness in the last 2 years	30	12	9	0.417	-0.083	0.712	0.703
Achenbach DSM scale Attention Deficit/Hyperactivity problems score	30	12	9	51.42	2.250	0.028	0.068
Achenbach DSM scale Antisocial Personality problems score	30	12	9	52.33	4.556	0.078	0.129
Achenbach ASR problem scale Attention Problems score	30	12	9	51.58	1.861	0.098	0.131
Achenbach ASR problem scale Rule Breaking score	30	12	9	52.42	4.583	0.046	0.069

Notes: This table shows the differences in baseline and age 30 characteristics between attrited and non-attrited at the biomedical sweep for the control group of the Abecedarian males who participated in the physical exam (PE). For males there is a difference in the sample size between those who participated in the physical exam and in the blood test collection. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** age of the participant when the data was collected; **Col.3:** attrited sample size; **Col.4:** non-attrited sample size; **Col.5:** attrited arithmetic mean; **Col.6:** unconditional difference in means across attrited and non-attrited; **Col.7:** the asymptotic *p*-value of the two-sided single hypothesis based on the *t*-statistic associated with the unconditional difference in means; **Col.8:** single hypothesis two-sided permutation *p*-value based on an unconstrained permutation scheme (30,000 permutation draws).

Table S8: Differences between Attrited and Non-Attrited: (Treated, Lab, Males)

Variable	Age	Sample Sizes		Attrited Mean	Difference in Means	Asy. <i>p</i> -value	Permutation <i>p</i> -value
		# Attrited	# Non-Attrited				
Baseline							
HRI component 1: Father absent	0	8	20	0.625	0.225	0.198	0.275
HRI component 2: No maternal relatives in the area	0	8	20	0.000	0.0000	-	-
HRI component 10: Other special circumstances	0	8	20	0.375	-0.025	0.905	0.913
Mother WAIS performance IQ score	0	9	20	87.89	-0.139	0.976	0.976
Mother's WAIS comprehension scale score	0	9	20	8.778	-1.278	0.279	0.383
Mother's WAIS digit symbol score	0	9	20	9.667	0.283	0.824	0.834
Mother's WAIS block design score	0	9	20	7.444	-0.894	0.363	0.350
Mother's age at intake	0	9	20	21.44	-2.544	0.135	0.276
Father's age at intake	0	8	20	23.62	-0.725	0.689	0.740
Father working at intake	0	8	20	0.375	0.475	0.008	0.030
Mother married at intake	0	9	20	0.333	-0.233	0.127	0.182
Number of siblings	0	9	20	1.111	-0.811	0.057	0.200
Length at birth	0	8	19	47.75	1.934	0.220	0.288
Low birth weight	0	8	20	0.125	-0.075	0.501	0.557
Prematurity (born before <37 weeks of gestation)	0	7	20	0.286	-0.186	0.244	0.340
Age 30							
Ever tried smoking	30	7	20	0.714	0.236	0.087	0.294
Number of cigarettes smoked per day in the past 30 days	30	7	20	1.714	2.561	0.259	0.163
Achenbach substance use scale tobacco score	30	6	19	52.17	1.570	0.353	0.281
Number of times used marijuana during life	30	7	20	20.71	39.26	0.032	0.025
Achenbach substance use scale substance abuse score	30	6	19	51.83	2.219	0.247	0.133
Achenbach adult self-report "Drink too much"	30	7	20	0.000	0.400	0.130	0.040
Achenbach adult self-report indicator "Do you have any illness"	30	7	19	0.000	0.053	0.552	0.937
Achenbach adult self-report indicator for having a headache	30	7	20	0.000	0.250	0.147	0.073
High blood pressure (self-reported)	30	7	20	0.000	0.100	0.400	0.485
No health insurance	30	7	20	0.286	0.014	0.946	0.925
No routine physical exam in the last 2 years	30	7	20	0.000	0.250	0.147	0.072
No exam for injury or illness in the last 2 years	30	7	20	0.857	-0.357	0.099	0.069
Achenbach DSM scale Attention Deficit/Hyperactivity problems score	30	7	20	51.57	3.829	0.095	0.022
Achenbach DSM scale Antisocial Personality problems score	30	7	20	52.00	5.600	0.069	0.014
Achenbach ASR problem scale Attention Problems score	30	7	20	50.86	5.043	0.075	0.005
Achenbach ASR problem scale Rule Breaking score	30	7	20	51.43	6.171	0.042	0.005

Notes: This table shows the differences in baseline and age 30 characteristics between attrited and non-attrited at the biomedical sweep for the treatment group of the Abecedarian males who participated in the lab tests. For males there is a difference in the sample size between those who participated in the physical exam and in the blood test collection. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** age of the participant when the data was collected; **Col.3:** attrited sample size; **Col.4:** non-attrited sample size; **Col.5:** attrited arithmetic mean; **Col.6:** unconditional difference in means across attrited and non-attrited; **Col.7:** the asymptotic *p*-value of the two-sided single hypothesis based on the *t*-statistic associated with the unconditional difference in means; **Col.8:** single hypothesis two-sided permutation *p*-value based on an unconstrained permutation scheme (30,000 permutation draws).

Table S9: Differences between Attrited and Non-Attrited: (Control, Lab, Males)

Variable	Age	Sample Sizes		Attrited Mean	Difference in Means	Asy. <i>p</i> -value	Permutation <i>p</i> -value
		# Attrited	# Non-Attrited				
Baseline							
HRI component 1: Father absent	0	11	12	0.545	0.205	0.319	0.361
HRI component 2: No maternal relatives in the area	0	11	12	0.000	0.167	0.162	0.167
HRI component 10: Other special circumstances	0	11	12	0.182	0.152	0.426	0.449
Mother WAIS performance IQ score	0	11	12	84.18	7.402	0.040	0.063
Mother's WAIS comprehension scale score	0	11	11	9.182	-1.273	0.244	0.261
Mother's WAIS digit symbol score	0	11	12	8.455	1.712	0.046	0.053
Mother's WAIS block design score	0	11	12	6.273	1.811	0.038	0.064
Mother's age at intake	0	11	12	25.18	-6.932	0.014	0.027
Father's age at intake	0	11	12	28.73	-5.477	0.109	0.138
Father working at intake	0	11	12	0.909	-0.076	0.606	0.594
Mother married at intake	0	11	12	0.455	-0.205	0.319	0.353
Number of siblings	0	11	12	1.364	-1.114	0.032	0.045
Length at birth	0	11	12	49.64	-0.303	0.817	0.838
Low birth weight	0	11	12	0.091	0.076	0.606	0.587
Prematurity (born before <37 weeks of gestation)	0	11	12	0.091	0.076	0.606	0.588
Age 30							
Ever tried smoking	30	9	11	0.667	0.152	0.456	0.485
Number of cigarettes smoked per day in the past 30 days	30	9	11	1.333	3.258	0.153	0.168
Achenbach substance use scale tobacco score	30	7	8	51.00	2.500	0.155	0.185
Number of times used marijuana during life	30	9	11	39.61	12.89	0.524	0.524
Achenbach substance use scale substance abuse score	30	7	7	52.14	2.286	0.233	0.280
Achenbach adult self-report "Drink too much"	30	9	12	0.000	0.167	0.207	0.224
Achenbach adult self-report indicator "Do you have any illness"	30	9	11	0.222	-0.040	0.832	0.878
Achenbach adult self-report indicator for having a headache	30	9	12	0.111	0.056	0.733	0.724
High blood pressure (self-reported)	30	9	12	0.111	0.139	0.442	0.424
No health insurance	30	9	12	0.333	0.333	0.133	0.182
No routine physical exam in the last 2 years	30	8	12	0.250	0.250	0.278	0.275
No exam for injury or illness in the last 2 years	30	9	12	0.556	-0.306	0.159	0.185
Achenbach DSM scale Attention Deficit/Hyperactivity problems score	30	9	12	51.33	1.833	0.367	0.380
Achenbach DSM scale Antisocial Personality problems score	30	9	12	52.89	2.444	0.367	0.380
Achenbach ASR problem scale Attention Problems score	30	9	12	51.67	1.250	0.281	0.280
Achenbach ASR problem scale Rule Breaking score	30	9	12	52.44	3.389	0.156	0.173

Notes: This table shows the differences in baseline and age 30 characteristics between attrited and non-attrited at the biomedical sweep for the control group of the Abecedarian males who participated in the lab tests. For males there is a difference in the sample size between those who participated in the physical exam and in the blood test collection. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** age of the participant when the data was collected; **Col.3:** attrited sample size; **Col.4:** non-attrited sample size; **Col.5:** attrited arithmetic mean; **Col.6:** unconditional difference in means across attrited and non-attrited; **Col.7:** the asymptotic *p*-value of the two-sided single hypothesis based on the *t*-statistic associated with the unconditional difference in means; **Col.8:** single hypothesis two-sided permutation *p*-value based on an unconstrained permutation scheme (30,000 permutation draws).

E Estimation results

In this section we provide additional results on the prevalence of behavioral risk factors, and then we report all the complete estimation results presented in the paper. Results on behavioral risk factors are displayed in Table S10. While the effects of the intervention on health are pervasive across multiple domains, the effects on behavioral risk factors are less pronounced and are not always in beneficial directions. Treated females are less likely to have ever been regular smokers by age 30 (diff.=0.243), although the treatment effect is only marginally significant ($p = 0.071$). However, if they are smokers, on average they have an earlier age of onset of smoking, although the difference is not statistically significant at conventional levels. In contrast, treated males, if smokers, have a significantly delayed onset—by almost 3 years (diff.=2.829; $p = 0.051$). Gender differences are also found with respect to alcohol use. Treated males show a trend towards early and more intensive use of alcohol, although these differences never achieve statistical significance at conventional levels. Treated females, instead, are significantly less likely to have experienced early onset drinking (0.280 vs. 0.571; $p = 0.011$).

There are no statistically significant differences between treatments and controls with respect to marijuana use, apart from a higher age of onset for treated males (17.5 vs. 15.5; $p = 0.052$). Finally, treated women are also more likely to engage in physical activity (0.320 vs. 0.071; $p = 0.004$) and to have a diet richer in fruit (0.800 vs. 0.286; $p = 0.004$) at age 21. Information is not available in the latest surveys.

Tables S11 and S12 display the full results for Tables 1 and 2 (Biomedical Sweep). Tables S13 and S14 display the full results for Tables 3, 4 and S10 (Health Care, Physical Development and Behavioral Risk Factors). In addition to the information reported in Tables 1-4, these supplementary tables provide sample sizes for control and treatment groups, asymptotic and

permutation-based p -values which do not account for attrition, and a test for gender difference in the effect of the treatment.

The availability of a wider set of results allows us to make two observations. First, in general the asymptotic p -values are too small, so that the treatment effects in a few cases are no longer significant at conventional levels, when permutation-based inference is applied. This occurs, for example, in the case of pre-hypertension (Table S12) and physical development (Table S14) for females, or dyslipidemia and severe obesity for males (Table S11). Second, accounting for attrition by means of IPW often leads to lower p -values as compared to the cases in which only block permutation is applied. This is evident especially for males, for the outcomes of blood pressure, hypertension, and for the Framingham Risk Score (Table S11).

Then, the tests for gender differences in the treatment effects reveal that, for the biomedical sweep, such differences emerge with respect to hypertension-related outcomes and assessment of nutritional status. In these cases, treated males are significantly better off than controls, while treated females are not—although in the latter case the differences between the two treatment arms never achieve statistical significance. Another gender-related difference emerges in relation to health care coverage at age 30: different from males, treated females are significantly less likely to have health insurance coverage. Finally, treated males are less likely to exercise 4 or more days per week at age 21, although this effect does not achieve statistical significance at conventional levels.

Table S10: Inference for the Abecedarian Intervention, Behavioral Risk Factors

Variable	Control Mean	Treatment Mean	Difference in Means	Conditional Treatment Effect	Block <i>p</i> -value	Step-Down <i>p</i> -value	Control Mean	Treatment Mean	Difference in Means	Conditional Treatment Effect	Block <i>p</i> -value	Step-Down <i>p</i> -value
<i>Males</i>						<i>Females</i>						
<i>Behavioral Risk Factors</i>												
Smoking												
Never a Regular Smoker by age 30	0.391	0.357	-0.034	-0.042	0.571	0.571	0.357	0.600	0.243	0.221	0.071	0.195
Age of Onset of Regular Smoking	16.893	19.722	2.829	2.660	0.051	0.143	17.861	17.050	-0.811	-1.544	0.848	0.848
Cigarettes Smoked per Day Past 30 Days at ages 21&30	11.167	8.685	2.482	3.820	0.199	0.344	10.294	10.450	-0.156	-0.939	0.567	0.807
Alcohol Use												
Early Onset Drinker (before age 17)	0.609	0.539	0.070	0.074	0.336	0.762	0.571	0.280	0.291	0.344	0.011	0.046
Drank Driving	0.391	0.538	-0.147	-0.090	0.663	0.877	0.222	0.120	0.102	0.100	0.172	0.358
Age of Onset of Alcohol Use	17.045	15.910	-1.135	-1.306	0.840	0.840	16.685	17.727	1.042	1.178	0.099	0.292
Days At Least 1 Drink Past 30 Days at ages 21&30	9.348	11.982	-2.634	-1.880	0.498	0.860	8.732	5.960	2.772	1.105	0.211	0.292
Days 5+ Drinks Past 30 Days at ages 21&30	2.457	3.429	-0.972	-0.946	0.585	0.900	2.321	1.260	1.061	-0.095	0.366	0.366
Marijuana Use												
Never Used Marijuana by age 30	0.130	0.107	-0.023	-0.019	0.476	0.476	0.250	0.400	0.150	0.089	0.169	0.402
Age of Onset of Marijuana Use	15.475	17.500	2.025	1.763	0.052	0.136	17.190	17.866	0.676	0.828	0.274	0.471
Times Used Marijuana Past 30 Days at ages 21&30	49.565	48.535	1.030	2.699	0.357	0.527	19.982	21.580	-1.598	-2.089	0.417	0.417
Physical Activity and Diet												
Exercise 4 or more Days per Week at age 21	0.391	0.307	-0.084	-0.094	0.858	0.858	0.071	0.320	0.249	0.287	0.004	0.008
No. of Fruit Servings per Day at age 21	0.826	0.846	0.020	-0.015	0.524	0.745	0.286	0.800	0.514	0.518	0.004	0.005

Notes: This table presents the inference and descriptive statistics of selected outcomes of the Abecedarian Intervention. The first column describes the outcome analyzed. The remaining twelve columns present the statistical analysis for males (first six columns) and females (remaining six columns). The columns present the following information for each gender: (1) control mean; (2) treatment mean; (3) unconditional difference in means across treatment and control groups. We multiply the difference in means by (-1) when a higher value of the variable in the raw data represents a worse outcome so that all outcomes are *normalized* in a favorable direction (but are not restricted to be positive). (4) conditional treatment effect controlling for cohort, number of siblings, mother’s IQ and high-risk index at birth, and accounting for attrition using Inverse Probability Weighting (IPW). The selection of covariates for IPW is age- and gender-specific. It is based on the lowest Akaike Information Criteria (AIC) among models examining all combinations of covariates that present statistically significant imbalance between attriters and non-attriters. See Supplementary Material Section C and Tables S1-S2 for details. (5) one-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate. By block permutation we mean that permutations are done within strata defined by the pre-program variables used in the randomization protocol: cohort, gender, number of siblings, mother’s IQ and high-risk index. (6) Multiple Hypothesis stepdown *p*-values associated with (5). The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines. See Tables S13-S14 for complete estimation results.

Table S11: Inference for the Abecedarian Intervention, Main Health Results, Biomedical Sweep (Males)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. <i>p</i> -value	Naïve <i>p</i> -value	Block Permutation		Block IPW Permutation		Gender Difference <i>p</i> -value
			# Control	# Treat.						Single <i>p</i> -value	Step-Down <i>p</i> -value	Single <i>p</i> -value	Step-Down <i>p</i> -value	
Blood Pressure														
Diastolic Blood Pressure (mmHg)	Yes	35	9	19	-92.000	13.474	19.220	0.017	0.042	0.075	0.075	0.024	0.024	0.248
Systolic Blood Pressure (mmHg)	Yes	35	9	19	-143.333	17.544	24.828	0.022	0.056	0.054	0.080	0.018	0.029	0.278
Pre-Hypertension (systolic bp \geq 120 & diastolic bp \geq 80)	Yes	35	9	19	-0.667	0.246	0.321	0.117	0.099	0.227	0.316	0.119	0.172	0.937
Pre-Hypertension (systolic bp \geq 120 or diastolic bp \geq 80)	Yes	35	9	19	-0.778	0.094	0.096	0.311	0.316	0.351	0.351	0.235	0.235	0.471
Hypertension (systolic bp \geq 140 & diastolic bp \geq 90)	Yes	35	9	19	-0.444	0.339	0.537	0.019	0.040	0.066	0.066	0.010	0.018	0.264
Hypertension (systolic bp \geq 140 or diastolic bp \geq 90)	Yes	35	9	19	-0.556	0.345	0.404	0.033	0.045	0.061	0.087	0.038	0.038	0.074
Lab Tests														
High-Density Lipoprotein (HDL) Cholesterol (mg/dL)	No	35	12	19	42.000	11.211	11.720	0.012	0.007	0.012	0.021	0.066	0.110	0.385
Dyslipidemia (HDL < 40 mg/dL)	Yes	35	12	19	-0.417	0.311	0.255	0.020	0.033	0.045	0.045	0.179	0.179	0.544
Pre-Diabetes (HbA1C \geq 5.7%)	Yes	35	12	19	-0.583	0.110	0.043	0.283	0.247	0.199	0.199	0.426	0.426	0.690
Vitamin D Deficiency (< 20 ng/mL)	Yes	35	12	19	-0.750	0.382	0.435	0.018	0.016	0.034	0.034	0.021	0.021	0.102
Obesity														
Overweight (BMI \geq 25)	Yes	35	8	18	-0.750	0.028	0.190	0.444	0.462	0.396	0.472	0.239	0.239	0.837
Obese (BMI \geq 30)	Yes	35	8	18	-0.625	0.069	0.211	0.376	0.384	0.454	0.454	0.233	0.345	0.972
Severely Obese (BMI \geq 35)	Yes	35	8	18	-0.375	0.264	0.404	0.059	0.088	0.126	0.261	0.115	0.232	0.582
Waist-Hip Ratio (WHR)	Yes	35	8	17	-0.962	0.025	0.045	0.218	0.231	0.240	0.240	0.293	0.293	0.549
Abdominal Obesity (WHR > 0.9)	Yes	35	8	17	-0.875	0.228	0.294	0.124	0.099	0.147	0.225	0.137	0.218	0.908
Multiple Risk Factors														
Obesity & Hypertension	Yes	35	8	18	-0.500	0.389	0.529	0.012	0.028	0.041	0.041	0.016	0.016	0.184
Severe Obesity & Hypertension	Yes	35	8	18	-0.375	0.375	0.502	0.001	0.000	0.001	0.004	0.005	0.012	0.019
Hypertension & Dyslipidemia	Yes	35	9	18	-0.333	0.333	0.435	0.003	0.001	0.002	0.004	0.006	0.012	0.081
Metabolic Syndrome (NCEP Definition)	Yes	35	8	17	-0.250	0.250	0.465	0.013	0.003	0.004	0.009	0.007	0.014	0.478
Framingham Risk Score (35)	Yes	35	9	18	-7.043	2.154	3.253	0.050	0.078	0.112	0.112	0.038	0.038	0.245

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment. This table examines selected outcomes for males. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** shows if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, number of siblings, mother’s IQ and high-risk index at birth, and accounting for attrition using Inverse Probability Weighting (IPW). Probabilities of IPW are estimated using the following variables: prematurity (gestational age <37 weeks), a dichotomous indicator for not having an exam for illness or injury in the past two years at age 30, Achenbach DSM Attention-Deficit/Hyperactivity (AD/H) problems scale at age 30, and Achenbach substance abuse scale at age 30. **Col.9:** the asymptotic *p*-value of 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 20,000 permutation draws. **Col.10:** single hypothesis one-sided naïve permutation *p*-value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** one-sided single hypothesis block permutation *p*-value. By block permutation we mean that permutations are done within strata defined by the variables used in the randomization protocol. **Col.12:** presents the multiple hypothesis testing (step-down) *p*-values associated with (10). **Col.13:** presents the one-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate **Col.14:** presents the the multiple hypothesis testing (step-down) *p*-values associated with the IPW inference. **Col.15:** double-sided *p*-value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S12: Inference for the Abecedarian Intervention, Main Health Results, Biomedical Sweep (Females)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. <i>p</i> -value	Naïve <i>p</i> -value	Block Permutation		Block IPW Permutation		Gender Difference <i>p</i> -value
			# Control	# Treat.						Single <i>p</i> -value	Step-Down <i>p</i> -value	Single <i>p</i> -value	Step-Down <i>p</i> -value	
Blood Pressure														
Diastolic Blood Pressure (mmHg)	Yes	35	22	18	-89.227	3.894	1.204	0.233	0.240	0.330	0.330	0.446	0.446	0.248
Systolic Blood Pressure (mmHg)	Yes	35	22	18	-135.636	5.970	2.185	0.180	0.184	0.214	0.280	0.300	0.380	0.278
Pre-Hypertension (systolic bp \geq 120 & diastolic bp \geq 80)	Yes	35	22	18	-0.727	0.227	0.101	0.072	0.073	0.102	0.102	0.222	0.222	0.937
Pre-Hypertension (systolic bp \geq 120 or diastolic bp \geq 80)	Yes	35	22	18	-0.909	0.242	0.244	0.028	0.040	0.029	0.047	0.042	0.069	0.471
Hypertension (systolic bp \geq 140 & diastolic bp \geq 90)	Yes	35	22	18	-0.318	0.096	-0.003	0.255	0.258	0.260	0.367	0.375	0.499	0.264
Hypertension (systolic bp \geq 140 or diastolic bp \geq 90)	Yes	35	22	18	-0.409	-0.091	-0.181	0.712	0.707	0.628	0.628	0.721	0.721	0.074
Lab Tests														
High-Density Lipoprotein (HDL) Cholesterol (mg/dL)	No	35	22	18	55.318	5.126	6.002	0.132	0.138	0.116	0.170	0.143	0.143	0.385
Dyslipidemia (HDL < 50 mg/dL)	Yes	35	22	18	-0.455	0.177	0.201	0.130	0.132	0.121	0.121	0.099	0.147	0.544
Pre-Diabetes (HbA1C \geq 5.7%)	Yes	35	22	17	-0.364	0.011	0.070	0.473	0.476	0.568	0.568	0.580	0.580	0.690
Vitamin D Deficiency (< 20 ng/mL)	Yes	35	22	18	-0.727	0.005	0.048	0.486	0.483	0.458	0.458	0.303	0.303	0.102
Obesity														
Overweight (BMI \geq 25)	Yes	35	22	18	-0.955	0.066	0.054	0.222	0.178	0.388	0.598	0.482	0.690	0.837
Obese (BMI \geq 30)	Yes	35	22	18	-0.727	0.061	-0.112	0.343	0.347	0.715	0.715	0.790	0.790	0.972
Severely Obese (BMI \geq 35)	Yes	35	22	18	-0.364	0.141	0.143	0.171	0.166	0.361	0.653	0.354	0.653	0.582
Waist-Hip Ratio (WHR)	Yes	35	21	16	-0.933	0.057	0.053	0.059	0.060	0.081	0.126	0.063	0.101	0.549
Abdominal Obesity (WHR > 0.85)	Yes	35	21	16	-0.762	0.199	0.198	0.103	0.119	0.113	0.113	0.080	0.080	0.908
Multiple Risk Factors														
Obesity & Hypertension	Yes	35	22	18	-0.364	0.086	-0.028	0.287	0.269	0.311	0.527	0.501	0.641	0.184
Severe Obesity & Hypertension	Yes	35	22	18	-0.136	-0.030	-0.066	0.602	0.605	0.654	0.654	0.696	0.696	0.019
Hypertension & Dyslipidemia	Yes	35	22	18	-0.182	0.015	-0.043	0.452	0.455	0.441	0.608	0.486	0.725	0.081
Metabolic Syndrome (NCEP Definition)	Yes	35	21	16	-0.190	0.128	0.057	0.134	0.131	0.148	0.329	0.184	0.393	0.478
Framingham Risk Score (35)	Yes	35	22	18	-1.482	0.339	0.331	0.116	0.112	0.049	0.049	0.070	0.070	0.245

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment. This table examines selected outcomes for females. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** shows if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, number of siblings, mother’s IQ and high-risk index at birth, and accounting for attrition using Inverse Probability Weighting (IPW). Probabilities of IPW are estimated using the following variables: prematurity (gestational age <37 weeks), mother WAIS digit symbol score at recruitment, Achenbach rule breaking problem scale at age 30, and Achenbach substance abuse scale at age 30. **Col.9:** the asymptotic *p*-value of 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 20,000 permutation draws. **Col.10:** single hypothesis one-sided naïve permutation *p*-value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** one-sided single hypothesis block permutation *p*-value. By block permutation we mean that permutations are done within strata defined by the variables used in the randomization protocol. **Col.12:** presents the multiple hypothesis testing (step-down) *p*-values associated with (10). **Col.13:** presents the one-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate **Col.14:** presents the the multiple hypothesis testing (step-down) *p*-values associated with the IPW inference. **Col.15:** double-sided *p*-value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S13: Inference for the Abecedarian Intervention, Behavioral Risk Factors and Physical Development (Males)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. <i>p</i> -value	Naïve <i>p</i> -value	Block Permutation		Block IPW Permutation		Gender Difference <i>p</i> -value
			# Control	# Treat.						Single <i>p</i> -value	Step-Down <i>p</i> -value	Single <i>p</i> -value	Step-Down <i>p</i> -value	
Health Care														
Health Insurance Coverage at age 30	No	30	21	27	0.476	0.228	0.226	0.057	0.058	0.077	0.077	0.039	0.039	0.077
Buys Health Insurance at age 30	No	30	21	27	0.333	0.296	0.248	0.021	0.018	0.032	0.044	0.035	0.080	0.200
Hospital or Doctor Office Care When Sick at age 30	No	30	21	27	0.524	0.291	0.265	0.015	0.018	0.027	0.060	0.037	0.068	0.011
Behavioral Risk Factors: Smoking														
Never a Regular Smoker by age 30	No	21-30	23	28	0.391	-0.034	-0.042	0.597	0.594	0.622	0.622	0.571	0.571	0.161
Age of Onset of Regular Smoking	No	21-30	14	18	16.893	2.829	2.660	0.017	0.019	0.031	0.087	0.051	0.143	0.042
Cigarettes Smoked per Day Past 30 Days at ages 21&30	Yes	21-30	12	19	-11.167	2.482	3.820	0.214	0.229	0.278	0.453	0.199	0.344	0.534
Behavioral Risk Factors: Alcohol Use														
Early Onset Drinker (before age 17)	Yes	21	23	26	-0.609	0.070	0.074	0.314	0.312	0.311	0.742	0.336	0.762	0.265
Drank Driving	Yes	21	23	26	-0.391	-0.147	-0.090	0.844	0.846	0.804	0.957	0.663	0.877	0.188
Age of Onset of Alcohol Use	No	21-30	22	28	17.045	-1.135	-1.306	0.837	0.831	0.858	0.858	0.840	0.840	0.140
Days At Least 1 Drink Past 30 Days at ages 21&30	Yes	21-30	23	28	-9.348	-2.634	-1.880	0.778	0.780	0.664	0.920	0.498	0.860	0.233
Days 5+ Drinks Past 30 Days at ages 21&30	Yes	21-30	23	28	-2.457	-0.972	-0.946	0.743	0.726	0.653	0.937	0.585	0.900	0.301
Behavioral Risk Factors: Marijuana Use														
Never Used Marijuana by age 30	No	21-30	23	28	0.130	-0.023	-0.019	0.599	0.616	0.515	0.515	0.476	0.476	0.301
Age of Onset of Marijuana Use	No	21-30	20	25	15.475	2.025	1.763	0.019	0.022	0.018	0.047	0.052	0.136	0.323
Times Used Marijuana Past 30 Days at ages 21&30	Yes	21-30	23	28	-49.565	1.030	2.699	0.467	0.465	0.529	0.678	0.357	0.527	0.879
Behavioral Risk Factors: Physical Activity and Diet														
Exercise 4 or more Days per Week at age 21	No	21	23	26	0.391	-0.084	-0.094	0.726	0.730	0.806	0.806	0.858	0.858	0.062
No. of Fruit Servings per Day at age 21	No	21	23	26	0.826	0.020	-0.015	0.470	0.462	0.434	0.618	0.524	0.745	0.139
Physical Development														
At Risk Overweight (CDC) at 3 months	Yes	0	22	27	-0.227	0.190	0.206	0.022	0.022	0.042	0.173	0.026	0.121	0.212
At Risk Overweight (CDC) at 6 months	Yes	1	20	25	-0.250	0.170	0.205	0.061	0.068	0.148	0.335	0.074	0.182	0.620
At Risk Overweight (CDC) at 9 months	Yes	1	17	16	-0.412	0.412	0.446	0.001	0.001	0.005	0.025	0.004	0.023	0.285
At Risk Overweight (CDC) at 12 months	Yes	1	21	27	-0.429	0.429	0.408	0.000	0.000	0.000	0.002	0.001	0.009	0.373
At Risk Overweight (CDC) at 18 months	Yes	2	18	26	-0.389	0.389	0.385	0.000	0.000	0.000	0.003	0.000	0.004	0.149
Overweight (CDC) at 24 months	Yes	2	15	27	-0.333	0.333	0.343	0.000	0.000	0.000	0.001	0.001	0.011	0.619
Overweight (CDC) at 36 months	Yes	3	19	25	-0.158	0.078	0.094	0.215	0.212	0.168	0.255	0.194	0.194	0.804
Overweight (CDC) at 48 months	Yes	4	20	24	-0.300	0.133	0.133	0.151	0.155	0.180	0.180	0.150	0.235	0.060
Overweight (CDC) at 60 months	Yes	5	20	24	-0.300	0.175	0.187	0.079	0.084	0.066	0.196	0.058	0.179	0.311
Overweight (CDC) at 96 months	Yes	8	19	25	-0.421	0.301	0.286	0.010	0.015	0.028	0.130	0.030	0.117	0.011
Weight-for-Length Change 0-24 months (CDC)	Yes	2	15	24	-0.858	0.963	1.176	0.057	0.056	0.058	0.058	0.058	0.058	0.219
Weight-for-Length Change 0-24 months (WHO)	Yes	2	15	24	-1.265	1.100	1.397	0.046	0.050	0.047	0.054	0.049	0.057	0.177

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment. This table examines selected outcomes for males. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** shows if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, number of siblings, mother's IQ and high-risk index at birth, and accounting for attrition using Inverse Probability Weighting (IPW). The selection of covariates for IPW is age- and gender-specific. It is based on the lowest Akaike Information Criteria (AIC) among models examining all combinations of covariates that present statistically significant imbalance between attriters and non-attriters. See Supplementary Material, Section C, and Table S1 for details. **Col.9:** the asymptotic *p*-value of 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 20,000 permutation draws. **Col.10:** single hypothesis one-sided naïve permutation *p*-value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** one-sided single hypothesis block permutation *p*-value. By block permutation we mean that permutations are done within strata defined by the variables used in the randomization protocol. **Col.12:** presents the multiple hypothesis testing (step-down) *p*-values associated with (10). **Col.13:** presents the one-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate **Col.14:** presents the the multiple hypothesis testing (step-down) *p*-values associated with the IPW inference. **Col.15:** double-sided *p*-value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S14: Inference for the Abecedarian Intervention, Behavioral Risk Factors and Physical Development (Females)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. <i>p</i> -value	Naïve <i>p</i> -value	Block Permutation		Block IPW Permutation		Gender Difference <i>p</i> -value
			# Control	# Treat.						Single <i>p</i> -value	Step-Down <i>p</i> -value	Single <i>p</i> -value	Step-Down <i>p</i> -value	
Health Care														
Health Insurance Coverage at age 30	No	30	28	25	0.857	-0.097	-0.159	0.812	0.812	0.925	0.925	0.943	0.943	0.077
Buys Health Insurance at age 30	No	30	28	25	0.357	0.043	-0.027	0.377	0.379	0.497	0.806	0.511	0.810	0.200
Hospital or Doctor Office Care When Sick at age 30	No	30	28	25	0.929	-0.129	-0.131	0.914	0.912	0.903	0.978	0.875	0.964	0.011
Behavioral Risk Factors: Smoking														
Never a Regular Smoker by age 30	No	21-30	28	25	0.357	0.243	0.221	0.039	0.045	0.083	0.238	0.071	0.195	0.161
Age of Onset of Regular Smoking	No	21-30	18	10	17.861	-0.811	-1.544	0.755	0.772	0.855	0.855	0.848	0.848	0.042
Cigarettes Smoked per Day Past 30 Days at ages 21&30	Yes	21-30	17	10	-10.294	-0.156	-0.939	0.521	0.522	0.506	0.753	0.567	0.807	0.534
Behavioral Risk Factors: Alcohol Use														
Early Onset Drinker (before age 17)	Yes	21	28	25	-0.571	0.291	0.344	0.016	0.016	0.023	0.102	0.011	0.046	0.265
Drank Driving	Yes	21	27	25	-0.222	0.102	0.100	0.170	0.169	0.179	0.389	0.172	0.358	0.188
Age of Onset of Alcohol Use	No	21-30	27	22	16.685	1.042	1.178	0.129	0.120	0.135	0.370	0.099	0.292	0.140
Days At Least 1 Drink Past 30 Days at ages 21&30	Yes	21-30	28	25	-8.732	2.772	1.105	0.164	0.159	0.269	0.366	0.211	0.292	0.233
Days 5+ Drinks Past 30 Days at ages 21&30	Yes	21-30	28	25	-2.321	1.061	-0.095	0.191	0.195	0.315	0.315	0.366	0.366	0.301
Behavioral Risk Factors: Marijuana Use														
Never Used Marijuana by age 30	No	21-30	28	25	0.250	0.150	0.089	0.125	0.133	0.123	0.339	0.169	0.402	0.301
Age of Onset of Marijuana Use	No	21-30	21	15	17.190	0.676	0.828	0.236	0.240	0.286	0.493	0.274	0.471	0.323
Times Used Marijuana Past 30 Days at ages 21&30	Yes	21-30	28	25	-19.982	-1.598	-2.089	0.560	0.556	0.541	0.541	0.417	0.417	0.879
Behavioral Risk Factors: Physical Activity and Diet														
Exercise 4 or more Days per Week at age 21	No	21	28	25	0.071	0.249	0.287	0.010	0.013	0.009	0.009	0.004	0.008	0.062
No. of Fruit Servings per Day at age 21	No	21	28	25	0.286	0.514	0.518	0.005	0.007	0.002	0.005	0.005	0.005	0.139
Physical Development														
At Risk Overweight (CDC) at 3 months	Yes	0	26	21	-0.192	0.002	-0.036	0.494	0.482	0.406	0.712	0.418	0.757	0.212
At Risk Overweight (CDC) at 6 months	Yes	1	26	24	-0.423	0.256	0.212	0.024	0.025	0.031	0.209	0.040	0.237	0.620
At Risk Overweight (CDC) at 9 months	Yes	1	25	14	-0.360	0.217	0.181	0.076	0.058	0.142	0.457	0.169	0.548	0.285
At Risk Overweight (CDC) at 12 months	Yes	1	23	24	-0.478	0.270	0.141	0.025	0.028	0.040	0.203	0.055	0.276	0.373
At Risk Overweight (CDC) at 18 months	Yes	2	25	22	-0.440	0.122	0.118	0.201	0.200	0.250	0.571	0.311	0.669	0.149
Overweight (CDC) at 24 months	Yes	2	17	23	-0.412	0.238	0.195	0.049	0.049	0.091	0.357	0.143	0.517	0.619
Overweight (CDC) at 36 months	Yes	3	23	21	-0.261	0.118	-0.020	0.171	0.162	0.244	0.614	0.202	0.556	0.804
Overweight (CDC) at 48 months	Yes	4	26	22	-0.192	-0.217	-0.247	0.949	0.945	0.965	0.965	0.944	0.944	0.060
Overweight (CDC) at 60 months	Yes	5	23	22	-0.261	-0.012	-0.050	0.535	0.538	0.503	0.708	0.554	0.781	0.311
Overweight (CDC) at 96 months	Yes	8	23	20	-0.174	-0.176	-0.230	0.904	0.894	0.900	0.953	0.943	0.985	0.011
Weight-for-Length Change 0-24 months (CDC)	Yes	2	15	21	-0.857	-0.062	-0.052	0.544	0.537	0.580	0.611	0.658	0.688	0.219
Weight-for-Length Change 0-24 months (WHO)	Yes	2	15	21	-1.129	-0.085	-0.006	0.559	0.556	0.611	0.611	0.660	0.660	0.177

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment. This table examines selected outcomes for females. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** shows if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, number of siblings, mother’s IQ and high-risk index at birth, and accounting for attrition using Inverse Probability Weighting (IPW). The selection of covariates for IPW is age- and gender-specific. It is based on the lowest Akaike Information Criteria (AIC) among models examining all combinations of covariates that present statistically significant imbalance between attriters and non-attriters. See Supplementary Material, Section C, and Table S2 for details. **Col.9:** the asymptotic *p*-value of 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 20,000 permutation draws. **Col.10:** single hypothesis one-sided naïve permutation *p*-value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** one-sided single hypothesis block permutation *p*-value. By block permutation we mean that permutations are done within strata defined by the variables used in the randomization protocol. **Col.12:** presents the multiple hypothesis testing (step-down) *p*-values associated with (10). **Col.13:** presents the one-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate **Col.14:** presents the the multiple hypothesis testing (step-down) *p*-values associated with the IPW inference. **Col.15:** double-sided *p*-value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

F Inference about the Treatment Effect for the Second Stage School-Age Component

This supplement investigates whether the second stage of the program had any favorable effects on health outcomes analyzed in the main paper. As described in Supplement A, the ABC experiment had two stages of intervention for children age 0 to 5 and a second stage intervention for children from age 5 to 8. In the second stage, treated children were provided need-based and individualized tutoring and bi-weekly home visits. The second stage treatment assignment was determined by a new randomization implemented at the end of first stage intervention. In the main paper, we define the treatment group to include all children who received first stage treatment irregardless of whether they received second stage treatment. Defining the treatment group in this fashion may bias downward the estimated treatment effect if controls received benefits from the second stage randomization. It biases the treatment effect upward if second stage treatment has a positive effect for those who received treatment in the first stage. We investigate these potential sources of bias in this paper.

In analyzing the second stage treatment effect, we redefine the treatment and the control groups. Different from the analysis in the main paper, the “treatment group” in the second stage consists of people who participated in both of first and second stages plus those who were randomized out from the first stage but were randomized into the second stage. The control group consists of those denied access to both stages by virtue of randomization plus those who had access to the first stage but were randomized out in the second stage. In this appendix, persons randomly assigned into the second stage of the ABC experiment are compared with those who did not participate in the second stage irregardless of their first stage treatment assignments. As evident in Tables S15a—S16b, our analysis shows very few statistically significant second-stage treatment

effects (e.g. Pre-Hypertension at age 35 for females; Hypertension and Severe Obesity at age 35 for males). Even fewer survive correction for multiple hypothesis testing. This implies that in the ABC experiment, the second stage intervention was not effective in terms of its effects on adult health.

Tables S17—S18 investigate this issue more deeply. In these tables we split the entire sample into four group where each group is defined by participant’s treatment assignment at the first and the second stage jointly. Those four groups are TT(treatment-treatment), TC(treatment-control), CT(control-treatment), and CC(control-control) where the first letter indicates program assignment at the first stage and the second corresponds to assignment at the second stage. We test three hypotheses:

H1 : $\mu_{TT} = \mu_{TC}$ (i.e. No difference in means between TT and TC),

H2 : $\mu_{CT} = \mu_{CC}$ (i.e. No difference in means between CT and CC), and

H3 : $\mu_{TT} = \mu_{TC}$ and $\mu_{CT} = \mu_{CC}$ (i.e. Joint Hypothesis).

Hypothesis 1 (H1) tests whether the second stage is statistically significant from first stage for those who were treated in the first stage. Hypothesis 2 (H2) tests whether the second stage is statistically significantly different from the first stage for those who did not receive treatment in the first stage. Hypothesis 3 (H3) tests these two hypotheses jointly. As summarized in Tables S17—S18, these alternative tests confirm the findings from the previous tables, that is, the second stage program was not effective in terms of its effects on adult health. For both males and females there is no consistent pattern favoring those who received second period treatment. On some outcomes, those receiving second period treatment are worse although the differences are not statistically significantly different from zero.

Table S15(a): Second-Stage (School-Age) Component of the Abecedarian Intervention, Main Health Results, Biomedical Sweep (Males)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. <i>p</i> -value	Naïve <i>p</i> -value	Block Permutation		Block IPW Permutation		Gender Difference
			# Control	# Treat.						Single <i>p</i> -value	Step-Down <i>p</i> -value	Single <i>p</i> -value	Step-Down <i>p</i> -value	
Blood Pressure														
Diastolic Blood Pressure (mmHg)	Yes	35	13	13	-129.462	-4.308	-1.929	0.682	0.659	0.659	0.659	0.586	0.586	0.544
Systolic Blood Pressure (mmHg)	Yes	35	13	13	-84.231	1.077	2.264	0.435	0.446	0.417	0.491	0.403	0.463	0.907
Pre-Hypertension (systolic bp \geq 120 & diastolic bp \geq 80)	Yes	35	13	13	-0.462	-0.077	-0.116	0.646	0.604	0.557	0.689	0.714	0.714	0.490
Pre-Hypertension (systolic bp \geq 120 or diastolic bp \geq 80)	Yes	35	13	13	-0.692	-0.077	0.003	0.664	0.659	0.659	0.659	0.603	0.746	0.210
Hypertension (systolic bp \geq 140 & diastolic bp \geq 90)	Yes	35	13	13	-0.231	0.000	0.063	0.500	0.501	0.481	0.481	0.359	0.359	0.593
Hypertension (systolic bp \geq 140 or diastolic bp \geq 90)	Yes	35	13	13	-0.385	0.154	0.147	0.206	0.205	0.203	0.267	0.242	0.291	0.875
Lab Tests														
HDL Cholesterol (mg/dL)	No	35	14	14	49.214	0.786	-2.158	0.444	0.440	0.484	0.625	0.677	0.677	0.366
Dyslipidemia (HDL < 40 mg/dL)	Yes	35	14	14	-0.143	-0.071	-0.068	0.682	0.691	0.677	0.677	0.671	0.819	0.733
Pre-Diabetes (HbA1C \geq 5.7%)	Yes	35	14	14	-0.500	0.000	-0.062	0.500	0.504	0.531	0.531	0.747	0.747	0.358
Vit. D Deficiency (< 20 ng/mL)	Yes	35	14	14	-0.500	0.000	-0.113	0.500	0.500	0.511	0.511	0.802	0.802	0.659
Obesity														
Overweight (BMI \geq 25)	Yes	35	13	11	-0.692	-0.126	-0.106	0.752	0.751	0.766	0.881	0.648	0.776	0.795
Obese (BMI \geq 30)	Yes	35	13	11	-0.462	-0.266	-0.280	0.901	0.897	0.896	0.896	0.891	0.891	0.709
Severely Obese (BMI \geq 35)	Yes	35	13	11	-0.231	0.049	-0.034	0.390	0.397	0.376	0.641	0.524	0.796	0.909
Waist-Hip Ratio (WHR)	Yes	35	12	11	-0.942	-0.015	-0.028	0.681	0.679	0.694	0.694	0.685	0.685	0.273
Abdominal Obesity (WHR > 0.90)	Yes	35	12	11	-0.750	0.023	-0.141	0.453	0.451	0.458	0.601	0.583	0.697	0.300

Notes: This table examines the causal effects of the second-stage randomization (school-age component) of the Abecedarian intervention. We present descriptive statistics, treatment effect estimates, and inference for selected outcomes for males. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, high-risk index at birth, and gender, and accounting for attrition using Inverse Probability Weighting (IPW). For IPW covariates we use High-Risk Index at birth and Mother WAIS IQ; **Col.9:** asymptotic *p*-value of 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 20,000 permutation draws. **Col.10:** Single hypothesis one-sided naïve permutation *p*-value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** One-sided single hypothesis block permutation *p*-value. By block permutation we mean that permutations are done within strata defined by cohort and gender. **Col.12:** Multiple hypothesis testing (step-down) *p*-values associated with column (10). **Col.13:** One-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate. **Col.14:** Multiple hypothesis testing (step-down) *p*-values associated with the IPW inference. **Col.15:** Double-sided *p*-value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S15(b): Second-Stage (School-Age) Component of the Abecedarian Intervention, Main Health Results, Biomedical Sweep (Females)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. p -value	Naïve p -value	Block Permutation		Block IPW Permutation		Gender Difference
			# Control	# Treat.						Single p -value	Step-Down p -value	Single p -value	Step-Down p -value	
Blood Pressure														
Diastolic Blood Pressure (mmHg)	Yes	35	16	20	-136.438	2.287	1.836	0.370	0.371	0.373	0.453	0.394	0.479	0.544
Systolic Blood Pressure (mmHg)	Yes	35	16	20	-88.875	0.075	-0.615	0.495	0.494	0.488	0.488	0.535	0.535	0.907
Pre-Hypertension (systolic bp \geq 120 & diastolic bp \geq 80)	Yes	35	16	20	-0.750	0.100	0.080	0.265	0.252	0.263	0.263	0.301	0.301	0.490
Pre-Hypertension (systolic bp \geq 120 or diastolic bp \geq 80)	Yes	35	16	20	-0.938	0.188	0.209	0.070	0.069	0.081	0.120	0.086	0.134	0.210
Hypertension (systolic bp \geq 140 & diastolic bp \geq 90)	Yes	35	16	20	-0.375	0.125	0.125	0.216	0.224	0.222	0.322	0.244	0.244	0.593
Hypertension (systolic bp \geq 140 or diastolic bp \geq 90)	Yes	35	16	20	-0.563	0.112	0.142	0.258	0.263	0.243	0.243	0.193	0.269	0.875
Lab Tests														
HDL Cholesterol (mg/dL)	No	35	16	20	62.250	-6.350	-5.202	0.900	0.902	0.914	0.914	0.864	0.864	0.366
Dyslipidemia (HDL < 50 mg/dL)	Yes	35	16	20	-0.250	-0.150	-0.115	0.822	0.846	0.826	0.918	0.759	0.863	0.733
Pre-Diabetes (HbA1C \geq 5.7%)	Yes	35	16	19	-0.188	-0.234	-0.241	0.928	0.925	0.931	0.931	0.924	0.924	0.358
Vit. D Deficiency (< 20 ng/mL)	Yes	35	16	20	-0.813	0.113	0.119	0.226	0.210	0.216	0.216	0.224	0.224	0.659
Obesity														
Overweight (BMI \geq 25)	Yes	35	16	20	-0.875	-0.075	-0.111	0.784	0.816	0.788	0.918	0.882	0.882	0.795
Obese (BMI \geq 30)	Yes	35	16	20	-0.625	-0.175	-0.177	0.873	0.877	0.866	0.866	0.880	0.957	0.709
Severely Obese (BMI \geq 35)	Yes	35	16	20	-0.375	0.075	0.103	0.323	0.351	0.356	0.614	0.321	0.588	0.909
Waist-Hip Ratio (WHR)	Yes	35	15	18	-0.869	-0.075	-0.075	0.973	0.976	0.970	0.970	0.966	0.966	0.273
Abdominal Obesity (WHR > 0.85)	Yes	35	15	18	-0.533	-0.244	-0.240	0.928	0.923	0.919	0.973	0.911	0.965	0.300

Notes: This table examines the causal effects of the second-stage randomization (school-age component) of the Abecedarian intervention. We present descriptive statistics, treatment effect estimates, and inference for selected outcomes for females. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, high-risk index at birth, and gender, and accounting for attrition using Inverse Probability Weighting (IPW). For IPW covariates we use High-Risk Index at birth and Mother WAIS IQ; **Col.9:** asymptotic p -value of 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 20,000 permutation draws. **Col.10:** Single hypothesis one-sided naïve permutation p -value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** One-sided single hypothesis block permutation p -value. By block permutation we mean that permutations are done within strata defined by cohort and gender. **Col.12:** Multiple hypothesis testing (step-down) p -values associated with column (10). **Col.13:** One-sided single hypothesis block permutation p -value associated with the IPW treatment effect estimate. **Col.14:** Multiple hypothesis testing (step-down) p -values associated with the IPW inference. **Col.15:** Double-sided p -value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S16(a): Second-Stage (School-Age) Component of the Abecedarian Intervention: (Males)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. <i>p</i> -value	Naïve <i>p</i> -value	Block Permutation		Block IPW Permutation		Gender Difference
			# Control	# Treat.						Single <i>p</i> -value	Step-Down <i>p</i> -value	Single <i>p</i> -value	Step-Down <i>p</i> -value	
Risk Factors														
Metabolic Syndrome at age 35 (NCEP)	Yes	35	12	11	-0.167	0.167	0.143	0.081	0.106	0.081	0.171	0.121	0.178	0.683
Hypertension and Obesity (age 35)	Yes	35	13	11	-0.308	0.217	0.209	0.101	0.094	0.089	0.149	0.099	0.215	0.431
Hypertension and Severe Obesity (age 35)	Yes	35	13	11	-0.231	0.231	0.188	0.044	0.053	0.051	0.132	0.074	0.190	0.682
Hypertension and Dyslipidemia (age 35)	Yes	35	13	12	-0.154	0.071	0.065	0.302	0.312	0.288	0.288	0.333	0.333	0.675
Framingham														
Framingham Probability	Yes	35	13	12	-6.009	0.952	548	0.242	0.243	0.269	0.269	0.385	0.385	0.516
Health Care Coverage at age 30														
Health Insurance Coverage at age 30	No	30	24	20	0.542	0.158	0.070	0.147	0.155	0.116	0.208	0.304	0.475	0.153
Buys Health Insurance (age 30)	No	30	24	20	0.417	0.183	0.131	0.117	0.120	0.110	0.226	0.185	0.357	0.218
Hospital or doctor office care when sick	No	30	24	20	0.750	-0.100	-0.064	0.760	0.757	0.745	0.745	0.697	0.697	0.620
Smoking														
Never a regular smoker by age 35	No	30	26	21	0.385	-0.004	0.051	0.510	0.514	0.482	0.482	0.368	0.368	0.866
Age Subject Began Smoking Regularly (Never=.)	No	30	16	12	18.063	0.854	0.852	0.293	0.308	0.263	0.394	0.269	0.471	0.872
Sum of Cigarettes per Day in Past 30 Days at 21 and 30	Yes	30	13	15	-12.115	5.049	3.207	0.051	0.066	0.069	0.179	0.159	0.388	0.741
Alcohol Use														
Did you begin drinking before 17	Yes	21	24	21	-0.583	0.012	0.118	0.469	0.461	0.437	0.782	0.219	0.525	0.927
Do you drink and drive?	Yes	21	24	21	-0.458	-0.018	-0.049	0.546	0.542	0.547	0.822	0.592	0.838	0.931
Age Subject Began Drinking (Never=.)	No	30	25	21	16.100	0.710	1.325	0.282	0.279	0.275	0.625	0.117	0.353	0.661
Sum of Days Subject Had at Least One Drink in Past 30 Days at 21 – 30	Yes	30	26	21	-9.423	-3.291	-1.499	0.812	0.818	0.812	0.812	0.712	0.810	0.808
Sum of Days Subject Drank 5+ Drinks in Past 30 Days at 21 – 30	Yes	30	26	21	-2.885	-0.520	-0.868	0.626	0.622	0.625	0.740	0.771	0.771	0.761
Marijuana														
Never Used Marijuana by age 30	No	30	26	21	0.423	0.148	0.206	0.161	0.180	0.176	0.472	0.103	0.253	0.807
Age of Onset of Marijuana Use	No	30	23	19	16.283	0.349	0.047	0.371	0.374	0.342	0.509	0.440	0.623	0.716
Times Subject Used Marijuana Past 30 Days (21 – 30 years old)	Yes	30	26	21	-45.846	-6.916	-11.110	0.701	0.687	0.671	0.671	0.804	0.804	0.963
Physical Activity and Diet														
Exercise 4 or more Days per Week (age 21)	No	21	24	21	0.292	0.185	0.151	0.105	0.100	0.109	0.196	0.150	0.262	0.100
No. of Fruit Servings per Day (age 21)	No	21	24	21	0.833	-0.024	0.016	0.533	0.537	0.500	0.500	0.473	0.473	0.815

Notes: This table examines the causal effects of the second-stage randomization (school-age component) of the Abecedarian intervention. We present descriptive statistics, treatment effect estimates, and inference for selected outcomes for males. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, high-risk index at birth, and gender, and accounting for attrition using Inverse Probability Weighting (IPW). For IPW covariates we use High-Risk Index at birth and Mother WAIS IQ; **Col.9:** asymptotic *p*-value of 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 20,000 permutation draws. **Col.10:** Single hypothesis one-sided naïve permutation *p*-value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** One-sided single hypothesis block permutation *p*-value. By block permutation we mean that permutations are done within strata defined by cohort and gender. **Col.12:** Multiple hypothesis testing (step-down) *p*-values associated with column (10). **Col.13:** One-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate. **Col.14:** Multiple hypothesis testing (step-down) *p*-values associated with the IPW inference. **Col.15:** Double-sided *p*-value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S16(b): Second-Stage (School-Age) Component of the Abecedarian Intervention: (Females)

Variable	Reversed	Age	Sample Sizes		Control Mean	Difference in Means	IPW Treatment Effect	Asy. <i>p</i> -value	Naïve <i>p</i> -value	Block Permutation		Block IPW Permutation		Gender Difference
			# Control	# Treat.						Single <i>p</i> -value	Step-Down <i>p</i> -value	Single <i>p</i> -value	Step-Down <i>p</i> -value	
Risk Factors														
Metabolic Syndrome at age 35 (NCEP)	Yes	35	15	18	-0.200	0.089	0.131	0.246	0.245	0.264	0.428	0.201	0.347	0.683
Hypertension and Obesity (age 35)	Yes	35	16	20	-0.375	0.025	0.039	0.440	0.438	0.432	0.602	0.400	0.563	0.431
Hypertension and Severe Obesity (age 35)	Yes	35	16	20	-0.250	0.150	0.172	0.120	0.131	0.139	0.319	0.112	0.261	0.682
Hypertension and Dyslipidemia (age 35)	Yes	35	16	20	-0.188	-0.013	0.032	0.536	0.537	0.524	0.524	0.428	0.428	0.675
Framingham														
Framingham Probability	Yes	35	16	20	-1.243	-0.123	-0.074	0.654	0.637	0.595	0.595	0.516	0.516	0.516
Health Care Coverage at age 30														
Health Insurance Coverage at age 30	No	30	20	25	0.850	-0.130	-0.129	0.846	0.850	0.837	0.837	0.809	0.809	0.153
Buys Health Insurance (age 30)	No	30	20	25	0.400	-0.080	-0.103	0.706	0.715	0.718	0.863	0.745	0.902	0.218
Hospital or doctor office care when sick	No	30	20	25	0.850	-0.010	-0.000	0.536	0.536	0.431	0.702	0.422	0.717	0.620
Smoking														
Never a regular smoker by age 35	No	30	20	25	0.450	0.030	-0.007	0.423	0.387	0.437	0.437	0.519	0.519	0.866
Age Subject Began Smoking Regularly (Never=.)	No	30	10	13	17.600	0.515	0.456	0.342	0.340	0.333	0.518	0.402	0.658	0.872
Sum of Cigarettes per Day in Past 30 Days at 21 and 30	Yes	30	12	10	-11.625	3.575	2.756	0.124	0.121	0.133	0.327	0.142	0.376	0.741
Alcohol Use														
Did you begin drinking before 17	Yes	21	20	25	-0.350	0.030	-0.007	0.418	0.409	0.390	0.831	0.507	0.845	0.927
Do you drink and drive?	Yes	21	19	25	-0.158	-0.002	0.001	0.507	0.522	0.437	0.835	0.452	0.861	0.931
Age Subject Began Drinking (Never=.)	No	30	20	21	17.825	-0.015	-0.183	0.507	0.514	0.479	0.767	0.572	0.862	0.661
Sum of Days Subject Had at Least One Drink in Past 30 Days at 21 – 30	Yes	30	20	25	-6.350	-2.050	-2.916	0.731	0.728	0.761	0.761	0.814	0.814	0.808
Sum of Days Subject Drank 5+ Drinks in Past 30 Days at 21 – 30	Yes	30	20	25	-2.150	0.170	-0.238	0.453	0.454	0.501	0.626	0.610	0.730	0.761
Marijuana														
Never Used Marijuana by age 30	No	30	20	25	0.700	0.100	0.081	0.225	0.240	0.210	0.522	0.295	0.607	0.807
Age of Onset of Marijuana Use	No	30	14	15	17.786	-0.219	-0.389	0.580	0.586	0.547	0.783	0.621	0.845	0.716
Times Subject Used Marijuana Past 30 Days (21 –30 years old)	Yes	30	20	25	-14.300	-7.820	-8.728	0.756	0.760	0.772	0.772	0.769	0.769	0.963
Physical Activity and Diet														
Exercise 4 or more Days per Week (age 21)	No	21	20	25	0.250	-0.130	-0.144	0.867	0.853	0.896	0.896	0.899	0.899	0.100
No. of Fruit Servings per Day (age 21)	No	21	20	25	0.500	0.060	0.049	0.397	0.412	0.411	0.635	0.432	0.682	0.815

Notes: This table examines the causal effects of the second-stage randomization (school-age component) of the Abecedarian intervention. We present descriptive statistics, treatment effect estimates, and inference for selected outcomes for females. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4:** control sample size; **Col.5:** treatment sample size; **Col.6:** control arithmetic mean; **Col.7:** unconditional difference in means across treatment and control groups; **Col.8:** conditional treatment effect controlling for cohort, high-risk index at birth, and gender, and accounting for attrition using Inverse Probability Weighting (IPW). For IPW covariates we use High-Risk Index at birth and Mother WAIS IQ; **Col.9:** asymptotic *p*-value of 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 20,000 permutation draws. **Col.10:** Single hypothesis one-sided naïve permutation *p*-value. By naïve we mean that the test is based on an unconstrained permutation scheme. **Col.11:** One-sided single hypothesis block permutation *p*-value. By block permutation we mean that permutations are done within strata defined by cohort and gender. **Col.12:** Multiple hypothesis testing (step-down) *p*-values associated with column (10). **Col.13:** One-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate. **Col.14:** Multiple hypothesis testing (step-down) *p*-values associated with the IPW inference. **Col.15:** Double-sided *p*-value for the test of gender difference on treatment effect. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S17: Tests of H1, H2, and H3: (Males)

Variable	Reversed	Age	Sample Sizes				Means				Conditional Permutation <i>p</i> -values		
			TT	TC	CT	CC	TT	TC	CT	CC	TT=TC	CT=CC	TT=TC=CT=CC
Blood Pressure													
Systolic Blood Pressure at age 35	Yes	35	11	7	2	6	77.000	80.429	117.000	88.667	0.422	0.364	0.433
Diastolic Blood Pressure at age 35	Yes	35	11	7	2	6	125.727	123.000	178.000	137.000	0.719	0.365	0.624
Pre-Hypertension													
Pre-Hypertension at age 35 (Sys. 120 and Dias. 80)	Yes	35	11	7	2	6	0.455	0.286	1.000	0.667	0.798	0.722	0.606
Pre-Hypertension at age 35 (Sys. 120 or Dias. 80)	Yes	35	11	7	2	6	0.727	0.571	1.000	0.833	0.644	0.666	0.918
Hypertension													
Hypertension at age 35 (Sys. 140 and Dias. 90)	Yes	35	11	7	2	6	0.091	0.143	1.000	0.333	0.552	0.351	0.085
Hypertension at age 35 (Sys. 140 or Dias. 90)	Yes	35	11	7	2	6	0.091	0.286	1.000	0.500	0.320	0.536	0.158
Lab Tests (Dyslipidemia)													
HDL Cholesterol at age 35	No	35	10	8	4	6	53.400	53.250	41.500	43.833	0.851	0.949	0.756
Dyslipidemia at age 35 (HDL 40 or HDL 50)	Yes	35	10	8	4	6	0.200	0.000	0.250	0.333	0.292	0.993	0.574
Lab Tests (Pre-Diabetes)													
Pre-Diabetes at age 35 (HbA1C 5.7)	Yes	35	10	8	4	6	0.500	0.375	0.500	0.667	0.598	0.770	0.730
Nutrition at 35													
Vit. D Deficiency (age 35)	Yes	35	10	8	4	6	0.500	0.250	0.500	0.833	0.282	0.969	0.452
Obesity													
Overweight at age 35 (BMI>25)	Yes	35	10	7	1	6	0.800	0.571	1.000	0.833	0.417	0.667	0.563
Obese at age 35 (BMI >30)	Yes	35	10	7	1	6	0.700	0.286	1.000	0.667	0.258	0.300	0.457
Severely Obese at age 35 (BMI>35)	Yes	35	10	7	1	6	0.200	0.000	0.000	0.500	0.308	0.569	0.114
Waist-Hip Ratio													
Waist-Hip Ratio at age 35	Yes	35	9	7	2	5	0.956	0.921	0.963	0.971	0.475	0.025	0.545
Abdominal Obesity at age 35	Yes	35	9	7	2	5	0.667	0.714	1.000	0.800	0.801	0.683	0.582
Risk Factors													
Metabolic Syndrome at age 35 (NCEP)	Yes	35	10	7	1	6	0.000	0.143	1.000	0.500	0.434	0.417	0.231
Hypertension and Obesity (age 35)	Yes	35	10	7	1	6	0.000	0.000	0.000	0.500	0.999	0.553	0.309
Hypertension and Severe Obesity (age 35)	Yes	35	10	7	2	6	0.000	0.000	0.500	0.333	0.999	0.741	0.688
Hypertension and Dyslipidemia (age 35)	Yes	35	9	7	2	5	0.000	0.000	0.000	0.400	0.999	0.773	0.307
Framingham													
Framingham Probability	Yes	35	10	7	2	6	4.730	5.211	6.694	6.941	0.757	0.835	0.960

Notes: This table examines the effect of the second-stage randomization (school-age component) of the Abecedarian intervention. We present descriptive statistics and *p*-values from three hypothesis testing for selected outcomes. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4—Col.7:** sample sizes of each group. **Col.8—Col.11:** group means of the variable of interest. **Col.12—Col.14:** two-sided *p*-values for three hypotheses described in the text. **Col. 12:** H1 : $\mu_{TT} = \mu_{TC}$ (i.e. No program effect between TT and TC). **Col.13:** H2 : $\mu_{CT} = \mu_{CC}$ (i.e. No Program effect between CT and CC), and **Col.14:** H3 : $\mu_{TT} = \mu_{TC}$ and $\mu_{CT} = \mu_{CC}$. The testing was conducted by a conditional permutation procedure using cohort, mother's IQ, High Risk Index (HRI) at birth, and the number of siblings as conditioning variables. For each testing, 5,000 permutations were drawn to construct the distribution of test statistics.

Table S18: Tests of H1, H2 and H3: (Females)

Variable	Reversed	Age	Sample Sizes				Means				Conditional Permutation <i>p</i> -values		
			TT	TC	CT	CC	TT	TC	CT	CC	TT=TC	CT=CC	TT=TC=CT=CC
Blood Pressure													
Systolic Blood Pressure at age 35	Yes	35	10	7	10	9	83.300	88.571	94.300	89.111	0.530	0.630	0.536
Diastolic Blood Pressure at age 35	Yes	35	10	7	10	9	125.700	138.571	142.600	134.778	0.270	0.474	0.278
Pre-Hypertension													
Pre-Hypertension at age 35 (Sys. 120 and Dias. 80)	Yes	35	10	7	10	9	0.500	0.571	0.800	0.889	0.738	0.682	0.942
Pre-Hypertension at age 35 (Sys. 120 or Dias. 80)	Yes	35	10	7	10	9	0.500	0.857	1.000	1.000	0.295	0.999	0.333
Hypertension													
Hypertension at age 35 (Sys. 140 and Dias. 90)	Yes	35	10	7	10	9	0.100	0.429	0.400	0.333	0.087	0.881	0.396
Hypertension at age 35 (Sys. 140 or Dias. 90)	Yes	35	10	7	10	9	0.500	0.571	0.400	0.556	0.998	0.376	0.760
Lab Tests (Dyslipidemia)													
HDL Cholesterol at age 35	No	35	10	7	10	9	55.900	67.857	55.900	57.889	0.262	0.746	0.335
Dyslipidemia at age 35 (HDL 40 or HDL 50)	Yes	35	10	7	10	9	0.400	0.143	0.400	0.333	0.366	0.720	0.558
Lab Tests (Pre-Diabetes)													
Pre-Diabetes at age 35 (HbA1C 5.7)	Yes	35	9	7	10	9	0.444	0.143	0.400	0.222	0.483	0.922	0.344
Nutrition at 35													
Vit. D Deficiency (age 35)	Yes	35	10	7	10	9	0.600	1.000	0.800	0.667	0.087	0.644	0.111
Obesity													
Overweight at age 35 (BMI>25)	Yes	35	10	7	10	9	1.000	0.714	0.900	1.000	0.210	0.436	0.264
Obese at age 35 (BMI >30)	Yes	35	10	7	10	9	0.800	0.571	0.800	0.667	0.641	0.909	0.552
Severely Obese at age 35 (BMI>35)	Yes	35	10	7	10	9	0.100	0.429	0.500	0.333	0.161	0.804	0.509
Waist-Hip Ratio													
Waist-Hip Ratio at age 35	Yes	35	8	7	10	8	0.898	0.866	0.982	0.872	0.708	0.222	0.146
Abdominal Obesity at age 35	Yes	35	8	7	10	8	0.750	0.429	0.800	0.625	0.606	0.979	0.336
Risk Factors													
Metabolic Syndrome at age 35 (NCEP)	Yes	35	10	7	10	9	0.300	0.286	0.400	0.444	0.946	0.518	0.972
Hypertension and Obesity (age 35)	Yes	35	10	7	10	9	0.100	0.286	0.100	0.222	0.531	0.205	0.534
Hypertension and Severe Obesity (age 35)	Yes	35	10	7	10	9	0.200	0.143	0.200	0.222	0.830	0.508	0.805
Hypertension and Dyslipidemia (age 35)	Yes	35	8	7	10	8	0.000	0.143	0.200	0.250	0.349	0.436	0.704
Framingham													
Framingham Probability	Yes	35	10	7	10	9	1.239	0.978	1.493	1.449	0.592	0.631	0.675

Notes: This table examines the effect of the second-stage randomization (school-age component) of the Abecedarian intervention. We present descriptive statistics and *p*-values from three hypothesis testing for selected outcomes. We list here the information of each column. **Col.1:** variable of interest; **Col.2:** if the variable was reversed or not. By reversed we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; **Col.3:** age of the participant when the data was collected; **Col.4—Col.7:** sample sizes of each group. **Col.8—Col.11:** group means of the variable of interest. **Col.12—Col.14:** two-sided *p*-values for three hypotheses described in the text. **Col. 12:** H1 : $\mu_{TT} = \mu_{TC}$ (i.e. No program effect between TT and TC). **Col.13:** H2 : $\mu_{CT} = \mu_{CC}$ (i.e. No Program effect between CT and CC), and **Col.14:** H3 : $\mu_{TT} = \mu_{TC}$ and $\mu_{CT} = \mu_{CC}$. The testing was conducted by a conditional permutation procedure using cohort, mother's IQ, High Risk Index (HRI) at birth, and the number of siblings as conditioning variables. For each testing, 5,000 permutations were drawn to construct the distribution of test statistics.

G A Simple Mediation Analysis

Tables S19 and S20 present a simple mediation analysis. As discussed in Section A, the Abecedarian intervention has educational, health care, and nutritional components. Hence, there might be different channels through which the treatment effects arise. In this section we examine an extensive set of potential mediators, which have been shown in previous work to be affected by the intervention: cognitive ability in childhood (60), infant temperament (88), body mass index at one year of age and the availability of health insurance at age 30 (Table 3), attained educational level by age 30, employment status and earnings at age 30 (89). These mediators were chosen because both they were affected by the intervention and because they have been shown in the literature to affect the outcomes of interest: obese children are more likely to display a variety of metabolic conditions in adulthood (90, 91), and in particular obesity at 12 months of age has been shown to be strongly correlated with later obesity (92); lower childhood intelligence test scores have been associated with higher prevalence of obesity, blood pressure and the metabolic syndrome (93-95); child temperament has also been shown to be predictive of precursors of obesity and of the metabolic syndrome (96, 97); and finally, education significantly affects health (98), as do income (99, 100) and health insurance, especially in vulnerable populations (101). Additionally, we examine as mediators indicators for the presence of frequent health problems and hospitalizations in the last 3 years at age 15, since it has been shown that health in adolescence is affected by early life health (102). Adolescence is considered by some to be a window of opportunity for treating emerging health problems and for preventing adolescent ill-health (103). On the other hand, behavioral risk factors such as smoking, drinking, physical activity and diet, which are usually targeted to reduce Non-Communicable Diseases (104), were not significantly affected by the intervention (Table S13), hence they are not considered plausible mediators.

We present two tables of results. Table S19 is based on a model where the coefficients of the effects of the mediating variables are restricted to be the same across treatment and controls. Table S20 is based on a model where the coefficients are not so restricted. We test and fail to reject the equality of coefficients across treatment and control group for almost all the mediators considered (Table S20).

Our mediation analysis is based on a Blinder-Oaxaca decomposition (105, 106) which is an application of a Laspeyres index. Consider the following model:

$$Y_d = \alpha_d X_d + \tau_d + \varepsilon_d,$$

where Y is the outcome, X is the mediator, τ is a constant and $d \in \{0,1\}$, where 1 denotes the treatment group, and 0 denotes the control group. The difference in mean outcomes across treatment and control groups $E(\Delta Y) = E(Y_1 - Y_0)$ can be decomposed into a part which is explained by group differences in the mediating variables $\tilde{\alpha}E(\Delta X) = \tilde{\alpha}E(X_1 - X_0)$ (where $\tilde{\alpha}$ is defined below), and into an unexplained component. Hence, the share of the treatment effect explained by each mediator is $\frac{\tilde{\alpha}E(\Delta X)}{E(\Delta Y)}$.

We report two sets of results. In a first set (Table S19), we restrict the effect of the coefficients on the mediating variables to be the same for the treatment and control groups: $\tilde{\alpha} = \alpha_1 = \alpha_0 = \alpha$. In this case, we choose α based on the following pooled model:

$$Y = \alpha X + \gamma D + \tau + \varepsilon,$$

where D is a dummy for treatment status (107).

In a second set (Table S20), we do not restrict the effect of the coefficients on the mediating variables to be the same for the treatment and control groups. In this case, we choose $\tilde{\alpha}$ as the average of the coefficients estimated for each group: $\tilde{\alpha} = 0.5\alpha_1 + 0.5\alpha_0$, following reference (108). When we restrict the coefficients of the effects on the mediating variables to be the same

across treatment and control groups, the results reported in Table S19 show that the effect of the treatment on hypertension and obesity is significantly mediated by the body mass index of the child at around one year of age. The estimated shares range from 0.49 to 0.67. Instead, the shares of the treatment effect mediated by cognition and temperament in childhood, and by education and labor market outcomes and the availability of health care in adulthood are much smaller and in general not statistically significant. We also test (Table S20) and in very few cases we reject equality of the coefficients of the effects of the mediating variables across treatment and control groups. When we do not restrict the regression coefficients for the treatment and control group to be the same, the estimated shares lose statistical significance (Table S20). The only notable exception appears the occurrence of hospitalizations in the last 3 years at age 15, which appears to explain a large share of the effect of the treatment on multiple risk factors. Given the small sample size, the numerous possible channels of intervention and the imprecision of the estimates, these findings are necessarily speculative at this point.

Table S19: Decompositions of Treatment Effects on Health Outcomes (Biomedical Sweep), Males

Outcome	Statistic	BMI 12m	IBR Task 12m	MDI 12m	SB IQ 48m	WPPSI 60m	Health Probl 15y	Hospital 15y	Education 30y	Employment 30y	Earnings 30y	Health Insurance 30y
Blood Pressure												
Systolic Blood Pressure	effect	0.473	0.025	0.000	-0.344	0.063	0.068	-0.061	-0.076	-0.013	0.139	0.023
	<i>p</i> -value	0.125	0.414	0.499	0.098	0.322	0.444	0.458	0.365	0.487	0.199	0.457
Diastolic Blood Pressure	effect	0.564	0.025	-0.002	-0.307	0.061	-0.068	-0.040	-0.043	0.123	0.080	0.229
	<i>p</i> -value	0.099	0.407	0.494	0.098	0.363	0.437	0.470	0.409	0.367	0.272	0.175
Hypertension												
Hypertension (Sys. > 140 and Dias. > 90)	effect	0.491	0.018	-0.005	-0.301	0.028	0.040	0.245	0.075	-0.056	0.139	0.062
	<i>p</i> -value	0.099	0.422	0.486	0.173	0.462	0.461	0.295	0.356	0.439	0.174	0.413
Hypertension (Sys. > 140 or Dias. > 90)	effect	0.534	0.040	0.000	0.010	0.142	0.198	0.193	0.070	0.462	0.178	0.201
	<i>p</i> -value	0.094	0.393	0.499	0.484	0.265	0.306	0.332	0.367	0.107	0.207	0.279
Nutrition												
Vit. D Deficiency	effect	0.465	0.021	-0.022	0.106	0.212	-0.262	-0.164	0.072	0.144	-0.022	0.107
	<i>p</i> -value	0.150	0.417	0.430	0.410	0.277	0.187	0.258	0.330	0.226	0.465	0.315
Risk Factors												
Hypertension and Obesity	effect	0.587	0.000	0.013	0.094	0.171	0.110	0.209	0.114	0.113	0.013	-0.130
	<i>p</i> -value	0.069	-	0.464	0.305	0.128	0.360	0.298	0.318	0.357	0.469	0.238
Hypertension and Severe Obesity	effect	0.315	0.000	0.015	0.344	0.208	0.262	0.421	0.267	0.264	0.016	-0.051
	<i>p</i> -value	0.124	-	0.444	0.041	0.101	0.229	0.140	0.098	0.222	0.365	0.359
Hypertension and Dyslipidemia	effect	0.144	0.002	0.000	0.070	0.152	0.244	0.473	0.084	0.117	0.015	-0.081
	<i>p</i> -value	0.355	0.481	0.498	0.385	0.146	0.257	0.142	0.304	0.356	0.384	0.317
Metabolic Syndrome (NCEP)	effect	0.070	0.000	0.002	0.236	0.241	0.111	0.627	0.228	0.072	-0.008	-0.328
	<i>p</i> -value	0.459	0.497	0.490	0.189	0.117	0.398	0.171	0.134	0.450	0.443	0.112
Framingham												
Framingham Probability	effect	-0.149	0.010	0.001	0.249	0.221	0.449	1.233	0.017	0.135	0.164	-0.037
	<i>p</i> -value	0.326	0.473	0.495	0.262	0.194	0.219	0.044	0.470	0.375	0.205	0.456

Note: This table presents an analysis of linear decomposition of health effects in the mid-30s based on a set of mediators affected by the intervention: body mass index (BMI) at 12 months, the Task Orientation scale of the Infant Behavior Record (IBR) at 12 months, the Bailey Mental Development Index (MDI) at 12 months, the Stanford-Binet IQ score at 48 months, the Weschler Preschool and Primary Scale of Intelligence (WPPSI) Block Design Subtest scale at 60 months, the presence of frequent health problems and hospitalizations in the last 3 years at age 15, educational attainment, employment status, earnings and health insurance at age 30. Unconditional shares of the treatment effect explained by each mediator (models using one mediator at a time) are shown, based on Blinder-Oaxaca decomposition (105, 106): $\alpha E(\Delta X)/E(\Delta Y)$, where $\Delta X = X_1 - X_0$, $\Delta Y = Y_1 - Y_0$, X is the mediator, Y is the outcome, 0 is control and 1 is treatment group. Regression coefficients for the treatment and control group are restricted to be the same (α), and are based on the following pooled model: $Y = \alpha X + \gamma D + \tau + \epsilon$, where D is a dummy for treatment status and τ is a constant (107). One sided *p*-values are associated with the inference that tests if the estimated explained share is statistically different from zero and are based on 1000 bootstrap draws. *p*-values below 10% are in bold. BMI = Body Mass Index. M=months.

Table S20: Unrestricted Decompositions of Treatment Effects on Health Outcomes (Biomedical Sweep), Males

Outcome	Statistic	BMI 12m	IBR Task 12m	MDI 12m	SB IQ 48m	WPPSI 60m	Health Probl 15y	Hospital 15y	Educational 30y	Employment 30y	Earnings 30y	Health Insurance 30y
Blood Pressure												
Systolic Blood Pressure	effect	0.397	0.060	-0.003	-0.346	0.078	0.130	-0.159	-0.086	0.026	0.863	0.009
	<i>p</i> -value	0.402	0.422	0.496	0.147	0.442	0.326	0.333	0.363	0.456	0.442	0.483
	$H_0: \alpha_1 = \alpha_0$	0.482	0.435	0.481	0.650	0.130	0.466	0.164	0.631	0.540	0.383	0.878
Diastolic Blood Pressure	effect	0.349	0.036	-0.007	-0.310	0.084	-0.035	-0.058	-0.037	0.174	1.786	0.206
	<i>p</i> -value	0.432	0.444	0.490	0.171	0.435	0.445	0.429	0.435	0.238	0.348	0.233
	$H_0: \alpha_1 = \alpha_0$	0.032	0.791	0.120	0.524	0.029	0.672	0.792	0.727	0.368	0.021	0.837
Hypertension												
Hypertension (Sys. > 140 and Dias. > 90)	effect	0.494	0.058	-0.003	-0.301	0.065	-0.010	0.159	0.068	-0.070	1.168	0.012
	<i>p</i> -value	0.374	0.413	0.494	0.268	0.453	0.483	0.285	0.400	0.402	0.325	0.485
	$H_0: \alpha_1 = \alpha_0$	0.976	0.185	0.795	0.988	0.009	0.532	0.138	0.727	0.824	0.229	0.716
Hypertension (Sys. > 140 or Dias. > 90)	effect	0.419	0.029	-0.008	0.012	0.178	0.065	0.089	0.090	0.549	1.859	0.168
	<i>p</i> -value	0.381	0.455	0.487	0.487	0.166	0.395	0.373	0.354	0.044	0.251	0.318
	$H_0: \alpha_1 = \alpha_0$	0.284	0.724	0.011	0.571	0.000	0.077	0.073	0.310	0.143	0.031	0.816
Nutrition												
Vit. D Deficiency	effect	0.591	0.012	-0.037	0.113	0.171	-0.337	-0.233	0.071	0.138	0.425	0.149
	<i>p</i> -value	0.132	0.462	0.382	0.430	0.301	0.064	0.114	0.341	0.251	0.380	0.256
	$H_0: \alpha_1 = \alpha_0$	0.362	0.593	0.338	0.708	0.000	0.377	0.234	0.823	0.819	0.078	0.247
Risk Factors												
Hypertension and Obesity	effect	0.534	0.000	0.026	0.098	0.199	0.058	0.133	0.155	0.061	1.288	-0.089
	<i>p</i> -value	0.391	-	0.459	0.420	0.112	0.383	0.284	0.281	0.378	0.338	0.362
	$H_0: \alpha_1 = \alpha_0$	0.541	0.769	0.029	0.576	0.000	0.285	0.134	0.147	0.293	0.076	0.699
Hypertension and Severe Obesity	effect	0.229	0.000	0.028	0.361	0.235	0.197	0.325	0.321	0.199	0.951	-0.093
	<i>p</i> -value	0.457	-	0.460	0.244	0.357	0.174	0.050	0.111	0.151	0.413	0.367
Hypertension and Dyslipidemia	effect	0.105	0.006	0.000	0.071	0.172	0.174	0.366	0.095	0.088	0.776	-0.143
	<i>p</i> -value	0.483	0.489	0.499	0.429	0.371	0.235	0.053	0.325	0.344	0.407	0.318
Metabolic Syndrome (NCEP)	effect	0.073	0.001	0.004	0.246	0.280	0.083	0.533	0.273	0.055	-0.869	-0.895
	<i>p</i> -value	0.495	0.499	0.496	0.363	0.411	0.389	0.076	0.262	0.439	0.443	0.002
Framingham												
Framingham Probability	effect	-0.018	0.010	0.002	0.257	0.236	0.190	1.265	0.019	0.094	-0.988	-0.225
	<i>p</i> -value	0.496	0.483	0.495	0.294	0.345	0.276	0.024	0.478	0.372	0.440	0.206
	$H_0: \alpha_1 = \alpha_0$	0.105	0.956	0.581	0.105	0.329	0.001	0.465	0.929	0.586	0.213	0.062

Note: This table presents an analysis of linear decomposition of health effects in the mid-30s based on a set of mediators affected by the intervention: body mass index (BMI) at 12 months, the Task Orientation scale of the Infant Behavior Record (IBR) at 12 months, the Bailey Mental Development Index (MDI) at 12 months, the Stanford-Binet IQ score at 48 months, the Weschler Preschool and Primary Scale of Intelligence (WPPSI) Block Design Subtest scale at 60 months, the presence of frequent health problems and hospitalizations in the last 3 years at age 15, educational attainment, employment status, earnings and health insurance at age 30. Unconditional shares of the treatment effect explained by each mediator (models using one mediator at a time) are shown, based on Blinder-Oaxaca decomposition (105, 106): $\tilde{\alpha}E(\Delta X)/E(\Delta Y)$, where $\Delta X = X_1 - X_0$, $\Delta Y = Y_1 - Y_0$, X is the mediator, Y is the outcome, 0 is control and 1 is treatment group. Regression coefficients for the treatment and control groups are not restricted to be the same ($\tilde{\alpha}$), and are the average of the coefficients estimated for each group: $\tilde{\alpha} = 0.5\alpha_1 + 0.5\alpha_0$ (108), based on the following models: $Y_d = \alpha_d X_d + \tau_d + \varepsilon_d$, where τ is a constant and $d \in \{0,1\}$. One sided *p*-values are associated with the inference that tests if the estimated explained share is statistically different from zero and are based on 1000 bootstrap draws. For each outcome, the third row reports two-sided *p*-values associated with the inference that the coefficients of the effects of the mediating variables are statistically significantly different across treatment and control groups, i.e. $H_0: \alpha_1 = \alpha_0$. For the outcomes ‘‘Hypertension and Severe Obesity’’, ‘‘Hypertension and Dyslipidemia’’ and ‘‘Metabolic Syndrome’’, these cannot be computed since the prevalence of these conditions is 0 in the treatment group. *p*-values below 10% are in bold. M=months.

H Methodology for Permutation-based Inference

The standard model of program evaluation describes the observed outcome Y_i of participant $i \in J$ by

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

where $J = \{1, \dots, N\}$ denotes the sample space indexing set, D_i denotes the treatment assignment for participant $i \in J$, ($D_i = 1$ 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.

Randomized experiments solve possible 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-program variables X used in the randomization protocol. All variables are defined in the common probability space (Ω, \mathcal{F}, P) . In our notation, a randomized experiment satisfies

Assumption 1 $Y(d) \perp\!\!\!\perp D | X; d \in \text{supp}(D)$,

where variables $X = (X_i; i \in J), D = (D_i; i \in J)$ are N -dimensional vectors of treatment assignments and pre-program variables, and $Y(d) = (Y_i(d_i); i \in J, d_i \in \{0, 1\})$ and $d \in \text{supp}(D) = \{0, 1\}^{|J|}$ denotes the vector of counterfactual outcomes. In the same fashion, we represent the vector of observed outcomes of Equation (1) by $Y = (Y_i; i \in J)$. The hypothesis of no treatment effect is equivalent to the statement that the conditional counterfactual outcome vectors have the same distribution:

Hypothesis 1 $Y(d) \stackrel{d}{=} Y(d') | X; d, d' \in \text{supp}(D)$,

Hypothesis 1 can be restated in a more tractable form:

Hypothesis 1 *Under Assumption 1 and Hypothesis 1, we have that $Y \perp\!\!\!\perp D | X$.*

Testing Hypothesis 1 poses some statistical challenges for the study of the ABC experiment. First, small sample sizes cast doubt on inferences that rely on the asymptotic behavior of test statistics. We address the problem of small sample size by generating exact test statistics conditioned on data. Second, the presence of multiple outcomes allows for the arbitrary selection of statistically significant outcomes. The selective reporting of “statistically significant” outcomes is often termed *cherry picking* and generates a downward biased inference with distorted, undersized p -values. We solve the problem of cherry picking by implementing a multiple-hypothesis testing based on the stepdown procedure of Romano and Wolf(24). They explain that the stepdown procedure strongly controls for family-wise error rate (FWER), while traditional tests do not. Also, Romano and Wolf (24) shows that the strong FWER control can be obtained by ensuring a certain monotonicity condition on the test statistics. This requirement is weaker than the assumption of subset pivotality, used in the various methods of resampling outcomes presented in Westfall and Young(109).

Our method is based on three steps. First, we seek to characterise the exact conditional distribution of $D | X$. Specifically, we characterize a multiset $D_x(d)$, defined by:

$$D_x(d) = \{d' \in \{0,1\}^{|J|}; P(D = d | X = x) = P(D = d' | X = x)\},$$

such that the distribution of D conditioned on the realised data is uniform among elements of $D_x(d)$. Next we use the assumption of null hypothesis of no treatment effects, i.e. $H_0: Y \perp\!\!\!\perp D | X$, to generate the exact conditional distribution of a test statistic $T(Y, D) | X$. Under it, we can construct inferences that control for the probability of falsely rejecting the null

hypothesis. We control for this probability in two ways: (1) in the case of single (joint) null hypothesis, we control for the standard Type-I error; (2) in the case of multiple hypothesis inference, we control for the family-wise error rate (FWER).

More notation is necessary to describe the method. Let K represent the indexing set for all available outcomes $Y_k; k \in K$. We represent the single (joint) null hypothesis that a set $L \subset K$ of outcomes $Y_k; k \in L$ are jointly independent of treatment status D conditional on pre-program variables X by

$$H_L : Y_L \perp\!\!\!\perp D \mid X, \text{ where } Y_L = (Y_k : k \in L). \quad (2)$$

When L is a singleton, say $L = \{k\}$, then the null hypothesis is given by $H_{\{k\}} : Y_k \perp\!\!\!\perp D \mid X$. In this notation, we can write the joint Hypothesis H_L as $H_L = \bigcap_{k \in L} H_{\{k\}}$.

Our goal is to tests a single (joint) null hypothesis controlling for the probability of a Type I error at level α , that is, $P(\text{reject } H_L \mid H_L \text{ is true}) \leq \alpha$. To do so, we rely on the fact that, under Hypothesis (2),

$$(Y_L, D) \mid X \stackrel{d}{=} (Y_L, gD) \mid X \quad \forall g \in \mathbf{G}_X, \quad (3)$$

where \mathbf{G}_X consists of all of the permutations within strata of X , that is,

$$\mathbf{G}_X = \{g; g : J \rightarrow J \text{ is a bijection and } g(j) = j' \Rightarrow (X_j) = (X_{j'})\},$$

and gD is a vector defined by:

$$gD = (\tilde{D}_i \in \text{supp}(D); i \in J \text{ and } \tilde{D}_i = D_{g(i)}).$$

We use relationship (3) to generate a statistical test. The exact distribution of the test statistic $T_L(Y_L, gD)$ is obtained by re-evaluating $T_L(Y_L, gD)$ as g varies in \mathbf{G}_X . Note that the

inference on Hypothesis 2 is depend on the choice of statistics. That is to say that even though any statistics $T_L(Y_L, D)$ whose values provide evidence against the null hypothesis can be used, the inference depends on this choice of statistic. An example of such statistics is the maximum of the t -statistic associated with the difference in means between treated and control groups over outcomes Y_k such that $k \in L$. Formally,

$$T_L(Y_L, D) = \max_{k \in L} T_k(Y_k, D), \quad (4)$$

where $T_k(Y_k, D)$ is the t -statistics for outcome Y_k . Relationship (3) implies that $T_L(Y_L, D) | X = x \stackrel{d}{=} T_L(Y_L, gD) | X = x$ for any $g \in \mathbf{G}_x$. Let $d \in \{0, 1\}^J$ satisfy $P(D = d | X = x) > 0$. The distribution of D conditioned on $X = x$ is uniform across elements of $\mathbf{D}_x(d)$ (see Lehmann and Romano (110)). Thus a critical value $c_{L,x}(Y_L, d, \alpha)$ such that $P(T_L(Y_L, d') > c_{L,x}(Y_L, d, \alpha) | X = x, H_L \text{ is true}) \leq \alpha$ can be computed as:

$$c_{L,x}(Y_L, d, \alpha) = \inf_{t \in \mathbf{R}} \left\{ \sum_{d' \in \mathbf{D}_x(d)} 1\{T_L(Y_L, d') \leq t\} \geq (1 - \alpha) |\mathbf{D}_x| \right\},$$

where $1\{\cdot\}$ is the indicator function and $|\mathbf{D}_x|$ denotes the cardinality of the set \mathbf{D}_x . The following notation is useful to further characterize $c_{L,x}(Y_L, d, \alpha)$. Let $T_{L,x}^{(1)}, \dots, T_{L,x}^{(|\mathbf{D}_x(d)|)}$ be the sequence of increasing ordered statistics $T_L(Y_L, d')$ for $d' \in \mathbf{D}_x(d)$. In this notation we can write the critical value as

$$c_{L,x}(Y_L, d, \alpha) = T_{L,x}^{(\lceil (1-\alpha)|\mathbf{D}_x \rceil)} \quad (5)$$

where $\lceil a \rceil$ stands for the smallest integer bigger or equal than a .

Under the null hypothesis H_L , the probability that a test statistic is bigger or equal than

the statistic $T_L(Y_L, d)$ actually observed, (i.e. the p -value), is given by:

$$p_{L,x}(d) = \inf_{\alpha \in [0,1]} \{c_{L,x}(Y_L, d, \alpha) \leq T_L(Y_L, d)\}. \quad (6)$$

Now let $r_{L,x} \in \{1, \dots, |\mathbf{D}_x(d)|\}$ be the lowest rank that the value of the observed test statistic

$T_L(Y_L, d)$ takes in the sequence $T_{L,x}^{(1)}, \dots, T_{L,x}^{(|\mathbf{D}_x(d)|)}$, that is to say:

$$r_{L,x} = 1 + \sum_{d' \in \mathbf{D}_x(d)} 1\{T_L(Y_L, d') < T_L(Y_L, d)\}.$$

Thus:

$$T_{L,x}^{(r_{L,x})} = T_L(Y_L, d). \quad (7)$$

By the ordered property of $T_{L,x}^{(r)}; r \in \{1, \dots, |\mathbf{D}_x(d)|\}$ and the definition of $r_{L,x}$, it follows that:

$$p_{L,x}(d) = 1 - \frac{r_{L,x}}{|\mathbf{D}_x(d)|}. \quad (8)$$

Moreover, p -value $p_{L,x}(d)$ has the following property:

$$P(p_{L,x}(d) \leq \phi | X = x) \leq \phi \forall \phi \in [0, 1].$$

We implement an inference method that tests for the multiple null hypothesis that each outcome $Y_k; k \in L$ is independent of treatment status D conditional on pre-program variables X . The representation of these multiple hypotheses in the same fashion as the single (joint) null hypothesis, is:

$$H_L = \bigcap_{k \in L} H_{\{k\}}; H_{\{k\}} : Y_k \perp\!\!\!\perp D | (Z, U).$$

The multiple hypothesis testing differs from the single (joint) hypothesis testing in the way it controls for the probability of false rejection. Specifically, let the subset L_0 be the set of true Hypothesis $H_{\{k\}}$ such that $k \in L_0 \subset L$. Our multiple hypothesis testing controls for the

familywise error rate (FWER), that is, the probability of even one false rejection among the set of true hypothesis L_0 . Formally, we control for:

$$P(\text{reject at least one } H_{\{k\}}; k \in L_0 \mid H_{L_0} \text{ is true}) \leq \alpha,$$

while single (joint) hypothesis testing controls for $P(\text{reject } H_L \mid H_L \text{ is true}) \leq \alpha$.

Bonferroni or Holm procedures are examples of inference methods that test multiple hypothesis controlling for FWER. These methods rely upon a “least favorable” dependence structure among the p -values. The stepdown procedure of Romano and Wolf (24) is less conservative as it accounts for the information about the dependence structure of p -values. The method is based on a monotonicity assumption, which, in our case, can be stated as:

$$c_{K,x}(Y_K, d, \alpha) \geq c_{L_0,x}(Y_{L_0}, d, \alpha) \text{ for any subset } K \text{ of } L \text{ containing } L_0 \text{ i.e. } L_0 \subset K \subset L. \quad (9)$$

Inequality (9) is satisfied by our choice of test statistic (4) and the fact that $L_0 \subset K$.

The stepdown procedure given in Romano and Wolf (24) is a stepwise method summarized in the following algorithm:

Algorithm 1.

Step 1: Set $L_1 = L$. If

$$\max_{k \in L_1} T_k(Y_k, d) \leq c_{L_1,x}(Y_{L_1}, d, \alpha), \quad (10)$$

then stop and reject no null hypotheses; otherwise, reject any $H_{\{k\}}$ with

$$T_k(Y_k, d) > c_{L_1,x}(Y_{L_1}, d, \alpha)$$

and go to Step 2.

⋮

Step j : Let L_j denote the indices of remaining null hypotheses. If

$$\max_{k \in L_j} T_k(Y_k, d) \leq c_{L_j, x}(Y_{L_j}, d, \alpha), \quad (11)$$

then stop and reject no further null hypotheses; otherwise, reject any $H_{\{k\}}$ with

$$T_k(Y_k, d) > c_{L_j, x}(Y_{L_j}, d, \alpha)$$

and go to Step $j + 1$.

⋮

We can compute the multiplicity-adjusted p -values of Equations(10)–(11) in the same fashion described by Equations (6)–(8).

I Bootstrap Analysis of Test Statistics and an Analysis of Confidence Intervals

This section presents the details of the estimation and inference of the Abecedarian treatment effects using the bootstrap method. Our analysis is based on Romano and Wolf(24). We also refer to Romano (111) for a theoretical background.

Some notation is necessary to define our inference problem. We use $J = \{1, \dots, N\}$ for the sample space indexing where D_i denotes the treatment assignment for participant $i \in J$, ($D_i = 1$ if treatment occurs, $D_i = 0$ otherwise). We use $Y_{k,i}; i \in J, k \in L$ for the k -th outcome of participant i . We use $X_i; i \in J$ for the pre-program variables X and of participant i that were used in the randomization protocol. Likewise, we use Z_i for the pre-program variables not used in the randomization protocol that we wish to control for. We use $Y_k = (Y_{k,i}; i \in J)$ for the outcome vector comprising all participants in J and $D = (D_i; i \in J)$ for the vector of treatment status. In the same fashion, we use $X = (X_i; i \in J)$ and $Z = (Z_i; i \in J)$ for the vector of pre-program variables X_i and Z_i across participants in J . All variables are defined in the common probability space (Ω, F, P) .

We model outcome $Y_k; k \in L$ through the following linear regression:

$$Y_{k,i} = \kappa_k + \Delta_k D_i + X_i \beta_k + Z_i \zeta_k + \varepsilon_i. \quad (12)$$

Let $\hat{\Delta}_{k,N}(Y_k, D, X, Z)$, or simply $\hat{\Delta}_{k,N}$, be the conditional average treatment effect estimate of the Abecedarian intervention for outcome Y_k . For the case of the least squares estimator,

$$\hat{\Delta}_{k,N} = (D' M_B D)^{-1} D' M_B Y,$$

where $B = [tXZ]$ is the matrix that comprises a vector of ones (t) and the matrices of the realized variables X, Z and M_B is the linear operator of the orthogonal projection on the sample space generate by the columns of B , that is,

$$M_B = I - B(B'B)^{-1}B',$$

where I is the identity matrix. This is our estimator for the parameter Δ_k of Equation (12). Our analysis does not apply exclusively to this choice of estimated parameter and holds for any other estimated parameter of interest. In particular, we also investigate the following objects: (1) the t -statistic associated with the treatment effect estimate $\hat{\Delta}_{k,N}$ cited above; (2) the weighted average treatment effect estimate whose weights are given by inverse probability weighting (IPW); or (3) the t -statistic associated with this IPW estimate.

Our goal is to test the single (joint) null hypothesis $H_L : \bigcap_{k \in L} H_k$ such that $H_k : \Delta_k = 0$ against the alternative hypothesis that $\Delta_k > 0; k \in L$. To do so, we will use the bootstrap method described in Romano and Wolf(24). Bootstrap relies on approximating the distribution of $\hat{T}_{L,N} = \max_{k \in L}(\hat{T}_{k,N})$, where $\hat{T}_{k,N} = \tau_N(\hat{\Delta}_{k,N} - \Delta_k)$. Parameter τ_N is introduced for purposes of asymptotic analysis so that a limiting distribution for $\tau_N \hat{\Delta}_{k,N}(Y, X, Z)$ is not degenerate. Typically, $\tau_N = N^{1/2}$. Moreover, when L is a singleton, say $L = \{k\}$, then the null hypothesis is the standard single-hypothesis testing of $H_k : \Delta_k = 0$.

We denote the distribution of $\hat{T}_{k,N}$ by $F_{N,k}(t) = P(\hat{T}_{k,N} < t)$ which we assume converges to a non-degenerate limiting law $F_k(t)$ when N tends to infinity. Formally, we write:

Assumption 2 $F_{N,k}(t) \xrightarrow{L} F_k(t) \forall k \in L$ such that H_k is true and $t \in \text{supp}(\hat{T}_{k,N})$.

Let the distribution of $\hat{T}_{L,N}$ be $F_{N,L}(t) = P(\hat{T}_{L,N} < t)$. Assumption 2 implies that the limiting distribution of $F_{N,L}(t)$, i.e. $F_L(t)$, exists. Also let the critical value for the α -quantile of variable $\hat{T}_{k,N}$ be given by:

$$c_{N,L}(\alpha, F_{N,L}) = \inf_{t \in \mathbf{R}} \{F_{N,L}(t) \geq (1-\alpha)\}. \quad (13)$$

If we assume the regularity condition that $F_L(P)$ is continuous and strictly increasing in its support, then $c_{N,L}(\alpha, F_{N,L}) \rightarrow c_L(\alpha, F_L)$.

The bootstrap approximation of the distribution $F_{N,k}(t); k \in L$ is denoted by $\hat{F}_{N,k}(t); k \in L$ and is obtained by re-sampling the data with replacement. Formally let the re-sampling function $h \in \mathbf{H} = \{h; h: J \rightarrow J\}$ be the group action that transform the data in the following manner:

$$hX = (\tilde{X}_i; i \in J; \tilde{X}_i = X_{h(i)}).$$

Then the distribution of $\hat{\Delta}_{k,N}(Y_k, X, Z); k \in L$ is given by the values $\hat{\Delta}_{k,N}^h = \hat{\Delta}_{k,N}(hY_k, hX, hZ); k \in L$ takes as h varies in \mathbf{H} . We can then use the bootstrap method to approximate of the distribution of $\hat{T}_{k,N}$ by the values that $\hat{T}_{k,N}^h = \tau_N(\hat{\Delta}_{k,N} - \hat{\Delta}_{k,N}^h); h \in \mathbf{H}$ takes.

Then the bootstrap approximation of $\hat{F}_{N,L}(t)$ is given by:

$$\hat{F}_{N,L}(t) = \sum_{i=1}^{|\mathbf{H}|} I\{\hat{T}_{L,N}^h < t\}, \quad (14)$$

where $\hat{T}_{L,N}^h = \max_{k \in L} \{\hat{T}_{k,N}^h\}$ and $I\{\cdot\}$ is the indicator function. Equation (14) allow us to compute

the bootstrap critical value $c_{N,L}(\alpha, \hat{F}_{N,L}(t))$ according to the definition (13). Finally, if we define $\hat{T}_{L,N}^{(1)}, \dots, \hat{T}_{L,N}^{(|\mathbf{H}|)}$ to be the sequence of increasing ordered statistics $\hat{T}_{L,N}^h$ for $h \in \mathbf{H}$, then the critical value $c_{N,L}(\alpha, \hat{F}_{N,L}(t))$ is given by:

$$c_{N,L}(\alpha, \hat{F}_{N,L}) = \hat{T}_{L,N}^{(\lceil (1-\alpha)|\mathbf{H}| \rceil)}, \quad (15)$$

where $\lceil a \rceil$ stands for the smallest integer bigger or equal than a . In this notation, the bootstrap p -value associated with the null hypothesis H_L is given by:

$$p_{N,L}(\hat{F}_{N,L}) = \inf_{\alpha \in [0,1]} \{c_{N,L}(\alpha, \hat{F}_{N,L}) \leq \hat{T}_{L,N}^h\} = 1 - \frac{\hat{r}_L}{|\mathbf{H}|}, \quad (16)$$

where

$$\hat{r}_L = 1 + \sum_{h \in \mathbf{H}} I\{\hat{T}_{L,N}^h < \hat{T}_{L,N}\}.$$

We use these parameters to perform the stepdown procedure in the same fashion discussed in Supplemental Appendix H. Romano and Wolf (24) also suggest the following $(1-\alpha)$ -level bootstrap confidence region:

$$(\hat{\Delta}_{L,N} - \tau_N^{-1} \hat{T}_{L,N}^{(\lceil (1-\alpha)|\mathbf{H}| \rceil)}, \hat{\Delta}_{L,N} - \tau_N^{-1} \hat{T}_{L,N}^{(\lceil (\alpha)|\mathbf{H}| \rceil)}). \quad (17)$$

We present the inference from the bootstrap procedure in Tables S21—S24 for males (S21 and S22) and females (S23 and S24). We compare the analyses of treatment effects obtained from the bootstrap and the permutation methods in Tables S25—S26. Bootstrap confidence intervals are

also presented in the same tables. p -values are corrected for attrition and imbalance of pre-program variables in both methods. p -values from multiple hypothesis testing are also displayed for each of methods.

Each of Tables S25-S26 consists of 13 columns. Column 1 names the variable of interest. Column 2 shows if the variable was reversed or not. By "reversed", we mean that the value from raw variable is multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome. Column 3 indicates the age of the participant when the data was collected. Column 4 and Column 5 displays the sample sizes for the control and the treatment groups. Column 6 and Column 7 displays the control group mean and the difference in means. Column 8 and Column 9 are the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level. Column 10 presents the one-sided single hypothesis block permutation where p -values are corrected for attrition bias via the method of Inverse Probability Weighting (IPW). Column 11 presents its corresponding multiple hypothesis testing (step-down) p -values. Column 12 and Column 13 present the p -values obtained from the bootstrap inference method corresponding to Column 10 and Column 11, respectively. Table S25 is for males and Table S26 is for females.

As shown in those tables, the inferences from the bootstrap are essentially the same as that from the permutation approach. And, the bootstrap confidence intervals are in agreement with this finding. Based on these findings, we conclude that although the bootstrap and the permutation approaches are based on different statistical ideas, they produce the same inference in the case of the health effects of ABC intervention.

Table S21(a): Unconditional Bootstrap Inference for the Abecedarian Intervention, Main Health Results, Biomedical Sweep (Males)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Mean	Treat. Effect	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
				Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Blood Pressure													
Diastolic Blood Pressure at age 35	Yes	-92.000	13.474	3.946	23.561	2.171	0.729	3.819	2.171	0.483	3.338	0.034	0.047
Systolic Blood Pressure at age 35	Yes	-143.333	17.544	4.200	31.847	2.055	0.632	3.463	2.055	0.632	3.463	0.038	0.038
Pre-Hypertension at age 35 (Sys. > 120 and Dias. > 80)	Yes	-0.667	0.246	-0.013	0.504	1.202	-0.058	2.713	1.202	-0.297	2.180	0.109	0.164
Pre-Hypertension at age 35 (Sys. > 120 or Dias. > 80)	Yes	-0.778	0.094	-0.143	0.320	0.496	-0.744	1.778	0.496	-0.744	1.778	0.284	0.284
Hypertension at age 35 (Sys. > 140 and Dias. > 90)	Yes	-0.444	0.339	0.097	0.580	2.134	0.593	4.099	2.134	0.399	3.432	0.037	0.050
Hypertension at age 35 (Sys. > 140 or Dias. > 90)	Yes	-0.556	0.345	0.083	0.590	1.874	0.430	3.658	1.874	0.430	3.658	0.046	0.046
Lab Tests													
HDL Cholesterol at age 35	No	42.000	11.211	5.698	16.733	2.338	1.322	3.437	2.338	0.817	3.233	0.004	0.020
Dyslipidemia at age 35 (HDL < 40)	Yes	-0.417	0.311	0.098	0.533	2.096	0.656	3.895	2.096	0.656	3.895	0.030	0.030
Pre-Diabetes at age 35 (HbA1C > 5.7)	Yes	-0.583	0.110	-0.142	0.353	0.579	-0.741	1.932	0.579	-0.741	1.932	0.279	0.279
Vit. D Deficiency (age 35)	Yes	-0.750	0.382	0.166	0.592	2.158	0.862	3.841	2.158	0.393	2.790	0.013	0.013

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S21(b): Unconditional Bootstrap Inference for the Abecedarian Intervention, Main Health Results, Biomedical Sweep (Males)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Ctr.	Treat.	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
		Mean	Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Obesity													
Overweight at age 35 (BMI ≥ 25)	Yes	-0.750	0.028	-0.222	0.285	0.142	-1.140	1.454	0.142	-1.141	1.454	0.424	0.424
Obese at age 35 (BMI ≥ 30)	Yes	-0.625	0.069	-0.206	0.349	0.318	-0.949	1.633	0.318	-1.235	1.282	0.368	0.481
Severely Obese at age 35 (BMI ≥ 35)	Yes	-0.375	0.264	0.000	0.526	1.592	0.049	3.452	1.592	-0.368	2.240	0.093	0.160
Waist-Hip Ratio at age 35	Yes	-0.962	0.025	-0.017	0.067	0.784	-0.524	2.387	0.784	-0.524	2.387	0.227	0.227
Abdominal Obesity at age 35	Yes	-0.875	0.228	0.000	0.438	1.169	0.000	2.340	1.169	-0.305	2.069	0.098	0.163
Multiple Risk Factors													
Hypertension and Obesity (age 35)	Yes	-0.500	0.389	0.123	0.655	2.307	0.712	4.451	2.307	0.706	3.198	0.032	0.032
Hypertension and Severe Obesity (age 35)	Yes	-0.375	0.375	0.143	0.600	3.157	1.857	5.226	3.157	1.400	3.728	0.001	0.007
Hypertension and Dyslipidemia (age 35)	Yes	-0.333	0.333	0.125	0.545	2.887	1.724	4.559	2.887	1.215	3.566	0.000	0.013
Metabolic Syndrome at age 35 (NCEP)	Yes	-0.250	0.250	0.000	0.500	2.283	1.447	3.843	2.283	1.447	3.843	0.001	0.001
Framingham Probability	Yes	-7.043	2.154	0.276	4.033	1.671	0.198	3.717	1.671	0.197	3.717	0.071	0.071

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S21(c): Unconditional Bootstrap Inference for the Abecedarian Intervention (Males)

Variable	Rev.	Mean	Treatment Effect Estimates			Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
			Ctr.	Treat.	Conf. Inter.	test	Conf. Inter.	Max	Conf. Inter.	<i>p</i> -values			
			Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Health Care													
Coverage at age 30													
Health Insurance Coverage at age 30	No	0.476	0.228	0.048	0.407	1.609	0.328	3.051	1.609	0.328	3.051	0.056	0.056
Buys Health Insurance (age 30)	No	0.333	0.296	0.119	0.474	2.086	0.809	3.591	2.086	0.581	3.187	0.018	0.033
Hospital or doctor office care when sick	No	0.524	0.291	0.122	0.466	2.223	0.915	3.847	2.223	0.485	2.864	0.014	0.034
Smoking													
Never a regular smoker by age 35	No	0.391	-0.070	-0.249	0.108	-0.511	-1.894	0.807	-0.511	-1.894	0.807	0.691	0.691
Age Subject Began Smoking Regularly (Never=.)	No	16.893	2.829	1.235	4.467	2.171	0.992	3.491	2.171	0.272	2.386	0.014	0.056
Cigarettes per Day in Past 30 Days at 21 and 30	Yes	-11.167	2.482	-1.542	6.800	0.798	-0.551	2.163	0.798	-0.939	1.377	0.219	0.379
Alcohol Use													
Did you begin drinking before 17	Yes	-0.609	0.070	-0.109	0.252	0.487	-0.769	1.807	0.487	-1.436	0.612	0.318	0.690
Do you drink and drive?	Yes	-0.391	-0.147	-0.338	0.042	-1.020	-2.464	0.293	-1.020	-2.464	0.293	0.844	0.844
Age Subject Began Drinking (Never=.)	No	17.045	-1.135	-2.626	0.302	-0.989	-2.202	0.292	-0.989	-2.681	-0.394	0.840	0.965
Days Subject Drank in Past 30 Days at 21 - 30	Yes	-9.348	-2.634	-7.008	1.348	-0.770	-1.958	0.442	-0.770	-2.558	-0.406	0.798	0.963
Days Subject Drank 5+ Drinks in Past 30 Days at 21 - 30	Yes	-2.457	-0.972	-2.828	0.820	-0.657	-1.870	0.629	-0.657	-2.508	-0.401	0.755	0.964
Marijuana Use													
Never Used Marijuana by age 30	No	0.130	-0.023	-0.142	0.093	-0.252	-1.571	1.046	-0.252	-1.571	1.046	0.607	0.607
Age of Onset of Marijuana Use	No	15.475	2.025	0.837	3.262	2.133	0.907	3.495	2.133	0.351	2.460	0.013	0.048
Times Subject Used Marijuana Past 30 Days (21 - 30 years old)	Yes	-49.565	1.030	-14.539	16.917	0.083	-1.190	1.390	0.083	-1.547	0.821	0.467	0.657

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S21(d): Unconditional Bootstrap Inference for the Abecedarian Intervention (Males)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Ctr.	Treat.	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
		Mean	Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Physical Activity and Diet													
Exercise 4 or more Days per Week (age 21)	No	0.391	-0.084	-0.259	0.101	-0.603	-1.924	0.737	-0.603	-1.925	0.737	0.718	0.718
No. of Fruit Servings per Day (age 21)	No	0.826	0.020	-0.312	0.369	0.075	-1.191	1.418	0.075	-1.583	0.723	0.462	0.675
Physical Develop													
At Risk Overweight (CDC) at 3 mo.	Yes	-0.227	0.190	0.063	0.316	2.067	0.814	3.297	2.067	0.068	2.173	0.022	0.087
At Risk Overweight (CDC) at 6 mo.	Yes	-0.250	0.170	0.033	0.316	1.572	0.304	2.899	1.572	-0.359	1.748	0.056	0.193
At Risk Overweight (CDC) at 9 mo.	Yes	-0.412	0.412	0.263	0.565	3.244	2.205	4.538	3.244	1.164	3.128	0.000	0.004
At Risk Overweight (CDC) at 12 mo.	Yes	-0.429	0.429	0.292	0.571	4.405	3.206	5.914	4.405	2.254	4.110	0.000	0.000
At Risk Overweight (CDC) at 18 mo.	Yes	-0.389	0.389	0.235	0.533	3.974	2.742	5.422	3.974	1.836	3.719	0.000	0.001
At Risk Overweight (CDC) at 24 mo.	Yes	-0.333	0.333	0.171	0.500	3.586	2.310	5.036	3.586	1.478	3.365	0.000	0.002
At Risk Overweight (CDC) at 36 mo.	Yes	-0.158	0.078	-0.050	0.209	0.794	-0.597	2.116	0.794	-0.597	2.116	0.218	0.218
At Risk Overweight (CDC) at 48 mo.	Yes	-0.300	0.133	-0.034	0.298	1.040	-0.276	2.425	1.040	-0.606	1.882	0.150	0.246
At Risk Overweight (CDC) at 60 mo.	Yes	-0.300	0.175	0.011	0.333	1.434	0.093	2.845	1.434	-0.365	2.010	0.084	0.194
At Risk Overweight (CDC) at 96 mo.	Yes	-0.421	0.301	0.132	0.471	2.377	1.040	3.989	2.377	0.298	2.377	0.010	0.053
Weight-for-Length Change (CDC) at 24 mo.	Yes	-0.858	0.963	0.210	1.717	1.609	0.364	2.909	1.609	0.364	2.909	0.051	0.051
Weight-for-Length Change (WHO) at 24 m.	Yes	-1.265	1.100	0.275	1.912	1.714	0.442	3.044	1.714	0.365	2.936	0.041	0.049

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S22(a): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)

Variable	Rev.	Mean	Treatment Effect Estimates			Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
			Ctr.	Treat.	Conf. Inter.	test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
			Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Blood Pressure													
Diastolic Blood Pressure at age 35	Yes	-58.788	17.413	5.335	27.190	2.759	0.998	4.772	2.759	0.998	4.772	0.030	0.030
Systolic Blood Pressure at age 35	Yes	-82.701	24.320	6.191	37.004	2.869	0.871	4.468	2.869	0.652	4.118	0.038	0.050
Pre-Hypertension at age 35 (Sys. > 120 and Dias. > 80)	Yes	1.853	0.299	0.008	0.554	1.533	0.039	3.155	1.533	-0.292	2.452	0.096	0.159
Pre-Hypertension at age 35 (Sys. > 120 or Dias. > 80)	Yes	1.061	0.142	-0.134	0.371	0.804	-0.722	2.202	0.804	-0.722	2.202	0.254	0.254
Hypertension at age 35 (Sys. > 140 and Dias. > 90)	Yes	-1.144	0.513	0.227	0.701	3.599	1.461	5.715	3.599	1.391	4.878	0.021	0.022
Hypertension at age 35 (Sys. > 140 or Dias. > 90)	Yes	1.568	0.380	0.115	0.635	2.190	0.594	4.097	2.190	0.594	4.097	0.039	0.039
Lab Tests													
HDL Cholesterol at age 35	No	32.863	11.711	6.009	17.509	2.442	1.420	3.760	2.442	0.794	3.244	0.004	0.025
Dyslipidemia at age 35 (HDL < 50)	Yes	0.901	0.243	-0.011	0.474	1.491	-0.078	3.089	1.491	-0.078	3.089	0.111	0.111
Pre-Diabetes at age 35 (HbA1C > 5.7)	Yes	0.010	0.053	-0.228	0.410	0.258	-1.136	2.163	0.258	-1.136	2.163	0.370	0.370
Vit. D Deficiency (age 35)	Yes	-0.788	0.443	0.182	0.649	2.766	1.042	4.602	2.766	1.042	4.602	0.017	0.017

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression the uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to:(1) number of siblings, (2) high risk index birth, (3) mother wais full scale iq score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significance imbalance between attrited and non-attrited groups. Models selection in based of the Akaike information criterion. Variables selected for males at 35 years old are: (1) Mother wais digit symbol, (2) Gestational age indicatoe (less than 37 weeks), (3) Substance abuse score at age 30, (4) Rule breaking score at age 30. Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis bootstrap *p* -value; Col.15: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S22(b): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Ctr.	Treat.	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
		Mean	Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Obesity													
Overweight at age 35 (BMI ≥ 25)	Yes	-0.924	0.151	-0.110	0.348	0.895	-0.627	2.209	0.895	-0.627	2.210	0.226	0.226
Obese at age 35 (BMI ≥ 30)	Yes	1.303	0.210	-0.112	0.462	1.017	-0.498	2.357	1.017	-0.776	1.990	0.197	0.269
Severely Obese at age 35 (BMI ≥ 35)	Yes	0.553	0.352	-0.026	0.597	1.957	-0.137	4.746	1.957	-0.388	2.628	0.114	0.159
Waist-Hip Ratio at age 35	Yes	-1.060	0.038	-0.029	0.089	1.070	-0.767	2.893	1.070	-0.767	2.893	0.217	0.217
Abdominal Obesity at age 35	Yes	-2.784	0.297	0.040	0.540	1.624	0.206	2.958	1.624	-0.398	2.654	0.076	0.154
Multiple Risk Factors													
Hypertension and Obesity (age 35)	Yes	1.320	0.501	0.247	0.702	3.371	1.474	5.443	3.371	1.135	4.298	0.010	0.013
Hypertension and Severe Obesity (age 35)	Yes	0.002	0.484	0.229	0.688	4.270	2.352	6.963	4.270	1.752	4.985	0.001	0.003
Hypertension and Dyslipidemia (age 35)	Yes	-0.475	0.409	0.144	0.608	3.742	1.932	5.940	3.742	1.342	4.540	0.000	0.006
Metabolic Syndrome at age 35 (NCEP)	Yes	-0.677	0.406	0.000	0.606	3.370	1.720	5.592	3.370	1.720	5.592	0.004	0.004
Framingham Probability	Yes	-10.062	3.080	0.605	4.883	2.424	0.448	4.632	2.424	0.448	4.632	0.055	0.055

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression that uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to: (1) number of siblings, (2) high risk index birth, (3) mother was full scale iq score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significant imbalance between attrited and non-attrited groups. Model selection is based on the Akaike information criterion. Variables selected for males at 35 years old are: (1) Mother was digit symbol, (2) Gestational age indicator (less than 37 weeks), (3) Substance abuse score at age 30, (4) Rule breaking score at age 30. Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis bootstrap *p* -value; Col.15: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S22(c): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)

Variable	Rev.	Mean	Treatment Effect Estimates			Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
			Effect	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
				Low	High		t-stat.	Low		High	t-stat.	Low	High
Health Care Coverage at age 30													
Health Insurance Coverage at age 30	No	1.065	0.253	0.103	0.447	1.926	0.782	3.617	1.926	0.494	3.159	0.021	0.033
Buys Health Insurance (age 30)	No	1.278	0.260	0.088	0.457	1.825	0.601	3.564	1.825	0.601	3.564	0.028	0.028
Hospital or doctor office care when sick	No	1.442	0.278	0.088	0.480	2.053	0.622	3.883	2.053	0.337	2.897	0.033	0.052
Smoking													
Never a regular smoker by age 35	No	0.496	-0.071	-0.244	0.106	-0.511	-1.797	0.782	-0.511	-1.797	0.783	0.685	0.685
Age Subject Began Smoking Regularly (Never=.)	No	22.066	2.802	0.908	4.535	2.000	0.658	3.332	2.000	-0.008	2.155	0.027	0.103
Cigarettes per Day in Past 30 Days at 21 and 30	Yes	-22.403	3.447	-0.699	7.291	1.276	-0.315	2.868	1.276	-0.586	1.762	0.154	0.237
Alcohol Use													
Did you begin drinking before 17	Yes	-1.308	0.093	-0.113	0.270	0.643	-0.777	1.901	0.643	-1.416	0.797	0.296	0.622
Do you drink and drive?	Yes	0.336	-0.108	-0.307	0.113	-0.720	-2.127	0.766	-0.720	-2.478	-0.114	0.728	0.921
Age Subject Began Drinking (Never=.)	No	15.110	-1.309	-2.965	0.067	-1.139	-2.525	0.062	-1.139	-2.525	0.062	0.889	0.889
Days Subject Drank in Past 30 Days at 21 - 30	Yes	-14.363	-1.350	-5.974	3.168	-0.390	-1.652	1.046	-0.390	-2.350	-0.099	0.634	0.919
Days Subject Drank 5+ Drinks in Past 30 Days at 21 - 30	Yes	3.866	-0.737	-2.416	1.148	-0.518	-1.780	0.876	-0.518	-2.415	-0.138	0.678	0.925
Marijuana Use													
Never Used Marijuana by age 30	No	0.286	-0.015	-0.123	0.099	-0.159	-1.421	1.111	-0.159	-1.421	1.111	0.556	0.556
Age of Onset of Marijuana Use	No	16.315	1.756	0.470	3.009	1.810	0.503	3.154	1.810	0.020	2.150	0.039	0.096
Times Subject Used Marijuana Past 30 Days (21 - 30 years old)	Yes	-2.240	3.994	-11.664	19.572	0.320	-0.949	1.607	0.320	-1.252	1.027	0.366	0.537

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression that uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to: (1) number of siblings, (2) high risk index birth, (3) mother was full scale iq score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significant imbalance between attrited and non-attrited groups. Models selection is based on the Akaike information criterion. Variables selected for males at 30 years old are: (1) Siblings IQ is below 85? (yes = 1, no = 0), (2) Mother was full scale iq score, (3) Subject is a firstborn, (4) Father last grade completed. Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis bootstrap *p* -value; Col.15: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S22(d): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Males)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Mean	Treat. Effect	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
				Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Physical Activity and Diet													
Exercise 4 or more Days per Week (age 21)	No	2.062	-0.097	-0.272	0.090	-0.715	-2.063	0.658	-0.715	-2.063	0.658	0.744	0.744
No. of Fruit Servings per Day (age 21)	No	1.097	-0.000	-0.375	0.384	-0.001	-1.401	1.419	-0.001	-1.750	0.724	0.496	0.693
Physical Develop. (Overweight)													
At Risk Overweight (CDC) at 3 mo.	Yes	-0.047	0.216	0.066	0.346	2.292	0.819	3.495	2.292	0.169	2.527	0.019	0.072
At Risk Overweight (CDC) at 6 mo.	Yes	-0.761	0.231	0.084	0.355	2.382	0.901	3.748	2.382	0.140	2.385	0.018	0.075
At Risk Overweight (CDC) at 9 mo.	Yes	-0.033	0.442	0.268	0.623	3.465	2.272	4.915	3.465	1.203	3.329	0.000	0.006
At Risk Overweight (CDC) at 12 mo.	Yes	-0.213	0.415	0.273	0.555	4.328	3.002	5.844	4.328	1.979	4.014	0.000	0.001
At Risk Overweight (CDC) at 18 mo.	Yes	0.323	0.375	0.204	0.504	3.715	2.205	4.983	3.715	1.384	3.435	0.000	0.004
At Risk Overweight (CDC) at 24 mo.	Yes	0.301	0.329	0.175	0.467	3.537	2.172	4.988	3.537	1.249	3.320	0.000	0.006
At Risk Overweight (CDC) at 36 mo.	Yes	0.147	0.093	-0.066	0.230	0.899	-0.704	2.342	0.899	-0.895	1.754	0.248	0.330
At Risk Overweight (CDC) at 48 mo.	Yes	0.872	0.115	-0.063	0.296	0.859	-0.493	2.284	0.859	-0.493	2.285	0.203	0.203
At Risk Overweight (CDC) at 60 mo.	Yes	1.358	0.171	0.017	0.329	1.441	0.139	2.952	1.441	-0.480	2.034	0.080	0.209
At Risk Overweight (CDC) at 96 mo.	Yes	0.949	0.283	0.113	0.461	2.203	0.928	3.779	2.203	0.169	2.638	0.014	0.071
Weight-for-Length Change (CDC) at 24 mo.	Yes	-3.658	0.924	0.091	1.601	1.615	0.159	2.921	1.615	0.159	2.921	0.075	0.075
Weight-for-Length Change (WHO) at 24 m.	Yes	-5.029	1.093	0.224	1.820	1.790	0.364	3.135	1.790	0.249	3.008	0.054	0.063

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. This table examines selected outcomes for males. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression that uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to: (1) number of siblings, (2) high risk index birth, (3) mother was full scale iq score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significant imbalance between attrited and non-attrited groups. Model selection is based on the Akaike information criterion. Variables selected for males at 2 years old are: (1) Father Absent (yes = 1, no = 0), (2) Other special circumstances? (yes = 1, no = 0), (3) Gestational age indicator (less than 37 weeks), (4) Siblings IQ is below 85? (yes = 1, no = 0). Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis bootstrap *p* -value; Col.15: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S23(a): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)

Variable	Rev.	Mean	Treatment Effect Estimates			Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
			Ctr.	Treat.	Conf. Inter.	test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
			Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Blood Pressure													
Diastolic Blood Pressure at age 35	Yes	-89.227	3.894	-2.820	10.683	0.734	-0.597	1.974	0.734	-0.597	1.974	0.229	0.229
Systolic Blood Pressure at age 35	Yes	-135.636	5.970	-1.922	14.103	0.921	-0.323	2.152	0.921	-0.584	1.849	0.162	0.231
Pre-Hypertension at age 35 (Sys. > 120 and Dias. > 80)	Yes	-0.727	0.227	0.033	0.428	1.481	0.207	2.968	1.481	0.207	2.968	0.066	0.066
Pre-Hypertension at age 35 (Sys. > 120 or Dias. > 80)	Yes	-0.909	0.242	0.077	0.410	1.949	0.650	3.431	1.949	0.404	2.881	0.028	0.046
Hypertension at age 35 (Sys. > 140 and Dias. > 90)	Yes	-0.318	0.096	-0.083	0.275	0.663	-0.584	1.959	0.663	-0.879	1.560	0.240	0.356
Hypertension at age 35 (Sys. > 140 or Dias. > 90)	Yes	-0.409	-0.091	-0.289	0.121	-0.563	-1.878	0.747	-0.563	-1.878	0.747	0.712	0.712
Lab Tests													
HDL Cholesterol at age 35	No	55.318	5.126	-0.745	11.008	1.126	-0.165	2.546	1.126	-0.165	2.546	0.132	0.132
Dyslipidemia at age 35 (HDL < 50)	Yes	-0.455	0.177	-0.020	0.368	1.139	-0.127	2.519	1.139	-0.403	2.145	0.125	0.189
Pre-Diabetes at age 35 (HbA1C > 5.7)	Yes	-0.364	0.011	-0.193	0.209	0.067	-1.245	1.350	0.067	-1.245	1.351	0.478	0.478
Vit. D Deficiency (age 35)	Yes	-0.727	0.005	-0.178	0.195	0.035	-1.256	1.361	0.035	-1.613	0.505	0.486	0.486

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S23(b): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Mean	Treat. Effect	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
				Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Obesity													
Overweight at age 35 (BMI ≥ 25)	Yes	-0.955	0.066	-0.043	0.176	0.770	-0.664	1.992	0.770	-0.911	1.505	0.219	0.350
Obese at age 35 (BMI ≥ 30)	Yes	-0.727	0.061	-0.131	0.250	0.406	-0.875	1.739	0.406	-0.875	1.739	0.344	0.344
Severely Obese at age 35 (BMI ≥ 35)	Yes	-0.364	0.141	-0.040	0.324	0.958	-0.276	2.309	0.958	-0.860	1.353	0.160	0.359
Waist-Hip Ratio at age 35	Yes	-0.933	0.057	0.012	0.100	1.582	0.322	3.040	1.582	0.032	2.566	0.054	0.093
Abdominal Obesity at age 35	Yes	-0.762	0.199	0.000	0.401	1.277	0.000	2.707	1.277	0.000	2.707	0.100	0.100
Multiple Risk Factors													
Hypertension and Obesity (age 35)	Yes	-0.364	0.086	-0.105	0.275	0.565	-0.691	1.863	0.565	-1.222	1.166	0.284	0.496
Hypertension and Severe Obesity (age 35)	Yes	-0.136	-0.030	-0.179	0.118	-0.260	-1.614	1.084	-0.260	-1.614	1.084	0.593	0.593
Hypertension and Dyslipidemia (age 35)	Yes	-0.182	0.015	-0.142	0.168	0.122	-1.200	1.411	0.122	-1.535	0.973	0.449	0.598
Metabolic Syndrome at age 35 (NCEP)	Yes	-0.190	0.128	-0.003	0.261	1.117	-0.038	2.212	1.117	-0.711	1.609	0.105	0.293
Framingham Probability	Yes	-1.482	0.339	-0.015	0.684	1.206	-0.057	2.401	1.206	-0.057	2.401	0.113	0.113

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S23(c): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)

Variable	Rev.	Mean	Treatment Effect Estimates			Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
			Treat.	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
			Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Health Care Coverage at age 30													
Health Insurance Coverage at age 30	No	0.857	-0.097	-0.237	0.043	-0.892	-2.229	0.399	-0.892	-2.526	-0.228	0.814	0.935
Buys Health Insurance (age 30)	No	0.357	0.043	-0.128	0.218	0.316	-0.952	1.644	0.316	-1.494	0.701	0.377	0.663
Hospital or doctor office care when sick	No	0.929	-0.129	-0.253	-0.009	-1.379	-2.668	-0.107	-1.379	-2.668	-0.107	0.921	0.921
Smoking													
Never a regular smoker by age 35	No	0.357	0.203	0.034	0.377	1.484	0.248	2.910	1.484	-0.301	1.651	0.064	0.172
Age Subject Began Smoking Regularly (Never=.)	No	17.861	-0.811	-2.200	0.509	-0.694	-1.810	0.497	-0.694	-1.810	0.497	0.776	0.776
Cigarettes per Day in Past 30 Days at 21 and 30	Yes	-10.294	-0.156	-3.980	3.490	-0.053	-1.395	1.250	-0.053	-1.632	0.410	0.514	0.764
Alcohol Use													
Did you begin drinking before 17	Yes	-0.571	0.291	0.121	0.465	2.193	0.878	3.761	2.193	0.228	2.225	0.013	0.058
Do you drink and drive?	Yes	-0.222	0.102	-0.031	0.235	0.963	-0.298	2.264	0.963	-0.693	1.488	0.165	0.293
Age Subject Began Drinking (Never=.)	No	16.685	1.042	-0.097	2.173	1.142	-0.098	2.684	1.142	-0.771	1.264	0.118	0.346
Days Subject Drank in Past 30 Days at 21 – 30	Yes	-8.732	2.772	-0.704	6.352	0.985	-0.260	2.188	0.985	-0.784	1.377	0.152	0.334
Days Subject Drank 5+ Drinks in Past 30 Days at 21 – 30	Yes	-2.321	1.061	-0.414	2.549	0.880	-0.400	2.041	0.880	-0.400	2.041	0.185	0.185
Marijuana Use													
Never Used Marijuana by age 30	No	0.250	0.150	-0.015	0.314	1.161	-0.112	2.550	1.161	-0.600	1.402	0.122	0.298
Age of Onset of Marijuana Use	No	17.190	0.676	-0.494	1.775	0.722	-0.459	2.435	0.722	-0.858	1.343	0.218	0.396
Times Subject Used Marijuana Past 30 Days (21 –30 years old)	Yes	-19.982	-1.598	-15.318	12.145	-0.152	-1.463	1.189	-0.152	-1.463	1.189	0.571	0.571

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S23(d): Unconditional Bootstrap Inference for the Abecedarian Intervention (Females)

Variable	Rev.	Mean	Treatment Effect Estimates			Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
			Ctr.	Treat.	Conf. Inter.	test	Conf. Inter.	Max	Conf. Inter.	<i>p</i> -values			
			Effect	Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Physical Activity and Diet													
Exercise 4 or more Days per Week (age 21)	No	0.071	0.249	0.113	0.384	2.388	1.146	3.735	2.388	1.146	3.735	0.010	0.010
No. of Fruit Servings per Day (age 21)	No	0.286	0.514	0.250	0.771	2.632	1.282	4.152	2.632	0.973	3.169	0.004	0.009
Physical Develop.													
At Risk Overweight (CDC) at 3 mo.	Yes	-0.192	0.002	-0.155	0.155	0.016	-1.312	1.357	0.016	-1.981	0.388	0.490	0.794
At Risk Overweight (CDC) at 6 mo.	Yes	-0.423	0.256	0.085	0.412	2.017	0.649	3.407	2.017	-0.313	1.805	0.025	0.170
At Risk Overweight (CDC) at 9 mo.	Yes	-0.360	0.217	0.039	0.388	1.447	0.257	2.711	1.447	-0.763	1.432	0.065	0.320
At Risk Overweight (CDC) at 12 mo.	Yes	-0.478	0.270	0.093	0.444	1.993	0.667	3.514	1.993	-0.298	1.875	0.027	0.165
At Risk Overweight (CDC) at 18 mo.	Yes	-0.440	0.122	-0.067	0.306	0.845	-0.455	2.189	0.845	-1.226	1.101	0.199	0.501
At Risk Overweight (CDC) at 24 mo.	Yes	-0.412	0.238	0.050	0.423	1.683	0.342	3.213	1.683	-0.577	1.608	0.054	0.249
At Risk Overweight (CDC) at 36 mo.	Yes	-0.261	0.118	-0.042	0.275	0.957	-0.340	2.290	0.957	-1.201	1.111	0.171	0.492
At Risk Overweight (CDC) at 48 mo.	Yes	-0.192	-0.217	-0.391	-0.049	-1.659	-3.168	-0.364	-1.659	-3.168	-0.364	0.952	0.952
At Risk Overweight (CDC) at 60 mo.	Yes	-0.261	-0.012	-0.185	0.153	-0.088	-1.407	1.147	-0.088	-1.931	0.623	0.542	0.725
At Risk Overweight (CDC) at 96 mo.	Yes	-0.174	-0.176	-0.349	-0.004	-1.316	-2.733	-0.029	-1.316	-3.088	-0.400	0.905	0.954
Weight-for-Length Change (CDC) at 24 mo.	Yes	-0.857	-0.062	-0.751	0.626	-0.112	-1.450	1.146	-0.112	-1.527	1.049	0.562	0.593
Weight-for-Length Change (WHO) at 24 m.	Yes	-1.129	-0.085	-0.811	0.617	-0.150	-1.492	1.081	-0.150	-1.492	1.081	0.573	0.573

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. Col.3: unconditional outcome control mean; Col.4: unconditional treatment effect, i.e. the outcome difference in means between treatment and control groups; Col.7 and Col.8 present the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the unconditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.13: one-sided single hypothesis bootstrap *p* -value; Col.14: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S24(a): Conditional IPW Bootstrap Inference for the Abecedarian Intervention (Females)

Variable	Rev.	Mean	Treatment Effect Estimates			Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
			Effect	Low	High	test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
						t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Blood Pressure													
Diastolic Blood Pressure at age 35	Yes	-84.821	1.413	-7.411	10.372	0.224	-1.354	1.559	0.224	-1.355	1.559	0.436	0.436
Systolic Blood Pressure at age 35	Yes	-123.120	2.381	-8.430	13.214	0.323	-1.391	1.690	0.323	-1.561	1.415	0.418	0.490
Pre-Hypertension at age 35 (Sys. > 120 and Dias. > 80)	Yes	-0.083	0.072	-0.220	0.298	0.404	-1.229	1.780	0.404	-1.229	1.781	0.416	0.416
Pre-Hypertension at age 35 (Sys. > 120 or Dias. > 80)	Yes	-0.698	0.241	-0.012	0.442	1.699	-0.085	3.411	1.699	-0.339	2.716	0.114	0.157
Hypertension at age 35 (Sys. > 140 and Dias. > 90)	Yes	-0.120	-0.009	-0.224	0.209	-0.056	-1.453	1.350	-0.056	-1.779	1.011	0.526	0.625
Hypertension at age 35 (Sys. > 140 or Dias. > 90)	Yes	-0.797	-0.180	-0.418	0.060	-1.043	-2.578	0.333	-1.043	-2.578	0.333	0.829	0.829
Lab Tests													
HDL Cholesterol at age 35	No	49.955	6.384	0.109	12.820	1.223	0.020	2.661	1.223	-0.317	2.304	0.097	0.162
Dyslipidemia at age 35 (HDL < 40 or HDL < 50)	Yes	-0.705	0.197	-0.049	0.449	1.099	-0.284	2.713	1.099	-0.285	2.713	0.154	0.154
Pre-Diabetes at age 35 (HbA1C > 5.7)	Yes	-1.474	0.049	-0.205	0.330	0.282	-1.216	2.027	0.282	-1.216	2.027	0.399	0.399
Vit. D Deficiency (age 35)	Yes	-1.392	0.062	-0.178	0.268	0.397	-1.169	1.784	0.397	-1.714	0.973	0.388	0.388

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression the uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to:(1) number of siblings, (2) high risk index birth, (3) mother wais full scale iq score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significance imbalance between attrited and non-attrited groups. Models selection in based of the Akaike information criterion. Variables selected for females at 35 years old are: (1) Mother wais digit symbol, (2) Gestational age indicatoe (less than 37 weeks), (3) Substance abuse score at age 30, (4) Rule breaking score at age 30. Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis bootstrap *p* -value; Col.15: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S24(b): Conditional IPW Bootstrap Inference for the Abecedarian Daycare Intervention (Females)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Mean	Treat. Effect	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
				Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Obesity													
Overweight at age 35 (BMI ≥ 25)	Yes	-0.287	0.039	-0.064	0.141	0.460	-0.855	1.689	0.460	-1.251	1.230	0.310	0.469
Obese at age 35 (BMI ≥ 30)	Yes	-0.849	-0.120	-0.319	0.107	-0.794	-2.234	0.706	-0.794	-2.234	0.706	0.749	0.749
Severely Obese at age 35 (BMI ≥ 35)	Yes	-1.207	0.155	-0.095	0.375	0.905	-0.563	2.337	0.905	-1.054	1.336	0.217	0.417
Waist-Hip Ratio at age 35	Yes	-1.105	0.048	-0.000	0.093	1.363	-0.002	3.025	1.363	-0.327	2.348	0.101	0.160
Abdominal Obesity at age 35	Yes	-2.484	0.178	-0.020	0.390	1.190	-0.132	2.723	1.190	-0.132	2.723	0.124	0.124
Multiple Risk Factors													
Hypertension and Obesity (age 35)	Yes	-0.456	-0.037	-0.269	0.198	-0.221	-1.641	1.214	-0.221	-2.055	0.477	0.580	0.781
Hypertension and Severe Obesity (age 35)	Yes	-0.508	-0.051	-0.231	0.123	-0.373	-1.735	0.954	-0.373	-1.735	0.954	0.632	0.632
Hypertension and Dyslipidemia (age 35)	Yes	-0.965	-0.033	-0.207	0.137	-0.246	-1.598	1.068	-0.246	-1.852	0.701	0.592	0.706
Metabolic Syndrome at age 35 (NCEP)	Yes	-0.626	0.063	-0.097	0.210	0.515	-0.834	1.811	0.515	-1.372	1.154	0.310	0.524
Framingham Probability	Yes	-3.654	0.255	-0.102	0.589	0.940	-0.441	2.026	0.940	-0.441	2.026	0.190	0.190

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression that uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to: (1) number of siblings, (2) high risk index birth, (3) mother was full scale iq score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significant imbalance between attrited and non-attrited groups. Model selection is based on the Akaike information criterion. Variables selected for females at 35 years old are: (1) Mother was digit symbol, (2) Gestational age indicator (less than 37 weeks), (3) Substance abuse score at age 30, (4) Rule breaking score at age 30. Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis bootstrap *p* -value; Col.15: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S24(c): Conditional IPW Bootstrap Inference for the Abecedarian Daycare Intervention (Females)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Mean	Treat. Effect	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
				Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Health Care Coverage at age 30													
Health Insurance Coverage at age 30	No	0.616	-0.095	-0.233	0.044	-0.838	-2.278	0.378	-0.838	-2.595	-0.051	0.811	0.910
Buys Health Insurance (age 30)	No	0.692	-0.001	-0.201	0.190	-0.005	-1.402	1.335	-0.005	-1.993	0.380	0.519	0.790
Hospital or doctor office care when sick	No	0.828	-0.120	-0.247	0.000	-1.117	-2.546	0.000	-1.117	-2.546	0.000	0.900	0.900
Smoking													
Never a regular smoker by age 35	No	1.136	0.170	-0.026	0.376	1.157	-0.170	2.727	1.157	-0.779	1.287	0.131	0.332
Age Subject Began Smoking Regularly (Never=.)	No	26.697	-1.779	-3.830	0.366	-1.136	-2.510	0.255	-1.136	-2.510	0.255	0.857	0.857
Cigarettes per Day in Past 30 Days at 21 and 30	Yes	2.751	0.574	-3.965	5.959	0.155	-1.093	1.739	0.155	-1.580	0.722	0.380	0.656
Alcohol Use													
Did you begin drinking before 17	Yes	0.228	0.299	0.105	0.485	2.195	0.737	3.872	2.195	0.016	2.174	0.026	0.097
Do you drink and drive?	Yes	-0.466	0.071	-0.099	0.232	0.593	-0.838	2.002	0.593	-1.251	1.024	0.303	0.535
Age Subject Began Drinking (Never=.)	No	23.780	0.981	-0.625	2.528	0.920	-0.564	2.744	0.920	-1.185	0.967	0.216	0.516
Days Subject Drank in Past 30 Days at 21 – 30	Yes	18.314	1.173	-2.189	4.934	0.415	-0.865	1.676	0.415	-1.154	1.295	0.335	0.441
Days Subject Drank 5+ Drinks in Past 30 Days at 21 – 30	Yes	8.547	0.023	-1.222	1.448	0.020	-1.225	1.219	0.020	-1.225	1.219	0.454	0.454
Marijuana Use													
Never Used Marijuana by age 30	No	0.233	0.118	-0.064	0.309	0.862	-0.460	2.385	0.862	-1.027	1.117	0.192	0.460
Age of Onset of Marijuana Use	No	18.762	0.784	-0.937	2.192	0.704	-0.738	2.785	0.704	-0.988	1.357	0.273	0.421
Times Subject Used Marijuana Past 30 Days (21 –30 years old)	Yes	43.289	-0.871	-13.972	13.696	-0.078	-1.290	1.247	-0.078	-1.290	1.247	0.512	0.512

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression that uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to: (1) number of siblings, (2) high risk index birth, (3) mother was full scale iq score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significant imbalance between attrited and non-attrited groups. Model selection is based on the Akaike information criterion. Variables selected for females at 30 years old are: (1) Siblings IQ is below 85? (yes = 1, no = 0), (2) Mother was full scale iq score, (3) Subject is a firstborn, (4) Father last grade completed. Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis bootstrap *p* -value; Col.15: stepdown one-sided multiple-hypothesis bootstrap *p* -value. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S24(d): Conditional IPW Bootstrap Inference for the Abecedarian Daycare Intervention (Females)

Variable	Rev.	Treatment Effect Estimates				Treat. Effect (t-statistics)			Max. of Treat. Eff. t-statistics			Bootstrap Inference	
		Mean	Treat. Effect	Conf. Inter.		test	Conf. Inter.		Max	Conf. Inter.		<i>p</i> -values	
				Low	High	t-stat.	Low	High	t-stat.	Low	High	Single	S.D.
Physical Activity and Diet													
Exercise 4 or more Days per Week (age 21)	No	0.678	0.279	0.125	0.419	2.488	1.167	3.914	2.488	0.740	3.005	0.007	0.023
No. of Fruit Servings per Day (age 21)	No	-0.602	0.546	0.229	0.849	2.465	1.004	4.203	2.465	1.004	4.203	0.017	0.017
Physical Develop. (Overweight)													
At Risk Overweight (CDC) at 3 mo.	Yes	0.303	-0.046	-0.202	0.118	-0.361	-1.675	0.970	-0.361	-2.426	0.075	0.636	0.887
At Risk Overweight (CDC) at 6 mo.	Yes	-0.119	0.251	0.073	0.446	1.831	0.528	3.517	1.831	-0.541	1.569	0.041	0.240
At Risk Overweight (CDC) at 9 mo.	Yes	1.146	0.219	0.032	0.400	1.418	0.224	2.765	1.418	-0.906	1.225	0.070	0.375
At Risk Overweight (CDC) at 12 mo.	Yes	1.806	0.122	-0.052	0.310	0.885	-0.388	2.361	0.885	-1.344	0.824	0.183	0.590
At Risk Overweight (CDC) at 18 mo.	Yes	0.997	0.057	-0.135	0.260	0.373	-0.896	1.818	0.373	-1.803	0.367	0.343	0.785
At Risk Overweight (CDC) at 24 mo.	Yes	1.119	0.189	0.019	0.396	1.351	0.144	2.902	1.351	-0.937	1.249	0.080	0.385
At Risk Overweight (CDC) at 36 mo.	Yes	1.708	0.027	-0.091	0.149	0.231	-0.784	1.344	0.231	-1.909	0.272	0.394	0.819
At Risk Overweight (CDC) at 48 mo.	Yes	0.027	-0.227	-0.424	-0.019	-1.559	-3.132	-0.137	-1.559	-3.465	-0.537	0.922	0.962
At Risk Overweight (CDC) at 60 mo.	Yes	-0.280	0.010	-0.169	0.210	0.067	-1.205	1.471	0.067	-2.051	0.284	0.450	0.823
At Risk Overweight (CDC) at 96 mo.	Yes	-0.421	-0.244	-0.434	-0.048	-1.737	-3.328	-0.349	-1.737	-3.328	-0.349	0.939	0.939
Weight-for-Length Change (CDC) at 24 mo.	Yes	3.111	-0.282	-1.040	0.527	-0.488	-1.894	0.932	-0.488	-1.894	0.932	0.639	0.639
Weight-for-Length Change (WHO) at 24 m.	Yes	2.364	-0.264	-1.006	0.531	-0.454	-1.836	0.895	-0.454	-1.950	0.829	0.636	0.661

Notes: This table presents the Abecedarian Treatment Effects estimates of the Day-care treatment center. We examine selected outcomes for females. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable were reversed or not. By reversed we mean multiplied by -1. The estimates of the remaining columns are based on a weighted linear regression that uses inverse probability of attrition as weighting. Covariates of the linear regression are the treatment effect indicator in addition to: (1) number of siblings, (2) high risk index birth, (3) mother's full scale IQ score, (4) Other special circumstances? (yes = 1, no = 0). The selection of covariates for IPW is age and gender specific. We select the most predictive model examining all combinations of covariates that present statistically significant imbalance between attrited and non-attrited groups. Model selection is based on the Akaike information criterion. Variables selected for females at 2 years old are: (1) Father Absent (yes = 1, no = 0), (2) Other special circumstances? (yes = 1, no = 0), (3) Gestational age indicator (less than 37 weeks), (4) Siblings IQ is below 85? (yes = 1, no = 0). Col.3: conditional outcome control mean; Col.4: conditional treatment effect; Col.7 and Col.8 present the bootstrap confidence interval of the conditional treatment effect estimate for 80% confidence level; Col.9 and Col.10 and Col.11 present the t-statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region; Col.12 and Col.13 and Col.14 present the stepdown test statistic associated with the conditional treatment effect estimate and the associated bootstrap confidence region adjusted for multiple hypothesis; Col.14: one-sided single hypothesis. The multiple hypothesis testing is applied to blocks of outcomes. Blocks of variables that are tested jointly using the stepdown algorithm are delineated by horizontal lines.

Table S25(a): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Males)

Variable	Rev.	Age	Sample Sizes		Cont. Mean	Diff. in Means	Bootstrap Confidence Interval (at 80%)		Conditional Permutation Inference		Conditional Bootstrap Inference	
			# Cont.	# Treat.			Low	High	Single p-value	Step-Down p-value	Single p-value	Step-Down p-value
Blood Pressure												
Diastolic Blood Pressure (mmHg)	Yes	35	9	19	-92,000	13,474	3.946	23.561	0.024	0.024	0.030	0.030
Systolic Blood Pressure (mmHg)	Yes	35	9	19	-143,333	17,544	4.200	31.847	0.018	0.029	0.038	0.050
Pre-Hypertension (systolic bp \geq 120 & diastolic bp \geq 80)	Yes	35	9	19	-0.667	0.246	-0.013	0.504	0.119	0.172	0.096	0.159
Pre-Hypertension (systolic bp \geq 120 or diastolic bp \geq 80)	Yes	35	9	19	-0.778	0.094	-0.143	0.320	0.235	0.235	0.254	0.254
Hypertension (systolic bp \geq 140 & diastolic bp \geq 90)	Yes	35	9	19	-0.444	0.339	0.097	0.580	0.010	0.018	0.021	0.022
Hypertension (systolic bp \geq 140 or diastolic bp \geq 90)	Yes	35	9	19	-0.556	0.345	0.083	0.590	0.038	0.038	0.039	0.039
Lab Tests												
High-Density Lipoprotein (HDL) Cholesterol (mg/dL)	No	35	12	19	42,000	11,211	5.698	16.733	0.066	0.110	0.004	0.025
Dyslipidemia (HDL < 40 mg/dL for males, HDL < 50 mg/dL for females)	Yes	35	12	19	-0.417	0.311	0.098	0.533	0.179	0.179	0.111	0.111
Pre-Diabetes (HbA1C \geq 5.7%)	Yes	35	12	19	-0.583	0.110	-0.142	0.353	0.426	0.426	0.370	0.370
Vitamin D Deficiency (< 20 ng/mL)	Yes	35	12	19	-0.750	0.382	0.166	0.592	0.021	0.021	0.017	0.017
Obesity												
Overweight (BMI \geq 25)	Yes	35	8	18	-0.750	0.028	-0.222	0.285	0.239	0.239	0.226	0.226
Obese (BMI \geq 30)	Yes	35	8	18	-0.625	0.069	-0.206	0.349	0.233	0.345	0.197	0.269
Severely Obese (BMI \geq 35)	Yes	35	8	18	-0.375	0.264	0.000	0.526	0.115	0.232	0.114	0.159
Waist-Hip Ratio (WHR)	Yes	35	8	17	-0.962	0.025	-0.017	0.067	0.293	0.293	0.217	0.217
Abdominal Obesity (WHR > 0.9 for males, WHR > 0.85 for females)	Yes	35	8	17	-0.875	0.228	-0.017	0.067	0.137	0.218	0.076	0.154
Multiple Risk Factors												
Obesity & Hypertension	Yes	35	8	18	-0.500	0.389	0.000	0.438	0.016	0.016	0.010	0.013
Severe Obesity & Hypertension	Yes	35	8	18	-0.375	0.375	0.123	0.655	0.005	0.012	0.001	0.003
Hypertension & Dyslipidemia	Yes	35	9	18	-0.333	0.333	0.143	0.600	0.006	0.012	0.000	0.006
Metabolic Syndrome (NCEP Definition)	Yes	35	8	17	-0.250	0.250	0.123	0.655	0.007	0.014	0.004	0.004
Framingham Risk Score (35)	Yes	35	9	18	-7,043	2,154	0.143	0.600	0.038	0.038	0.055	0.055

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment using the permutation and bootstrap inference methods. This table examines selected outcomes. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable was reversed or not. By "reversed", we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; Col.3: age of the participant when the data was collected; Col.4: control sample size; Col.5: treatment sample size; Col.6: control arithmetic mean; Col.7: unconditional difference in means across treatment and control groups; Col.8 and Col.9: the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.10: presents the one-sided single hypothesis block permutation -value associated with the IPW treatment effect estimate Col.11: presents the multiple hypothesis testing (step-down) -values associated with the IPW inference; Col. 12 and Col. 13 : the p-values obtained from the bootstrap inference method corresponding to Col.10 and Col.11, respectively.

Table S25(b): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Males, cont'd)

Variable	Rev.	Age	Sample Sizes		Cont. Mean	Diff. in Means	Bootstrap Confidence Interval (at 80%)		Conditional Permutation Inference		Conditional Bootstrap Inference	
			# Cont.	# Treat			Low	High	Single p-value	Step-Down p-value	Single p-value	Step-Down p-value
Health Care												
Health Insurance Coverage at age 30	No	30	21	27	0.476	0.228	0.048	0.407	0.039	0.039	0.021	0.033
Buys Health Insurance at age 30	No	30	21	27	0.333	0.296	0.119	0.474	0.035	0.080	0.028	0.028
Hospital or Doctor Office Care When Sick at age 30	No	30	21	27	0.524	0.291	0.122	0.466	0.037	0.068	0.033	0.052
Behavioral Risk Factors: Smoking												
Never a Regular Smoker by age 30	No	21-30	23	28	0.391	-0.034	-0.249	0.108	0.571	0.571	0.685	0.685
Age of Onset of Regular Smoking	No	21-30	14	18	16,893	2,829	1.235	4.467	0.051	0.143	0.027	0.103
Cigarettes Smoked per Day Past 30 Days at ages 21&30	Yes	21-30	12	19	-11,167	2,482	-1.542	6.800	0.199	0.344	0.154	0.237
Behavioral Risk Factors: Alcohol Use												
Early Onset Drinker (before age 17)	Yes	21	23	26	-0.609	0.070	-0.109	0.252	0.336	0.762	0.296	0.622
Drank Driving	Yes	21	23	26	-0.391	-0.147	-0.338	0.042	0.663	0.877	0.728	0.921
Age of Onset of Alcohol Use	No	21-30	22	28	17,045	-1,135	-2.626	0.302	0.840	0.840	0.889	0.889
Days At Least 1 Drink Past 30 Days at ages 21&30	Yes	21-30	23	28	-9,348	-2,634	-7.008	1.348	0.498	0.860	0.634	0.919
Days 5+ Drinks Past 30 Days at ages 21&30	Yes	21-30	23	28	-2,457	-0.972	-2.828	0.820	0.585	0.900	0.678	0.925
Behavioral Risk Factors: Marijuana Use												
Never Used Marijuana by age 30	No	21-30	23	28	0.130	-0.023	-0.142	0.093	0.476	0.476	0.556	0.556
Age of Onset of Marijuana Use	No	21-30	20	25	15,475	2,025	0.837	3.262	0.052	0.136	0.039	0.096
Times Used Marijuana Past 30 Days at ages 21&30	Yes	21-30	23	28	-49,565	1,030	-14.539	16.917	0.357	0.527	0.366	0.537
Behavioral Risk Factors: Physical Activity and Diet												
Exercise 4 or more Days per Week at age 21	No	21	23	26	0.391	-0.084	-0.259	0.101	0.858	0.858	0.744	0.744
No. of Fruit Servings per Day at age 21	No	21	23	26	0.826	0.020	-0.312	0.369	0.524	0.745	0.496	0.693
Physical Development												
At Risk Overweight (CDC) at 3 months	Yes	0	22	27	-0.227	0.190	0.063	0.316	0.026	0.121	0.019	0.072
At Risk Overweight (CDC) at 6 months	Yes	1	20	25	-0.250	0.170	0.033	0.316	0.074	0.182	0.018	0.075
At Risk Overweight (CDC) at 9 months	Yes	1	17	16	-0.412	0.412	0.263	0.565	0.004	0.023	0.000	0.006
At Risk Overweight (CDC) at 12 months	Yes	1	21	27	-0.429	0.429	0.292	0.571	0.001	0.009	0.000	0.001
At Risk Overweight (CDC) at 18 months	Yes	2	18	26	-0.389	0.389	0.235	0.533	0.000	0.004	0.000	0.004
Overweight (CDC) at 24 months	Yes	2	15	27	-0.333	0.333	0.171	0.500	0.001	0.011	0.000	0.006
Overweight (CDC) at 36 months	Yes	3	19	25	-0.158	0.078	-0.050	0.209	0.194	0.194	0.248	0.330
Overweight (CDC) at 48 months	Yes	4	20	24	-0.300	0.133	-0.034	0.298	0.150	0.235	0.203	0.203
Overweight (CDC) at 60 months	Yes	5	20	24	-0.300	0.175	0.011	0.333	0.058	0.179	0.080	0.209
Overweight (CDC) at 96 months	Yes	8	19	25	-0.421	0.301	0.132	0.471	0.030	0.117	0.014	0.071
Weight-for-Length Change 0-24 months (CDC)	Yes	2	15	24	-0.858	0.963	0.210	1.717	0.058	0.058	0.075	0.075
Weight-for-Length Change 0-24 months (WHO)	Yes	2	15	24	-1,265	1,100	0.000	2.000	0.049	0.057	0.054	0.063

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment using the permutation and bootstrap inference methods. This table examines selected outcomes. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable was reversed or not. By "reversed", we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; Col.3: age of the participant when the data was collected; Col.4: control sample size; Col.5: treatment sample size; Col.6: control arithmetic mean; Col.7: unconditional difference in means across treatment and control groups; Col.8 and Col.9: the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.10: presents the one-sided single hypothesis block permutation -value associated with the IPW treatment effect estimate Col.11: presents the multiple hypothesis testing (step-down) -values associated with the IPW inference; Col. 12 and Col. 13 : the p-values obtained from the bootstrap inference method corresponding to Col.10 and Col.11, respectively.

Table S26(a): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Females)

Variable	Rev.	Age	Sample Sizes		Cont. Mean	Diff. in Means	Bootstrap Confidence Interval (at 80%)		Conditional Permutation Inference		Conditional Bootstrap Inference	
			# Cont.	# Treat.			Low	High	Single <i>p</i> -value	Step-Dow n <i>p</i> -value	Single <i>p</i> -value	Step-Dow n <i>p</i> -value
Blood Pressure												
Diastolic Blood Pressure (mmHg)	Yes	35	22	18	-89,227	3,894	-2.820	10.683	0.446	0.446	0.436	0.436
Systolic Blood Pressure (mmHg)	Yes	35	22	18	-135,636	5,970	-1.922	14.103	0.300	0.380	0.418	0.490
Pre-Hypertension (systolic bp \geq 120 & diastolic bp \geq 80)	Yes	35	22	18	-0.727	0.227	0.033	0.428	0.222	0.222	0.416	0.416
Pre-Hypertension (systolic bp \geq 120 or diastolic bp \geq 80)	Yes	35	22	18	-0.909	0.242	0.077	0.410	0.042	0.069	0.114	0.157
Hypertension (systolic bp \geq 140 & diastolic bp \geq 90)	Yes	35	22	18	-0.318	0.096	-0.083	0.275	0.375	0.499	0.526	0.625
Hypertension (systolic bp \geq 140 or diastolic bp \geq 90)	Yes	35	22	18	-0.409	-0.091	-0.289	0.121	0.721	0.721	0.829	0.829
Lab Tests												
High-Density Lipoprotein (HDL) Cholesterol (mg/dL)	No	35	22	18	55,318	5,126	-0.745	11.008	0.143	0.143	0.097	0.162
Dyslipidemia (HDL < 40 mg/dL for males, HDL < 50 mg/dL for females)	Yes	35	22	18	-0.455	0.177	-0.020	0.368	0.099	0.147	0.154	0.154
Pre-Diabetes (HbA1C \geq 5.7%)	Yes	35	22	17	-0.364	0.011	-0.193	0.209	0.580	0.580	0.399	0.399
Vitamin D Deficiency (< 20 ng/mL)	Yes	35	22	18	-0.727	0.005	-0.178	0.195	0.303	0.303	0.388	0.388
Obesity												
Overweight (BMI \geq 25)	Yes	35	22	18	-0.955	0.066	-0.043	0.176	0.482	0.690	0.310	0.469
Obese (BMI \geq 30)	Yes	35	22	18	-0.727	0.061	-0.131	0.250	0.790	0.790	0.749	0.749
Severely Obese (BMI \geq 35)	Yes	35	22	18	-0.364	0.141	-0.040	0.324	0.354	0.653	0.217	0.417
Waist-Hip Ratio (WHR)	Yes	35	21	16	-0.933	0.057	0.012	0.100	0.063	0.101	0.101	0.160
Abdominal Obesity (WHR > 0.9 for males, WHR > 0.85 for females)	Yes	35	21	16	-0.762	0.199	0.000	0.401	0.080	0.080	0.124	0.124
Multiple Risk Factors												
Obesity & Hypertension	Yes	35	35	22	18	-0.364	0.086	-0.105	0.275	0.501	0.641	0.580
Severe Obesity & Hypertension	Yes	35	35	22	18	-0.136	-0.030	-0.179	0.118	0.696	0.696	0.632
Hypertension & Dyslipidemia	Yes	35	35	22	18	-0.182	0.015	-0.142	0.168	0.486	0.725	0.592
Metabolic Syndrome (NCEP Definition)	Yes	35	35	21	16	-0.190	0.128	-0.003	0.261	0.184	0.393	0.310
Framingham Risk Score (35)	Yes	35	35	22	18	-1,482	0.339	-0.015	0.684	0.070	0.070	0.190

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment using the permutation and bootstrap inference methods. This table examines selected outcomes. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable was reversed or not. By "reversed", we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; Col.3: age of the participant when the data was collected; Col.4: control sample size; Col.5: treatment sample size; Col.6: control arithmetic mean; Col.7: unconditional difference in means across treatment and control groups; Col.8 and Col.9: the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.10: presents the one-sided single hypothesis block permutation *p*-value associated with the IPW treatment effect estimate Col.11: presents the the multiple hypothesis testing (step-down) *p*-values associated with the IPW inference; Col. 12 and Col. 13 : the *p*-values obtained from the bootstrap inference method corresponding to Col.10 and Col.11, respectively.

Table S26(b): Comparison of Bootstrap and Permutation Inferences for the Abecedarian Daycare Intervention (Females, cont'd)

Variable	Rev.	Age	Sample Sizes		Cont. Mean	Diff. in Means	Bootstrap Confidence Interval (at 80%)		Conditional Permutation Inference		Conditional Bootstrap Inference	
			# Cont.	# Treat			Low	High	Single p-value	Step-Down p-value	Single p-value	Step-Down p-value
Health Care												
Health Insurance Coverage at age 30	No	30	28	25	0.857	-0.097	-0.237	0.043	0.943	0.943	0.811	0.910
Buys Health Insurance at age 30	No	30	28	25	0.357	0.043	-0.128	0.218	0.511	0.810	0.519	0.790
Hospital or Doctor Office Care When Sick at age 30	No	30	28	25	0.929	-0.129	-0.253	-0.009	0.875	0.964	0.900	0.900
Behavioral Risk Factors: Smoking												
Never a Regular Smoker by age 30	No	21-30	28	25	0.357	0.243	0.034	0.377	0.071	0.195	0.131	0.332
Age of Onset of Regular Smoking	No	21-30	18	10	17,861	-0.811	-2.200	0.509	0.848	0.848	0.857	0.857
Cigarettes Smoked per Day Past 30 Days at ages 21&30	Yes	21-30	17	10	-10,294	-0.156	-3.980	3.490	0.567	0.807	0.380	0.656
Behavioral Risk Factors: Alcohol Use												
Early Onset Drinker (before age 17)	Yes	21	28	25	-0.571	0.291	0.121	0.465	0.011	0.046	0.026	0.097
Drank Driving	Yes	21	27	25	-0.222	0.102	-0.031	0.235	0.172	0.358	0.303	0.535
Age of Onset of Alcohol Use	No	21-30	27	22	16,685	1.042	-0.097	2.173	0.099	0.292	0.216	0.516
Days At Least 1 Drink Past 30 Days at ages 21&30	Yes	21-30	28	25	-8,732	2,772	-0.704	6.352	0.211	0.292	0.335	0.441
Days 5+ Drinks Past 30 Days at ages 21&30	Yes	21-30	28	25	-2,321	1,061	-0.414	2.549	0.366	0.366	0.454	0.454
Behavioral Risk Factors: Marijuana Use												
Never Used Marijuana by age 30	No	21-30	28	25	0.250	0.150	-0.015	0.314	0.169	0.402	0.192	0.460
Age of Onset of Marijuana Use	No	21-30	21	15	17,190	0.676	-0.494	1.775	0.274	0.471	0.273	0.421
Times Used Marijuana Past 30 Days at ages 21&30	Yes	21-30	28	25	-19,982	-1,598	-15.318	12.145	0.417	0.417	0.512	0.512
Behavioral Risk Factors: Physical Activity and Diet												
Exercise 4 or more Days per Week at age 21	No	21	28	25	0.071	0.249	0.113	0.384	0.004	0.008	0.007	0.023
No. of Fruit Servings per Day at age 21	No	21	28	25	0.286	0.514	0.250	0.771	0.005	0.005	0.017	0.017
Physical Development												
At Risk Overweight (CDC) at 3 months	Yes	0	26	21	-0.192	0.002	-0.155	0.155	0.418	0.757	0.636	0.887
At Risk Overweight (CDC) at 6 months	Yes	1	26	24	-0.423	0.256	0.085	0.412	0.040	0.237	0.041	0.240
At Risk Overweight (CDC) at 9 months	Yes	1	25	14	-0.360	0.217	0.039	0.388	0.169	0.548	0.070	0.375
At Risk Overweight (CDC) at 12 months	Yes	1	23	24	-0.478	0.270	0.093	0.444	0.055	0.276	0.183	0.590
At Risk Overweight (CDC) at 18 months	Yes	2	25	22	-0.440	0.122	-0.067	0.306	0.311	0.669	0.343	0.785
Overweight (CDC) at 24 months	Yes	2	17	23	-0.412	0.238	0.050	0.423	0.143	0.517	0.080	0.385
Overweight (CDC) at 36 months	Yes	3	23	21	-0.261	0.118	-0.042	0.275	0.202	0.556	0.394	0.819
Overweight (CDC) at 48 months	Yes	4	26	22	-0.192	-0.217	-0.391	-0.049	0.944	0.944	0.922	0.962
Overweight (CDC) at 60 months	Yes	5	23	22	-0.261	-0.012	-0.185	0.153	0.554	0.781	0.450	0.823
Overweight (CDC) at 96 months	Yes	8	23	20	-0.174	-0.176	-0.349	-0.004	0.943	0.985	0.939	0.939
Weight-for-Length Change 0-24 months (CDC)	Yes	2	15	21	-0.857	-0.062	-0.751	0.626	0.658	0.688	0.639	0.639
Weight-for-Length Change 0-24 months (WHO)	Yes	2	15	21	-1,129	-0.085	-0.811	0.617	0.660	0.660	0.636	0.661

Notes: This table shows the full results of the small-sample inference for Abecedarian Treatment Effects of Day-care treatment using the permutation and bootstrap inference methods. This table examines selected outcomes. We list here the information of each column. Col.1: variable of interest; Col.2: shows if the variable was reversed or not. By "reversed", we mean multiplied by -1 in order to scale the effect so that a higher score is associated with a favorable outcome; Col.3: age of the participant when the data was collected; Col.4: control sample size; Col.5: treatment sample size; Col.6: control arithmetic mean; Col.7: unconditional difference in means across treatment and control groups; Col.8 and Col.9: the bootstrap confidence interval of the unconditional treatment effect estimate for 80% confidence level; Col.10: presents the one-sided single hypothesis block permutation -value associated with the IPW treatment effect estimate Col.11: presents the the multiple hypothesis testing (step-down) -values associated with the IPW inference; Col. 12 and Col. 13 : the p-values obtained from the bootstrap inference method corresponding to Col.10 and Col.11, respectively.