THE EFFECTS OF TWO INFLUENTIAL EARLY CHILDHOOD INTERVENTIONS ON HEALTH AND HEALTHY BEHAVIOUR*

Gabriella Conti, James J. Heckman and Rodrigo Pinto

This article examines the long-term impacts on health and healthy behaviour of two of the oldest and most widely cited US early childhood interventions evaluated by the method of randomisation with long-term follow-up: the Perry Preschool Project (PPP) and the Carolina Abecedarian Project (ABC). There are pronounced gender effects strongly favouring boys, although there are also effects for girls. Dynamic mediation analyses show a significant role played by improved childhood traits, above and beyond the effects of experimentally enhanced adult socioeconomic status. These results show the potential of early life interventions for promoting health.

Discussions of ways to control the soaring costs of the health care system in the US and elsewhere largely focus on the provision of health care (Emanuel, 2012; Jamison *et al.*, 2013). However, treatment of disease is only part of the story. Prevention has a substantial role to play. Most medical care costs in developed countries like the US arise from a minority of individuals with multiple chronic conditions, like cardiovascular and metabolic diseases, and cancer (Cohen and Yu, 2012).[1,2] Such conditions are the main causes of premature death and managing them effectively requires that patients make lifestyle changes by adhering to healthy behaviour (Mokdad *et al.*, 2004; Ford *et al.*, 2012; Kontis *et al.*, 2014). While prevention holds the key for lifelong health, changing behaviour in adulthood is challenging (Marteau *et al.*, 2012).[3]

* Corresponding author: James J. Heckman, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA. Email: jjh@uchicago.edu.

[1] In the United States, in 2008, 1% of the population accounted for 20% of total health care expenditures. These are older patients with cancer, diabetes, heart disease and other multiple chronic conditions. In contrast, the bottom half of the expenditure distribution accounted for 3.1% of spending.

[2] The United Nations in 2011 has set a goal of reducing the probability of premature mortality due to these diseases by 25% by the year 2025.

[3] One potentially promising approach uses insights from behavioural economics to design effective programmes implemented by employers, insurers, and health care providers, to increase patient engagement and to encourage individuals to take better care of themselves (Loewenstein *et al.*, 2007, 2013). These chronic conditions can indeed be prevented, or, at least, their onset can be substantially delayed (Sherwin *et al.*, 2004; Ezzati and Riboli, 2012).

A substantial body of evidence shows that adult illnesses are more prevalent and more problematic among those who have experienced adverse early life conditions (Danese *et al.*, 2007; Galobardes *et al.*, 2008). At present, the exact pathways through which early life experiences translate into health over the life cycle are not well known, although there is increasing understanding of the role that might be played by biological embedding of social and economic adversity. Many recent papers discuss the effect of early life experiences on adult health. See, for example, Hertzman (1999); Knudsen *et al.* (2006); Heijmans *et al.* (2008); Gluckman *et al.* (2009); Entringer *et al.* (2012); Garner *et al.* (2012). The evidence on the social determinants of health (Marmot and Wilkinson, 2005) suggests that a strategy of prevention rather than later life treatment may be more effective. Such an approach recognises the dynamic nature of health capital formation and views policies that shape early life environments as effective tools for promoting health (Conti and Heckman, 2014). Following this path, a recent interdisciplinary literature points to the role that might be played by early childhood interventions targeted to disadvantaged children in promoting adult health (Di Cesare *et al.*, 2013; Black and Hurley, 2014; Campbell *et al.*, 2014).

This article contributes to this literature by examining the effects on health and healthy behaviour of the two most influential, high-quality, US-based early childhood interventions – the Perry Preschool Project (PPP) and the Abecedarian Project (ABC). Both interventions used the method of randomisation to assign enriched environments to disadvantaged children. Participants are followed into adulthood. PPP was conducted in Ypsilanti, Michigan, starting in 1962; ABC in Chapel Hill, North Carolina, starting in 1972. PPP provided preschool education at ages 3–5 and home-based parenting guidance. ABC also included a health care and a nutritional component, and lasted from birth until age 8.[4] Data from PPP and ABC enable analysts to learn about the health benefits of early life interventions for disadvantaged populations. Since children are generally in good health and reliable early life biomarkers predictive of later disease have yet to be discovered, it is challenging to demonstrate health effects of early life interventions in the absence of long-term follow-ups.

The PPP data have rich information on healthy behaviour but not objective measures of health. ABC has a survey of health at age 34 in addition to measures of healthy behaviour. For both studies, we perform analyses by gender and find substantial differences in the effects of treatment between males and females. We present evidence that both the Perry and the Abecedarian interventions have statistically and substantively significant effects on the health and healthy behaviour of their participants. The specific outcomes affected vary across studies, although for both interventions, treatment effects are much stronger and more precisely determined for males. The Perry male participants have significantly fewer behavioural risk factors (especially smoking) by the time they have reached age 40, while the Abecedarian male participants are in better physical health by their mid-30s.

---

[4] The Abecedarian Project had a second-stage intervention at ages 5–8 via another randomised experimental design. Campbell *et al.* (2008) show that the early educational intervention had far stronger effects than the school-age treatment on the majority of the outcomes studied. Campbell *et al.* (2014) also show that the second-stage intervention had no effects on health. Hence, in this article, we only analyse the first-stage intervention.

We use robust statistical methods and apply the frameworks developed in Heckman *et al.* (2010) and Campbell *et al.* (2014) to account systematically for small sample sizes of the experiments, the effects of multiple hypothesis testing and non-random panel attrition to analyse these studies. We adjust for departures from randomisation protocols when appropriate. We show that accounting for multiple hypotheses affects inference from these studies, but the effects of accounting for small sample sizes are minor in the ABC sample but of greater consequence in the Perry study.

Rather than using arbitrarily constructed aggregates of health indicators and other outcomes as employed in previous analyses of these experiments, we use more interpretable disaggregated measures. We examine the mechanisms through which treatment effects arise using dynamic mediation analyses. We use as mediators both early child developmental traits and adult socioeconomic outcomes. We document the important role played by enhancements in childhood traits, above and beyond educational attainment and adult socioeconomic status, as mechanisms producing treatment effects.

We address the challenges that analysts face when comparing results across experiments. The baseline characteristics of the populations treated differ. Treatments vary. Follow-up periods and questions asked are not strictly comparable. Many treatment effects across programmes are not comparable because different outcomes are measured, different survey instruments are used, and different ages are sampled. Where outcome measures are comparable, estimated treatment effects are stronger for ABC males compared to PPP males. The imprecise estimates for women found in each programme translate into imprecise estimates of differences in female programme effects. Our analysis suggests that simple comparisons of treatment effects across programmes as featured in commonly reported meta-analyses (Karoly *et al.*, 2005; Camilli *et al.*, 2010) are potentially very misleading guides to policy.

This article proceeds as follows. Section 1 describes the ABC and PPP interventions. Section 2 discusses the statistical challenges addressed in this article and presents our econometric procedures. Section 3 presents and discusses our estimates of treatment effects and the results of our mediation analyses. Section 4 concludes.

1. The ABC and PPP Interventions

Both the ABC and the PPP interventions were centre-based small-scale programmes designed to enrich the early environments of disadvantaged children. The main characteristics of both interventions are displayed in Table 1. The Perry Preschool Project (PPP) took place in the mid-1960s in the district of the Perry Elementary School, a public school in Ypsilanti, Michigan (a small city near Detroit). The Carolina Abecedarian Project (ABC) took place one decade later at the Frank Porter Graham Child Development Institute at the University of North Carolina's Chapel Hill campus. Eligibility was based on weighted scales which included multiple indicators of socioeconomic disadvantage, although the specific items and weights differed across the two interventions.[5] ABC enrolled children soon after birth[6] until

---

[5] The specific ABC and PPP items and the PPP weights are reported in Table 1; the weights used for the ABC scale are reported in Table 1 of Ramey *et al.* (2000).
[6] The average age at entry for the treated was 8.8 weeks, and it ranged between 6 and 21 weeks.

Table 1

ABC and PPP: Main Characteristics and Eligibility Criteria

Abecedarian	Perry
Main characteristics	
Location: Chapel Hill, NC	Location: Ypsilanti, MI
Racial composition: 98% African American	Racial composition: All African American
Age of child: 0–5	Age of child: 3–5
Sample size: 111 (57T, 54C)	Sample size: 123 (58T, 65C)
Intervention year: 1972–83	Intervention year: 1962–7
Follow-up: up to mid-30s (2010–2)	Follow-up: up to early 40s (2000–2)
Intensity: 40 hours/week (8 hours/day for 5 days/week) for 50 weeks/year	Intensity: 12.5–15 hours/week (2.5–3 hours/day for 5 days/week) for 30 weeks/year (mid-Oct. through May) + 1.5 hours/week of home visits + 1 monthly parent group meeting
Number of years: 5 years at ages 0 to age 5	Number of years: 2 years at ages 3 to age 5 for Waves 1–4; 1 year for Wave 0
Cost per child/year: 12,955 (2010$)*	Cost per child/year: 9,604 (2010$)
Eligibility criteria	
Requirement: No apparent biological conditions	Requirement: Child IQ< 85 ('educably mentally retarded')
Weighted scale: High Risk Index:[†]	Weighted scale: cultural deprivation scale:[‡] parents' average years of schooling at entry/2 + father's occupational status at entry *2 + 2*(rooms/persons in home at entry)
(1) Mother's educational level (last grade completed)	
(2) Father's educational level (last grade completed)	
(3) Family income (dollars per year)	
(4) Father absent for reasons other than health or death	
(5) Absence of maternal relatives in local area	
(6) Siblings of school age one or more grades behind age-appropriate level or with equivalently low scores on school-administered achievement tests	
(7) Payments received from welfare agencies in past 3 years	
(8) Record of father's work indicates unstable or unskilled semi-skilled labour	
(9) Mother's or father's IQ ≤ 90	
(10) Sibling's IQ ≤90	
(11) Relevant social agencies in the community indicate the family is in need of assistance	
(12) One or more members of the family has sought counselling or professional help the past 3 years	
(13) Special circumstances not included in any of the above likely contributors to cultural or social disadvantage	

Notes. *This figure is inclusive of the health care costs (the figure reported in Barnett and Masse (2007) is not). Estimated from cost–benefit analysis conducted on both PPP and ABC projects. †See Ramey and Smith (1977). ‡See Weikart *et al.* (1978).

5 years of age[7] for a very intensive 6.5–10 hours per day programme. PPP enrolled children at 3 years of age

---

[7] As previously mentioned, the intervention consisted of a two-stage treatment: a preschool stage (0–5) and a school-age stage (5–8). In this article, we only study the effects of the preschool treatment, both for comparability with PPP, and because previous work has reported negligible or no effects from the second-stage treatment (see Campbell *et al.*, 2014).

Table 2

Descriptive Statistics of ABC and PPP Pre-programme Variables

	IQ at 3 years		Birth weight		Mother's age		Father's age	
	ABC	PPP	ABC	PPP	ABC	PPP	ABC	PPP
Mean	92.65	79.02	3.19	3.10	19.78	25.56	23.21	32.81
SD	15.95	6.44	0.61	0.47	4.77	6.53	5.84	6.88
Skewness	0.04	−0.76	−0.59	−0.05	2.16	0.52	1.29	0.52

	Mother's education		Father's education		Number of siblings	
	ABC	PPP	ABC	PPP	ABC	PPP
Mean	10.17	9.42	10.89	8.60	0.64	4.28
SD	1.84	2.20	1.78	2.40	1.09	2.59
Skewness	−0.28	−0.78	−0.38	−0.32	2.15	0.90

	Mother's working status		Father's working status		Father present	
	ABC	PPP	ABC	PPP	ABC	PPP
Mean	0.36	0.20	0.73	0.14	0.29	0.53
SD	0.48	0.40	0.45	0.35	0.45	0.50
Skewness	0.58	1.47	−1.03	2.09	0.94	−0.11

Notes. This Table provides some descriptive statistics of the ten pre-programme variables which were collected in both the Abecedarian and Perry interventions: the Stanford-Binet IQ score at 3 years of age (we only use data from the control group for the ABC intervention, since it started at birth); weight at birth in kilograms; mother's and father's age at the time of the participant's birth; mother's and father's last grade completed; number of participant's siblings; mother's and father's working status (this variable takes value 1 if the parent is employed and 0 otherwise); presence of the father (a binary indicator which takes value 1 if the participant's father is a current resident of the household). The descriptive statistics reported are the arithmetic mean, the standard deviation, and the skewness. Those are respectively measured by $\bar{Y} = (\sum_{i=1}^{N} Y_i)/N$, $\hat{\sigma} = \sqrt{[\sum_{i=1}^{N}(Y_i - \bar{Y})^2]/N}$ and $\hat{s} = \{[\sum_{i=1}^{N}(Y_i - \bar{Y})^3]/N\} \times 1/\hat{\sigma}^3$, where N denotes sample size and Y_i denotes the outcome for participant i.

for 2 years[8] for a less intensive 2.5–3 hours per day programme.[9] Details of the randomisation protocol are presented in online Appendix A. In this Section, we report: the background characteristics of the two populations (subsection 1.1); the interventions administered (subsection 1.2); and the data collections carried out (subsection 1.3).

1.1. *The Background Characteristics of the Two Populations*

While both ABC and PPP targeted disadvantaged populations, the background characteristics of the participants differed. We summarise them in Table 2 and Figures 1 and 2.[10]

---

[8] Wave 0 experienced only one year of treatment, starting at age 4.

[9] Note that, if we compute the hourly cost per child, the PPP intervention was more expensive than the ABC.

[10] See Hojman *et al.* (2013) for a comparison of the background characteristics of the ABC, PPP, CARE (Carolina Approach to Responsive Education), IHDP (the Infant Health and Development Program) and ETP (Early Training Project).
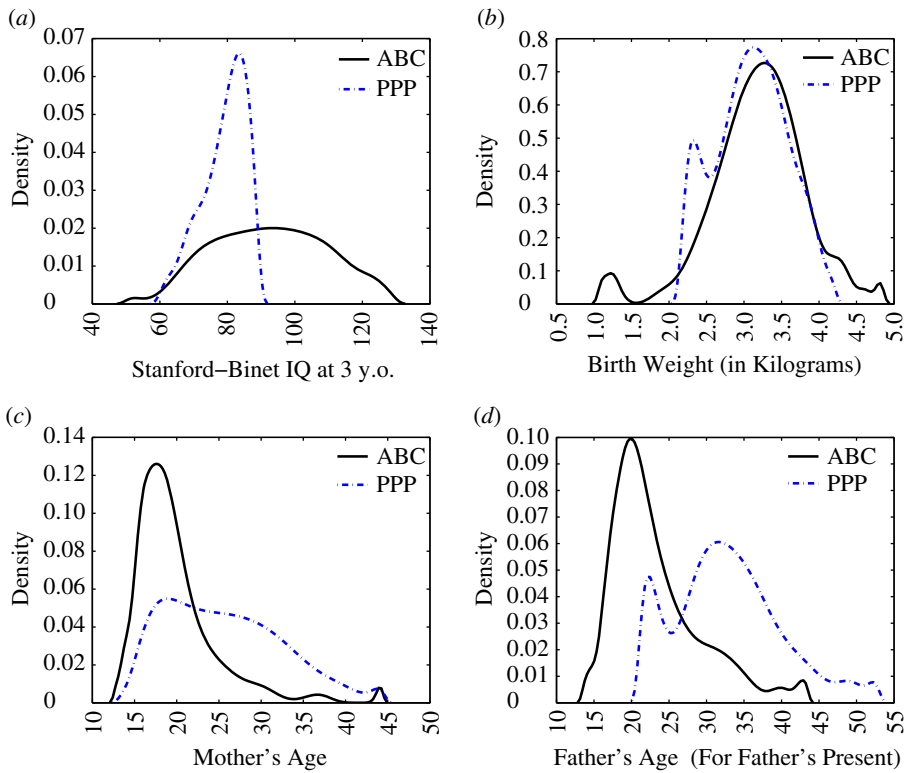
Fig. 1. *Comparison Between Pre-programme Variables of ABC and PPP. (a) IQ at 3 Years. (b) Birth Weight. (c) Mother's Age. (d) Father's Age*

Notes. These Figures present the density estimation of four pre-programme variables collected in both the Perry and Abecedarian interventions. (*a*) The Stanford-Binet IQ score at 3 years of age (we only use data for the control group for the ABC intervention, since it started at birth). (*b*) Weight at birth in kilograms. (*c* and *d*) Mother's and father's age at the time of the participant's birth. These estimates are based on a normal kernel function with optimal bandwidth for normal densities.

The first substantial difference that emerges is in the IQs of participants. While the average Stanford-Binet score at 3 years of age is 79 points in PPP, it is 14 points higher at the same age in the control group of ABC.[11] This difference is also visible in panel (*a*) of Figure 1, which shows that the region of common support is limited to the bottom half of the density of ABC. The partial overlap in the IQ distributions across the two interventions arises because PPP required an IQ smaller than 85 to be eligible to participate in the programme.

There is no significant difference in average health at birth (Table 2). However, a slightly higher proportion of ABC participants are born with low (<2,500 g) or high birthweights, (>4,000 g), as shown in panel (*b*) of Figure 1.[12]

_____

[11] We only use data from the control group for ABC, since it started at birth, hence by age 3 the treatment group would have already received three years of the programme.

[12] Parenthetically, the median birthweight for PPP was 3.14 kg, compared to a national population average of 3.29 kg in 1964. For ABC, the median birthweight was 3.24, compared to a national population average of 3.34.
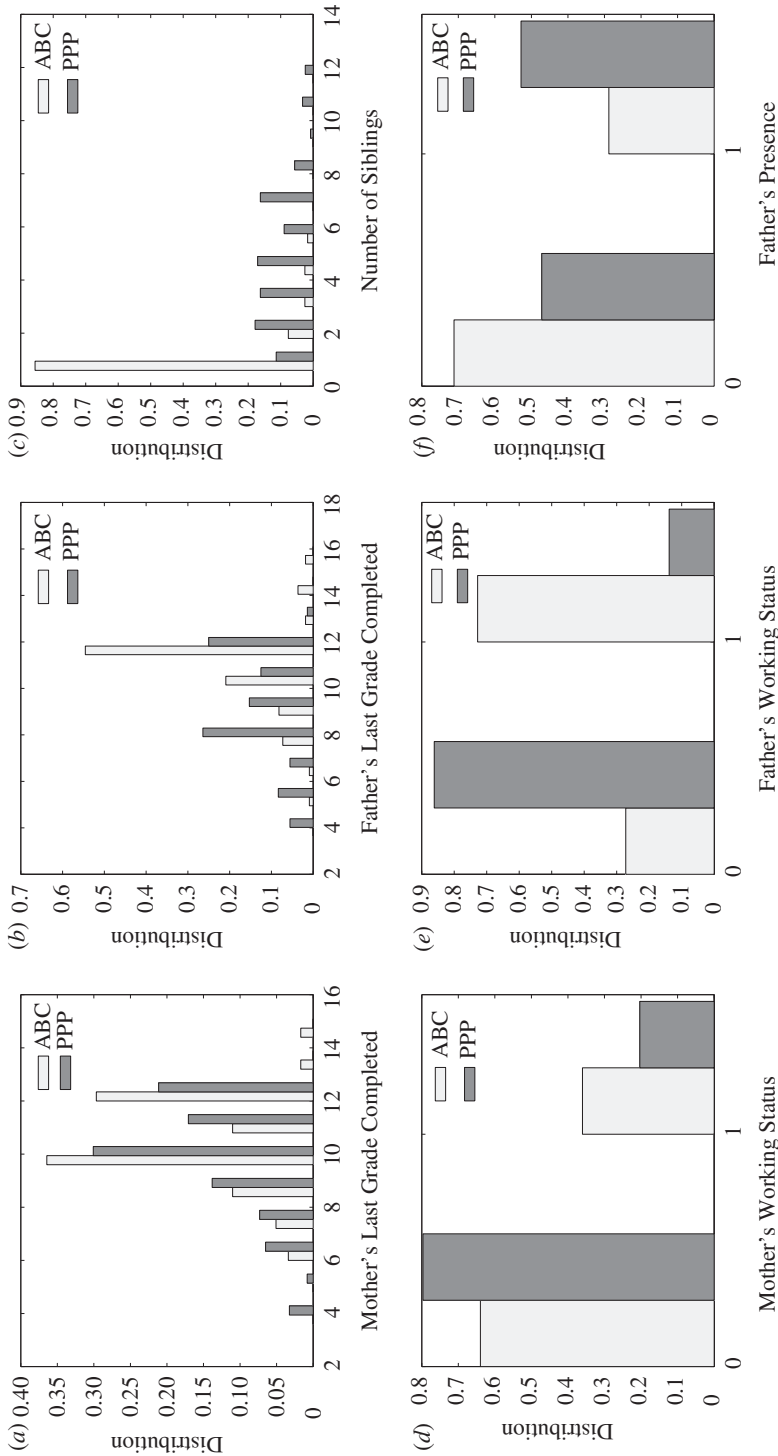
Fig. 2. *Comparison Between the Baseline Variables of ABC and PPP.* (*a*) *Mother's Education.* (*b*) *Father's Education.* (*c*) *Number of Siblings.* (*d*) *Mother's Working Status.* (*e*) *Father's Working Status.* (*f*) *Father's Presence*

Notes. These Figures present estimates of the empirical distributions of three categorical and three binary variables collected in both the Perry and Abecedarian interventions. All variables were collected at the onset of each intervention. (*a* and *b*) Mother's and father's last grade completed. (*c*) The number of participant's siblings. (*d* and *e*) Mother's and father's working status (a binary indicator, which takes value 1 if the parent is employed and 0 otherwise). (*f*) Presence of the father (a binary indicator, which takes value 1 if the participant's father is a current resident of the household).

Turning to the parental demographic characteristics, we see that the parents in PPP are older than those in ABC, with the age difference amounting to six years for the mothers and to nine years for the fathers (when fathers are present). The density reported in panel (*d*) of Figure 1 shows that the region of common support for paternal age only extends between the ages 20 and 45. In line with the older parental age, the participants of the PPP intervention also have, on average, a greater number of siblings (4, up to a maximum of 12, as shown in panel (*c*) of Figure 2), while ABC children are more likely to be first born. Additionally, ABC participants are more likely to be born to single mothers, with the father being present almost twice as often in PPP households than in ABC households (53% *versus* 29%, Table 2). Finally, the parents of ABC participants have higher socio-economic backgrounds, higher levels of education, and are more likely to be employed (as shown in Table 2 and panels (*a–b*) and (*d–e*) of Figure 2, respectively).

In sum, while more children in Perry are from two-parent homes,[13] many other socioeconomic characteristics are more favourable for ABC participants, especially for those with fathers present.[14] However, as shown in Table A1 of online Appendix A, controlling for these background characteristics does not substantially change estimated treatment effects.

1.2. *The Interventions*[15]

1.2.1. *Intervention strategies*
From 1962 to 1967, the Perry Preschool Project (PPP) recruited disadvantaged children three to four years of age on the basis of two selection criteria: 'cultural deprivation' and evidence of being 'educably mentally retarded' based on the Stanford-Binet Intelligence score (mean = 79). Mid-intervention and follow-up summaries describe a programme that operated for 2.5–3 hours each morning, 5 days per week over the course of a school year (Weikart, 1966, 1967, 1970). Except for the first treatment group that participated for one year only, four treatment groups experienced two years of the instructional programme. In addition to a monthly parent group meeting hosted by social work staff, PPP further incorporated a 60–90-minute weekly home visit, designed to offer individualised instruction as needed, establish teacher–primary caregiver relationship and involve the latter in their child's development (Weikart, 1964, 1967, 1970).

Weikart's descriptions of the programme change significantly throughout the course of the intervention, including its length and format for both children and parents, the intervention methodologies and learning activities, the role of the teacher, the role of the child as a learner, and even his/her understanding of cognitive development (Weikart, 1964, 1967, 1970). This reflects both experimentation within the programme and the changed framing of it as the literature on child development evolved while the

---

[13] See, e.g., Lopoo and DeLeire (2014) for a recent study on the long-term outcomes of children born to single mothers.

[14] See, e.g., Carneiro *et al.* (2013) and Dickson *et al.* (2016) on the intergenerational effects of maternal education on cognitive and behavioural outcomes for a sample of children from the US and the UK, respectively.

[15] See Heckman *et al.* (2014) for a more detailed description of the ABC and PPP interventions.

programme was being implemented. What remains consistent, however, are Weikart's stated primary goals of cognitive development with an emphasis on language development, the use of developmental theory in guiding curriculum framework and intervention methods, and a combined approach of a morning centre-based preschool programme and a weekly afternoon home visit by the child's teacher (Weikart, 1964, 1967, 1970). The learning programme implemented in PPP from 1962 to early 1965 included unit-based instruction, intentional adult-child interactive language, and a rich set of learning materials including Montessori tools, movement/dancing and an emphasis on caregiver-planned large and small-group activities. In the final year of PPP, the learning programme more closely resembled the later developed HighScope curriculum including 'Plan, Do, Review.' Individual instruction was not a specific feature of the Perry centre-based programme (Weikart *et al.*, 1978; Kuperman, 2014–2015; Kuperman and Cheng, 2015), whereas in ABC, it was a key component of the learning programme.

Ten years after PPP began, ABC recruited four cohorts of infants born between 1972 and 1977 at hospitals near Chapel Hill, NC, for an intensive early childhood intervention designed to prevent retardation for low-income, multi-risk populations. Treated children were transported by programme staff from their homes to the newly built Frank Porter Graham Center (FPGC) for up to 9 hours each day for 50 weeks/year (Ramey *et al.*, 1976).

What is now known as the 'Abecedarian Approach' emerged from a process of distinctive curriculum development. The number of teaching and learning activities expanded through formal testing and evaluation with each successive ABC cohort. The *LearningGames for the First Three Years* were designed by both Joseph Sparling and Isabelle Lewis as play-based adult-child activities for the expressed purposes of minimising infants' maladaptive, high-risk behaviour, and enhancing adult-infant interactions that support children's language, motor, and cognitive development and socio-emotional competence, including task orientation (Sparling and Lewis, 1979). Influenced by Piaget's theory of developmental stages, each individual activity included a stated learning objective thought to be developmentally appropriate, specification of needed materials, directions for teacher behaviour, and expected child outcome. In addition to tracking and dating activity assignments, these records enabled staff to prescribe a specific instructional programme every 2–3 weeks for each child by rotating learning activities and to note developmental progress or its lack thereof from programme entry to approximately age 36 months (Ramey *et al.*, 1976). During preschool, ABC supplemented the original *LearninGames* with a programme for three and four year olds, thought to be developmentally appropriate and developed together by staff and caregiving professionals with assistance from outside consultants. The *Abecedarian Approach to Social Competence* encouraged cognitive development, sociolinguistic and communicative competence, and reinforced socially adaptive behaviour involved in task orientation, peer–peer relations, adult–child relationships, and emotional self-awareness (Ramey *et al.*, 1976, 1982). Language intervention remained the critical ABC vehicle for supporting cognition and social skills (McGinness and Ramey, 1981).

The two randomised controlled trials share many features, including an emphasis on language and cognitive development in the intervention for disadvantaged children, the background influence of developmental theory on the design of the curriculum

but with plenty of room for individual adaptation, and general similarities such as the use of field trips as a learning tool, organisation of the learning environment during preschool years and ongoing professional development for staff. However, a comparison of reports drafted by the directors of PPP and ABC concurrently with their own interventions also reveals some key differences.

The programmes differed in the way they perceived their treated children and designed their intervention goals and conceptual approaches. Perry was motivated by a 'deficits' model, and the intervention was perceived as remediating cultural deprivation and mental retardation. PPP was launched in an era when cognitive psychology was in ascendance and shaped educational policy (Heckman and Kautz, 2014). This conceptual approach initially led Weikart to prioritise cognitive over socio-emotional learning in his reporting of the Perry programme, which he described as a key feature of a traditional middle class nursery school. However, in practice, PPP teachers modified this agenda and intentionally fostered the child's socio-emotional development, including self-regulation and the capacity of making judgments.[16] The middle-class teachers who initiated the programme did for the disadvantaged children in Perry what middle-class parents do for their own children (Heckman *et al.*, 2014) and effectively prevented the programme from being focused solely on cognition. Indeed, in reporting the first findings from the study, Weikart (1967, p. 5) wrote:

> Preschool must demonstrate ability to affect the general development of children in three areas. These are intellectual growth, academic achievement, and school behavior.

In contrast, ABC aspired to prevent retardation and thus recruited its sample from birth. By the time it was launched, the literature on child development had evolved beyond a sole focus on cognition. It benefited from an enhanced understanding of the work of child development psychologists Piaget and Vygotsky. For ABC, socio-emotional learning and cognitive development were intertwined and embedded within adult-child interactions and adult-mediated activities that incorporated an intentional use of language as a teaching tool to elicit children's emerging social competence and ability to reason.

ABC and PPP differ on a number of programme elements. In addition to the difference in intensity and duration, ABC and PPP involved the family in different ways. PPP incorporated weekly home visits designed to offer opportunities for individualised instruction as needed, to establish a relationship between the child's centre-based teacher and the mother/primary caregiver, and to involve her in the child's education. Weekly home visits lasted approximately 60–90 minutes (Weikart, 1964, 1970). In addition, PPP offered an opportunity for parents to participate in monthly group meetings hosted by social work staff (Weikart, 1964, 1967). In ABC, while there were no home visits, parents were invited to be actively involved in preschool classrooms and to participate in parent-teacher conferences to share updates about the treated child. Both treatment and control groups in ABC received family

---

[16] Source: meeting held at the University of Chicago on 26 July 2013 with the former Perry teachers Louise Derman-Sparks, Constance Kamii, and Evelyn Moore (Heckman *et al.*, 2014).

Table 3

The Health Care Components of ABC for the Treated Children

Component	Content
Well-child care	
Well-child Visits	Assessments were made at 2, 4, 6, 9, 12, 18, and 24 months, and yearly thereafter
	A health history and a social history were obtained and a complete physical examination was performed
Immunisations	Appropriate immunisations (diphteria, pertussis, tetanus, polio, measles, mumps, and rubella) as recommended by the American Academy of Pediatrics were given
Lab tests	A sickle cell preparation was obtained at 9 and 12 months from all black children
	A skin test for tuberculosis was given yearly, and a hematocrit was done at 9 and 18 months and yearly thereafter
	During symptom-free periods, the children were cultured for bacteria at two-week intervals, and for viruses and mycoplasmas every four weeks
Health Education	The parents were present at the child well-care visits. They were taught and counseled in the areas of: feeding and nutrition, weaning, cleanliness, skin care, child growth and development, behaviour, toilet training, accident prevention, and dental hygiene
	They were also encouraged to express their concerns and to discuss the problems that they were facing
Vision	Routine screening for vision was provided annually
Hearing	During symptom-free periods, the children underwent pneumatic otoscopy and tympanometry once a month
	If any tympanogram was abnormal, the child was seen for repeat otoscopy and tympanometry after two weeks
Ill-child care (for treated children only after the first year); for control children only in the first year	
Sick-care	Daily surveillance of all children in the centre for illness: the licensed practical nurse visited the classroom daily to review the health status of the children and receive reports from the parents
	Children who were unwell were promptly seen by a member of the health care staff
	A history was obtained and a physical examination done; appropriate laboratory tests and cultures were performed
	Children had their upper respiratory secretions cultured by throat swab and a saline nasal wash for isolation of viruses and bacteria
	A computer form was completed each time the child was examined, listing pertinent history, physical findings, diagnosis, and culture results
	Parents were informed of the nature of the child's ailment, and given prescriptions, but were responsible for buying medicines
	The family nurse practitioner made sure that half of the prescriptions were sent home and half to the centre
	The children were followed through the illness to recovery. They were allowed to attend the centre when ill except in case of chickenpox
	These referrals were made to specialists and hospitals but specialised visits and hospitalisations were not paid for

Notes. *In the first year of the study, the control children received medical care from the FPGC. After the first year, they were left on their own. Free medical care for the control children was offered by FPGC and two university-affiliated hospitals to control families, and reports suggest that this incentive was discontinued after the first year (Ramey *et al.*, 1976; Heckman *et al.*, 2016). *Sources.* Sanyal *et al.* (1980), Ramey *et al.* (1982), F. Campbell, personal communication.

support in the form of social work services on a request basis to obtain family planning and legal help.

Early reports of parental involvement in ABC suggest that each nursery and classroom staff member was assigned four treatment families to contact in order to establish individualised and open communication between parents and the centre.

Teachers were directed to plan an afternoon for each family to visit FPGC, observe their child and meet other teachers and medical staff. Families were provided photographs of their children engaging in programme activities that served to further strengthen the connection between home and school. Reports indicate that family holiday parties were well attended (Ramey *et al.*, 1977).

1.2.2. *The health care and nutritional components*

ABC differed significantly from PPP because it also included health care and nutritional components. Table 3 displays the treatments and exams included in the health care component of the ABC. Free paediatric care was provided to all the treated children who attended the Frank Porter Graham Center (FPGC) (Ramey *et al.*, 1982). The on-site medical staff had two paediatricians, a family nurse practitioner, and a licensed practical nurse.[17] 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 examination was performed and parents of the treated children were counselled about child health care, nutrition, growth, and development.[18] The *ill-child care* component included daily surveillance of all the treated children in the FPGC for illness.[19]

When ill, 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 (Ramey *et al.*, 1982). The cost of medicines was not covered; the parents were responsible for buying them, but the on-site staff ensured they were taken. If children were referred to a hospital, hospitalisation costs were not covered. Only the treated children received the free paediatric care. Free medical care for the control children had been initially offered at the FPGC and two university-affiliated hospitals. However, this incentive was discontinued after the first year (Ramey *et al.*, 1976; Heckman *et al.*, 2014), and the control families were left with the other sources of health care that were available at the time: community clinics for visits (mostly crowded and with rotating doctors), the local office of the health department for well-baby check-ups and immunisations, and the hospital ER for emergencies (F. Campbell. personal communication). Hence, an important difference was the continuity of early health care provided to the treated as compared to the control group.

In addition to primary paediatric care, the treated children also received breakfast, lunch and an afternoon snack at the centre. Food was prepared in 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.

On the other hand, PPP did not provide any form of health care or nutrition. The ABC curriculum incorporated weekly Cook to Learn activities focused on recipes with healthy ingredients (Harms, 1981; Veitch and Harms, 1981). ABC utilised meal times as educational experiences, complementing the rest of the learning programme for

---

[17] Active research on respiratory tract infections in children was also ongoing (Sanyal *et al.*, 1980; Roberts *et al.*, 1986).

[18] Apart from this health counselling, there was no parenting component in the ABC intervention.

[19] The licensed practical nurse visited the classroom daily to review the health status of the children and receive reports from the parents (Sanyal *et al.*, 1980).

promoting self-help, motor skill development, cognition and social behaviour, self-regulation, language development and specifically, for knowledge of healthy eating behaviours. Not only were meals and snacks at ABC prepared according to state nutritional guidelines, but a formal educational structure was in place for meals and eating at FPGC before ABC started collecting data. In contrast, in Perry, there was no formal activity supporting healthy nutrition or eating behaviour. The PPP curriculum also incorporated cooking activities, however, nutritional quality was low and healthy behaviour were not reinforced (Moore *et al.*, 1965; Kuperman and Cheng, 2015). The teachers provided healthy snacks in the form of crackers and juice. Perry used snack time to support language and social development (Moore *et al.*, 1965; Hall and Holmberg, 1974; Ramey *et al.*, 1977; Kuperman, 2014–2015).

1.2.3. *Child care experiences of the control group*

The PPP was launched before Head Start and the push for early childhood interventions. The control group was in home care or in neighbourhood home-care settings with neighbours, friends, and relatives. Things had changed ten years later. Children in the control group of the ABC intervention attended various types of out-of-home care before age 5, for periods of time varying between 0 and 60 months (Pungello *et al.*, 2010). This article does not account for control contamination, which is dealt with extensively in García *et al.* (2016). They find that doing so shows that estimated programme effects depend on gender and the choices of alternatives to treatment.

1.3. *The Data Collected*

Both the ABC and PPP interventions followed participants over time and collected a substantial amount of information about their lives. In PPP, data were collected annually from age 3 (the entry age) until the fourth grade (measures of intelligence and academic aptitude, achievement tests, assessments of socio-emotional development and information from school records starting at kindergarten through to secondary education). We know if participants went to post-secondary education but do not know teacher ratings or performance there, apart from information on graduation. Four follow-ups with interviews were conducted at ages 15, 19, 27, and 40. The retention rate has been high throughout: 91% of the original participants were re-interviewed at age 40.[20] Information on the health of the subjects was collected only at ages 27 and 40, all based on self-reports.[21]

Richer data were collected for the Abecedarian intervention than for the Perry intervention. Background characteristics were collected at the beginning of the programme and include parental attributes, family structure, socioeconomic status and the health of the mother and of the baby. Anthropometric measures were collected and a wide variety of assessments of the cognitive and socio-emotional development of the child and of both the family and the classroom environment were conducted, from

---

[20] Among those lost at follow-up, five controls and two treated were dead, and two controls and two treated had gone missing.
[21] An age 50 follow-up is virtually complete, including collection of an extensive set of biomarkers.

soon after the start of the preschool programme until the end of the school year. Four follow-ups with interviews were carried out at ages 12, 15, 21, and 30. A biomedical sweep was conducted when the participants were in their mid-30s, for the purpose of collecting indicators of cardiovascular and metabolic disease risk (Campbell *et al.*, 2014).

Many measures taken are not strictly comparable across programmes. Online Appendix C gives details on the exact survey questions asked and on the construction of the variables examined. Online Appendix C, Table C1 summarises their comparability. The lack of comparability poses several challenges for meta-analyses, often reported in the literature on child development.

We focus our empirical analysis on a set of outcomes of public health relevance according to the following categories:

   (*i*) physical health;
  (*ii*) health insurance and demand for health care and
 (*iii*) behavioural risk factors/lifestyles (diet and physical activity, smoking and drinking).

2. Methodology

Randomised Controlled Trials (RCTs) are sometimes touted as the 'gold standard' of programme evaluation (see, e.g., Ludwig *et al.*, 2011). A major benefit of randomisation is that, when properly executed, it solves the problem of selection bias for mean outcomes. RCTs can render treatment assignments statistically independent of unobserved characteristics that affect the choice of participation in a programme and that might also affect outcomes. As a consequence, a perfectly implemented randomised experiment enables analysts to evaluate mean treatment effects by using simple differences-in-means between treatment and control groups.[22]

In spite of their potential benefits, RCTs are often plagued by a range of statistical problems that require careful attention. They often have small sample sizes and many outcomes. They are often implemented through complex randomisation protocols that depart from an idealised random experiment (Heckman *et al.*, 2010). A compromised randomisation protocol is not an issue for the ABC experiment. It is a substantial issue in PPP. Heckman *et al.* (2010) discuss this point in detail. We apply their methods in this article and refer the reader to that article for details of the procedure and how application of it affects estimated treatment effects.

In addition to these challenges, the small sample sizes of the PPP and ABC interventions suggest that standard applications of large sample statistical inference procedures, which rely on the asymptotic behaviour of test statistics, may be inappropriate. The large number of outcomes poses the additional danger of arbitrarily selecting 'statistically significant' treatment effects for which high values of test statistics arise by chance. Indeed, for any particular treatment parameter, the probability of rejecting a true null hypothesis of no treatment effect, i.e., the type-I

---

[22] As noted by Heckman (1992), experiments only identify means and not distributions and so do not directly address many important policy questions without making assumptions beyond the validity of randomisation. See also Heckman *et al.* (1997).

error, grows exponentially as the number of tested outcomes increases. This phenomenon leads to cherry picking of significant results. Finally, non-random attrition can generate spurious inferences.

We account for all of these issues in our statistical analysis. We address the common criticism of analyses of the Perry and Abecedarian data regarding the validity of large sample inferential procedures. We examine if statistically significant results survive after accounting for small sample sizes, multiple hypothesis testing, non-random attrition, and departures from the intended randomisation protocols. It turns out that accounting for small sample size using permutation testing has little effect on the inferences obtained from large sample approaches for the ABC study. (See Campbell *et al.*, 2014 and García *et al.*, 2016). For Perry, the inferences are affected and treatment effects are more precisely determined using small sample permutation methods. Additionally, adjusting for multiple–hypothesis testing affects inference in both PPP and ABC. Our more careful statistical analysis of multiple hypotheses makes a substantial difference in the inference about the effectiveness of early childhood programmes that is often not fully appreciated in the advocacy-driven early childhood literature. Adjustments for attrition and compromised randomisation are implemented but not discussed in this article.[23]

The rest of this Section is organised as follows. We discuss our method for small sample inference in subsection 2.1. Subsection 2.2 explains how we address the problem of multiple-hypothesis testing. Subsection 2.3 describes our correction for attrition. Subsection 2.4 describes our method for decomposing statistically significant adult treatment effects into interpretable components associated with inputs that are enhanced by the treatment.[24] A more detailed description of our methodology is presented in online Appendix D.

2.1. *Small Sample Inference*

We address the potential problem of small sample size by using exact permutation tests which are tailored to the randomisation protocol implemented in each intervention, following the analysis of Heckman *et al.* (2010). Permutation tests are distribution free. They are valid in small samples since they do not rely on the asymptotic behaviour of the test statistics. Permutation-based inference gives accurate p-values even when the sampling distribution is skewed (Lehmann and Romano, 2005). It is often used when sample sizes are small and sample statistics are unlikely to be normal. In order to discuss our methodology more formally, we first introduce some notation.

Let $Y = (Y_i : i \in \mathcal{I})$ denote the vector of outcomes Y_i for participant i in sample $\mathcal{I}$. Let $D = (D_i : i \in \mathcal{I})$ be the binary vector of treatment assignments, $D_i = 1$ if participant i is assigned to the treatment group, and $D_i = 0$ otherwise. We use $X = (X_i : i \in \mathcal{I})$ for the set of covariates used in the randomisation protocol. Our method exploits the invariance of the joint distribution (Y, D) under permutations that swap the elements of the vector of treatment status D.

[23] We refer the reader to Campbell *et al.* (2014) for a discussion of attrition in the health wave of ABC and to Heckman *et al.* (2010) for compromised randomisation in PPP. Attrition is not an issue for PPP, nor is compromised randomisation an issue for ABC. See also García *et al.* (2016).

[24] This approach is called 'mediation analysis' in the applied statistics literature.

The invariance of the joint distribution (Y, D) stems from two statistical properties. First, randomised trials guarantee that D is exchangeable for the set of permutations that swap elements in D within the strata formed by the values taken by X – see Heckman *et al.* (2010) for a discussion. This exchangeability property comes from the fact that under the null hypothesis of no treatment effect, scrambling the treatment status of the participants sharing the same values of X does not change the underlying distribution of the vector of treatment assignments D. Second, the hypothesis of no treatment effect implies that the joint distribution of (Y, D) is invariant under these selected permutations of the vector D. As a consequence, a statistic based on assignments D and outcomes Y is distribution-invariant under reassignments based on the class of admissible permutations. Lehmann and Romano (2005) show that under the null hypothesis and conditional on the data, the exact distribution of such statistics is given by the collection of its values generated by all admissible permutations.

An important feature of the exchangeability property is that it relies on limited information on the randomisation protocol. It does not require a full specification of the distribution D nor of the assignment mechanism but only the knowledge of which variables are used as covariates X in implementing the randomisation protocol. Moreover, the exchangeability property remains valid under compromises of the randomisation protocol that are based on the information contained in observed variables X. In PPP, the assignment variables X used in the randomisation protocol are cohort, gender, child IQ, socio-economic status (SES, as measured by the cultural deprivation scale) and maternal employment status. Treatment assignment was randomised for each family on the basis of strata defined by these variables. In the ABC study, the assignment variables X are cohort, gender, maternal IQ, high risk index and number of siblings. The participants were matched in pairs on the basis of strata defined by the X variables.

2.2. *Correcting for Multiple Hypothesis Testing*

The presence of multiple outcomes in these studies creates the potential problem of cherry picking by analysts who report significant estimates. This generates a downward-biased inference with p-values smaller than the true ones. To see why, suppose that a single-hypothesis test statistic rejects a true null hypothesis at significance level α. Thus, the probability of rejecting a single null hypothesis out of K null hypotheses is $1 - (1 - \alpha)^K$ even if there are no significant treatment effects. As the number of outcomes K increases without bound, the likelihood of rejecting a null hypothesis becomes 1.

One approach that avoids these problems is to form arbitrarily equally weighted indices of outcomes (Muennig *et al.*, 2009, 2011). Doing so, however, produces estimates that are difficult to interpret. Instead, we analyse disaggregated outcomes. We correct for the possibility of arbitrarily selecting statistically significant treatment effects by conducting tests of multiple hypotheses. We adopt the familywise error rate (FWER) as the Type-I error. FWER is the probability of rejecting any true null hypothesis in a joint test of a set of hypotheses. The stepdown algorithm of Lehmann and Romano (2005) exhibits strong FWER control, that is to say that FWER is held at or below a specified level regardless of which individual hypotheses are true within a set of hypotheses.

The Lehmann and Romano (2005) stepdown method achieves better statistical properties than traditional Bonferroni and Holm methods by exploiting the statistical dependence of the distributions of test statistics. By accounting for the correlation among single hypothesis p-values, we are able to create less conservative multiple hypothesis tests. In addition, the stepdown method generates as many adjusted p-values as there are hypotheses, which facilitates examination of which sets of hypotheses are rejected. There is some arbitrariness in defining the blocks of hypotheses that are jointly tested in a multiple-hypothesis testing procedure. In an effort to avoid this arbitrariness, we define blocks of independent interest that are selected on interpretable *a priori* grounds (e.g. unhealthy lifestyles such as smoking and drinking). We also report the p-values obtained with the traditional Bonferroni method to compare it with the stepdown results.

2.3. *Correcting for Attrition*

Non-random attrition is also a potential source of bias in the estimation and inference of treatment effects. While the treatment status D and pre-programme variables X are observed for all participants, outcomes Y are not observed for some participants due to panel attrition. As a consequence, this may induce correlation between the treatment status and the unobserved characteristics that affect sample retention.

We address this issue by implementing an Inverse Probability Weighting (IPW) procedure that identifies features of the full outcome distribution by reweighting non-missing observations by their probability of being non-attrited, which is modelled as a function of observed covariates.[25] The IPW method relies on matching on observed variables to generate weights that are used to adjust the treatment effects for the probability of retention. These probability weights are estimated using a logit model, following the approach used in Campbell *et al.* (2014).[26] Small sample IPW inference is performed by recalculating these probabilities for each draw used to construct permutations. In PPP, attrition rates are below 10% at age 30 follow-up. For ABC, attrition rates are lower – roughly 6%. However, for the health component, there was substantial attrition, and we replicate the analysis of Campbell *et al.* (2014) to correct for it.

2.4. *Mediation Analysis*

We also conduct a dynamic mediation analysis to decompose the effects of the treatment into components associated with the experimentally induced enhancement of inputs at different ages in the production of health.[27] The observed outcome Y_i of participant $i \in \mathcal{I}$ is:

---

[25] For a recent review, see Huber (2012).

[26] We use a logit specification that models attrition as function of pre-programme variables for PPP and for ABC at ages 21 and 30, and also as function of variables collected in the previous sweep for ABC at mid 30s, given the severity of attrition in the biomedical sweep. We follow the procedure applied in Campbell *et al.* (2014), which is described in greater detail there.

[27] See Heckman and Pinto (2015) for a theoretical framework for dynamic mediation analysis. A fully comparable mediation analysis for both the ABC sample and the PPP sample is difficult. Different measurements have been collected in the two interventions (e.g., the Pupil Behaviour Inventory has only been used in PPP, while height and weight have only been measured in ABC), and the data collection was carried out at different ages. See online Appendix C.

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

where D_i denotes the treatment assignment for participant i ($D_i = 1$ if treatment occurs, $D_i = 0$ otherwise) and $[Y_i(0), Y_i(1)]$ are potential outcomes for participant i when treatment is fixed at control and treatment status, respectively. The goal of our mediation analysis is to decompose a average treatment effect $E[Y_i(1) - Y_i(0)]$ into components associated with previous health-related measures that were enhanced by the intervention.

We use $\mathbf{I}_i$ for the vector of health-related variables of participant i that are affected by the treatment. Those are the inputs in our health production function. $[\mathbf{I}_i(0), \mathbf{I}_i(1)]$ stands for the counterfactual inputs for participant i under control and treatment assignments. In the same fashion of (1), we have that $\mathbf{I}_i = D_i\mathbf{I}_i(1) + (1 - D_i)\mathbf{I}_i(0)$. The vector of inputs is indexed by the set $\mathcal{J}$ that is $\mathbf{I}_i(d) = [I_{i,j}(d) : j \in \mathcal{J}]$ where $d \in \{0,1\}$. Our analysis is based on the following linear health production function:

$$Y_i(d) = \kappa_d + \boldsymbol{\alpha}_d \mathbf{I}_i(d) + \boldsymbol{\beta}_d \mathbf{X}_i + \tilde{\epsilon}_i(d); \ d \in \{0,1\}, \tag{2}$$

where κ_d is an intercept; $\boldsymbol{\alpha}_d = (\alpha_{d,j} : j \in \mathcal{J})$ is the vector of parameters that multiply the inputs, $\mathbf{X}_i$ are pre-programme variables assumed not to be affected by the treatment; $\boldsymbol{\beta}_d$ is the vector of pre-programme parameters; and $\tilde{\epsilon}_i(d)$ is a zero-mean error term. Following Heckman *et al.* (2013), we partition the indexing set $\mathcal{J}$ of all inputs into $\mathcal{J}_M$ that stands for inputs that are measured and $\mathcal{J} \backslash \mathcal{J}_M$ for inputs that are not measured. We decompose the term $\boldsymbol{\alpha}_d \mathbf{I}_i(d)$ in (2) into components due to the measured inputs ($\mathcal{J}_M$) and unmeasured inputs ($\mathcal{J} \backslash \mathcal{J}_M$):

$$Y_i(d) = \kappa_d + \sum_{j \in \mathcal{J}_M} \alpha_{d,j} I_{i,j}(d) + \sum_{j \in \mathcal{J} \backslash \mathcal{J}_M} \alpha_{d,j} I_{i,j}(d) + \boldsymbol{\beta}_d \mathbf{X}_i + \tilde{\epsilon}_i(d), \tag{3}$$

$$= \tau_d + \sum_{j \in \mathcal{J}_M} \alpha_{d,j} I_{i,j}(d) + \boldsymbol{\beta}_d \mathbf{X}_i + \epsilon_i(d), \tag{4}$$

where $\tau_d = \kappa_d + \Sigma_{j \in \mathcal{J} \backslash \mathcal{J}_M} \alpha_{d,j} E[I_{i,j}(d)]$ and $\epsilon_i(d) = \tilde{\epsilon}_i(d) + \Sigma_{j \in \mathcal{J} \backslash \mathcal{J}_M} \alpha_{d,j} \{I_{i,j}(d) - E[I_{i,j}(d)]\}$. Clearly, unobserved inputs may also be changed by the experiment. Those changes may be correlated with the observed input changes.

Heckman *et al.* (2013) discuss these issues and propose and implement methods for addressing this potential endogeneity problem. Under assumptions specified in that paper, they test and do not reject the null hypothesis that increments in unobservables are independent of increments of observables. We apply their test for both interventions and we also fail to reject this null hypothesis.[28] We also test if $\alpha_{0,j} = \alpha_{1,j}; j \in \mathcal{J}_M$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$. We do not reject these hypotheses either. test and do not reject these hypotheses for the ABC data. They show that both ABC and Perry programme treatment effects are generated by changes of measured and proxied inputs along a stable production function. Thus, we can safely simplify the notation and write (4) as:

$$Y_i(d) = \tau_d + \sum_{j \in \mathcal{J}_M} \alpha_j I_{i,j}(d) + \boldsymbol{\beta} \mathbf{X}_i + \epsilon_i(d). \tag{5}$$

[28] The results are displayed in Tables F1 and F5 of online Appendix F.

Equation (1) can thus be rewritten as:

$$Y_i = \tau_0 + \tau D_i + \sum_{j \in \mathcal{J}_M} \alpha_j I_{i,j} + \boldsymbol{\beta} \mathbf{X}_i + \epsilon_i, \tag{6}$$

where $\tau = \tau_1 - \tau_0$ is the contribution of unmeasured inputs to mean treatment effects, $\epsilon_i = D_i \epsilon_i(1) + (1 - D_i) \epsilon_i(0)$ is a zero-mean error term, and $I_{i,j} = D_i I_{i,j}(1) + (1 - D_i) I_{i,j}(0); j \in \mathcal{J}_M$ are the measured inputs. García *et al.* (2016) effectively show that one cannot reject the null hypothesis $\tau = 0$ for all outcomes studied.

Our aim is to decompose treatment effects into components attributable to changes in measurable inputs. On the basis of (6), we can decompose the effects of the intervention on health as:

$$\mathrm{E}[Y_i(1) - Y_i(0)|\mathbf{X}] = (\tau_1 - \tau_0) + \sum_{j \in \mathcal{J}_M} \alpha_j \mathrm{E}\{[I_{i,j}(1) - I_{i,j}(0)] \mid \mathbf{X}\}, \tag{7}$$

where the second term on the right hand side is the contribution of measured inputs to the treatment effect.

We next expand this framework by partitioning the indexing set of measured inputs $\mathcal{J}_M$ into a set of child inputs $\mathcal{J}_M^C$ and a set of adult inputs $\mathcal{J}_M^A$, i.e. $\mathcal{J}_M = \mathcal{J}_C \cup \mathcal{J}_A$ and $\mathcal{J}_M^C \cap \mathcal{J}_M^A = \emptyset$. Child inputs of participant i are given by $I_{i,j}^C$ such that $j \in \mathcal{J}_M^C$. Adult inputs are denoted by $I_{i,j'}^A$ such that $j' \in \mathcal{J}_M^A$. In this notation, (6) can be rewritten as:

$$Y_i = \tau_0 + \tau D_i + \sum_{j \in \mathcal{J}_M^C} \alpha_j^C I_{i,j}^C + \sum_{j' \in \mathcal{J}_M^A} \alpha_{j'}^A I_{i,j'}^A + \boldsymbol{\beta} \mathbf{X}_i + \epsilon_i, \tag{8}$$

where α_j^C for $j \in \mathcal{J}_M^C$ and $\alpha_{j'}^A$ for $j' \in \mathcal{J}_M^A$ denote childhood and adulthood parameters respectively. We apply the same rationale of (6) to express adult inputs as the following linear production function:

$$I_{i,j'}^A = \mu_{0,j'} + \mu_{j'} D_i + \sum_{j \in \mathcal{J}_M^C} \gamma_{j,j'} I_{i,j}^C + \boldsymbol{\delta}_{j'} \mathbf{X}_i + \eta_{i,j'} \quad \text{for each } j' \in \mathcal{J}_M^A, \tag{9}$$

where $\mu_{j'} = \mu_{1,j'} - \mu_{0,j'}$, $\eta_{i,j'} = D_i \eta_{i,j'}(1) + (1 - D_i) \eta_{i,j'}(0)$, and $I_{i,j}^C = D_i I_{i,j}^C(1) + (1 - D_i) I_{j,j'}^C(0)$ such that $j \in \mathcal{J}_M^C$ and $j' \in \mathcal{J}_M^A$. According to (8) and (9), the effect of the intervention on health can be then decomposed as:

$$\mathrm{E}[Y_i(1) - Y_i(0)|\mathbf{X}] = \underbrace{(\tau_1 - \tau_0)}_{\text{treatment effect due to unmeasured inputs}} + \underbrace{\sum_{j \in \mathcal{J}_M^C} \alpha_j^C \mathrm{E}[I_{i,j}^C(1) - I_{i,j}^C(0)]}_{\text{treatment effect due to early inputs (direct effect)}} \tag{10}$$

$$+ \underbrace{\sum_{j' \in \mathcal{J}_M^A} \alpha_{j'}^A \mathrm{E}(\mu_{1,j'} - \mu_{0,j'})}_{\text{treatment effect due to late inputs}} + \underbrace{\sum_{j \in \mathcal{J}_M^C} \sum_{j' \in \mathcal{J}_M^A} \alpha_{j'}^A \gamma_{j,j'} \mathrm{E}[I_{i,j}^C(1) - I_{i,j}^C(0)]}_{\text{treatment effect due to early inputs through late inputs (indirect effect)}}.$$

We call this mediation analysis 'dynamic,' since we consider inputs at different ages, where the early inputs can have both direct effects on health outcomes and indirect effects operating through the late stage inputs. In our empirical application, we also

compare it with the results obtained from two 'static' mediation analyses, i.e. a first one based on the following health production function in which only childhood inputs are included:

$$Y_i = \tau_0 + \tau D_i + \sum_{j \in \mathcal{J}_M^C} \alpha_j^C I_{i,j}^C + \boldsymbol{\beta}^C \mathbf{X}_i + \epsilon_i, \qquad (11)$$

as done in Heckman *et al.* (2013) – and a second one based on the following health production function in which only adulthood inputs are included:

$$Y_i = \tau_0 + \tau D_i + \sum_{j' \in \mathcal{J}_M^A} \alpha_j^A I_{i,j'}^A + \boldsymbol{\beta}^A \mathbf{X}_i + \epsilon_i. \qquad (12)$$

as done in Muennig *et al.* (2009).[29] We show that accounting for both early and late inputs and for the dynamics in the process of formation of human capital makes a substantial difference. Excluding early inputs leads to overestimation of the role played by late ones in explaining the mechanisms through which the ABC and PPP interventions produced health impacts.

3. Empirical Results

This Section presents the results of our empirical analysis. We discuss the mean treatment effects in subsection 3.1, and the dynamic mediation analysis results in subsection 3.2.

Departing from the previous literature in child development,[30] we conduct our analysis by gender. The rationale for this choice is based on both biological and behavioural considerations. It is well established in both animal and human studies that males are more greatly affected by stress (Kudielka and Kirschbaum, 2005). Gender differences in growth, health, and mortality have been reported in the medical literature, starting *in utero* (Case and Paxson, 2005; Eriksson *et al.*, 2009). In addition, differences between men and women in the propensity to engage in unhealthy behaviour and in developing cardiovascular disease in the presence of common risk factors have been well documented. These behavioural differences have led some scholars to propose gender-based interventions (Courtenay *et al.*, 2002; Juutilainen *et al.*, 2004; Wardle *et al.*, 2004; Marino *et al.*, 2011). Despite the large body of interdisciplinary evidence, substantial gaps remain in our understanding of the sources of gender differences, especially in relation to the interconnections between social and biological processes (Rieker and Bird, 2005; Short *et al.*, 2013). The magnitude of, and explanations for, gender differences likely vary depending on the specific stage of the life cycle and the particular health measure considered (Matthews *et al.*, 1999). The existing literature does not provide a definitive answer as to why men and women have differential responses to environments. Nonetheless, our analysis confirms the importance of taking the gender dimension into account when analysing the impacts of interventions. For the outcomes we study, we find much stronger effects of these programmes for boys than for girls.

---

[29] However, they do not control for omitted inputs.
[30] Heckman *et al.* (2010), Campbell *et al.* (2014) and García *et al.* (2016) are exceptions.

Table 4

Inference Results: Perry Preschool Intervention

Variable (1)	No C (2)	No T (3)	Ctr. M. (4)	Treat. M. (5)	Diff. Ms. (6)	Asy. p-val. (7)	Naive p-val. (8)	Blk. p-val. (9)	Per. S.D. (10)	Blk. p-val. (11)	IPW p. S.D. (12)	Bonf. p-val. (13)
Lifestyles: diet and physical activity at 40 y.o. – males												
Physical activity	35	30	0.457	0.367	0.090	0.766	0.779	0.584	0.584	0.545	0.545	1.000
Healthy diet	35	29	0.229	0.379	0.151	*0.097*	0.113	*0.015*	*0.033*	*0.020*	*0.072*	*0.040*
Lifestyles: smoking at 27 y.o. – males												
Not a daily smoker	39	31	0.462	0.581	0.119	0.164	0.160	*0.092*	*0.092*	*0.089*	*0.089*	0.267
Not a heavy smoker	39	31	0.615	0.903	0.288	*0.003*	*0.002*	*0.004*	*0.005*	*0.004*	*0.005*	*0.012*
No. of cigarettes	39	31	8.744	4.291	4.453	*0.011*	*0.010*	*0.008*	*0.009*	*0.006*	*0.011*	*0.018*
Lifestyles: smoking at 40 y.o. – males												
Never smoker	36	30	0.444	0.600	0.156	0.107	0.109	*0.042*	*0.042*	*0.040*	*0.040*	0.160
Not a daily smoker	36	30	0.472	0.667	0.194	*0.058*	*0.063*	*0.014*	*0.042*	*0.010*	*0.035*	*0.040*
Not a heavy smoker	35	28	0.743	0.929	0.186	*0.027*	*0.027*	*0.013*	*0.023*	*0.011*	*0.021*	*0.044*
No. of cigarettes	35	28	6.543	3.714	2.829	*0.080*	*0.082*	*0.043*	*0.057*	*0.035*	*0.049*	0.140
Lifestyles: diet and physical activity at 40 y.o. – females												
Physical activity	22	24	0.045	0.375	0.330	*0.003*	*0.003*	*0.002*	*0.005*	*0.002*	*0.012*	*0.004*
Healthy diet	22	24	0.227	0.375	0.148	0.143	0.144	0.238	0.238	0.283	0.283	0.566
Lifestyles: drinking at 27 y.o. – females												
Not a frequent drinker	22	25	0.773	0.880	0.107	0.169	0.193	*0.004*	*0.019*	*0.015*	*0.028*	*0.030*
Alcohol consumption	22	25	3.818	3.200	0.618	0.314	0.320	*0.085*	*0.085*	*0.094*	*0.094*	0.188
Lifestyles: drinking at 40 y.o. – females												
Not a frequent drinker	22	23	0.909	0.870	0.040	0.659	0.663	0.600	0.600	0.698	0.698	1.000
Alcohol consumption	22	23	4.227	2.826	1.401	0.248	0.256	0.406	0.406	0.467	0.469	0.920

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

3.1. *Estimates and Inference*

Our main results are displayed in Tables 4 (for PPP) and 5 (for ABC). A complete set of results is displayed in online Appendix E. The general pattern reported there is that for most blocks of outcomes, there are fewer statistically significant health and/or health lifestyle outcomes for girls, although there are numerous statistically significant health and/or health lifestyle outcomes for boys. In each Table, we present simple differences in means between the treatment and control groups, and different p-values. These range from the traditional large-sample p-value for the one-sided single hypothesis that treatment had a positive effect to the constrained permutation p-value based on the Inverse Probability Weighting (IPW) t-statistic associated with the difference in means between the treatment groups, and its corresponding multiple hypothesis testing (stepdown) p-value. Column (11) of each Table reports p-values which account for all the statistical challenges addressed in this article. Finally, column (13) reports conservative Bonferroni p-values that adjust for multiple hypothesis testing for comparison. We find statistically significant health effects for males in both interventions. PPP promoted healthy behaviours. ABC improved biomarkers for cardiovascular and metabolic health.

We first examine the treatment effects for PPP. It is evident from Table 4 that there is a substantial and significant reduction in both smoking prevalence and intensity among the males in the treatment group, with effects already present at age 27 and sustained through to age 40. Muennig *et al.* (2009) also examine the impact of the intervention on smoking but were unable to detect any impact, since they pool male and female samples. A separate analysis by gender is justified on *a priori* grounds, on the basis of the interdisciplinary literature documenting differences in both determinants of smoking behaviour (Waldron, 1991; Hamilton *et al.*, 2006) and responses to interventions (Bjornson *et al.*, 1995; McKee *et al.*, 2005).

Males in the treatment group have a lower lifetime prevalence (0.40 *versus* 0.56 in the control group). They also have significantly lower rates of daily smoking than the controls, with the proportion of daily smokers declining from 0.42 to 0.33 between age 27 and the age 40 follow-up for the treated, while remaining stable at just above 50% for the controls, so that the difference between the treated and the controls doubles in a decade. This difference – 19 percentage points (p.p.) – amounts to the gap in smoking prevalence between men with an undergraduate degree (11.9%) and those with low education (29.5%) in the US in 2005 (Centers for Disease Control and Prevention (CDC), 2010). Additionally, while the smoking prevalence among the treated aligns with US-wide figures for men below the poverty level in 2005 (34.3%, CDC), the comparable figure for the controls is 20 p.p. higher. Another finding is that the biggest difference between the two groups emerges in relation to the intensity of smoking, which is only partly reduced between the ages 27 and 40 due to a decline in intensity among the controls: the average number of cigarettes smoked per day falls from 8.7 at age 27 to 6.5 at age 40.[31]

---

[31] Instead, the ABC intervention seems not to have affected smoking behaviour to the same extent. The only statistically significant impact is a delay in the age of onset of smoking by approximately three years, from 17 years old for the controls to 20 years old for the treated males (Table E3 in online Appendix E). However, this effect loses statistical significance once we account for multiple hypothesis testing. One plausible explanation for the lack of impact of the ABC on smoking could be the much lower smoking prevalence experienced by the two cohorts, who lived at two different phases of the smoking epidemics.

This is consistent with the decreasing trend in smoking behaviour which has been experienced in the US after the release of the Surgeon's General Report in 1964, as documented in the literature (Fiore *et al.*, 1989) – an opposite to that trend documented for obesity.

These estimates have substantial relevance for public health. Tobacco use is considered the leading preventable cause of early death in the US and about half of all long-term smokers are expected to die from a smoking-related illness (Centers for Disease Control and Prevention, 2010). In two major studies carried out for the US, one estimated that lifetime male smokers have a reduced life expectancy of 11 years as compared to non-smokers and that, although male smokers who quit at younger ages have greater gains in life expectancy (by 6.9–8.5 years for those who quit by age 35), even those who quit much later in life gain some benefits (Taylor *et al.*, 2002). Typical male smokers at age 24 have a reduced lifetime expectancy of up to 6 years as compared to non-smokers (Sloan *et al.*, 2004); this includes those who subsequently quit. Hence, we would expect this reduction in smoking to translate into improved health among the treated participants relative to the controls as they age.

Additionally, the treated males at age 40 in PPP are more likely than the controls to report having made dietary changes in the last 15 years for health reasons (38% *versus* 23%, see Table 4): most of these changes are related to reductions in the amount of fat and salt in the diet and in the intake of junk food. Hence, we would expect these changes in dietary habits to also translate into substantial health improvements – see, for example, Sacks *et al.* (2001) for the effects of diet on blood pressure.

Finally, the PPP intervention also substantially improved the healthy habits of the women who were randomised to the treatment group: by age 40, they are 33 percentage points more likely to engage in regular physical activity than those randomly assigned to the control group (Table 4); they also report to drink significantly less frequently in the age 27 sweep, although this difference is no longer significant by the time they reach age 40.

We next turn to analyse the impacts of the Abecedarian intervention, where anthropometric and cardiovascular biomarkers were collected during a physician's visit when the subjects were in their mid-30s. We first examine three outcomes not previously reported: weight, height and BMI. For each of them, the treated males perform better than the controls: they are, on average, 7 kg lighter, 5 cm taller, and with a BMI 4 points lower–just below the obesity threshold. However, the statistical significance of these differences vanishes once we account for multiple hypothesis testing. A comparison with nationwide figures for 2011–2 (Ogden *et al.*, 2014) reveals that ABC male participants are more likely to be both overweight and obese than 20–39 year old African-Americans: the prevalence of being overweight is 72% for the treated and 75% for the controls, against a nationwide figure of 63%, while that of obesity is 56% for the treated and 62.5% for the controls, against a US average of 35%.

Substantial differences are also found for all the reported outcomes related to blood pressure. Treated males have, on average, lower values of both systolic and diastolic blood pressure, and are less likely to fall into the Stage I hypertension category, according to the definition of the American Heart Association.[32] The magnitude of

---

[32] A more extensive set of health outcomes from the biomedical sweep is analysed in Campbell *et al.* (2014).

Table 5

Inference Results: Abecedarian Intervention

Variable (1)	No C (2)	No T (3)	Ctr. M. (4)	Treat. M. (5)	Diff. Ms. (6)	Asy. p-val. (7)	Naive p-val. (8)	Blk. p-val. (9)	Per. S.D. (10)	Blk. p-val. (11)	IPW p. S.D. (12)	Bonf. p-val. (13)
Physical health in the 30s – males												
Measured weight	9	18	100.6	93.80	6.850	0.242	0.274	0.320	0.320	0.154	0.303	0.462
Measured height	9	18	1.739	1.790	0.050	*0.044*	*0.061*	*0.083*	0.187	0.215	0.215	0.645
BMI	8	18	33.29	29.22	4.075	*0.076*	0.108	0.141	0.175	*0.093*	0.204	0.279
Overweight (BMI ≥ 25)	8	18	0.750	0.722	0.028	0.444	0.455	0.391	0.466	0.234	0.234	0.468
Obese (BMI ≥ 30)	8	18	0.625	0.556	0.069	0.376	0.378	0.448	0.448	0.227	0.335	0.454
Diastolic blood pressure	9	19	92.00	78.53	13.47	*0.017*	*0.046*	*0.075*	*0.075*	*0.025*	*0.025*	*0.050*
Systolic blood pressure	9	19	143.3	125.8	17.54	*0.022*	*0.059*	*0.057*	*0.085*	*0.019*	*0.031*	*0.038*
Hypertension I	9	19	0.444	0.105	0.339	*0.019*	*0.043*	*0.063*	*0.063*	*0.010*	*0.018*	*0.020*
Hypertension II	9	19	0.556	0.211	0.345	*0.033*	*0.049*	*0.061*	*0.095*	*0.037*	*0.037*	*0.074*
Health insurance at 30 y.o. – males												
Health care coverage	21	27	0.476	0.704	0.228	*0.057*	*0.062*	*0.080*	*0.080*	*0.040*	*0.040*	*0.080*
Employer-provided/bought	21	27	0.333	0.444	0.296	*0.021*	*0.018*	*0.034*	*0.048*	*0.035*	*0.055*	*0.070*
Demand for health care in the 30s – males												
Hospitalised	9	19	0.556	0.211	0.345	*0.033*	*0.039*	*0.042*	*0.042*	*0.100*	*0.100*	0.200
Scheduled treatment/exam	21	27	0.476	0.222	0.254	*0.033*	*0.040*	*0.026*	*0.051*	*0.043*	*0.080*	*0.086*
Lifestyles: diet and physical activity at 21 y.o. – females												
Physical activity	28	25	0.071	0.320	0.249	*0.010*	*0.013*	*0.009*	*0.009*	*0.004*	*0.004*	*0.008*
No. of fruit servings	28	25	0.286	0.800	0.514	*0.005*	*0.009*	*0.002*	*0.004*	*0.003*	*0.006*	*0.006*
Lifestyles: drinking at 30 y.o. – females												
Not a frequent drinker	28	25	0.857	0.880	0.023	0.405	0.414	0.493	0.586	0.547	0.547	1.000
Alcohol consumption	28	25	3.536	3.180	0.356	0.422	0.430	0.536	0.536	0.516	0.586	1.000
Age of onset < 17	28	25	0.571	0.280	0.291	*0.016*	*0.018*	*0.023*	*0.061*	*0.009*	*0.023*	*0.018*

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

these impacts is both statistically and medically significant. The estimated reductions in blood pressure are at least twice as large as those obtained from the most successful randomised control trials based on changing multiple behaviour risk factors (see Ebrahim *et al.*, 2011). The greatest reduction reported in meta-analyses of these trials is

−8.5 and −10 for diastolic and systolic blood pressure, respectively, found in Cakir and Pinar (2006), against the −13.5 and −17.5 reported in the ABC.

The superior health status of the males in the ABC treatment group is confirmed when we analyse the use of health care (Table 5). The treated are significantly less likely to have ever been hospitalised (21% *versus* 56% in the control group) and also to have had a scheduled treatment or examination in the past 12 months (22% *versus* 48% in the control group). They also enjoy higher health insurance coverage than those in the control group, especially if provided by the employer.

Finally, although they do not appear to be in better health than the controls, the females in ABC benefit from the treatment in terms of improved health habits. Interestingly, we find an improvement in healthy behaviour for PPP and ABC. The treated females in ABC and PPP are more likely to engage in physical activity, although the measures are not strictly comparable. ABC treatment women are more likely to eat fresh fruit than controls. They start drinking alcohol later. Perry treatment women are less likely to drink than controls.

For comparable outcomes, we find significant differences in the effects of the treatment between the ABC and PPP interventions for males. For females, reflecting the imprecision of the estimates for them within each programme, the there are no sharp differences in treatment effect across programmes.[33] Table E5 of online Appendix E reports tests of equality of comparably measured treatment effects by gender across the two studies. The relatively stronger effects found in ABC are consistent with (but do not definitively establish) the efficacy of the early health care given to participants in that programme.

3.1.1. *Methodological issues*
As noted in Section 2, both the ABC and PPP studies are plagued by several problems. We deal with these problems using methods tailored to the characteristics of each intervention. They make a substantial difference in inference, especially in the case of PPP. For many outcomes in that intervention, statistical significance is gained (e.g., for the probability of never being a smoker by age 40) or increases as we move from a large-sample analysis to a permutation-based analysis. In contrast is the effect of applying more refined methods to the Abecedarian sample. In that sample, no conclusion about any outcome is affected by the choice of or asymptotic inference. For a few outcomes the treatment effects do not survive the multiple hypothesis testing correction (height and BMI). This suggests that using large-sample methods does no harm in analysing the Abecedarian sample. However, accounting for multiple hypotheses makes a difference. This is evident when we compare the stepdown p-values with the more conservative ones obtained using the Bonferroni method (column (13)). The analysis of the Perry intervention requires more sophisticated methods to obtain reliable inference due, in part, to the greater complexity – and compromise – in its randomisation protocol.[34] As reported in Campbell *et al.* (2014) and García *et al.* (2016), adjusting for attrition from ABC makes a difference.

---

[33] Table C1 of online Appendix C summarises the comparability of the measures available in PPP and ABC.
[34] See Heckman *et al.* (2010), where this is discussed in depth.

3.2. *Mechanisms Producing the Treatment Effects*

We next investigate the mechanisms through which estimated treatment effects arise using the mediation analysis described in subsection 2.4. The literature suggests both direct and indirect mechanisms through which early childhood experiences might affect later health. Inadequate levels of stimulation and nutrition, the lack of a nurturing environment and of a secure attachment relationship, are all inputs which have been shown to play important roles in retarding development, by altering the stress response and metabolic systems, and leading to changes in brain architecture (Taylor, 2010).[35] On the one hand, child development might directly affect adult health, both because early health conditions are quite persistent throughout the lifecycle (as, in the case of obesity, see Millimet and Tchernis (2013)) and because early traits are determinants of lifestyles (Conti and Heckman, 2010).[36] On the other hand, child development might also affect adult health indirectly, by improving socio-economic determinants such as education, employment and income (Heckman *et al.*, 2010)–factors which might also have an independent effect on health, as documented in a large interdisciplinary literature (Smith, 1999; Deaton, 2001; Marmot, 2002; Lochner, 2011; Heckman *et al.*, 2016).

As described in subsection 2.4, we use a dynamic mediation analysis developed in Heckman and Pinto (2015) to examine the role of childhood and adult inputs in explaining treatment effects. We allow early childhood developmental traits to have both a direct impact on outcomes and an indirect one through educational attainment and adult socioeconomic status. We then compare the results obtained from a dynamic mediation analysis with those obtained by performing two static mediation analyses, where only child and adult inputs in turn are included in the health capital production function. The rationale for this exercise is to show the bias researchers might encounter by not considering a dynamic model of human capital formation.

Differences in both the timing and the content of the data collected do not allow us to use exactly the same childhood mediators. Nonetheless, we can analyse the role played by cognitive and behavioural traits for both interventions. Additionally, we include comparable mediators for educational attainment and adult socioeconomic status. In particular, for PPP, as early childhood mediators we consider (following Heckman *et al.*, 2013): IQ (the Stanford-Binet scale), externalising behaviour (reduced aggression) and academic motivation (constructed from selected items of the Pupil Behavior Inventory available in Perry). All are measured at ages 7–9. Heckman *et al.* (2013) show the powerful role of reduced externalising behaviour in producing a variety of beneficial outcomes in PPP. For adult inputs, we use high school graduation as a measure of educational attainment and unemployment (number of months unemployed in the last two years) and monthly income at age 27 as measures of socioeconomic status. Heckman *et al.* (2010) show that these measures were significantly affected by treatment. For the ABC, the childhood mediators represent the three different domains of development of the child: the Bayley Mental

---

[35] Given the lack of brain scans and measures of cortisol, we use proxies related to the underlying biological systems, such as cognitive and behavioural test scores.

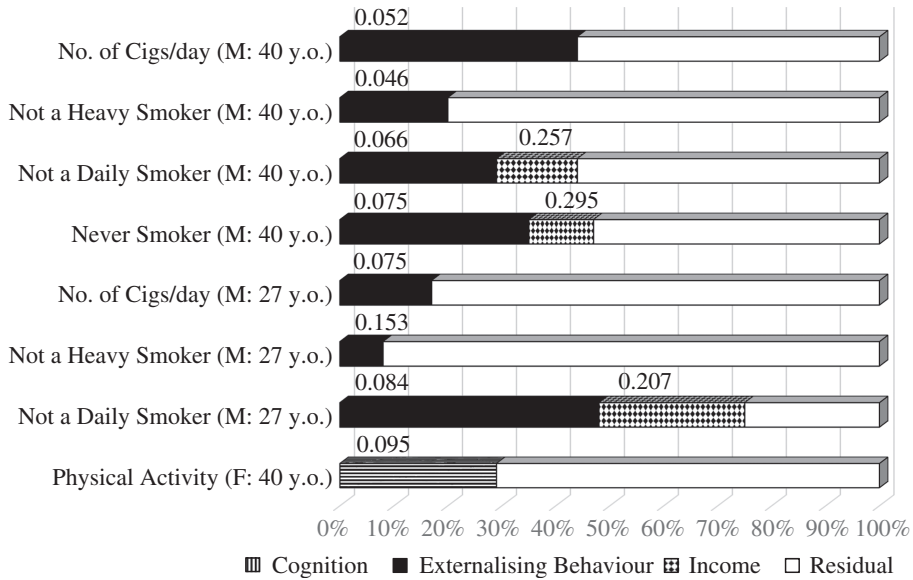[36] See also D'Onise *et al.* (2010) for a review of the literature on the health effects of ECIs.

Fig. 3. *PPP Dynamic Mediation Analysis of Treatment Effects on Male Outcomes*

Notes. This graph provides a simplified representation of the results of the dynamic mediation analysis of the statistically significant outcomes for the PPP intervention. Each bar represents the total treatment effect normalised to 100%. One-sided p-values that test if the share is statistically significantly different from zero are shown above each component of the decomposition. The mediators displayed are: cognition and externalising behaviour, as in Heckman *et al.* (2013) among the early childhood inputs; and income as in Heckman *et al.* (2010) among the adult inputs. The complete mediation results are reported in Tables F2 and F3 in online Appendix F. The definition of each outcome is reported in online Appendix C. The sample the outcomes refer to (M = males; F = females) and the age at which they have been measured (y.o. = years old) are shown in parentheses to the left of each bar, after the description of the variable of interest.

Development and the Stanford-Binet Scales for cognition, the Infant Behavior Record (IBR) Task Orientation Scale for behavioural development[37] and the Body Mass Index of the child for physical health. All are averages of standardised measurements taken at ages 1–2. All of these measures have been shown in previous work to be significantly affected by the treatment (Burchinal *et al.*, 1997; Campbell *et al.*, 2014). For adult inputs, we use college graduation as a measure of educational attainment, and employment status and earnings at age 30 as measures of socioeconomic status. García *et al.* (2016) document a significant impact of the intervention on these outcomes.

Complete results on mediation analyses are reported in online Appendix F. The main results for PPP are displayed in Figure 3. We decompose the treatment effects for the outcomes which survive the multiple hypothesis testing correction, and display the results for those for which we find that the mediators explain statistically significant shares of the treatment effects. Consistent with the evidence in Heckman *et al.* (2013),

---

[37] As seen in subsection 1.2, task orientation was one of the adaptive behaviour emphasised in the Abecedarian curriculum.

(a)

Early Child Development Mediators

0   10   20   30   40   50   60

□ Externalising-Static   ■ Externalising-Dynamic
⊞ Cognition-Static   ⊠ Cognition-Dynamic

(b)

Adult Socioeconomic Status Mediators

0   10   20   30   40   50   60   70   80   90
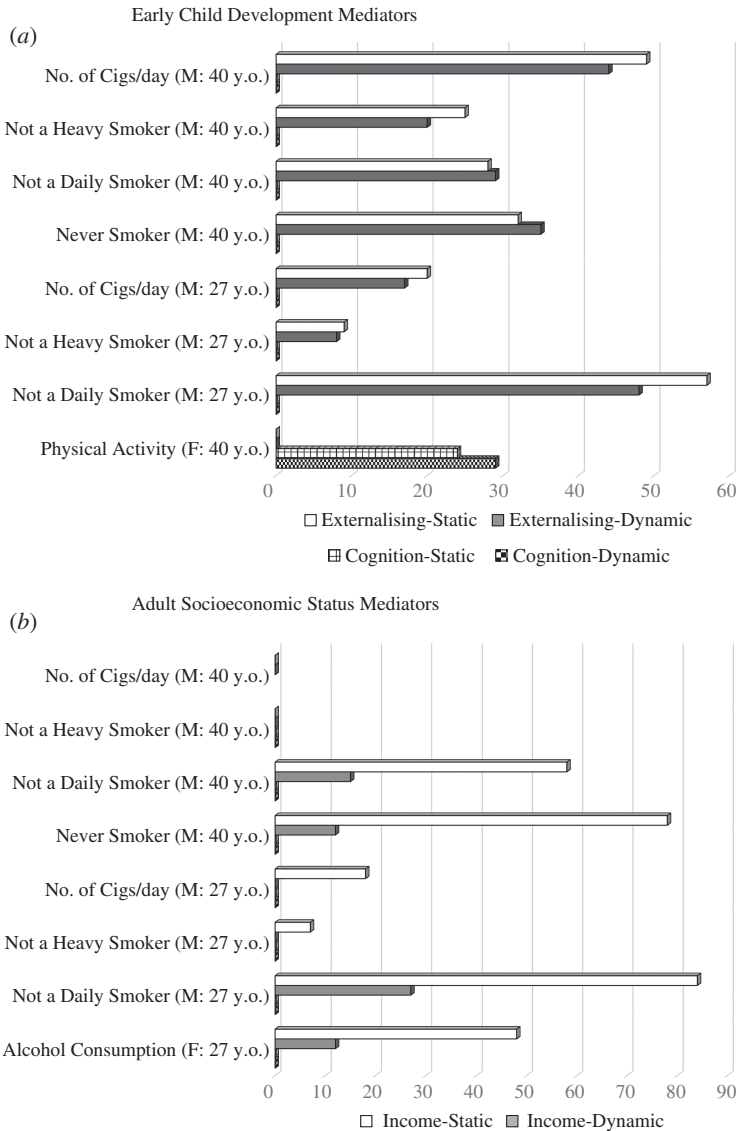
□ Income-Static   ■ Income-Dynamic

Fig. 4. *PPP: Static Versus Dynamic Mediation Analysis of Treatment Effects on Statistically Significant Male and Female Outcomes (Denoted 'M' and 'F')*

Notes. This Figure consists of two panels. Each panel compares the decompositions obtained from using the childhood (panel (*a*)) or adult (panel (*b*)) mediators alone (static) and the effects when both are used together (dynamic) for the results of the statistically significant outcomes for the PPP intervention. For each outcome and mediator, the lighter-coloured bars display the static mediation analysis results, while the darker-coloured bars display the dynamic mediation analysis results (as shown in Figure 3). Complete mediation results are reported in Tables F2, F3 and F4 in online Appendix F. The definition of each outcome is reported of online Appendix C. The sample the outcomes refer to (M = males; F = females) and the age at which they have been measured (y.o. = years old) are shown in parentheses to the left of each bar, after the description of the variable of interest.

we find that externalising behaviour is the main mediator of the effect of the intervention on smoking for males. Its mediating role survives even when later educational attainment or socioeconomic status is entered and its effects on adult behaviour are accounted for. It accounts for shares of the treatment effects ranging between 17% and 48%. It explains almost half of the treatment effect on the probability of not being a daily smoker at 27 years (p = 0.084), and 43% on the number of cigarettes smoked per day at age 40 (p = 0.052). The contribution of later life mediators is much smaller and fails to reach statistical significance. The role played by childhood behavioural traits is consistent with evidence reported in Conti and Heckman (2010), who show that improvements in child self-regulation are associated with a significantly lower probability of being a daily smoker at age 30, above and beyond its effect on education and the effect of boosts in education attainment on outcomes. This finding also contributes to the recent but flourishing literature on the importance of personality and preferences for healthy behaviour (Moffitt *et al.*, 2011; Conti and Hansman, 2013; Cobb-Clark *et al.*, 2014; Heckman *et al.*, 2016). For females, we find that enhancements in cognition are the main mediators of the effect of the intervention on physical activity. This is in line with evidence reported in Conti and Heckman (2010), who show that improvements in cognition are associated with better health for women but not for men.

Figure 4 compares the results from the dynamic mediation analysis with those obtained from the two static mediation analyses, including, respectively, those using only childhood mediators (panel (*a*)) and those using only adult mediators (panel (*b*)). They show that the decomposition components for the childhood mediators are unchanged in the static and dynamic mediation analysis (both in case of externalising behaviour for males, and of cognition for females). However, only including adult socioeconomic factors as inputs overestimates their importance. Indeed, while the shares explained by income are large and statistically significant in the static model, they are substantially reduced in magnitude and driven to insignificance when childhood factors are accounted for. Childhood factors have an impact on health behaviour above and beyond their effects on socioeconomic status in adulthood.

We now turn to the results for the Abecedarian Program, which are displayed in Figure 5.[38] We only report the results for men. Analysis of the female data from ABC shows few treatment effects. The mediators are clearly not comparable with those used in the analysis of Perry. We confirm the PPP results that early childhood traits mediate the health effects of treatment, above and beyond any induced improvement in adult socioeconomic status. The shares explained by task orientation and the body mass index of the child range between 17% and 28% for blood pressure, and between 20% and 31% for hypertension. Together, they explain half of the treatment effect. This is consistent with

---

[38] We only report mediation results for the males in case of ABC, since the dynamic mediation analysis and the static mediation analysis with late inputs cannot be performed for females, because the only statistically significant outcomes for this sample are those at age 21, and the late mediators are measured at age 30. The results for the static mediation analysis with early inputs for ABC are shown in the lower panel of Table F7 in online Appendix F. In contrast to the case for males, no mediator appears to explain a statistically significant share of the treatment effect. IQ explains 42% of the effect of the treatment on physical activity–a mechanism similar to the one uncovered for PPP–although it fails to achieve statistical significance at conventional levels.
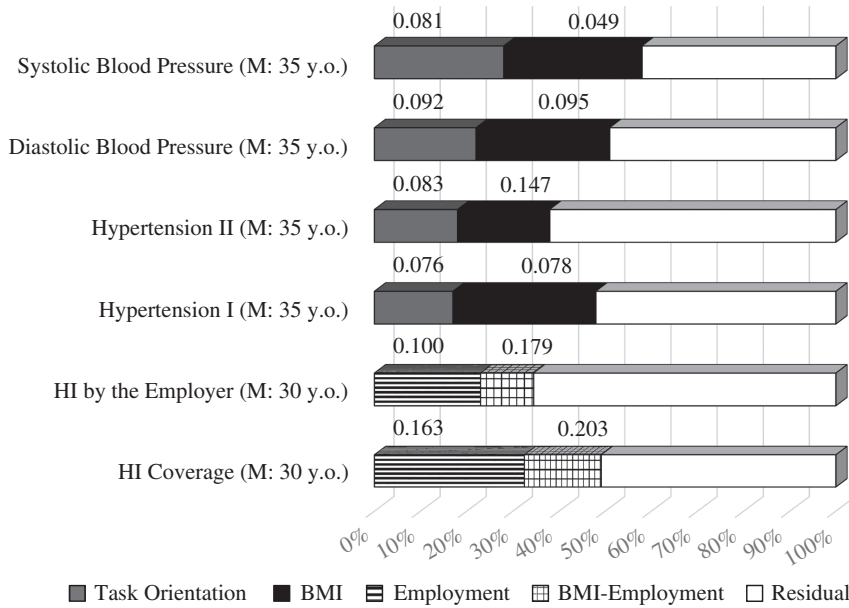
Fig. 5. *ABC Dynamic Mediation Analysis of Treatment Effects on Outcomes for Males*
Notes. This graph provides a simplified representation of the results of the dynamic mediation analysis of the statistically significant outcomes for the ABC intervention. Each bar represents the total treatment effect normalised to 100%. One-sided p-values that test if the share is statistically significantly different from zero are shown above each component of the decomposition. The mediators displayed are: task orientation as in Burchinal *et al.* (1997) and BMI as in Campbell *et al.* (2014) among the early childhood inputs; and employment as in García *et al.* (2016) among the adult inputs. The complete mediation results are reported in Table F6 in online Appendix F. The definition of each outcome is reported in online Appendix C. The sample refers to males. The Figure gives the ages at which the outcomes have been measured (y.o. = years old) as shown in parentheses to the left of each bar, after the description of the variable of interest (HI = Health Insurance). BMI-Employment is the share of the treatment effect which can be attributed to the indirect effect of experimentally induced changes in BMI affecting health insurance coverage through its impact on employment (see (10)).

existing evidence on both the role of child temperament[39] and that of physical development in the early years as key predictors for the risk of later obesity (Pulkki-Råback *et al.*, 2005; Conti and Heckman, 2010; Park *et al.*, 2012). Interventions to fight the obesity epidemic starting in the childhood years are increasingly being advocated, both to promote healthy dietary and exercise patterns (Deckelbaum and Williams, 2001) and to improve parental knowledge of proper nutrition and recognition of the child being overweight (Etelson *et al.*, 2003). As described in Section 1, the Abecedarian intervention included all of these components. Treated children enjoyed better nutrition and time for exercise while they were in the childcare centre. These features of the intervention could

---

[39] Specifically, task orientation has been associated with increased physical activity (Boyd *et al.*, 2002); this seems a plausible mechanism through which this trait might have by itself affected obesity, although data limitations prevent us from testing this formally.
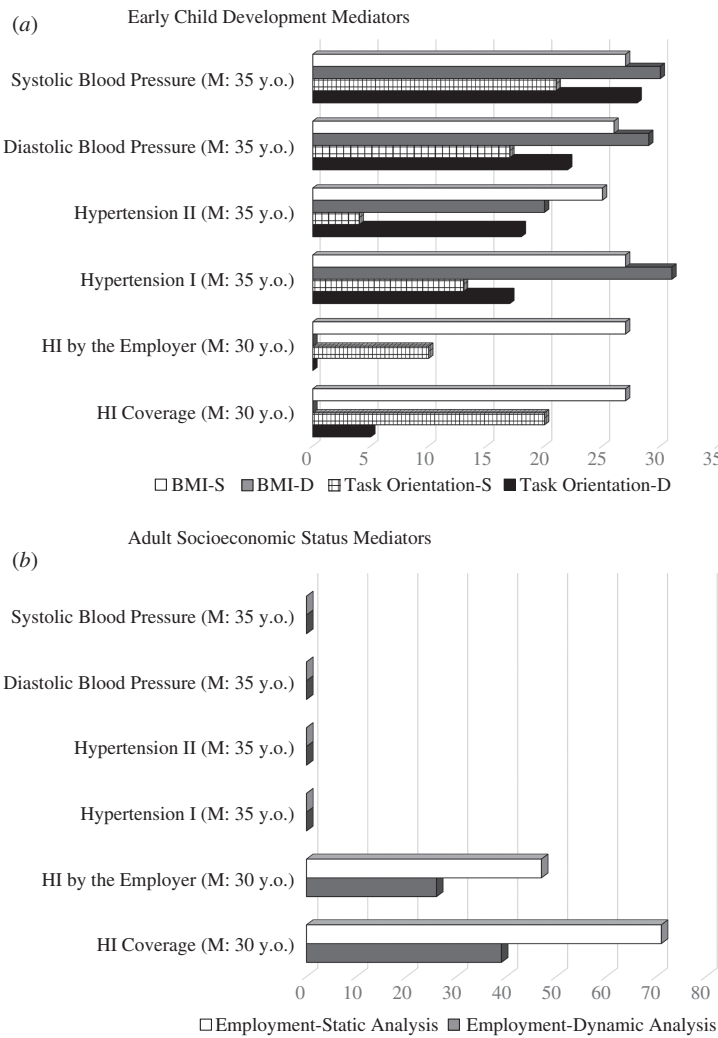
(a)

Early Child Development Mediators

(b)

Adult Socioeconomic Status Mediators

Fig. 6. *ABC: Static (S) Versus Dynamic (D) Mediation Analysis of Treatment Effects on Statistically Significant Outcomes for Males*

Notes. This Figure consists of two panels. Each panel compares the decompositions from the static and the dynamic mediation analyses of the statistically significant outcomes for the ABC intervention, respectively by comparing the results for the early child development mediators task orientation and BMI (panel (*a*)) and for the adult socioeconomic input employment (panel (*b*)). For each outcome and mediator, the lighter-coloured bars display the static mediation analysis results, while the darker-coloured bars display the dynamic mediation analysis results (as shown in Figure 5). The complete mediation results are reported in Tables F6 and F7 in online Appendix F. The definition of each outcome is reported in online Appendix C. The sample is for males and the age at which outcomes have been measured (y.o. = years old) is shown in parentheses to the left of each bar, after the description of the variable of interest (HI = Health Insurance). The term BMI-Employment in Figure 5 does not appear here since the static mediation analyses do not account for the indirect effects of early inputs affecting health outcomes through their impacts on late inputs. S = static mediation analysis; D = dynamic mediation analysis.

have had both a direct effect on their fat mass composition, and an indirect effect through a change in their preferences and behaviour. Additionally, participants were not allowed to eat outside meals and had to clean up the table once they were finished. This feature might have further contributed to the development of their self-regulatory skills. Finally, the counselling provided to the parents during the child well-care visits might have also improved the eating environment at home. Unfortunately, the data at our disposal do not allow us to disentangle the roles of these different channels.

On the other hand, the role of childhood traits in explaining the effect of the treatment on the greater availability of health insurance is much reduced when adult mediators are introduced. Consistent with the fact that the provision of health insurance is tied to a job, we find that employment status is the main mediator of the effect of treatment, with explained shares of 39% in case of health care coverage and 26% in case of employment-provided health insurance, respectively. Additionally, we also uncover evidence of a dynamic interaction between child and adult factors, with 20% and 13% of the effect of the treatment on the health insurance outcomes being mediated by the indirect effect of child BMI on adult employment.[40]

We also compare the dynamic mediation analysis results with those obtained from the two static mediation analyses (Figure 6). As for PPP, we find that the shares explained by the childhood mediators are comparable in the static and in the dynamic model for the physical health outcomes. However, for health insurance outcomes they are substantially reduced in the dynamic model (from 25% to 0% in the case of BMI) and driven to insignificance. In other words, the effects of early traits on health care coverage work entirely through their impact on adult socioeconomic status. Conversely, the small and insignificant shares of the treatment effects on the physical health outcomes explained by employment in the static model are reduced to zero in the dynamic model. Employment status still explains a significant share of the treatment effect on the health insurance outcomes in the dynamic model (panel (b) of Figure 6).[41] For females, income appears to explain half of the treatment effect on alcohol consumption in the static mediation model. This share is reduced to 12% and driven to statistical insignificance in the dynamic model (as shown in panel (b) of Figure 4).

In sum, our analysis shows the powerful role of enhanced early childhood traits in explaining the effect of treatment on adult health and health behaviour, above and beyond any effects of adult socioeconomic status. This is consistent with the framework of Cunha and Heckman (2009) and Cunha et al. (2010), as reviewed and extended by Heckman and Mosso (2014), in which early investments promote later life skills by boosting the base of capabilities that shape performance on a variety of tasks. Our analysis shows the importance of developing the child in her entirety, going beyond purely cognitive traits, within an integrated approach that also promotes behavioural and health development.

---

[40] As expected, higher child BMI at ages 1–2 is associated with a lower probability of being employed at age 30.

[41] It should also be noticed that in the case of the static mediation analysis we do not pass the specification test we apply following Heckman et al. (2013). See Table F5 in online Appendix F.

4. Conclusions

This article analyses the long-term impacts on healthy behaviour and health of two of the oldest and most cited US early childhood interventions: the High Scope Perry Preschool Program and the Carolina Abecedarian Project. We address some of the major limitations of previous work analysing these data. That research does not account for the variety of statistical challenges that arise in analysing these studies.[42] For many outcomes, these corrections make a substantial difference.[43] We also demonstrate differences across interventions in:

(*i*) characteristics of the treated populations;
(*ii*) the nature of the treatment; and
(*iii*) the data collected.

These differences create serious challenges for the meta-analyses routinely conducted in the literature on child development.

There are strong differences in the impact of the interventions by gender. Treatment effects are particularly strong for men. Both the Perry and the Abecedarian interventions have statistically significant effects on the health behaviour and health of their participants. The specific health outcomes affected vary by intervention. The Perry participants have significantly fewer behavioural risk factors (in particular smoking) by the time they reach age 40. The Abecedarian participants are in better physical health in their mid-30s. When strictly comparable outcomes are compared across the two programmes, for the same gender, estimated treatment effects are stronger for male ABC participants. This is broadly consistent with the emphasis on early health embedded in ABC. We find no statistically significant differences across the programmes for women.

In an attempt to shed light on the mechanisms through which these treatment effects emerge, we conduct dynamic mediation analyses following Heckman and Pinto (2015). Despite the lack of overlap in the measurements taken in the two interventions, the outcomes significantly affected by them, and the imperfect comparability of the mediators, we have uncovered an important role of enhanced early childhood traits as sources of adult treatment effects, above and beyond adult enhancements in socioeconomic status. This evidence is broadly consistent with the models of dynamic capability formation reviewed in Heckman and Mosso (2014). Skills developed early in life enhance the capabilities of persons to effectively perform a variety of lifetime tasks.

As the cohorts we have studied age and diseases start becoming more prevalent and manifest, it will be valuable to assess the contribution of behavioural risk factors and health insurance as additional mechanisms explaining the health effects of early childhood interventions. Plans for collecting and analysing later life data for both interventions are underway. Our results contribute to an emerging body of evidence that shows the potential of early life interventions for preventing disease and promoting health.

---

[42] Compromised randomisation is not an issue with the ABC programme. For Perry, where it is an issue, we apply the methods discussed in Heckman *et al.* (2010), where they make a difference in the reported estimates.

[43] Heckman *et al.* (2010) show that correcting for compromised randomisation in Perry as we do in this article makes a difference. Correcting for attrition from the medical wave of ABC has substantial impacts on estimates. (See Campbell *et al.*, 2014.)

University College London
University of Chicago
University of California — Los Angeles

Additional Supporting Information may be found in the online version of this article:

Appendix A. Eligibility Criteria and Randomisation Protocol in the ABC and PPP.
Appendix B. Moderators.
Appendix C. Description of the Variables Used in the Article.
Appendix D. Methodology.
Appendix E. Results.
Appendix F. Mediation Analyses.

References

Barnett, W.S. and Masse, L.N. (2007). 'Comparative benefit-cost analysis of the Abecedarian program and its policy implications', *Economics of Education Review*, vol. 26(1), pp. 113–25.

Bjornson, W., Rand, C., Connett, J.E., Lindgren, P., Nides, M., Pope, F., Buist, A.S., Hoppe-Ryan, C. and O'Hara, P. (1995). 'Gender differences in smoking cessation after 3 years in the Lung Health Study', *American Journal of Public Health*, vol. 85(2), pp. 223–30.

Black, M.M. and Hurley, K.M. (2014). 'Investment in early childhood development', *The Lancet*, vol. 384 (9950), pp. 1244–5.

Boyd, M.P., Weinmann, C. and Yin, Z. (2002). 'The relationship of physical self-perceptions and goal orientations to intrinsic motivation for exercise', *Journal of Sport Behavior*, vol. 25(1), pp. 1–18.

Burchinal, M.R., Campbell, F.A., Bryant, D.M., Wasik, B.H. and Ramey, C.T. (1997). 'Early intervention and mediating processes in cognitive performance of children of low-income African American families', *Child Development*, vol. 68(5), pp. 935–54.

Cakir, H. and Pinar, R. (2006). 'Randomised controlled trial on lifestyle modification in hypertensive patients', *Western Journal of Nursing Research*, vol. 28(2), pp. 190–209.

Camilli, G., Vargas, S., Ryan, S. and Barnett, W.S. (2010). 'Meta-analysis of the effects of early education interventions on cognitive and social development', *The Teachers College Record*, vol. 112(3), pp. 579–620.

Campbell, F.A., Conti, G., Heckman, J.J., Moon, S.H., Pinto, R., Pungello, E.P. and Pan, Y. (2014). 'Early childhood investments substantially boost adult health', *Science*, vol. 343(6178), pp. 1478–85.

Campbell, F.A., Wasik, B., Pungello, E., Burchinal, M., Barbarin, O., Kainz, K., Sparling, J. and Ramey, C. (2008). 'Young adult outcomes of the Abecedarian and CARE early childhood educational interventions', *Early Childhood Research Quarterly*, vol. 23(4), pp. 452–66.

Carneiro, P., Meghir, C. and Parey, M. (2013). 'Maternal education, home environments, and the development of children and adolescents', *Journal of the European Economic Association*, vol. 11(S1), pp. 123–60.

Case, A. and Paxson, C. (2005). 'Sex differences in morbidity and mortality', *Demography*, vol. 42(2), pp. 189–214.

Centers for Disease Control and Prevention (CDC). (2010). 'How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease', US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, GA.

Cobb-Clark, D.A., Kassenboehmer, S.C. and Schurer, S. (2014). 'Healthy habits: the connection between diet, exercise, and locus of control', *Journal of Economic Behavior and Organization*, vol. 98, pp. 1–28.

Cohen, S.B. and Yu, W. (2012). 'The concentration and persistence in the level of health expenditures over time: Estimates for the US population, 2008–2009', Agency for Healthcare Research and Quality, Rockville, MD.

Conti, G. and Hansman, C. (2013). 'Personality and the education–health gradient: a note on "understanding differences in health behaviors by education"', *Journal of Health Economics*, vol. 32(2), pp. 480–5.

Conti, G. and Heckman, J.J. (2010). 'Understanding the early origins of the education-health gradient: A framework that can also be applied to analyze gene-environment interactions', *Perspectives on Psychological Science*, vol. 5(5), pp. 585–605.

Conti, G. and Heckman, J.J. (2014). 'The economics of child well-being', in (A. Ben-Arieh, F. Casas, I. Frønes and J.E. Korbin, eds.), *Handbook of Child Well-Being*, pp. 363–401, Dordrecht: Springer.

Courtenay, W.H., McCreary, D.R. and Merighi, J.R. (2002). 'Gender and ethnic differences in health beliefs and behaviors', *Journal of Health Psychology*, vol. 7(3), pp. 219–31.

Cunha, F. and Heckman, J.J. (2009). 'The economics and psychology of inequality and human development', *Journal of the European Economic Association*, vol. 7(2–3), pp. 320–64.

Cunha, F., Heckman, J.J. and Schennach, S.M. (2010). 'Estimating the technology of cognitive and noncognitive skill formation', *Econometrica*, vol. 78(3), pp. 883–931.

Danese, A., Pariante, C.M., Caspi, A., Taylor, A. and Poulton, R. (2007). 'Childhood maltreatment predicts adult inflammation in a life-course study', *Proceedings of the National Academy of Sciences*, vol. 104(4), pp. 1319–24.

Deaton, A. (2001). 'Inequalities in income and inequalities in health', in (F. Welch, ed.), *The Causes and Consequences of Increasing Inequality*, pp. 285–313, Chicago, IL: University of Chicago Press.

Deckelbaum, R.J. and Williams, C.L. (2001). 'Childhood obesity: The health issue', *Obesity Research*, vol. 9 (S11), pp. 239S–243S.

Di Cesare, M., Khang, Y.H., Asaria, P., Blakely, T., Cowan, M.J., Farzadfar, F., Guerrero, R., Ikeda, N., Kyobutungi, C., Msyamboza, K.P., Oum, S., Lynch, J.W., Marmot, M.G. and Ezzati, M. (2013). 'Inequalities in non-communicable diseases and effective responses', *The Lancet*, vol. 381(9866), pp. 585–97.

Dickson, M., Gregg, P. and Robinson, H. (2016). 'Early, late or never? When does parental education impact child outcomes?', ECONOMIC JOURNAL, vol. 126(596), pp. F184–231.

D'Onise, K., Lynch, J.W., Sawyer, M.G. and McDermott, R.A. (2010). 'Can preschool improve child health outcomes? A systematic review', *Social Science and Medicine*, vol. 70(9), pp. 1423–40.

Ebrahim, S., Taylor, F., Ward, K., Beswick, A., Burke, M. and Smith, G.D. (2011). 'Multiple risk factor interventions for primary prevention of coronary heart disease', Cochrane Database of Systematic Reviews.

Emanuel, E.J. (2012). 'Where are the health care cost savings?', *Journal of the American Medical Association*, vol. 307(1), pp. 39–40.

Entringer, S., Buss, C. and Wadhwa, P.D. (2012). 'Prenatal stress, telomere biology, and fetal programming of health and disease risk', *Science Signaling*, vol. 5(248), pp. pt12.

Eriksson, J.G., Kajantie, E., Osmond, C., Thornburg, K. and Barker, D.J. (2009). 'Boys live dangerously in the womb', *American Journal of Human Biology*, vol. 22(3), pp. 330–5.

Etelson, D., Brand, D.A., Patrick, P.A. and Shirali, A. (2003). 'Childhood obesity: do parents recognize this health risk?', *Obesity Research*, vol. 11(11), pp. 1362–8.

Ezzati, M. and Riboli, E. (2012). 'Can noncommunicable diseases be prevented? Lessons from studies of populations and individuals', *Science*, vol. 337(6101), pp. 1482–7.

Fiore, M.C., Novotny, T.E., Pierce, J.P., Hatziandreu, E.J., Patel, K.M. and Davis, R.M.(1989). 'Trends in cigarette smoking in the United States: The changing influence of gender and race', *Journal of the American Medical Association*, vol. 261(1), pp. 49–55.

Ford, E.S., Bergmann, M.M., Boeing, H., Li, C. and Capewell, S. (2012). 'Healthy lifestyle behaviors and all-cause mortality among adults in the United States', *Preventive Medicine*, vol. 55(1), pp. 23–7.

Galobardes, B., Lynch, J.W. and Smith, G.D. (2008). 'Is the association between childhood socioeconomic circumstances and cause-specific mortality established? Update of a systematic review', *Journal of Epidemiology and Community Health*, vol. 62(5), pp. 387–90.

García, J., Heckman, J.J., Hojman, A., Leaf, D.E. and Prados, M.J. (2016). 'The life-cycle benefits of an influential early childhood program', Working Paper, University of Chicago, Department of Economics.

Garner, A.S., Shonkoff, J.P., Siegel, B.S., Dobbins, M.I., Earls, M.F., Garner, A.S., McGuinn, L., Pascoe, J., Wood, D.L. and Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, and Section on Developmental and Behavioral Pediatrics. (2012). 'Early childhood adversity, toxic stress, and the role of the pediatrician: translating developmental science into lifelong health', *Pediatrics*, vol. 129(1), pp. e224–31.

General Assembly of the United Nations (2011). 'Draft resolution submitted by the president of the general assembly: political declaration of the high-level meeting of the general assembly on the prevention and control of non-communicable diseases', United Nations.

Gluckman, P.D., Hanson, M.A., Buklijas, T., Low, F.M. and Beedle, A.S. (2009). 'Epigenetic mechanisms that underpin metabolic and cardiovascular diseases', *Nature Reviews: Endocrinology*, vol. 5(7), pp. 401–8.

Hall, J.S. and Holmberg, M.C. (1974). 'The effect of teacher behaviors and food serving arrangements on young children's eating in a day care center', *Child Care Quarterly*, vol. 3(2), pp. 97–108.

Hamilton, A.S., Lessov-Schlaggar, C.N., Cockburn, M.G., Unger, J.B., Cozen, W. and Mack, T.M. (2006). 'Gender differences in determinants of smoking initiation and persistence in California twins', *Cancer Epidemiology Biomarkers and Prevention*, vol. 15(6), pp. 1189–97.

Harms, T. (1981). *Learning from Cooking Experiences: A Teacher's Guide to Accompany Cook and Learn*, Menlo Park, CA: Addison-Wesley Publishing Company.

Heckman, J.J. (1992). 'Randomization and social policy evaluation', in (C.F. Manski and I. Garfinkel, eds.), *Evaluating Welfare and Training Programs*, pp. 201–30, Cambridge, MA: Harvard University Press.

Heckman, J.J., Humphries, J.E. and Veramendi, G. (2016). 'Returns to education: The causal effects of education on earnings, health and smoking', *Journal of Political Economy*, forthcoming.

Heckman, J.J. and Kautz, T. (2014). 'Achievement tests and the role of character in American life', in (J.J. Heckman, J.E. Humphries and T. Kautz, eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, pp. 3–56, Chicago, IL: University of Chicago Press.

Heckman, J.J., Kuperman, S. and Cheng, C. (2014). 'Comparing the programs of high quality early childhood interventions with long-term evidence as they were implemented', Working Manuscript, Department of Economics, University of Chicago.

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

Heckman, J.J. and Mosso, S. (2014). 'The economics of human development and social mobility', *Annual Review of Economics*, vol. 6(1), pp. 689–733.

Heckman, J.J. and Pinto, R. (2015). 'Econometric mediation analyses: identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs', *Econometric Reviews*, vol. 34(1–2), pp. 6–31.

Heckman, J.J., Pinto, R. and Savelyev, P.A. (2013). 'Understanding the mechanisms through which an influential early childhood program boosted adult outcomes', *American Economic Review*, vol. 103(6), pp. 2052–86.

Heckman, J.J., Smith, J.A. and Clements, N. (1997). 'Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts', *Review of Economic Studies*, vol. 64(4), pp. 487–535.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E. and Lumey, L.H. (2008). 'Persistent epigenetic differences associated with prenatal exposure to famine in humans', *Proceedings of the National Academy of Sciences*, vol. 105(44), pp. 17046–9.

Hertzman, C. (1999). 'The biological embedding of early experience and its effects on health in adulthood', *Annals of the New York Academy of Sciences (Socioeconomic Status and Health in Industrial Nations: Social, Psychological, and Biological Pathways)*, vol. 896, pp. 85–95.

Hojman, A., Schmerold, J. and Shea, J. (2013). 'A comparison of the populations of five experimental preschool programs', Unpublished manuscript, University of Chicago, Department of Economics.

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

Jamison, D.T., Summers, L.H., Alleyne, G., Arrow, K.J., Berkley, S., Binagwaho, A., Bustreo, F., Evans, D., Feachem, R.G.A., Frenk, J., Ghosh, G., Goldie, S.J., Guo, Y., Gupta, S., Horton, R., Kruk, M.E., Mahmoud, A., Mohohlo, L.K., Ncube, M., Pablos-Mendez, A., Reddy, K.S., Saxenian, H., Soucat, A., Ulltveit-Moe, K.H. and Yamey, G. (2013). 'Global health 2035: a world converging within a generation', *The Lancet*, vol. 382(9908), pp. 1898–955.

Juutilainen, A., Kortelainen, S., Lehto, S., Rönnemaa, T., Pyörälä, K. and Laakso, M. (2004). 'Gender difference in the impact of type 2 diabetes on coronary heart disease risk', *Diabetes Care*, vol. 27(12), pp. 2898–904.

Karoly, L.A., Kilburn, M.R. and Cannon, J.S. (2005). *Early Childhood Interventions: Proven Results, Future Promise*, Santa Monica, CA: RAND Corporation.

Knudsen, E.I., Heckman, J.J., Cameron, J.L. and Shonkoff, J.P. (2006). 'Economic, neurobiological, and behavioral perspectives on building America's future workforce', *Proceedings of the National Academy of Sciences*, vol. 103(27), pp. 10155–62.

Kontis, V., Mathers, C.D., Rehm, J., Stevens, G.A., Shield, K.D., Bonita, R., Riley, L.M., Poznyak, V., Beaglehole, R. and Ezzati, M. (2014). 'Contribution of six risk factors to achieving the 25×25 non-communicable disease mortality reduction target: a modelling study', *The Lancet*, vol. 384(9941), pp. 427–37.

Kudielka, B.M. and Kirschbaum, C. (2005). 'Sex differences in HPA axis responses to stress: a review', *Biological Psychology*, vol. 69(1), pp. 113–32.

Kuperman, S. (2014–2015). 'Interviews and reflections of Louise Derman-Sparks, Evelyn Moore, and Constance Kamii', University of Chicago's Center for the Economics of Human Development, Chicago, IL.

Kuperman, S. and Cheng, C. (2015). 'Comparison of randomized early childhood interventions as they were implemented: Perry, ABC, CARE, and IHDP', Working Paper, University of Chicago, Department of Economics.

Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd edn, New York: Springer-Verlag.

Lochner, L. (2011). 'Nonproduction benefits of education: crime, health, and good citizenship', in (E.A. Hanushek, S. Machin and L. Wößmann, eds.), *Handbook of the Economics of Education*, pp. 183–282, vol. 4, Amsterdam: Elsevier B. V.

Loewenstein, G.F., Asch, D.A. and Volpp, K.G. (2013). 'Behavioral economics holds potential to deliver better results for patients, insurers, and employers', *Health Affairs*, vol. 32(7), pp. 1244–50.

Loewenstein, G.F., Brennan, T.A. and Volpp, K.G. (2007). 'Asymmetric paternalism to improve health behaviors', *Journal of the American Medical Association*, vol. 298(20), pp. 2415–7.

Lopoo, L.M. and DeLeire, T. (2014). 'Family structure and the economic wellbeing of children in youth and adulthood', *Social Science Research*, vol. 43, pp. 30–44.

Ludwig, J., Kling, J.R. and Mullainathan, S. (2011). 'Mechanism experiments and policy evaluations', *Journal of Economic Perspectives*, vol. 25(3), pp. 17–38.

Marino, M., Masella, R., Bulzomi, P., Campesi, I., Malorni, W. and Franconi, F. (2011). 'Nutrition and human health from a sex–gender perspective', *Molecular Aspects of Medicine*, vol. 32(1), pp. 1–70.

Marmot, M. (2002). 'The influence of income on health: views of an epidemiologist', *Health Affairs*, vol. 21 (2), pp. 31–46.

Marmot, M.G. and Wilkinson, R.G. (2005). *Social Determinants of Health*, 2nd edn., Oxford: Oxford University Press.

Marteau, T.M., Hollands, G.J. and Fletcher, P.C. (2012). 'Changing human behavior to prevent disease: the importance of targeting automatic processes', *Science*, vol. 337(6101), pp. 1492–5.

Matthews, S., Manor, O. and Power, C. (1999). 'Social inequalities in health: are there gender differences?', *Social Science and Medicine*, vol. 48(1), pp. 49–60.

McGinness, G.D. and Ramey, C.T. (1981). 'Developing sociolinguistic competence in children', *Canadian Journal of Early Childhood Education*, vol. 1(2), pp. 22–43.

McKee, S.A., O'Malley, S.S., Salovey, P., Krishnan-Sarin, S. and Mazure, C.M. (2005). 'Perceived risks and benefits of smoking cessation: gender-specific predictors of motivation and treatment outcome', *Addictive Behaviors*, vol. 30(3), pp. 423–35.

Millimet, D.L. and Tchernis, R. (2013). 'The origins of early childhood anthropometric persistence', National Bureau of Economic Research.

Moffitt, T.E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R., Roberts, B.W., Ross, S., Sears, M.R., Thomson, W.M. and Caspi, A. (2011). 'A gradient of childhood self-control predicts health, wealth, and public safety', *Proceedings of the National Academy of Sciences*, vol. 108 (7), pp. 2693–8.

Mokdad, A.H., Marks, J.S., Stroup, D.F. and Gerberding, J.L. (2004). 'Actual causes of death in the United States, 2000', *Journal of the American Medical Association*, vol. 291(10), pp. 1238–45.

Moore, E., Derman-Sparks, L., Huth, C., Anderson, E., McClelland, D. and Hamilton, M. (1965). 'Lesson plans from the Perry Preschool Project, 1963–1964 and 1964–1965', Unpublished manuscript, High/Scope Educational Research Foundation.

Muennig, P., Robertson, D., Johnson, G., Campbell, F., Pungello, E.P. and Neidell, M. (2011). 'The effect of an early education program on adult health: the Carolina Abecedarian Project randomized controlled trial', *American Journal of Public Health*, vol. 101(3), pp. 512–6.

Muennig, P., Schweinhart, L., Montie, J. and Neidell, M. (2009). 'Effects of a prekindergarten educational intervention on adult health: 37-year follow-up results of a randomized controlled trial', *American Journal of Public Health*, vol. 99(8), pp. 1431–7.

Ogden, C.L., Carroll, M.D., Kit, B.K. and Flegal, K.M. (2014). 'Prevalence of childhood and adult obesity in the United States 2011–2012', *Journal of the American Medical Association*, vol. 311(8), pp. 806–14.

Park, M.H., Falconer, C., Viner, R.M. and Kinra, S. (2012). 'The impact of childhood obesity on morbidity and mortality in adulthood: a systematic review', *Obesity Reviews*, vol. 13(11), pp. 985–1000.

Pulkki-Råback, L., Elovainio, M., Kivimäki, M., Raitakari, O.T. and Keltikangas-Järvinen, L. (2005). 'Temperament in childhood predicts body mass in adulthood: the Cardiovascular Risk in Young Finns Study', *Health Psychology*, vol. 24(3), pp. 307.

Pungello, E.P., Kainz, K., Burchinal, M., Wasik, B.H., Sparling, J.J. and Ramey, C.T. (2010). 'Early educational intervention, early cumulative risk, and the early home environment as predictors of young adult outcomes within a high-risk sample', *Child Development*, vol. 81(1), pp. 410–26.

Ramey, C.T., Campbell, F.A., Burchinal, M., Skinner, M.L., Gardner, D.M. and Ramey, S.L. (2000). 'Persistent effects of early childhood education on high-risk children and their mothers', *Applied Developmental Science*, vol. 4(1), pp. 2–14.

Ramey, C.T., Collier, A.M., Sparling, J.J., Loda, F.A., Campbell, F.A., Ingram, D.A. and Finkelstein, N.W. (1976). 'The Carolina Abecedarian Project: A longitudinal and multidisciplinary approach to the prevention of developmental retardation', in (T. Tjossem, ed.), *Intervention Strategies for High-Risk Infants and Young Children*, pp. 629–55, Baltimore, MD: University Park Press.

Ramey, C.T., Holmberg, M.C., Sparling, J.H. and Collier, A.M. (1977). 'An introduction to the Carolina Abecedarian Project', in (B.M. Caldwell and D.J. Stedman, eds.), *Infant Education: A Guide for Helping Handicapped Children in the First Three Years*, pp. 101–121, New York: Walker and Company.

Ramey, C.T., McGinness, G.D., Cross, L., Collier, A.M. and Barrie-Blackley, S. (1982). 'The Abecedarian approach to social competence: Cognitive and linguistic intervention for disadvantaged preschoolers', in (K.M. Borman, ed.), *The Social Life of Children in a Changing Society*, pp. 145–74, Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Rieker, P.P. and Bird, C.E. (2005). 'Rethinking gender differences in health: why we need to integrate social and biological perspectives', *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 60 (Special Issue 2), pp. S40–S47.

Roberts, J.E., Sanyal, M.A., Burchinal, M.R., Collier, A.M., Ramey, C.T. and Henderson, F.W. (1986). 'Otitis media in early childhood and its relationship to later verbal and academic performance', *Pediatrics*, vol. 78(3), pp. 423–30.

Sacks, F.M., Svetkey, L.P., Vollmer, W.M., Appel, L.J., Bray, G.A., Harsha, D., Obarzanek, E., Conlin, P.R., Miller, E.R., Simons-Morton, D.G., Karanja, N., Lin, P.H., Aickin, M., Most-Windhauser, M.M., Moore, T.J., Proschan, M.A. and Cutler, J.A. (2001). 'Effects on blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (DASH) diet', *New England Journal of Medicine*, vol. 344(1), pp. 3–10.

Sanyal, M.A., Henderson, F.W., Stempel, E.C., Collier, A.M. and Denny, F.W. (1980). 'Effect of upper respiratory tract infection on Eustachian tube ventilatory function in the preschool child', *Journal of Pediatrics*, vol. 97(1), pp. 11–5.

Sherwin, R.S., Anderson, R.M., Buse, J.B., Chin, M.H., Eddy, D., Fradkin, J., Ganiats, T.G., Ginsberg, H.N., Kahn, R., Nwankwo, R., Rewers, M., Schlessinger, L., Stern, M., Vinicor, F. and Zinman, B. (2004). 'Prevention or delay of type 2 diabetes', *Diabetes Care*, vol. 27(Supplement 1), pp. S47–54.

Short, S.E., Yang, Y.C. and Jenkins, T.M. (2013). 'Sex, gender, genetics, and health', *American Journal of Public Health*, vol. 103(S1), pp. S93–S101.

Sloan, F.A., Ostermann, J., Conover, C., Taylor, D.H. and Picone, G. (2004). *The Price of Smoking*, Cambridge, MA: MIT Press.

Smith, J.P. (1999). 'Healthy bodies and thick wallets: the dual relation between health and economic status', *Journal of Economic Perspectives*, vol. 13(2), pp. 145–66.

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

Taylor, D.H., Hasselblad, V., Henley, S.J., Thun, M.J. and Sloan, F.A. (2002). 'Benefits of smoking cessation for longevity', *American Journal of Public Health*, vol. 92(6), pp. 990–6.

Taylor, S.E. (2010). 'Mechanisms linking early life stress to adult health outcomes', *Proceedings of the National Academy of Sciences*, vol. 107(19), pp. 8507–12.

Veitch, B. and Harms, T. (1981). *Cook and Learn: Pictorial Single Portion Recipes. A child's Cook Book*, revised and enlarged edn, Menlo Park, CA: Addison-Wesley Publishing Company.

Waldron, I. (1991). 'Patterns and causes of gender differences in smoking', *Social Science and Medicine*, vol. 32 (9), pp. 989–1005.

Wardle, J., Haase, A.M., Steptoe, A., Nillapun, M., Jonwutiwes, K. and Bellisie, F. (2004). 'Gender differences in food choice: the contribution of health beliefs and dieting', *Annals of Behavioral Medicine*, vol. 27(2), pp. 107–16.

Weikart, D.P. (1964). *Perry Preschool Project: Progress Report*, Ypsilanti, MI: Ypsilanti Public Schools.

Weikart, D.P. (1966). 'Preschool programs: preliminary findings', *Journal of Special Education*, vol. 1(2), pp. 163–81.

Weikart, D.P. (1967). 'Preliminary results from a longitudinal study of disadvantaged preschool children', ERIC No. ED 030 490. Presented at the 1967 convention of the Council for Exceptional Children, St. Louis, MO.

Weikart, D.P. (1970). *Longitudinal Results of the Ypsilanti Perry Preschool Project*, vol. 1 of *Monographs of the High/Scope Educational Research Foundation*, Ypsilanti, MI: High/Scope Educational Research Foundation.

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